Advances in Experimental Medicine and Biology 1163

Jian Zhang
Ruth Nussinov *Editors*

Protein Allostery in Drug Discovery

Springer

Advances in Experimental Medicine and Biology

Volume 1163

Advances in Experimental Medicine and Biology presents multidisciplinary and dynamic findings in the broad fields of experimental medicine and biology. The wide variety in topics it presents offers readers multiple perspectives on a variety of disciplines including neuroscience, microbiology, immunology, biochemistry, biomedical engineering and cancer research.

Advances in Experimental Medicine and Biology has been publishing exceptional works in the field for over 30 years and is indexed in Medline, Scopus, EMBASE, BIOSIS, Biological Abstracts, CSA, Biological Sciences and Living Resources (ASFA-1), and Biological Sciences. The series also provides scientists with up to date information on emerging topics and techniques. 2018 Impact Factor: 2.126

Content published in this book series is peer reviewed.

More information about this series at http://www.springer.com/series/5584

Jian Zhang • Ruth Nussinov
Editors

Protein Allostery in Drug Discovery

Springer

Editors
Jian Zhang
Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education
Shanghai Jiao Tong University, School of Medicine
Shanghai, China

Ruth Nussinov
Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research
National Cancer Institute
Frederick, MD, USA

Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine
Tel Aviv University
Tel Aviv, Israel

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-981-13-8718-0 ISBN 978-981-13-8719-7 (eBook)
https://doi.org/10.1007/978-981-13-8719-7

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

The book series focuses on protein allostery in drug discovery. Allosteric regulation, "the second secret of life," fine-tunes virtually most biological processes and controls physiological activities. Allostery can both cause human diseases and contribute to the development of new therapeutics. Allosteric drugs exhibit unparalleled advantages compared to conventional orthosteric drugs, rendering the development of allosteric modulators as an appealing strategy to improve selectivity and pharmacodynamic properties in drug leads. The series delineates the immense significance of protein allostery—as demonstrated by recent advances in the repertoires of the concept, its mechanistic mechanisms, and networks; characteristics of allosteric proteins, modulators, and sites; development of computational and experimental methods to predict allosteric sites; small-molecule allosteric modulators of protein kinases and G-protein-coupled receptors; engineering allostery; and the underlying role of allostery in precise medicine. A comprehensive understanding of protein allostery is expected to guide the rational design of allosteric drugs for the treatment of human diseases.

School of Medicine
Shanghai Jiao Tong University
Shanghai, China

Jian Zhang

Contents

Contributors

Hazem Abdelkarim Department of Biochemistry and Molecular Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA

Taaha Arshad State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Lindy Astl Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA

Avik Banerjee Department of Chemistry, University of Illinois at Chicago, Chicago, IL, USA

David D. Boehr Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

Souvik Chattopadhaya School of Applied Sciences, Republic Polytechnic, Singapore, Singapore

Xi Cheng State Key Laboratory of Drug Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China

Hai-Feng Chen State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
Shanghai Center for Bioinformation Technology, Shanghai, China

Anthony G. Coyne University Chemical Laboratory, University of Cambridge, Cambridge, UK

Rebecca N. D'Amico Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

Xavier Daura Catalan Institution for Research and Advanced Studies (ICREA) and Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain

Kang Ding iHuman Institute, ShanghaiTech University, Shanghai, China

Vadim Gaponenko Department of Biochemistry and Molecular Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA

Scott D. Gorman Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

Xinheng He Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

Ben Hitchinson Department of Biochemistry and Molecular Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA

Hyunbum Jang Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA

Hualiang Jiang State Key Laboratory of Drug Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China

Takayoshi Kinoshita Graduate School of Science, Osaka Prefecture University, Sakai, Osaka, Japan

Luhua Lai BNLMS, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, China
Center for Quantitative Biology, AAIS, Peking University, Beijing, China

Na Liu School of Pharmacy, Second Military Medical University, Shanghai, China

Shaoyong Lu Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

Duan Ni Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

Ruth Nussinov Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA
Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

Resmi C. Panicker School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, People's Republic of China

Mueed Ur Rahman State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Ashfaq Ur Rehman State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China
Department of Biochemistry, Abdul Wali Khan University Mardan, Mardan, Pakistan

Chunquan Sheng School of Pharmacy, Second Military Medical University, Shanghai, China

Kun Song Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

Rajavel Srinivasan School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, People's Republic of China

Jiahui Tong iHuman Institute, ShanghaiTech University, Shanghai, China
School of Life Science and Technology, ShanghaiTech University, Shanghai, China

Chung-Jung Tsai Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA

Amanda Tse Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA

Gennady M. Verkhivker Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA
Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA, USA
Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA

Dennis S. Winston Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

Yiran Wu iHuman Institute, ShanghaiTech University, Shanghai, China

Juan Xie BNLMS, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, China
Center for Quantitative Biology, AAIS, Peking University, Beijing, China

Jian Zhang Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

Weilin Zhang BNLMS, Peking-Tsinghua Center for Life Sciences at the College of Chemistry and Molecular Engineering, Peking University, Beijing, China
Center for Quantitative Biology, AAIS, Peking University, Beijing, China

Suwen Zhao iHuman Institute, ShanghaiTech University, Shanghai, China
School of Life Science and Technology, ShanghaiTech University, Shanghai, China

Qingtong Zhou iHuman Institute, ShanghaiTech University, Shanghai, China

Chapter 1
Allostery in Drug Development

Xi Cheng and Hualiang Jiang

Abstract Allosteric regulation is a ubiquitous strategy employed in nature to control cellular processes by regulating the affinities of biomolecules. Allosteric modulators are able to tune the protein/substrate affinity in a highly predictable way, suggesting that such modulators may represent safe drugs. Tremendous advances have been made in the development of allosteric modulators and the characterization of their therapeutic targets. Here, we briefly introduce several representative allosteric modulators of important drug targets, such as the G protein-coupled receptor family. We also review the state-of-the-art experimental and computational approaches used in allosteric drug development. The accumulated knowledge of allosteric regulation and the technical progress made in drug development will lead to an explosion of promising allosteric drugs in the near future.

Keywords Allosteric perturbation · Biased agonists · Conformational changes · Protein dynamics · Rational drug design

Abbreviations

ASD	AlloSteric Database
BRET	Bioluminescence resonance energy transfer
cAMP	Cyclic adenosine monophosphate
CCR5	C-C chemokine receptor type 5
DAG	Diacylglycerol
DNA	Deoxyribonucleic acid
EC50	Half maximum effect concentration
EF	Bacillus anthracis adenylyl cyclase toxin

X. Cheng · H. Jiang (✉)
State Key Laboratory of Drug Research, Drug Discovery and Design Center, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China
e-mail: hljiang@simm.ac.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_1

Emax	Maximum effect concentration
FDA	Food and Drug Administration
FRET	Fluorescence resonance energy transfer
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GPCR	G protein-coupled receptor
GRK	G protein-coupled receptor kinase
GTP	Guanosine-5′-triphosphate
HIV/AIDS	Human immunodeficiency virus infection and acquired immune deficiency syndrome
IP3	Inositol triphosphate
MD	Molecular dynamics
MSA	Multiple sequence alignment
NMA	Normal mode analysis
NMR	Nuclear magnetic resonance
RET	Resonance energy transfer
RNA	Ribonucleic acid
SCA	Statistical coupling analysis

1.1 Allostery in Nature

Billions of years of evolution have equipped organisms with complex molecular mechanism networks that enable their highly efficient responses to environmental changes. Most of these mechanisms rely on structure-switching biomolecules that undergo specific conformational changes upon stimulation. Allostery is one of the most important naturally occurring mechanisms in which a stimulus induces the conformational changes and functional modulation of the target biomolecule [102]. It arises from noncovalent events, such as binding of ions, small molecules, ribonucleic acid (RNA), deoxyribonucleic acid (DNA), or proteins [26, 27, 94]; from covalent events, such as phosphorylation, glycation, and nitration [4, 42, 106, 113]; and from light absorption [112]. Typically, the allosteric perturbation (such as the binding of an effector) occurs at a site distal from the active site of the target biomolecule [23] (Fig. 1.1). Allostery takes place in all dynamic proteins, in multimolecular assemblies, and in RNA and DNA polymers [71].

Allosteric regulation is a ubiquitous strategy employed in nature to control cellular processes by regulating the affinities of biomolecules. Many essential proteins take advantage of allosteric regulation to control their activities. One of the best examples is the negative allosteric modulation of hemoglobin by 2,3-bisphosphoglycerate [85]. Hemoglobin is the major oxygen-transport protein in the red blood cells of all vertebrates. The 2,3-bisphosphoglycerate binds to hemoglobin and decreases its affinity for oxygen, thus enhancing oxygen transport efficiency. On the other hand, the oxygen also positively allosterically modulates the

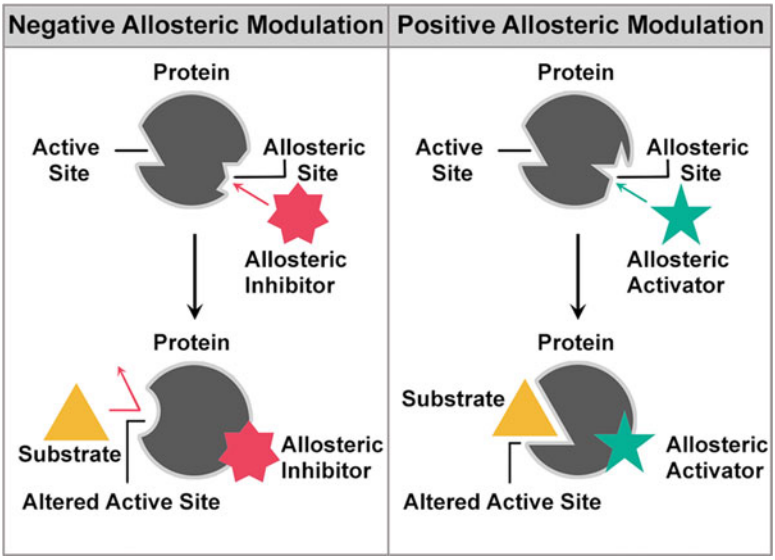


Fig. 1.1 Ligand-induced allosteric regulation. In negative allosteric modulation (left panel), an allosteric inhibitor (red) modifies the active site of a protein (gray) so that substrate (yellow) binding is reduced or prevented. In contrast, in positive allosteric modulation (right panel), an allosteric activator (green) modifies the active site of the protein (gray) so that the affinity for the substrate (yellow) increases

hemoglobin. The binding of oxygen to one subunit of hemoglobin induces a conformational change that enhances the oxygen affinity of the remaining active sites and allows hemoglobin to carry more oxygen. In addition to the participation in transportation of biomolecules, allosteric control is also strongly involved in numerous cell signaling and enzyme activities.

1.2 Allosteric Drugs

Allosteric modulators may represent novel and safe drugs. One main advantage of an allosteric modulator is its ability to tune the receptor/substrate affinity in a highly predictable manner. Instead of completely eliminating all activity, the allosteric modulator can selectively tune the receptor activity up (or down) only when its endogenous substrate is present. As a result, large doses of allosteric modulators can be administered with a low propensity toward target-based toxicity. Moreover, allosteric modulation does not affect the specificity of the biomolecule for its substrate because the allosteric site is distal from the receptor/substrate interface. Additionally, because they have not faced the same evolutionary pressure as endogenous substrate binding sites (orthosteric sites), the allosteric binding sites are less conserved. By targeting these diverse allosteric sites, specific drugs with greater

selectivity across homologous receptors may be obtained, which avoid off-target side effects. Allosteric modulators can efficiently govern the conversion of a specific signal input into a desired output, which determines its significance in the therapeutic developments for disease treatment.

The US Food and Drug Administration (FDA) has approved several efficient allosteric modulators as marketed drugs (Fig. 1.2 and 1.3a). Those allosteric drugs

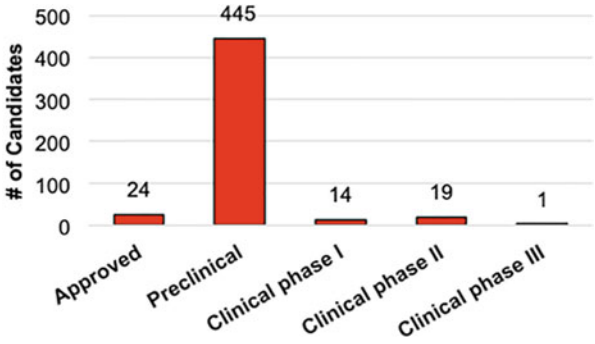


Fig. 1.2 Development status of allosteric drugs. (Data were collected from the Thomson Reuters Integrity database http://integrity.thomson-pharam.com)

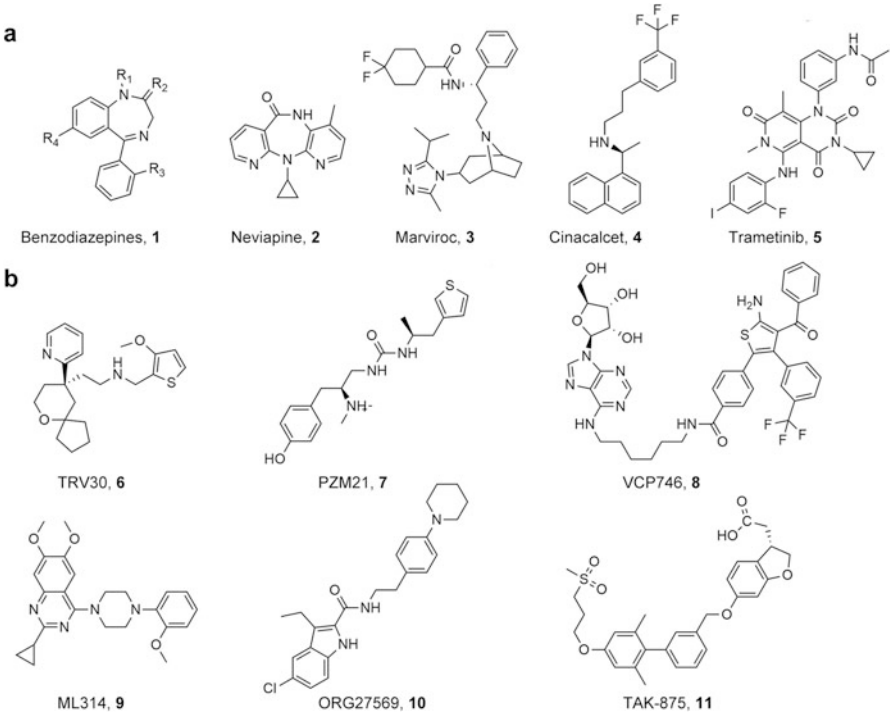


Fig. 1.3 Chemical structures of allosteric modulators described in this chapter. (**a**) FDA-approved allosteric drugs. "R" labels denote common locations of side chains, which give different benzodiazepines their unique properties. (**b**) Allosteric compounds undergoing clinical trials

interact with key proteins, including but not limited to ion channels, enzymes, and G protein-coupled receptors (GPCRs). One class of the earliest allosteric drugs with a major impact on society is the benzodiazepines (compound **1**). They are allosteric modulators of the γ-aminobutyric acid receptor, which is a ligand-gated ion channel regulating brain functions. Benzodiazepines effectively treat anxiety and sleep disorders without inducing the potentially lethal effects of direct-acting agonists [84]. Nevirapine (compound **2**) is another famous allosteric drug. This drug prevents human immunodeficiency virus infection and acquired immune deficiency (HIV/AIDS) via inhibiting the reverse transcriptase enzyme, which transcribes viral RNA into DNA. Unlike nucleoside reverse transcriptase inhibitors binding at the polymerase active site, nevirapine binds allosterically at a hydrophobic pocket 10 Å away from the active site [104]. In triple-combination therapy, nevirapine has been shown to effectively suppress viral load when used as initial antiretroviral therapy [86]. Maraviroc (compound **3**) is another antiretroviral drug used in the treatment of HIV infection, which works through a negative modulation mechanism inhibiting a GPCR, i.e., C-C chemokine receptor type 5 (CCR5). As the co-receptor of HIV envelopes glycoprotein gp120, CCR5 is necessary for the entrance of the virus into the host cell for most HIV strains. Maraviroc can allosterically induce conformational changes of CCR5 that block the association of gp120 with the receptor and prevent the entry process of the virus into the host cell [72]. This allosteric drug has been proven for use in the treatment of HIV patients. Cinacalcet (compound **4**) is another GPCR drug targeting calcium-sensing receptor. In the treatment of secondary hyperparathyroidism, it allosterically increases the sensitivity of calcium-sensing receptor on parathyroid cells to reduce parathyroid hormone levels and decrease serum calcium levels [118]. Trametinib (compound **5**) is a selective allosteric inhibitor of mitogen-activated protein kinases. This allosteric drug is the first oral chemotherapy regimen approved in 2013 [37]. In addition to these marketed drugs, a number of compounds working through allosteric mechanisms are under active development at the time of writing (Figs. 1.2 and 1.3b).

Focusing on the most fruitful drug targets, i.e., GPCRs, tremendous advances have been made in allosteric modulator development and therapeutic profile characterization. Previous studies have demonstrated that GPCR signaling is primarily controlled via interacting with three protein families: G proteins, G protein-coupled receptor kinases (GRK), and β-arrestins (Fig. 1.4). Upon stimulation, a GPCR activates a corresponding heterotrimeric G protein and catalyzes the exchange of guanosine-5′-triphosphate (GTP) for guanosine diphosphate (GDP). Then, the heterotrimeric G protein dissociates into Gα and Gβγ subunits to promote the formation of second-messenger effectors. After G protein activation, the receptor is phosphorylated by GRKs, which increases β-arrestin binding to the receptor [9]. β-arrestins mediate desensitization, tracking, and internalization of GPCRs [40, 76]. In addition to acting as negative regulators, β-arrestins also couple to many signaling mediators, including mitogen-activated protein kinases, serine/threonine kinases, tyrosine kinases, nuclear factor κB, and phosphoinositide 3-kinase [8, 35, 77]. These signaling pathways are different from classical G protein signaling pathways and regulate distinct aspects of receptor activity. GPCRs present multiple

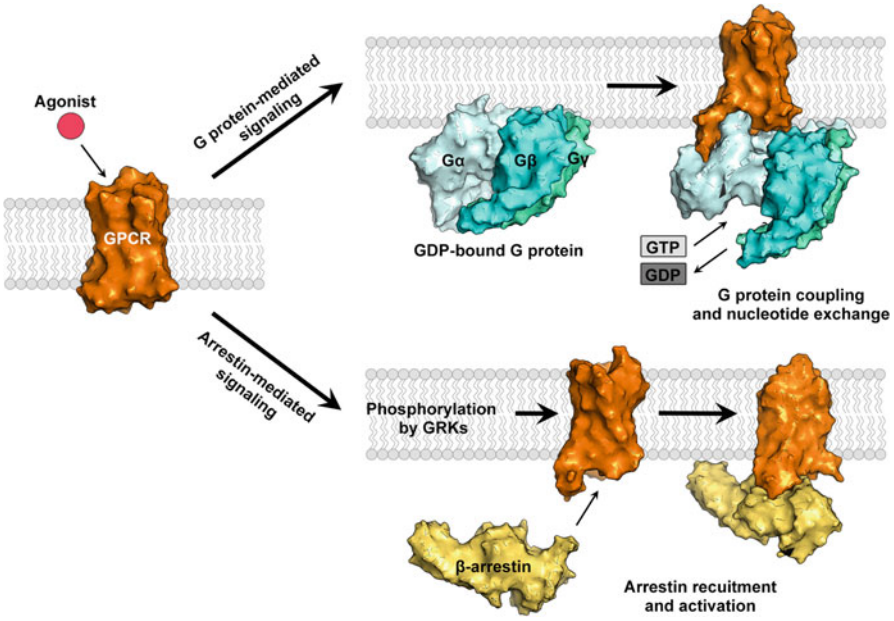


Fig. 1.4 Schematic of a heterotrimeric G protein, a G protein-coupled kinase, and an arrestin interaction with a GPCR

functional states and conformations to interact with G proteins, GRKs, and β-arrestins upon different stimuli, e.g., agonist binding. This phenomenon is known as biased signaling or biased agonism. The development of biased agonists targeting a particular signaling pathway and pharmacological output is suggested to increase drug efficacy with reduced side effects. Here, we introduce several biased agonists of GPCRs with improved therapeutic profiles.

1.2.1 G Protein-Biased Agonists

μ-Opioid receptor is a well-known drug target providing analgesia. However, traditional μ-opioid receptor agonists (such as morphine and fentanyl) are limited by side effects that include addiction, respiratory depression, constipation, and tolerance. Early studies of μ-opioid receptor indicated that activation of G protein signaling and blockage of β-arrestin signaling might enhance analgesia with reduced side effects [13, 14]. DeWire et al. reported TRV130 (compound **6**) as a G protein-biased agonist of μ-opioid receptor. This agonist elicits robust G protein signaling with potency and efficacy similar to morphine, but it has less β-arrestin recruitment and receptor internalization [31]. Compared with morphine, TRV130 increased both analgesia and pain relief and reduced on-target adverse effects in a randomized double-blind

controlled trial [110]. PZM21 (compound **7**) is another G protein-biased agonist of the μ-opioid receptor [81]. Discovered via computational modeling and structure-based optimization, this agonist is structurally distinct from previously reported μ-opioid receptor modulators. PZM21 has a potent Gαi signaling profile, decreased β-arrestin recruitment, improved analgesia efficacy, and fewer side effects compared with morphine in a preclinical pain model.

Another representative case of biased agonist development involves adenosine A1 receptor. The adenosine A1 receptor plays a major role in regulation of myocardial oxygen consumption and coronary blood flow [122]. Stimulation of adenosine A1 receptor has a myocardial depressant effect by decreasing the conduction of electrical impulses and suppressing pacemaker cell function, resulting in a decrease in heart rate. This makes its agonists useful medications for treating tachyarrhythmia. However, unbiased stimulation of adenosine A1 receptor results in several on-target side effects, including bradycardia, atrioventricular conduction blockade, and hypotension. Screening a series of 4-substituted 2-amino-3-benzoylthiophenes, Aurelio et al. identified a G protein-biased allosteric modulator of adenosine A1 receptor [3]. Using this allosteric modulator as a backbone, Valant et al. designed the G protein-biased allosteric agonist VCP746 (compound **8**) [121]. By modulating the adenosine A1 receptor, VCP746 provided a cytoprotective benefit to rat cardiomyocytes and stimulated anti-fibrotic signaling pathways in cardiac-derived cell lines without hemodynamic side effects, such as bradycardia [123]. This finding demonstrates the superiority of G protein-biased agonists of the adenosine A1 receptor.

1.2.2 β-Arrestin-Biased Agonists

In addition to G protein-biased modulators of GPCRs, several β-arrestin-biased allosteric agonists with unique therapeutic profiles have also been developed. Targeting neurotensin 1 receptor, ML314 (compound **9**) is a biased agonist operating via the β-arrestin pathway [46]. Compared with traditional agonists of the neurotensin 1 receptor, ML314 showed low potential for promiscuity and improved pharmacological data for the treatment of methamphetamine abuse [6, 97]. Another example of a β-arrestin-biased agonist is ORG27569 (compound **10**). As an allosteric modulator of cannabinoid receptor 1, ORG27569 enhances binding affinity of orthosteric agonists to receptors, inhibits G protein signaling efficacy, and promotes β-arrestin 2 activities [1]. In animal behavior tests, pretreatment with ORG27569 attenuated reinstatement of cocaine- and methamphetamine-seeking behavior, suggesting that ORG27569 is a potential pharmacotherapy for drug addiction [57]. TAK-875 (compound **11**) is a potent allosteric agonist of the free fatty acid receptor 1. This compound induces interactions between the receptor and β-arrestins 1 and 2 [80]. This agonist-induced β-arrestin 2 signaling contributes to insulin secretion in pancreatic β cells, indicating a new clinical strategy for treating type II diabetes.

1.3 Strategies in Allosteric Drug Development

1.3.1 Structural Study Regarding Allostery

The molecular mechanism of allosteric modulation relies on structural switches between particular conformations in distinct functional states. Atomic-level protein structure provides the molecular basis of allosteric drug development. Recent breakthroughs in crystallography have led to the widespread adoption of structure-based drug design methodologies [55, 61, 67, 101, 108]. X-ray crystallography can provide detailed structural information for a protein before and after perturbation. Moreover, some crystal structures were solved in complex with modulators, revealing a wide variety of allosteric binding sites [48, 56, 65, 74, 87, 93, 111, 115, 130, 133, 134]. Docking to such structures has proven to be highly effective in the discovery of novel ligands [88, 100, 127]. Many drug candidates developed through structure-based methods have entered preclinical assays and clinical trials [1, 3, 6, 31, 46, 57, 81, 97, 110, 121, 123]. However, the lack of dynamic information on the static crystal structure and possible bias related to conformational change due to crystal packing set limits on X-ray crystallography in allostery studies.

Nuclear magnetic resonance (NMR) spectroscopy provides experimental probes for studying protein dynamics on a wide range of timescales, ranging from picoseconds to milliseconds and beyond. Recent NMR methodological advancements have enabled the characterization of allosteric modulation mechanisms by increasing the proportion of the transient conformational ensembles that can be experimentally observed [15]. For example, using fluorine-19 NMR spectroscopy, Liu et al. monitored the local chemical environment on the intracellular face of the β2 adrenergic receptor and captured conformational changes of the receptor induced by various agonists [73]. Subsequent NMR studies offered plenty of structural data and further characterized the allosteric activation pathways of GPCRs involving crucial downstream effectors, such as β-arrestin [52, 131].

1.3.2 Allosteric Site Identification

Identification of allosteric sites in proteins of pharmaceutical interest is the first step in allosteric drug discovery. Despite many potential advantages of developing drugs targeting allosteric sites, identification of such sites is challenging. The physical and chemical properties of allosteric sites are not necessarily similar to those of conserved orthosteric sites. Some allosteric sites exist exclusively in intermediate functional states of proteins. In these scenarios, it is difficult to sufficiently detect allosteric sites from crystallographic or NMR structures. Alternatively, computational approaches are useful in helping researchers analyze and select potential allosteric sites for drug discovery. Here, we briefly introduce the computational approaches developed for the prediction of allosteric sites.

1.3.2.1 Sequence-Based Prediction Approaches

Statistical coupling analysis (SCA) [75] is a sequence-based approach that uses a multiple sequence alignment (MSA) to identify networks of coevolving residues in a protein family. The family members of a target protein are used to construct an MSA. Pairwise correlations were calculated between all pairs of residues in the alignment. The networks of coevolving residues are uncovered to provide a molecular basis for allosteric communications between functional and allosteric sites. Mutational analysis can be employed to validate the predicted allosteric site and coevolving network. Recently, Novinec et al. used SCA to analyze the family of papain-like cysteine peptidases [90]. They computationally identified and experimentally validated 14 residues and 8 allosteric sites that were involved in the allosteric communications of cathepsin K. An active compound, NSC13345, was identified in the virtual screening against these allosteric sites. Further X-ray crystallographic investigation of the cathepsin K-NSC13345 complex showed a novel allosteric site in the protein. These results prove that it is feasible to predict allosteric sites by exploiting the SCA method.

1.3.2.2 Dynamics-Based Prediction Approaches

Complementing experimental approaches, computational methods can provide massive structural information and dynamic details regarding allostery. Molecular dynamics (MD) simulation is the most commonly used method in protein dynamics characterization. In MD simulations, Newton's equations of motion are integrated over time to generate the trajectories followed by each particle in molecular systems [58]. The large number of snapshots generated from MD simulations captures the motion of the proteins, thus providing structural information of allosteric transitions. In a previous study of an anti-anthrax target, i.e., *Bacillus anthracis* adenylyl cyclase toxin (EF), Laine et al. used MD simulations to generate a conformational pathway connecting the known inactive and active structures of the target protein [68]. The resulting intermediate conformations revealed the presence of a pocket on the surface of EF, which underwent significant structural changes during the inactive-to-active transition. Laine et al. suggested this pocket as an allosteric site that regulates the calmodulin-induced activation of EF and then discovered EF inhibitors binding to this site using in vitro assays [68].

While traditional MD simulations have been used to identify allosteric sites and elucidate protein conformational diversity, Markov state models can further characterize the probability of these conformations, guiding rational design against dynamic proteins [25, 79, 95, 105, 109]. Briefly, a Markov state model is a stochastic kinetic model that describes the probability of transitioning between discrete states at a fixed interval. The probability of transitioning between these states is independent of previous transitions. By clustering protein structures extracted from a simulation trajectory, it is possible to identify discrete conformational states required in Markov

state models. In a recent study of the β2 adrenergic receptor, Kohlhoff et al. used Markov state models to aggregate independent MD simulations and revealed multiple activation pathways of the receptor. The resulting intermediate conformations were successfully exploited in a subsequent virtual screening campaign [60]. Additionally, researchers have also used Markov state models to accurately determine the ligand-protein binding kinetics (k_{on} and k_{off}) [20, 117]. While this approach is currently too computationally demanding for virtual screening, Markov state models provided a strong foundation regarding protein-ligand interactions and allosteric mechanisms for rational drug discovery.

Compared with MD simulations, the coarse-grained and lattice modeling need much fewer computational resources. Instead of explicitly representing every atom of the system, coarse-grained simulations use "pseudo-atoms" approximating groups of atoms, such as a whole amino acid residue. The coarse-grained two-state model can be used to identify allosteric sites [99]. Based on the concept that allostery is a conformation population shift process, an ensemble including two functional states of a protein is constructed. When perturbations are added to a site, the ensemble is monitored. If the conformation population redistribution is observed, the perturbed site is predicted as an allosteric site. Applying this approach, Qi et al. successfully identified new allosteric sites of *Escherichia coli* phosphoglycerate dehydrogenase and discovered inhibitors regulating the *Escherichia coli* serine synthesis pathway [99].

Normal mode analysis (NMA) is another method that characterizes protein dynamics. It was developed based on the hypothesis that the largest movements in a protein are functionally relevant. This method captures the functional motions of a protein and unearths sites participating in allosteric communications [5]. Compared with MD simulations, NMA is better suited to study large structural rearrangements of biomolecules with fewer computational costs in terms of CPU time (although more expensive in terms of memory). The low computing cost of NMA contributed to the development of several web servers performing normal mode calculations. PARS (http://bioinf.uab.cat/cgi-bin/pars-cgi/pars.pl) [96] is an NMA web server predicting allosteric sites based on the alteration of protein flexibility upon ligand binding. Given a protein submitted by a user, NMA is performed for the apoprotein and protein-ligand complex, in which ligands are simplistically represented by dummy atoms. The differences in NMA-derived B factors between the apo and ligand-bound states of the protein are compared. If a significant change is observed, the place where the ligand is positioned is proposed to be an allosteric site. This web server successfully predicted 44% of known allosteric sites from the benchmark set [96]. SPACER (http://allostery.bii.a-star.edu.sg) is a web server integrating NMA with Monte Carlo simulations to predict biologically active sites, including allosteric sites [39]. It uses simulations to produce putative active sites, which are further described by low-frequency normal modes. This method can identify latent allosteric sites in structurally homologous proteins [119].

There are a number of open tools allowing users to analyze protein conformations from simulations or modeling trajectories. They were developed based on geometry- or/and energy-based methods. Geometry-based site detection methods, such as

POCASA [132], CASTp [11], Fpocket [70], Q-SiteFinder [69], and POVME [32], identify ligand-binding sites by considering only the atomic coordinates of a target protein. All predicted sites are ranked by simple geometric metrics, such as volume and shape. A site is rejected when it is shallow or partially collapsed. Energy-based site detection methods evaluate the interactions between small organic probes and the target protein to identify binding sites. For example, SiteMap [44] and SITEHOUND [45] dock a methane or water molecule to the target protein to evaluate the likelihood of small-molecule binding. FTMap [18, 63, 64], FTSite [89], and GRID [83] consider multiple chemically diverse probes in the identification of druggable hot spots. Compared with geometry-based methods, energetic-based methods are more computationally demanding, which limits their applicability to large-scale structural data sets. Both geometry- and energy-based tools are efficient in identifying allosteric sites that only occasionally manifest themselves over time.

1.3.2.3 Knowledge-Based Prediction Approaches

Knowledge-based site detection methods query existing databases to determine ligand-binding sites. For example, 3DLigandSite [125], FINDSITE [19], and Pocketome [66] predict ligand-binding sites by searching databases for ligand-bound proteins that are structurally similar to the target protein. To identify allosteric sites, other methods, such as FRpred [36], ConSeq [10], and ConCavity [22], use hybrid multiple-dimensional information, including sequence conservation, presence of surface residues, and structural homology. Evolutionarily important individual protein residues or residue groups are identified and further examined with additional knowledge, such as structural homology.

Recently, a specialized database dedicated to allostery was developed to serve as a source of knowledge on allosteric sites. To systemically interpret the increasing allostery information, Huang et al. built the AlloSteric Database (ASD, http://mdl.shsmu.edu.cn/ASD/) describing hundreds of allosteric proteins and thousands of allosteric modulators [50]. The ASD has been developed to provide comprehensive information characterizing allosteric regulation, ranging from allosteric proteins and modulators to interactions, sites, pathways, functions, and associated diseases. In the recently updated version of this database, two of the largest human allosteromes of protein kinases and GPCRs were constructed, and more than 1600 allosteric actions were identified in allosteric networks [107]. These data have the benefit of prompting an investigation of allosteric mechanisms, allosteric-related diseases, and allosteric drugs. Chemists can use allosteric modulators in ASD to implement structural modifications for rational drug design. Integrating the allosteric protein information in ASD enables the allosteric site identification and characterization. Based on ASD, a server-based model called Allosite [49] was developed to predict allosteric sites. In a fivefold cross-validation test, Allosite showed high sensitivity and high specificity. Because it is capable of discovering novel allosteric sites, Allosite is a powerful tool in rational drug design targeting allosteric sites.

1.3.3 Experimental Assays for Screening Allosteric Modulators

Experimental screening and validation of allosteric modulators are uniquely challenging in drug development. Traditionally, experimental assays measure orthosteric function rather than ligand binding at the allosteric site. Efficient development of allosteric drugs requires qualitative and quantitative characterization of ligand-binding affinity and allosteric efficacy, including conformational changes and downstream response [91, 128]. Here, we describe several pharmacological approaches that can be used to screen allosteric modulators.

1.3.3.1 Binding Assays

Binding assays are important tools for studying protein allostery, as they can often directly validate whether a ligand has an allosteric mode of action. At the early stages of the drug discovery process, the radioligand assay can be used to identify allosteric modulators. In this assay, a potent ligand bearing a radionuclide is used as a probe [59]. In the presence of candidate modulators, the allosteric activity is detected by monitoring the probe dissociation kinetics (k_{on} and k_{off}) [41]. This approach is applicable to different experimental preparation types, such as whole cells, tissue sections, or cellular membrane preparations. However, technical constraints and radioactive waste management make it poorly amenable to high-throughput screening.

Fluorescent ligand assays overcome the shortages of radioligand assays and offer the significant advantage of having a real-time readout. Many fluorescent ligand-based technologies have been developed, including but not limited to time-resolved fluorescence intensity, fluorescence resonance energy transfer (FRET), time-resolved FRET, and bioluminescence resonance energy transfer (BRET) [53]. Among these techniques, approaches based on resonance energy transfer (RET) are currently widely used in drug screening. In RET assays, the probes can be endogenous ligands, synthetic small molecules, and antibodies labeled with fluorophore, quencher, or lanthanide ions, while the proteins can be labeled using fluorescent proteins, enzyme tags, or antibodies [51, 78]. These assays can serve as first filters in screening allosteric modulators targeting transmembrane proteins. For example, the DTect-All™ FRET-based binding platform can identify allosteric effectors interacting with the transmembrane domain of GPCRs. In this platform, a fluorophore is linked to a probe, while a green fluorescent protein (GFP) is fused to the amino terminal part of the target receptor [51, 114]. Because the amino terminal domain is truncated, candidate modulators only bind to the extracellular and transmembrane domains, where most allosteric sites are located. DTect-All™ assays have been successfully used to identify several allosteric modulators of glucagon-like peptide 1 receptor [129] and metabotropic glutamate receptors [103]. However,

fluorescent ligand assays still have drawbacks. As a label to a probe or a target protein, a fluorophore is generally large, which may alter allosteric interactions.

NMR techniques can also be employed in allosteric binding assays. For example, saturation transfer difference NMR methods are commonly utilized to differentiate between competitive and allosteric inhibitors [21, 28]. This method is implemented as a competitive binding experiment in the presence of excess orthosteric ligands. Another efficient NMR method is the target immobilized NMR screening, which is a fragment-based approach [24]. In this assay, the target protein and a reference protein are immobilized on a solid support. Chemical fragments that bind specifically to the target will be detected in NMR spectra. This approach is particularly robust for identifying both allosteric inhibitors and agonists of unstable membrane proteins. Last but not least, fluorine chemical shift anisotropy and exchange is a highly sensitive and reliable screening method for identifying allosteric effectors [29, 30, 124]. In this assay, the use of fluorine labels is not limited to ligands but is applicable to the target protein [2, 38, 98]. In a study of master coactivator CBP/p300, the binding of allosteric modulators and conformational changes of functionally important KIX domain were examined by labeling the aromatic amino acids at protein-biomolecule interfaces [98].

1.3.3.2 Functional Assays

Because the binding affinity of an allosteric modulator does not always correlate with its efficacy on modulating protein functions, functional assays are required in the screening of allosteric modulators. One major advantage of functional assays is that the desired functional output can be directly observed in screening. In the pharmacological study of important membrane proteins, such as GPCRs, functional assays remain the current standard for high-throughput screening [82]. To detect and quantify allosteric efficacy, a typical functional assay usually examines the ability of the allosteric modulator to alter the half maximum effect concentration (EC_{50}) and/or the maximum effect concentration (E_{max}) of the orthosteric agonist. The choice of functional assays is highly dependent on the functional roles of the target protein. It is impossible to provide a full description of all functional assays. Rather, we focus on the well-known drug target GPCR family and describe several mature functional assays in screening GPCR allosteric modulators.

GPCR functional assays may provide a direct measure of receptor activation by monitoring stimulation and/or recruitment of receptor primary effectors, such as β-arrestins and G proteins. β-arrestins not only regulate receptor desensitization, interaction, and recycling but also serve as signaling scaffolds to regulate various signaling networks downstream of the corresponding receptors. As a result, functional assays measuring ligand-induced β-arrestin engagement have become increasingly important in GPCR drug discovery programs. In typical microscopic imaging assays, ligand-induced translation of GFP-tagged β-arrestins from the cytosol to the activated GPCR-containing pits is monitored with epifluorescence microscopy, confocal microscopy, or microplate laser scanning cytometry-based imaging

systems [16, 33, 43]. These imaging assays not only quantitatively characterize the β-arrestin engagement but also provide information on cytotoxic effects of allosteric modulators. As well-studied primary effectors of GPCRs, G proteins are activated by the conformational changes of the corresponding receptors. Energy transfer assays such as FRET and BRET can be used to directly measure the recruitment of different G protein subtypes to a given receptor in real time in living cells [54]. In G protein engagement assays, energy donor and acceptor are fused to the subunits of a G protein. Agonist-induced activation of a GPCR produces intramolecular structural switch of the G protein. Energy transfer efficiency between the tagged G protein subunits is correlated with the allosteric efficacy of the receptor. These BRET-based assays can also be used to detect β-arrestin, which allows comprehensive characterization of the allosteric modulator behaviors.

Other functional assays of GPCR focus on the downstream signaling cascade, such as second-messenger production. Cyclic adenosine monophosphate (cAMP) is a second messenger of many GPCRs. The change in cellular cAMP level has widely been used to estimate the biological activity of various GPCR modulators. From a variety of cellular cAMP assays, FRET-based biosensors allow real-time detection of changes in cellular cAMP levels [47]. For example, fluorescence lifetime imaging microscopy is a technique measuring the changes in cAMP concentration. This measurement relies on the relative fluorescence of donor and acceptor fluorophores fused to a cAMP-binding protein moiety. The sensor proteins are distributed in the cell cytosol. Binding of cAMP results in a decrease in acceptor fluorescence and an increase in donor fluorescence. The calculated acceptor/donor emission ratio correlates with the changes in intracellular cAMP. In addition to FRET-based assays, immunoassays can also help characterize cAMP levels. In a typical competitive cAMP immunoassay [7, 12, 34, 126], labeled cAMP and cellular cAMP compete for binding to an anti-cAMP antibody. As binding of labeled cAMP to an anti-cAMP antibody will send a signal, a decreased signal is observed with increasing intracellular cAMP levels. The cAMP assays described above could also be adapted to the analysis of another second messenger, such as cyclic guanosine monophosphate.

If cAMP accumulation is observed, most high-throughput screening campaigns use calcium mobilization assays to identify GPCR modulators. Calcium mobilization results from the activation of certain GPCRs, which bind to a specific Gα subunit, Gq. Gq-coupled receptors can active phospholipase C pathways and lead to an increase of inositol triphosphate (IP_3) level. IP_3 activates IP_3-sensitive calcium channels, which results in a calcium release from the endoplasmic reticulum. If the receptor is not naturally coupled to Gq, a promiscuous or chimeric G protein co-expressed with the receptor can be employed to modulate calcium mobilization by redirecting the signaling pathway [92]. In these assays, cells are loaded with a dye that fluoresces upon calcium ion binding. The fluorescence magnitude emission is directly correlated with calcium response magnitude. This type of assay is particularly interesting because it enables screening of GPCR modulators without knowledge of specific receptor signaling pathways. However, the usage of calcium assays in detecting slow-binding ligands is not recommended because calcium efflux transiently occurs within seconds. An alternative method to monitor calcium flux

involves the measurement of IP_3 and/or diacylglycerol (DAG). In a scintillation proximity assay of inositol phosphates, the label is incorporated into cellular phosphatidylinositol 4,5-bisphosphate, which is hydrolyzed to IP_3 and DAG upon receptor activation [17]. The mass of soluble inositol phosphates is a quantitative readout of receptor activation, which is amenable to high-throughput screening. In functional assays monitoring DAG levels, genetically encoded biosensors of DAG are used to detect the DAG binding to protein kinase C after receptor activation [116]. The change in fluorescence indicates the intracellular DAG levels.

1.3.3.3 Bias in Assays

In the past decades, applications of binding assays and functional assays have not only led to the discovery of many promising allosteric modulators of GPCR but have also provided crucial information on the corresponding signaling pathway. However, the discovery of allosteric modulators remains challenging because the allosteric efficacy may be not always readily apparent. The data interpretation might depend on the assay conditions, such as the probe used and the protein expression level. For a target protein having multiple orthosteric ligands, the use of different ligands as probes can give opposite activities for the same allosteric modulators. This phenomenon was illustrated with the characterization of glucagon-like peptide 1 receptor [62, 129]. To avoid such "probe bias," it is necessary to test allosteric modulators with various probes, including endogenous ligands and surrogate agonists. The efficacy of an allosteric modulator also depends on the expression level of its target protein or/and its downstream transducer elements in a given tissue or system, which is known as "system bias" [135]. For example, due to the distinct expression of β-arrestins and G protein-coupled receptor kinases across brain regions, agonists targeting dopamine receptor 2 show different effects in the striatum and prefrontal cortex [120]. System bias may be observed in a situation where all measurements of a specific assay are amplified or diminished. To identify efficient allosteric modulators, it is crucial to remove the effects of system bias from the cellular response.

1.4 Concluding Remarks and Future Perspectives

The tremendous technical progresses in the basic and translational studies of allosteric regulation have led to the discovery of many promising allosteric modulators. By offering high selectivity and improved physiochemical properties, allosteric modulators allow researchers to dissect the functional roles and therapeutic potential of a protein in its highly conserved family. Advanced experimental and computational methods detecting transient allosteric sites even offer new therapeutic opportunities for many proteins that were previously considered "undruggable." Numerous binding assays and functional assays have been developed for

high-throughput and routine screening of allosteric modulators, suggesting that this field is maturing.

Although impressive progress has been made, there are still considerable challenges in allosteric drug development. First, allosteric drug development is restricted by the small number of experimentally observed allosteric sites. Second, more effort must be spent on improving the performance of computational methods predicting allosteric models and active sites. Simplified modeling, such as coarse-grained simulations, might miss essential interactions modulating allosteric communications. All-atom molecular dynamics simulation is more accurate but computationally expensive. Moreover, all structure- or topology-based prediction models suffer from biased force fields. In the near future, the soaring computation power and the rapid development of force fields may help solve these problems. Last but not least, current studies of allostery have largely focused on single proteins. Proteins function through highly interconnected cellular pathways. To fully describe the allosteric effects of a modulator, it should be put into an allosteric network including multiple proteins in the framework of a cell system. Benefiting from the massive accumulated information of allosteric modulation, researchers in this field have constructed allosteric networks of many important drug targets using specialized databases such as ASD. We expect the comprehensive understanding of allostery in such a network manner may lead to new therapeutic developments and disease treatments.

References

1. Ahn KH, Mahmoud MM, Shim JY, Kendall DA (2013) Distinct roles of beta-arrestin 1 and beta-arrestin 2 in ORG27569-induced biased signaling and internalization of the cannabinoid receptor 1 (CB1). J Biol Chem 288(14):9790–9800
2. Assemat O, Antoine M, Fourquez JM, Wierzbicki M, Charton Y, Hennig P, Perron-Sierra F, Ferry G, Boutin JA, Delsuc MA (2015) ^{19}F nuclear magnetic resonance screening of glucokinase activators. Anal Biochem 477:62–68
3. Aurelio L, Valant C, Flynn BL, Sexton PM, Christopoulos A, Scammells PJ (2009) Allosteric modulators of the adenosine A1 receptor: synthesis and pharmacological evaluation of 4-substituted 2-amino-3-benzoylthiophenes. J Med Chem 52(14):4543–4547
4. Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE, Forman-Kay JD (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature 519(7541):106–109
5. Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 110(3):1463–1497
6. Barak LS, Bai Y, Peterson S, Evron T, Urs NM, Peddibhotla S, Hedrick MP, Hershberger P, Maloney PR, Chung TD, Rodriguiz RM, Wetsel WC, Thomas JB, Hanson GR, Pinkerton AB, Caron MG (2016) ML314: a biased neurotensin receptor ligand for methamphetamine abuse. ACS Chem Biol 11(7):1880–1890
7. Beaudet L, Bedard J, Breton B, Mercuri RJ, Budarf ML (2001) Homogeneous assays for single-nucleotide polymorphism typing using AlphaScreen. Genome Res 11(4):600–608
8. Beaulieu JM, Sotnikova TD, Marion S, Lefkowitz RJ, Gainetdinov RR, Caron MG (2005) An Akt/beta-arrestin 2/PP2A signaling complex mediates dopaminergic neurotransmission and behavior. Cell 122(2):261–273

9. Benovic JL, Strasser RH, Caron MG, Lefkowitz RJ (1986) Beta-adrenergic receptor kinase: identification of a novel protein kinase that phosphorylates the agonist-occupied form of the receptor. Proc Natl Acad Sci U S A 83(9):2797–2801
10. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 20(8):1322–1324
11. Binkowski TA, Naghibzadeh S, Liang J (2003) CASTp: computed atlas of surface topography of proteins. Nucleic Acids Res 31(13):3352–3355
12. Blackburn GF, Shah HP, Kenten JH, Leland J, Kamin RA, Link J, Peterman J, Powell MJ, Shah A, Talley DB et al (1991) Electrochemiluminescence detection for development of immunoassays and DNA probe assays for clinical diagnostics. Clin Chem 37(9):1534–1539
13. Bohn LM, Gainetdinov RR, Lin FT, Lefkowitz RJ, Caron MG (2000) Mu-opioid receptor desensitization by beta-arrestin-2 determines morphine tolerance but not dependence. Nature 408(6813):720–723
14. Bohn LM, Lefkowitz RJ, Gainetdinov RR, Peppel K, Caron MG, Lin FT (1999) Enhanced morphine analgesia in mice lacking beta-arrestin 2. Science 286(5449):2495–2498
15. Boulton S, Melacini G (2016) Advances in NMR methods to map allosteric sites: from models to translation. Chem Rev 116(11):6267–6304
16. Bowen WP, Wylie PG (2006) Application of laser-scanning fluorescence microplate cytometry in high content screening. Assay Drug Dev Technol 4(2):209–221
17. Brandish PE, Hill LA, Zheng W, Scolnick EM (2003) Scintillation proximity assay of inositol phosphates in cell extracts: high-throughput measurement of G-protein-coupled receptor activation. Anal Biochem 313(2):311–318
18. Brenke R, Kozakov D, Chuang GY, Beglov D, Hall D, Landon MR, Mattos C, Vajda S (2009) Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. Bioinformatics 25(5):621–627
19. Brylinski M, Skolnick J (2008) A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc Natl Acad Sci U S A 105(1):129–134
20. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc Natl Acad Sci U S A 108(25):10184–10189
21. Cala O, Guilliere F, Krimm I (2014) NMR-based analysis of protein-ligand interactions. Anal Bioanal Chem 406(4):943–956
22. Capra JA, Laskowski RA, Thornton JM, Singh M, Funkhouser TA (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol 5(12):e1000585
23. Changeux JP (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 41:103–133
24. Chen D, Errey JC, Heitman LH, Marshall FH, Ijzerman AP, Siegal G (2012) Fragment screening of GPCRs using biophysical methods: identification of ligands of the adenosine A (2A) receptor with novel biological activity. ACS Chem Biol 7(12):2064–2073
25. Chodera JD, Noe F (2014) Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol 25:135–144
26. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35(10):539–546
27. Cui Q, Karplus M (2008) Allostery and cooperativity revisited. Protein Sci 17(8):1295–1307
28. Dalvit C, Caronni D, Mongelli N, Veronesi M, Vulpetti A (2006) NMR-based quality control approach for the identification of false positives and false negatives in high throughput screening. Curr Drug Discov Technol 3(2):115–124
29. Dalvit C, Fagerness PE, Hadden DT, Sarver RW, Stockman BJ (2003) Fluorine-NMR experiments for high-throughput screening: theoretical aspects, practical considerations, and range of applicability. J Am Chem Soc 125(25):7696–7703

30. Dalvit C, Mongelli N, Papeo G, Giordano P, Veronesi M, Moskau D, Kummerle R (2005) Sensitivity improvement in 19F NMR-based screening experiments: theoretical considerations and experimental applications. J Am Chem Soc 127(38):13380–13385
31. DeWire SM, Yamashita DS, Rominger DH, Liu G, Cowan CL, Graczyk TM, Chen XT, Pitis PM, Gotchev D, Yuan C, Koblish M, Lark MW, Violin JD (2013) A G protein-biased ligand at the mu-opioid receptor is potently analgesic with reduced gastrointestinal and respiratory dysfunction compared with morphine. J Pharmacol Exp Ther 344(3):708–717
32. Durrant JD, de Oliveira CA, McCammon JA (2011) POVME: an algorithm for measuring binding-pocket volumes. J Mol Graph Model 29(5):773–776
33. Eggeling C, Brand L, Ullmann D, Jager S (2003) Highly sensitive fluorescence detection technology currently available for HTS. Drug Discov Today 8(14):632–641
34. Eglen RM (2002) Enzyme fragment complementation: a flexible high throughput screening assay technology. Assay Drug Dev Technol 1(1 Pt 1):97–104
35. Eichel K, Jullie D, von Zastrow M (2016) Beta-Arrestin drives MAP kinase signalling from clathrin-coated structures after GPCR dissociation. Nat Cell Biol 18(3):303–310
36. Fischer JD, Mayer CE, Soding J (2008) Prediction of protein functional residues from sequence by probability density estimation. Bioinformatics 24(5):613–620
37. Flaherty KT, Infante JR, Daud A, Gonzalez R, Kefford RF, Sosman J, Hamid O, Schuchter L, Cebon J, Ibrahim N, Kudchadkar R, Burris HA 3rd, Falchook G, Algazi A, Lewis K, Long GV, Puzanov I, Lebowitz P, Singh A, Little S, Sun P, Allred A, Ouellet D, Kim KB, Patel K, Weber J (2012) Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. N Engl J Med 367(18):1694–1703
38. Ge X, MacRaild CA, Devine SM, Debono CO, Wang G, Scammells PJ, Scanlon MJ, Anders RF, Foley M, Norton RS (2014) Ligand-induced conformational change of Plasmodium falciparum AMA1 detected using 19F NMR. J Med Chem 57(15):6419–6427
39. Goncearenco A, Mitternacht S, Yong T, Eisenhaber B, Eisenhaber F, Berezovsky IN (2013) SPACER: server for predicting allosteric communication and effects of regulation. Nucleic Acids Res 41(web server issue):W266–W272
40. Goodman OB Jr, Krupnick JG, Santini F, Gurevich VV, Penn RB, Gagnon AW, Keen JH, Benovic JL (1996) Beta-arrestin acts as a clathrin adaptor in endocytosis of the beta2-adrenergic receptor. Nature 383(6599):447–450
41. Goupil E, Laporte SA, Hebert TE (2013) A simple method to detect allostery in GPCR dimers. Methods Cell Biol 117:165–179
42. Gustavsson M, Verardi R, Mullen DG, Mote KR, Traaseth NJ, Gopinath T, Veglia G (2013) Allosteric regulation of SERCA by phosphorylation-mediated conformational shift of phospholamban. Proc Natl Acad Sci U S A 110(43):17338–17343
43. Haasen D, Wolff M, Valler MJ, Heilker R (2006) Comparison of G-protein coupled receptor desensitization-related beta-arrestin redistribution using confocal and non-confocal imaging. Comb Chem High Throughput Screen 9(1):37–47
44. Halgren TA (2009) Identifying and characterizing binding sites and assessing druggability. J Chem Inf Model 49(2):377–389
45. Hernandez M, Ghersi D, Sanchez R (2009) SITEHOUND-web: a server for ligand binding site identification in protein structures. Nucleic Acids Res 37(Web Server issue):W413–W416
46. Hershberger P, Hedrick M, Peddibhotla S, Maloney P, Li Y, Milewski M, Gosalia P, Gray W, Mehta A, Sugarman E, Hood B, Suyama E, Nguyen K, Heynen-Genel S, Vasile S, Salaniwal S, Stonich D, Su Y, Mangravita-Novo A, Vicchiarelli M, Smith LH, Roth G, Diwan J, Chung TDY, Caron MG, Thomas JB, Pinkerton AB, Barak LR (2010) Small molecule agonists for the neurotensin 1 receptor (NTR1 agonists). Probe reports from the NIH Molecular Libraries Program. Bethesda (MD),
47. Hill SJ, Williams C, May LT (2010) Insights into GPCR pharmacology from the measurement of changes in intracellular cyclic AMP; advantages and pitfalls of differing methodologies. Br J Pharmacol 161(6):1266–1275
48. Hollenstein K, Kean J, Bortolato A, Cheng RK, Dore AS, Jazayeri A, Cooke RM, Weir M, Marshall FH (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. Nature 499(7459):438–443

49. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T, Zhang J (2013) Allosite: a method for predicting allosteric sites. Bioinformatics 29(18):2357–2359
50. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39(Database issue):D663–D669
51. Ilien B, Franchet C, Bernard P, Morisset S, Weill CO, Bourguignon JJ, Hibert M, Galzi JL (2003) Fluorescence resonance energy transfer to probe human M1 muscarinic receptor structure and drug binding properties. J Neurochem 85(3):768–778
52. Isogai S, Deupi X, Opitz C, Heydenreich FM, Tsai CJ, Brueckner F, Schertler GF, Veprintsev DB, Grzesiek S (2016) Backbone NMR reveals allosteric signal transduction networks in the beta1-adrenergic receptor. Nature 530(7589):237–241
53. Jaeger WC, Armstrong SP, Hill SJ, Pfleger KD (2014) Biophysical detection of diversity and bias in GPCR function. Front Endocrinol (Lausanne) 5:26
54. Janetopoulos C, Jin T, Devreotes P (2001) Receptor-mediated activation of heterotrimeric G-proteins in living cells. Science 291(5512):2408–2411
55. Jazayeri A, Dias JM, Marshall FH (2015) From G protein-coupled receptor structure resolution to rational drug design. J Biol Chem 290(32):19489–19495
56. Jazayeri A, Dore AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, Bortolato A, Koglin M, Robertson NJ, Errey JC, Andrews SP, Teobald I, Brown AJ, Cooke RM, Weir M, Marshall FH (2016) Extra-helical binding site of a glucagon receptor antagonist. Nature 533(7602):274–277
57. Jing L, Qiu Y, Zhang Y, Li JX (2014) Effects of the cannabinoid CB (1) receptor allosteric modulator ORG 27569 on reinstatement of cocaine- and methamphetamine-seeking behavior in rats. Drug Alcohol Depend 143:251–256
58. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652
59. Knudsen LB, Kiel D, Teng M, Behrens C, Bhumralkar D, Kodra JT, Holst JJ, Jeppesen CB, Johnson MD, de Jong JC, Jorgensen AS, Kercher T, Kostrowicki J, Madsen P, Olesen PH, Petersen JS, Poulsen F, Sidelmann UG, Sturis J, Truesdale L, May J, Lau J (2007) Small-molecule agonists for the glucagon-like peptide 1 receptor. Proc Natl Acad Sci U S A 104(3):937–942
60. Kohlhoff KJ, Shukla D, Lawrenz M, Bowman GR, Konerding DE, Belov D, Altman RB, Pande VS (2014) Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. Nat Chem 6(1):15–21
61. Kooistra AJ, Leurs R, de Esch IJ, de Graaf C (2014) From three-dimensional GPCR structure to rational ligand discovery. Adv Exp Med Biol 796:129–157
62. Koole C, Wootten D, Simms J, Valant C, Sridhar R, Woodman OL, Miller LJ, Summers RJ, Christopoulos A, Sexton PM (2010) Allosteric ligands of the glucagon-like peptide 1 receptor (GLP-1R) differentially modulate endogenous and exogenous peptide responses in a pathway-selective manner: implications for drug screening. Mol Pharmacol 78(3):456–465
63. Kozakov D, Grove LE, Hall DR, Bohnuud T, Mottarella SE, Luo L, Xia B, Beglov D, Vajda S (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. Nat Protoc 10(5):733–755
64. Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE, Beglov D, Pelletier J, Whitty A, Vajda S (2011) Structural conservation of druggable hot spots in protein-protein interfaces. Proc Natl Acad Sci U S A 108(33):13528–13533
65. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hubner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature 504(7478):101–106
66. Kufareva I, Ilatovskiy AV, Abagyan R (2012) Pocketome: an encyclopedia of small-molecule binding sites in 4D. Nucleic Acids Res 40(Database issue):D535–D540
67. Kumari P, Ghosh E, Shukla AK (2015) Emerging approaches to GPCR ligand screening for drug discovery. Trends Mol Med 21(11):687–701

68. Laine E, Goncalves C, Karst JC, Lesnard A, Rault S, Tang WJ, Malliavin TE, Ladant D, Blondel A (2010) Use of allostery to identify inhibitors of calmodulin-induced activation of Bacillus anthracis edema factor. Proc Natl Acad Sci U S A 107(25):11277–11282
69. Laurie AT, Jackson RM (2005) Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21(9):1908–1916
70. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. BMC Bioinf 10:168
71. Lechtenberg BC, Freund SM, Huntington JA (2012) An ensemble view of thrombin allostery. Biol Chem 393(9):889–898
72. Levy JA (2009) HIV pathogenesis: 25 years of progress and persistent challenges. AIDS 23(2):147–160
73. Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012a) Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR. Science 335(6072):1106–1110
74. Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, AP IJ, Cherezov V, Stevens RC (2012b) Structural basis for allosteric regulation of GPCRs by sodium ions. Science 337(6091):232–236
75. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286(5438):295–299
76. Lohse MJ, Benovic JL, Codina J, Caron MG, Lefkowitz RJ (1990) Beta-arrestin: a protein that regulates beta-adrenergic receptor function. Science 248(4962):1547–1550
77. Luttrell LM, Ferguson SS, Daaka Y, Miller WE, Maudsley S, Della Rocca GJ, Lin F, Kawakatsu H, Owada K, Luttrell DK, Caron MG, Lefkowitz RJ (1999) Beta-arrestin-dependent formation of beta2 adrenergic receptor-Src protein kinase complexes. Science 283(5402):655–661
78. Ma Z, Du L, Li M (2014) Toward fluorescent probes for G-protein-coupled receptors (GPCRs). J Med Chem 57(20):8187–8203
79. Malmstrom RD, Lee CT, Van Wart A, Amaro RE (2014) On the application of molecular-dynamics based Markov state models to functional proteins. J Chem Theory Comput 10(7):2648–2657
80. Mancini AD, Bertrand G, Vivot K, Carpentier E, Tremblay C, Ghislain J, Bouvier M, Poitout V (2015) Beta-arrestin recruitment and biased agonism at free fatty acid receptor 1. J Biol Chem 290(34):21131–21140
81. Manglik A, Lin H, Aryal DK, McCorvy JD, Dengler D, Corder G, Levit A, Kling RC, Bernat V, Hubner H, Huang XP, Sassano MF, Giguere PM, Lober S, Da D, Scherrer G, Kobilka BK, Gmeiner P, Roth BL, Shoichet BK (2016) Structure-based discovery of opioid analgesics with reduced side effects. Nature 537(7619):185–190
82. May LT, Leach K, Sexton PM, Christopoulos A (2007) Allosteric modulation of G protein-coupled receptors. Annu Rev Pharmacol Toxicol 47:1–51
83. Miranker A, Karplus M (1991) Functionality maps of binding sites: a multiple copy simultaneous search method. Proteins 11(1):29–34
84. Mohler H, Fritschy JM, Rudolph U (2002) A new benzodiazepine pharmacology. J Pharmacol Exp Ther 300(1):2–8
85. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
86. Montaner JS, Reiss P, Cooper D, Vella S, Harris M, Conway B, Wainberg MA, Smith D, Robinson P, Hall D, Myers M, Lange JM (1998) A randomized, double-blind trial comparing combinations of nevirapine, didanosine, and zidovudine for HIV-infected patients: the INCAS trial. Italy, The Netherlands, Canada and Australia study. JAMA 279(12):930–937
87. Nawaratne V, Leach K, Felder CC, Sexton PM, Christopoulos A (2010) Structural determinants of allosteric agonism and modulation at the M4 muscarinic acetylcholine receptor: identification of ligand-specific and global activation mechanisms. J Biol Chem 285(25):19012–19021
88. Negri A, Rives ML, Caspers MJ, Prisinzano TE, Javitch JA, Filizola M (2013) Discovery of a novel selective kappa-opioid receptor agonist using crystal structure-based virtual screening. J Chem Inf Model 53(3):521–526

89. Ngan CH, Hall DR, Zerbe B, Grove LE, Kozakov D, Vajda S (2012) FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics 28(2):286–287
90. Novinec M, Korenc M, Caflisch A, Ranganathan R, Lenarcic B, Baici A (2014) A novel allosteric mechanism in the cysteine peptidase cathepsin K discovered by computational methods. Nat Commun 5:3287
91. Nussinov R, Tsai CJ (2012) The different ways through which specificity works in orthosteric and allosteric drugs. Curr Pharm Design 18(9):1311–1316
92. Offermanns S, Simon MI (1995) G alpha 15 and G alpha 16 couple a wide variety of receptors to phospholipase C. J Biol Chem 270(25):15175–15180
93. Oswald C, Rappas M, Kean J, Dore AS, Errey JC, Bennett K, Deflorian F, Christopher JA, Jazayeri A, Mason JS, Congreve M, Cooke RM, Marshall FH (2016) Intracellular allosteric antagonism of the CCR9 receptor. Nature 540(7633):462–465
94. Pan Y, Tsai CJ, Ma B, Nussinov R (2010) Mechanisms of transcription factor selectivity. Trends Genet: TIG 26(2):75–83
95. Pande VS, Beauchamp K, Bowman GR (2010) Everything you wanted to know about Markov state models but were afraid to ask. Methods 52(1):99–105
96. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of protein allosteric and regulatory sites. Bioinformatics 30(9):1314–1315
97. Peddibhotla S, Hedrick MP, Hershberger P, Maloney PR, Li Y, Milewski M, Gosalia P, Gray W, Mehta A, Sugarman E, Hood B, Suyama E, Nguyen K, Heynen-Genel S, Vasile S, Salaniwal S, Stonich D, Su Y, Mangravita-Novo A, Vicchiarelli M, Roth GP, Smith LH, Chung TD, Hanson GR, Thomas JB, Caron MG, Barak LS, Pinkerton AB (2013) Discovery of ML314, a brain penetrant non-peptidic beta-arrestin biased agonist of the neurotensin NTR1 receptor. ACS Med Chem Lett 4(9):846–851
98. Pomerantz WC, Wang N, Lipinski AK, Wang R, Cierpicki T, Mapp AK (2012) Profiling the dynamic interfaces of fluorinated transcription complexes for ligand discovery and characterization. ACS Chem Biol 7(8):1345–1350
99. Qi Y, Wang Q, Tang B, Lai L (2012) Identifying allosteric binding sites in proteins with a two-state go model for novel allosteric effector discovery. J Chem Theory Comput 8(8):2962–2971
100. Rodriguez D, Gao ZG, Moss SM, Jacobson KA, Carlsson J (2015a) Molecular docking screening using agonist-bound GPCR structures: probing the A2A adenosine receptor. J Chem Inf Model 55(3):550–563
101. Rodriguez D, Ranganathan A, Carlsson J (2015b) Discovery of GPCR ligands by molecular docking screening: novel opportunities provided by crystal structures. Curr Top Med Chem 15(24):2484–2503
102. Rossetti M, Porchetta A (2018) Allosterically regulated DNA-based switches: from design to bioanalytical applications. Anal Chim Acta 1012:30–41
103. Schann S, Mayer S, Franchet C, Frauli M, Steinberg E, Thomas M, Baron L, Neuville P (2010) Chemical switch of a metabotropic glutamate receptor 2 silent allosteric modulator into dual metabotropic glutamate receptor 2/3 negative/positive allosteric modulators. J Med Chem 53(24):8775–8779
104. Schauer GD, Huber KD, Leuba SH, Sluis-Cremer N (2014) Mechanism of allosteric inhibition of HIV-1 reverse transcriptase revealed by single-molecule and ensemble fluorescence. Nucleic Acids Res 42(18):11687–11696
105. Schwantes CR, McGibbon RT, Pande VS (2014) Perspective: Markov models for long-timescale biomolecular dynamics. J Chem Phys 141(9):090901
106. Sevcsik E, Trexler AJ, Dunn JM, Rhoades E (2011) Allostery in a disordered protein: oxidative modifications to alpha-synuclein act distally to regulate membrane binding. J Am Chem Soc 133(18):7152–7158
107. Shen Q, Wang G, Li S, Liu X, Lu S, Chen Z, Song K, Yan J, Geng L, Huang Z, Huang W, Chen G, Zhang J (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res 44(D1):D527–D535

108. Shoichet BK, Kobilka BK (2012) Structure-based drug screening for G-protein-coupled receptors. Trends Pharmacol Sci 33(5):268–272
109. Shukla D, Hernandez CX, Weber JK, Pande VS (2015) Markov state models provide insights into dynamic modulation of protein function. Acc Chem Res 48(2):414–422
110. Soergel DG, Subach RA, Burnham N, Lark MW, James IE, Sadler BM, Skobieranda F, Violin JD, Webster LR (2014) Biased agonism of the mu-opioid receptor by TRV130 increases analgesia and reduces on-target adverse effects versus morphine: a randomized, double-blind, placebo-controlled, crossover study in healthy volunteers. Pain 155(9):1829–1835
111. Srivastava A, Yano J, Hirozane Y, Kefala G, Gruswitz F, Snell G, Lane W, Ivetac A, Aertgeerts K, Nguyen J, Jennings A, Okada K (2014) High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. Nature 513(7516):124–127
112. Strickland D, Moffat K, Sosnick TR (2008) Light-activated DNA binding in a designed allosteric protein. Proc Natl Acad Sci U S A 105(31):10709–10714
113. Subedi GP, Hanson QM, Barb AW (2014) Restricted motion of the conserved immunoglobulin G1 N-glycan is essential for efficient FcgammaRIIIa binding. Structure 22(10):1478–1488
114. Tahtaoui C, Guillier F, Klotz P, Galzi JL, Hibert M, Ilien B (2005) On the use of nonfluorescent dye labeled ligands in FRET-based receptor binding studies. J Med Chem 48 (24):7847–7859
115. Tan Q, Zhu Y, Li J, Chen Z, Han GW, Kufareva I, Li T, Ma L, Fenalti G, Li J, Zhang W, Xie X, Yang H, Jiang H, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu B (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science 341(6152):1387–1390
116. Tewson P, Westenberg M, Zhao Y, Campbell RE, Quinn AM, Hughes TE (2012) Simultaneous detection of Ca2+ and diacylglycerol signaling in living cells. PLoS One 7(8):e42791
117. Tiwary P, Limongelli V, Salvalaglio M, Parrinello M (2015) Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. Proc Natl Acad Sci U S A 112(5):E386–E391
118. Torres PU (2006) Cinacalcet HCl: a novel treatment for secondary hyperparathyroidism caused by chronic kidney disease. J Ren Nutr 16(3):253–258
119. Udi Y, Fragai M, Grossman M, Mitternacht S, Arad-Yellin R, Calderone V, Melikian M, Toccafondi M, Berezovsky IN, Luchinat C, Sagi I (2013) Unraveling hidden regulatory sites in structurally homologous metalloproteases. J Mol Biol 425(13):2330–2346
120. Urs NM, Gee SM, Pack TF, McCorvy JD, Evron T, Snyder JC, Yang X, Rodriguiz RM, Borrelli E, Wetsel WC, Jin J, Roth BL, O'Donnell P, Caron MG (2016) Distinct cortical and striatal actions of a beta-arrestin-biased dopamine D2 receptor ligand reveal unique antipsychotic-like properties. Proc Natl Acad Sci U S A 113(50):E8178–E8186
121. Valant C, May LT, Aurelio L, Chuo CH, White PJ, Baltos JA, Sexton PM, Scammells PJ, Christopoulos A (2014) Separation of on-target efficacy from adverse effects through rational design of a bitopic adenosine receptor agonist. Proc Natl Acad Sci U S A 111(12):4614–4619
122. Vecchio EA, Baltos JA, Nguyen ATN, Christopoulos A, White PJ, May LT (2018) New paradigms in adenosine receptor pharmacology: allostery, oligomerization and biased agonism. Br J Pharmacol 175:4036
123. Vecchio EA, Chuo CH, Baltos JA, Ford L, Scammells PJ, Wang BH, Christopoulos A, White PJ, May LT (2016) The hybrid molecule, VCP746, is a potent adenosine A2B receptor agonist that stimulates anti-fibrotic signalling. Biochem Pharmacol 117:46–56
124. Vulpetti A, Hommel U, Landrum G, Lewis R, Dalvit C (2009) Design and NMR-based screening of LEF, a library of chemical fragments with different local environment of fluorine. J Am Chem Soc 131(36):12949–12959
125. Wass MN, Kelley LA, Sternberg MJ (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res 38(Web Server issue):W469–W473
126. Weber M, Ferrer M, Zheng W, Inglese J, Strulovici B, Kunapuli P (2004) A 1536-well cAMP assay for Gs- and Gi-coupled receptors using enzyme fragmentation complementation. Assay Drug Dev Technol 2(1):39–49

127. Weiss DR, Ahn S, Sassano MF, Kleist A, Zhu X, Strachan R, Roth BL, Lefkowitz RJ, Shoichet BK (2013) Conformation guides molecular efficacy in docking screens of activated beta-2 adrenergic G protein coupled receptor. ACS Chem Biol 8(5):1018–1026
128. Wenthur CJ, Gentry PR, Mathews TP, Lindsley CW (2014) Drugs for allosteric sites on receptors. Annu Rev Pharmacol Toxicol 54:165–184
129. Wootten D, Simms J, Koole C, Woodman OL, Summers RJ, Christopoulos A, Sexton PM (2011) Modulation of the glucagon-like peptide-1 receptor signaling by naturally occurring and synthetic flavonoids. J Pharmacol Exp Ther 336(2):540–550
130. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC (2010) Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. Science 330(6007):1066–1071
131. Yang F, Yu X, Liu C, Qu CX, Gong Z, Liu HD, Li FH, Wang HM, He DF, Yi F, Song C, Tian CL, Xiao KH, Wang JY, Sun JP (2015) Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and ^{19}F-NMR. Nat Commun 6:8202
132. Yu J, Zhou Y, Tanaka I, Yao M (2010) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics 26(1):46–52
133. Zhang D, Gao ZG, Zhang K, Kiselev E, Crane S, Wang J, Paoletta S, Yi C, Ma L, Zhang W, Han GW, Liu H, Cherezov V, Katritch V, Jiang H, Stevens RC, Jacobson KA, Zhao Q, Wu B (2015) Two disparate ligand-binding sites in the human P2Y1 receptor. Nature 520(7547):317–321
134. Zheng Y, Qin L, Zacarias NV, de Vries H, Han GW, Gustavsson M, Dabros M, Zhao C, Cherney RJ, Carter P, Stamos D, Abagyan R, Cherezov V, Stevens RC, AP IJ, Heitman LH, Tebben A, Kufareva I, Handel TM (2016) Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. Nature 540(7633):458–461
135. Zidar DA, Violin JD, Whalen EJ, Lefkowitz RJ (2009) Selective engagement of G protein coupled receptor kinases (GRKs) encodes distinct functions of biased ligands. Proc Natl Acad Sci U S A 106(24):9649–9654

Chapter 2
Dynamic Protein Allosteric Regulation and Disease

Ruth Nussinov, Chung-Jung Tsai, and Hyunbum Jang

Abstract Allostery is largely associated with conformational and functional transitions in individual proteins. All dynamic proteins are allosteric. This concept can be extended to consider the impact of conformational perturbations on cellular function and disease states. In this section, we will illuminate how allostery can control physiological activities and cause disease, aiming to increase the awareness of the linkage between disease symptoms on the cellular level and specific aberrant allosteric actions on the molecular level.

Keywords Cell organization · Cell signaling · Signaling pathways · Signal transduction · Cell structure · Signaling modules · Diffusion

2.1 Introduction

In the cell, a protein can be present in the active or more frequently in the inactive state. Function requires that it populates a distinct conformational state. In 1991, Frauenfelder et al. [28] transformed our views of proteins and broadly biological macromolecules. They described them as existing in *not* only two – active and inactive – states but in an ensemble of states that reflect their relative energy levels which can be mapped on the free energy landscape. This concept profoundly influenced the understanding of protein folding and prompted the development of

R. Nussinov (✉)
Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA

Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
e-mail: NussinoR@mail.nih.gov

C.-J. Tsai · H. Jang
Cancer and Inflammation Program, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD, USA

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_2

numerous experimental and computational methodologies to identify the states, the transitions between them, and the foundational problem of protein folding. The free energy landscape is forceful since it is able to capture the mapping of all possible states, native and nonnative, that the molecule can populate as a function of their corresponding energy levels. The description on a two (or three)-dimensional Cartesian coordinate systems is simple and powerful, since it captures *all* possible physical states of the protein around the native basin. This landscape idea is compelling: it shows that the protein may populate any of these conformations under certain conditions.

However, the picture is a static snapshot of the possible states; it is unable to explain biological function, which is based on cooperativity. In the late 1990s, this inspired us to offer a *dynamic ensemble landscape* concept [65, 83, 147, 148]. In our description, the relative distributions of the conformational states are *dynamic* and depend on the conditions in the cell, which reflect functional cell states following intra- and extra-molecular events. We suggested that the shifts of the population between the states, which are triggered by such events, are the origin of cooperativity, i.e., allostery [37, 65]. Changes in conditions can be reflected by binding events whether non-covalent, such as ligands, lipids, proteins, water, and nucleic acids, or covalent. Changes in covalent linkages include mutational events, posttranslational modifications (PTMs) [107], as well as alterations in temperature, pH, ion concentrations, interactions with lipids such as signaling lipids or the membrane, and even water molecules [148]. Any of these events can shift the relative population of the states, which we termed "population shift." In our description this redistribution of the populations of the states is the allosteric effect. Binding of an allosteric effector promotes a population shift from a highly populated inactive state to an active state, with an observable (or a very subtle) conformational change at the active site [104]. We further proposed that all possible conformations preexist. Allosteric (or non-allosteric) events do not create new conformations; they only shift the populations between the conformations as they get stabilized (or destabilized) [4, 19, 147, 148]. We had a great difficulty in publishing those ideas. At the time they were considered heretic, with the prevailing concept being of only two states, active (bound or closed) and inactive (unbound or open) since those were the structures captured by crystallography and deposited in the PDB. It took us 2 years to finally get them accepted by journals.

Allostery is of paramount importance for biological processes. The unified view of allostery that we formulated recently merges three components: thermodynamics, free energy landscape of population shift, and structural; critically, all are with exactly the same allosteric descriptors [144]. Our unified view innovates by linking these views or elements. The thermodynamic view quantified the binding of the ligand to the active (or inactive) states by employing experimentally measurable quantities. The free energy landscape describes the population shift in terms of energy, which captures the relative stabilization (or destabilization) of the two states. The structural view connects the first two views, thereby capturing the allosteric pathway between the active and allosteric sites on the protein. If there is no population shift, on its own, a preexisting propagation pathway does not imply

allostery; it only indicates coupling between the active and another site on the protein [155]. As we discuss below, allosteric propagation pathways are often sought between allosteric driver mutations and functional sites [15, 24, 93, 136], and driver mutations are sometimes suggested to result in new states, which is *not* the case. They simply increase (or decrease) the relative population of a state, because the mutation changes its relative stability. For most proteins, the inactive state is the more stable. An allosteric mutation may act by destabilizing the inactive state, stabilizing the active state or both [101].

Early on, protein binding was postulated to take place through a "lock-and-key" mechanism. Binding was proposed to involve matching of two rigid shapes, which precisely fit each other similar to a jigsaw puzzle. This was followed by an "induced fit" mechanism where the shape of the protein binding site may not fit perfectly that of its incoming ligand; however, the ligand would induce a conformational change to optimize it. These mechanisms are unable to explain how allosteric driver mutations can promote oncogenic consequences. By contrast, the dynamic free energy landscape, which postulates a shift in the population between states – elicited by a driver mutation where a now populated state can favorably select a ligand – can explain the oncogenic outcome. We dubbed this mechanism "conformational selection and population shift" [65, 83, 84]. In this mechanism new functions do not involve newly created conformations; instead, new functions can involve a shift of the population between preexisting conformations, populating a conformation that was only sparsely populated earlier [102]. Our "conformational selection and population shift" concept posited that because ensembles are vast, encompassing many possible states with varying degrees of energy, binding will take place between the most compatible conformations, followed by a minor conformational change, largely of the side chains, for optimal fit [16]. Since the binding conformations will be removed from solution, the ensemble will shift ("a population shift") in their favor to maintain the equilibrium.

The conformational selection and population shift paradigm which argues that the closed state already exists in solution contrasts the induced fit. The kinetic differences between the two models [154] reflect ligand concentrations [59]: if very high (typically a nonphysiological occurrence), induced fit may prevail. Because usually the complementary conformer has lower population, which will require crossing a barrier to switch between the states, induced fit has faster timescales. The slower timescales make conformational selection more difficult to detect experimentally [4], which explains the often mistaken conclusion that the studied mechanism follows an induced fit process. The timescales reflect a dynamic free energy landscape [65], where substates populate shallow wells and binding alters their relative stabilities [87]. The elegant NMR data of Gardino et al. [33] provided detail how transitions between high energy states can take place. Population shift is the origin of allostery and regulation and can explain how constitutive allosteric driver mutations can rewire signaling in cancer [101, 109, 142].

Allostery can provoke disease in many ways. The most common among these are allosteric driver mutations and pathogen proteins or metabolites. Mutational effects can be direct, or they can involve alterations in patterns of allosteric PTMs, either

through direct substitution or via mutational effects in the enzymes tagging the protein with the PTMs or those cleaving them [132]. From the allosteric standpoint, the mechanisms are unchanged.

Here we discuss dynamic conformational ensembles focusing on allosteric events in signaling and drug discovery. We focus on allosteric driver and discuss also latent driver mutations [105].

2.2 Not All Driver Mutations Are Allosteric and Not All Are Frequent, yet They May Be Associated with Disease

Mutations are classified as drivers or passengers. *Passenger* mutations are unconnected to the disease; driver mutations can elicit it. Driver mutations can be at active or functional sites, in which case they are orthosteric; but they can also be allosteric. Driver mutations are typically detected through their high frequencies of occurrence in sequence analysis. Allosteric driver mutations can be predicted through their activation/inactivation effects. Both can alter signaling pathways. However, not all driver mutations are statistically significant; some are rare, at the tail of the distribution. Notwithstanding, their characteristics and mechanisms are unchanged: some are orthosteric, some allosteric. Detecting rare driver mutations is challenging. One way is via their clustering: if the mutations occur in clusters, chances are higher that they are driver mutations, even if rare [32]. Such clusters were recently identified in *RAC1* and *MAP 2K1*. Proximal associations can suggest cooperativity, similar to the tendency of residue hot spots to cluster into "hot regions" [61, 62]. Clusters can include established frequent drivers such as $KRas^{G12D}$, but often they are rare mutations, such as $KRas^{D33E}$ (defined as <0.1% of the sample) [64], as well as others in KRas (*KRAS*), B-Raf (*BRAF*), and p53 (*TP53*). Most of these rare mutations are likely to be passengers, but some may be drivers. Cluster analysis in protein structures also identified rare drivers in tumor suppressors including *PTEN*, *CDH1*, and *KEAP1*. 15 clusters with 48 residues were detected in *PTEN* tumor suppressor (with 2 known frequent mutations and 46 rare, some of which may be drivers) – all around the phosphatase catalytic core motif [20].

By definition, allosteric driver mutations elicit an observable conformational change, whereas passenger mutations do not. However, we proposed that under certain conditions or combinations, rare presumably passenger mutations may transform into driver mutations. We dubbed these mutations "latent drivers" [103]. The free energy landscape suggests that a binary yes/no driver/passenger classification of mutations into allosteric and non-allosteric may not be accurate [105]. This is important since pharmacology may overlook relevant mutational targets [110]. These silent latent driver mutations may not cause a switch from the inactive to the active state (or vice versa); however, should another cooperative silent mutation emerge during cancer development, their combined effect on the ensemble can be significant, turning on (or off) a signaling pathway [102]. Together they may

work via AND (i.e., additive) or graded logic gate integration mechanisms (but not OR) [7]. Signaling often reflects a combination of events [109], such as binding of several proteins, small molecules, ions, lipids, water molecules, as well as PTMs [37, 108], and with millions of possible genetic variants and cellular events such allosteric effects are at play [26, 75, 79, 80, 114, 126, 143, 145, 157, 163].

Below, we provide examples of proteins that control key cellular functions with allosteric driver mutations that are linked to cancer: protein and lipid kinases and the small GTPase Ras protein.

2.3 Allosteric Driver Mutations: Theory and Examples

2.3.1 Protein Kinases: Introduction

Protein kinases [85] exemplify how allosteric driver mutations shift the ensemble from the inactive to the constitutively active state [142]. Catalysis involves binding of ATP and a substate, transfer of the γ-phosphate of ATP to a hydroxyl group, and release of the phosphorylated substate and ADP. Optimal phosphate transfer dictates the almost identical kinases' active conformations [47, 98], with hydrophobic contacts and electrostatic interactions accurately orienting and coordinating the catalytic residues [135]. However, the inactive conformations vary, which permit distinct kinases to bind to and activate diverse different ligands. The active and inactive states differ in the conformation of the activation loop (in the active conformation, it is often in an extended (or less favorable disordered) conformation, whereas in the inactive state, it is collapsed, folding back to block substate binding), a specific salt bridge (between the β3 lysine and the αC glutamate), and the structurally coupled and organized R-spine [105] which occurs only in the active αC-helix-In, but not in the inactive αC-helix-out kinase state, where the salt bridge is broken and the R-spine distorted. Activation involves a rotation and shift of the αC helix which is controlled by an allosteric switch. Normally, the catalytic domain populates the inactive state [57].

2.3.2 EGFR

Oncogenic driver mutations can shift the ensemble toward the active conformation by stabilizing it, as in the case of Leu858 driver mutation, or destabilizing the inactive state by breaking interactions that stabilize it [101, 125], as in the cases of T790M mutation in EGFR, T315I in Bcr-abl, T334I in c-Abl, T341I in Src, T670I in c-Kit (a.k.a CD117), and T674I in PDGFRα. All allosterically shift the kinase populations toward the activated kinase states. The T790M driver mutation stabilizes the active monomer by increasing the population of the asymmetric dimer. The L858R driver mutation disrupts interactions that stabilize the inactive symmetric

dimer interface versus the asymmetric dimer. However, like T790M driver mutation, it can also stabilize the active state and asymmetric dimer formation by stabilizing the αC-helix-In conformation. Thus, driver mutations can adopt one (or two) of three mechanisms: destabilize the inactive state, stabilize the active state, or both. From the standpoint of the energy landscape, these can be described by the alterations in the relative depths of the minima [13, 21, 23, 48, 56, 70, 88, 121, 124, 125, 152, 158–162, 166].

2.3.3 Raf

Here, we exploit Raf to provide a unified scheme for kinase allosteric activation under physiological conditions and in cancer [146]. We emphasize that the population shift from the inactive to the active state is relative and that allosteric events alter the relative populations additively. We further indicate that the structural features of the active conformation are coupled with the regulatory and catalytic spines for accurately positioning the catalytic sequence motifs. These illuminate the necessity for Raf dimerization and clarify how the V600E driver mutation activates Raf.

Raf's inactive autoinhibited state is an ensemble of "closed" monomeric states [112], with the N-terminal tail, Ras-binding domain (RBD), cysteine-rich domain (CRD), and linker, with its Ser/Thr-rich segment, preventing kinase domain dimerization [8, 12, 17, 30, 40, 50, 51, 68, 92, 123, 130, 131, 137, 139, 140, 164, 165]. The oncogenic V600E mutation of B-Raf weakens the autoinhibition [151].

In response to external stimuli, the Ras/Raf/MEK/ERK signaling pathway initiates physiological processes including growth and proliferation. Under physiological conditions, via a series of events, the activated receptor stimulates conversion of Ras-GDP to Ras-GTP, with subsequent dimerization, and nanoclustering [10, 55, 79, 94, 96, 146]. Coupled with the high binding affinity of Ras-GTP to Raf's RBD, the spatial proximity of Ras dimers/nanoclusters [167] increases the effective local concentration of Raf, promoting Raf activation via a side-by-side (or face-to-face) symmetric Raf dimerization [2, 111]. The increase of the effective local concentration by Ras dimers/nanoclusters and the high Ras-Raf affinity is critical to Raf's activation in the cell, since the weak binding affinity of Raf-Raf interactions [69] limits full-length Raf dimerization under normal Raf expression levels. The recruitment to the membrane through active Ras interaction, which is further strengthened by its CRD attachment to the membrane [72, 73, 141], effectively reduces Raf's distribution space from three-dimensional to two-dimensional. In this scenario Raf activation depends on activated Ras oligomer, in contrast to the Ras-independent scenario, which may take place in oncogenic Raf activation.

Above we described our global picture of "how allostery works" [103, 144]. Here we describe how allostery can work in activation, illustrating it with population shift in the Raf kinase through structural features that destabilize the inactive state and/or stabilize the active state to explain how oncogenic drivers activate Raf [101, 143]. Even though not discussed here, paradoxical Raf activation by asymmetric Raf dimerization by an inhibitor also fits this description.

Monomeric Raf [138] is populated in the inactive state. Physiological activation involves dimerization as a homodimer or as a Raf family heterodimer [29, 42, 52, 53, 95, 120, 151]. Dimerized active Raf monomers can catalyze transphosphorylation of their partner's activation loop, with binding and phosphorylation [164] being additive allosteric events [22]. How does Raf allosteric asymmetric activation via symmetric dimerization take place? According to our nomenclature, Arg509 and Trp450 are allosteric *driver* residues. Allosteric drivers clash with other atoms/residues of the protein partner (here the backbone of Arg506 and Lys507 in the inactive state). The clash exerts a "pull" and/or "push" action that shifts the receptor population from the inactive to the active state [103]. The rest of the residues in the activator interface are *anchor* residues; that is, they are responsible for docking into an allosteric pocket. The conformation with which they interact is unchanged during the transition between the inactive and active states. Anchor residues provide the foundation that allows the driver to exert its action. These residues are responsible primarily for the stability of the dimer. The side chains of Arg509 and Trp450 *destabilize* the *inactive* monomer by forcing the αC helix into an active state. This clarifies why R509H and W450A abrogate Raf's transactivation [42]. Oncogenic B-Raf harboring the V600E mutation is active as a monomer; thus, it is Ras- and activation loop phosphorylation-independent. It constitutes 90% of *BRAF* oncogenic mutations [18, 151]. Recently we clarified in detail how V600E can switch the B-Raf population toward an active monomeric state [146]. How does paradoxical activation take place? Orthosteric, ATP-competitive inhibitor binds to one monomer in the dimer [39, 41, 117], allosterically reducing its affinity for the other monomer [146].

2.3.4 PI3Kα Lipid Kinase

Allosteric oncogenic mutations in PI3Kα lipid kinase [44, 60] promote activation by relieving the autoinhibition by acting to release the nSH2 domain or by boosting nSH2-release-induced allosteric motions. Two driver mutations in the helical domain, E542K and E545K, which reverse the charges act in this manner. Another, less frequent mutation of Gln546 (to Lys or Arg) in the helical domain, is at the nSH2–helical domain interface does not interact favorably with the nSH2 domain. In the autoinhibited state it is close to nSH2 Lys382 (under 5 Å). The repulsive force created by the mutation facilitates nSH2 release. Two frequent oncogenic mutations (Cys420 and Asn345) in the C2 domain occur at the iSH2-C2 interface. They are often mutated to basic residues (Arg or Lys), which destabilize the iSH2-C2 interface. The disruption of the interface results in a conformational change by nSH2 release, promoting iSH2 rotation and the C-terminal portion of the kinase domain exposure in PI3Kα activation. The rotation is coupled with the interacting ABD domain in the p110α catalytic subunit. There are eight frequent driver mutations in the ABD domain: at the ABD-kinase domain interface and at the ABD-RBD linker. They either break the ABD-kinase domain interface or contacts in the linker, which can also promote iSH2 rotation and thus the C-terminal portion of the kinase domain

exposure. Several driver mutations promote the C-terminal portion of the kinase domain movement away from C2 and the helical domains or modulate the C-terminal portion of the kinase domain dynamics, thereby contributing to the PI3Kα membrane interactions in PI3Kα activation [unpublished data].

2.3.5 Ras

Under normal physiological conditions, Ras proteins are also in the inactive state. They are activated at the membrane [5, 54, 127] by an incoming signal from RTKs via an exchange of GDP by GTP [11, 80, 90]. Their activation promotes tumor proliferation pathways, especially MAPK (via Raf) and PI3K (via PI3k/Akt/mTOR) [26, 43, 94, 97, 113]. They are inactivated by hydrolysis of GTP to GDP, assisted by the GAP proteins. Driver mutations suppress hydrolysis, rendering Ras constitutively active. Conformational analysis revealed that mutations can change the landscape of Ras isoforms [1, 3, 14, 34, 35, 38, 45, 49, 58, 66, 76–78, 82, 110, 115, 119, 122, 133, 134, 143]. Substitutions of Gly12, Gly13, and Gln61 are common and block GAP-catalyzed GTP hydrolysis [9, 27, 63, 81]. Notably, even mutations at functional sites as here or ligand binding at orthosteric sites have allosteric effects. Simulations of wild-type KRas4B-GTP bound to GAP suggested that GAP not only inserts its arginine finger for GTP hydrolysis but also further stabilizes the active state [80] by reducing the Switch II fluctuations, promoting formation of the H-bonds between the GTP γ-phosphate and Gln61. Driver mutations block the coordination of GAP's Arg789 and the catalytic Gln61 and thus GTP hydrolysis. NMR of GppNHp-bound HRas observed two interconverting conformations, inactive and active. Simulation of GTP-bound KRas also witnessed that in solution, GTP-bound KRas4B exists in the inactive and active states. The simulations also detail how distinct oncogenic mutations abolish catalysis by altering the catalytically competent Gln61-Arg789 organization.

Wild-type KRas4B-GDP is autoinhibited by its farnesylated HVR [54]. At the membrane, when GTP-bound, the attachment of the HVR to the catalytic domain is released [36, 54, 55, 89, 129, 156, 167]. The catalytic domain fluctuates, exposing the effector-binding surface. NMR residue chemical shift perturbation analysis and simulations showed that oncogenic driver mutations shift the landscape toward the exposed effector-binding site. Thus, driver mutations can work via diverse mechanisms, to abolish GTP hydrolysis, ease effector binding, and weaken the HVR interactions with the catalytic domain, thus increasing the population of the exposed site, versus the wild-type and importantly promote GDP to GTP exchange through the SOS (a GEF) protein [74].

2.4 Allosteric Effects in the Cellular Network

Above we highlighted allosteric driver mutations in kinases and in Ras in cancer that act by turning the signaling pathway ON constitutively. They can also act by rewiring the network by switching the signaling from one partner to another, when both partners bind at the same shared site. How can proteins bind specifically more than one partner at the same site? Several strategies can be at play [150], including the partners sharing a binding site motif or conserved interactions, as well as different conformations of the receptor preferentially selecting different partners. This may depend on prior allosteric events, relative ligand concentrations, PTMs, and more [4, 16]. This may be especially at play for disordered proteins or regions, which have broad ensembles with the energy levels (or minima on the landscape) not too different and with low barriers between them [149]. The multiple receptor states allow conformational selection events to (somewhat) different ligand surfaces. Prior allosteric events taking place on the receptor can be a key factor in partner ligand selection. An allosteric event at one site may alter the binding site conformation and dynamics, promoting a population shift in turn leading to a conformational change in the receptor binding site, which may then prefer a distinct ligand partner. Membrane-anchored Eph receptor tyrosine kinases and their ephrin ligands provide examples [67, 99, 109, 118]. The consequent different binding events which mirror cell cycle events in an altered environment are a powerful theme in evolution. The cellular network is dynamic and sensitive to its surroundings. Because all conformations preexist, function will always be performed at some level, even though it may be too low for executing the required protein action. An allosteric event, such as a mutation, absence or presence of a PTM, or overexpression, may disrupt normal regulation, altering the conformational distribution with far-reaching (disease) consequences by promoting alternative (parallel) pathways or constitutive activation.

Key to successful drug discovery is accounting for dynamic changes in the cellular environment; such changes and cellular response are largely reflected in signaling. Signal transduction relays extracellular information to the interior of cells. Cell surface membrane-bound receptors are stimulated by ions, small-molecule morphogens, hormones, neurotransmitters, and covalent modification events. This initiates a cascade of intracellular protein interactions which eventually transmit to genome activation (or suppression) resulting in altered cellular expression. Proteins are dynamic, and this is particularly the case for signaling proteins, which frequently contain disordered domains/segments [31, 91]. Here, we described proteins in terms of their dynamically fluctuating conformational ensembles and their distributions; how these distributions change upon structural perturbations, such as those caused by binding of proteins, cofactors, DNA, small molecules or drugs, or changes in the cellular environment; and the consequences of these fundamental phenomena, called allosteric effects, for beneficial drug discovery.

2.5 Principles of Allosteric Drug Action

Drug discovery aims to identify protein targets whose inhibition (or activation) can treat disease. Diseases are complex, and identifying druggable, disease-relevant proteins and accounting for their role in the network is a challenging task. Modeling of the structural proteome can provide the complete static structural network; on the other hand, the free energy landscape complements this description by helping to understand the dynamic changes in the distributions of the substates in vivo and thus the species to be targeted.

Allosteric drugs work through the same mechanism as any allosteric effectors. They bind elsewhere on the protein surface and allosterically change the conformation of the protein binding site. Similar to orthosteric drugs, their mechanisms present potential pitfalls which call for considerations in their design; however, these are distinct from those of their orthosteric counterparts. Different than orthosteric drugs which often cause toxicity through binding to active sites of related proteins and consequently high affinity to the target to permit low dosage is the primary concern, this is not the case for allosteric sites, which are more likely to be unique. Mechanistically, allosteric drugs act by shifting the free energy landscape. Effective allosteric drugs have atoms in good contact with protein atoms that elicit propagation optimally reaching the active site. Affinity is important; however, designs exploring the protein conformational ensemble and the preferred propagation states are expected to selectively and more potently achieve the goal of advantageous, target-only tunable binding [100]. Different from drugs binding at active sites, allosteric drugs allow modulation of responses, thus holding great promise for the future [59].

Several tools have been published for prediction of allosteric sites based on their characterization (e.g., Huang et al. [45, 46, 71, 128]). Recently, we unraveled the structural mechanisms of allosteric drug action [103].

2.6 Conclusions

Proteins can exist in multiple conformations around their native states which can be portrayed by an energy landscape. Their emerging picture would feature individual valleys, which correspond to the conformational substates. All functional substates preexist, and their relative populations will change upon allosteric events, which perturb the structure. The resulting energetic frustration elicits population shifts, where the energy is distributed in the structures and the conformational changes shift the energetics [25, 116], which may be observed by changes in the shapes and properties of the protein active sites.

Allostery reflects conformational and functional transitions in macromolecules. It may emerge through the actions of multiple factors, including small molecules as diverse as the allosteric inhibition of GTP-bound H-Ras protein by a small-molecule

compound carrying a naphthalene ring [86] and not surprisingly even a curved membrane [6]. Conformational perturbations are the essence of function and life [153]; thus, it is expected that they are harnessed for combating dysfunctional diseased states. In this section, we illuminated how allostery can control physiological activities and cause disease, aiming to increase the awareness of the linkage between disease symptoms on the cellular level and specific aberrant allosteric actions on the molecular level. Eventually, all reflect or harness fundamental laws of quantum mechanics and structural chemistry [106].

Acknowledgments This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government. This research was supported (in part) by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

1. Alred EJ, Scheele EG, Berhanu WM, Hansmann UH (2014) Stability of Iowa mutant and wild type Abeta-peptide aggregates. J Chem Phys 141(17):175101
2. Aoki K, Yamada M, Kunida K, Yasuda S, Matsuda M (2011) Processive phosphorylation of ERK MAP kinase in mammalian cells. Proc Natl Acad Sci U S A 108(31):12675–12680
3. Blacklock K, Verkhivker GM (2014) Computational modeling of allosteric regulation in the hsp90 chaperones: a statistical ensemble analysis of protein structure networks and allosteric communications. PLoS Comput Biol 10(6):e1003679
4. Boehr DD, Nussinov R, Wright PE (2009) The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol 5(11):789–796
5. Bos JL, Rehmann H, Wittinghofer A (2007) GEFs and GAPs: critical elements in the control of small G proteins. Cell 129(5):865–877
6. Bozelli JC Jr, Jennings W, Black S, Hou YH, Lameire D, Chatha P, Kimura T, Berno B, Khondker A, Rheinstadter MC, Epand RM (2018) Membrane curvature allosterically regulates the phosphatidylinositol cycle, controlling its rate and acyl-chain composition of its lipid intermediates. J Biol Chem 293(46):17780–17791
7. Bradshaw JM (2010) The Src, Syk, and Tec family kinases: distinct types of molecular switches. Cell Signal 22(8):1175–1184
8. Bruder JT, Heidecker G, Rapp UR (1992) Serum-, TPA-, and Ras-induced expression from Ap-1/Ets-driven promoters requires Raf-1 kinase. Genes Dev 6(4):545–556
9. Buhrman G, Wink G, Mattos C (2007) Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. Structure 15(12):1618–1629
10. Chen M, Peters A, Huang T, Nan X (2016) Ras dimer formation as a new signaling mechanism and potential cancer therapeutic target. Mini Rev Med Chem 16(5):391–403
11. Cherfils J, Zeghouf M (2013) Regulation of small GTPases by GEFs, GAPs, and GDIs. Physiol Rev 93(1):269–309
12. Chong H, Guan KL (2003) Regulation of Raf through phosphorylation and N terminus-C terminus interaction. J Biol Chem 278(38):36269–36276
13. Chung I, Akita R, Vandlen R, Toomre D, Schlessinger J, Mellman I (2010) Spatial control of EGF receptor activation by reversible dimerization on living cells. Nature 464(7289):783–787

14. Clausen R, Ma B, Nussinov R, Shehu A (2015) Mapping the conformation space of wildtype and mutant H-Ras with a memetic, cellular, and multiscale evolutionary algorithm. PLoS Comput Biol 11(9):e1004470
15. Collier G, Ortiz V (2013) Emerging computational approaches for the study of protein allostery. Arch Biochem Biophys 538(1):6–15
16. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35(10):539–546
17. Cutler RE Jr, Stephens RM, Saracino MR, Morrison DK (1998) Autoregulation of the Raf-1 serine/threonine kinase. Proc Natl Acad Sci U S A 95(16):9214–9219
18. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA (2002) Mutations of the BRAF gene in human cancer. Nature 417(6892):949–954
19. del Sol A, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 17(8):1042–1050
20. Di Cristofano A, Pesce B, Cordon-Cardo C, Pandolfi PP (1998) Pten is essential for embryonic development and tumour suppression. Nat Genet 19(4):348–355
21. Dixit A, Yi L, Gowthaman R, Torkamani A, Schork NJ, Verkhivker GM (2009) Sequence and structure signatures of cancer mutation hotspots in protein kinases. PLoS One 4(10):e7485
22. Dodson CA, Bayliss R (2012) Activation of Aurora-A kinase by protein partner binding and phosphorylation are independent and synergistic. J Biol Chem 287(2):1150–1157
23. Endres NF, Engel K, Das R, Kovacs E, Kuriyan J (2011) Regulation of the catalytic activity of the EGF receptor. Curr Opin Struct Biol 21(6):777–784
24. Feher VA, Durrant JD, Van Wart AT, Amaro RE (2014) Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol 25:98–103
25. Ferreiro DU, Komives EA, Wolynes PG (2018) Frustration, function and folding. Curr Opin Struct Biol 48:68–73
26. Fetics SK, Guterres H, Kearney BM, Buhrman G, Ma B, Nussinov R, Mattos C (2015) Allosteric effects of the oncogenic RasQ61L mutant on Raf-RBD. Structure 23(3):505–516
27. Ford B, Hornak V, Kleinman H, Nassar N (2006) Structure of a transient intermediate for GTP hydrolysis by ras. Structure 14(3):427–436
28. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254(5038):1598–1603
29. Freeman AK, Ritt DA, Morrison DK (2013) Effects of Raf dimerization and its inhibition on normal and disease-associated Raf signaling. Mol Cell 49(4):751–758
30. Fukui M, Yamamoto T, Kawai S, Mitsunobu F, Toyoshima K (1987) Molecular cloning and characterization of an activated human c-raf-1 gene. Mol Cell Biol 7(5):1776–1781
31. Fuxreiter M (2018) Towards a stochastic paradigm: from fuzzy ensembles to cellular functions. Molecules 23(11):3008
32. Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, Zhang H, Solit DB, Taylor BS, Schultz N, Sander C (2017) 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. Genome Med 9(1):4
33. Gardino AK, Villali J, Kivenson A, Lei M, Liu CF, Steindel P, Eisenmesser EZ, Labeikovsky W, Wolf-Watz M, Clarkson MW, Kern D (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. Cell 139(6):1109–1118
34. Gkeka P, Evangelidis T, Pavlaki M, Lazani V, Christoforidis S, Agianian B, Cournia Z (2014) Investigating the structure and dynamics of the PIK3CA wild-type and H1047R oncogenic mutant. PLoS Comput Biol 10(10):e1003895

35. Gorfe AA, Grant BJ, McCammon JA (2008) Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins. Structure 16(6):885–896
36. Gulyas G, Radvanszki G, Matuska R, Balla A, Hunyady L, Balla T, Varnai P (2017) Plasma membrane phosphatidylinositol 4-phosphate and 4,5-bisphosphate determine the distribution and function of K-Ras4B but not H-Ras proteins. J Biol Chem 292(46):18862–18877
37. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57(3):433–443
38. Hall BE, Bar-Sagi D, Nassar N (2002) The structural basis for the transition from Ras-GTP to Ras-GDP. Proc Natl Acad Sci U S A 99(19):12138–12142
39. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S, Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS, Malek S (2010) RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. Nature 464(7287):431–435
40. Heidecker G, Huleihel M, Cleveland JL, Kolch W, Beck TW, Lloyd P, Pawson T, Rapp UR (1990) Mutational activation of c-raf-1 and definition of the minimal transforming sequence. Mol Cell Biol 10(6):2503–2512
41. Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, Dhomen N, Hussain J, Reis-Filho JS, Springer CJ, Pritchard C, Marais R (2010) Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. Cell 140(2):209–221
42. Hu J, Stites EC, Yu H, Germino EA, Meharena HS, Stork PJS, Kornev AP, Taylor SS, Shaw AS (2013) Allosteric activation of functionally asymmetric RAF kinase dimers. Cell 154(5):1036–1046
43. Huang L, Hofer F, Martin GS, Kim SH (1998) Structural basis for the interaction of Ras with RalGDS. Nat Struct Biol 5(6):422–426
44. Huang CH, Mandelker D, Gabelli SB, Amzel LM (2008) Insights into the oncogenic effects of PIK3CA mutations from the structure of p110alpha/p85alpha. Cell Cycle 7(9):1151–1156
45. Huang W, Wang G, Shen Q, Liu X, Lu S, Geng L, Huang Z, Zhang J (2015) ASBench: benchmarking sets for allosteric discovery. Bioinformatics 31(15):2598–2600
46. Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, Gao J, Hong Y, Li Q, Ni D, Xu J, Chen G, Zhang J (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. Nucleic Acids Res 46(W1):W451–W458
47. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. Cell 109(3):275–282
48. Hynes NE, MacDonald G (2009) ErbB receptors and signaling pathways in cancer. Curr Opin Cell Biol 21(2):177–184
49. Invernizzi G, Tiberti M, Lambrughi M, Lindorff-Larsen K, Papaleo E (2014) Communication routes in ARID domains between distal residues in helix 5 and the DNA-binding loops. PLoS Comput Biol 10(9):e1003744
50. Ishikawa F, Takaku F, Hayashi K, Nagao M, Sugimura T (1986) Activation of rat c-raf during transfection of hepatocellular carcinoma DNA. Proc Natl Acad Sci U S A 83(10):3209–3212
51. Ishikawa F, Sakai R, Ochiai M, Takaku F, Sugimura T, Nagao M (1988) Identification of a transforming activity suppressing sequence in the c-raf oncogene. Oncogene 3(6):653–658
52. Jambrina PG, Bohuszewicz O, Buchete NV, Kolch W, Rosta E (2014) Molecular mechanisms of asymmetric RAF dimer activation. Biochem Soc Trans 42(4):784–790
53. Jambrina PG, Rauch N, Pilkington R, Rybakova K, Nguyen LK, Kholodenko BN, Buchete NV, Kolch W, Rosta E (2016) Phosphorylation of RAF kinase dimers drives conformational changes that facilitate transactivation. Angew Chem Int Ed Engl 55(3):983–986
54. Jang H, Banerjee A, Chavan TS, Lu S, Zhang J, Gaponenko V, Nussinov R (2016a) The higher level of complexity of K-Ras4B activation at the membrane. FASEB J 30(4):1643–1655

55. Jang H, Muratcioglu S, Gursoy A, Keskin O, Nussinov R (2016b) Membrane-associated Ras dimers are isoform-specific: K-Ras dimers differ from H-Ras dimers. Biochem J 473(12):1719–1732
56. Jura N, Endres NF, Engel K, Deindl S, Das R, Lamers MH, Wemmer DE, Zhang X, Kuriyan J (2009) Mechanism for activation of the EGF receptor catalytic domain by the juxtamembrane segment. Cell 137(7):1293–1307
57. Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J (2011) Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. Mol Cell 42(1):9–22
58. Kapoor A, Travesset A (2015) Differential dynamics of RAS isoforms in GDP- and GTP-bound states. Proteins 83(6):1091–1106
59. Kar G, Keskin O, Gursoy A, Nussinov R (2010) Allostery and population shift in drug discovery. Curr Opin Pharmacol 10(6):715–722
60. Karakas B, Bachman KE, Park BH (2006) Mutation of the PIK3CA oncogene in human cancers. Br J Cancer 94(4):455–459
61. Keskin O, Ma B, Nussinov R (2005) Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. J Mol Biol 345(5):1281–1294
62. Keskin O, Gursoy A, Ma B, Nussinov R (2008) Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev 108(4):1225–1244
63. Khrenova MG, Mironov VA, Grigorenko BL, Nemukhin AV (2014) Modeling the role of G12V and G13V Ras mutations in the Ras-GAP-catalyzed hydrolysis reaction of guanosine triphosphate. Biochemistry 53(45):7093–7099
64. Kim E, Ilic N, Shrestha Y, Zou L, Kamburov A, Zhu C, Yang X, Lubonja R, Tran N, Nguyen C, Lawrence MS, Piccioni F, Bagul M, Doench JG, Chouinard CR, Wu X, Hogstrom L, Natoli T, Tamayo P, Horn H, Corsello SM, Lage K, Root DE, Subramanian A, Golub TR, Getz G, Boehm JS, Hahn WC (2016) Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. Cancer Discov 6(7):714–726
65. Kumar S, Ma B, Tsai CJ, Sinha N, Nussinov R (2000) Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 9(1):10–19
66. Kumar A, Glembo TJ, Ozkan SB (2015) The role of conformational dynamics and allostery in the disease development of human ferritin. Biophys J 109(6):1273–1281
67. Lackmann M, Boyd AW (2008) Eph, a protein family coming of age: more confusion, insight, or complexity? Sci Signal 1(15):re2
68. Lavoie H, Therrien M (2015) Regulation of RAF protein kinases in ERK signalling. Nat Rev Mol Cell Biol 16(5):281–298
69. Lavoie H, Thevakumaran N, Gavory G, Li JJ, Padeganeh A, Guiral S, Duchaine J, Mao DY, Bouvier M, Sicheri F, Therrien M (2013) Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization. Nat Chem Biol 9(7):428–436
70. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141(7):1117–1134
71. Li S, Shen Q, Su M, Liu X, Lu S, Chen Z, Wang R, Zhang J (2016) Alloscore: a method for predicting allosteric ligand-protein interactions. Bioinformatics 32(10):1574–1576
72. Li S, Jang H, Zhang J, Nussinov R (2018a) Raf-1 cysteine-rich domain increases the affinity of K-Ras/Raf at the membrane, promoting MAPK signaling. Structure 26(3):513–525. e512
73. Li ZL, Prakash P, Buck M (2018b) A “Tug of War” maintains a dynamic protein-membrane complex: molecular dynamics simulations of C-Raf RBD-CRD bound to K-Ras4B at an anionic membrane. ACS Cent Sci 4(2):298–305
74. Liao TJ, Jang H, Fushman D, Nussinov R (2018) Allosteric KRas4B can modulate SOS1 fast and slow ras activation cycles. Biophys J 115(4):629–641
75. Liu J, Nussinov R (2008) Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design. Proc Natl Acad Sci U S A 105(3):901–906

76. Liu JY, Chen XE, Zhang YL (2015) Insights into the key interactions between human protein phosphatase 5 and cantharidin using molecular dynamics and site-directed mutagenesis bioassays. Sci Rep 5:12359
77. Lu S, Li S, Zhang J (2014) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34(6):1242–1285
78. Lu S, Deng R, Jiang H, Song H, Li S, Shen Q, Huang W, Nussinov R, Yu J, Zhang J (2015) The mechanism of ATP-dependent allosteric protection of Akt kinase phosphorylation. Structure 23(9):1725–1734
79. Lu S, Jang H, Muratcioglu S, Gursoy A, Keskin O, Nussinov R, Zhang J (2016a) Ras conformational ensembles, allostery, and signaling. Chem Rev 116(11):6607–6665
80. Lu S, Jang H, Nussinov R, Zhang J (2016b) The structural basis of oncogenic mutations G12, G13 and Q61 in small GTPase K-Ras4B. Sci Rep 6:21949
81. Lukman S, Grant BJ, Gorfe AA, Grant GH, McCammon JA (2010) The distinct conformational dynamics of K-Ras and H-Ras A59G. PLoS Comput Biol 6(9):e1000922
82. Ma J, Karplus M (1997) Molecular switch in signal transduction: reaction paths of the conformational changes in ras p21. Proc Natl Acad Sci U S A 94(22):11905–11910
83. Ma B, Kumar S, Tsai CJ, Nussinov R (1999) Folding funnels and binding mechanisms. Protein Eng 12(9):713–720
84. Ma B, Shatsky M, Wolfson HJ, Nussinov R (2002) Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 11(2):184–197
85. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934
86. Matsumoto S, Hiraga T, Hayashi Y, Yoshikawa Y, Tsuda C, Araki M, Neya M, Shima F, Kataoka T (2018) Molecular basis for allosteric inhibition of GTP-bound H-Ras protein by a small-molecule compound carrying a naphthalene ring. Biochemistry 57(36):5350–5358
87. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP (2009) Quantifying correlations between allosteric sites in thermodynamic ensembles. J Chem Theory Comput 5(9):2486–2502
88. Mi LZ, Lu C, Li Z, Nishida N, Walz T, Springer TA (2011) Simultaneous visualization of the extracellular and cytoplasmic domains of the epidermal growth factor receptor. Nat Struct Mol Biol 18(9):984–989
89. Michael JV, Goldfinger LE (2017) Concepts and advances in cancer therapeutic vulnerabilities in RAS membrane targeting. Semin Cancer Biol 54:121–130
90. Milburn MV, Tong L, deVos AM, Brunger A, Yamaizumi Z, Nishimura S, Kim SH (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. Science 247(4945):939–945
91. Miskei M, Gregus A, Sharma R, Duro N, Zsolyomi F, Fuxreiter M (2017) Fuzziness enables context dependence of protein interactions. FEBS Lett 591(17):2682–2695
92. Molders H, Defesche J, Muller D, Bonner TI, Rapp UR, Muller R (1985) Integration of transfected LTR sequences into the c-raf proto-oncogene: activation by promoter insertion. EMBO J 4(3):693–698
93. Morra G, Verkhivker G, Colombo G (2009) Modeling signal propagation mechanisms and ligand-based conformational dynamics of the Hsp90 molecular chaperone full-length dimer. PLoS Comput Biol 5(3):e1000323
94. Muratcioglu S, Chavan TS, Freed BC, Jang H, Khavrutskii L, Freed RN, Dyba MA, Stefanisko K, Tarasov SG, Gursoy A, Keskin O, Tarasova NI, Gaponenko V, Nussinov R (2015) GTP-dependent K-Ras dimerization. Structure 23(7):1325–1335
95. Nan X, Collisson EA, Lewis S, Huang J, Tamguney TM, Liphardt JT, McCormick F, Gray JW, Chu S (2013) Single-molecule superresolution imaging allows quantitative analysis of RAF multimer formation and signaling. Proc Natl Acad Sci U S A 110(46):18519–18524
96. Nan X, Tamguney TM, Collisson EA, Lin LJ, Pitt C, Galeas J, Lewis S, Gray JW, McCormick F, Chu S (2015) Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway. Proc Natl Acad Sci U S A 112(26):7996–8001

97. Nassar N, Horn G, Herrmann C, Block C, Janknecht R, Wittinghofer A (1996) Ras/Rap effector specificity determined by charge reversal. Nat Struct Biol 3(8):723–729
98. Nolen B, Taylor S, Ghosh G (2004) Regulation of protein kinases; controlling activity through activation segment conformation. Mol Cell 15(5):661–675
99. Nussinov R, Ma B (2012) Protein dynamics and conformational selection in bidirectional signal transduction. BMC Biol 10:2
100. Nussinov R, Tsai CJ (2012) The different ways through which specificity works in orthosteric and allosteric drugs. Curr Pharm Des 18(9):1311–1316
101. Nussinov R, Tsai CJ (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305
102. Nussinov R, Tsai CJ (2014a) Free energy diagrams for protein function. Chem Biol 21(3):311–318
103. Nussinov R, Tsai CJ (2014b) Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol Sci 35(5):256–264
104. Nussinov R, Tsai CJ (2015a) Allostery without a conformational change? Revisiting the paradigm. Curr Opin Struct Biol 30:17–24
105. Nussinov R, Tsai CJ (2015b) 'Latent drivers' expand the cancer mutational landscape. Curr Opin Struct Biol 32:25–32
106. Nussinov R, Wolynes PG (2014) A second molecular biology revolution? The energy landscapes of biomolecular function. Phys Chem Chem Phys 16(14):6321–6322
107. Nussinov R, Tsai CJ, Xin F, Radivojac P (2012) Allosteric post-translational modification codes. Trends Biochem Sci 37(10):447–455
108. Nussinov R, Ma B, Tsai CJ, Csermely P (2013a) Allosteric conformational barcodes direct signaling in the cell. Structure 21(9):1509–1521
109. Nussinov R, Tsai CJ, Ma B (2013b) The underappreciated role of allostery in the cellular network. Annu Rev Biophys 42:169–189
110. Nussinov R, Jang H, Tsai CJ (2014) The structural basis for cancer treatment decisions. Oncotarget 5(17):7285–7302
111. Nussinov R, Tsai CJ, Jang H (2018a) Oncogenic Ras Isoforms signaling specificity at the membrane. Cancer Res 78(3):593–602
112. Nussinov R, Zhang M, Tsai CJ, Liao TJ, Fushman D, Jang H (2018b) Autoinhibition in Ras effectors Raf, PI3Kalpha, and RASSF5: a comprehensive review underscoring the challenges in pharmacological intervention. Biophys Rev 10(5):1263–1282
113. Pacold ME, Suire S, Perisic O, Lara-Gonzalez S, Davis CT, Walker EH, Hawkins PT, Stephens L, Eccleston JF, Williams RL (2000) Crystal structure and functional analysis of Ras binding to its effector phosphoinositide 3-kinase gamma. Cell 103(6):931–943
114. Park MJ, Shen H, Spaeth JM, Tolvanen JH, Failor C, Knudtson JF, McLaughlin J, Halder SK, Yang Q, Bulun SE, Al-Hendy A, Schenken RS, Aaltonen LA, Boyer TG (2018) Oncogenic exon 2 mutations in Mediator subunit MED12 disrupt allosteric activation of cyclin C-CDK8/19. J Biol Chem 293(13):4870–4882
115. Parker JA, Mattos C (2018) The K-Ras, N-Ras, and H-Ras isoforms: unique conformational preferences and implications for targeting oncogenic mutants. Cold Spring Harb Perspect Med 8(8):a031427
116. Parra RG, Schafer NP, Radusky LG, Tsai MY, Guzovsky AB, Wolynes PG, Ferreiro DU (2016) Protein frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. Nucleic Acids Res 44(W1):W356–W360
117. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N (2010) RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature 464(7287):427–430
118. Qin H, Lim L, Song J (2012) Protein dynamics at Eph receptor-ligand interfaces as revealed by crystallography, NMR and MD simulations. BMC Biophys 5:2
119. Raimondi F, Portella G, Orozco M, Fanelli F (2011) Nucleotide binding switches the information flow in ras GTPases. PLoS Comput Biol 7(3):e1001098
120. Rajakulendran T, Sahmi M, Lefrancois M, Sicheri F, Therrien M (2009) A dimerization-dependent mechanism drives RAF catalytic activation. Nature 461(7263):542–545

121. Red Brewer M, Choi SH, Alvarado D, Moravcevic K, Pozzi A, Lemmon MA, Carpenter G (2009) The juxtamembrane region of the EGF receptor functions as an activation domain. Mol Cell 34(6):641–651
122. Scarabelli G, Grant BJ (2014) Kinesin-5 allosteric inhibitors uncouple the dynamics of nucleotide, microtubule, and neck-linker binding sites. Biophys J 107(9):2204–2213
123. Schultz AM, Copeland T, Oroszlan S, Rapp UR (1988) Identification and characterization of c-raf phosphoproteins in transformed murine cells. Oncogene 2(2):187–193
124. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, Jensen MO, Dror RO, Kuriyan J, Shaw DE (2009) A conserved protonation-dependent switch controls drug binding in the Abl kinase. Proc Natl Acad Sci U S A 106(1):139–144
125. Shan Y, Eastwood MP, Zhang X, Kim ET, Arkhipov A, Dror RO, Jumper J, Kuriyan J, Shaw DE (2012) Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. Cell 149(4):860–870
126. Shen Q, Cheng F, Song H, Lu W, Zhao J, An X, Liu M, Chen G, Zhao Z, Zhang J (2017) Proteome-scale investigation of protein allosteric regulation perturbed by somatic mutations in 7,000 cancer genomes. Am J Hum Genet 100(1):5–20
127. Sondermann H, Soisson SM, Boykevisch S, Yang SS, Bar-Sagi D, Kuriyan J (2004) Structural analysis of autoinhibition in the Ras activator Son of sevenless. Cell 119(3):393–405
128. Song K, Liu X, Huang W, Lu S, Shen Q, Zhang L, Zhang J (2017) Improved method for the identification and validation of allosteric sites. J Chem Inf Model 57(9):2358–2363
129. Sperlich B, Kapoor S, Waldmann H, Winter R, Weise K (2016) Regulation of K-Ras4B membrane binding by calmodulin. Biophys J 111(1):113–122
130. Stanton VP Jr, Cooper GM (1987) Activation of human raf transforming genes by deletion of normal amino-terminal coding sequences. Mol Cell Biol 7(3):1171–1179
131. Stanton VP Jr, Nichols DW, Laudano AP, Cooper GM (1989) Definition of the human raf amino-terminal regulatory region by deletion mutagenesis. Mol Cell Biol 9(2):639–647
132. Steklov M, Pandolfi S, Baietti MF, Batiuk A, Carai P, Najm P, Zhang M, Jang H, Renzi F, Cai Y, Abbasi Asbagh L, Pastor T, De Troyer M, Simicek M, Radaelli E, Brems H, Legius E, Tavernier J, Gevaert K, Impens F, Messiaen L, Nussinov R, Heymans S, Eyckerman S, Sablina AA (2018) Mutations in LZTR1 drive human disease by dysregulating RAS ubiquitination. Science 362(6419):1177–1182
133. Sun H, Li Y, Tian S, Wang J, Hou T (2014) P-loop conformation governed crizotinib resistance in G2032R-mutated ROS1 tyrosine kinase: clues from free energy landscape. PLoS Comput Biol 10(7):e1003729
134. Sun H, Tian S, Zhou S, Li Y, Li D, Xu L, Shen M, Pan P, Hou T (2015) Revealing the favorable dissociation pathway of type II kinase inhibitors via enhanced sampling simulations and two-end-state calculations. Sci Rep 5:8457
135. Taylor SS, Kornev AP (2011) Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci 36(2):65–77
136. Tehver R, Chen J, Thirumalai D (2009) Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. J Mol Biol 387(2):390–406
137. Terrell EM, Morrison DK (2018) Ras-mediated activation of the Raf family kinases. Cold Spring Harb Perspect Med 9:a033746
138. Thevakumaran N, Lavoie H, Critton DA, Tebben A, Marinier A, Sicheri F, Therrien M (2015) Crystal structure of a BRAF kinase domain monomer explains basis for allosteric regulation. Nat Struct Mol Biol 22(1):37–43
139. Tran NH, Frost JA (2003) Phosphorylation of Raf-1 by p21-activated kinase 1 and Src regulates Raf-1 autoinhibition. J Biol Chem 278(13):11221–11226
140. Tran NH, Wu X, Frost JA (2005) B-Raf and Raf-1 are regulated by distinct autoregulatory mechanisms. J Biol Chem 280(16):16244–16253
141. Travers T, Lopez CA, Van QN, Neale C, Tonelli M, Stephen AG, Gnanakaran S (2018) Molecular recognition of RAS/RAF complex at the membrane: role of RAF cysteine-rich domain. Sci Rep 8(1):8461

142. Tsai CJ, Nussinov R (2013) The molecular basis of targeting protein kinases in cancer therapeutics. Semin Cancer Biol 23(4):235–242
143. Tsai CJ, Nussinov R (2014a) The free energy landscape in translational science: how can somatic mutations result in constitutive oncogenic activation? Phys Chem Chem Phys 16(14):6332–6341
144. Tsai CJ, Nussinov R (2014b) A unified view of "how allostery works". PLoS Comput Biol 10(2):e1003394
145. Tsai CJ, Nussinov R (2017) Allostery modulates the beat rate of a cardiac pacemaker. J Biol Chem 292(15):6429–6430
146. Tsai CJ, Nussinov R (2018) Allosteric activation of RAF in the MAPK signaling pathway. Curr Opin Struct Biol 53:100–106
147. Tsai CJ, Kumar S, Ma B, Nussinov R (1999a) Folding funnels, binding funnels, and protein function. Protein Sci 8(6):1181–1190
148. Tsai CJ, Ma B, Nussinov R (1999b) Folding and binding cascades: shifts in energy landscapes. Proc Natl Acad Sci U S A 96(18):9970–9972
149. Tsai CJ, Ma B, Sham YY, Kumar S, Nussinov R (2001) Structured disorder and conformational selection. Proteins 44(4):418–427
150. Tsai CJ, Ma B, Nussinov R (2009) Protein-protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci 34(12):594–600
151. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R, Cancer Genome P (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 116(6):855–867
152. Wang Z, Longo PA, Tarrant MK, Kim K, Head S, Leahy DJ, Cole PA (2011) Mechanistic insights into the activation of oncogenic forms of EGF receptor. Nat Struct Mol Biol 18(12):1388–1393
153. Wei G, Xi W, Nussinov R, Ma B (2016) Protein ensembles: how does nature harness thermodynamic fluctuations for life? The Diverse functional roles of conformational ensembles in the cell. Chem Rev 116(11):6516–6551
154. Weikl TR, von Deuster C (2009) Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. Proteins 75(1):104–110
155. Weinkam P, Pons J, Sali A (2012) Structure-based model of allostery predicts coupling between distant sites. Proc Natl Acad Sci U S A 109(13):4875–4880
156. Wright LP, Philips MR (2006) Thematic review series: lipid posttranslational modifications. CAAX modification and membrane targeting of Ras. J Lipid Res 47(5):883–891
157. Xu Q, Tang Q, Katsonis P, Lichtarge O, Jones D, Bovo S, Babbi G, Martelli PL, Casadio R, Lee GR, Seok C, Fenton AW, Dunbrack RL Jr (2017) Benchmarking predictions of allostery in liver pyruvate kinase in CAGI4. Hum Mutat 38(9):1123–1131
158. Yarden Y, Schlessinger J (1987a) Epidermal growth factor induces rapid, reversible aggregation of the purified epidermal growth factor receptor. Biochemistry 26(5):1443–1451
159. Yarden Y, Schlessinger J (1987b) Self-phosphorylation of epidermal growth factor receptor: evidence for a model of intermolecular allosteric activation. Biochemistry 26(5):1434–1442
160. Yarden Y, Sliwkowski MX (2001) Untangling the ErbB signalling network. Nat Rev Mol Cell Biol 2(2):127–137
161. Yun CH, Boggon TJ, Li Y, Woo MS, Greulich H, Meyerson M, Eck MJ (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell 11(3):217–227
162. Yun CH, Mengwasser KE, Toms AV, Woo MS, Greulich H, Wong KK, Meyerson M, Eck MJ (2008) The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. Proc Natl Acad Sci U S A 105(6):2070–2075
163. Zhan C, Qi R, Wei G, Guven-Maiorov E, Nussinov R, Ma B (2016) Conformational dynamics of cancer-associated MyD88-TIR domain mutant L252P (L265P) allosterically tilts the landscape toward homo-dimerization. Protein Eng Des Sel 29(9):347–354

164. Zhang BH, Guan KL (2000) Activation of B-Raf kinase requires phosphorylation of the conserved residues Thr598 and Ser601. EMBO J 19(20):5429–5439
165. Zhang BH, Guan KL (2001) Regulation of the Raf kinase by phosphorylation. Exp Lung Res 27(3):269–295
166. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125(6):1137–1149
167. Zhou Y, Prakash P, Liang H, Cho KJ, Gorfe AA, Hancock JF (2017) Lipid-sorting specificity encoded in K-Ras membrane anchor regulates signal output. Cell 168(1–2):239–251. e216

Chapter 3
Protein Allostery in Rational Drug Design

Takayoshi Kinoshita

Abstract This chapter focuses on protein kinases that transfer the phosphate group of ATP to the hydroxyl group of a substrate protein. Five hundred eighteen human protein kinases are classified into serine/threonine kinases and tyrosine kinases and individually or synergistically transduce physiologic stimuli into cell to promote cell proliferation or apoptosis, etc. Protein kinases are identified as drug targets because dysfunction of kinases leads to severe diseases such as cancers and autoimmune diseases. A large number of the crystal structures of the protein kinase inhibitor complex are available in Protein Data Bank and facilitated the drug discovery targeting protein kinases. The protein kinase inhibitors are classified into categories, Type-I, Type-II, Type-III, Type-IV, and Type-V, and as a separate class, covalent-type inhibitors. In any type, a protein kinase inhibitor bound to the allosteric region is advantageous in terms of selectivity compared to the traditional ATP-competitive one. In the following sections, the successful and promising examples of the partially or fully allosteric protein kinase inhibitors are illustrated in the following pages.

Keywords Protein kinase · Allosteric inhibitor · Structure-based drug discovery · High selectivity

3.1 Introduction

Protein kinases (PKs) catalyze the γ-phosphate group transfer from ATP to the hydroxyl group of substrate proteins requiring magnesium ions. Five hundred eighteen protein kinases had been identified in human genome [23] and divided into tyrosine kinases (TK), serine/threonine kinases (STK), and pseudo-kinases. PK transduces the cellular signals started with the extracellular stimulations to regulate the complicated physiological functions involving cell differentiation, proliferation,

T. Kinoshita (✉)
Graduate School of Science, Osaka Prefecture University, Sakai, Osaka, Japan
e-mail: kinotk@b.s.osakafu-u.ac.jp

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_3

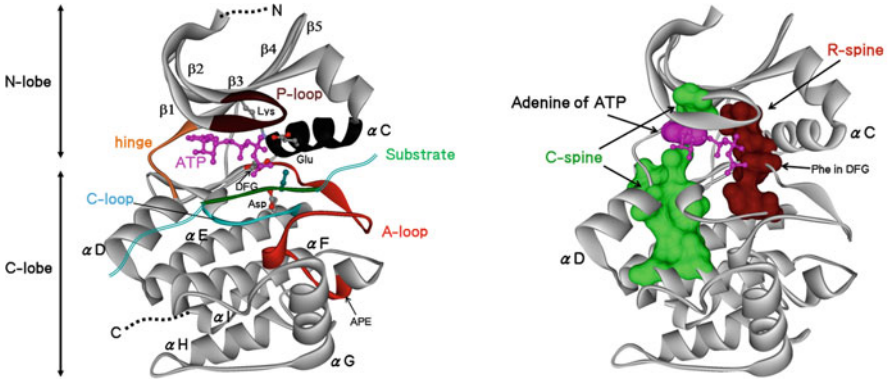


Fig. 3.1 A typical protein kinase fold. (**a**) The numbering and alphabeting of the secondary structure and the nomenclatures of the sub-domains are shown on the protein kinase A. (**b**) Hydrophobic intramolecular spines. The regulatory spine (R-spine, brown) and catalytic spine (C-spine, green) consisting of the hydrophobic amino acid residues stabilize the protein kinase fold. The phenylalanine residue in the DFG motif is jointed to the R-spine

and apoptosis. Therefore, collapse of the vital valance sophisticatedly regulated by PKs results in the serious diseases such as cancers. Approximately 30 kinase inhibitors including imatinib and gefitinib have been approved as molecular target drugs by the Food and Drug Administration (FDA) of the USA to date [43]. Modern scientific technologies allow to discover high-potency inhibitors for a target protein kinase associated with the diseases. However, it remains a serious challenge to develop inhibitors to be drugs. Promiscuous kinase inhibitors would suppress a variety of the off-target protein kinases as well as the target kinase and exert the adverse effects on vital. X-ray crystal analyses of the premature inhibitor with the off-target PKs as well as with the target kinase would promote drug discovery process.

PK involves a wide variety of combinations of the functional domains or subunits. About 200 kinds of human protein kinase domain structures are available from the Protein Data Bank (PDB). These crystal structures reveal that the catalytic domain of PK configures a typical kinase fold consisting of an N-lobe involving a α-helix (αC-helix) and five β-strands, a C-lobe with high helicity, and a hinge region connecting these lobes (Fig. 3.1a). The numbering and alphabeting of the β-strands and α-helices have been defined upon the structure of the protein kinase A (Fig. 3.1a) [11]. The hinge region composed of a single strand contributes to the plasticity of the relative positions of the N- and C-lobes and the recognition of the adenine ring of ATP. The flexible regions such as P-loop, A-loop, N- and C-terminals, and insertion loop singly or collaboratively regulate the protein kinase activity, concerted with the regulatory subunit (or protein) binding or detaching and/or chemical modifications such as phosphorylation. The phosphate binding loop (P-loop) with a glycine-rich sequence in the N-lobe binds ATP or ATP-competitive inhibitors. The activation loop (A-loop), which is bracketed by the highly conserved DFG (Asp-Phe-Gly) and

APE (Ala-Pro-Glu) motifs, functions as a molecular switch via phosphorylation by the client kinase or itself (autophosphorylation). The A-loop works as a platform for substrate binding in the active state and as a molecular brake in the inactive state. The aspartate residue in the DFG motif binds a magnesium ion that is essential for enzyme activity. Activation by promoting factors such as phosphorylation provokes the configuration changes in the N-lobe position relative to the C-lobe, A-loop, and αC-helix. This structural transition results in the salt-bridge formation between the catalytic lysine residue in the β3-strand and glutamate residue in the αC-helix, which can accommodate the ATP molecule in the transition-state conformation. The active conformation is structurally stabilized by two intramolecular spines, regulatory and catalytic spines (R-spine and C-spine), which consist of the hydrophobic residues, respectively (Fig. 3.1b). The C-spine involves the adenine ring of the ATP molecule, and the R-spine does the benzene ring in the DFG motif (Fig. 3.1b). In case that these structural requirements are prepared, the protein kinase is ready for the catalytic reaction. The catalytic loop (C-loop) contains the catalytic aspartate residue which eliminates the proton from the hydroxyl group in the substrate. Subsequently, the activated hydroxyl group of the substrate makes a nucleophilic attack to the phosphorous atom in the γ-phosphate group of ATP. The N- and C-terminal regions and/or insertion loop represents remarkable diversity in amino acid sequence and size and allosterically regulates the protein kinase activity involving the autoinhibition, substrate recognition, and subcellular localization. The structural clarification of the allosteric molecular switches eventually provides the valuable clues for producing highly selective protein kinase inhibitors.

3.2 Classification of Protein Kinase Inhibitors

Approximately 3000 crystal structures of the protein kinase-inhibitor complex have been registered in the Protein Data Bank (PDB) and facilitated the anti-protein kinase drug discovery. Kinase inhibitors are classified into categories Type-I, Type-II, Type-III, Type-IV, and Type-V and, as a separate class, covalent-type inhibitors as shown in Table. 3.1 [24]. Type-I inhibitor binds mainly to the ATP binding region, which is located at the hydrophobic slot between the N- and C-lobes. The structural requirements regarding ATP binding in this region are well-conserved

Table 3.1 Classification of kinase inhibitors

Type	Binding region	Approved drug
Type-I	ATP site	Gefitinib, erlotinib
Type-II	ATP site + DFG-out region	Imatinib, nilotinib
Type-III	DFG-out or near-allosteric region	Trametinib
Type-IV	Far-allosteric region	–
Type-V	ATP and substrate sites	–
Covalent	ATP site	Afatinib, ibrutinib

among protein kinases. The main chain atoms in the hinge region make the hydrogen bonds with the adenine moiety of ATP. The hydrophobic residues participated in the R-spine bracket the adenine ring of ATP (Fig. 3.1b). Thus, a number of Type-I inhibitors involve the hetero-aromatic ring as a substitute for the adenine moiety. The lysine and magnesium ion ligated at the aspartate residue in the DFG motif, which are conserved in the ATP binding site, largely contribute to recognize the phosphate groups. The hydrogen bonding with these lysine and aspartate residues is essential for the potential inhibitors. Type-II inhibitor binds to the so-called DFG-out conformation, which is caused by the flipping motion of the DFG motif in the N-terminal end of the A-loop. This motion results in the emergence of the hydrophobic pocket, referred to as DFG-out region, a poorly conserved region among protein kinases. Type-II inhibitors bind to the DFG-out region as well as the ATP binding site and thus present the rather high selectivity against the off-target kinases. Type-III inhibitor binds to the near-allosteric region such as the DFG-out region but not to the ATP binding region. Type-IV inhibitor binds to the far-allosteric region, distant to the ATP site, involving the substrate binding or autoregulatory regions. Type-V inhibitor is composed of the ATP analogue and peptide derived from substrate protein and covers almost all of the enzyme reaction sites [31]. However, Type-V inhibitor is apt to display the low inhibitory activity because the peptide moiety causes the entropy loss due to its conformational flexibility. Further, the peptide part contains not less than ten residues by which the molecular weight of the Type-V inhibitors greatly exceeds to 1000 daltons. The peptide sequence around the phosphorylation site configures the extended formation and forms the β-sheet structure with the A-loop. Namely, protein kinases request the rather low sequence specificity for the substrate recognition. These obstacles likely put a damper on the development of the Type-V inhibitors. Recently, it had reported that the covalent-type inhibitor was effective for a kind of protein kinases that possessed the cysteine residues around the ATP binding site [22]. The cysteine residues are observed in the hinge region, gatekeeper, P-loop, A-loop, and the other allosteric sites in protein kinases, but not conserved among all of them. In practice, these cysteine residues were significant for developing the highly selective drugs such as afatinib and ibrutinib (Table. 3.1). Crystal structure showed that the α,β-unsaturated ketone moiety of afatinib covalently bound to Cys797 in the hinge region end of epidermal growth factor receptor (EGFR) [37]. In addition of the covalent bond, the 2-chloro-3-fluorobenzene moiety bound into the deep ATP site which was occupied by the methylbenzene moiety of erlotinib (described in detail in the next section), and the quinazoline ring forms the hydrogen bonds with the main chain atoms in the hinge region. Afatinib is classified into the Type-I inhibitor with the covalent bonding. Finally, the covalent bonding character could be introduced to all types of protein kinase inhibitors.

The comparison in the amino acid sequence and crystal structure suggests that the protein kinase inhibitor bound to the allosteric region is advantageous in terms of selectivity compared to the traditional ATP-competitive one such as staurosporine. The successful and promising examples of the partially or fully allosteric kinase inhibitor are illustrated in the following pages.

3.3 Utilization of the ATP Back Pocket, a Near-Allosteric Site/Type-I

Due to binding to the highly conserved ATP binding region, the Type-I inhibitors display a low selectivity against the off-target kinases. However, the extensive structural dissections in the deep-inside and/or peripheral region of the ATP binding site clarified the subtle but unique structural differences among protein kinases and promoted to produce the highly selective Type-I inhibitors [21]. The gatekeeper residue and N-terminal neighboring one of the DFG motif (DFG-1) in front of the hydrophobic pocket in the deep ATP site, referred to as ATP back pocket, are the key residues for the definition of pharmacophore for protein kinase inhibitors (Fig. 3.2) [46]. This ATP back pocket is divergent among protein kinases and has been utilized for drug discovery of selective protein kinase inhibitors such as gefitinib and erlotinib (Table. 3.1). Erlotinib, an anticancer drug, targets epidermal growth factor receptor (EGFR). The threonine residues as the gatekeeper and DFG-1 cause the large hydrophobic back pocket deep in the ATP binding site, thus defined as a near-allosteric site (Fig. 3.2). The acetylenyl benzene moiety of erlotinib binds to the ATP back pocket (Fig. 3.2) [32]. This characteristic interaction allows erlotinib to be highly selective for EGFR. However, the clinical efficacy of gefitinib and erlotinib is ultimately reduced by the emergence of the acquired drug resistance such as by mutation of the gatekeeper Thr790 residue of EGFR (T790M), which is detected in half of the drug-administrated patients. The resulting Met790 buried the back pocket in the deep ATP site essential for the recognition of the acetylenyl benzene moiety of erlotinib and significantly decreased the binding affinity of erlotinib with the T790M mutant. Zhou et al. discovered a novel compound effective to the T790M mutant and showed that this inhibitor fitted into the modified ATP binding site [45]. On the other hand, the clinical activity of the Type-I inhibitor entrectinib was greatly limited by the acquired resistance by the mutation at the DFG-1 position of tropomyosin

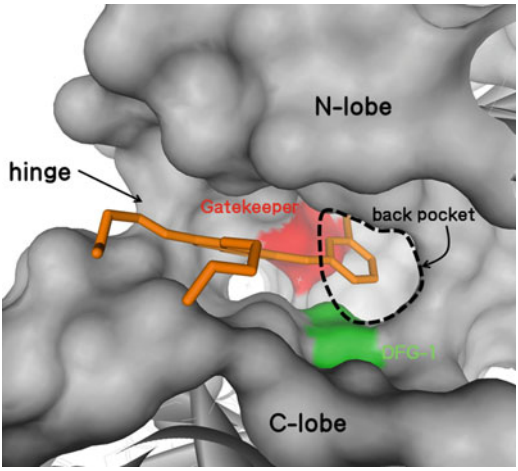


Fig. 3.2 The ATP back pocket of the epidermal growth factor receptor (EGFR) kinase domain. The acetylenyl benzene group of erlotinib (orange) intrudes into the ATP back pocket (enclosed black dot line). The gatekeeper (Thr766, shown in red) and DFG-1 (Thr830, shown in green) residues are located in front of the ATP back pocket

receptor kinase A (TrkA) (G667C). The resulting Cys667 works as a steric hindrance for entrectinib binding to the ATP back pocket of the G667C mutant, while foretinib suppressed both of the wild-type and G667C mutant [28]. The molecular modification matched for the drug-resistant mutant could be implemented, but cancer cells would acquire the resistance against the new drug. The cat-and-mouse game between the cancer cells and human wisdom will be permanently continued. Nevertheless, the ATP back pocket plays a key role in producing highly selective Type-I protein kinase inhibitors.

3.4 Utilization of the DFG-out Region, a Near-Allosteric Site/Type-II

The DFG-out region of protein kinase has appeared adjacent to the ATP binding site along with the flip-flop motion in the DFG motif. In the innate DFG-out conformation (DFG-in), the side chain of the aspartate residue is faced into the ATP site and that of the phenylalanine residue is involved in the R-spine outside of the ATP site. In the DFG-out conformation, the side chains of aspartate and phenylalanine residues moved outside and inside of the ATP binding site, respectively. Consequently, the DFG-out region retains the hydrophobic slot occupied by phenylalanine in the DFG-in conformation. The DFG-out region represents the structural divergence and connects to the ATP back pocket. Type-II inhibitor binds to the DFG-out region as well as the ATP binding site via the ATP back pocket and thus tends to be highly selective compared with Type-I inhibitors. Imatinib, an approved drug for chronic myeloid leukemia, binds to the ATP binding site and DFG-out region of Bcr-Abl, a chimera protein kinase resulting from the gene translocation [35]. Therefore, imatinib is classified as a Type-II kinase inhibitor. The pyridine, pyrimidine, and methyl benzene moieties of imatinib bound to the ATP binding site and the other portion of the inhibitor occupied the DFG-out region. The nitrogen atom in the pyridine ring forms a hydrogen bond with the main chain NH of the hinge region in the ATP binding site. The three rings are bracketed by the hydrophobic residues in the N- and C-lobes. These hydrophobic interactions complete the C-spine. The interactions like these are observed in binding of the typical Type-I inhibitor to the ATP site. The pyridine moiety of imatinib fitted into the hydrophobic space surrounded in the DFG-out region of Bcr-Abl. The pyridine ring inserts and completes the R-spine of Bcr-Abl. However, the DFG-out region was not the ultimate target for gaining the complete selectivity. The post-marketing surveillance studies revealed that imatinib also inhibited c-Kit, a receptor-type tyrosine kinase, although it was at first identified as a selective Bcr-Abl inhibitor. To date, the DFG-out configuration has been observed in a large number of protein kinases. The crystal structures of protein kinases indicated that the size of the gatekeeper residue correlates with the propensity to be DFG-out. The small gatekeeper residue probably

tolerates DFG-out motion but the large one does not. Imatinib also binds to the DFG-in conformation of spleen tyrosine kinase (Syk) that has the methionine residue at the gatekeeper position although it binds to the DFG-out conformation of the target kinases with the threonine gatekeeper [3]. This gatekeeper rule is applied to the other protein kinases such as mitogen-activated protein kinases (MAPKs). The DFG-out configuration was often observed in p38α MAPK but little or no in extracellular-regulated kinase 2 (ERK2) nor c-Jun N-terminal kinase 1 (JNK1) to date. The gatekeeper residues are threonine, glutamine, and methionine in p38α MAPK, ERK2, and JNK1, respectively.

Finally, Type-II inhibitor utilizing the ATP site and DFG-out region as a near-allosteric region could not achieve the complete selectivity. The following illustrates about the unique allosteric binding sites distinct from the regions utilized by the Types-I and Type-II inhibitors.

3.5 Characteristic p38α MAPK Inhibitors/Type-I½

p38α mitogen-activated protein kinase (MAPK) plays a crucial role in the regulation of pro-inflammatory cytokine production such as IL-6 and is an attractive drug target for inflammatory diseases including rheumatoid arthritis and psoriasis. The Type-I½ inhibitors for p38α MAPK displayed high potency and selectivity [41]. The classification of Type-I½ was descended from the combination of the Type-I and Type-II features. The most deterministic factor in the discrimination between Type-I and Type-II is the inhibitor-binding conformation in the DFG motif. The Type-I inhibitor binds to the DFG-in conformation and has no involvement in the R-spine, while the Type-II inhibitor binds to the DFG-out conformation and the hydrophobic moiety of the inhibitor takes part in forming the R-spine. The Type-I½ inhibitor binds to the DFG-in conformation, but the hydrophobic moiety of the inhibitor interacts with and stabilizes the R-spine, moving the phenylalanine residue of the DFG motif a little. Consequently, the near-allosteric pharmacophore around the ATP back pocket of p38α MAPK represents a quite unique shape among protein kinases and is pivotal for producing highly selective inhibitors. Furthermore, the Type-I½ configuration of p38α MAPK is often accompanied with the flip motion at Gly110 in the hinge region and resultantly possesses the unique shape in the ATP binding site, which accommodates atypical moieties dissimilar to the adenine moiety. The flipping point observed in p38α MAPK is not conserved among the MAPK family: ERK2 and JNK1 possess the glutamate and aspartate residues at the Gly110 position, respectively. Our structural inspections for MAPKs revealed that the Type-I½ configuration of p38α MAPK was quite unique and thus a beneficial target for elaborating highly selective inhibitors. Moreover, the structural flexibility of p38α MAPK offers the great potential to configure an unidentified conformation involving a novel allosteric pocket useful for producing potent and selective inhibitors.

3.6 Potential of the Allosteric Regions for the CK2α Inhibition/Type-I and Type-IV

Casein kinase 2 (CK2) is a serine/threonine kinase that promotes cellular growth, proliferation, and survival. CK2 is an important target protein for cancer and glomerulonephritis therapies. The CK2 holoenzyme is constitutively active without phosphorylation and consists of two catalytic subunits (CK2α) and two regulatory subunits (CK2β). CK2 uses either ATP or GTP as a phosphate group donor in the protein kinase reaction. This unusual dual specificity for nucleosides was underpinned by the particularly large space near the ribose binding site, as well as the flexibility in the αD-helix connected to the hinge region (Fig. 3.3a) [27]. CK2α displays a constitutive enzyme activity regardless of whether or not it is phosphorylated. Instead, the N-terminal segment stabilizes the active conformation of CK2α via the hydrogen bonds and hydrophobic interaction. The salt bridge between Lys68 in the β3-strand and Glu81 in the αC-helix is complete in the constitutive active conformation. The N-terminal segment, the phenylalanine residue (Phe113) at the gatekeeper position, and the tryptophan residue (Trp176) replaced by phenylalanine in the DFG motif of CK2α structurally stabilize the DFG-in configuration, i.e., likely impede transition to the DFG-out configuration. The phenylalanine residue (Phe113)

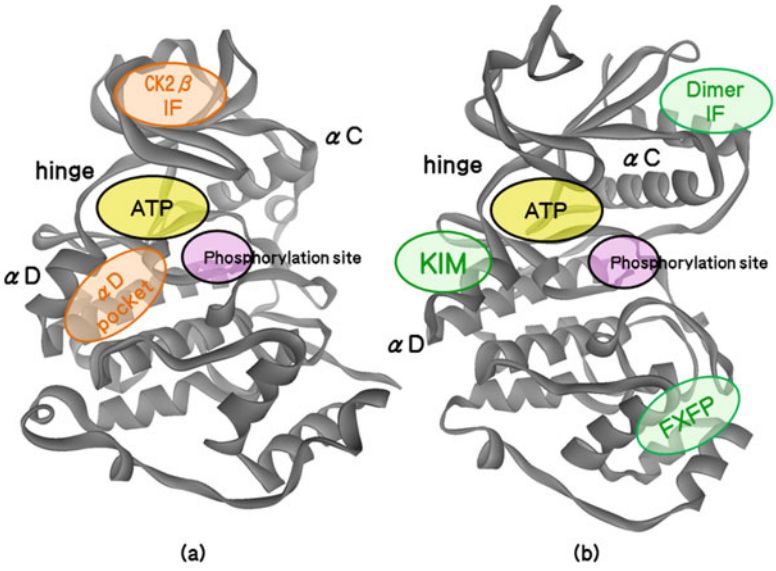


Fig. 3.3 Near- and far-allosteric sites in the protein kinase domains of CK2α and ERK2. The ATP binding and phosphorylation sites are shown in yellow and pink, respectively. (**a**) CK2α equips the αD-helix pocket and CK2β binding interface for the allosteric binding (orange circles). The αD-helix pocket is located adjacent to the ATP binding site, and the CK2β binding interface is far from the ATP site. (**b**) ERK2 involves the KIM and FXFP sites essential for the substrate recognition and dimer interface essential for activity in cytosol (green circles)

at the gatekeeper position limits the accessibility to the ATP back pocket. The crystal structures of the inhibitor-bound CK2α conferred a highly conserved water molecule hydrogen bonding with the side chains of Glu81 in the ATP back pocket and main chain NH of Trp176 in the DWG motif. This structural insight is helpful for developing selective Type-I CK2α inhibitors. The carboxylic group of the several CK2α inhibitors forms a hydrogen bond with the conserved water molecule [18]. Further, the ATP binding site is narrow compared with the other kinases owing to the bulky residues involving Phe113, Leu45, Val53, Met163, and Ile174. Several planar compounds have been identified as the Type-I inhibitor for CK2 by the structural analyses. Among them, silmitasertib (CX-4945) is in preclinical development for cancer therapy. The low flexibility except for the αD-helix constrains the pharmacophore for CK2α and thus enabled to easily develop potent inhibitors based upon the crystal structures.

The high-resolution crystal structure manifested two novel allosteric regions, the hydrophobic pocket near the αD-helix and CK2β binding interface (Fig. 3.3a) [17]. An ethylene glycol molecule, a precipitant reagent for the crystallization, bound in the αD-helix pocket, and the other ethylene glycol molecules occupied the CK2β interface. Recently, several compounds bound to the αD-helix pocket were discovered and merged with the low-selective Type-I inhibitors on the basis of several crystal structures [14]. The αD-helix pocket located in the C-lobe is a unique character of CK2 and large enough to accommodate fused-ring heterocyclic compounds such as the indole moiety. Therefore, this fragment-merge strategy is useful for producing the highly selective CK2α inhibitors. On the other hand, Raaf et al. indicated that 5,6-dichloro-1-b-d-ribofuranosyl benzimidazol (DRB), an ATP-competitive CK2 inhibitor, binds to the CK2β interface as well as the ATP binding site, and this conferred the potential to interfere with the CK2α/CK2β interaction [33]. Several compounds involving DRB bind to the CK2α/CK2β interface located at the top of the N-lobe (Fig. 3.3a) and inhibit the enzyme activity with a non-ATP competitive manner [6]. These structurally unique regions, the αD-helix pocket and CK2α/CK2β interface, are beneficial to develop CK2α-specific drugs as Type-I or Type-IV.

3.7 Unique Allosteric Binding Sites of MAP 2K/Type-III and Type-IV

Mitogen-activated protein kinase kinase 1 (MAP 2K1) functions as cell proliferation and survival and is defined as an attractive drug target of cancers. The crystal structure of the MAP 2K1 with U0126, a non-ATP-competitive inhibitor, unveiled a highly unique region adjacent to the ATP binding site [30]. In this structure, the inhibitor is bound to the unique site without interfering the ATP binding. Further investigation indicated that the unique region appeared in an auto-inhibition state adopted by the N-terminal regulatory domain (NRD) consisting of a single helix at

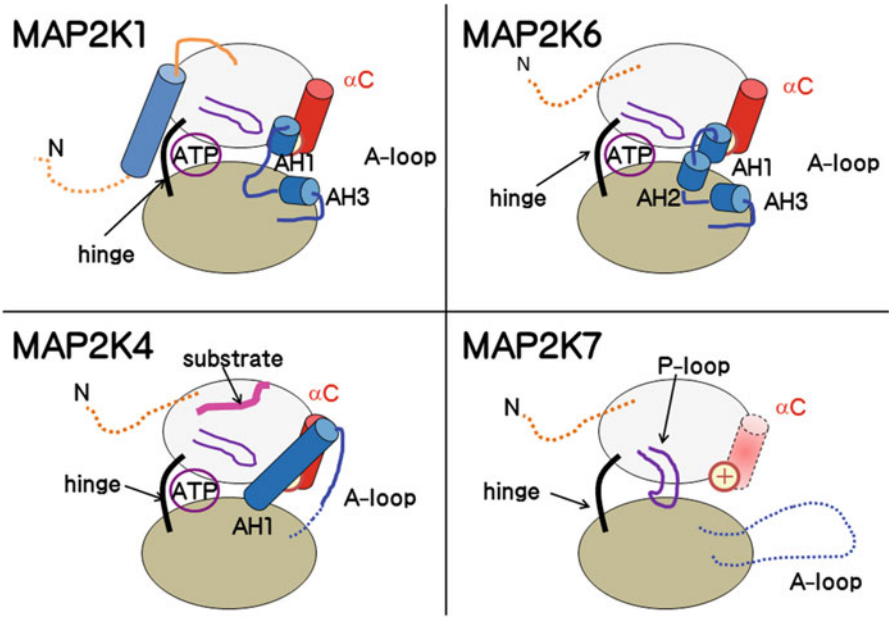


Fig. 3.4 Auto-inhibition mechanisms of the MAP 2K family kinases. MAP 2K1 is autoregulated by the N-terminal helix, which fixes the hinge motion between the N- and C-lobes. MAP 2K4 is autoregulated by binding the substrate in the N-lobe groove and configuring a long helix in the A-loop region. MAP 2K6 is autoregulated by configuring the three helices in the A-loop region. Commonly, the A-loop configurations of MAP 2K1, MAP 2K4, and MAP 2K6 in auto-inhibition state are not ready for substrate binding. The auto-inhibition state of MAP 2K7 is formed by an interaction of P-loop with the C-lobe, interfering the ATP binding

the N-terminal of the kinase domain (Fig. 3.4) ([8]). The NRD region bound to the back side of the hinge region and pulled the N-lobe toward the hinge region. The gain-of-function mutations in the NRD region found in the cancer cells likely cancel the hydrogen bond network between the NRD region and N- and C-lobes in the kinase domain. Subsequently, the αC-helix moved out from the ATP site in the auto-inhibition state, and this αC-out motion resulted in the emergence of the unique region, which accommodates a variety of non-ATP-competitive inhibitors involving U0126. Consequently, these structural insights indicated that this unique region was available for producing the Type-III MAP 2K1 inhibitor, defined as a near-allosteric site binder. Trametinib had been approved as an anticancer drug and displayed a high selectivity for MAP 2K1 and MAP 2K2 against the other MAP 2K [40]. MAP 2K2 presents a similar physiological role with MAP 2K1 and conserves the NRD and trametinib binding region [30]. The trametinib binding region of auto-inhibited MAP 2K1 is structurally unique and likely has potential to produce highly selective Type-III inhibitors. The strong point of the Type-III inhibitor is that this binds to the target kinase collaboratively with the ATP molecule, while the Type-I or Type-II inhibitor must eliminate the ATP molecule from the ATP binding site, resulting in a large

enthalpy loss. Several Type-III inhibitors for MAP 2K1 bound to the unique region, making the hydrogen bonds with the phosphate group of ATP. The allosteric pockets similar to MAP 2K1 have been observed in with-no-lysine kinase 1 (WNK1) and cyclin-dependent kinase 2 (CDK2) [44] [5], although these are distinguishable based upon their detailed structures. In the near future, the structural mechanism for forming the MAP 2K1-type allosteric pocket along with the αC-out motion would be clarified based upon the crystal structure analyses, and this likely facilitates the development of selective inhibitors for these protein kinases.

On the other hand, crystal structures of the MAP 2K1 homologues, MAP 2K4, MAP 2K6, and MAP 2K7, in the auto-inhibition state indicated that these three MAP 2Ks owned no NRD, but each has a characteristic auto-inhibition mechanism (Fig. 3.4) [25, 26, 36]. MAP 2K4 is a client kinase of p38 mitogen-activated protein kinases (p38 MAPKs) and c-Jun N-terminal kinases (JNKs), and its dysregulation occurs in the diseases such as cancers. The substrate binding to a deep cleft in the top of the N-lobe induces an auto-inhibition conformation of MAP 2K4, in which the A-loop region configures a long helix, interfering substrate binding (Fig. 3.4) [25]. Therefore, the N-lobe cleft and vacant space caused by the structural rearrangement in A-loop are likely available for producing allosteric inhibitors of MAP 2K4. MAP 2K6 is a client kinase of p38 MAPKs and an attractive drug target for inflammation and autoimmune diseases. The ATP binding triggers an auto-inhibition conformation of MAP 2K6, in which the A-loop segment was converted into three short helices, interfering substrate binding (Fig. 3.4) [26]. The hydrophobic space located at the back side of these helices is perhaps available as an allosteric site for discovering novel MAP 2K6 inhibitors. MAP 2K7 is involved in the JNK signal cascade and a significant drug discovery target against arthritis, cardiac hypertrophy, and so on. The n-$\sigma*$ interaction between Cys218 at the N-terminal of the αD-helix and Gly145 in the P-loop of MAP 2K7 shaped a closed structure, interfering with the ATP molecule binding (Fig. 3.4) [36]. The resulting structure presents the large and unique hydrophobic space in the DFG-out region sequestered by the ATP binding site. These allosteric spaces discovered in the auto-inhibition state of MAP 2K4, MAP 2K6, and MAP 2K7 are not conserved among the MAP 2K family kinases and thus noteworthy for the development of highly selective Type-III inhibitors.

High-resolution X-ray analysis, which was achieved through the experiments under the microgravity environment, conferred the alternative auto-inhibition state of MAP 2K7, regulating with an intermolecular manner [16]. The C-terminal fragment of Gly424-His430, following to the kinase domain, presented an extended configuration and bound to the groove in the N-lobe of the neighbor MAP 2K7 molecule via the hydrogen bonds and hydrophobic interaction and likely worked as an intermolecular brake of MAP 2K7. The side chain of Leu426 in the middle of the C-terminal fragment intruded into the deep hydrophobic pocket involving Asn138, Trp151, and Val164, forming the side and bottom of the N-lobe groove. The intermolecular interaction likely induces the shift of the N-lobe toward the hinge region and subsequently the DFG-out conformation without forming the C-spine. Further structural inspections revealed that this N-lobe pocket of MAP 2K7 is unique

Table 3.2 Amino acid residues significant for the N-lobe pocket of MAP 2Ks

MAP2K7	Asn138	Trp151	Val164
2K1	Lys	Phe	Arg
2K2	Arg	Thr	Arg
2K3	Thr	Glu	Val
2K4	Asp	Asn	Val
2K5	Tyr	Tyr	Val
2K6	Pro	Glu	Val

among MAP 2Ks. The other MAP 2K could not accommodate the leucine residue with the region similar to the Leu426 binding pocket of MAP 2K7 because of the bulky or hydrophilic residue replaced by that of MAP 2K7 (Table. 3.2). Actually, the synthesized C-terminal peptide moderately but concentrate dependently inhibited the MAP 2K7 activity and thus could be defined as a seed compound of the Type-IV drug. The top of the N-lobe is used for the protein-protein interaction and autoregulation in several kinases such as MAP 2K4 and 3-phosphoinositide-dependent protein kinase (PDK1) and useful for the development of highly selective inhibitors. Inactive MAP 2K4 accommodates the substrate in the N-terminal groove as abovementioned [25]. PDK1 equips the PDK1-interacting fragment (PIF) pocket in the N-lobe, which primary recruits the substrate proteins. Rettenmaier et al. discovered the Type-IV inhibitors, which inhibit PDK1 in cells, based upon this structural insight [34]. Fortunately, these allosteric pockets of MAP 2K4, MAP 2K7, and PDK1 are nonoverlapping in the N-lobe and recognize the distinguished sequence.

Together all, we have the great chances to produce the near- and far-allosteric MAP 2K inhibitors (Type-III and Type-IV), which represent the tendency to be highly selective against the off-target kinases. However, we cannot quantitatively estimate the potential of the obtained inhibitor to date.

3.8 Far-Allosteric Inhibition of ERK2/Type-IV

Extracellular-regulated kinase 2 (ERK2) is a member of mitogen-activated protein kinases (MAPKs), regulated by MAP 2K1 via phosphorylation and defined as a drug discovery target for serious diseases such as cancer and type 2 diabetes. The ATP-competitive Type-I inhibitors for ERK2 have been discovered, and some of them are in clinical study [29, 42]. In addition to these, the far-allosteric inhibitors were discovered on the basis of the unique substrate recognition mechanism [15]. ERK2 anchors the substrate proteins via their consensus sequence of Lys $(Arg)_{2\text{-}3}\text{-}X_{1\text{-}6}\text{-}\Phi_a\text{-}X\text{-}\Phi_b$ (Φ, hydrophobic amino acid; X, any amino acid) at the kinase interacting motif binding site (KIM site) which is distinct from the phosphorylation site recognizing the Ser(Thr)-Pro sequence of the substrate (Fig. 3.3b). The far-allosteric KIM site is located in the back side of the hinge region but inaccessible

to the ATP binding site. Therefore, the KIM binder is classified as the Type-IV inhibitor. A synthesized peptide based on the consensus sequence in signal transducer and activator of transcription 3 (STAT3) inhibited the ERK2 activity with the IC_{50} value of 9.8 μM and had a significant effect on the model mice without the serious adverse effects [15]. The X-ray analysis conferred the separated sub-pockets in the KIM site. The two aliphatic residues of the peptide inhibitor are accommodated in the shallow pocket configured by the hydrophobic residues in the KIM region. The positive-charged residues of the peptide inhibitor were located in the negative-charged pool involving three aspartates and a glutamate residue in the KIM region of ERK2. The hydrophobic pocket and negative-charged pool are completely sequestered in the KIM region. The arbitrary sequence in the middle of the consensus peptide likely works as a sole linker of these characteristic regions. The KIM region forms a shallow groove and is disadvantageous for the gain of binding affinity compared with the ATP binding site. Nevertheless, the drug discovery studies indicated the potential to realize the far-allosteric inhibition of ERK2. The KIM binders discovered by the in silico screening displayed the moderate inhibitory activity and competitive behavior to the STAT3 peptide [19]. JNKs and p38 MAPKs, members of the MAP kinases involving ERK2, possess the common substrate recognition mechanism using the KIM region. MAPKs discriminate their own substrates with high fidelity through the recognition in the KIM region. Therefore, the KIM binding inhibitors are expected to possess high selectivity among MAPKs. However, the KIM inhibitors reported to date displayed the rather low activity. To solve this problem, the bivalent inhibitors were invented by merging the Type-I ATP competitive and KIM peptide inhibitors via the long linker segment [20]. The bivalent inhibitors exhibited high inhibitory potency against ERK2.

The FXFP region found in the MAP kinase insert region of MAPKs works as a second substrate anchoring site utilized by the transcription factor substrates (Fig. 3.3b) [1]. Structural studies revealed that the FXFP site was observed in the active ERK2 (doubly phosphorylated state) but inaccessible to the inactive state of ERK2 (non-phosphorylated state). In the inactive state, Phe181 and Leu182 in the A-loop occupied the FXFP site of ERK2. Double phosphorylation likely induces the conformation change in the A-loop and opens the FXFP site. The second anchoring system is also conserved in all MAPKs but divergent enough to discriminate their own substrates like the KIM site.

Upon activation, the dimer state is essential for the ERK2 extranuclear signaling. Herrero et al. reported that a small molecule inhibitor for ERK2 dimerization forestalls tumorigenesis [13]. Biochemical experiments revealed that this inhibitor interfered with the dimerization of ERK2 and probably bound to the dimerization interface of ERK2. Crystal structure indicated that the tip of the A-loop involving Asp175 intruded into the small pocket was located in the back side of the connecting loop between the β3-strand and αC-helix [13] (Fig. 3.3b).

To the best of our knowledge, three far-allosteric pockets involving the substrate recognition sites of KIM and FXFP and dimer interface are available for producing highly selective Type-IV inhibitors of ERK2.

3.9 Allosteric Inhibition of Receptor Tyrosine Kinases/ Type-III

The receptor tyrosine kinase (RTK) family represents a typical auto-inhibition configuration in which accompanying the DFG-out transition, the phenylalanine residue in the DFG motif intrudes deeply into the ATP site (Fig. 3.5). In addition to that, the tyrosine residue as a phosphorylate acceptor in the A-loop is located at the substrate binding position in this auto-inhibition state (Fig. 3.5). The resulting allosteric pocket involving the DFG-out region is probably available for producing selective Type-III inhibitors (Fig. 3.5). FMS-like tyrosine kinase (FLT3), a member of the RTK family, is essential for the normal function of stem cells and the immune system and potential target for therapies of a variety of leukemia. The allosteric pocket in the typical RTK auto-inhibition structure of FLT3 is fully occupied by Tyr572 and its marginal residues in the juxtamembrane region (Fig. 3.5) [10] and is likely available for developing selective FLT3 inhibitors. Insulin-like growth factor-1 receptor (IGF-1R) is also a member of RTK family and associates to the signaling to tumor. Heinrich et al. showed that the indolealkylamine class inhibitors, which were discovered by a high-throughput screening, bound to the near-allosteric pocket involving the DFG-out region in the typical RTK auto-inhibition conformation of IGF-1R [12]. Thus, these inhibitors are defined as Type-III inhibitors and stabilize the RTK auto-inhibition conformation, concealing the phosphorylating tyrosine residue in the A-loop.

Tropomyosin receptor kinase A (TrkA) is a receptor for nerve growth factor (NGF) that mediates neuronal differentiation and cellular survival. TrkA belongs to the receptor tyrosine kinase family, consisting of an extracellular ligand-binding

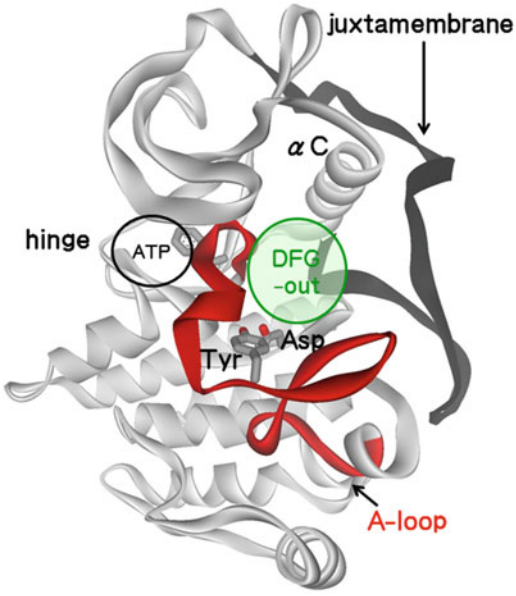


Fig. 3.5 An auto-inhibition configuration of FLT3 on behalf of the receptor tyrosine kinase. The phenylalanine residue in the DFG motif intrudes into the ATP binding site, resulting in the narrow ATP binding site. The tyrosine residue in the A-loop occupies the phosphorylation site and forms a hydrogen bond with the aspartate residue as a proton acceptor. In this conformation, the A-loop sequesters the DFG-out region from the ATP binding site

domain, a transmembrane helix, an intracellular juxtamembrane region (JM), and a tyrosine kinase domain. The TrkA signaling pathway plays a crucial role in the physiological function of chronic pain and is an attractive drug target for chronic pain. TrkA is a member of the Trk subfamily, which also includes TrkB and TrkC. Each Trk plays a distinct function. Thus, selective TrkA inhibitors are desired for the chronic pain therapy. Undesirable side effects have been reported for pan-Trk inhibitors. The several Types-I and Type-II inhibitors for TrkA have been developed based upon the crystal structures but disappointingly exerted the pan-activity. The crystal structures depicted that the inhibitor-binding environments in the ATP and DFG-out sites were completely conserved among the Trk family. These structural insights coincide with the low selectivity of the above inhibitors. However, some research organizations have reported the highly selective TrkA inhibitors, which were expected to be classified into Type-III, binding to the RTK allosteric pocket as observed in IGF-1R. The crystal structures of TrkA with the highly selective inhibitor unveiled that the inhibitor-binding pocket were assembled from the allosteric RTK pocket and JM region and completely sequestered from the ATP site [9, 38]. Therefore, these selective inhibitors were confirmed as the Type-III inhibitor. One-third of the pocket is composed of amino acid residues in the JM region. Owing to the low sequence homology of the JM region among Trks, these inhibitors exhibit high selectivity. Trks share significant sequence homology in the kinase domain involving the DFG-out region, whereas the JM region is highly divergent among the Trk family. The selective Type-III inhibitor binds to the JM region via hydrogen bonds, van der Waals interactions, and CH-π interactions at Leu486, His489, Ile490, etc. These residues of TrkA are not conserved in TrkB and TrkC and should contribute to the high binding affinity and selectivity.

Collectively, the receptor tyrosine kinases have a typical auto-inhibition structure and its close relatives useful for producing potent Type-III inhibitors, and the flexible region such as JM occasionally plays a definitive role for the high selectivity and potency of the inhibitors.

3.10 Molecular Switch Regulation of RSKs and MSKs/ Type-IV

Dimethyl fumarate (DMF) has been used for the oral therapy of psoriasis and multiple sclerosis. More recently, Anderson et al. suggested that DMF covalently bound to the hydrophobic pocket in the C-lobe of p90 ribosomal S6 kinases (RSKs) and mitogen- and stress-activated kinases (MSKs) as its potential target proteins [2]. The crystal structure of the DMF-RSK2 complex showed that the Michael acceptor of DMF made a covalent bond with the Cys599 in the bottom of the C-lobe pocket formed by the hydrophobic residues Trp602, Val662, Ile633, and Leu710 [2]. The C-lobe pocket was estimated to be an allosteric regulator that bound the tip of the phosphorylated A-loop and stabilized the active conformation. The

C-lobe pocket containing the cysteine residue is conserved among RSKs and MSKs. Therefore, the DMF-binding pocket is noteworthy for producing highly selective Type-IV inhibitors for RSKs and MSKs. Further structural dissections probably allow to discriminating the C-lobe pocket of these homologous kinases.

3.11 The Pursuit of Unidentified Allosteric Pockets in Protein Kinases

In the near future, all the protein kinase structures would be available in Protein Data Bank and consequently accelerate the production of highly selective inhibitors. Bioinformatics technology and computational chemistry using artificial intelligence (AI) presumably play central roles in dissecting and characterizing the inhibitor-binding pockets of all protein kinases. The structural flexibility of protein kinases confers the innumerable conformations in solution, involving some energetically stable those such as the auto-inhibition state. The conformational selection by a protein kinase inhibitor results in the population shift in the protein kinase configuration. The DFG-out configuration was firstly identified as a result of the Type-II inhibitor binding but not observed as an apo state. Molecular dynamics studies facilitated the comprehensive understanding of the structural flexibility of protein kinases [7, 39]. With the upsurge in computational power, it would be possible to develop drug discovery techniques involving in silico screening against the conformational ensemble of a target protein kinase and the clarification of unprecedented conformations that could not been identified without crystal analyses. Meanwhile, several recent studies showed that the propensity of the slow dissociation kinetics of the protein kinase inhibitor was beneficial for drug efficacy and associated with conformational change. Ayaz et al. indicated that roniciclib, a Type-I inhibitor, exerted the kinetic selectivity for CDK2, indicating sustainable effect on retinoblastoma protein phosphorylation [4]. Furthermore, the derivatives of roniciclib prolonged the resident time by the inhibitor-induced conformational change in binding to the CDK2 [4]. Finally, the computation development involving molecular dynamics would dissect the conformational behavior and binding kinetics of protein kinases to produce highly selective protein kinase inhibitors.

3.12 Conclusion

It is necessary to resolve a large number of the significant challenges to develop inhibitors to be drugs. Protein kinase inhibitors with low selectivity suppress a variety of the off-target protein kinases as well as the target one and arguably exert the adverse effects on vital. To pursue the production of highly selective protein kinase inhibitors, a variety of the druggable allosteric sites involving the ATP back

Table 3.3 Allosteric sites available for producing highly selective protein kinase inhibitors

Protein kinase	Allosteric site	Target type	Section
EGFR	ATP back pocket	Type-I	3
TrkA	ATP back pocket	Type-I	3
	DFG-out region	Type-III	9
Bcr-Abl	DFG-out region	Type-II	4
c-Kit	DFG-out region	Type-II	4
p38α MAPK	ATP back pocket	Type-I½	5
CK2α	αD-helix pocket	Type-I	6
	CK2β interface	Type-IV	6
MAP2K1	αC-out pocket	Type-III	7
MAP2K2	αC-out pocket	Type-III	7
MAP2K4	N-lobe pocket	Type-IV	7
MAP2K6	DFG-out region	Type-III	7
MAP2K7	DFG-out region	Type-III	7
	N-lobe pocket	Type-IV	7
WNK1	αC-out pocket	Type-III	7
CDK2	αC-out pocket	Type-III	7
PDK1	N-lobe pocket	Type-IV	7
ERK2	KIM site	Type-IV	8
	FXFP site	Type-IV	8
	Dimer interface	Type-IV	8
FLT3	DFG-out region	Type-III	9
IGF-1R	DFG-out region	Type-III	9
RSK2	C-lobe pocket	Type-IV	10
MSKs	C-lobe pocket	Type-IV	10

pocket (Sect. 3.3), the DFG-out region (Sect. 3.4), the αD-helix pocket and CK2β binding interface of CK2α (Sect. 3.6), the KIM and FXFP sites of ERK2 (Sect. 3.8), the N-lobe pockets (Sects. 3.6, 3.7 and 3.8), the C-lobe pocket (Sect. 3.10), the physiologically essential dimer interface (Sect. 3.8), and so on are available in the protein kinase domains (Table 3.3). These allosteric sites are primarily utilized for the diverse physiological function such as the activity regulation, substrate recognition, and subcellular localization of protein kinases. These allosteric molecular switches are quite unique among protein kinases. Therefore, the allosteric sites equipped in the protein kinase domain are noteworthy for gaining high selectivity. The approved kinase inhibitors gained the high selectivity by more or less using the allosteric site. Erlotinib, a Type-I EGFR inhibitor for the anticancer therapy, bound to the ATP back pocket (Sect. 3.3). Imatinib, a Type-II Bcr-Abl inhibitor for the chronic leukemia therapy, occupied the DFG-out region (Sect. 3.4). Trametinib, a Type-III MAP 2K1 inhibitor as anticancer drug, used only the near-allosteric region induced by the auto-inhibition (Sect. 3.7). The Type-IV kinase inhibitor binds to the far-allosteric region and thus is expected to be highly selective but displays the low inhibitory potency. The Type-V kinase inhibitor presents a considerable burden in gaining the high selectivity because of little or no use of the allosteric region.

The structural flexibility of protein kinases confers the innumerable conformations in solution, involving the unidentified allosteric pockets as discovered in the auto-inhibition state. Finally, the computation development involving molecular dynamics would dissect the conformational behavior of protein kinases and promote to produce highly selective protein kinase inhibitors.

References

1. Akella R, Moon TM, Goldsmith EJ (2008) Unique MAP Kinase binding sites. Biochim Biophys Acta 1784(1):48–55
2. Andersen JL, Gesser B, Funder ED, Nielsen CJF, Gotfred-Rasmussen H, Rasmussen MK, Toth R, Gothelf KV, Arthur JSC, Iversen L, Nissen P (2018) Dimethyl fumarate is an allosteric covalent inhibitor of the p90 ribosomal S6 kinases. Nat Commun 9(1):4344
3. Atwell S, Adams JM, Badger J, Buchanan MD, Feil IK, Froning KJ, Gao X, Hendle J, Keegan K, Leon BC, Muller-Dieckmann HJ, Nienaber VL, Noland BW, Post K, Rajashankar KR, Ramos A, Russell M, Burley SK, Buchanan SG (2004) A novel mode of Gleevec binding is revealed by the structure of spleen tyrosine kinase. J Biol Chem 279(53):55827–55832
4. Ayaz P, Andres D, Kwiatkowski DA, Kolbe CC, Lienau P, Siemeister G, Lucking U, Stegmann CM (2016) Conformational adaption may explain the slow dissociation kinetics of roniciclib (BAY 1000394), a type I CDK inhibitor with kinetic selectivity for CDK2 and CDK9. ACS Chem Biol 11(6):1710–1719
5. Betzi S, Alam R, Martin M, Lubbers DJ, Han H, Jakkaraj SR, Georg GI, Schonbrunn E (2011) Discovery of a potential allosteric ligand binding site in CDK2. ACS Chem Biol 6(5):492–501
6. Brear P, North A, Iegre J, Hadje Georgiou K, Lubin A, Carro L, Green W, Sore HF, Hyvonen M, Spring DR (2018) Novel non-ATP competitive small molecules targeting the CK2 alpha/beta interface. Bioorg Med Chem 26(11):3016–3020
7. De Vivo M, Masetti M, Bottegoni G, Cavalli A (2016) Role of molecular dynamics and related methods in drug discovery. J Med Chem 59(9):4035–4061
8. Fischmann TO, Smith CK, Mayhood TW, Myers JE, Reichert P, Mannarino A, Carr D, Zhu H, Wong J, Yang RS, Le HV, Madison VS (2009) Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. Biochemistry 48(12):2661–2674
9. Furuya N, Momose T, Katsuno K, Fushimi N, Muranaka H, Handa C, Ozawa T, Kinoshita T (2017) The juxtamembrane region of TrkA kinase is critical for inhibitor selectivity. Bioorg Med Chem Lett 27(5):1233–1236
10. Griffith J, Black J, Faerman C, Swenson L, Wynn M, Lu F, Lippke J, Saxena K (2004) The structural basis for autoinhibition of FLT3 by the juxtamembrane domain. Mol Cell 13(2):169–178
11. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. FASEB J 9(8):576–596
12. Heinrich T, Gradler U, Bottcher H, Blaukat A, Shutes A (2010) Allosteric IGF-1R inhibitors. ACS Med Chem Lett 1(5):199–203
13. Herrero A, Pinto A, Colon-Bolea P, Casar B, Jones M, Agudo-Ibanez L, Vidal R, Tenbaum SP, Nuciforo P, Valdizan EM, Horvath Z, Orfi L, Pineda-Lucena A, Bony E, Keri G, Rivas G, Pazos A, Gozalbes R, Palmer HG, Hurlstone A, Crespo P (2015) Small molecule inhibition of ERK dimerization prevents tumorigenesis by RAS-ERK pathway oncogenes. Cancer Cell 28(2):170–182
14. Iegre J, Brear P, De Fusco C, Yoshida M, Mitchell SL, Rossmann M, Carro L, Sore HF, Hyvonen M, Spring DR (2018) Second-generation CK2alpha inhibitors targeting the alphaD pocket. Chem Sci 9(11):3041–3049

15. Kinoshita T, Doi K, Sugiyama H, Kinoshita S, Wada M, Naruto S, Tomonaga A (2011a) Knowledge-based identification of the ERK2/STAT3 signal pathway as a therapeutic target for type 2 diabetes and drug discovery. Chem Biol Drug Des 78(3):471–476
16. Kinoshita T, Hashimoto T, Sogabe Y, Fukada H, Matsumoto T, Sawa M (2017) High-resolution structure discloses the potential for allosteric regulation of mitogen-activated protein kinase kinase 7. Biochem Biophys Res Commun 493(1):313–317
17. Kinoshita T, Nakaniwa T, Sekiguchi Y, Sogabe Y, Sakurai A, Nakamura S, Nakanishi I (2013) Crystal structure of human CK2alpha at 1.06 A resolution. J Synchrotron Radiat 20(Pt 6):974–979
18. Kinoshita T, Sekiguchi Y, Fukada H, Nakaniwa T, Tada T, Nakamura S, Kitaura K, Ohno H, Suzuki Y, Hirasawa A, Nakanishi I, Tsujimoto G (2011b) A detailed thermodynamic profile of cyclopentyl and isopropyl derivatives binding to CK2 kinase. Mol Cell Biochem 356(1-2):97–105
19. Kinoshita T, Sugiyama H, Mori Y, Takahashi N, Tomonaga A (2016) Identification of allosteric ERK2 inhibitors through in silico biased screening and competitive binding assay. Bioorg Med Chem Lett 26(3):955–958
20. Lechtenberg BC, Mace PD, Sessions EH, Williamson R, Stalder R, Wallez Y, Roth GP, Riedl SJ, Pasquale EB (2017) Structure-guided strategy for the development of potent bivalent ERK inhibitors. ACS Med Chem Lett 8(7):726–731
21. Liao JJ (2007) Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. J Med Chem 50(3):409–424
22. Liu Q, Sabnis Y, Zhao Z, Zhang T, Buhrlage SJ, Jones LH, Gray NS (2013) Developing irreversible inhibitors of the protein kinase cysteinome. Chem Biol 20(2):146–159
23. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934
24. Marcie AG, Gregory DC, Min L, Brittany D, Kenneth A, Ross LS, Kenneth SK (2007) New approaches to the discovery of cdk5 inhibitors. Current Alzheimer Research 4(5):547–549
25. Matsumoto T, Kinoshita T, Kirii Y, Yokota K, Hamada K, Tada T (2010) Crystal structures of MKK4 kinase domain reveal that substrate peptide binds to an allosteric site and induces an auto-inhibition state. Biochem Biophys Res Commun 400(3):369–373
26. Matsumoto T, Kinoshita T, Matsuzaka H, Nakai R, Kirii Y, Yokota K, Tada T (2012) Crystal structure of non-phosphorylated MAP 2K6 in a putative auto-inhibition state. J Biochem 151(5):541–549
27. Niefind K, Putter M, Guerra B, Issinger OG, Schomburg D (1999) GTP plus water mimic ATP in the active site of protein kinase CK2. Nat Struct Biol 6(12):1100–1103
28. Nishiyama A, Yamada T, Kita K, Wang R, Arai S, Fukuda K, Tanimoto A, Takeuchi S, Tange S, Tajima A, Furuya N, Kinoshita T, Yano S (2018) Foretinib overcomes entrectinib resistance associated with the NTRK1 G667C mutation in NTRK1 fusion-positive tumor cells in a brain metastasis model. Clin Cancer Res 24(10):2357–2369
29. Ohori M, Kinoshita T, Okubo M, Sato K, Yamazaki A, Arakawa H, Nishimura S, Inamura N, Nakajima H, Neya M, Miyake H, Fujii T (2005) Identification of a selective ERK inhibitor and structural determination of the inhibitor-ERK2 complex. Biochem Biophys Res Commun 336(1):357–363
30. Ohren JF, Chen H, Pavlovsky A, Whitehead C, Zhang E, Kuffa P, Yan C, McConnell P, Spessard C, Banotai C, Mueller WT, Delaney A, Omer C, Sebolt-Leopold J, Dudley DT, Leung IK, Flamme C, Warmus J, Kaufman M, Barrett S, Tecle H, Hasemann CA (2004) Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. Nat Struct Mol Biol 11(12):1192–1197
31. Parang K, Till JH, Ablooglu AJ, Kohanski RA, Hubbard SR, Cole PA (2001) Mechanism-based design of a protein kinase inhibitor. Nat Struct Biol 8(1):37–41
32. Park JH, Liu Y, Lemmon MA, Radhakrishnan R (2012) Erlotinib binds both inactive and active conformations of the EGFR tyrosine kinase domain. Biochem J 448(3):417–423

33. Raaf J, Brunstein E, Issinger OG, Niefind K (2008) The CK2 alpha/CK2 beta interface of human protein kinase CK2 harbors a binding pocket for small molecules. Chem Biol 15(2):111–117
34. Rettenmaier TJ, Sadowsky JD, Thomsen ND, Chen SC, Doak AK, Arkin MR, Wells JA (2014) A small-molecule mimic of a peptide docking motif inhibits the protein kinase PDK1. Proc Natl Acad Sci U S A 111(52):18590–18595
35. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 289(5486):1938–1942
36. Sogabe Y, Hashimoto T, Matsumoto T, Kirii Y, Sawa M, Kinoshita T (2016) A crucial role of Cys218 in configuring an unprecedented auto-inhibition form of MAP 2K7. Biochem Biophys Res Commun 473(2):476–481
37. Solca F, Dahl G, Zoephel A, Bader G, Sanderson M, Klein C, Kraemer O, Himmelsbach F, Haaksma E, Adolf GR (2012) Target binding properties and cellular activity of afatinib (BIBW 2992), an irreversible ErbB family blocker. J Pharmacol Exp Ther 343(2):342–350
38. Su HP, Rickert K, Burlein C, Narayan K, Bukhtiyarova M, Hurzy DM, Stump CA, Zhang X, Reid J, Krasowska-Zoladek A, Tummala S, Shipman JM, Kornienko M, Lemaire PA, Krosky D, Heller A, Achab A, Chamberlin C, Saradjian P, Sauvagnat B, Yang X, Ziebell MR, Nickbarg E, Sanders JM, Bilodeau MT, Carroll SS, Lumb KJ, Soisson SM, Henze DA, Cooke AJ (2017) Structural characterization of nonactive site, TrkA-selective kinase inhibitors. Proc Natl Acad Sci U S A 114(3):E297–E306
39. Tong M, Seeliger MA (2015) Targeting conformational plasticity of protein kinases. ACS Chem Biol 10(1):190–200
40. Uitdehaag JC, de Roos JA, van Doornmalen AM, Prinsen MB, de Man J, Tanizawa Y, Kawase Y, Yoshino K, Buijsman RC, Zaman GJ (2014) Comparison of the cancer gene targeting and biochemical selectivities of all targeted kinase inhibitors approved for clinical use. PLoS One 9(3):e92146
41. Walter NM, Wentsch HK, Buhrmann M, Bauer SM, Doring E, Mayer-Wrangowski S, Sievers-Engler A, Willemsen-Seegers N, Zaman G, Buijsman R, Lammerhofer M, Rauh D, Laufer SA (2017) Design, synthesis, and biological evaluation of novel type I(1)/2 p38alpha MAP kinase inhibitors with excellent selectivity, high potency, and prolonged target residence time by interfering with the R-Spine. J Med Chem 60(19):8027–8054
42. Ward RA, Colclough N, Challinor M, Debreczeni JE, Eckersley K, Fairley G, Feron L, Flemington V, Graham MA, Greenwood R, Hopcroft P, Howard TD, James M, Jones CD, Jones CR, Renshaw J, Roberts K, Snow L, Tonge M, Yeung K (2015) Structure-guided design of highly selective and potent covalent inhibitors of ERK1/2. J Med Chem 58(11):4790–4801
43. Wu P, Nielsen TE, Clausen MH (2016) Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov Today 21(1):5–10
44. Yamada K, Zhang JH, Xie X, Reinhardt J, Xie AQ, LaSala D, Kohls D, Yowe D, Burdick D, Yoshisue H, Wakai H, Schmidt I, Gunawan J, Yasoshima K, Yue QK, Kato M, Mogi M, Idamakanti N, Kreder N, Drueckes P, Pandey P, Kawanami T, Huang W, Yagi YI, Deng Z, Park HM (2016) Discovery and characterization of allosteric WNK kinase inhibitors. ACS Chem Biol 11(12):3338–3346
45. Zhou W, Ercan D, Chen L, Yun CH, Li D, Capelletti M, Cortot AB, Chirieac L, Iacob RE, Padera R, Engen JR, Wong KK, Eck MJ, Gray NS, Janne PA (2009) Novel mutant-selective EGFR kinase inhibitors against EGFR T790M. Nature 462(7276):1070–1074
46. Zuccotto F, Ardini E, Casale E, Angiolini M (2010) Through the "Gatekeeper Door": exploiting the active kinase conformation. J Med Chem 53(7):2681–2694

Chapter 4
Progress in Allosteric Database

Kun Song, Jian Zhang, and Shaoyong Lu

Abstract An allosteric mechanism refers to the biological regulation process wherein macromolecules propagate the effect of ligand binding at one site to a spatially distant orthosteric locus, thus affecting activity. The theory has remained a trending topic in biology research for over 50 years, since the understanding of allostery is fundamental for gleaning numerous biological processes and developing new drug therapies. In the past two decades, the allosteric paradigm has evolved into more descriptive models, with ever-expanding amounts of experimental data pertaining to newly identified allosteric molecules. The AlloSteric Database (ASD, accessible at http://mdl.shsmu.edu.cn/ASD), which is a comprehensive knowledge repository, has provided the public with integrated information encompassing allosteric proteins, modulators, sites, pathways, and networks to investigate allostery since 2009. In this chapter, we introduce the history and usage of the ASD and give attention to specific applications that have benefited from the ASD.

Keywords ASD · Allostery · Drug development

4.1 Introduction

Allostery is the most common and efficient means that enables the exquisite control of macromolecule function. Nature leverages allosteric modulation in a wide variety of cellular processes such as signal transduction, transcription activation, and feedback inhibition [5, 16, 33]. The biological occurrence of allosteric regulation is characterized by conformational or dynamical changes at the active site induced by the stimuli-like binding of molecules or covalent modification at another site via the transmittal cascade of residue-residue fluctuations throughout the structure [4].

K. Song · J. Zhang · S. Lu (✉)
Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_4

Early experimental evidence has resulted in the development of two theoretical allosteric models, a concerted model raised by Monod, Wyman, and Changeux (MWC) [35] and a sequential model proposed by Koshland, Némethy, and Filmer (KNF) [25], both of which consider two conformation states for different protein domains. The MWC model emphasizes that the chemical equilibrium shift of two pre-existing states upon ligand binding leads to cooperative regulation. In contrast, the KNF model assumes the "induced-fit" allosteric transition mechanism that ligand binding on one subunit will trigger conformational changes in other subunits. Undoubtedly, both models have been successful in depicting allosteric mechanisms for a long time. However, how the structure facilitates allosteric coupling between the allosteric site and the orthosteric site is still unclear. Considering the ensemble nature of biomolecule systems, a modern perspective of allostery supports that the manipulation is derived from the population redistribution of the entire conformational state [36, 39].

Advance in deciphering allosteric mechanisms drives the field of developing allosteric drugs, which holds the promise for producing therapeutic molecules with higher selectivity and lower toxicity compared to conventional orthosteric drugs [20, 40, 57]. Allosteric modulators generally have a saturable effect that does not fully activate or inhibit protein function, resulting in an improved safety profile [22]. Additionally, the modulation effect often occurs in the presence of endogenous orthosteric ligands, allowing for natural cellular function to proceed [42]. Since allosteric drugs boast many potential advantages, great efforts have been made by the pharmaceutical society, which is evidenced by the approval of 11 allosteric drugs.

In the past two decades, improvements in experimental methods such as X-ray diffraction, nuclear magnetic resonance, and high-throughput screening in conjunction with computational approaches such as molecular dynamics simulation have produced numerous data regarding allosteric cocomplexes, compounds, and pathways. These improvements have led to a singular opportunity for propelling the understanding of the allosteric mode of action, the disclosure of allosteric proteins, and the exploiting of novel compounds targeting allosteric sites. In contrast, most allosteric compounds are discovered coincidentally through virtual screening, as the resources and strategies that can promote rational and systematic development of these agents are still lacking [7, 55]. Therefore, the presence of an open-accessible database designed for searching, visualizing, and analyzing allosteric data may bridge this gap. The AlloSteric Database (ASD) has provided a versatile resource for structure, function, disease, and related annotation for the well-established allosteric macromolecules and allosteric modulators since 2009. Data in the ASD are free for biologists and medicinal chemists interested in allosteric regulation mechanisms and allosteric drug discovery.

4.2 History of ASD

In this section, three released versions of the ASD will be introduced. The deluge of allosteric data and feedback from the community enable us to append new modules for deciphering allostery and facilitating allosteric drug development. Figure 4.1 shows the main page of the ASD.

Fig. 4.1 Screenshot of the main page of ASD v3.0

4.2.1 ASD v1.0

ASD v1.0, the first version of the database, was constructed as an integrated resource that focuses on allosteric information describing the specific structure, function, and mechanism of 336 allosteric proteins and 8095 allosteric modulators [18]. All allosteric data are deposited in MySQL (Fig. 4.2).

The original allosteric data were collected from scientific literature, relevant web servers, and United States Patent and European Patent files. Three hundred and thirty-six proteins comprising 101 species deposited in the ASD were manually verified in which their functional variation was evoked by ligand binding at a site that was topographically distinct from the orthosteric site. These proteins were annotated with gene information, biological function, natural mutations, and related diseases. The structural information for allosteric proteins were also collected and annotated, incorporating crystal structures from the Protein Data Bank (PDB) or theoretical structure models predicted by I-TASSER [46], the structural classifications SCOP [38] and CATH [9], and allosteric sites validated by biochemical assays or structural details from the literature.

Likewise, allosteric modulators for 336 allosteric proteins were searched for in PubMed and the United States Patent and European Patent files, which yielded 8095 chemical allosteric modulators. All allosteric ligands were classified into three categories, namely, allosteric activators for those increasing a protein function, allosteric inhibitors for

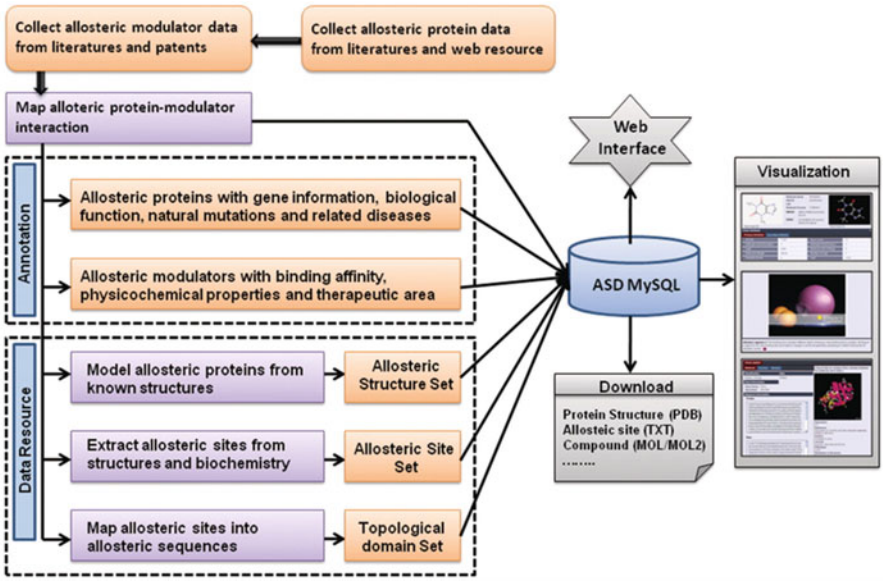


Fig. 4.2 The architecture of the ASD. Users can visualize and analyze a great range of allosteric information by interfacing the web page

those decreasing a protein function, and allosteric regulators for those regulating enzyme activity or controlling cooperativity of multisubunit proteins indirectly via allostery, etc. Additionally, allosteric modulators were labeled as "Endogenous" or "Druggable" to distinguish molecules produced in vivo from compounds leveraged for drug design. Physicochemical properties such as the logP, PSA, and number of rotatable bonds together with the binding affinity available from respective references for each allosteric modulator were provided for chemical optimization.

In the first release of the ASD, the goal was to focus on integrating identified allosteric proteins and modulators in the past 50 years to facilitate systematic allosteric investigation. However, owing to the increasing number of allosteric proteins being identified, the whole repertoire of allostery needed to be further unraveled.

4.2.2 ASD v2.0

For the second version of the ASD, in addition to increasing the amount of allosteric proteins and modulators, the allosteric sites and allosteric pathways were constructed to recapitulate the regulatory mechanisms in various biological processes from the structure view [19]. Moreover, medical chemists could design compounds with high selectivity targeting less conserved allosteric pockets. Figure 4.3 displays the new features of ASD v2.0.

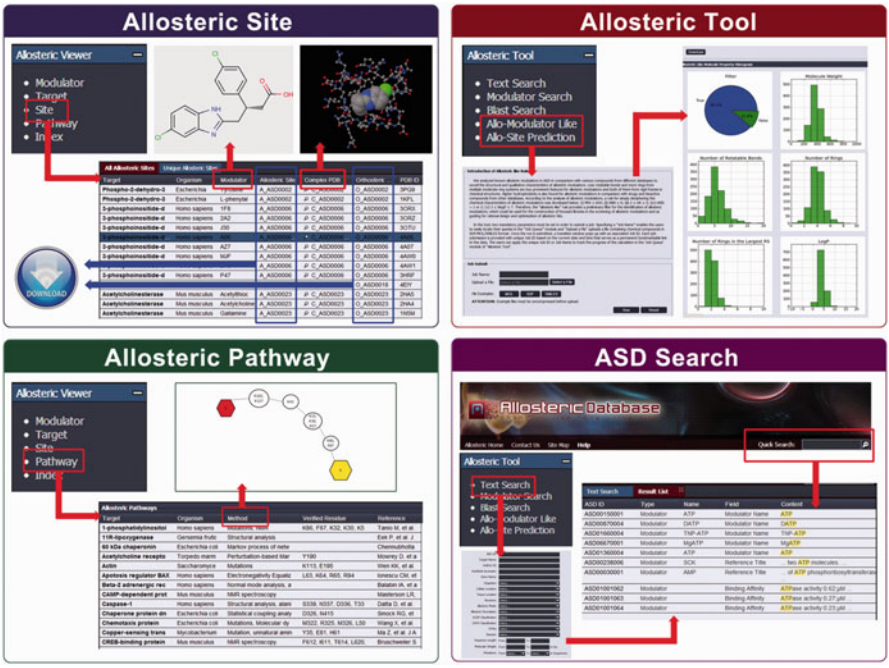


Fig. 4.3 The new features in ASD v2.0

Using the same collection method in ASD v1.0, 1286 allosteric proteins across various species, including kinases, GPCRs, ion channels, peptidases, phosphatases, transcription factors, and nuclear receptors, were curated and deposited. Accordingly, 565 diseases reported to be relevant to the dysregulation of those allosteric proteins were also collected. In this version, the database comprises 22,008 allosteric modulators that are categorized as 15,140 activators, 6207 inhibitors, and 850 regulators.

The presence of druggable, spatially distinct allosteric sites in a wide range of protein families offered a new paradigm for compounds to regulate receptor function [32]. Original allosteric cocrystals encompassing 907 allosteric protein-ligand complexes of 23,009 experimentally verified interactions were acquired from PDB. Afterward, 218 unique allosteric sites with or without bound modulators were visualized to investigate the key interactions and perturbations on the pocket elicited by allosteric binding. In addition, the paired orthosteric sites can also be queried to explore coupling with regulatory sites.

An important aspect underlying allosteric mechanisms is the communication-mediating allosteric and functional sites. The identification of potential allosteric pathways could guide a rational design of physiopathological activity for therapy [12, 14]. Thus, in ASD v2.0, we collected 48 allosteric pathways identified by experimental methods including site-directed mutagenesis, NMR, and theoretical methods such as Markov models and MD simulation. The key residues constituting the signaling pathways were listed.

In addition to the two features, Allo-Modulator Like and Allo-Site Prediction are deployed as two standalones to predict allo-like traits for small molecules and allosteric sites for a protein structure, respectively. In this version of the ASD, the repository was enhanced with structural perspectives on illuminating allosteric regulation, which may boost the development of computational tools for allosteric proteins and their modulators.

4.2.3 ASD v3.0

In this update, we focused on the characterization of allosteric drug action, allosteric cellular networks, and well-established therapeutic protein families such as protein kinases and GPCRs [49].

Building on the previous two versions, with increasing interest in the field of allostery, we expanded the data volume of allosteric proteins and modulators considerably (Table. 4.1). ASD v3.0 contains 1473 allosteric proteins and 71,538 allosteric modulators consisting of 25,339 activators, 33,604 inhibitors, and 13,462 regulators. The increase in modulators is mainly derived from a significant interest in the drug discovery community. Other allosteric annotations and features, including allosteric interactions, sites, pathways, and related diseases, were also augmented.

As a key step to characterize the allosteric control of a molecule system, the desired compound activity can be examined in the detected allosteric site via adjusting the degree of allosteric action exerted on the surrounding residues [41]. To observe the origin of the conformational transition, comparison of the allosteric site before and after binding the modulator is critical. To this end, 1688 allosteric *apo/holo* paired structures in 308 proteins from 107 organisms were constructed in ASD v3.0 to determine the allosteric modulator action. As shown in Fig. 4.4, the superimposition of apo/holo structures of glutamate suggests pronounced backbone displacement induced by binding of the allosteric effector. This module also revealed that the opposite mechanism originated from one allosteric site by triggering different hotspots, raising challenges for allosteric drug design.

Cellular functions can be decoded in terms of signaling networks such as gene transcription networks, metabolic networks, etc. The propagation of allosteric signals is not limited to a single protein structure. The binding of an allosteric modulator to one protein may induce conformational and dynamical changes on adjacent proteins. Moreover, as allosteric signals are transmitted inside the protein structures and over the entire network in an anisotropic and stepwise manner, proteins largely distant from the origin of propagation can also be perturbed [43]. To integrate the dynamic regulation of allosteric proteins and reveal their relationships in functional networks, 261 allosteric networks, in which allosteric proteins function as perturbed nodes and allosteric modulators act as perturbants, were built from the literature through cross-linking and manual calibration. Figure 4.5 displays the “MAPK signaling pathway” allosteric network.

Analysis of collected allosteric information indicates that kinases and GPCRs are two of the largest allosteric protein families that have been extensively exploited

Table 4.1 Total data in different versions of ASD

Data category	ASD v3.0	ASD v2.0	ASD v1.0
Number of all modulators	71,538	22,003	8095
Number of activators	25,339	15,144	4784
Number of inhibitors	33,604	6205	3035
Number of regulators	13,462	854	386
Number of dual activators/regulators	334	50	16
Number of dual inhibitors/regulators	320	55	16
Number of dual activators/inhibitors	257	125	87
Number of multiple activators/inhibitors/regulators	44	30	9
Number of all allosteric sites	1930	907	
Number of all proteins	1473	1261	336
Number of kinases	207	187	46
Number of GPCRs	118	109	48
Number of ion channels	134	119	21
Number of peptidases	59	55	0
Number of phosphatases	30	27	0
Number of transcription factors	55	46	0
Number of nuclear receptors	26	24	0
Number of E-proteins	5	5	2
Number of other proteins	839	689	219
Number of protein-modulator interactions	75,462	23,120	8680
Number of all protein pathways	56	48	
Number of allosteric-related diseases	3350	565	248

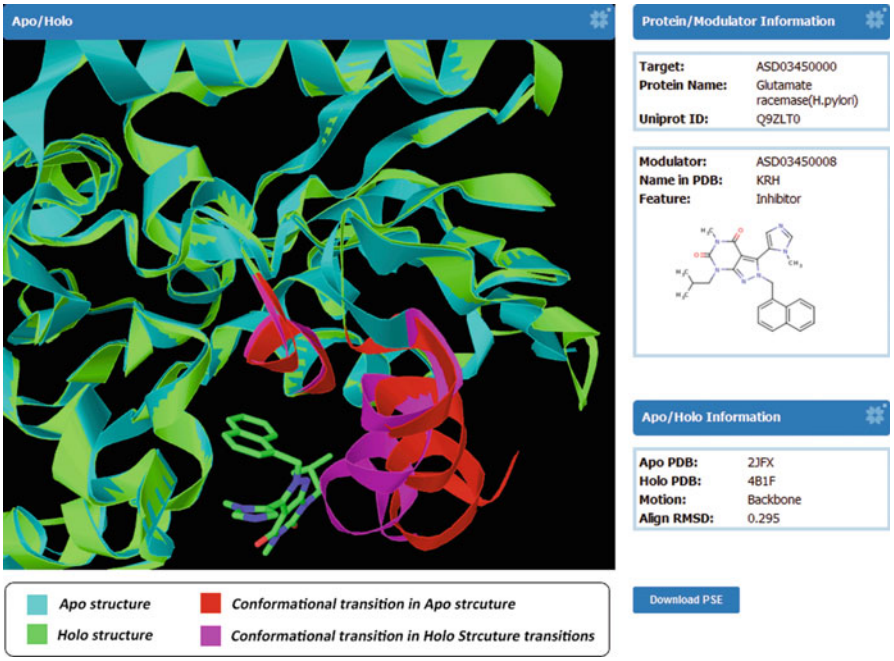


Fig. 4.4 Conformational rearrangements at the allosteric site on the basis of aligning the glutamate structure of *apo* and *holo*

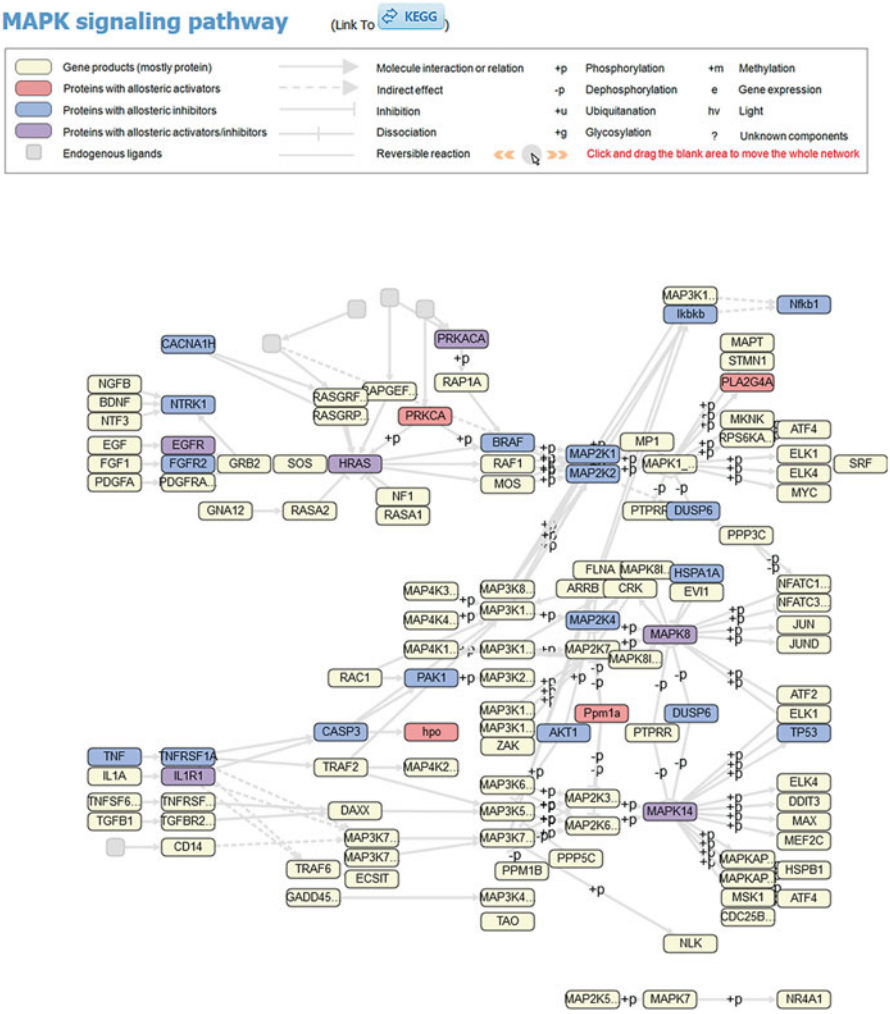


Fig. 4.5 A screenshot of allosteric network "MAPK signaling pathway"

with drug discovery. Unraveling the evolutionary relationship among allosteric sites of these two families could provide a promising paradigm for novel site identification and could help design potent compounds with high subtype selectivity and improved biased signaling. To this end, two allosterome maps in humans, protein kinases, and GPCRs were completely established in terms of multiple sequence alignments and phylogenetic trees. As illustrated in Fig. 4.6, among 76 kinases, 51 members with discovered allosteric modulators are labeled with white circles and 12 members with solved allosteric complex structures are tagged with red flags.

In this ASD update, great efforts were made to describe allosteric action on the structural view, map allosteric cellular networks, and construct protein kinase and GPCR allosteromes.

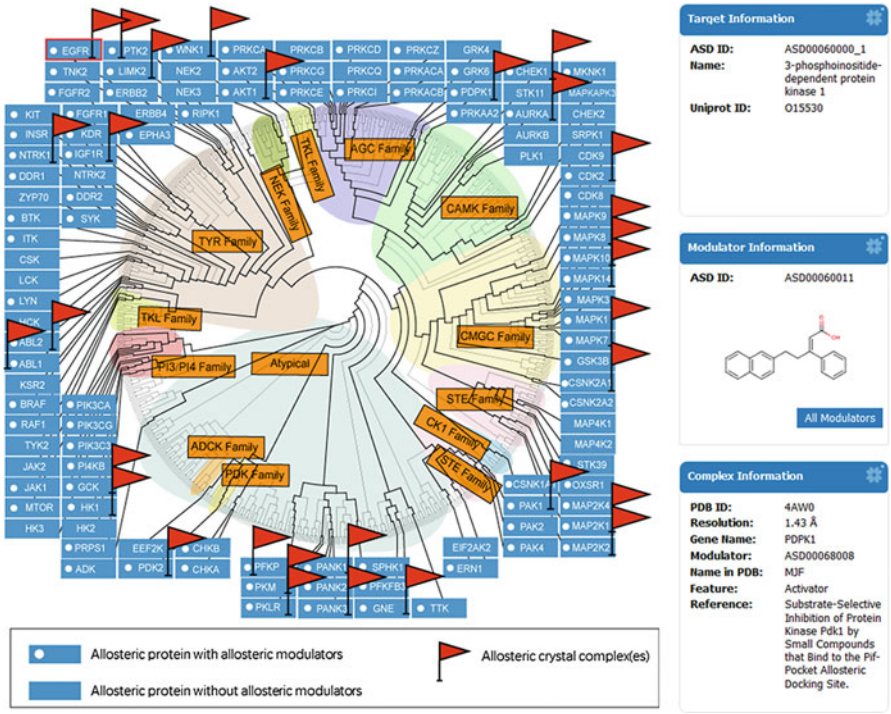


Fig. 4.6 The illustration of protein kinase allosterome

4.3 Usage of the ASD

The main page contains seven items on the top menu: "HOME," "MOLECULES," "FEATURES," "TOOLS," "MISC," "DOWNLOAD," and "HELP," "MOLECULE" contains two submenus, "PROTEIN" and "MODULATOR," and "FEATURES" contains four submenus, "SITE," "PATHWAY," "NETWORK," and "ALLOSTEROME." The overall organization of ASD v3.0 is depicted in Fig. 4.7.

Users can click the "PROTEIN" menu to access Level 1 Protein Browse in the Main Frame. It groups all macromolecules into multilevel tab panels according to the selected class and subclass. All allosteric proteins are classified as Kinase, GPCR, Channel, Peptidase, Phosphatase, Transcription Factor, Nuclear Receptor, E-Protein, and Other. Each page of the tab lists all targets within the subclass. Double-clicking the target name jumps to show Level 2 information of the target. Level 2 of the Protein module encompasses an integrated description of allosteric proteins ranging from sequence, structures, mutations, and modifications. Visualizations of allosteric actions for a protein are available in the Allosteric Mechanism section of the module. Clicking the modulator button shows the details of the apo/holo aligned structures with basic information of the target, the bound

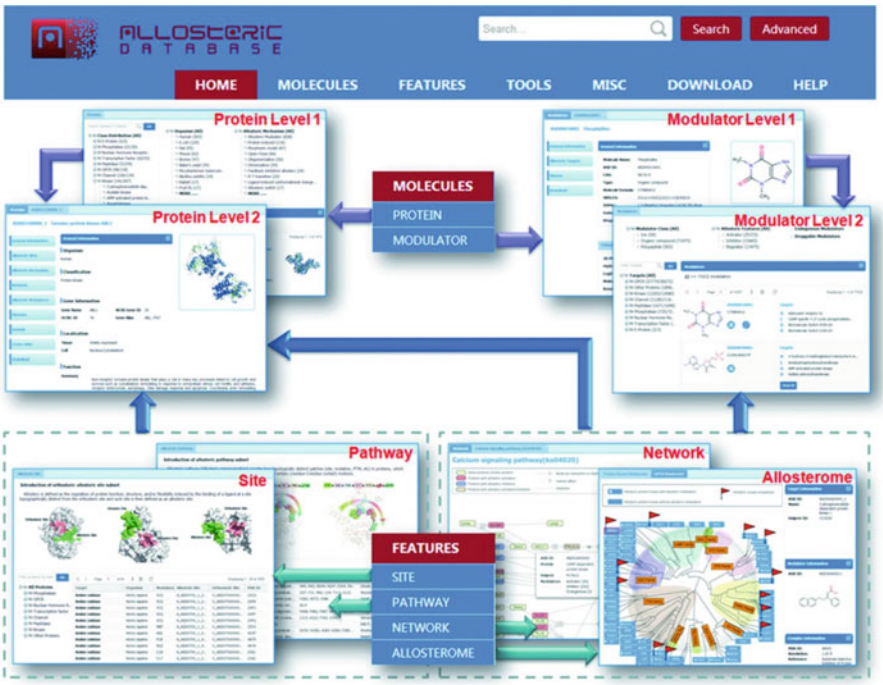


Fig. 4.7 Organization of the web interface in ASD v3.0

modulator, and the extent of conformational rearrangement. All proteins deposited in the ASD are cross-linked to other widely used biological databases including UniProt, GenBank, and PDB.

Clicking the "MODULATOR" menu displays the Level 1 Modulator module in the Main Frame. All modulator entries in four tab panels are defined by the modulator type, namely, All, Activator, Inhibitor, and Regulator. Each page of the tab panel illustrates 10 modulator records with their ASD IDs, 2D structures, and modulating effects for different targets. Clicking the ASD ID jumps to Level 2 information of the modulator, which contains standard molecule formula, 2D/3D structures, physicochemical attributes, and related targets with original references.

The Site module can be accessed by clicking the "SITE" menu under the "FEATURE" menu. The site data are listed in the grid table where the target name, organism, modulator, allosteric site and orthosteric site, and PDB ID are shown. The Pathway module lists all allosteric pathways stored and the information for each pathway includes target, organism, method, verified residues, and references. The Network module is organized as two levels. Level 1 lists all allosteric networks stored in ASD v3.0. Each page of the grid shows five records with network titles. Double clicking a record brings the user to Level 2 of the module. In this page, users can query the details of allosteric networks, in which allosteric proteins are marked with different colors. Users can view the protein details by suspending the

mouse on the node. The two allosterome results can be accessed by clicking the "ALLOSTEROME" menu.

Clicking the gene name reveals detailed information of the target, modulators, and complex structures. Users can download all archive files of modulators and proteins in the download page under the "DOWNLOAD" menu.

4.4 Application Cases

In this section, we focus on four application studies that adopt allosteric information in the ASD to investigate allosteric mechanisms or allosteric drug discovery and design.

4.4.1 GNM Analysis of Hemoglobin and iGluR

The normal mode analysis (NMA) method has become one of the standard technologies to probe protein dynamics [1]. The theory of normal modes is based on the assumption that a given conformation is at the energy minimum and all other conformations can be considered as thermally induced harmonic fluctuations around the minimum. Thus, the energy surface in the vicinity of the minimum is a multidimensional parabola. NMA is commonly applied to detect slow modes with large conformational fluctuations, namely, low-frequency modes, which are representative of global motion and are relevant to protein function. The Gaussian network model (GNM), which is the simplest coarse-grained model based on NMA, regards the entire protein structure as a set of Cα atoms connected through spring-like interactions within a certain cutoff distance [13]. The fluctuations around the equilibrium are distributed in a Gaussian manner. The topology of the network is fully defined by a Kirchhoff matrix, and its diagonalization can acquire the frequency and direction of the mode. Due to its simplicity and efficiency, GNM has been applied to allosteric transitions and allosteric coupling [28, 30, 58].

4.4.1.1 Case Study 1: Hemoglobin

Perhaps the most thoroughly studied example of allostery is hemoglobin (Hb), a tetrameric protein consisting of four subunits (α1, α2, β1, and β2) systematically arranged as a dimer of dimers (α1β1 and α2β2) (Fig. 4.8) [45]. Hemoglobin can bind O_2 reversibly at four heme sites on four subunits and act as an O_2 carrier to transport oxygen from the lungs to all tissues in the body. In the 1960s, it was believed that the allosteric transition of hemoglobin had been well-understood by the determination of the deoxygenated and oxygenated structures, which correspond to the tense (T) and relaxed (R) states of the two-state allosteric model [35]. The structural transition

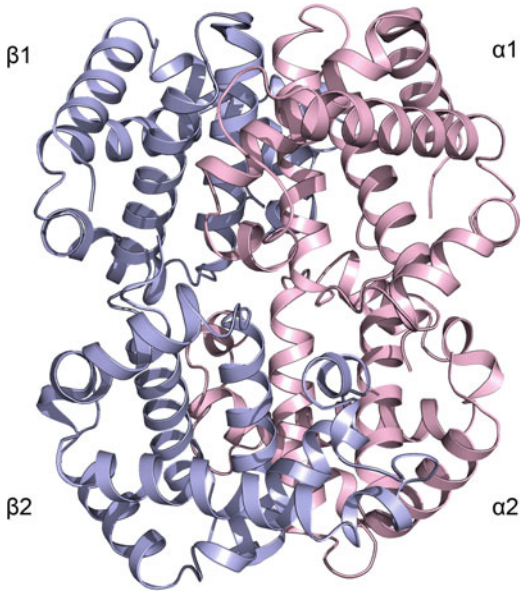


Fig. 4.8 The tetramer structure of hemoglobin contains two α-subunits (light pink) and two β-subunits (blue-white)

from T to R is manipulated in a cooperative manner, as the binding of O_2 to the first subunit increases the O_2-binding affinity of the second and successive subunits. In recent decades, a growing body of evidence suggests the existence of several states in conformational change. Silva et al. detected an alternate quaternary state, R2, for ligated hemoglobin [51]. Maria et al. confirmed that R2 is an oxygenated hemoglobin end state, rather than an R state [47]. Recently, Stefan et al. used the conjugate peak refinement (CPR) algorithm to identify the minimum energy pathway in the transition. They found that the protein was present in an intermediate quaternary state between two distinctive quaternary events, Q1 and Q2 [15]. Naoya et al. determined the structures and functions of the coexisting states and identified a new intermediate form with an intermediate O_2 affinity [50]. These findings suggest that the dynamic paradigm of hemoglobin remains a trending topic in exploring allosteric mechanisms.

To identify regions that are likely to undergo structural change in allosteric transition, Dror and his coworker calculated the GNM for the hemoglobin dataset retrieved from the ASD, including 169 allosteric structures. They found that the α1β1/α2β2 interface is a dynamically variable region, but not at the α1β2/α2β1 and the α1α2/β1β2 interfaces [10]. They represented the GNM modes as curves and utilized a neighbor-joining approach, similar to the Clustal W multiple sequence alignment algorithm, to align multiple curves. The standard deviation was calculated as an indicator of how well the curves were aligned at a specific point. From the alignment of the four slowest modes of the two end states (PDB ID: 1A3N, 1BBB) and all other structures in the dataset, the majority of hemoglobin structures were dynamically similar to the T state rather than the R2 state. Next, the standard deviation was adopted to evaluate variable regions with a threshold of 5.5. All

dynamically variable residues according to the alignment of the first mode were within the α1β1/α2β2 interface. In addition, the second and the third modes accounting for the α1β2/α2β1 and α1α2/β1β2 movements did not reveal dynamical variability. The most variable residues identified on the fourth mode were enriched at the proximity of heme groups at the α1β1/α2β2 and α1/β1 interfaces. Collectively, they concluded that the α1β1/α2β2 interface serves an important role in the allosteric transition from T to R2.

4.4.1.2 Case Study 2: iGluR

Ionotropic glutamate receptors (iGluRs) are ligand-gated cation channels that mediate the majority of excitatory synaptic transmissions in the central nervous system and are involved in brain development and function, including learning and memory formation [63]. iGluR malfunctions are correlated with devastating chronic neurodegenerative symptoms such as Alzheimer's and Parkinson's diseases. iGluRs include three major subfamilies, namely, AMPA receptors, kainate receptors, and N-methyl-D-aspartate (NMDA) receptors, which have different pharmacological profiles and kinetic properties but share similar structural features. Structurally, iGluRs have a unique molecular architecture and contain four domains, namely, the distant amino-terminal domain (ATD); the clamshell-shaped ligand-binding domain (LBD), which binds glutamate or other agonists; the transmembrane domain (TMD), which harbors a cation channel; and the intracellular C-terminal domain (CTD), which is implicated in synaptic targeting [24]. Recent X-ray crystallography studies have revealed the receptor architecture and allosteric modulation of gating [62]. The iGluR structure is organized like a capital "Y," in which the ATD, LBD, and TMD are arranged in layers. The distal ATD forms the upside, the TMD corresponds to the downside, and the LBD is between them (Fig. 4.9, left). The LBD consists of upper lobes and lower lobes, which face the ATD and TMD, respectively (Fig. 4.9, right). Armstrong et al. determined that modulator binding at the LBD evokes the closure of the two lobes that lead to an opening in the channel [2].

Dror explored allosteric effects on the dynamics of the LBD using an improved GNM method on the iGluR dataset in the ASD [53]. Given the homology among the iGluR family, they optimized the aforementioned GNM mode alignment approach using a combination of sequence alignments. Specifically, gaps or insertions were introduced in GNM modes according to the sequence alignment and the sign of all modes were aligned to the first one. The standard deviation, a measurement of how well the fluctuation values are conserved, was calculated for points that presented half of the residues. Initially, they calculated GNM modes for the 13 LBD structures with the endogenous ligand and various allosteric effectors collected from the ASD. The results indicated that the second mode describing the clamshell-like motion of the upper and lower lobes is affected by allosteric modulators. Two regions, T91-P105 and D216-L236, which are located at the interface between the upper and lower lobes, showed pronounced dynamical variability and were characterized

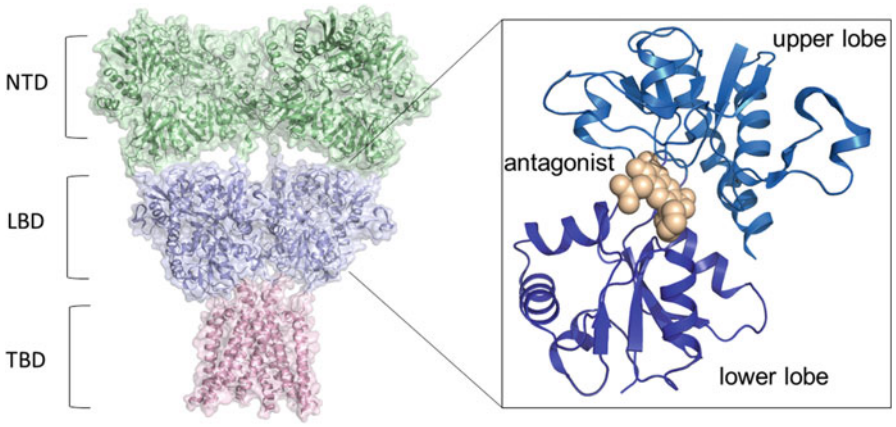


Fig. 4.9 Left: The crystal structure of iGluR (PDB ID: 3KG2) incorporates ATD (pale green), LBD (blue-white), and TBD (light pink). Right: The zoomed-in version of the LBD. The upper lobe and lower lobe are displayed with cartoons in marine and purple, respectively. The antagonist is highlighted with a sphere in wheat

with an SD $\geq$ 5.9. Moreover, as the two regions are characterized by experimentally low B-factors, the dynamical variability cannot result from flexibility or thermal fluctuations. Therefore, the clamshell-like movement may elicit dynamical changes at these regions. The results of the GNM analysis performed on structures without accounting for the ligand and effectors are very similar. Taken together, the allosteric effect induced by agonist/antagonist binding at the LBD causes dynamical variability at the interfacial region between the upper and lower lobes.

Both studies used an elegant GNM analysis to probe the allosteric mechanisms of hemoglobin and iGluRs, which provide new insight into the dynamic paradigm of allostery. The increase of allosteric protein information deposited in the ASD is useful for the application of computational approaches that examine allostery.

4.4.2 Fragment-Based Drug Design of Allosteric Modulators of $mGlu_5$

G-protein-coupled receptors (GPCRs) are the largest superfamily of transmembrane proteins encoded by the human genome. GPCRs can be stimulated by a diverse variety of endogenous ligands including odors, neurotransmitters, hormones, and pheromones and can be coupled with sensor molecules to elicit signaling transduction cascades [59]. The GPCR superfamily, comprising more than 800 proteins, can be further classified into four major subfamilies, A, B, C, and F (frizzled), in terms of sequence homology [31]. The metabotropic glutamate receptors, a family of class C GPCRs, are activated by the endogenous neurotransmitter glutamate. The receptor

falls into three clusters based on protein structure and physiological activity, namely, group I, group II, and group III. mGlu5 is a member of group I and plays an important role in modulating neuron activity and plasticity [29]. $mGlu_5$ represents a valuable therapeutic target, and negative allosteric modulators (NAMs) that could alleviate receptor activation by endogenous glutamates are under clinical evaluation for the treatment of diseases ranging from fragile X syndrome to depression [17].

A substantial argument with respect to allosteric modulators for targets deposited into the ASD is that it could facilitate and lead optimization using conventional structure-based medicinal approaches and have potential to generate novel compounds by applying state-of-the-art in silico technologies. Fragment-based drug design (FBDD) is an attractive strategy that is increasingly applied in the pharmaceutical community and is used for reducing the detrition rate in the early phase of the process and provides novel compounds [6, 11, 23]. Fragments, which are generally low molecular weight and moderately hydrophobic organic molecules, can be grown, linked, and merged with other fragments to form large molecules with improved potency [26]. Some studies have suggested that the FBDD campaign could result in drug-like molecules, where traditional drug development approaches have struggled [37].

Recently, Bian et al. harnessed a purely computational fragment-based method to synthesize novel allosteric modulators for $mGlu_5$ on the basis of a fragment library built in ASD v2.0 [3] (Fig. 4.10). As a result, they identified synthetic compounds that shared high similarity with patented modulators of $mGlu_5$, which indicates the feasibility of the methodology. The discovery of other compounds without reported activities could be further exploited.

Initially, 27,262 allosteric modulators of GPCRs in ASD v2.0 were broken into fragments using retrosynthetic combinatorial analysis procedure (RECAP). The "Rule of Three" instead of "Rule of Five" was used as a guideline to select ideal fragments. Next, fragments were docked into the allosteric binding site defined by the $mGlu_5$-mavoglurant complex structure (PDB ID: 4OO9) (Fig. 4.11). The fragment docking study revealed that fragments were enriched at two distinct subpockets

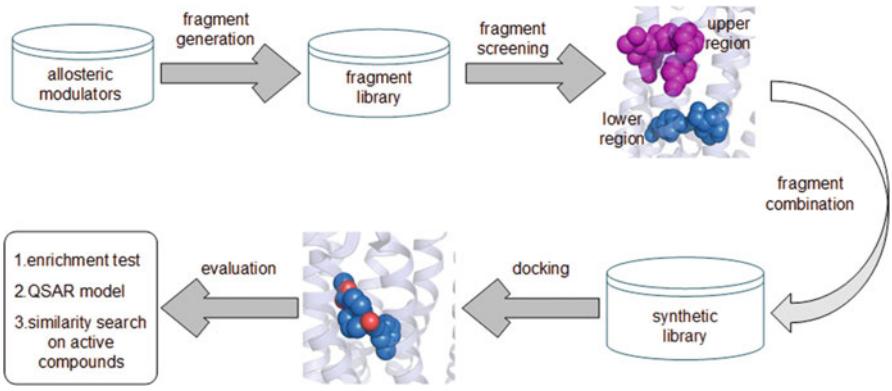


Fig. 4.10 Workflow of fragment-based method to generate novel allosteric modulators for $mGlu_5$

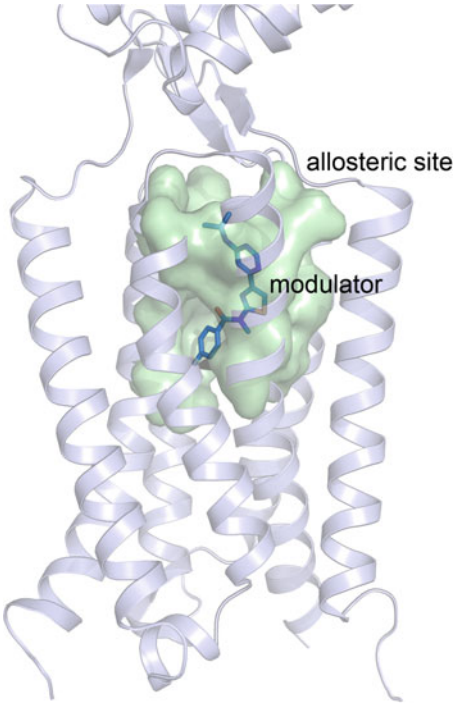


Fig. 4.11 The allosteric site of $mGlu_5$ (PDB ID: 4OO9). The $mGlu_5$, allosteric site, and allosteric compound are displayed in cartoon, surface, and stick modes, respectively

at the allosteric site, the upper and bottom regions. Subsequently, RECAP synthesis was adopted to generate novel compounds by combining the fragments with the highest score in both regions for increasing binding affinity.

The authors conducted a series of evaluation procedures for the top-ranked synthetics. First, in order to verify the docking algorithm for ranking the lead compounds, a benchmark set containing 50 decoys against NAM mavoglurant was generated for enrichment analysis. The docking evaluation for the top 20 synthesized compounds, 50 decoys, and mavoglurant indicates that newly generated compounds can be enriched with NAM and were distinct from the challenging decoys. Next, a QSAR model, which uses the relationship between LogKi and theoretical descriptors of molecules, was built with a training set of 66 analogs of 1,2-diphenylethyne acquired from the ASD. The model accuracy was cross-validated, and both the predicted LogKi of the 20 top-ranked synthesized compounds and the docking scores indicated that these compounds were allosteric modulators of $mGlu_5$. In addition, the top 20 in silico hits included structurally diverse compounds with reported and unreported scaffolds in patents. A structure similarity search revealed that a series of patented scaffolds with reported mGluR allosteric activities could reemerge through computational design.

One of the goals of constructing the ASD was to facilitate allosteric drug design, especially in promising targets including GPCRs and kinases. The above study highlights that the combination of the ASD and fragment-based approaches can be used to design novel allosteric compounds for $mGlu_5$ and other targets. Additionally,

this study shows the potential for computational breakthroughs in propelling the ongoing process of allosteric drug design.

4.4.3 *Pharmacophore Modeling and Screening for Allosteric Modulators of $mGlu_1$*

Metabotropic glutamate receptor 1, another member of group I mGlu receptors, has been proposed to be an attractive drug target for the treatment of cerebellar ataxia and schizophrenia [8]. Endeavors in designing negative allosteric modulators of $mGlu_1$ have revealed the complex crystal structure of the receptor with a small-molecule NAM 4-fluoro-N-(4-(6-(isopropylamino)pyrimidin-4-yl)thiazol-2-yl)-N-methylbenzamide (FITM) [60]. The allosteric cocrystal, together with other verified allosteric modulators in the ASD, could be instrumental in guiding the discovery of new hits.

A pharmacophore model can be defined as an ensemble of chemical features that is essential for specific protein-ligand binding. Generally, elucidating the pharmacophore hypothesis involves either exploring the possible interactions of the receptor/complex structure or aligning a set of active ligands and extracting common features responsible for the activity [61]. Once the pharmacophore model is generated, it can be used as a query to search chemical structure libraries for potential hits. Pharmacophore-based virtual screening has represented an extremely efficient method owing to reduced search time and the discovery of novel compounds with different scaffolds and functional groups [27, 48].

Recently, Jiang et al. leveraged NAMs of $mGlu_1$ in the ASD to generate pharmacophore hypotheses and applied them to discover potential Chinese herbs [21] (Fig. 4.12). First, they collected 74 NAMs of $mGlu_1$ from the ASD and

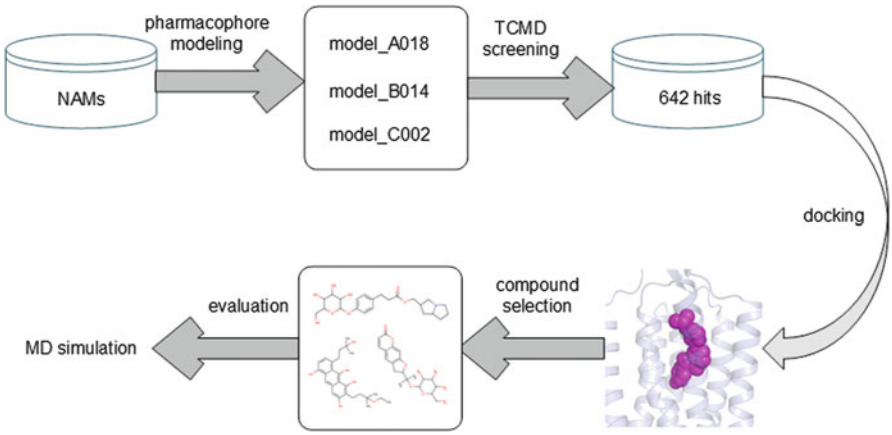


Fig. 4.12 Workflow of pharmacophore modeling and screening for discovering Chinese herbs targeting $mGlu_1$

clustered them into three groups based on structure similarity. Six highly active molecules from each group were selected as a training set to create up to 20 pharmacophore models per group. Then, 74 active compounds and 222 inactive compounds formed a test set to validate the pharmacophore models mapped for the three structure types. Three pharmacophores achieving improved specificity were chosen to screen the Traditional Chinese Medicine Database (TCMD), which contains 233,033 natural compounds from 6735 medicinal plants. Consequently, 642 potential hits were retrieved that met the pharmacophore queries and had drug-like traits. To elucidate the accommodation of these compounds for $mGlu_1$, 642 hits were docked into the FITM allosteric binding site in the crystal complex (PDB ID: 4OR2) (Fig. 4.13). Three compounds, thesinine-4′-O-β-d-glucoside, nigrolineaxanthone-P, and nodakenin, with a high docking score and high pharmacophore fit score were considered as potential novel compounds binding to $mGlu_1$. Additionally, a molecular dynamics simulation revealed three hits that form similar interactions at the binding pocket with FITM under dynamic conditions, further indicating that the compounds have negative allosteric effects on $mGlu_1$.

Although the hits were not verified as active by experimental methods, the case demonstrates that knowledge-driven allosteric compound discovery is an efficient and rational alternative to the "high-throughput" docking-based screening approach. Moreover, the application of knowledge-based methods to disease-related targets appears particularly worthwhile, given the existence of assay and structure data in the ASD and the challenges of drug candidates in the preclinical phase.

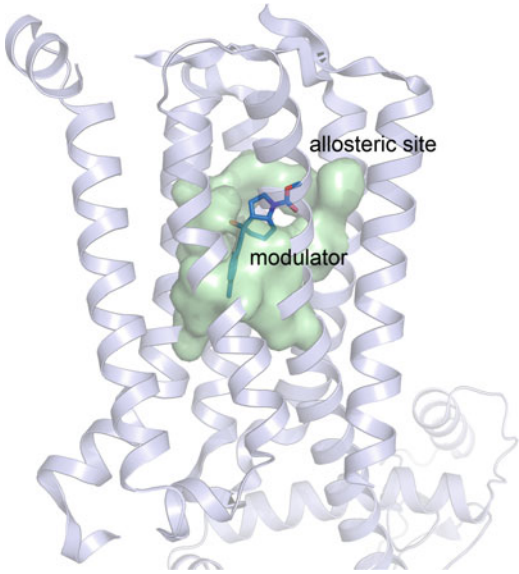


Fig. 4.13 The allosteric site of $mGlu_1$ (PDB ID: 4OR2). The figure scheme follows that of Fig. 4.10

4.4.4 Allosteric Modulator Characterization

Orthosteric compounds block the active site and hinder protein function activity. Consequently, allosteric drugs bind at the allosteric site to remodel the energy landscape and regulate protein function [41]. Due to different modes of action, allosteric modulators have formed a conceptually distinct class of compounds for targets of interest [54]. In the past decade, allosteric drug discovery has gained momentum because it not only provides highly selective molecules but also presents a feasible strategy for designing compounds targeting therapeutic proteins. Unfortunately, specific proteins of which the active site has undesirable physicochemical properties make them potentially untargetable [34, 44]. Arguably, the amount of allosteric modulators in ASD v3.0 has increased ~800% compared to that in ASD v1.0. The current deluge of allosteric modulators could foster the characterization of methodology for designing allosteric drugs.

Wang et al. performed a systematic analysis on the structural and physicochemical traits of allosteric modulators by the comparison of molecules from various chemical databases [56]. They selected 3916 diverse allosteric modulators after applying fingerprint clustering to the original modulator set obtained from the ASD. Other compounds were collected from the Available Chemicals Directory (ACD), Comprehensive Medicinal Chemistry (CMC) dataset, Chinese Natural Product Database (CNPD), DrugBank, MDDR, and NCI Open Database. Eighteen metrics widely used for drug discovery were calculated to evaluate compounds in the ASD and the other datasets. Next, an association rule was used to find an optimal combination of the most highly correlated attributes for retrieving allosteric molecules. The rule was established as follows: (i) molecular weight ≤ 600; (ii) number of rotatable bonds ≤ 6; (iii) $2 \leq$ number of rings ≤ 5; (iv) number of rings in the largest ring system $= 1$ or 2; and (v) $3 \leq$ SlogP ≤ 7.

These results indicate that fewer rotatable bonds and more rings in moderate ring systems comprise the primary features of allosteric small molecules, which suggests the rigid configuration of chemical structures. In addition, higher hydrophobicity was also found to be a distinct character for allosteric-likeness.

Recently, Richard D. Smith and his collaborators proposed that allosteric ligands tend to be more aromatic and rigid than competitive ligands based on a comparison of molecule data from the recently released version of the ASD and the CHEMBL database [52]. However, they contradicted the above study that had indicated allosteric ligands tended to be more hydrophobic.

The authors started by collecting allosteric and competitive ligand data. The allosteric set consisted of 70,219 unique allosteric ligands including 67,749 modulators from ASD v3.0 and 2470 from CHEMBL, covering 1048 receptors. The competitive set had 9511 unique ligands from CHEMBL, targeting 860 proteins. To remove redundancy, proteins were grouped by sequence similarity and within each protein family; all ligands were clustered by structure similarity using the extended connectivity chemical fingerprints. The clustering process was performed at four levels of thresholds, namely, protein similarity/ligand similarity $= 100\%$,

90%, 75%, and 60%. Twenty-nine physicochemical properties including atom counts, bond counts, physical properties, and drug−/lead-like characteristics were calculated. Statistically significant differences in physical properties had Wilcoxon p-values <0.0001 and no overlap in the 95% confidence intervals of the medians determined from 100,000 bootstrapped samples.

They found that the combination of increased aromatic atoms and decreased number of rotatable single bonds in allosteric ligands indicates that these molecules are more rigid. Additionally, the difference in the typical hydrophobicity metric, SlogP, was statistically marginal in the allosteric and competitive sets.

These insights into the intrinsic nature of allosteric modulators could allow researchers in the drug discovery field to estimate whether a hit compound has a potential to allosterically bind to the receptor and emerge as an empirical principle to guide the optimization of the chemical structure. With additional curated allosteric modulators deposited into the ASD, the depiction of those modulators could be more informative, which would benefit the whole community.

4.5 Conclusion

The ASD has experienced three updates and is continuously maintained as an integrated public repository of curated information incorporating allosteric proteins, modulators, sites, pathways, and networks. The abundance of allosteric protein data in the ASD has allowed the community to exploit allosteric phenomena at various levels, from atom movement in a single structure to the cascades of a complex network. Furthermore, the ASD offers an increased body of allosteric modulator and site data, which, coupled with computational tactics, would reveal new opportunities for rational allosteric drug design, especially for undruggable proteins. In subsequent releases of the ASD, we will focus on tracing the evolutionary origin of allostery across species. With the rapid improvement of next-generation sequencing, investigating the effect of mutations occurring at allosteric sites may provide new insight into clinical syndromes, physiological abnormalities, and disorders.

References

1. Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M (2005) Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. Protein Sci 14(3):633–643
2. Armstrong N, Gouaux E (2000) Mechanisms for activation and antagonism of an AMPA-sensitive glutamate receptor: crystal structures of the GluR2 ligand binding core. Neuron 28(1):165–181
3. Bian Y, Feng Z, Yang P, Xie XQ (2017) Integrated in silico fragment-based drug design: case study with allosteric modulators on metabotropic glutamate receptor 5. AAPS J 19(4):1235–1248

4. Changeux JP, Christopoulos A (2016) Allosteric modulation as a unifying mechanism for receptor function and regulation. Cell 166(5):1084–1102
5. Changeux JP, Edelstein SJ (2005) Allosteric mechanisms of signal transduction. Science 308(5727):1424–1428
6. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate–a historical perspective. Drug Discov Today 14(13–14):668–675
7. Congreve M, Oswald C, Marshall FH (2017) Applying structure-based drug design approaches to allosteric modulators of GPCRs. Trends Pharmacol Sci 38(9):837–847
8. Conn PJ, Pin JP (1997) Pharmacology and functions of metabotropic glutamate receptors. Annu Rev Pharmacol Toxicol 37:205–237
9. Cuff AL, Sillitoe I, Lewis T, Redfern OC, Garratt R, Thornton J, Orengo CA (2009) The CATH classification revisited–architectures reviewed and new ways to characterize structural divergence in superfamilies. Nucleic Acids Res 37(Database issue):D310–D314
10. Davis M, Tobi D (2014) Multiple Gaussian network modes alignment reveals dynamically variable regions: the hemoglobin case. Proteins 82(9):2097–2105
11. de Kloe GE, Bailey D, Leurs R, de Esch IJ (2009) Transforming fragments into candidates: small becomes big in medicinal chemistry. Drug Discov Today 14(13–14):630–646
12. Dokholyan NV (2016) Controlling allosteric networks in proteins. Chem Rev 116(11):6463–6487
13. Erman B (2006) The gaussian network model: precise prediction of residue fluctuations and application to binding problems. Biophys J 91(10):3589–3599
14. Feher VA, Durrant JD, Van Wart AT, Amaro RE (2014) Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol 25:98–103
15. Fischer S, Olsen KW, Nam K, Karplus M (2011) Unsuspected pathway of the allosteric transition in hemoglobin. Proc Natl Acad Sci U S A 108(14):5608–5613
16. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. Nat Chem Biol 4(8):474–482
17. Gregory KJ, Dong EN, Meiler J, Conn PJ (2011) Allosteric modulation of metabotropic glutamate receptors: structural insights and therapeutic potential. Neuropharmacology 60(1):66–81
18. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39(Database issue):D663–D669
19. Huang Z, Mou L, Shen Q, Lu S, Li C, Liu X, Wang G, Li S, Geng L, Liu Y, Wu J, Chen G, Zhang J (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42(Database issue):D510–D516
20. Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, Gao J, Hong Y, Li Q, Ni D, Xu J, Chen G, Zhang J (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. Nucleic Acids Res 46(W1):W451–W458
21. Jiang L, Zhang X, Chen X, He Y, Qiao L, Zhang Y, Li G, Xiang Y (2015) Virtual screening and molecular dynamics study of potential negative allosteric modulators of mGluR1 from Chinese herbs. Molecules 20(7):12769–12786
22. Johnstone S, Albert JS (2017) Pharmacological property optimization for allosteric ligands: a medicinal chemistry perspective. Bioorg Med Chem Lett 27(11):2239–2258
23. Joseph-McCarthy D, Campbell AJ, Kern G, Moustakas D (2014) Fragment-based lead discovery and design. J Chem Inf Model 54(3):693–704. https://doi.org/10.1021/ci400731w
24. Kew JN, Kemp JA (2005) Ionotropic and metabotropic glutamate receptor structure and pharmacology. Psychopharmacology 179(1):4–29
25. Koshland DE Jr, Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5(1):365–385
26. Kumar A, Voet A, Zhang KY (2012) Fragment based drug design: from experimental to computational approaches. Curr Med Chem 19(30):5128–5147

27. Leach AR, Gillet VJ, Lewis RA, Taylor R (2010) Three-dimensional pharmacophore methods in drug discovery. J Med Chem 53(2):539–558
28. Li H, Chang YY, Yang LW, Bahar I (2016) iGNM 2.0: the Gaussian network model database for biomolecular structural dynamics. Nucleic Acids Res 44(D1):D415–D422
29. Lindsley CW, Emmitte KA, Hopkins CR, Bridges TM, Gregory KJ, Niswender CM, Conn PJ (2016) Practical strategies and concepts in GPCR allosteric modulator discovery: recent advances with metabotropic glutamate receptors. Chem Rev 116(11):6707–6741
30. Liu Y, Bahar I (2010) Toward understanding allosteric signaling mechanisms in the ATPase domain of molecular chaperones. Pac Symp Biocomput 2010:269–280
31. Lu S, Zhang J (2018) Small molecule allosteric modulators of G-protein-coupled receptors: drug-target interactions. J Med Chem 62(1):24–45. https://doi.org/10.1021/acs.jmedchem.7b01844
32. Lu S, Huang W, Zhang J (2014a) Recent computational advances in the identification of allosteric sites in proteins. Drug Discov Today 19(10):1595–1600
33. Lu S, Li S, Zhang J (2014b) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34(6):1242–1285
34. Lu S, Jang H, Muratcioglu S, Gursoy A, Keskin O, Nussinov R, Zhang J (2016) Ras conformational ensembles, allostery, and signaling. Chem Rev 116(11):6607–6665
35. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
36. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508(7496):331–339
37. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1(3):187–192
38. Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 247(4):536–540
39. Nussinov R (2016) Introduction to protein ensembles and allostery. Chem Rev 116(11):6263–6266
40. Nussinov R, Tsai CJ (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305
41. Nussinov R, Tsai CJ (2014) Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol Sci 35(5):256–264
42. Nussinov R, Tsai CJ (2015) The design of covalent allosteric drugs. Annu Rev Pharmacol Toxicol 55:249–267
43. Nussinov R, Ma B, Tsai CJ, Csermely P (2013) Allosteric conformational barcodes direct signaling in the cell. Structure 21(9):1509–1521
44. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM (2013) K-Ras (G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 503(7477):548–551
45. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North AC (1960) Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. Nature 185(4711):416–422
46. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5(4):725–738
47. Schumacher MA, Zheleznova EE, Poundstone KS, Kluger R, Jones RT, Brennan RG (1997) Allosteric intermediates indicate R2 is the liganded hemoglobin end state. Proc Natl Acad Sci U S A 94(15):7841–7844
48. Seidel T, Ibis G, Bendix F, Wolber G (2010) Strategies for 3D pharmacophore-based virtual screening. Drug Discov Today Technol 7(4):e203–e270
49. Shen Q, Wang G, Li S, Liu X, Lu S, Chen Z, Song K, Yan J, Geng L, Huang Z, Huang W, Chen G, Zhang J (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res 44(D1):D527–D535
50. Shibayama N, Sugiyama K, Tame JR, Park SY (2014) Capturing the hemoglobin allosteric transition in a single crystal form. J Am Chem Soc 136(13):5097–5105

51. Silva MM, Rogers PH, Arnone A (1992) A third quaternary structure of human hemoglobin A at 1.7-A resolution. J Biol Chem 267(24):17248–17256
52. Smith RD, Lu J, Carlson HA (2017) Are there physicochemical differences between allosteric and competitive ligands? PLoS Comput Biol 13(11):e1005813
53. Tobi D (2016) Dynamics and allostery of the ionotropic glutamate receptors and the ligand binding domain. Proteins 84(2):267–277
54. van Westen GJ, Gaulton A, Overington JP (2014) Chemical, target, and bioactive properties of allosteric modulation. PLoS Comput Biol 10(4):e1003559
55. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE (2016) Emerging computational methods for the rational discovery of allosteric drugs. Chem Rev 116(11):6370–6390
56. Wang Q, Zheng M, Huang Z, Liu X, Zhou H, Chen Y, Shi T, Zhang J (2012) Toward understanding the molecular basis for chemical allosteric modulator design. J Mol Graph Model 38:324–333
57. Wenthur CJ, Gentry PR, Mathews TP, Lindsley CW (2014) Drugs for allosteric sites on receptors. Annu Rev Pharmacol Toxicol 54:165–184
58. Williams G (2010) Elastic network model of allosteric regulation in protein kinase PDK1. BMC Struct Biol 10:11
59. Wootten D, Christopoulos A, Sexton PM (2013) Emerging paradigms in GPCR allostery: implications for drug discovery. Nat Rev Drug Discov 12(8):630–644
60. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science 344(6179):58–64
61. Yang SY (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15(11–12):444–450
62. Yelshanskaya MV, Li M, Sobolevsky AI (2014) Structure of an agonist-bound ionotropic glutamate receptor. Science 345(6200):1070–1074
63. Zhu S, Gouaux E (2017) Structure and symmetry inform gating principles of ionotropic glutamate receptors. Neuropharmacology 112(Pt A):11–15

Chapter 5
Correlation Between Allosteric and Orthosteric Sites

Weilin Zhang, Juan Xie, and Luhua Lai

Abstract Correlation between an allosteric site and its orthosteric site refers to the phenomenon that perturbations like ligand binding, mutation, or posttranslational modifications at the allosteric site leverage variation in the orthosteric site. Understanding this kind of correlation not only helps to disclose how information is transmitted in allosteric regulation but also provides clues for allosteric drug discovery. This chapter starts with an overview of correlation studies on allosteric and orthosteric sites and then introduces recent progress in evolutionary and simulation-based dynamic studies. Discussions and perspectives on future directions are also given.

Keywords Correlation · Evolutionary analysis · Statistical coupling analysis · Elastic network model · Molecular dynamics · Two state Go model · Rigid-body simulation · Community analysis · Mutual information

5.1 Overview

Allostery can be described as the phenomenon where the function of a protein is tuned by perturbation at a distant site [9]. Despite many studies, how the allosteric site communicates and influences the biologically functional (orthosteric) site remains elusive. Many efforts have been carried out toward this direction, aiming to disclose the correlation between allosteric site and orthosteric site. Here, "correlation" refers to the incident that ligand binding, mutation, or posttranslational modifications at an allosteric site leverage variation in the orthosteric site.

Two general types of approaches have been used to study the correlation (Fig. 5.1). The first one is based on evolutionary analysis. As allosteric effects are

W. Zhang · J. Xie · L. Lai (✉)
College of Chemistry and Molecular Engineering, Peking University, Beijing, China

Center for Quantitative Biology, AAIS, Peking University, Beijing, China
e-mail: lhlai@pku.edu.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_5

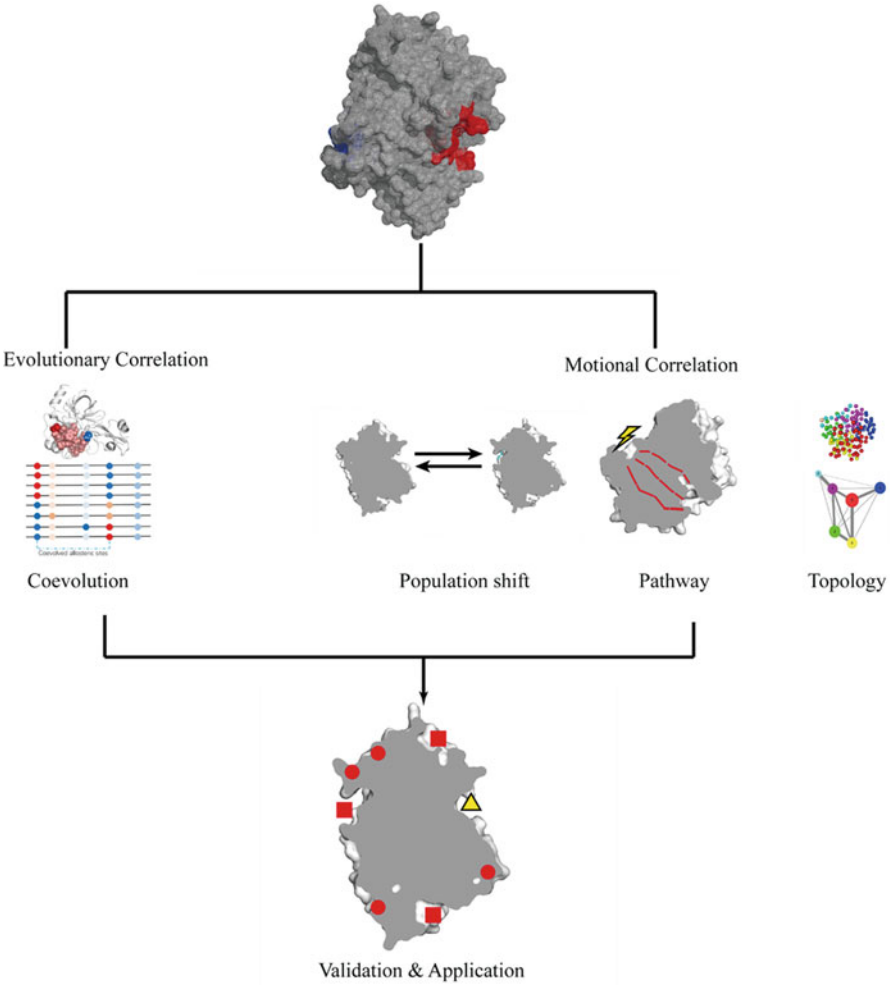


Fig. 5.1 Overview of correlation studies on allosteric and orthosteric sites

evolved through time [50], evolutionary correlation could be deduced by examining the related sequences. The second one is based on dynamic analysis as the motions of allosteric site and orthosteric site are believed to be highly coupled. These motions could be obtained by computer simulations or inferred from experimental studies. Either equilibrium-based or non-equilibrium-based simulation methods have been used at atomic or coarse-grain level. Based on the simulations, motion correlation, network analysis, and information propagation pathways can be deduced. In the present chapter, recent progresses in both evolutionary and dynamic correlation analysis will be introduced.

5.2 Evolutionary-Based Analysis on the Correlation Between Orthogonal Site and Allosteric Site

The accumulation of protein sequence data provides opportunity to understand how fitness factors like the key constraints in maintaining correct folding, stable structure, and specific functions shape the evolution. Practically, an ensemble of homologous sequences are chosen to make a multiple sequence alignment (MSA), and a correlation matrix between positions in sequences are computed. Different algorithms for information extraction from MSA have been developed. For example, the spatial constraints that reflect direct contacts are used in the prediction of protein structures [27]. For allostery which includes certain cooperative action of a number of residues in the protein, deduction of the global constraints is not straightforward. Ranganathan and coworkers developed the statistical coupling analysis (SCA) method, which represents the most successful evolutionary approach for allostery study in the past decade [8, 20, 23, 48, 50].

As in other sequence-based methods, two types of data from the MSA are processed. Position-specific conservation belongs to the first-order statistics. The evolutionary conservation at one position is in the form of the Kullback-Leibler relative entropy, which is estimated from the deviation of the computed distribution of amino acids at this position from a random distribution expected by neutral drift. When mapping the positional conservation result to protein structure, the most conserved residues tend to be positioned in the hydrophobic core or on the surface as functional sites. The conserved correlations between positions belong to the second-order statistics. Based on the alignment, the natural measure of the correlation of two types of residues at two distinct positions could be given by the subtraction of their independent frequency from their joint frequency. A covariance matrix will be generated in this way and could be studied by mutual information for their statistical dependency. However, this kind of treatment assumed the independence from the calculated frequency of each site and therefore did not properly treat the evolutionary relevance of observing these frequencies. To better incorporate such effect, in the current version of SCA [49], the positional conservation, which is in the form of the relative entropies, is used to compute conserved correlation over an ensemble of alignments. In practices, such alignments could be made by bootstrap sampling of the original alignment. Finally, compared to the original covariance matrix which is based on one MSA, a weighted covariance matrix is obtained with the weight functions which are representations of conservation of each amino acid type at each position. Highly conserved correlations are emphasized by such multiplication. However, unlike the interpretation of position-specific conservation, direct mapping of such correlation is not obvious on protein structures.

In the current version of SCA, a spectral decomposition is further performed on the weighted covariance matrix and then compared with the result of many trials of randomized alignments. Only the top eigenmodes that outperform the random are kept. As suggested by the authors, the allosteric effect may include higher-order correlations [49]. Thus, unlike previous heuristical clustering, independent

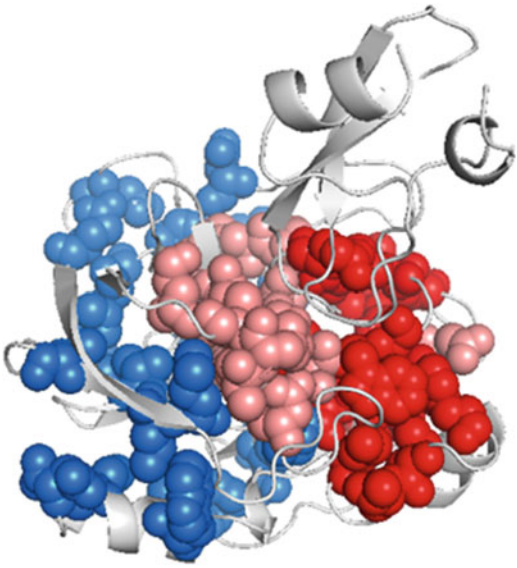


Fig. 5.2 Sectors in S1A proteases discovered by SCA. The figure was prepared according to reference [8] using PyMOL (the PyMOL Molecular Graphics System). The red sector comprises the S1 pocket, the salmon sector comprises the catalytic mechanism of the protease, and the blue sector comprises a ring of residues including a calcium-binding region

component analysis (ICA) is applied to make the top eigenmodes into maximally independent components (ICs). When mapping to the sequence space with evolutionary information, these ICs are related to conserved, differentially evolving functional units in proteins. Ranganathan and coworkers defined "sectors" in their 2009 article as a group of residues each displaying strong intragroup correlations and weak intergroup correlations which are nonuniformly spatially connected [8] (Fig. 5.2).

A significant IC does not need to correspond to an independent sector. In the most recent version of SCA [49], key residues contributed to each IC are first identified via a heuristic statistical approach, and their correlations with other key residues are extracted from the weighted covariance matrix. Compared with previously identified sectors, ICs with more intercorrelations are likely to be grouped together as a sector. Overall, these key ICs could be either a subpart of a single sector or a real sector with a distinct function. As suggested by SCA, sectors are composed of physically connected and coevolving residues. These spatially organized residues link active sites to many surface sites distributed throughout the protein structure. Sectors may be coincided with evolutionarily conserved mechanism that transfers perturbations at the surface to the functional site in allosteric regulation. In mutation experiments, residues within sectors are sensitive compared to those outside of sectors.

In practice, the sequence-analysis results are interpreted by integrating with other forms of data such as structures, phylogenetic tree information, and mutation experiments. Reynolds and coworkers recently proposed a protein engineering approach, the rational engineering of allostery at conserved hotspots (REACH), to introduce novel regulation into a protein by examining potential allosteric site from SCA and have been successfully applied in protein kinase studies [4, 46].

Similar to other sequence-based analysis methods, the evolutionary analysis approach is limited by the available sequences and the selection process. The user should select the input sequences carefully and make a reasonable multiple sequence alignment. Considering the statistical significance, the number of the sequences should be at least five times the length of the protein. Although the evolutionary method can give clues of key residues that are important for the allosteric effect, the signal propagation process cannot be reconstructed.

5.3 Dynamic Analysis on the Correlation Between Orthogonal Site and Allosteric Site

Allosteric effects happen when dynamic correlation patterns between residues in the orthogonal site and allosteric site are altered either through conformational or dynamic changes. The allosteric information can be transmitted through different ways, namely, the "domino" and "violin" style. In the "domino" style, such information is transferred through relay nodes. The binding of an effector will trigger local structural changes that sequentially propagate to the functional site through one dominant pathway which has been observed in the PDZ domain family and G protein-couple receptors, etc. In the "violin" style, the vibrational modes of the protein change without obvious configurational modification [16].

Both atom-based molecular simulations and coarse-grained models have been used to study allosteric communications. For systems with only one structure and no available information on allostery, simulations can be performed and analyzed to estimate the potential of allosteric regulation and identify the possible allosteric sites. Ligand discovery can then be carried out based on the predicted allosteric sites. For systems that are known to be allosterically regulated, simulations with and without allosteric effector binding provide information to understand allosteric regulation and information transmission.

5.3.1 Methods Using Only a Single Protein Structure

As allostery usually involves long-range regulations and global dynamic changes have major contributions, coarse-grained models have advantages in extracting correlation information related to allosteric effect, especially when only the ligand-free structure of a protein is known. The elastic network model (ENM) is a commonly used coarse-grained model to study global dynamics of proteins [3]. As a protein system can be regarded in the vicinity of a minimum, the potential energy is approximated with the sum of harmonic potentials. The simplest ENM is the Gaussian network model (GNM). In this model, each protein residue is represented

as a node and connected by harmonic springs. The interatomic potential energy of the system can be expressed as

$$U = \frac{\gamma}{2}\left[\sum_{i,j}^{N} \Delta R_i \, \Gamma_{ij} \, \Delta R_j\right]$$

where γ is a force constant uniform for all springs, ΔR_i is a vector that represents the displacement of the ith residue from its equilibrium position, and Γ_{ij} is the element of the Kirchhoff (or connectivity) matrix of inter-residue contacts, Γ, defined by

$$\Gamma_{ij} = \begin{cases} -1 & if \;\; i \neq j \;\; and \;\; r_{ij} \leq r_c \\ 0 & if \;\; i \neq j \;\; and \;\; r_{ij} > r_c \\ -\sum\limits_{i, i \neq j} \Gamma_{ij} & if \;\; i = j \end{cases}$$

where r_c is a cutoff distance for spatial interactions and usually taken to be 7 Å for amino acid pairs (represented by their α-carbons) [59]. Based on statistical mechanics, the expectation values of residue fluctuations (mean-square fluctuations, MSF) and cross-correlation could be given by the form:

$$< \Delta R_i \cdot \Delta R_j >= \frac{3k_B T}{\gamma}\left(\Gamma_{ij}^{-1}\right)$$

where k_B is Boltzmann's constant and T is the absolute temperature. The normal modes of GNM are found by diagonalization of the Kirchhoff matrix. The eigenvalues correspond to vibrational frequencies of the normal modes, while the eigenvectors represent the directions of the normal modes. Low-frequency modes (slow modes) reflect large collective or delocalized motions in the protein structure, while high-frequency modes reflect small vibrations in localized regions. The dynamical information is contained in the cross-correlations and can be expressed by a sum over the N-1 non-zero normal modes as:

$$\Gamma^{-1} = U\Lambda^{-1}U^T$$

$$< \Delta R_i \cdot \Delta R_j >= \frac{3k_B T}{\gamma}\sum_{k=2}^{N}\frac{u_{ik} \cdot u_{jk}}{\lambda_k}$$

Because the slow modes have large amplitudes, several top modes always dominate. In fact, the ENM slow modes have been shown to agree well with those predicted by detailed atomic-level force fields and with experimentally determined dynamics [18, 51]. Fast- and medium-frequency modes that contribute less to global motions are important for coordinating finer motions and with their incorporation with slow modes are necessary to analyze the correlated motions of allosteric and orthosteric sites [22, 26, 41, 44].

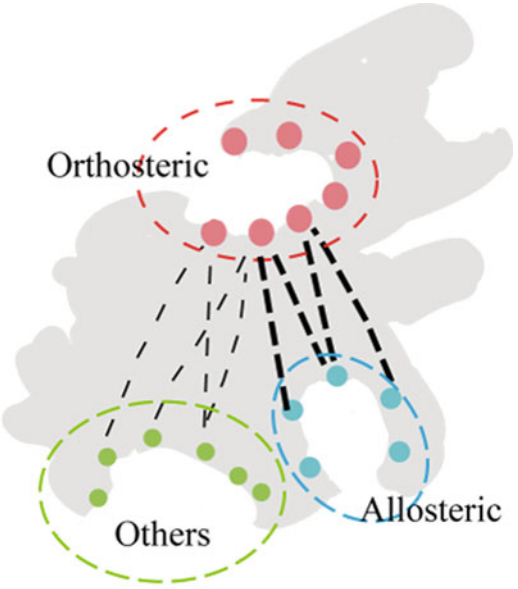


Fig. 5.3 Illustration of motion correlations between orthosteric site and allosteric site. Allosteric site tends to have higher total correlation than non-allosteric site as suggested by CorrSite [26]

To account for the strength of such correlations, Ma et al. proposed an algorithm that only depends on one existing protein structure, that is, the general protein topology [26]. The potential ligand-binding sites are identified by a pocket detection program, CAVITY [61]. The total correlation of each sites with the orthosteric site is the summation of the amplitude of all the cross-correlation terms which belong to these two sites (Fig. 5.3). Then all the total correlation calculated is normalized using the Z-score. Among the 24 known allosteric sites in the dataset used, 22 are correctly predict within the top 3 rank. This method has been incorporated into a web-server CavityPlus [58] with the name of CorrSite. Only the coordinates of the protein structure or the PDB code are needed as input to predict potential allosteric site. Some of the predicted allosteric sites have been verified in experimental studies (Li et al. 2018; [57]).

By using the GNM model, Su et al. used a spherical probe representing allosteric effector binding to sample the protein surface, and additional springs were attached between this effector point and the local residues. The free-energy difference ($\Delta\Delta G$) between the protein-ligand and protein-effector-ligand systems was calculated. In the Hsp70 and GluA2 AMPA receptor, regions with large $\Delta\Delta G$ values were found to correspond to experimental verified allosteric sites [54]. NMA has also been incorporated with machine learning methods for allosteric site prediction [35, 42], which will be covered in other chapters.

While these ENM-based studies have shown that the residue contact network topology is the dominant factor that defines the collective modes, more detailed structural properties are not considered. By adding terms related to distance, the nearest and the second nearest neighbor effects, and backbone hydrogen bond patterns, Lezon and Bahar proposed an improved model (mGNM) that could better

resemble the dynamics revealed by NMR studies [22]. In the original ENM, the correlation is calculated by dot multiplication, which favors collective motions in parallel direction. Yu et al. used a singular value decomposition (SVD) method that analyzes the correlation coefficient of fluctuation dynamics with an arbitrary angle between the correlated directions. This method also worked in cases where two correlated movements are perpendicular to each other which cannot be analyzed using dot product [60].

With atomic details, Amor et al. made an atomic graph with energy-weighted covalent bonds, hydrogen bonds, salt bridges, hydrophobic tethers, and electrostatic interactions [1]. For each bond, a bond-to-bond propensity is defined to account for the total perturbation effect of such bond on all the bonds formed between ligand atom and the residues in active site. Quantile regression was then used to identify significant interactions. Three biologically important proteins, caspase-1, CheY, and h-Ras, were tested, and key allosteric interactions were correctly predicted.

5.3.2 Simulation-Based Methods and Analysis

Atom-based molecular dynamics (MD) simulations let the system evolve in a near natural manner. Currently, simulations up to milliseconds are possible with enough computational resources and large amount of data will be generated from conformation sampling and subsequent analysis.

La Sala et al. examined pocket formation, dynamics, and allosteric communication embedded in microsecond-long MD simulations of three proteins, PNP, A2A, and Abl kinase [19]. As the system evolved, cryptic pockets emerged on protein surface. These hidden pockets were very dynamic at nanosecond timescale as their volume varied from zero to hundreds of cubic angstroms, disappeared, and then appeared again. The analysis algorithm checked temporal exchange of atoms between adjacent pockets along the MD trajectories as a fingerprint. The allosteric communication networks across the protein surface could be detected. These results suggest the prospective use of this dynamic analysis (or long-time MD simulation) to characterize transient binding pockets for structure-based drug design.

A straightforward approach for studying allosteric effect is to simulate the system in both apo and bounded state and compare the difference (Fig. 5.4a). The root mean square fluctuations of each residue is a classical descriptor to characterize the flexibility of the protein. After the binding of effectors, the stability of some regions will change significantly. By comparing the result from active state to apo state, key residues could be located [7].

The difference between apo state and bound state could also be explored by comparing the variation of interactions. In 2015, Ma et al. used the molecular mechanics generalized Born surface area (MM/GBSA) energy decomposition method to calculate the residue-residue interaction energies in the allosteric sites in 17 known allosteric proteins [25]. They found that the interactions in the allosteric effector binding state were different from those before binding, indicating that

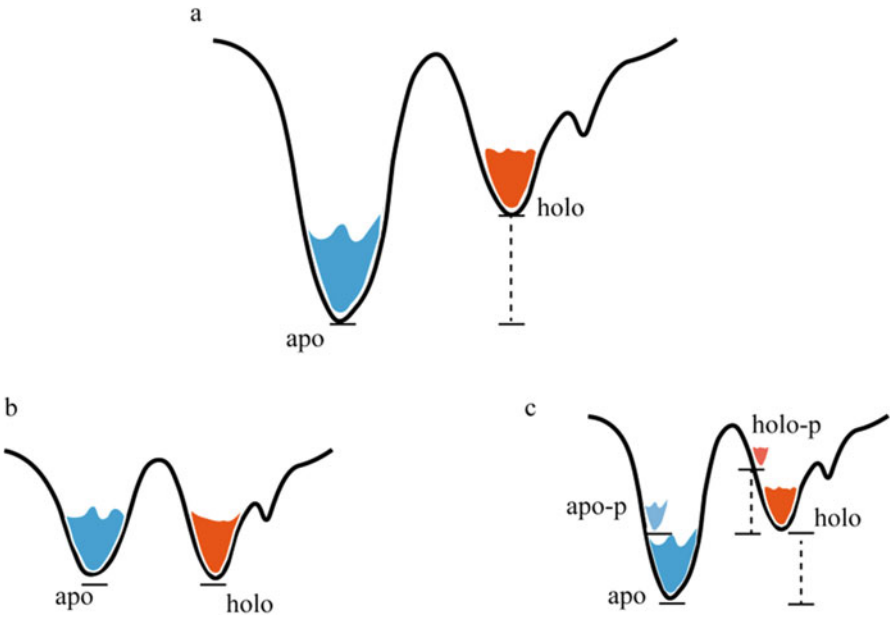


Fig. 5.4 Two-state simulation schemes. (**a**) The system is simulated at two separate states and difference is compared. (**b**) Control parameters are used to assist the population shift process. (**c**) Auxiliary states are introduced in the simulation

redistribution of interactions happened within the allosteric site. This is consistent with a more detailed simulation by Kumawat et al. in 2017 [17]. They studied the PDZ domains, in which ligand binding produces local energetic perturbation that propagates in inter-residue interaction pattern in a domino-like style. These significant changes in the nature of specific interactions and side-chain reorientations together drive the major redistribution of energy. The internal redistribution and rewiring of side-chain interactions here also led to large cancellations resulting in small change in the overall enthalpy of the protein.

Based on the population shift scheme [38], Qi et al. studied allosteric transitions by changing the interaction strength in the allosteric site based on a two-state Go model [39]. An ensemble of the two functional states of a protein was first constructed as reference. Then at a potential allosteric site, the native contact interactions are rescaled by a factor F and new ensembles are constructed (Fig. 5.4b). The population distribution of the new ensemble is monitored. For known allosteric sites, the population was shifted to the other state with a switch-like transition as F varied. As this method relies on the two-state Go potential, it can be applied for cases where structural rearrangements after the binding of the allosteric effector occur. Although not directly correlated with orthosteric site, this method verified a collective behavior of local stability with global configuration variation. Extending this observation to other sites in the proteins tested, novel allosteric sites were predicted. They further experimentally tested the novel allosteric

site predicted in the *Escherichia coli* phosphoglycerate dehydrogenase (PGDH) to see whether novel allosteric inhibitors can be discovered. Virtual screening by molecular docking was performed with a known chemical database and the top-ranking molecules were experimentally tested. Three compounds showed remarkable inhibition activities with the lowest IC_{50} value of 21.6 μM. Mutation studies confirmed that these molecules did bind to the predicted allosteric site. Compared to the known allosteric site used by L-serine, this new allosteric site is much larger in size and can be easily used to design novel chemical regulators [47, 56].

Allostery is considered to be a dynamical process with several residues sitting on the proposed major allosteric pathways. Protein could be considered as semirigid [15]. Changing the dynamics of a key residue may significantly disrupt allosteric pathways. Kalescky et al. introduced a direct way to change single-residue dynamics behavior by treating a specified residue as a single rigid body, i.e., the residue has no internal degrees of freedom which eliminate the reorientation of the side chain and energetic redistribution locally from this residue [10]. Therefore, two more auxiliary states are generated (Fig. 5.4c). The influence of a single residue on allostery can be investigated directly through such rigid-body simulations. The intrinsic dynamics and rigidity-based descriptor, the average residue correlation index (ARC), and the residue correlation similarity (RCS) index are defined to help characterize the difference between states. Two groups of key residues are identified. The first group of residues are proposed as "switches" that are needed to "turn on" the binding effect of protein allostery. When they are rigidified, the difference in cross-correlation patterns between both perturbed simulations of the unbound and bound states is nearly zero. Conceptually, the allosteric effect upon ligand binding in the form of motional dynamics is cancelled when these residues are "turned off." The second group is from a second-order comparison: the difference between both perturbed simulations of the unbound and bound proteins is similar to the difference between unperturbed simulations of the unbound and bound proteins. As the energy flow may cause significant dynamic changes of residues along allosteric pathways, the least disruptive residues for such effects of binding on intramolecular communication are mostly suitable to play a role of "wire" that let the energy or information propagate. They further combined this method with configurational entropy calculation, principal component analysis (PCA), and projection of ensembles onto coherent allosteric modes [11]. The prediction of key allosteric residues has good agreement with experimental studies.

Pfleger et al. constructed a model of dynamic allostery by integrating an ensemble-based perturbation approach with the analysis of biomolecular rigidity and flexibility. The system in complex with allosteric effector is simulated and conformation ensemble is generated from the trajectory. For each conformation which is regarded as a ground state, the corresponding perturbed state is constructed by removing the allosteric effector from the ground state or by mutations no allosteric effector is known. All short-range rigid contacts are counted for both ground state and perturbed state and a chemical potential difference between the two states is deduced. The free energy associated with the change in biomolecular

stability due to removal of a ligand or mutation is calculated by an ensemble average. The validation is done on three distinct biomolecular systems: eglin c, protein tyrosine phosphatase 1B, and the lymphocyte function-associated antigen 1 (LFA-1) domain. In all cases, it successfully identified key residues for signal transmission in agreement with experiments. It also correctly discriminated the positively from negatively cooperative effects in LFA-1 system.

The above MD simulations let the system evolve at equilibrium state and deduced the communication through collective motions. As biological systems are governed by numerous physical interactions, perturbation at one place may be propagated to other parts of the molecule. Nonequilibrium MD simulations could be used to trace the energy flow related to allosteric effect. In anisotropic thermal diffusion method, only the source atoms were coupled to a high-temperature bath. During the nonequilibrium MD simulations, the propagation of the root mean square deviation of residues was recorded, and then signaling pathways could be formulated [40]. In the pump-probe MD simulation method, the atoms were coupled with oscillating forces, and the motions of atoms were analyzed with Fourier transformation to detect energy transduction pathways [53]. Such energy transfer pathways could also be extracted from the evaluation of the cumulative response of the energy source by nonequilibrium correlation functions [14, 36]. Such methods could be suitable to explore fast events like energy transfer in a short timescale without large configuration variation.

5.3.3 Network- and Information-Based Analysis

When analyzing the simulation results, the protein motions are usually mapped to network representations. Each protein residue is depicted by a single node. The length of the edge between two nodes are usually related to their correlation. The metrics of the length could be defined based on correlated motions or the number of contacts. By doing such mapping, various network-analysis algorithms can be used to locate important pathways of communication or group of residues that have strong communications [2].

Community analysis is one of the widely used methods that provide a good visualization of the modular subpart of the whole protein [52]. Sethi et al. studied the allostery in tRNA-protein complexes by community analysis. The pre-transfer and post-transfer states are simulated. After simulating the system, amino acid residues, nucleotides, and the AMP substrate are treated as a single node. Consecutive residues are excluded in the network. Two nonconsecutive nodes are regarded as in contact if any heavy atoms from these two residues are within 4.5 Å of each other in at least 75% of the frames. The weight of an edge between two connected nodes is the probability of information transfer across that edge as measured by the correlation values between the nodes in the simulation. The Girvan-Newman algorithm is next used to identify communities of highly interconnected residues. The optimum community structure is found by maximizing the modularity. This analysis made a

good illustration of how protein dynamics changed when a modulator binds as the community number varied.

In most of the cases, the orthosteric site and allosteric site are located in different communities. The length of a path between distant residues is defined as the sum of the edge weights between the residues on the path. The shortest paths between pairs of residues in different communities are calculated with Floyd-Warshall algorithm and the residues connected by these edges are considered as critical residues for allosteric signal transduction. Meanwhile, the number of paths with length shorter than a certain limit above the shortest distance is a good measure of the path degeneracy in the network. The calculation of these near-optimal pathways is computationally more expensive than the shortest pathway when every residue in allosteric site and orthosteric site should be considered. Van Wart et al. later proposed an efficient program called Weighted Implementation of Suboptimal Paths (WISP) [55]. In this program, physically distant residues and residues that are unlikely to be in the optimal or near-optimal allosteric communication pathways are excluded. Then for the rest of the residues, the shortest path connecting to the source and sink node via this node is calculated by Dijkstra's algorithm. If this path is too long to be an optimal or near-optimal path, there is no way that other paths passing through this node could be an ideal one and this node is eliminated. By such simplification process, notable pathways of allosteric communication could be computed more efficiently. The result is illustrated in Fig. 5.5. Communications between the allosteric and orthosteric site are different in the presence and absence of an effector molecule. When hundreds of suboptimal paths are considered, a large effector-dependent shift in communication between the allosteric and allosteric site becomes apparent.

Correlated motion could also be quantified using mutual information which is related to second-order terms from the configurational entropy expansion. MutInf, first introduced by McClendon et al. [28], uses internal coordinates and torsion angles to derive the mutual information. With statistical correction, the result is robust from multiple short simulations. In the updated version, the Kullback-Leibler divergence metric has been incorporated to improve the sensitivity [29]. For the original example of interleukin-2, population-shift-mediated signal transmission is observed between the allosteric and orthosteric sites via the hydrophobic core. Meng et al. carried out MutInf analysis on 15-LOX and identified a novel allosteric site together with CAVITY. Both activators and inhibitors were discovered that bind to the predicted allosteric site [31, 32].

Higher-order correlations may also be important to address the allosteric effect. LeVine et al. described an N-body Information Theory (NbIT)-based method [21], which identifies communication channels between allosteric sites and orthosteric site through information theory-based analysis of N-body collective motions derived from the configurational entropy. Careful treatment to estimate entropy with clustering is necessary. Good sampling of the system is also required. The authors analyzed two MD simulations of the bacterial transporter LeuT and the results are consistent with previously reported experimental data.

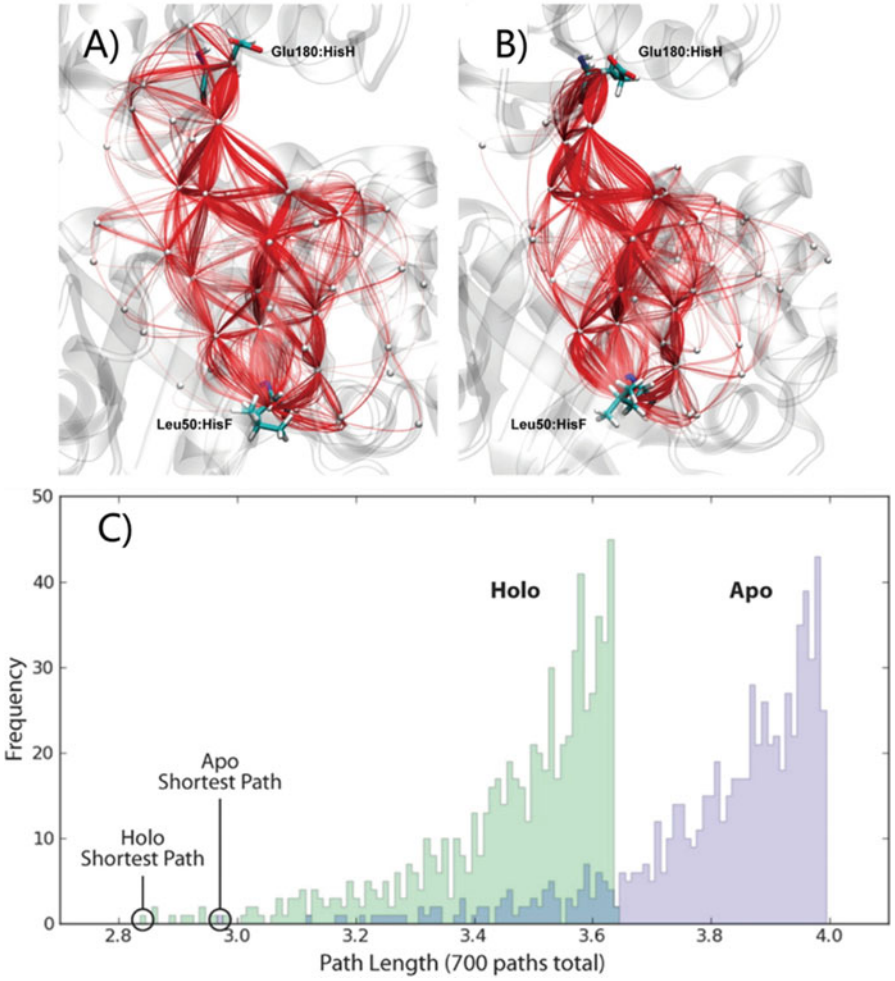


Fig. 5.5 WISP-generated signaling pathways. The 700 shortest paths between Leu50:HisF and Glu180:HisH, shown as red splines, derived from (**a**) the apo trajectory and (**b**) the holo trajectory. (**c**) The path distribution is largely shifted to the left for the holo (allosteric) state

To accelerate the sampling of dynamics, accelerated MD (aMD) could be used [34, 45]. The enhancement of the conformational sampling was done by adding a harmonic boost potential to smoothen the system potential energy surface. Within the timescale of normal MD simulation, the transitions between states could be sampled efficiently. Community-analysis workflow for allosteric effect analysis could also be applied to the trajectory generated by aMD as in the cases of alpha-thrombin and the M2 muscarinic receptor [5, 33].

5.4 Summary and Perspective

A quantitative and predictive description of allostery is fundamental for understanding biological processes. Methods introduced in this chapter can be used to evaluate the significance of individual residues in protein dynamics and allostery and narrow down the number of residues required for further experimental study. Based on correlation analysis between allosteric site and orthosteric site, significant progress has been made to locate potential allosteric sites without prior knowledge of allosteric ligand and key residues. Although successful examples of allosteric ligand discovery based on the identified allosteric sites have been reported, optimization of these ligands remains difficult. Simply increasing binding affinity may not improve its efficacy of allosteric regulation. A clear understanding of the allosteric mechanism is necessary. On the other hand, without enough binding potency, the allosteric effector cannot be specific enough. In the current allosteric regulation studies, the art of balance still depends on "trial and error." Better models need to be developed to quantify correlation, efficacy, and key sites.

Allostery is believed to exist in all the proteins as an adaptive nature of evolution [6]. Most of our current knowledge on allosteric mechanisms are derived from a limited number of model systems. More systems need to be studied to understand the diversified allosteric regulation mechanisms in the protein universe and to generate theoretical models. In addition to physical model-based simulations and statistical analysis, analysis of the large-scale sequencing and omics data provides another dimension for allosteric studies. Posttranslational modification and single nucleotide polymorphism data provide useful information to locate key residues and potential allosteric pathways. On the other hand, allostery provides useful tools to explore functions of proteins and to regulate the biological networks that they are involved in [24, 30, 37, 43].

With the rapid development of structural biology techniques, especially CryoEM, many important large biological assemblies have been solved in recent years. How allostery regulates the function of large biological machines can be carried out. Coarse-grained modeling can be used for such large systems [12]. The discovery that intrinsic disorder regions in protein impose entropy-driven allosteric effect on structured regions has put new challenges to methods developed for treating structured systems [13]. Artificial intelligence will certainly help, as it did in classification of states and identification of key residues [62]. Integration of physical models with large-scale experimental data will provide novel opportunities to address the remaining major questions in understanding allosteric mechanism and in designing better ways of allosteric regulation.

References

1. Amor BRC, Schaub MT, Yaliraki SN, Barahona M (2016) Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. Nat Commun 7:12477–12477
2. Atilgan C, Okan OB, Atilgan AR (2012) Network-based models as tools hinting at nonevident protein functionality. Ann Rev Biophys 41(1):205–225
3. Bahar I, Lezon TR, Yang L-W, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. Ann Rev Biophys 39(1):23–42
4. David P, Orna R, Kimberly AR (2017) An evolution-based strategy for engineering allosteric regulation. Phys Biol 14(2):025002
5. Gasper PM, Fuglestad B, Komives EA, Markwick PRL, McCammon JA (2012) Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. Proc Natl Acad Sci U S A 109(52):21216–21222
6. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57(3):433–443
7. Guo J, Zhou H-X (2016) Protein allostery and conformational dynamics. Chem Soc Rev 116(11):6503–6515
8. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138(4):774–786
9. Hilser VJ, Wrabl JO, Motlagh HN (2012) Structural and energetic basis of allostery. Ann Rev Biophys 41(1):585–609
10. Kalescky R, Liu J, Tao P (2015) Identifying key residues for protein allostery through rigid residue scan. Indian J Chem A 119(9):1689–1700
11. Kalescky R, Zhou H, Liu J, Tao P (2016) Rigid residue scan simulations systematically reveal residue entropic roles in protein allostery. Plos Comput Biol 12(4):e1004893
12. Katkar HH, Davtyan A, Durumeric AEP, Hocky GM, Schramm AC, De La Cruz EM, Voth GA (2018) Insights into the cooperative nature of ATP hydrolysis in actin filaments. Biophys J 115(8):1589–1602
13. Keul ND, Oruganty K, Schaper Bergman ET, Beattie NR, McDonald WE, Kadirvelraj R, Gross ML, Phillips RS, Harvey SC, Wood ZA (2018) The entropic force generated by intrinsically disordered segments tunes protein function. Nature 563(7732):584–588
14. Kobus M, Nguyen PH, Stock G (2011) Coherent vibrational energy transfer along a peptide helix. J Chem Phys 134(12):124518
15. Kornev AP (2018) Self-organization, entropy and allostery. Biochem Soc T 46(3):587–597
16. Kornev AP, Taylor SS (2015) Dynamics-driven allostery in protein kinases. Trends Biochem Sci 40(11):628–647
17. Kumawat A, Chakrabarty S (2017) Hidden electrostatic basis of dynamic allostery in a PDZ domain. Proc Natl Acad Sci U S A 114(29):E5825–E5834
18. Kundu S, Melton JS, Sorensen DC, Phillips GN (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J 83(2):723–732
19. La Sala G, Decherchi S, De Vivo M, Rocchia W (2017) Allosteric communication networks in proteins revealed through pocket crosstalk analysis. ACS Cent Sci 3(9):949–960
20. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R (2008) Surface sites for engineering allosteric control in proteins. Science 322(5900):438–442
21. LeVine MV, Weinstein H (2014) NbIT – a new information theory-based analysis of allosteric mechanisms reveals residues that underlie function in the leucine transporter LeuT. Plos Comput Biol 10(5):e1003603
22. Lezon TR, Bahar I (2010) Using entropy maximization to understand the determinants of structural dynamics beyond native contact topology. Plos Comput Biol 6(6):e1000816
23. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286(5438):295–299
24. Lu S, Li S, Zhang J (2014) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34(6):1242–1285

25. Ma X, Qi Y, Lai L (2015) Allosteric sites can be identified based on the residue–residue interaction energy difference. Proteins 83(8):1375–1384
26. Ma X, Meng H, Lai L (2016) Motions of allosteric and orthosteric ligand-binding sites in proteins are highly correlated. J Chem Inf Model 56(9):1725–1733
27. Marks DS, Hopf TA, Sander C (2012) Protein structure prediction from sequence variation. Nat Biotechnol 30:1072
28. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP (2009) Quantifying correlations between allosteric sites in thermodynamic ensembles. J Chem Theory Comput 5(9):2486–2502
29. McClendon CL, Hua L, Barreiro G, Jacobson MP (2012) Comparing conformational ensembles using the Kullback–Leibler divergence expansion. J Chem Theory Comput 8(6):2115–2126
30. Meng H, Liu Y, Lai L (2015) Diverse ways of perturbing the human arachidonic acid metabolic network to control inflammation. Acc Chem Res 48(8):2242–2250
31. Meng H, McClendon CL, Dai Z, Li K, Zhang X, He S, Shang E, Liu Y, Lai L (2016) Discovery of novel 15-lipoxygenase activators to shift the human arachidonic acid metabolic network toward inflammation resolution. Eur J Med Chem 59(9):4202–4209
32. Meng H, Dai Z, Zhang W, Liu Y, Lai L (2018) Molecular mechanism of 15-lipoxygenase allosteric activation and inhibition. Phys Chem Chem Phys 20(21):14785–14795
33. Miao Y, Nichols SE, Gasper PM, Metzger VT, McCammon JA (2013) Activation and dynamic network of the M2 muscarinic receptor. Proc Natl Acad Sci U S A 110(27):10982–10987
34. Miao Y, Feher VA, McCammon JA (2015) Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. J Chem Theory Comput 11(8):3584–3595
35. Mitternacht S, Berezovsky IN (2011) Binding leverage as a molecular basis for allosteric regulation. Plos Comput Biol 7(9):e1002148
36. Nguyen PH, Derreumaux P, Stock G (2009) Energy flow and long-range correlations in guanine-binding riboswitch: a nonequilibrium molecular dynamics study. Indian J Chem A 113(27):9340–9347
37. Nussinov R, Tsai C-J (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305
38. Nussinov R, Tsai C-J, Liu J (2014) Principles of allosteric interactions in cell signaling. J Am Chem Soc 136(51):17692–17701
39. Okazaki K-i, Koga N, Takada S, Onuchic JN, Wolynes PG (2006) Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: Structure-based molecular dynamics simulations. Proc Natl Acad Sci U S A 103(32):11844–11849
40. Ota N, Agard DA (2005) Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. Am J Respir Cell Mol 351(2):345–354
41. Panjkovich A, Daura X (2012) Exploiting protein flexibility to predict the location of allosteric sites. BMC Bioinformatics 13(1):273
42. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites. Bioinformatics 30(9):1314–1315
43. Pei J, Yin N, Ma X, Lai L (2014) Systems biology brings new dimensions for structure-based drug design. J Am Chem Soc 136(33):11556–11565
44. Petrone P, Pande VS (2006) Can conformational change be described by only a few normal modes? Biophys J 90(5):1583–1593
45. Pierce LCT, Salomon-Ferrer R, de Oliveira CAF, McCammon JA, Walker RC (2012) Routine access to millisecond time scale events with accelerated molecular dynamics. J Chem Theory Comput 8(9):2997–3002
46. Pincus D, Pandey JP, Feder ZA, Creixell P, Resnekov O, Reynolds KA (2018) Engineering allosteric regulation in protein kinases. Sci Signal 11(555). https://doi.org/10.1126/scisignal.aar3250
47. Qi Y, Wang Q, Tang B, Lai L (2012) Identifying allosteric binding sites in proteins with a two-state go model for novel allosteric effector discovery. J Chem Theory Comput 8(8):2962–2971

48. Reynolds Kimberly A, McLaughlin Richard N, Ranganathan R (2011) Hot spots for allosteric regulation on protein surfaces. Cell 147(7):1564–1575
49. Rivoire O, Reynolds KA, Ranganathan R (2016) Evolution-based functional decomposition of proteins. Plos Comput Biol 12(6):e1004817
50. Salinas VH, Ranganathan R (2018) Coevolution-based inference of amino acid interactions underlying protein function. eLife 7:e34300
51. Sen TZ, Feng Y, Garcia JV, Kloczkowski A, Jernigan RL (2006) The extent of cooperativity of protein motions observed with elastic network models is similar for atomic and coarser-grained models. J Chem Theory Comput 2(3):696–704
52. Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA: protein complexes. Proc Natl Acad Sci U S A 106(16):6620–6625
53. Sharp K, Skinner JJ (2006) Pump-probe molecular dynamics as a tool for studying protein motion and long range coupling. Proteins 65(2):347–361
54. Su JG, Qi LS, Li CH, Zhu YY, Du HJ, Hou YX, Hao R, Wang JH (2014) Prediction of allosteric sites on protein surfaces with an elastic-network-model-based thermodynamic method. Phys Rev E 90(2):022719
55. Van Wart AT, Durrant J, Votapka L, Amaro RE (2014) Weighted implementation of suboptimal paths (WISP): an optimized algorithm and tool for dynamical network analysis. J Chem Theory Comput 10(2):511–517
56. Wang Q, Qi Y, Yin N, Lai L (2014) Discovery of novel allosteric effectors based on the predicted allosteric sites for Escherichia coli D-3-phosphoglycerate dehydrogenase. PLOS One 9(4):e94829
57. Wang Q, Liberti MV, Liu P, Deng X, Liu Y, Locasale JW, Lai L (2017) Rational design of selective allosteric inhibitors of PHGDH and serine synthesis with anti-tumor activity. Cell Chem Biol 24(1):55–65
58. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46(W1):W374–W379
59. Yang L-W, Eyal E, Chennubhotla C, Jee J, Gronenborn AM, Bahar I (2007) Insights into equilibrium dynamics of proteins from comparison of NMR and X-ray data with computational predictions. Structure 15(6):741–749
60. Yu M, Ma X, Cao H, Chong B, Lai L, Liu Z (2018) Singular value decomposition for the correlation of atomic fluctuations with arbitrary angle. Proteins 86(10):1075–1087
61. Yuan Y, Pei J, Lai L (2013) Binding site detection and druggability prediction of protein targets for structure-based drug design. Curr Pharm Design 19(12):2326–2333
62. Zhou H, Dong Z, Tao P (2018) Recognition of protein allosteric states and residues: machine learning approaches. J Comput Chem 39(20):1481–1490
63. Li C, Deng X, Zhang W, Xie X, Conrad M, Liu Y, Angeli JPF, Lai L (2018) Novel allosteric activators for ferroptosis regulator glutathione peroxidase 4. J Med Chem 62(1):266–275

Chapter 6
Characteristics of Allosteric Proteins, Sites, and Modulators

Xinheng He, Duan Ni, Shaoyong Lu, and Jian Zhang

Abstract Allostery is considered one of the most direct and efficient ways to regulate biological macromolecule functions. Allostery is increasingly receiving attention in the field of drug discovery because of the unique advantages of allosteric modulators such as high selectivity and low toxicity. Because of technical breakthroughs in the allosteric studies, the understanding of the characteristics of allosteric entities such as allosteric proteins and their allosteric sites and modulators has made great strides. These features play a critical role in both the evolution of the allosteric concept and the prediction of allosteric interactions. In this chapter, we highlight the fundamental characteristics of allosteric proteins, allosteric sites, and allosteric modulators. Importantly, the applications of such principles in real cases are depicted in detail. Collectively, these characteristics are beneficial in aiding allosteric drug design and allosteric mechanism research.

Keywords Allostery · Allosteric modulation · Allosteric mechanism · Drug design

6.1 Introduction

Allostery, also referred to as allosteric regulation, is the process by which biological macromolecules (mostly proteins) transmit the effect of ligand binding at allosteric sites spatially and topographically remote from the functional sites, resulting in the regulation of macromolecule activity [26, 69, 115, 124, 182]. Allosteric signals are transmitted across the structure of macromolecules to the active site via allosteric signal pathways including atomic fluctuations, residue networks, or domain movements [45, 77, 82, 117, 136, 173, 188]. Allosteric interactions can be initiated by

X. He · D. Ni · S. Lu · J. Zhang (✉)
Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, Shanghai Jiao Tong University, School of Medicine, Shanghai, China
e-mail: jian.zhang@sjtu.edu.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_6

many factors ranging from noncovalent contacts (small molecules, ions, lipids, peptides, proteins, and nucleic acids) [33, 40, 95, 114, 140, 204] to covalent modifications (disulfide trapping, phosphorylation, and point mutations) [113, 131, 135, 142, 168], light absorption [175], and environmental fluctuations (ionic strength, pH, and temperature) [20, 43, 56]. Allosteric perturbation is an efficient mechanism that allows the regulation of the functions and activities of proteins and then fine-tunes a myriad of biological processes, including metabolism, signal transduction, enzyme activity, and gene regulation [3, 23, 68, 70, 107, 121, 139, 158, 204]. Allosteric perturbation has been recognized as the "second secret of life," next only to the genetic code [53, 182]. The dysfunction of allosteric networks has been significantly associated with the etiology of a broad spectrum of human diseases such as cancer, growth disorders, and neuron disease [27, 131, 143].

Allosteric regulation is always triggered by allosteric modulators that present a couple of advantages over orthosteric modulators that target the functional sites. Above all, coded by homologous genes, proteins in the same family always possess highly conserved orthosteric binding sites, especially in G-protein-coupled receptors (GPCRs) and kinases. However, allosteric sites are often less conserved because allosteric effects occur in the diverse structure of the protein but not the functional sites [18, 201]. Furthermore, even if the affinity of an allosteric modulator is the same among different subtypes of a protein, the modulator may still exert a selective effect by having different degrees of cooperativity that exists between orthosteric and allosteric sites at each subtype [35, 105]. Thus, allosteric modulators can distinguish the target protein from homologous proteins, causing better subtype selectivity and fewer side effects than orthosteric modulators [71, 100, 133]. Moreover, any binding sites on the protein surface remote from the functional site could, in theory, potentially serve as allosteric sites, resulting in an evident diversity of allosteric effectors [48, 66, 126, 189]. Allosteric modulators also have no direct competition with orthosteric ligands, so they can exert their influence with an orthosteric ligand bound to the receptor. However, competitive inhibitors play their role only with a vacant orthosteric site [94, 127]. Additionally, allosteric modulators regulate the protein activity rather than absolutely eliminating the activity, indicative of their moderate potential effects [130, 131]. In addition to direct binding to the receptor, allosteric modulators also produce indirect allosteric effects that influence proteins in one signaling unit so they have the potential to fine-tune tissue activities specifically [161, 176]. With these advantages, allosteric drug development has been regarded as a new paradigm and a prevailing trend in drug design at present [12, 88, 93].

Nevertheless, a battery of disadvantages cannot be ignored in allosteric regulation. Allosteric modulators may be less effective than orthosteric modulators due to the common flat sites [193]. For instance, allosteric modulators of GPCRs often show micromolar affinities to their receptor, which restrains their drug potential [184]. Furthermore, nonconserved sites might lead to an easier onset of resistance in the systems with rapid gene mutation and severe selection (e.g., antivirus and anticancer therapeutic areas) [103, 186]. Additionally, allosteric sites are usually different between species owing to the low evolutionary conservation. Therefore, it

is difficult to develop allosteric drugs because rodent model experiment results might totally mismatch the results in the human body [184, 193].

Additionally, there are several challenges in the field of allosteric drug development. First, although possible allosteric sites are widespread on the protein surface, it is difficult to discover highly effective allosteric sites [115, 127, 191]. Moreover, allosteric effector design is not the same as orthosteric drug design. The latter can refer to related ligands but the former mostly needs to create a new world [138]. With these challenges and devoid of relevant research methodologies, the number of allosteric modulator identifications was limited before the year 2000. Fortunately, the rapid development of structural biology and computational biology has recently provided a series of methods such as NMR, X-ray crystallography, and fluorescent labeling, which strongly promotes the research of allostery and fuels the interest in the discovery of allosteric effectors and allosteric sites [8, 14, 31, 49, 96, 110, 128, 162]. The trend is confirmed by the explosive growth of the number of allosteric modulators in the past 8 years. In the latest Allosteric Database (ASD), 77825 allosteric modulators and 1788 allosteric macromolecules were included [81, 84, 165]. There is no doubt that the rapid development of related technologies is the precondition of research into allostery features. Thus, a review of the characteristics of allostery is helpful in the field of new technology application and advanced research about allosteric regulation.

Here, we summarize the characteristics of allosteric regulation. We first generalize the regulating mechanism and distribution of allosteric proteins, the largest proportion of the allosteric biological macromolecules. Then, we state the characteristics of allosteric sites that are invoked for the prediction of new sites and the development of novel allosteric drugs. Finally, the commonalities and key physiological influential factors of allosteric modulators are discussed and quintessential examples of allosteric drug design are depicted to prove the utility of allostery in biological processes. These characteristics might help in future allosteric mechanism research and allosteric drug development.

6.2 Allosteric Proteins

6.2.1 Models of Allosteric Proteins

Multiple models such as the Monod–Wyman–Changeux (MWC) model [123], the Koshland–Nemethy–Filmer (KNF) model [102], the morpheein model [89, 90], the population shift model [92], and the dynamic-driven model [101, 112, 146] have been reported to elucidate the foundation of allostery (Fig. 6.1). The MWC and KNF models both state that allosteric proteins are symmetrical homo-oligomers. However, the MWC model indicates that modulator binding and conformational changes occur in a concerted manner, but the KNF model assumes that modulator binding is separated to cause conformation remodeling, as a sequential process. In addition, the MWC model supports that ligand binding shifts the equilibrium of structures so that

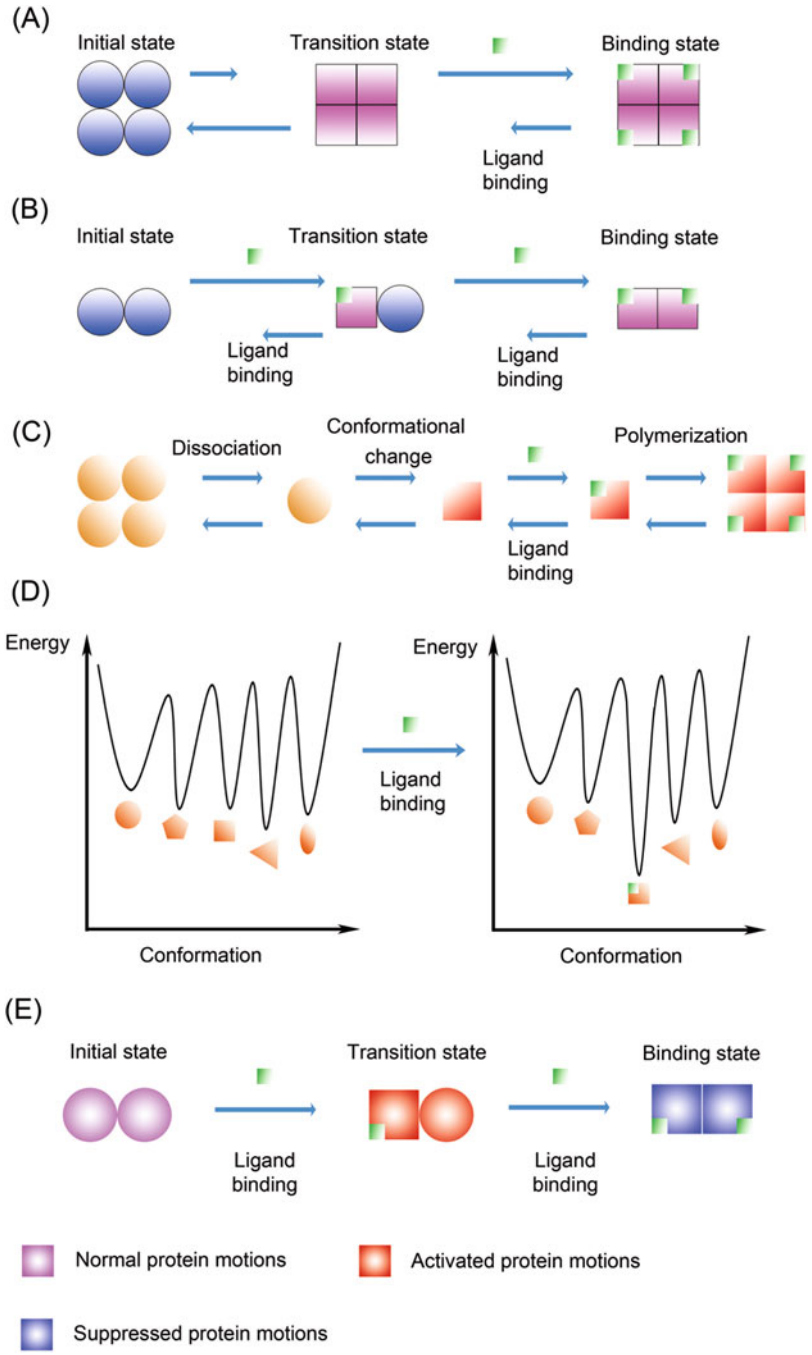


Fig. 6.1 Allostery models. (**a**) The two-state concerted model of Monod–Wyman–Changeux (MWC). (**b**) The sequential model of Koshland–Nemethy–Filmer (KNF). (**c**) The morpheein model. (**d**) The population shift model. (**e**) The dynamic-driven model

the transition state exists as a stable state with ligand binding (Fig. 6.1a, b). As groundbreaking theories, the two models are referenced extensively in early allostery research.

The morpheein model is based on homo-oligomeric protein allostery. In this model, subunits of the protein assemble as a specific multimer, while alternate subunits of the protein can assemble as a structurally and functionally different alternate multimer (Fig. 6.1c). Alternate subunits can bind with the allosteric ligand that stabilizes them and shifts the equilibrium to form the alternate multimer. As a result, the function of the protein is altered toward the targeted multimer, providing a mechanism for allostery [89, 90].

With further research, the consensus is that the molecular mechanisms of allostery are rooted in the dynamic nature of proteins, an inherent property of biomacromolecules [72, 73, 129]. Allosteric regulation exists whether small structures or complex oligomeric enzymes are involved, so the viewpoint from homo-oligomeric protein allostery needs to be supplemented. Two dynamic models are helpful in the explanation. In the population shift model, evolution has formulated a conformation ensemble composed of a series of stable energy basins in the protein structure [112]. Depicted as a selection process, an allosteric modulator binds with the conformation that has the site most suitable for it. As illustrated in Fig. 6.1d, binding of an allosteric effector reduces the energy of the corresponding protein structure and makes it a stable and dominant conformation. Then, the conformation ensemble undergoes a population shift, and the conformational redistribution is accomplished [62]. In the dynamic-driven model, allosteric ligands exert their action by transmitting changes in protein motions rather than the alteration of the protein conformation [146]. Importantly, changes in protein dynamics dominate the allosteric effect, but the shape of the protein structure is maintained. As shown in Fig. 6.1e, the first ligand binding does not induce structural changes in the unliganded subunit, but the motions of its residues are activated as well. Similarly, the second ligand binding incurs an obvious decrease in total atomic fluctuations (conformational entropic penalty) but not atomic mean position changes in other subunits, indicative of the allosteric interactions driven by changes in protein motions [112, 146].

6.2.2 Distribution of Allosteric Proteins

To determine if a protein is an allosteric protein, the most direct way is to solve the cocrystal structure between the protein and the allosteric effector with the help of X-ray crystallography. With a technological advance such as the cryo-electron microscope, the acquisition of crystal structure troubles researchers less [6, 16]. Otherwise, under the condition that cocrystal structures are not always easily available, indirect biochemical evidence is also applied to confirm allosteric interactions between the protein and modulator. For example, the point mutation at key residues, which causes a measurable decrease of allosteric ligand-binding affinity, can define a allosteric effect [39].

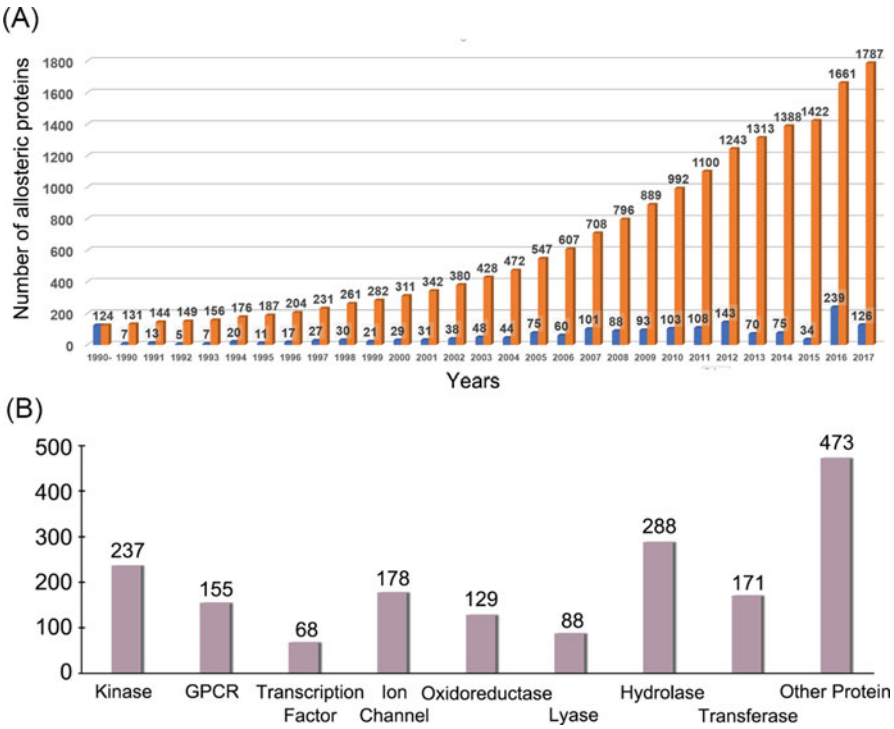


Fig. 6.2 (**a**) The overall growth of released allosteric protein structures per year (The 1990- represents structures released before 1990). The orange bars show the number of total available allosteric protein structures, and the blue bars show the number of structures released annually. Data come from ASD3.0 [165]. (**b**) The class distribution of all proteins in the latest ASD (Aug. 2018)

After being defined, allosteric proteins and allosteric modulators are deposited into databases to summarize and utilize them better. The Allosteric Database (ASD) serves as a relatively complete database of allosteric interactions [81, 84, 165]. Although the first version of the ASD contained only 316 allosteric proteins in Aug. 2010, the latest ASD (Nov. 2018) contains 1787 proteins [81, 165]. Figure 6.2a shows the growth trend of allosteric proteins in the past few years. A small fraction of allosteric proteins was evidently reported before 2000, but research related to allosteric proteins experienced rapid growth after the millennium. The reasons for this phenomenon stem not only from the attraction of the advantages of allosteric drugs to researchers but also from the new insights into mechanisms of allostery caused by the advances in structural biology [64, 116, 131]. In the recent two years, the application of cryo-electron microscopy and computational methods has promoted the discovery of allosteric proteins further [13, 69, 118, 119].

In the origin theory, proteins possessing an oligomeric structure with symmetry properties were defined as allosterically regulatory proteins [24]. A typical example is the first allosteric protein hemoglobin regulated by oxygen [123]. Allosteric perturbation transmission between topographically remote sites is accomplished by conformational changes established among several pre-established states with the conservation of symmetry, and the effect on receptor-binding sites is produced by this cooperativity [24]. However, the symmetrical oligomeric model is a common but not universal property in proteins [25, 108]. In the Protein Data Bank in August 2018, the number of monomers (64637) is not far from the number of oligomers (67138), and the universality of allostery makes it impossible to ignore monomers ([153]; Rose et al. [154]). Meanwhile, allosteric phenomena have been observed in monomer proteins such as GPCRs [156], thrombin [75], and proteases [187] with NMR and X-ray crystallography. For instance, type II protease-activated receptors (PAR2), a member of the rhodopsin-like family of class A GPCRs, can obviously be downregulated by an allosteric modulator AZ8838 binding to a pocket lined by extracellular loop 2, transmembranes 1–3 and 7, providing a possible selective PAR2 antagonist for a series of therapeutic uses (e.g., anti-inflammatory) [28, 202]. Recent research reveals that even intrinsically disordered proteins (IDPs) can experience allosteric regulation in signal transmission [181, 197]. Thus, the old definition needs to be supplemented.

As shown in Fig. 6.2b, allosteric phenomena are widely spread in various proteins such as kinases, ion channels, GPCRs, transcription factors, and kinds of enzymes. The existence of allostery in most proteins confirms that rooted in the nature of proteins, allosteric regulation is as ubiquitous as orthosteric modulation. Therefore, allosteric regulation gains increasing attention in academia and the pharmaceutical industry. To date, allosteric protein collection is focused mainly on kinases, ion channels, and GPCRs, which are the three largest categories of human drug targets. In these three categories, classic orthosteric effectors interact with the highly conserved sites bound by the endogenous modulators, so it is difficult to achieve great selectivity for individual subtypes [37, 116]. However, high selectivity, even toward closely related subtypes, can be provided by allosteric modulators due to their distinct binding sites, leading to relatively low off-target effects and side effects [30, 195]. Moreover, orthosteric modulators mimicking endogenous ligands have a limited effect on proteins that respond to special stimuli such as light, protons, and divalent cations, but allosteric modulators play their role by regulating the interaction between receptors and ligands [37]. Of note, allosteric drugs bind with the nonfunctional structure of proteins, so common mutations occurring on functional sites exclude the effect of orthosteric modulators but not allosteric effects [198]. Based on these advantages, the rapid development of allosteric proteins, especially kinases, ion channels, and GPCRs, can reasonably be explained.

6.3 Allosteric Sites

6.3.1 Definition of Allosteric Sites

The defined allosteric sites are some residues that are away from the catalytic center and able to interact with allosteric modulators to regulate the protein activity. Thus, the allosteric sites exercise a remote control in the regulation of protein activity. Conversely, orthosteric sites are always coincident with the catalytic center [25, 72]. Figure 6.3a shows the relative positions of the two sites in cyclin-dependent kinase 2 (CDK2), a serine/threonine kinase that is a therapeutic target in cancer

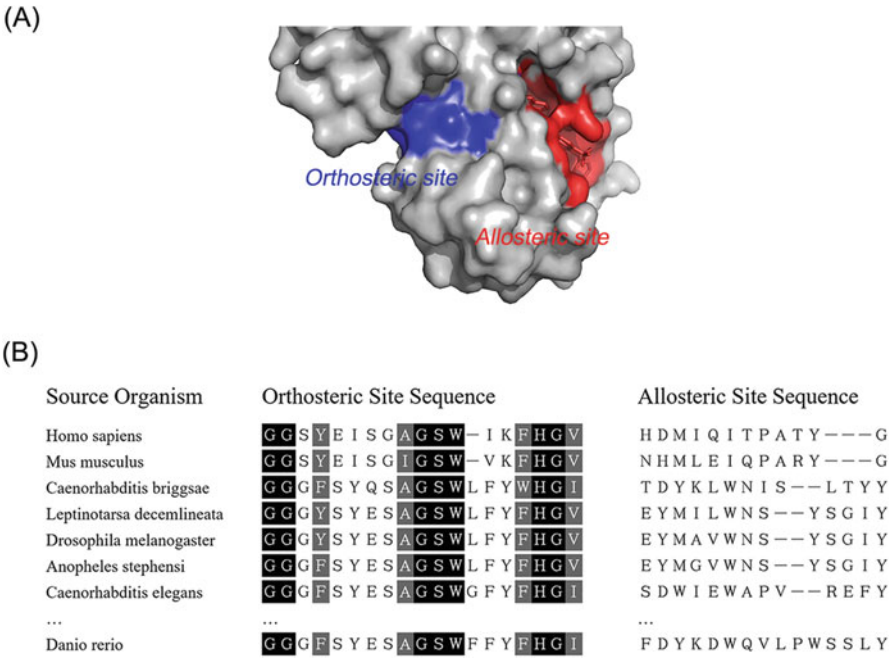


Fig. 6.3 Several features of allosteric sites. (**a**) The relative position of allosteric and orthosteric sites is colored in the surface representation, taking CDK2 as an example. The blue area represents the orthosteric site interacting with CCT068127, an inhibitor competing with the substrate ATP [194]. The red area represents the allosteric sites bound with two 8-anilino-1-naphthalene sulfonate molecules, the exclusive CDK2 allosteric inhibitor (PDB ID: 3PY1) [9]. (**b**) A sequence alignment comparison of residues of an allosteric site and an orthosteric site in AchE from 24 different species (partial sequences are omitted for clarity). Residues that maintain identity among the 24 sequences possess a black background, and those that are similar among the sequences have a gray background. The remaining residues keep a white background. The homologous protein data were collected from the ConSurf server [21, 67], and the sequence alignment was finished at the Sequence Manipulation Suite [174]. The residue numbers of allosteric sites and orthosteric sites were derived from ASD 3.0 [165]

treatment [9, 151, 194]. In addition to this key difference, allosteric sites also have a series of distinctions from orthosteric sites, as described below.

6.3.1.1 Nonconservation of Allosteric Sites

At first, allosteric sites were relatively nonconservative compared to the orthosteric sites. Yang et al. [201] proposed that allosteric site residues have an average conservation score of 0.58 that is calculated from homologous sequence alignments in 58 enzymes collected from various species. However, orthosteric site residues have an average conservation score of 0.94 ($P=1.36 \times 10^{-67}$). Using acetylcholinesterase (AChE) as a sequence comparison example in Fig. 6.3b, it is significant that the allosteric site of AchE shows negligible similarity among 24 different species. Nevertheless, the orthosteric site of AchE is highly conserved. Even more than one third of the residues of the orthosteric site of AchE are exactly the same among the species. Thus, allosteric sites are statistically unequivocally less conserved than orthosteric sites.

From an evolutionary perspective, orthosteric sites are responsible for substrate binding and further reaction, so mutations here inevitably demolish the normal function of the protein [7]. Rather, allosteric sites serve only as platforms for modulator binding instead of being involved in the catalytic effect toward substrates. Therefore, mutations on allosteric sites rarely influence the key functions of the proteins [72, 186]. Furthermore, different environments render allostery variables in different organisms and species, which provides an opportunity to evolve allosteric site residues differently to fit the environments [201]. For instance, AMP allosterically inhibits the porcine fructose-1,6-bisphosphatase (FBPase), but this effect is absent in *Escherichia coli* FBPase, which shares a 41% sequence identity and similar functions with the former FBPase [61, 78]. As a result, allosteric sites are not as conserved as orthosteric sites. Thus, the drugs targeting allosteric sites can be specific and prevent side effects [34, 100]. Meanwhile, low conservation makes it difficult to develop allosteric drugs by cross-species experiments [193].

6.3.1.2 Residue Features of Allosteric Sites

In addition to distinct evolutionary pathways, allosteric sites also have unique amino acid compositions that are related to the allosteric mechanism and helpful in allosteric site prediction. Song et al. [172], Li et al. [109], and Yang et al. [201] all revealed that allosteric sites tend to be more hydrophobic than orthosteric sites with data collected from different databases. Polar amino acid residues such as asparagine, histidine, aspartic acid, and glutamic acid are more abundant at orthosteric sites than at allosteric sites, whereas the hydrophobic residues (e.g., leucine, valine, isoleucine, phenylalanine, and proline) are highly enriched at allosteric sites. Notably, amino acids with hydroxyl groups such as serine, tyrosine, and threonine are abundant at orthosteric sites.

Different site compositions could be attributed to different functions undertaken by the sites. Orthosteric ligands commonly interact with the sites by a covalent effect (e.g., phosphorylation), but allosteric sites seldom form covalent bonds with corresponding ligands [41, 113, 199]. Thus, hydroxyl-containing amino acids are enriched at orthosteric sites to form covalent interactions. For the hydrophobicity possessed by allosteric sites, hydrophilic orthosteric site residues have the potential to form unique connections such as hydrogen bond networks to facilitate receptor-ligand interactions, but allosteric site residues indicate fewer similar interactions in this regard [5, 144]. Rather, the allosteric site residues drive changes in the hydrogen bond network instead of being involved in that network [97]. In addition, nonpolar residues at allosteric sites are useful in the formation of the ligand-binding pocket, with only a small part of polar residues facilitating interactions during ligand binding [201].

6.3.2 Prediction of Allosteric Sites

Essentially, allosteric regulations are due to allosteric sites receiving and transmitting perturbation signals to the functional sites through atomic fluctuations, amino acid residue networks, or domain movements, resulting in dynamic alterations in the protein conformation population and ultimately conversion between two or more functional states of the protein [19, 45, 92, 117, 124]. During this process, the allosteric sites are undoubtedly the trigger points and regulatory centers for allostery. Therefore, how to identify allosteric sites is regarded as an important issue for drug design related to allosteric regulation [85, 116, 200]. In the early period of allostery research, allosteric site identification was dependent mainly on experimental methods. With the rapid development of computational biology, methods in silico applied for the identification of allosteric sites have emerged generally and show a promising future [32, 115, 127]. Next, we summarize the basic concept and recent progress in the field of experimental and computational methods for the prediction of allosteric sites.

6.3.2.1 Prediction of Allosteric Sites Based on Experimental Methods

X-Ray Crystallography

As the most common experimental method in allosteric site identification research, X-ray crystallography continues to make contributions to the discovery of allosteric sites [110, 118, 166]. In the research for site discovery, the procedure is as follows. First, the allosteric modulators are screened by a high-throughput method, and then the cocrystal structure of the protein and modulator is cultured using structural biological methods. Next, the diffraction images of the structure are obtained by X-ray crystallography. Analysis of the complex crystal structure of the effector and

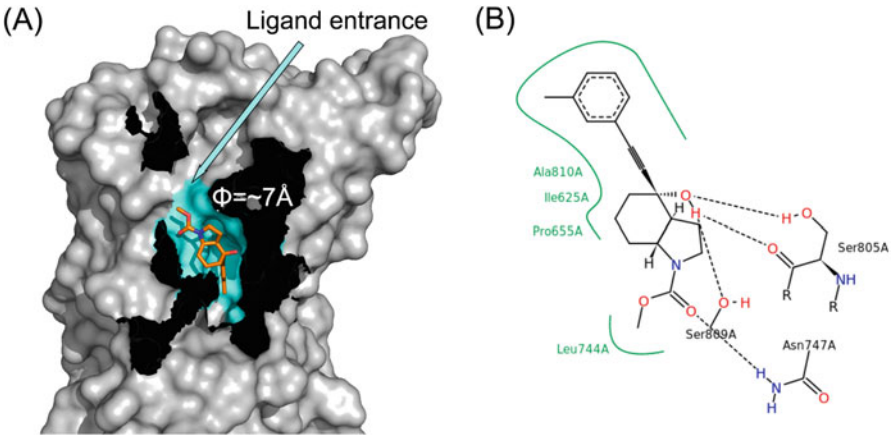


Fig. 6.4 The position of mavoglurant and interactions between $mGlu_5$ and the mavoglurant. (**a**) Surface representation of the $mGlu_5$–mavoglurant complex structure (PDB ID: 4OO9), viewed parallel to the membrane lipids. In a stick representation, mavoglurant is colored orange with the binding pocket colored in cyan. The arrow indicates the ligand entrance, marked with its diameter. (**b**) Interactions between pocket residues and mavoglurant. Hydrogen bonds are represented by dashed lines, and part of the site is represented by green lines

the protein directly explains the allosteric regulatory site and mechanism. In addition, by comparing the crystal structures between the apo state and the holo state, it is possible to find detailed changes in the protein after the allosteric perturbation. However, X-ray crystallography also has limitations. Allostery is caused not only by protein conformation changes but also by protein thermodynamics. The crystal structure is observed statically, but the allosteric process is dynamic. Therefore, the result of X-ray crystallography lacks kinetic data, which can be complemented by following molecular dynamic simulation [55, 157, 166].

Metabotropic glutamate receptor 5 ($mGlu_5$), a member of class C GPCRs, is of considerable interest as a promising allosteric therapeutic target in the treatment of a series of emotion and movement disorders [87, 203]. However, structural information on the seven-transmembrane domain (TMD) of $mGlu_5$ has remained elusive for a long time. This domain plays a particularly important role in the identification of allosteric sites and allosteric modulators bound to this region [35, 51]. Doré et al. [46] first determined the cocrystal structure of the $mGlu_5$ TMD domain in a complex with the allosteric modulator mavoglurant using X-ray crystallography based on a thermostabilized mutated receptor. As shown in Fig. 6.4a, X-ray crystallography data directly reveal that the allosteric site defined by residues from transmembrane helices is hidden in a pocket 8 Å from the receptor surface, and the entrance of the modulator is restrained to 7 Å by the transmembrane helix and the extracellular loop 2 of $mGlu_5$. Furthermore, the crystal structure also identifies the interactions between residues and ligand (Fig. 6.4b), which helps the explanation of its mechanism and drug optimization in the next step.

Isotope Tracer Labeling Method

Isotope labeling technology has recently been developed rapidly to discover allosteric sites. With the ^{13}C and ^{15}N labeling, nuclear magnetic resonance (NMR) spectroscopy provides not only high-resolution structural information for a complex in solution but also information about the dynamic process on a picosecond to millisecond time scale. Moreover, low-populated conformational states and residue conformational entropy can be provided by NMR [169, 183]. All of this information is helpful in the identification of an allosteric site.

Hydrogen-deuterium exchange mass spectrometry (HDX-MS) is another site revelation method that is useful with the help of isotope labeling. In this way, the protein is immersed in a heavy aqueous solution, and the hydrogen atoms of the protein are exchanged with the deuterium in the solution. During the exchange process, the exchange rate of hydrogens located outside the protein is fast, but the hydrogens inside or participating in the formation of hydrogen bonds exchange slowly. Then, by mass spectrometry, the exchange rate of different sequence fragments of the protein can be monitored to show the intrinsic fluctuations, local unfolding, rigid body motion, and conformation rearrangement of the protein, which can be summarized as kinetic parameters and energy parameters and finally infer the location of allosteric sites [22, 192].

Heat shock protein 90 (HSP90) is a kind of molecular chaperone that facilitates protein folding, maturation, and stabilization. Proteins chaperoned by Hsp90 are sometimes significant in the progression of various diseases. Thus, Hsp90 inhibition site discovery gains intense academic interest [86, 205]. Chandramohan et al. [22] combined both structural insights and dynamic information provided by HDX-MS, observed characteristic long-range conformational changes, and finally identified its allosteric site. The binding of a phenolic class ligand with low affinity (K_D = 570 μm) demonstrates the difference in the hydrogen–deuterium exchange rate between the apo state and the holo state at the allosteric site and regions affected by allostery (Fig. 6.5). Therefore, the site position and allosteric effect are determined.

Fluorescent Label Method

In addition to the isotope label, the fluorescent label can also be used in the field of allosteric site identification. As a commonly used protein tracer, fluorescent molecules bind covalently to the suitable residues of a target protein, and its motion dynamically expresses the movement or conformation changes of the corresponding region of the protein. Because of its high sensitivity and the convenience of detection, the fluorescent label is widely used in the discovery of allosteric sites [122].

Regarded as a new target for antibiotic discovery, FtsZ is a widely conserved tubulin-like GTPase that directs bacterial cell division [120]. A potent allosteric inhibitor PC190723 binds to FtsZ in a cleft-opening conformation and suppresses its function [50, 177]. Using fluorescent benzamide analogs of PC190723, Artola et al.

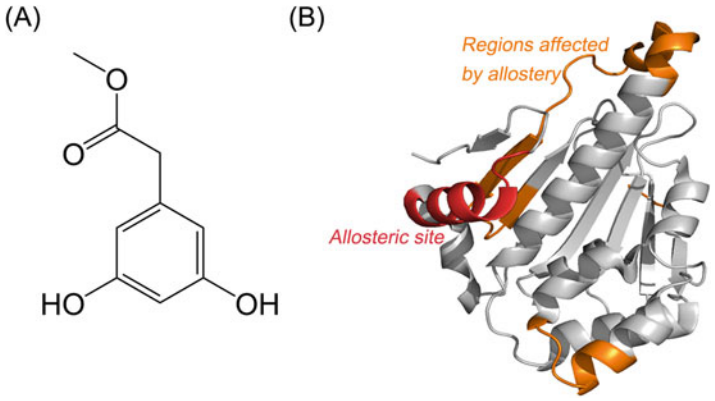


Fig. 6.5 Ligand influence observed by HDX-MS shows allosteric effect in Hsp90. (**a**) Modulator 2,4-dihydroxypropiophenone causes conformational motions. (**b**) Cartoon representation of Hsp90 crystal structure (PDB ID: 4EGK). Hydrogen–deuterium exchange rate alteration after ligand binding shows the positions of regions affected by allostery (colored in orange) and allosteric site (colored in red)

[2] demonstrated the binding site and elucidated the mechanism experimentally. With an open cleft, the probe fluorescent molecule interacts with the allosteric site and shows the position of the allosteric site. However, a closed cleft in FtsZ prevents the allosteric interaction between ligand and receptor. Combined with cytological methods, the fluorescent probe binding assays are also employed to screen for allosteric inhibitors of FtsZ in search of new antibiotics [1] (Fig. 6.6).

Disulfide Trapping Method

Disulfide trapping is a site-directed approach developed by Well et al. [76, 159], offering a strategy to detect hidden and potential allosteric sites on a given protein using small molecules. In this approach, a protein harboring cysteine adjacent to a site of interest is screened against a library of disulfide-containing compounds for the purpose of forming a disulfide bond between the appropriate small molecules and a cysteine in the proximity of the site under reducing conditions. The formation of a disulfide bond between the protein and the specific small molecule is then easily detected via mass spectrometry. Ultimately, X-ray crystallography is performed to demonstrate the binding mode between the protein and disulfide-trapped fragments. The disulfide bond capture method has a site-directed orientation, but its application scope can only cover proteins with cysteine next to the site. Overall, disulfide trapping has proven to be a valuable tool to verify suspected or orphan allosteric sites on the protein surfaces.

A study of caspase-5 shows the application of the disulfide trapping method. Caspase-5 is a member of the inflammatory aspartate-specific thiol proteases [17]. Working together with caspase-1, caspase-5 plays an important role in

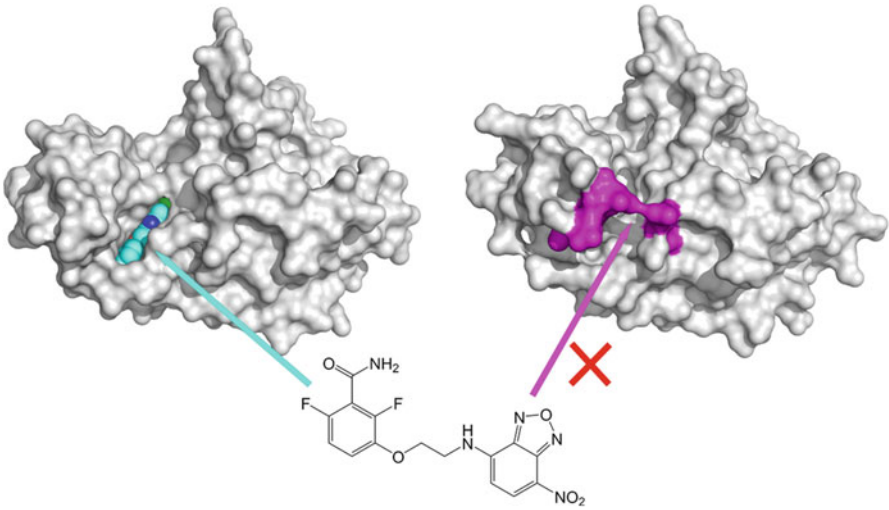


Fig. 6.6 Scheme interpreting the allosteric site and effect of the cleft. The cyan arrow represents the success of the probe fluorescent molecule (green spheres) in binding to the open state of FtsZ (PDB ID: 4DXD), resulting in the observed fluorescent. However, the magenta arrow and red cross indicate that the closed cleft structure (colored in magenta) prevents the effector from interacting with the receptor at the same allosteric site. In total, the fluorescent molecule identifies the allosteric site and the binding mechanism

inflammatory responses, but its mechanism remains obscure [10]. The development of caspase-5-specific inhibitors is beneficial to both elucidations of its mechanism and novel treatment of inflammatory disorders [190]. However, the orthosteric site is highly conserved among caspases, so the discovery of the allosteric site is imperative to address this issue. Gao and Wells [60] used the disulfide trapping approach to validate an allosteric site on caspase-5 and generate selective allosteric inhibitors of caspase-5. They left unchanged the Cys341 of caspase-5, which corresponds to Cys331 of caspase-1, located at its novel allosteric site [160]. Having screening approximately 15,000 thio-containing small molecule compounds, they identified via mass spectrometry 61 compounds that conjugate highly to Cys341. The 10 first-tier hits were selected to be resynthesized based on the structural attractiveness and the least similarity to known caspase-1 hits. The results revealed a naphthyl-thiazole-containing compound (Fig. 6.7a) that displays the most potent inhibition of caspase-5 but little inhibitory effect on caspase-1 (Fig. 6.7b). Further, a mass spectrometric study on wild-type caspase-5 confirmed that the compound specifically binds to the p10 subunit that is away from the orthosteric site and interferes with the substrate binding by inducing conformational changes. Therefore, a novel allosteric site on caspase-5 was discovered by the disulfide trapping approach, which provides a starting point for the design of selective allosteric modulators targeting caspase-5.

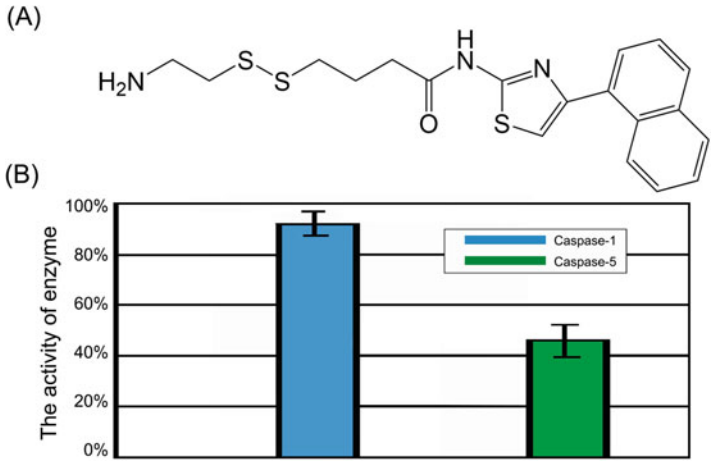


Fig. 6.7 (**a**) Chemical structure of the selective allosteric modulator. (**b**) Inhibition of wild-type caspase-1 and caspase-5 with the compound. The effector was incubated with either 50 nm caspase-1 or 200 nm caspase-5 at a concentration of 1 μm for 1 hour at room temperature and tested for inhibition of activity. The blue bar and green bar represent the activity of caspase-1 and caspase-5, respectively, after treating

6.3.2.2 Prediction of Allosteric Sites Based on Computational Methods

Prediction Based on Sequence Analysis

As the basic information for a protein, the sequence information has the potential to predict an allosteric site mainly through sequence alignment. Combining correlation information with sequence conservation, statistical coupling analysis (SCA) is one of the most representative methods for sequence-based prediction [74, 111, 179]. First, residues coevolving in a protein family are found through multiple sequence alignment techniques. Then, protein sectors, the networks of identified residues, are regarded as the structural basis of the allostery phenomenon. Finally, surface pockets that directly interact with the protein sectors are predicted to be allosteric sites [63, 185].

The discovery of allosteric hotspots for the metabolic enzyme dihydrofolate reductase (DHFR) provides solid support for the SCA prediction method. Widely spread in prokaryotes and eukaryotes, DHFR is an essential enzyme that is necessary for the biosynthesis of purines, pyrimidines, and amino acids [163]. However, the allosteric function of DHFR remains elusive. Using SCA to scan the homolog sequences of DHFR, Reynolds et al. [152] discovered a system of coevolving amino acids as a protein sector that forms a sparse and physically contiguous network. The network connects the DHFR orthosteric site with the substrate, the cofactor site, and several distant surface regions (Fig. 6.8) [106]. Indeed, the sector-connected surface sites are hotspots for the emergence of allostery, namely, possible allosteric sites. Thus, although there is no known allosteric function, DHFR contains

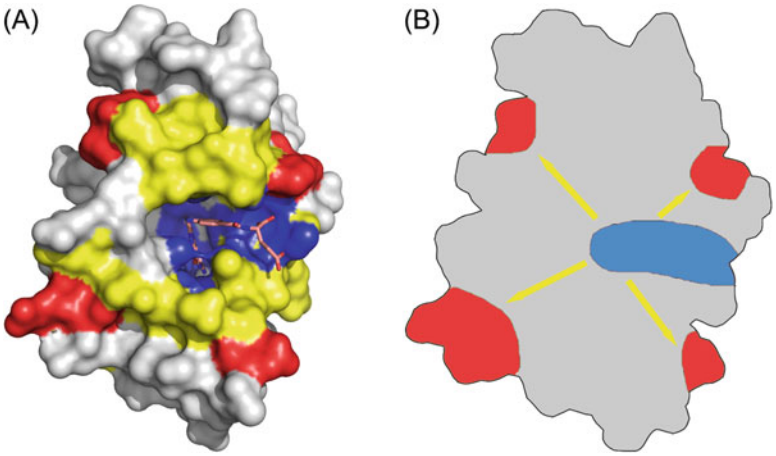


Fig. 6.8 The orthosteric site, connecting network, and predicted allosteric sites in DHFR. (**a**) A surface representation of the sparse but physically connected networks that links the active site to a few distant surface positions (PDB ID: 1RX2). The active site is colored blue. The sector comprising the network is colored yellow, and the possible allosteric sites are colored red. Other unrelated areas are colored gray. The orthosteric ligand (folic acid) is depicted by a stick model. (**b**) A flattened schematic representation of the slice through the protein core. Yellow arrows show the residue network linking the orthosteric site (blue) and a few distant surface positions, as well as the predicted allosteric sites (red)

a sparse sector structure that is the same as other allosteric proteins, providing a statistically preferable route for the initiation of allosteric control.

Prediction Based on Structure Analysis

Each protein has a unique 3D structure, and features of allosteric sites can be identified by machine learning that demands a large sum of allosteric structure data. Thus, the structures of a protein can also be applied to predict the allosteric sites. Based on data in ASD, Huang et al. [83] presented a highly efficient web platform, Allosite, to identify potential allosteric sites based on pocket-based analysis and a support vector machine (SVM) classifier. A broad array of allosteric and non-allosteric sites was employed to carry out the SVM learning process, and 21 site descriptions were performed to characterize the topological and physicochemical features of the sites. With the sensitivity and specificity meeting the need in the fivefold cross-validation test, the model was applied to validate 18 additional allosteric sites and 231 non-allosteric sites, and the rate of the correct results was as high as 96%. Although Allosite has been a highly sensitive and accurate allostery prediction tool, its performance can still be improved. Recently, Song et al. [172] presented AllositePro to upgrade the accuracy and effect of Allosite. A combination of structural features and approaches based on normal-mode analysis (NMA) was

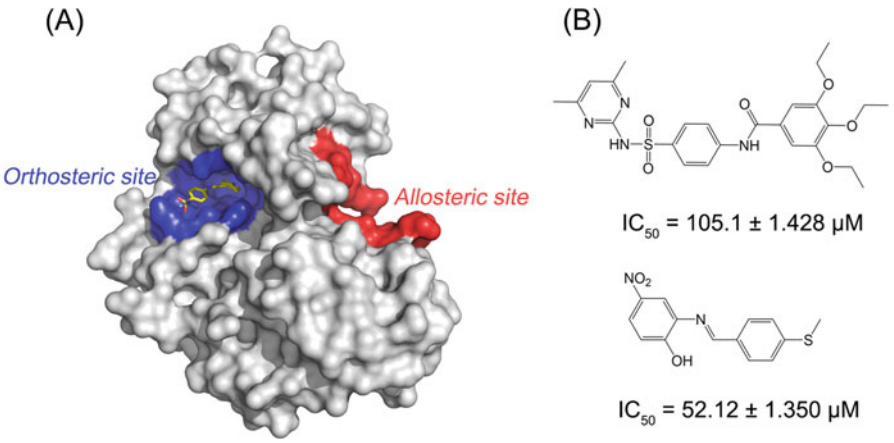


Fig. 6.9 The discovery of a novel allosteric site in CDK2 proved the usefulness of AllositePro. (**a**) The surface representation of the orthosteric site and allosteric site of CDK2. The novel allosteric site is represented in red, and the orthosteric site is colored in blue with DT2, an orthosteric inhibitor depicted by the stick model. (**b**) The chemical structure of novel allosteric modulators

used in the prediction of allosteric sites in AllositePro, making it a more accurate and reliable tool for allosteric site prediction.

A successful application of AllositePro is the discovery of a novel allosteric site on cyclin-dependent kinase 2 (CDK2). As a serine/threonine kinase, CDK2 is involved in the G1/S checkpoint and activates the S phase-specific gene expression [164]. Its overexpression leads to damage of the cell cycle and initiation of cancer, so CDK2 inhibition is regarded as a therapeutic target in cancer treatment [9, 151, 194]. In the prediction process, the crystal structure of CDK2 (PDB ID: 2C6K) was uploaded as the input. After the estimation of the site score, a pocket at least 5 Å from both the ATP-binding site and the substrate site was predicted to serve as a novel allosteric site (Fig. 6.9a). A mutation on residues R150 and Y180 at the predicted site significantly attenuates the catalytic effect of CDK2, indicative of the accuracy and potential use of the prediction results of AllositePro. Moreover, two new allosteric inhibitors were discovered based on the allosteric site (Fig. 6.9b) [80]. Thus, prediction methods based on structure analysis are useful in finding potential allosteric sites.

Prediction Based on Dynamic Information

Some proteins could have naturally cryptic ligand-binding sites. Although these sites have the potency to modulate the protein activities allosterically, they are usually difficult to detect in the ligand-free structures, even a single representative structure. However, powered by molecular dynamics (MD) simulations, a full map of receptor conformation space and representative samples of all structures with atomic-level

accuracy make the allosteric sites and molecular interaction processes visible [57, 145]. MD simulations reproduce the dynamic interaction process and offer an in-depth insight into the detailed interactions between proteins and modulators, without prior knowledge of the binding site position [118, 127]. Specifically, large-scale unbiased MD simulations start with the receptor protein and known modulator placed away from the orthosteric site. In the simulation process, the binding of allosteric effectors reveals several metastable hidden binding sites that can be confirmed by X-ray crystallography [47, 104]. Furthermore, based on the conformation clusters of the protein, virtual docking blindly can identify the possible sites [79]. In addition to a single MD simulation, massive MD simulations construct Markov state models (MSMs) that are able to discover allosteric sites hidden in minor conformations [15, 29]. With the pMD membrane newly developed, membrane-bound protein interaction can be simulated, as well as solvent proteins, and the prediction of allosteric sites on them becomes easy and accurate [147]. In short, dynamically predicting allosteric sites is a method used widely, especially in the field of the prediction of sites hidden in the intermediate states.

The nonreceptor tyrosine kinase c-Src is a well-known drug target because of its oncogenic properties in many cancers and importance in neuroinflammation [58, 171]. Extensive experimental data have revealed both inactive and active states of c-Src and showed large-scale distinctions between them [155, 178]. However, the absence of intermediate states along the activation pathway impedes control of the aberrant kinase activation [52, 148]. Using massively distributed MD simulation (totally 550 μs), Shukla et al. [167] revealed the presence and structures of key intermediates and analyzed the conformational landscape of the c-Src activation pathways. With the help of MsM, two intermediate states (I1 and I2) were discovered along the transition pathway. The C-helix and β-sheets in the N-terminal lobe of c-Src were observed to form an allosteric site in I2. Additionally, 8-anilino-1-naphthalene sulfonate (ANS) toward this site inhibits c-Src by trapping the kinase in this partially active intermediate state. The discovery of this site would be both beneficial in the development of selective allosteric kinase inhibitors and useful for research about the functions of a single kinase in complex cellular signaling pathways [98]. The research above indicates that large-scale MD simulations have the ability to detect an allosteric site hidden in intermediate states. In the former example of FtsZ, molecular dynamics simulations had also identified a correct allosteric site and revealed the regulation mechanism in silico before the experimental evidence obtained by the fluorescent label [149].

6.4 Allosteric Modulators

Allosteric modulators are protein allosteric site-binding molecules that have potentially higher specificity and less toxicity than traditional orthosteric compounds [72]. Due to the advantages and advanced technology, the interest in the allosteric modulator discovery has been fueled in the past 15 years, especially in the recent

5 years. In 2000, there were fewer than 1000 allosteric molecules in total. However, the number increased to 10,000 in 2010 and has reached 80,000 to date. In the year 2013 alone, 26,308 new allosteric modulators were discovered. In short, the discovery of allosteric modulators has grown explosively. In the following section, the characteristics and main applications of allosteric modulators are introduced.

6.4.1 Characteristics of Allosteric Modulators

6.4.1.1 Classification of Allosteric Modulators

Based on the different effects on the receptor, allosteric modulators can be divided into positive allosteric modulators (PAMs), negative allosteric modulators (NAMs), and silent allosteric modulators (SAMs, also named neutral allosteric modulators). Through binding to the corresponding allosteric site, a PAM improves the action of an orthosteric effector but has no intrinsic activity. The PAM may enhance the efficacy and/or affinity of the orthosteric effector. Meanwhile, the PAM may also enhance the coupling process to the G protein. Conversely, a NAM decreases the action of an orthosteric effector and inhibits the function of the orthosteric effector by binding to the corresponding allosteric site. The NAM is regarded as a noncompetitive antagonist in the field of enzymology. In addition, a SAM occupies an allosteric site fully but exerts no pharmacological function on its own. Nevertheless, it blocks the allosteric site of both PAMs and NAMs, resulting in the inhibition of the corresponding allosteric activities [51, 196]. The modes of these interactions are shown in Fig. 6.10.

6.4.1.2 Mechanism of Allosteric Modulator Action

From a unified perspective, allosteric signal propagation pathways pre-exist in proteins, and their activation is associated with events about protein topologies such as modulator binding and covalent modifications [132, 182]. The binding of an allosteric modulator shifts the ensemble of pre-existing pathways to a specific allosteric pathway, causing the population variation of the protein [124]. In this way, the modulator stabilizes the allosterically active conformation and/or destabilizes the allosterically inactive conformation. Hence, an allosteric modulator does not create new conformations of their receptors. An allosteric modulator only changes the conformation distributions within the ensemble [131]. Moreover, via modulation of protein structural dynamics, allosteric modulators are often non-competitive effectors that are always considered as prospective drugs [72].

However, it is troublesome that only subtle differences exist between some PAMs and NAMs toward the same protein. Even identical modulators in different environments lead to largely different, even opposing, effects. The determining factors of the population shift direction need to be revealed to explain or predict the

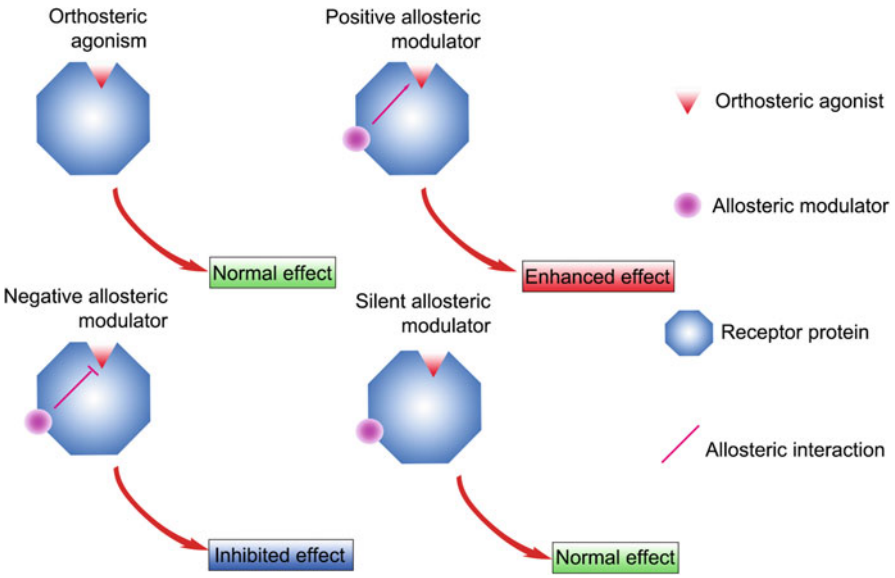


Fig. 6.10 The schematic illustration of the influences caused by different kinds of allosteric ligands

consequences of allosteric modulator binding [4, 125, 132, 159]. Faced with this problem, Nussinov and Tsai raised the concept of anchor and driver atoms [132, 137]. The atoms of allosteric modulators can be divided into anchor and driver. The anchor part docks into the allosteric site and is involved in favorable interactions with common conformations in the active and inactive states. These interactions between the anchor atoms and the receptor protein do not alter during the switch of the protein state. In contrast, the driver atoms interact with the receptor. They can be classified as attractive "pulling" atoms that stabilize the allosterically active conformations or repulsive "pushing" atoms that destabilize the allosterically inactive conformations. Thus, attractive interactions between drive atoms and allosteric site can pull the inactive conformations into active conformations or vice versa. Furthermore, it is understandable that subtle but important modifications in the drive atoms of allosteric modulators can lead to totally distinct effects. The concept of driver and anchor proposes an advanced allosteric drug design strategy based on the structural mechanisms of allosteric modulator action.

6.4.1.3 Physicochemical Properties and Chemical Features of Allosteric Modulators

Different from orthosteric ligands, allosteric modulators have some common structural and physicochemical properties that have been involved in research for years due to the importance of allosteric drug design and understanding of allostery phenomena [141, 186].

Guided by the massive information collected from ASD and ChEMBL, Smith et al. [170] investigated the features of allosteric modulators and compared them to orthosteric modulators. First, they discovered that allosteric modulators are more rigid and constrained. There are fewer bonds per heavy atom in allosteric ligands, which means fewer saturated bonds and less possibility to rotate the bonds. With the increase in aromatic structures in allosteric effectors, the allosteric effectors tend to be more rigid in allosteric interactions, though allosteric sites always undergo a conformational change [42, 141]. However, in terms of hydrophobicity, large-scale statistics revealed that only negligible differences exist between orthosteric ligands and allosteric ligands, while allosteric sites show more hydrophobicity than orthosteric sites. Meanwhile, modulators of both categories appear to cover similar chemical space such as molecule weight. The study provides insight into chemical features that can be considered in the design of allosteric drugs.

6.4.2 Allosteric Modulators in Drug Discovery

Since allostery is widespread in receptors, huge molecular machines, and protein kinases, effective drugs for multiple diseases can be developed [72]. Additionally, allosteric drugs are free from major drawbacks (e.g., off-target effect) of traditional orthosteric compounds [116]. Because of their specificity, they cause fewer side effects and are regarded as safer. They have no competition with the orthosteric ligand, so the regulation can be controlled but not produce the on/off effects [44, 132]. Moreover, the allostery-specific "effect ceiling" increases the level of drug safety under the overdose condition [64]. Thus, the development of allosteric drugs appears to be a promising trend in drug design. In the following part, we will discuss some typical cases that provide new insights into the therapeutic field, which is instructive for future drug design and the optimization of existing allosteric drugs.

6.4.2.1 GPCR Allosteric Modulators

As key cell-surface proteins, GPCRs undergo a complex conformation rearrangement upon ligand binding to transmit extracellular environmental changes into biochemical signals such as G protein activation across the membrane [99, 180]. GPCRs have been proved to be engaged in a broad array of human disorders. Thus, it is not surprising that GPCRs are the target of many therapeutic methods currently in use (e.g., benzodiazepines and γ-aminobutyric acid receptors) [35]. Since GPCRs have all evolved from the same ancestor, and the active site has to keep conserved to guarantee their function, it is difficult to discover selective orthosteric modulators for a specific GPCR subtype [34]. Nevertheless, by targeting the less conserved allosteric sites, selective modulators have been developed successfully [36]. Meanwhile, allosteric modulators do not cause receptor desensitization, which is a typical result of orthosteric modulators [193, 196]. Additionally, an

allosteric modulator can regulate specified downstream pathways but not influence others, which is depicted as functional selectivity [59]. As a result, the allosteric approach is increasingly popular in the field of GPCRs.

6.4.2.2 Kinase Allosteric Modulators

Implicated in various cellular and extracellular activities, kinases transfer the gamma phosphate of ATP onto the hydroxyl group possessed by proteins, lipids, and sugars. Kinases provide a central switching function in cellular transduction pathways such as cellular growth, proliferation, and differentiation [11, 91]. Overexpression of kinases is directly associated with many human diseases such as diabetes, Alzheimer's disease, and Parkinson's disease [150]. Thus, kinase inhibition has been indicated as a potential therapeutic target, but undesired selectivity profiles and the emergence of resistance undermine the clinical applications of kinase inhibitors [54, 198]. However, allosteric kinase inhibitors do not affect conserved ATP-binding sites, avoiding unnecessary inhibition and preventing drug toxicity. In contrast, orthosteric drugs interact with the ATP-binding sites directly, and multiple kinases are influenced due to the similarity of the substrate sites. Other advantages of allosteric modulators over ATP-competitive kinase inhibitors include the ability to overcome resistance caused by mutations in the ATP-binding site such as the common T315I mutations in the gatekeeper position of BCR-Abl [65]. In addition, allosteric kinase modulators may not need to possess affinity at the nM level to compete with the ATP with a high intracellular concentration, rendering it easier to develop and optimize allosteric inhibitors. With these attractive features, allosteric modulators are regarded as a new generation of kinase inhibitors [198].

6.4.2.3 Allosteric Network Modulators

In addition to the alteration of the single protein activity, the allosteric changes in a single protein may induce dynamic changes in proximal proteins and then have an effect on adjacent proteins. In this fashion, allosteric signals propagate in cellular networks and finally exert an effect on the distant target protein [116, 176]. Nussinov et al. ([134], Csermely et al. [38], Szilágyi et al. [176]) proposed a novel concept, "allo-network," to explain this behavior. A conventional modulator directly has an effect on an individual protein. However, an allo-network modulator initiates a perturbation signal in a protein. Then, with the help of the cellular network, the signal is enabled to reach the distant target protein. Thus, the allo-network modulator greatly expands the range of drug target proteins. Since allo-network modulators have the potential to generate a specific protein signaling pathway to change the abnormal cellular network, they possess fewer side effects and lower toxicity and have prospects for further drug development [176].

6.5 Conclusions

Defined by distant regulation in biological macromolecules, allostery has been regarded as an efficient and ubiquitous regulation of protein activities. Allostery is described dynamically, namely, the population shift of a macromolecule is caused by the binding of allosteric modulators. Allosteric regulation has many differences from orthosteric regulation, which can be applied to the prediction of allosteric interactions. Additionally, allosteric regulation is receiving increasing attention in drug development because of some unique advantages (e.g., subtype selectivity) of allostery above the differences. In the last 10 years, advanced technologies promote the identification of allosteric proteins, sites, and modulators, providing a premise for the research of allostery mechanism and the optimization of allosteric modulators. The application of allostery such as in kinase inhibitors and GPCR modulators has demonstrated the promising future of allosteric regulation. Because of the obvious merits of allosteric regulation and the considerable interest in the development, we believe that the repertoire of allostery will be burgeoning in the pharmaceutical industry and drug discovery program.

Acknowledgments This work was supported by the National Natural Science Foundation of China (21778037; Shanghai Health and Family Planning Commission (20154Y0058), the Fundamental Research Funds for the Central Universities, and the Chun-Tsung project (No.2018-02-04).

References

1. Araújo-Bazán L, Ruiz-Avila LB, Andreu D, Huecas S, Andreu JM (2016) Cytological profile of antibacterial FtsZ inhibitors and synthetic peptide MciZ. Front Microbiol 7:1558
2. Artola M, Ruíz-Avila LB, Ramírez-Aportela E, Martínez RF, Araujo-Bazán L, Vázquez-Villa H, Martín-Fontecha M, Oliva MA, Martín-Galiano AJ, Chacón P (2017) The structural assembly switch of cell division protein FtsZ probed with fluorescent allosteric inhibitors. Chem Sci 8:1525–1534
3. Babel H, Bischofs IB (2016) Molecular and cellular factors control signal transduction via switchable allosteric modulator proteins (SAMPs). BMC Syst Biol 10:35
4. Badireddy S, Yunfeng G, Ritchie M, Akamine P, Wu J, Kim CW, Taylor SS, Qingsong L, Swaminathan K, Anand GS (2010) Cyclic AMP analog blocks kinase activation by stabilizing inactive conformation: conformational selection highlights a new concept in allosteric inhibitor design. Mol Cell Proteomics 10. https://doi.org/10.1074/mcp.M110.004390
5. Bai X, Yan C, Yang G, Lu P, Ma D, Sun L, Zhou R, Scheres SHW, Shi Y (2015) An atomic structure of human γ-secretase. Nature 525:212–217
6. Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P (2016) 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. Science 351:871–875
7. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. Bioinformatics 20:1322–1324
8. Bessa LM, Launay H, Dujardin M, Cantrelle F-X, Lippens G, Landrieu I, Schneider R, Hanoulle X (2017) NMR reveals the intrinsically disordered domain 2 of NS5A protein as

an allosteric regulator of the hepatitis C virus RNA polymerase NS5B. J Biol Chem 292:18024–18043
9. Betzi S, Alam R, Martin M, Lubbers DJ, Han H, Jakkaraj SR, Georg GI, Schönbrunn E (2011) Discovery of a potential allosteric ligand binding site in CDK2. ACS Chem Biol 6:492–501
10. Bian Z-M, Elner SG, Khanna H, Murga-Zamalloa CA, Patil S, Elner VM (2011) Expression and functional roles of caspase-5 in inflammatory responses of human retinal pigment epithelial cells. Invest Ophthalmol Vis Sci 52:8646–8656
11. Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. Nature 411:355–365
12. Bock A, Schrage R, Mohr K (2018) Allosteric modulators targeting CNS muscarinic receptors. Neuropharmacology 136:427–437
13. Bonomi M, Pellarin R, Vendruscolo M (2018) Simultaneous determination of protein structure and dynamics using cryo-electron microscopy. Biophys J 114:1604–1613
14. Boulton S, Melacini G (2016) Advances in NMR methods to map allosteric sites: from models to translation. Chem Rev 116:6267–6304
15. Bowman GR, Bolin ER, Hart KM, Maguire BC, Marqusee S (2015) Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. Proc Natl Acad Sci U S A 112:2734–2739
16. Brignole EJ, Tsai K-L, Chittuluru J, Li H, Aye Y, Penczek PA, Stubbe J, Drennan CL, Asturias F (2018) 3.3-Å resolution cryo-EM structure of human ribonucleotide reductase with substrate and allosteric regulators bound. Elife 7:e31502
17. Bryant C, Fitzgerald KA (2009) Molecular mechanisms involved in inflammasome activation. Trends Cell Biol 19:455–464
18. Buzko O, Shokat KM (2002) A kinase sequence database: sequence alignments and family assignment. Bioinformatics 18:1274–1275
19. Campitelli P, Guo J, Zhou H-X, Ozkan SB (2018) Hinge-shift mechanism modulates allosteric regulations in human pin1. J Phys Chem B 122:5623–5629
20. Capdevila DA, Braymer JJ, Edmonds KA, Wu H, Giedroc DP (2017) Entropy redistribution controls allostery in a metalloregulatory protein. Proc Natl Acad Sci U S A 114:4424–4429
21. Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, Pupko T, Ben-Tal N (2013) ConSurf: using evolutionary data to raise testable hypotheses about protein function. Isr J Chem 53:199–206
22. Chandramohan A, Krishnamurthy S, Larsson A, Nordlund P, Jansson A, Anand GS (2016) Predicting allosteric effects from orthosteric binding in Hsp90-ligand interactions: implications for fragment-based drug design. PLoS Comput Biol 12:e1004840
23. Changeux J-P (2010) Allosteric receptors: from electric organ to cognition. Annu Rev Pharmacol Toxicol 50:1–38
24. Changeux J-P (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 41:103–133
25. Changeux J-P (2013) 50 years of allosteric interactions: the twists and turns of the models. Nat Rev Mol cell Biol 14:819–829
26. Chen I (2013) Allostery through DNA. Nat Struct & Mol Biol 20:410
27. Chen H, Marsiglia WM, Cho M-K, Huang Z, Deng J, Blais SP, Gai W, Bhattacharya S, Neubert TA, Traaseth NJ (2017) Elucidation of a four-site allosteric network in fibroblast growth factor receptor tyrosine kinases. Elife 6:e21137
28. Cheng RKY, Fiez-Vandal C, Schlenker O, Edman K, Aggeler B, Brown DG, Brown GA, Cooke RM, Dumelin CE, Doré AS, Geschwindner S, Grebner C, Hermansson N-O, Jazayeri A, Johansson P, Leong L, Prihandoko R, Rappas M, Soutter H, Snijder A, Sundström L, Tehan B, Thornton P, Troast D, Wiggin G, Zhukov A, Marshall FH, Dekker N (2017) Structural insight into allosteric modulation of protease-activated receptor 2. Nature 545:112–115
29. Chodera JD, Noé F (2014) Markov state models of biomolecular conformational dynamics. Curr Opin Struct Biol 25:135–144

30. Christopoulos A (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat Rev Drug Discov 1:198–210
31. Coleman JA, Green EM, Gouaux E (2016) X-ray structures and mechanism of the human serotonin transporter. Nature 532:334–339
32. Collier G, Ortiz V (2013) Emerging computational approaches for the study of protein allostery. Arch Biochem Biophys 538:6–15
33. Cong X, Liu Y, Liu W, Liang X, Laganowsky A (2017) Allosteric modulation of protein-protein interactions by individual lipid binding events. Nat Commun 8:2203
34. Congreve M, Oswald C, Marshall FH (2017) Applying structure-based drug design approaches to allosteric modulators of GPCRs. Trends Pharmacol Sci 38:837–847
35. Conn PJ, Christopoulos A, Lindsley CW (2009a) Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. Nat Rev Drug Discov 8:41–54
36. Conn PJ, Jones CK, Lindsley CW (2009b) Subtype-selective allosteric modulators of muscarinic receptors for the treatment of CNS disorders. Trends Pharmacol Sci 30:148–155
37. Conn PJ, Lindsley CW, Meiler J, Niswender CM (2014) Opportunities and challenges in the discovery of allosteric modulators of GPCRs for treating CNS disorders. Nat Rev Drug Discov 13:692–708
38. Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther 138:333–408
39. Csizmok V, Orlicky S, Cheng J, Song J, Bah A, Delgoshaie N, Lin H, Mittag T, Sicheri F, Chan HS, Tyers M, Forman-Kay JD (2017) An allosteric conduit facilitates dynamic multisite substrate recognition by the SCFCdc4 ubiquitin ligase. Nat Commun 8:13943
40. Das D, Krantz BA (2016) Peptide- and proton-driven allosteric clamps catalyze anthrax toxin translocation across membranes. Proc Natl Acad Sci U S A 113:9611–9616
41. De Cesco S, Kurian J, Dufresne C, Mittermaier AK, Moitessier N (2017) Covalent inhibitors design and discovery. Eur J Med Chem 138:96–114
42. Demerdash ONA, Daily MD, Mitchell JC (2009) Structure-based predictive models for allosteric hot spots. PLoS Comput Biol 5:e1000531
43. Di Russo NV, Martí MA, Roitberg AE (2014) Underlying thermodynamics of ph-dependent allostery. J Phys Chem B 118:12818–12826. https://doi.org/10.1021/jp507971v
44. Digby GJ, Noetzel MJ, Bubser M, Utley TJ, Walker AG, Byun NE, Lebois EP, Xiang Z, Sheffler DJ, Cho HP (2012) Novel allosteric agonists of M1 muscarinic acetylcholine receptors induce brain region-specific responses that correspond with behavioral effects in animal models. J Neurosci 32:8532–8544
45. Dokholyan NV (2016) Controlling Allosteric Networks in Proteins. Chem Rev 116:6463–6487
46. Doré AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B (2014) Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 511:557–561
47. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, Shan Y, Xu H, Shaw DE (2011) Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc Natl Acad Sci U S A 108:13118–13123
48. Dror RO, Green HF, Valant C, Borhani DW, Valcourt JR, Pan AC, Arlow DH, Canals M, Lane JR, Rahmani R, Baell JB, Sexton PM, Christopoulos A, Shaw DE (2013) Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature 503:295–299
49. Eddy MT, Lee M-Y, Gao Z-G, White KL, Didenko T, Horst R, Audet M, Stanczak P, McClary KM, Han GW, Jacobson KA, Stevens RC, Wüthrich K (2018) Allosteric coupling of drug binding and intracellular signaling in the A2A adenosine receptor. Cell 172:68–80.e12. https://doi.org/10.1016/j.cell.2017.12.004
50. Elsen NL, Lu J, Parthasarathy G, Reid JC, Sharma S, Soisson SM, Lumb KJ (2012) Mechanism of action of the cell-division inhibitor PC190723: modulation of FtsZ assembly cooperativity. J Am Chem Soc 134:12342–12345

51. Engers DW, Lindsley CW (2013) Allosteric modulation of Class C GPCRs: a novel approach for the treatment of CNS disorders. Drug Discov Today Technol 10:e269–e276
52. Fang Z, Grütter C, Rauh D (2012) Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features. ACS Chem Biol 8:58–70
53. Fenton AW (2008) Allostery: an illustrated definition for the 'second secret of life.'. Trends Biochem Sci 33:420–425
54. Foda ZH, Seeliger MA (2014) An allosteric add-on. Nat Chem Biol 10:796–797
55. Foda ZH, Shan Y, Kim ET, Shaw DE, Seeliger MA (2015) A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. Nat Commun 6:5939
56. Gafurov B, Chen Y-D, Chalovich JM (2004) Ca(2+) and ionic strength dependencies of S1-ADP binding to actin-tropomyosin-troponin: regulatory implications. Biophys J 87:1825–1835. https://doi.org/10.1529/biophysj.104.043364
57. Ganesan A, Coote ML, Barakat K (2017) Molecular dynamics-driven drug discovery: leaping forward with confidence. Drug Discov Today 22:249–269
58. Gangoso E, Talaverón R, Jaraíz-Rodríguez M, Domínguez-Prieto M, Ezan P, Koulakoff A, Medina JM, Giaume C, Tabernero A (2017) A c-Src inhibitor peptide based on connexin43 exerts neuroprotective effects through the inhibition of glial hemichannel activity. Front Mol Neurosci 10:418
59. Gao Z-G, Jacobson KA (2013) Allosteric modulation and functional selectivity of G protein-coupled receptors. Drug Discov Today Technol 10:e237–e243
60. Gao J, Wells JA (2012) Identification of specific tethered inhibitors for caspase-5. Chem Biol Drug Des 79:209–215
61. Gao Y, Shen L, Honzatko RB (2014) Central cavity of fructose-1,6-bisphosphatase and the evolution of AMP/fructose 2,6-bisphosphate synergism in eukaryotic organisms. J Biol Chem 289:8450–8461
62. Gardino AK, Villali J, Kivenson A, Lei M, Liu CF, Steindel P, Eisenmesser EZ, Labeikovsky W, Wolf-Watz M, Clarkson MW (2009) Transient non-native hydrogen bonds promote activation of a signaling protein. Cell 139:1109–1118
63. Gasper PM, Fuglestad B, Komives EA, Markwick PRL, McCammon JA (2012) Allosteric networks in thrombin distinguish procoagulant vs. anticoagulant activities. Proc Natl Acad Sci U S A 109:21216–21222
64. Gentry PR, Sexton PM, Christopoulos A (2015) Novel allosteric modulators of G protein-coupled receptors. J Biol Chem 290:19478–19488
65. Gibbons DL, Pricl S, Kantarjian H, Cortes J, Quintás-Cardama A (2012) The rise and fall of gatekeeper mutations? the BCR-ABL1 T315I paradigm. Cancer 118:293–299
66. Gill-Thind JK, Dhankher P, D'Oyley JM, Sheppard TD, Millar NS (2014) Structurally similar allosteric modulators of α7 nicotinic acetylcholine receptors exhibit five distinct pharmacological effects. J Biol Chem 290:3552–3562
67. Goldenberg O, Erez E, Nimrod G, Ben-Tal N (2008) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. Nucleic Acids Res 37:D323–D327
68. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. Nat Chem Biol 4:474–482
69. Greener JG, Sternberg MJE (2018) Structure-based prediction of protein allostery. Curr Opin Struct Biol 50:1–8
70. Grimm SS, Isacoff EY (2016) Allosteric substrate switching in a voltage sensing lipid phosphatase. Nat Chem Biol 12:261–267
71. Grover AK (2013) Use of allosteric targets in the discovery of safer drugs. Med Princ Pract 22:418–426
72. Guarnera E, Berezovsky IN (2016) Allosteric sites: remote control in regulation of protein activity. Curr Opin Struct Biol 37:1–8
73. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins Struct Funct Bioinforma 57:433–443

74. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138:774–786
75. Handley LD, Fuglestad B, Stearns K, Tonelli M, Fenwick RB, Markwick PRL, Komives EA (2017) NMR reveals a dynamic allosteric pathway in thrombin. Sci Rep 7:39575
76. Hardy JA, Wells JA (2004) Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 14:706–715
77. Hertig S, Latorraca NR, Dror RO (2016) Revealing atomic-level mechanisms of protein allostery with molecular dynamics simulations. PLoS Comput Biol 12:e1004746
78. Hines JK, Fromm HJ, Honzatko RB (2007) Structures of activated fructose-1, 6-bisphosphatase from Escherichia coli: coordinate regulation of bacterial metabolism and the conservation of the r-state. J Biol Chem 282:11696–11704
79. Hocker HJ, Cho K-J, Chen C-YK, Rambahal N, Sagineedu SR, Shaari K, Stanslas J, Hancock JF, Gorfe AA (2013) Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. Proc Natl Acad Sci U S A 110:10201–10206
80. Hu Y, Li S, Liu F, Geng L, Shu X, Zhang J (2015) Discovery of novel nonpeptide allosteric inhibitors interrupting the interaction of CDK2/cyclin A3 by virtual screening and bioassays. Bioorg Med Chem Lett 25:4069–4073
81. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39:D663–D669
82. Huang M, Lu S, Shi T, Zhao Y, Chen Y, Li X, Liu X, Huang Z, Zhang J (2013a) Conformational transition pathway in the activation process of allosteric glucokinase. PLoS One 8: e55857
83. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T (2013b) Allosite: a method for predicting allosteric sites. Bioinformatics 29:2357–2359
84. Huang Z, Mou L, Shen Q, Lu S, Li C, Liu X, Wang G, Li S, Geng L, Liu Y (2014) ASD v2. 0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42: D510–D516
85. Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, Gao J, Hong Y, Li Q, Ni D, Xu J, Chen G, Zhang J (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. Nucleic Acids Res 46:W451–W458
86. Isaacs JS, Xu W, Neckers L (2003) Heat shock protein 90 as a molecular target for cancer therapeutics. Cancer Cell 3:213–217
87. Isherwood SN, Robbins TW, Dalley JW, Pekcec A (2018) Bidirectional variation in glutamate efflux in the medial prefrontal cortex induced by selective positive and negative allosteric mGluR5 modulators. J Neurochem 145:111–124
88. Jacobson KA (2015) New paradigms in GPCR drug discovery. Biochem Pharmacol 98:541–555
89. Jaffe EK (2005) Morpheeins–a new structural paradigm for allosteric regulation. Trends Biochem Sci 30:490–497
90. Jaffe EK, Lawrence SH (2012) The morpheein model of allostery: evaluating proteins as potential morpheeins. In: Allostery. Springer, pp 217–231. https://doi.org/10.1007/978-1-61779-334-9
91. Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101:2209–2242
92. Kar G, Keskin O, Gursoy A, Nussinov R (2010) Allostery and population shift in drug discovery. Curr Opin Pharmacol 10:715–722
93. Kenakin T (2007) Collateral efficacy in drug discovery: taking advantage of the good (allosteric) nature of 7TM receptors. Trends Pharmacol Sci 28:407–415
94. Kenakin TP (2017) Chapter 5 – allosteric drug effects. In: Kenakin TPBT-P in DD and D, 2nd edn. Academic Press, Amsterdam, pp 101–129
95. Kim S, Broströmer E, Xing D, Jin J, Chong S, Ge H, Wang S, Gu C, Yang L, Gao YQ, Su X, Sun Y, Xie XS (2013) Probing Allostery Through DNA. Science (80-) 339:816–819

96. Kim JG, Kim TW, Kim J, Ihee H (2015) Protein structural dynamics revealed by time-resolved X-ray solution scattering. Acc Chem Res 48:2200–2208
97. Kityk R, Vogel M, Schlecht R, Bukau B, Mayer MP (2015) Pathways of allosteric regulation in Hsp70 chaperones. Nat Commun 6:8308
98. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology. Nat Rev Cancer 10:130–137
99. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. Trends Pharmacol Sci 28:397–406
100. Korczynska M, Clark MJ, Valant C, Xu J, Von Moo E, Albold S, Weiss DR, Torosyan H, Huang W, Kruse AC, Lyda BR, May LT, Baltos J-A, Sexton PM, Kobilka BK, Christopoulos A, Shoichet BK, Sunahara RK (2018) Structure-based discovery of selective positive allosteric modulators of antagonists for the M2 muscarinic acetylcholine receptor. Proc Natl Acad Sci U S A 115:E2419–E2428. https://doi.org/10.1073/pnas.1718037115
101. Kornev AP, Taylor SS (2015) Dynamics-driven allostery in protein kinases. Trends Biochem Sci 40:628–647
102. Koshland DE Jr, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5:365–385
103. Kruger FA, Overington JP (2012) Global analysis of small molecule binding to related protein targets. PLoS Comput Biol 8:e1002333
104. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hübner H, Pardon E, Valant C, Sexton PM (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature 504:101–106
105. Lazareno S, Doležal V, Popham A, Birdsall NJM (2004) Thiochrome enhances acetylcholine affinity at muscarinic M4 receptors: receptor subtype selectivity via cooperativity rather than affinity. Mol Pharmacol 65:257–266
106. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R (2008) Surface sites for engineering allosteric control in proteins. Science (80-) 322:438–442
107. Lee Y, Choi S, Hyeon C (2014) Mapping the intramolecular signal transduction of G-protein coupled receptors. Proteins Struct Funct Bioinforma 82:727–743
108. Levy ED, Teichmann SA (2013) Structural, evolutionary, and assembly principles of protein oligomerization. Prog Mol Biol Transl Sci 117:25–51
109. Li X, Chen Y, Lu S, Huang Z, Liu X, Wang Q, Shi T, Zhang J (2013) Toward an understanding of the sequence and structural basis of allosteric proteins. J Mol Graph Model 40:30–39
110. Liu X, Ahn S, Kahsai AW, Meng K-C, Latorraca NR, Pani B, Venkatakrishnan AJ, Masoudi A, Weis WI, Dror RO, Chen X, Lefkowitz RJ, Kobilka BK (2017) Mechanism of intracellular allosteric β2AR antagonist revealed by X-ray crystal structure. Nature 548:480–484
111. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science (80-) 286:295–299
112. Louet M, Seifert C, Hensen U, Gräter F (2015) Dynamic allostery of the catabolite activator protein revealed by interatomic forces. PLoS Comput Biol 11:e1004358
113. Lu S, Zhang J (2017) Designed covalent allosteric modulators: an emerging paradigm in drug discovery. Drug Discov Today 22:447–453. https://doi.org/10.1016/j.drudis.2016.11.013
114. Lu S, Zhang J (2018) Small molecule allosteric modulators of G-protein-coupled receptors: drug–target interactions. J Med Chem. https://doi.org/10.1021/acs.jmedchem.7b01844
115. Lu S, Huang W, Zhang J (2014a) Recent computational advances in the identification of allosteric sites in proteins. Drug Discov Today 19:1595–1600
116. Lu S, Li S, Zhang J (2014b) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34:1242–1285
117. Lu S, Jang H, Muratcioglu S, Gursoy A, Keskin O, Nussinov R, Zhang J (2016) Ras conformational ensembles, allostery, and signaling. Chem Rev 116:6607–6665

118. Lu S, Ji M, Ni D, Zhang J (2017) Discovery of hidden allosteric sites as novel targets for allosteric drug design. Drug Discov Today 23:359–365
119. Lü W, Du J, Goehring A, Gouaux E (2017) Cryo-EM structures of the triheteromeric NMDA receptor and its allosteric modulation. Science (80-) 355:eaal3729
120. Lutkenhaus J, Pichoff S, Du S (2012) Bacterial cytokinesis: from Z ring to divisome. Cytoskeleton 69:778–790
121. Macpherson JA, Anastasiou D (2017) Allosteric regulation of metabolism in cancer: endogenous mechanisms and considerations for drug design. Curr Opin Biotechnol 48:102–110
122. Modesti M (2011) Fluorescent labeling of proteins. In: Single molecule analysis. Springer, New York, pp 101–120
123. Monod J, Wyman J, Changeux J-P (1978) On the nature of allosteric transitions: a plausible model. In: Selected papers in molecular biology by Jacques Monod. Elsevier, Amsterdam, pp 593–623
124. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339
125. Muchmore DB (2000) Raloxifene: a selective estrogen receptor modulator (SERM) with multiple target system effects. Oncologist 5:388–392
126. Newcombe J, Chatzidaki A, Sheppard TD, Topf M, Millar NS (2018) Diversity of nicotinic acetylcholine receptor positive allosteric modulators revealed by mutagenesis and a revised structural model. Mol Pharmacol 93:128–140
127. Ni D, Lu S, Zhang J (2018) Methods applied for the allosteric site revelation. Encycl Anal Chem:1–13. https://doi.org/10.1002/9780470027318.a9614
128. Niu W, Wang J, Qian J, Wang M, Wu P, Chen F, Yan S (2018) Allosteric control of human cystathionine β-synthase activity by a redox active disulfide bond. J Biol Chem 293:2523–2533
129. Nussinov R (2016) Introduction to protein ensembles and allostery. Chem Rev 116:6263–6266
130. Nussinov R, Tsai C-J (2012) The different ways through which specificity works in orthosteric and allosteric drugs. Curr Pharm Des 18:1311–1316
131. Nussinov R, Tsai C-J (2013) Allostery in disease and in drug discovery. Cell 153:293–305
132. Nussinov R, Tsai C-J (2014) Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol Sci 35:256–264
133. Nussinov R, Tsai C-J (2015) The design of covalent allosteric drugs. Annu Rev Pharmacol Toxicol 55:249–267
134. Nussinov R, Tsai C-J, Csermely P (2011) Allo-network drugs: harnessing allostery in cellular networks. Trends Pharmacol Sci 32:686–693
135. Nussinov R, Tsai C-J, Xin F, Radivojac P (2012) Allosteric post-translational modification codes. Trends Biochem Sci 37:447–455
136. Nussinov R, Ma B, Tsai C-J, Csermely P (2013) Allosteric conformational barcodes direct signaling in the cell. Structure 21:1509–1521
137. Nussinov R, Tsai C-J, Liu J (2014) Principles of allosteric interactions in cell signaling. J Am Chem Soc 136:17692–17701
138. Omer A, Prasad CVSS (2012) Designing allosteric modulators for active conformational state of m-glutamate G-protein coupled receptors. Bioinformation 8:170–174
139. Otero LH, Rojas-Altuve A, Llarrull LI, Carrasco-López C, Kumarasiri M, Lastochkin E, Fishovitz J, Dawley M, Hesek D, Lee M, Johnson JW, Fisher JF, Chang M, Mobashery S, Hermoso JA (2013) How allosteric control of Staphylococcus aureus penicillin binding protein 2a enables methicillin resistance and physiological function. Proc Natl Acad Sci U S A 110:16808–16813
140. Pan Y, Tsai C-J, Ma B, Nussinov R (2010) Mechanisms of transcription factor selectivity. Trends Genet 26:75–83
141. Panjkovich A, Daura X (2012) Exploiting protein flexibility to predict the location of allosteric sites. BMC Bioinformatics 13:273

142. Patricelli MP, Janes MR, Li L-S, Hansen R, Peters U, Kessler LV, Chen Y, Kucharski JM, Feng J, Ely T, Chen JH, Firdaus SJ, Babbar A, Ren P, Liu Y (2016) Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. Cancer Discov 6:316–329
143. Pfaff SJ, Fletterick RJ (2010) Hormone and co-regulator binding to the glucocorticoid receptor are allosterically coupled. J Biol Chem 285:15256–15267
144. Pinney MM, Natarajan A, Yabukarski F, Sanchez DM, Liu F, Liang R, Doukov T, Schwans JP, Martinez TJ, Herschlag D (2018) Structural coupling throughout the active site hydrogen bond networks of ketosteroid isomerase and photoactive yellow protein. J Am Chem Soc 140:9827–9843
145. Plattner N, Doerr S, De Fabritiis G, Noé F (2017) Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nat Chem 9:1005–1011
146. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13:831–838
147. Prakash P, Sayyed-Ahmad A, Gorfe AA (2015) pMD-Membrane: a method for ligand binding site Identification in membrane-bound proteins. PLoS Comput Biol 11:e1004469
148. Rabiller M, Getlik M, Klüter S, Richters A, Tückmantel S, Simard JR, Rauh D (2010) Proteus in the world of proteins: conformational changes in protein kinases. Arch der Pharm An Int J Pharm Med Chem 343:193–206
149. Ramírez-Aportela E, López-Blanco JR, Andreu JM, Chacón P (2014) Understanding nucleotide-regulated FtsZ filament dynamics and the monomer assembly switch with large-scale atomistic simulations. Biophys J 107:2164–2176
150. Rask-Andersen M, Zhang J, Fabbro D, Schiöth HB (2014) Advances in kinase targeting: current clinical use and clinical trials. Trends Pharmacol Sci 35:604–620
151. Rescifina A, Scala A, Sciortino MT, Colao I, Siracusano G, Mazzaglia A, Chiacchio U, Grassi G (2015) Decorated 6, 6′, 7, 7′-tetrahydro-1 H, 1′ H-2, 3′-biindole scaffold as promising candidate for recognition of the CDK2 allosteric site. Medchemcomm 6:311–318
152. Reynolds KA, McLaughlin RN, Ranganathan R (2011) Hot spots for allosteric regulation on protein surfaces. Cell 147:1564–1575
153. Rose PW, Bi C, Bluhm WF, Christie CH, Dimitropoulos D, Dutta S, Green RK, Goodsell DS, Prlić A, Quesada M (2012) The RCSB Protein Data Bank: new resources for research and education. Nucleic Acids Res 41:D475–D482
154. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Green RK, Goodsell DS, Hudson B, Kalro T, Lowe R, Ezra Peisach SKB (2017) The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res 45:D271–D281
155. Roskoski R (2004) Src protein–tyrosine kinase structure and regulation. Biochem Biophys Res Commun 324:1155–1164
156. Roth BL, Irwin JJ, Shoichet BK (2017) Discovery of new GPCR ligands to illuminate new biology. Nat Chem Biol 13:1143–1151
157. Ruff EF, Muretta JM, Thompson AR, Lake EW, Cyphers S, Albanese SK, Hanson SM, Behr JM, Thomas DD, Chodera JD (2018) A dynamic mechanism for allosteric activation of Aurora kinase A by activation loop phosphorylation. Elife 7:e32766
158. Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ (2018) Dynamic allostery can drive cold adaptation in enzymes. Nature 558:324–328
159. Sadowsky JD, Burlingame MA, Wolan DW, McClendon CL, Jacobson MP, Wells JA (2011) Turning a protein kinase on or off from a single allosteric site via disulfide trapping. Proc Natl Acad Sci U S A 108:6056–6061
160. Scheer JM, Romanowski MJ, Wells JA (2006) A common allosteric site and mechanism in caspases. Proc Natl Acad Sci U S A 103:7595–7600

161. Schelshorn D, Joly F, Mutel S, Hampe C, Breton B, Mutel V, Lütjens R (2012) Lateral allosterism in the glucagon receptor family: glucagon-like peptide 1 induces G-protein-coupled receptor heteromer formation. Mol Pharmacol 81:309–318
162. Schiffmann A, Gimpl G (2018) Sodium functions as a negative allosteric modulator of the oxytocin receptor. Biochim Biophys Acta – Biomembr 1860:1301–1308
163. Schnell JR, Dyson HJ, Wright PE (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. Annu Rev Biophys Biomol Struct 33:119–140
164. Sebastian B, Kakizuka A, Hunter T (1993) Cdc25M2 activation of cyclin-dependent kinases by dephosphorylation of threonine-14 and tyrosine-15. Proc Natl Acad Sci U S A 90:3521–3524
165. Shen Q, Wang G, Li S, Liu X, Lu S, Chen Z, Song K, Yan J, Geng L, Huang Z, Huang W, Chen G, Zhang J (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res 44:D527–D535
166. Shi Y (2014) A glimpse of structural biology through X-ray crystallography. Cell 159:995–1014
167. Shukla D, Meng Y, Roux B, Pande VS (2014) Activation pathway of Src kinase reveals intermediate states as targets for drug design. Nat Commun 5:3397
168. Sinha N, Nussinov R (2001) Point mutations and sequence variability in proteins: redistributions of preexisting populations. Proc Natl Acad Sci U S A 98:3139–3144
169. Skora L, Mestan J, Fabbro D, Jahnke W, Grzesiek S (2013) NMR reveals the allosteric opening and closing of Abelson tyrosine kinase by ATP-site and myristoyl pocket inhibitors. Proc Natl Acad Sci U S A 110:E4437–E4445
170. Smith RD, Lu J, Carlson HA (2017) Are there physicochemical differences between allosteric and competitive ligands? PLoS Comput Biol 13:e1005813
171. Song GJ, Jung M, Kim J-H, Park H, Rahman MH, Zhang S, Zhang Z-Y, Park DH, Kook H, Lee I-K (2016) A novel role for protein tyrosine phosphatase 1B as a positive regulator of neuroinflammation. J Neuroinflammation 13:86
172. Song K, Liu X, Huang W, Lu S, Shen Q, Zhang L, Zhang J (2017) Improved method for the identification and validation of allosteric sites. J Chem Inf Model 57:2358–2363
173. Stetz G, Verkhivker GM (2017) Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: a community-hopping model of allosteric regulation and communication. PLoS Comput Biol 13:e1005299
174. Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102–1104
175. Strickland D, Moffat K, Sosnick TR (2008) Light-activated DNA binding in a designed allosteric protein. Proc Natl Acad Sci U S A 105:10709–10714
176. Szilágyi A, Nussinov R, Csermely P (2013) Allo-network drugs: extension of the allosteric drug concept to protein-protein interaction and signaling networks. Curr Top Med Chem 13:64–77
177. Tan CM, Therien AG, Lu J, Lee SH, Caron A, Gill CJ, Lebeau-Jacob C, Benton-Perdomo L, Monteiro JM, Pereira PM (2012) Restoring methicillin-resistant Staphylococcus aureus susceptibility to β-lactam antibiotics. Sci Transl Med 4:126ra35
178. Taylor SS, Kornev AP (2011) Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci 36:65–77
179. Teşileanu T, Colwell LJ, Leibler S (2015) Protein sectors: Statistical coupling analysis versus conservation. PLoS Comput Biol 11:e1004091
180. Thal DM, Glukhova A, Sexton PM, Christopoulos A (2018) Structural insights into G-protein-coupled receptor allostery. Nature 559:45–53
181. Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37:509–516
182. Tsai C-J, Nussinov R (2014) A Unified View of "How Allostery Works.". PLOS Comput Biol 10:e1003394

183. Tzeng S-R, Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. Curr Opin Struct Biol 21:62–67. https://doi.org/10.1016/j.sbi.2010.10.007
184. Urwyler S (2011) Allosteric modulation of family C G-protein-coupled receptors: from molecular insights to therapeutic perspectives. Pharmacol Rev 63:59–126
185. Van Wart AT, Durrant J, Votapka L, Amaro RE (2014) Weighted implementation of suboptimal paths (WISP): an optimized algorithm and tool for dynamical network analysis. J Chem Theory Comput 10:511–517
186. van Westen GJP, Gaulton A, Overington JP (2014) Chemical, target, and bioactive properties of allosteric modulation. PLOS Comput Biol 10:e1003559
187. Verespy S III, Mehta AY, Afosah D, Al-Horani RA, Desai UR (2016) Allosteric partial inhibition of monomeric proteases. Sulfated coumarins induce regulation, not just inhibition, of thrombin. Sci Rep 6:24043
188. Verkhivker GM (2016) Molecular dynamics simulations and modelling of the residue interaction networks in the BRAF kinase complexes with small molecule inhibitors: probing the allosteric effects of ligand-induced kinase dimerization and paradoxical activation. Mol Biosyst 12:3146–3165
189. Verkhivker GM (2017) Leveraging structural diversity and allosteric regulatory mechanisms of protein kinases in the discovery of small molecule inhibitors. Curr Med Chem 24:4838–4872
190. Viganò E, Diamond CE, Spreafico R, Balachander A, Sobota RM, Mortellaro A (2015) Human caspase-4 and caspase-5 regulate the one-step non-canonical inflammasome activation in monocytes. Nat Commun 6:8761
191. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE (2016) Emerging computational methods for the rational discovery of allosteric drugs. Chem Rev 116:6370–6390
192. Wei H, Mo J, Tao L, Russell RJ, Tymiak AA, Chen G, Iacob RE, Engen JR (2014) Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. Drug Discov Today 19:95–102
193. Wenthur CJ, Gentry PR, Mathews TP, Lindsley CW (2014) Drugs for allosteric sites on receptors. Annu Rev Pharmacol Toxicol 54:165–184
194. Whittaker SR, Barlow C, Martin MP, Mancusi C, Wagner S, Self A, Barrie E, Te Poele R, Sharp S, Brown N (2018) Molecular profiling and combinatorial activity of CCT068127: a potent CDK2 and CDK9 inhibitor. Mol Oncol 12:287–304
195. Wilhelm A, Lopez-Garcia LA, Busschots K, Fröhner W, Maurer F, Boettcher S, Zhang H, Schulze JO, Biondi RM, Engel M (2012) 2-(3-Oxo-1,3-diphenylpropyl)malonic acids as potent allosteric ligands of the PIF pocket of phosphoinositide-dependent Kinase-1: development and prodrug concept. J Med Chem 55:9817–9830. https://doi.org/10.1021/jm3010477
196. Wootten D, Christopoulos A, Sexton PM (2013) Emerging paradigms in GPCR allostery: implications for drug discovery. Nat Rev Drug Discov 12:630–644
197. Wright PE, Dyson HJ (2014) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18–29
198. Wu P, Clausen MH, Nielsen TE (2015) Allosteric small-molecule kinase inhibitors. Pharmacol Ther 156:59–68
199. Xu Y, Smith R, Vivoli M, Ema M, Goos N, Gehrke S, Harmer NJ, Wagner GK (2017) Covalent inhibitors of LgtC: a blueprint for the discovery of non-substrate-like inhibitors for bacterial glycosyltransferases. Bioorg Med Chem 25:3182–3194
200. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46:W374–W379
201. Yang J-S, Seo SW, Jang S, Jung GY, Kim S (2012) Rational engineering of enzyme allosteric regulation through sequence evolution analysis. PLoS Comput Biol 8:e1002612

202. Yau M-K, Liu L, Fairlie DP (2013) Toward drugs for protease-activated receptor 2 (PAR2). J Med Chem 56:7477–7497
203. Youssef EA, Berry-Kravis E, Czech C, Hagerman RJ, Hessl D, Wong CY, Rabbia M, Deptula D, John A, Kinch R (2018) Effect of the mGluR5-NAM basimglurant on behavior in adolescents and adults with Fragile X syndrome in a randomized, double-blind, placebo-controlled trial: fragXis phase 2 results. Neuropsychopharmacology 43:503–512
204. Zhu R, Song Y, Liu H, Yang Y, Wang S, Yi C, Chen PR (2017) Allosteric histidine switch for regulation of intracellular zinc(II) fluctuation. Proc Natl Acad Sci U S A 114:13661–13666
205. Zuehlke AD, Moses MA, Neckers L (2018) Heat shock protein 90: its inhibition and function. Phil Trans R Soc B 373:20160527

Chapter 7
Advances in the Computational Identification of Allosteric Sites and Pathways in Proteins

Xavier Daura

Abstract With the increasing difficulty to develop new drugs and the emergence of resistance to traditional orthosteric-site inhibitors, the search for alternatives is finally approaching the focus on allosteric sites. Allosteric sites offer opportunities to regulate many pharmacologically targeted pathways by inhibition or activation. In addition, allosteric sites tend to be less conserved than the functional site, which may facilitate the design of specific effectors in the protein families for which specific orthosteric inhibitors have proved difficult to design. Furthermore, recent evidence suggests that all proteins might be susceptible of allosteric regulation, increasing the space of druggable targets. Computational identification of allosteric sites has therefore become an active field of research. The problem can be approached from two sides: (1) the identification of allosteric-communication pathways between the functional site and potential allosteric sites and (2) the functional-site-independent identification of allosteric sites. While the first approach tends to be more laborious and thus restricted to a single protein, the second tends to be more amenable to larger-scale analysis, thus providing tools for the two drug discovery scenarios: the analysis of known targets and the screening for new potential targets. Here, I show some basic concepts and methods useful to the identification of allosteric sites and pathways, in line with these two approaches. I describe them in some detail to build a clear framework, at the risk of losing the interest of experts. Examples of recent studies involving these methods are also illustrated, focusing on the techniques rather than on their findings on allosterism.

Keywords Proteins · Allosteric site · allosteric pathway · computational models · molecular dynamics simulation · elastic network model · covariance · mutual information · statistical coupling analysis · energy coupling · perturbation methods

X. Daura (✉)
Catalan Institution for Research and Advanced Studies (ICREA) and Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain
e-mail: xavier.daura@uab.cat

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_7

7.1 Introduction

In the late 1950s, the discovery of feedback inhibition [1] meant that inhibitors could differ from substrates in terms of shape, size, and charge, questioning the binding to a common site in the protein. A few years later, Monod and collaborators described how a binding site distinct from the active site could regulate protein function and, together with Jacob, coined the term "allosteric" in the conclusions article of a Cold Spring Harbor Symposium in 1961 [2]. In its initial definition the term focused on the effector molecule—a molecule other than a steric analogue of the substrate. However, the focus quickly moved to the protein [3], as a structure-related property involving a reversible conformational alteration triggered by the binding of an allosteric effector to a site different of the active site. A crucial observation for rationalizing the allosteric effect was that allosteric regulation could be impaired through mutation while keeping the catalytic function intact. It was also observed that regulation mechanisms were more sensitive to mutations than the catalytic function, meaning that allosteric transitions involved not only the regulatory site but also relied on residues spread over the protein structure.

Models of allosterism initially focused on homomeric proteins. The MWC (Monod, Wyman, and Changeux) or concerted model [4] assumed the existence of at least two conformational states in equilibrium, for both the ligand-bound and ligand-unbound states. All monomers would be in the same state and the ligand would bind the different conformations with different affinities. Thus, upon binding there would be a shift in the equilibrium toward the most stable conformation for the bound state. It is a precursor of today's conformational selection model. On the other hand, the KNF (Koshland, Nemethy, and Filmer) or sequential model [5] represents an induced-fit model. It assumes two conformational states associated to the unbound and bound states, respectively. The binding of the effector ligand would cause a conformational change. The monomers would go through that conformational change independently, but the switch of one subunit would make the other subunits more likely to change.

The evolution toward our current understanding of allosterism was mainly triggered by emerging ideas on protein dynamics [6] and the realization that quaternary structure was not fundamental to allostery [7]. The current view of allosterism retains, nevertheless, the basic concepts of the two classical models [8–10], which can be recognized in different degrees in different allosteric systems and could be probably summarized in the following keywords: ensemble of states, conformational selection, population shift, induced fit, perturbation, and signal propagation. The largest digression from the classical concept was probably the proposal that allosteric signals may propagate by altering protein dynamics alone, without a detectable conformational change [11]. The baseline is that allosteric mechanisms are diverse and complex, as it is expected from a phenomenon affecting a large diversity of proteins and functions.

The presence of allosteric sites in a protein may not always be evident. Some allosteric sites are characterized by a transient nature and escape structural

studies—cryptic allosteric sites [12, 13]. In other cases, regions in a protein that could accommodate a ligand but lack natural ligands may become allosteric given the identification of an appropriate compound—serendipitous allosteric sites [14]—a property that has important implications in drug discovery. Indeed, it has been argued that in vivo there may be multiple effector sites and multiple signaling pathways within a single allosteric protein [15]. These findings and the notion that protein flexibility is a key requirement for allostery support the concept of allostery as a universal phenomenon [16]. An important difficulty when trying to identify allosteric sites and pathways in proteins, as already identified by Monod [3], is that the coding of allosteric properties in the protein structure appears to be subtle and fragile when compared to a catalytic activity. The picture is further complicated by the finding that intrinsically disordered proteins are no exception to allosterism [17].

Generally speaking, prediction algorithms may be based on (1) well-funded physical principles describing the system and property of interest (physical models); (2) a large number of previous recordings of the property of interest and possibly of other variables of the system that may correlate with it (statistical models); (3) exploratory knowledge, more often intuition, of other features of the system that may correlate with or determine the property of interest (heuristic models); or (4) a self-learning combination of the latter two (machine learning models). Often, prediction algorithms combine aspects of these different modeling strategies or derive a consensus from different models.

In the case at hand, the computational identification of allosteric sites and pathways in proteins, the development of predictors is hampered by both our still limited understanding of the physical determinants of allostery and the limited number and biased nature of well-characterized cases—1930 sites and 56 pathways described in the AlloSteric Database [18] at the time of this writing—that makes the use of statistical or machine learning models nonoptimal. The many facets of allostery make it also difficult to apply a unified prediction approach. For example, allosteric mechanisms may operate in a single protein domain, across domains in a protein, or across monomers in an oligomer. In addition, allosteric sites may map to protein cavities, surfaces, or interfaces. On the other hand, while we think that we understand some of the physical principles that govern allostery, we are currently unable to unambiguously identify allosteric sites based only on those and usually require the incorporation of heuristic knowledge. This comes in the form of candidate reporter properties—computable properties that presumably correlate with the presence of an allosteric site at a specific position in the protein's three-dimensional structure.

The distance between what we know about allostery and what we can actually predict is illustrated by a recent community experiment on the prediction of the effect of mutations on the allosteric regulation of human liver pyruvate kinase [19]. Although the number of participant groups was small and the result may thus be considered not significant—one could also argue that those who participated may represent a significant subpopulation of the still small set who have something to say—the experiment showed that the best predictions were marginally better than random. Clearly, there is still a lot to do before we are able to routinely identify allosteric sites and associated key signaling residues in proteins for drug discovery purposes.

7.2 Identification of Allosteric Sites and Pathways: A Methodological Framework

The computational identification of allosteric sites and pathways has been object of recent, excellent revisions [20–24]. For this reason, rather than trying to give a thorough overview that can be already found in these articles, here I present a personal selection of methods that can be useful to build a conceptual framework to study and predict allosteric sites and their communication with the functional site. The methods are described in some detail to give the nonexpert a basis to estimate their possibilities and understand the underlying assumptions. This methodological framework consists of two elements: the analysis of the dynamics of the system and the evaluation of correlations of different types between close and distant sites in a protein, eventually mapping the allosteric site and/or its communication with the functional site. Both interaction energy and entropy, as well as coevolution, are considered in this analysis.

7.2.1 Representation of Protein Structure and Dynamics

Protein structure and dynamics can be represented at different levels of resolution and using different approximations. The two methods that are most used to study allostery are molecular dynamics (MD) simulation and normal mode analysis (NMA) of elastic network models (ENM). Both can use an atomic or a coarse-grained representation of the protein structure, but it is more common to see the atomic level in MD simulation, including the protein environment, and one particle per residue in the ENM. Both make use of Newton's equation of motion to describe the dynamics of the system, solved numerically in MD simulation and, analytically, in the context of the harmonic approximation, in NMA-ENM. In MD simulation the information on the dynamics of the systems is extracted from the exploration of the potential energy surface, while in NMA-ENM it is extracted from a single minimum. For a main analysis considered here, both methods deliver a covariance matrix of the positions of the atoms (or other particles) in the system.

7.2.1.1 Molecular Dynamics Simulation

The method most often used to model proteins and their environment at high (atomic) resolution is molecular dynamics (MD) simulation [25]. MD simulation provides all the elements required to fully characterize a molecular system, thermodynamically and dynamically, provided the problem under study can be adequately approached from a classical mechanics basis. In the field of allostery, MD simulation is mainly used to investigate allosteric mechanisms of specific proteins.

Nevertheless, as we shall see, it can be also used for the identification of allosteric sites, albeit the technique is currently not suited for either fast or automatic identification studies.

MD simulations require the definition of (1) the molecular model, including the type of degrees of freedom and how these interact, (2) the equations of motion that will be used to propagate the system coordinates over time, and (3) the boundary conditions used to solve these equations, including both spatial and thermodynamic boundaries defining the theoretical macroscopic state of the system.

Model and Force Field When using a molecular model at the atomic level of resolution, the potential energy function describing the interactions between the N atoms of the system as a function of the $3N$ Cartesian coordinates $\mathbf{r}$ is typically given by [26, 27]:

$$\begin{aligned} V(\mathbf{r}) = \; & \frac{1}{2}\sum_{n=1}^{N_b} k_n^b (b_n - b_n^0)^2 + \frac{1}{2}\sum_{n=1}^{N_\theta} k_n^\theta (\theta_n - \theta_n^0)^2 \\ & + \frac{1}{2}\sum_{n=1}^{N_\xi} k_n^\xi (\xi_n - \xi_n^0)^2 + \sum_{n=1}^{N_\varphi} k_n^\varphi [1 + \cos(m_n \varphi_n - \delta_n)] \\ & + \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{1}{4\pi\varepsilon_0 \varepsilon_r} \frac{q_i q_j}{r_{ij}} \right) + V^s(\mathbf{r}) \end{aligned} \tag{7.1}$$

where the first three harmonic terms describe the interactions through bonds, bond angles, and so-called improper dihedral angles—used to enforce ring planarity as well as tetrahedral carbon geometry if aliphatic hydrogen atoms are not explicitly represented (united-atom models). The cosine function describes the interactions through torsional dihedral angles and, together with the harmonic terms, conforms the so-called bonded interactions. The last explicit term describes the van der Waals and electrostatic interactions—known together as nonbonded interactions—by means of the Lennard-Jones and Coulomb potentials, respectively. Because the number of atoms commonly considered in MD simulations of proteins makes the direct calculation of the sum in this term forbiddingly expensive (see "Boundary Conditions"), different methods have been developed to approximate the contribution from interactions beyond a cutoff radius or to effectively sum over periodic images, particularly for the electrostatic part [28, 29]. The special term $V^s(\mathbf{r})$ refers to purpose-specific functions adding restrains or a bias or incorporating a mean-force contribution from missing degrees of freedom.

The force constants k_n and the parameters b_n^0, θ_n^0, ξ_n^0, m_n, δ_n, A_{ij}, B_{ij}, q_i, and q_j constitute, together with Eq. (7.1), the (empirical) force field. These parameters are derived by fitting simulation data, generated with the force field that is being optimized, to quantum mechanical calculations and experimental data—gas and liquid phases—on simple molecules containing the chemical groups present in proteins and other biomolecules of interest. Importantly, these force fields have not been parameterized against protein experimental data and are therefore free of any bias toward the native folded structure of a protein. The selection of the force

field is a matter of choice, since there is currently no force field that outperforms every other one for every system and every property of interest.

Equations of Motion The system dynamics are obtained, in general, by integrating Newton's equations of motion. Recalling that in a conservative field, the force is equal to minus the gradient of the potential—which has a straightforward analytical form given Eq. (7.1)—the force acting on atom i is given by:

$$\mathbf{F}_i = m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial V(\mathbf{r})}{\partial \mathbf{r}_i} \tag{7.2}$$

The resulting system of equations must be integrated numerically, for which there exist different algorithms [26]. In order to avoid numerical instabilities, the integration time step must be significantly smaller than the highest-frequency motions in the system, i.e., those between atoms interacting through a bond. The treatment of bonds as classical harmonic oscillators, as in Eq. (7.1), represents a gross approximation to the notorious quantum mechanical nature of the bond. For this reason, and because of its little impact on system dynamics, the bond term in the potential may be replaced by an equally gross approximation consisting in the fixing of the bond length by imposing a constraint [30]. This allows an increase of the time step to the commonly used value of 2 fs, as well as a reduction of the range of frequencies coexisting in the system, leading to a faster redistribution of energy between modes.

The evaluation of the forces in Eq. (7.2) is, by far, the most time-consuming task in a MD simulation. While this cost can be reduced by using lower-resolution models—known as coarse-grained models—that map groups of atoms to the dynamic particles [31], this approach has significant limitations when applied to proteins, requiring the superposition to the standard force field of secondary structure restraints [32] or an elastic network [33] in order to maintain a correct fold. It is often argued that this level of modeling is sufficient to capture the main features of the protein's global dynamics, which are thought to be less sensitive to the fine details of the structure—"butterfly effects" such as the destabilizing action of some point mutations, often transparent to such force fields, would speak against it. Clearly, these models can be very useful for the study of large biomolecular systems, particularly if diffusion and binding are important components of the study. However, if one is interested in intramolecular dynamics, such as in the study of allosteric communication, it is not clear if a coarse-grained MD model would provide higher-quality information than the computationally far less demanding elastic network models discussed below.

Boundary Conditions Although there exist mean-field approaches that allow a relatively inexpensive incorporation of solvent effects in the dynamics of the protein [34, 35], the molecular environment of the protein is in most cases represented explicitly. This is because there are important aspects of solvation, such as an entropic component due to correlations between protein and solvent degrees of freedom, that cannot be easily incorporated in such mean-field approaches. When the protein environment is explicitly represented, the molecular system needs to be

given a finite size by defining appropriate three-dimensional boundaries. This is most often done by surrounding a central system having a biologically relevant and computationally tractable size by an infinite array of periodic images, in so-called periodic boundary conditions [26].

In addition to specifying the number of particles by physically delimiting the system, one needs to specify the degree of separation between the system and its surrounding—whether it is isolated or there is exchange of heat, work, or matter—and the value of the thermodynamic state variables (temperature, pressure, volume, etc.) that are constant, in order to define the macrostate of the system. In MD simulation, quantities such as temperature and pressure may be held "constant" by use of different algorithms that couple the system to an external reservoir [36].

Interpretation of MD Simulation Data Under the ergodic hypothesis we can assume that, when the simulation time approaches infinity—in practice, the time required to sample the system's phase space with the right probability density distribution—the average of an observable over time and its average over the corresponding equilibrium statistical ensemble are the same. This simple relation grounds molecular dynamics simulations in statistical mechanics. Clearly, in real applications the simulation time not only falls short of this target but in most cases does not even reach the biologically relevant time scale. However, the problem of insufficient sampling not only questions the use of statistical mechanics to interpret molecular dynamics simulations, it also questions its use to interpret native ensemble methods facing equivalent computational limitations.

We may represent the trajectory ϕ of position and velocity coordinates of the atoms of the system at discrete time points in the simulation as:

$$\phi = (\mathbf{r}(t_0), \mathbf{r}(t_1), \ldots, \mathbf{r}(t_M), \mathbf{v}(t_0), \mathbf{v}(t_1), \ldots, \mathbf{v}(t_M)) \tag{7.3}$$

where $\mathbf{r}(t_k)$ and $\mathbf{v}(t_k)$ describe the positions and velocities of the N atoms at time t_k:

$$\begin{aligned} \mathbf{r}(t_k) &= (\mathbf{r}_1(t_k), \mathbf{r}_2(t_k), \ldots, \mathbf{r}_N(t_k)) \\ \mathbf{v}(t_k) &= (\mathbf{v}_1(t_k), \mathbf{v}_2(t_k), \ldots, \mathbf{v}_N(t_k)) \end{aligned} \tag{7.4}$$

With the tools of statistical mechanics at hand, the possibilities opened by this data set in terms of characterizing the system and its properties—at the given macrostate—are almost unlimited. Again, the only factor that will reduce our options to succeed in that characterization is the amount of sampling performed, which will be unavoidably on the short side. And, of course, given sufficient sampling we will be able to fully characterize nothing more than the computational system, which will be still a model.

7.2.1.2 Elastic Network Models

Elastic network models (ENM) are used in combination with normal mode analysis (NMA) as an approximation to the dynamics of a protein around a minimum in the

potential energy surface, typically that represented by the crystallographic structure. While MD simulation integrates numerically Newton's equations of motion, NMA solves them analytically by assuming that the potential (and kinetic) energy of the system can be approximated by a quadratic form. Thus, NMA assumes that the motion of the atoms is limited to small fluctuations around equilibrium positions corresponding to a minimum of a given potential energy function. This is because the shape of the potential in the vicinity of the minimum may be assumed to be quadratic even if the underlying potential energy function is not.

Standard NMA, like MD simulation, makes use of potential energy functions such as the one in Eq. (7.1) and requires energy minimization of the initial structure to guarantee that the calculations are performed at an energy minimum. To ensure that the first derivative of the potential is zero with respect to the positions of all atoms and thus calculate meaningful second derivatives, which are required to solve the equations of motion, a careful energy-minimization protocol is mandatory in standard NMA. Instead, as we shall see below, in ENM any structure of choice of the protein is, by construction of the potential energy function, a global energy minimum.

Although the potential energy surface of a protein—and its environment—is, as a matter of fact, complex and characterized by multiple minima, ENM have shown remarkable success at reproducing main features of low-frequency, large-amplitude collective motions observed in molecular dynamics simulations using all-atom force fields [37, 38] as well as in the calculation of crystallographic B-factors [39], revealing an unsuspected robustness of the global modes of motion relative to details in atomic coordinates or specific atomic interactions. Indeed, it has been postulated that the global modes are an intrinsic property of the three-dimensional shape of the protein [40].

Remarkably, ENM have been also shown to outperform standard NMA in many applications [37]. This is not completely surprising, since the power of empirical atomistic force fields lies in their combination with techniques such as molecular dynamics or Monte Carlo simulation to sample the potential energy surface. This type of force fields has not been, for example, optimized to accurately reproduce the folded structure of proteins, but is based on thermodynamic, dynamic, and electrostatic properties of simpler systems. Thus, if one is to base a calculation, under the harmonic approximation, on one single structure of the protein, it is probably better to take as energy minimum the crystallographic structure, particularly if it has been solved at high resolution, than a relatively close structure that happens to be a minimum in an empirical force field. Focusing on a single minimum of an empirical interaction function will only make more evident the defects of this function.

As an advantage over more accurate methods such as MD simulation, the low computational cost of ENM means that they can be used in studies involving relatively large sets of proteins, for example, in the context discussed here, for the prediction of allosteric sites or the study of allosteric mechanisms.

We distinguish two types of ENM models, the anisotropic network model (ANM) [41], which is a three-dimensional model based on seminal work by Tirion [42], and the Gaussian network model (GNM) [43, 44], a one-dimensional model. I will describe in some detail the ANM and then mention main differences of the GNM.

Anisotropic Network Model It models the protein as a network of N atomic masses, in principle representing only the C_α atoms. Two atoms i and j of the network are linked by a Hookean spring if their distance $r_{ij}^0 = \|\mathbf{r}_j^0 - \mathbf{r}_i^0\|$ in the studied structure $\mathbf{r}^0$ is smaller than a cutoff R_c. The potential energy of the system is then described by the function:

$$V = \frac{1}{2}k \sum_{\substack{i,j \\ r_{ij}^0 \leq R_c}} (r_{ij} - r_{ij}^0)^2 \tag{7.5}$$

where the force constant k is in principle uniform for all pairs and has, in such case, the sole function of specifying the units—note that k does not modify the shape of the normal modes, but only scales uniformly their frequencies, which must be anyway interpreted in relative, rather than absolute, terms. R_c becomes then the only parameter in the model. Contrary to standard NMA, this model does not require energy minimization, since by construction the studied structure corresponds to the global minimum of Eq. (7.5).

The parameter R_c should be chosen sufficiently small that the network holds information on topology and shape, rather than only on size, and sufficiently large that no artificial subnetworks, i.e., subnetworks not corresponding to actual protein domains, are generated. In practice, if we use a too small cutoff distance, the eigen-decomposition of the Hessian (see below) will lead to additional zero eigenvalues [38]. An optimum R_c of 18 Å has been proposed for ANM using C_α atoms as nodes [45], although values as low as 10 Å are also commonly used [46].

By recalling Newton's and Hooke's laws, we can write the equations of motion for the system as:

$$\mathbf{F} = \mathbf{M}\frac{d^2}{dt^2}\Delta\mathbf{r} = -\mathbf{H}\Delta\mathbf{r} \tag{7.6}$$

where $\mathbf{F} = (\mathbf{F}_1, \ldots, \mathbf{F}_N)$ are the restoring forces on the N atoms, $\mathbf{M}$ is the $3N \times 3N$ mass diagonal matrix, $\mathbf{H}$ is the force constant or Hessian matrix, and $\Delta\mathbf{r} = (\Delta\mathbf{r}_1, \ldots, \Delta\mathbf{r}_N)$, with $\Delta\mathbf{r}_i = (\mathbf{r}_i - \mathbf{r}_i^0)$, are the atomic displacements from equilibrium.

It can be shown that the analytical solution to this equation for atom i is [38]:

$$\mathbf{r}_i(t) = \mathbf{r}_i^0 + \sum\nolimits_{n=1}^{3N} (\mathbf{u}_n)_i \left(\frac{2E_n}{\omega_n^2 m_i}\right)^{1/2} \cos(\omega_n t + \phi_n) \tag{7.7}$$

where E_n, ω_n, and $(\mathbf{u}_n)_i$ are the total energy, angular frequency ($= 2\pi\nu_n$), and ith coordinates of the so-called normal mode of vibration n, and m_i is the mass of atom i. The amplitude of the mode—the square root factor in Eq. (7.7)—depends, together with the phase ϕ_n, on initial positions and velocities. Note that all modes have on average the same vibrational energy, i.e., $k_B T$.

Geometrically, the normal modes are the particular directions **u** in which the force is aligned with the displacement. Therefore, when applied to **u**, the linear transformation **H** must result in a simple scaling of the vector, so that $\mathbf{Hu} = \lambda\mathbf{u}$, where λ and **u** are eigenvalues and corresponding eigenvectors of **H**.

The frequencies ω_n and directions $\mathbf{u}_n$ in Eq. (7.7) can then be obtained by using the slightly modified eigen equation:

$$\mathbf{Hu} = \lambda\mathbf{Mu} \tag{7.8}$$

where it can be shown that $\lambda = \omega^2$. Conventionally, this equation is solved in the simpler form $\tilde{\mathbf{H}}\tilde{\mathbf{u}} = \lambda\tilde{\mathbf{u}}$, where the tilde indicates that the coordinates and force constants have been mass weighted, i.e., $\tilde{u}_{x_i} = u_{x_i}\sqrt{m_i}$ and $\tilde{H}_{x_i y_j} = H_{x_i y_j}/\sqrt{m_i}\sqrt{m_j}$, with the subindex x_i indicating the x component of atom i. However, when all masses are equal, as in the case described here, Eq. (7.8) reduces to $\mathbf{Hu} = \lambda m\mathbf{u} = \tilde{\lambda}\mathbf{u}$, which can be readily solved.

The Hessian **H**, evaluated at $\mathbf{r}^0$, where $r_{ij} = r_{ij}^0$, can be represented as the block matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{N1} & \cdots & \mathbf{H}_{NN} \end{bmatrix} \tag{7.9}$$

with off-diagonal and diagonal blocks given by:

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V/\partial x_i \partial x_j & \partial^2 V/\partial x_i \partial y_j & \partial^2 V/\partial x_j \partial z_j \\ \partial^2 V/\partial y_i \partial x_j & \partial^2 V/\partial y_i \partial y_j & \partial^2 V/\partial y_i \partial z_j \\ \partial^2 V/\partial z_i \partial x_j & \partial^2 V/\partial z_i \partial y_j & \partial^2 V/\partial z_i \partial z_j \end{bmatrix} \tag{7.10}$$

$$\mathbf{H}_{ii} = \begin{bmatrix} \partial^2 V/\partial {x_i}^2 & \partial^2 V/\partial x_i \partial y_j & \partial^2 V/\partial x_i \partial z_i \\ \partial^2 V/\partial y_i \partial x_i & \partial^2 V/\partial {y_i}^2 & \partial^2 V/\partial y_i \partial z_i \\ \partial^2 V/\partial z_i \partial x_i & \partial^2 V/\partial z_i \partial y_i & \partial^2 V/\partial {z_i}^2 \end{bmatrix} \tag{7.11}$$

where the second derivatives are given by:

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{\partial^2 V}{\partial y_j \partial x_i} = \frac{\partial^2 V}{\partial y_i \partial x_j} = -k\frac{1}{r_{\mathrm{ij}}^{0^2}}(x_j - x_i)(y_j - y_i) \tag{7.12}$$

$$\mathbf{H}_{ii} = -\sum_{\substack{j \neq i \\ r_{ij}^0 \leq R_c}} \mathbf{H}_{ij} \tag{7.13}$$

The matrix $\mathbf{H}$ is real symmetric and can be decomposed in $3N-6$ nonzero eigenvalues (informing on the frequencies of vibration) and corresponding orthogonal eigenvectors (informing on the directions and relative amplitudes). The remaining six eigenvalues are equal to zero and correspond to the protein's rigid-body motion. As already mentioned, the vibrational energy is, on average, equal for all modes, which means that the average amplitude of oscillation along a given mode is inversely proportional to its frequency. Thus, the collective motions of largest amplitude, which tend to be most interesting in terms of protein function, are those along the lowest-frequency—lowest eigenvalue—modes.

Although the zero eigenvalues make $\mathbf{H}$ non-invertible, by discarding these six eigenvalues, we can define a pseudo-Hessian $\breve{\mathbf{H}}$ using the similarity transformation $\breve{\mathbf{H}} = \mathbf{PDP}^{-1}$, where $\mathbf{P}$ is the change-of-basis matrix with the $3N-6$ orthonormal eigenvectors as columns and $\mathbf{D}$ is the diagonal matrix with the $3N-6$ eigenvalues, such that we may obtain the inverse as $\breve{\mathbf{H}}^{-1} = \mathbf{PD}^{-1}\mathbf{P}^T$. The inverse pseudo-Hessian is also organized in $N \times N$ submatrices of size 3×3, with element $x_i y_j$ of submatrix $\left(\breve{\mathbf{H}}^{-1}\right)_{ij}$ given by:

$$\left(\breve{\mathbf{H}}^{-1}\right)_{x_i y_j} = \sum_{n=1}^{3N-6} \frac{1}{\tilde{\lambda}_n} (u_n)_{x_i} (u_n)_{y_j} \tag{7.14}$$

where $\tilde{\lambda}_n = \omega_n^2 m$ is the nth eigenvalue and $u_{n,xi}$ is the component x of atom i in the nth eigenvector.

The trace of submatrix $\left(\breve{\mathbf{H}}^{-1}\right)_{ij}$ defines the covariance between the positions of atoms i and j (or the mean square displacement if $i = j$; see Sect. 7.2.2.1 for a discussion on these quantities):

$$\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j \rangle = k_B T \operatorname{tr}\left(\left(\breve{\mathbf{H}}^{-1}\right)_{ij}\right) = \frac{k_B T}{m} \sum_{n=1}^{3N-6} \frac{(\mathbf{u}_n)_i (\mathbf{u}_n^T)_j}{{\omega_n}^2} \tag{7.15}$$

$$\left\langle {\Delta\mathbf{r}_i}^2 \right\rangle = k_B T \operatorname{tr}\left(\left(\breve{\mathbf{H}}^{-1}\right)_{ii}\right) = \frac{k_B T}{m} \sum_{n=1}^{3N-6} \frac{(\mathbf{u}_n)_i \left(\mathbf{u}_n^T\right)_i}{{\omega_n}^2} \tag{7.16}$$

In practice, since these quantities scale as the inverse of the frequency squared, a sum over the lowest-frequency normal modes of the system in Eqs. (7.15) and (7.16) is usually sufficient to obtain a reasonable approximation.

ANM exists with different flavors, including the inverse weighting of the force constant with the distance between atoms [45] or the original derivation including distinct atom types (masses) [42].

Gaussian Network Model The GNM assumes that the C_α atoms and corresponding interatomic distances undergo Gaussian-distributed, isotropic fluctuations around their equilibrium coordinates under the potential of their near neighbors in the network. The potential energy of the system takes then the form:

$$V = \frac{1}{2}k \sum_{\substack{i,j \\ r_{ij}^0 \le R_c}} (\Delta\mathbf{r}_j - \Delta\mathbf{r}_i)^2 = \frac{1}{2}k \sum_{\substack{i,j \\ r_{ij}^0 \le R_c}} (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0)^2 \tag{7.17}$$

where an optimum R_c value of 7.3 Å has been proposed [47]. Note that, contrary to Eq. (7.5) (ANM), the interaction energy of an atom pair $i\,j$ in Eq. (7.17) is only zero if both the magnitude ($i\,j$ distance) and direction of the vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{ij}^0$ are equal. This is assumed to be a main reason for the comparatively better agreement of the mean square displacements from GNM with experimental data, relative to ANM calculations [48].

In this model, the fluctuations of the atomic positions and their cross-correlations are determined by the adjacency-degree matrix, which describes the neighboring relations between all atoms:

$$\Gamma_{ij} = \begin{cases} -1 & \text{if} \quad j \neq i \quad \text{and} \quad r_{ij} \le R_c \\ 0 & \text{if} \quad j \neq i \quad \text{and} \quad r_{ij} > R_c \\ -\sum_{l \neq i} \Gamma_{ij} & \text{if} \quad j = i \end{cases} \tag{7.18}$$

The matrix Γ can be therefore viewed as the $N \times N$ counterpart of the Hessian. Note that Γ provides no information on directionality, which means that the GNM can be used to obtain information on the relative amplitudes of motion but not on their three-dimensional directions.

Within this framework, the covariances and mean square displacements can be calculated after decomposing the adjacency-degree matrix into its eigenvalues and eigenvectors to then invert it—discarding the smallest, zero-valued eigenvalue, which corresponds to translation—much as it is done with the Hessian in the ANM. Covariances and mean square displacement can be then computed as:

$$\langle \Delta\mathbf{r}_i \Delta\mathbf{r}_j \rangle = \frac{3k_B T}{k} \left(\breve{\Gamma}^{-1}\right)_{ij} \tag{7.19}$$

$$\langle \Delta \mathbf{r}_i^2 \rangle = \frac{3k_B T}{k} \left(\breve{\Gamma}^{-1} \right)_{ii} \tag{7.20}$$

Main allosteric-site prediction servers and freely available codes, such as AlloFinder [49], DynOmics [50], STRESS [51], AlloPred [52], PARS [53], and SPACER [54], have as main component a normal mode analysis on an elastic network model, typically within a perturbation approach (see Sect. 7.2.2.4).

7.2.2 Analysis of Coupled Sites in Proteins

For an effector molecule to modulate the function of a protein upon binding to an allosteric site, there must be a causal connection or coupling between the latter and the active site or region of the protein involved in function. It has been proposed that this coupling may not necessarily translate into a conformational change—be it induced or through the stabilization of a transient conformer—upon binding of the allosteric effector, as assumed in the early MWC (conformational-selection-like) [4] and KNF (induced-fit-like) [5] models, but could be due to changes in dynamics with no translation to conformation [11]. Examples supporting this latter possibility include experimental studies [55], and, as we will see in Sect. 7.2.2.4, it has given rise to models exploring allosteric communication from a purely entropic perspective, by assuming the absence of a conformational change as integral part of the model [56]. Nevertheless, the question of whether there can be an allosterically induced change of the functional state of a protein without a corresponding conformational change is, far from settled, a topic of active discussion [57]. In either way, the coupling between the allosteric site and the functional site must have an energy basis (interactions), an entropy basis (fluctuations and their correlations), or both.

The fact that allostery requires the communication between generally distant sites has given rise to the concept of allosteric pathway. This concept is in line with the observation that local perturbations do not propagate isotropically in proteins. However, it can be misleading because the causal connection between the allosteric and functional sites may not necessarily involve a contiguous string, or even an ensemble of strings, of coupled residues. Furthermore, the risk of finding an allosteric signaling pathway when actively searching for it is high. This is illustrated by a recent study that has estimated the statistical significance of contiguous pathways for the null model of selecting residues at random [58]. The authors found that when 20% of the residues in three proteins were randomly selected, contiguous pathways at the 6 Å cutoff level were found with success rates ranging from 51% to 3% depending on the protein. The pathways were found to be most significant (lower percentage of false positives) in the system exhibiting the greatest degree of complexity.

Under the conceptual framework presented here, the problem of identifying the allosteric site of a protein might be transformed into that of identifying significant energy and/or entropy couplings, which might map to evolutionary couplings, between the functionally relevant site and other regions in the protein. Since such couplings have been proposed to be ubiquitous in proteins [16, 59], we may not necessarily expect to come out with a single candidate site from such exercise. Indeed, the potential for multiple allosteric regulation—by the binding of effectors to different sites in a protein—might be a feature common to all proteins [15], whether these potential allosteric sites are orphan, having a yet unknown natural effector, or serendipitous, having no natural ligand [60].

In Sects. 7.2.2.1, 7.2.2.2, 7.2.2.3, and 7.2.2.4, I describe a personal selection of concepts and methods that are being used to evaluate the coupling between atoms and residues at close or far distances in a protein and that can, therefore, be used to identify communication pathways and, from them, potential allosteric sites given the knowledge of the functional site. They cover the aspects mentioned before, entropy and energy, linked by dynamics, and coevolution. Along their description, I give examples of recent studies making use of these concepts and methods or of variants of them.

7.2.2.1 Methods Based on the Covariances of Atomic Positions

One way to evaluate the coupling between pairs of residues at any distance in a protein is by analyzing the degree to which they move in a concerted way. This coupling or correlation can be measured for two residues at a given time point t or, in order to investigate a potential causality, one residue at a time t and another at a later time $t + \tau$. These analyses are described and discussed in the following paragraphs.

Given two jointly distributed random variables X and Y, their covariance (a measure of their joint variability) is defined as:

$$\mathrm{cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle \tag{7.21}$$

where the angle brackets denote an expected value (mean). The covariance is positive if the two variables increase and decrease together, negative if they have an opposite behavior, and zero if they are independent.

If $\mathbf{X}$ is a random vector, we define its covariance matrix as the matrix whose elements are given by the covariances $\mathrm{cov}(X_k, X_l)$ of the scalar components k and l of the vector. The covariance matrix is symmetric, with the diagonal elements being the variances.

For example, consider a trajectory of the Cartesian coordinates of the N atoms of a protein from a molecular dynamics (MD) simulation. After translational and

rotational fitting of the protein configurations sampled in a given time interval, to remove translational and—imperfectly—rotational motion, we could calculate the covariance cov(x_k, x_l) (c_{kl} in simplified notation) between coordinates x_k and x_l as:

$$c_{kl} = \langle (x_k - \langle x_k \rangle)(x_l - \langle x_l \rangle) \rangle = \langle \Delta x_k \Delta x_l \rangle \tag{7.22}$$

where k and l run over the $3N$ coordinates of the protein and the angle brackets are now time averages. The resulting matrix can then be used for the analysis of principal components (PCA) [61]—a projection onto the orthogonal directions of largest variance of the $3N$-dimensional space where the vectors representing the coordinates of all atoms of the protein as a function of time (the original dataset) are defined—or the calculation of configurational entropies [62, 63]. Although in principle both types of analysis could be further used to identify potentially coupled sites in proteins [64, 65], we will see that, in the study of allostery, entropy is generally approached from the information-theory side. Note that, formally, the top-ranking modes obtained by PCA should be comparable to the lowest-frequency modes derived by normal mode analysis (NMA), provided that the ensemble of conformers from MD simulation and the Hessian used in the NMA are representative for the same equilibrium.

The covariances given by Eq. (7.22) may be difficult to interpret if we are interested in the covariance between the positions $\mathbf{r}_i$ and $\mathbf{r}_j$ of atoms i and j, which can be also calculated as:

$$c_{ij} = \langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)(\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle = \langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \rangle \tag{7.23}$$

where i and j run now over the N atoms of the protein. Note, however, that the dot product means that this equation measures the positional covariance by projecting the displacement of one atom onto the direction of displacement of the other. Thus, two atoms moving in perfect synchrony but in perpendicular directions will have a covariance of zero. This can be further understood by noting that the linear transformation represented by this covariance matrix operates on a vector space of dimension N, which can be thought of as a projection of the three-dimensional space of atomic coordinates to one dimension.

The diagonal elements of this covariance matrix are the mean square displacements $\langle \Delta \mathbf{r}_i^2 \rangle$ of the atoms, which can be related to the isotropic temperature factors B_i (B-factors) [66]:

$$B_i = \frac{8\pi^2}{3} \langle \Delta \mathbf{r}_i^2 \rangle \tag{7.24}$$

Note that the mean square displacements are also available from NMA.

The magnitude of the covariance depends on the magnitudes of the atomic displacements from the mean, which makes it difficult to interpret in terms of the extent to which the atoms move in a concerted way (amount of correlation). The correlation or normalized covariance matrix is defined by elements C_{ij} such that:

$$C_{ij} = \frac{c_{ij}}{c_{ii}^{1/2} c_{jj}^{1/2}} = \frac{\langle \Delta \mathbf{r}_i \Delta \mathbf{r}_j \rangle}{\langle \Delta \mathbf{r}_i^2 \rangle^{1/2} \langle \Delta \mathbf{r}_j^2 \rangle^{1/2}} \tag{7.25}$$

The correlation coefficients C_{ij} have values from -1 (perfectly anticorrelated motions) to 1 (perfectly correlated motions), with zero corresponding to fully uncorrelated motions. The diagonal elements are equal to 1.

As the covariance matrix, the correlation matrix is symmetric and gives therefore no information on a potential directionality of the motional coupling of two atoms. One may also calculate the time covariance $c_{ij}(\tau)$ and correlation $C_{ij}(\tau)$ as a function of a lag time τ:

$$\begin{aligned} C_{ij}(\tau) &= \frac{c_{ij}(\tau)}{c_{ii}^{1/2} c_{jj}^{1/2}} = \frac{\langle (\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle)(\mathbf{r}_j(t+\tau) -, \langle \mathbf{r}_j \rangle) \rangle}{\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle^{1/2} \langle (\mathbf{r}_j - \langle \mathbf{r}_j \rangle)^2 \rangle^{1/2}} \\ &= \frac{\langle \Delta \mathbf{r}_i(t) \Delta \mathbf{r}_j(t+\tau) \rangle}{\langle \Delta \mathbf{r}_i^2 \rangle^{1/2} \langle \Delta \mathbf{r}_j^2 \rangle^{1/2}} \end{aligned} \tag{7.26}$$

This introduces directionality in the correlation, often interpreted as causality, and in general $C_{ij}(\tau) \neq C_{ji}(\tau)$.

Correlation and principal component analysis are, for example, used by the package MONETA [64] to localize the propagation of structural and dynamical effects of a perturbation throughout a protein structure, for the identification of allosteric pathways. Note that a correlation analysis can be performed for different quantities, for example, between the positional fluctuation of one residue and the stabilization energy of another (see Sect. 7.2.2.3), for all residue pairs in a protein [67].

7.2.2.2 Methods Based on Probabilities of States and Information Theory

A second way to investigate the coupling between residues in a protein is by the analysis of their joint probabilities. The framework for this analysis is information theory, from which some basic concepts will be introduced here. A key concept in this context is the mutual information between two variables, formally an entropy. Its overtime counterpart is the transfer entropy, which corresponds to a conditional mutual information.

In addition, the relative probabilities of different states of a system can be used to calculate their free energy differences. This has been exploited, as we will also see, to

study the coevolution of residues in a protein, indicative of functional links between them. Evolutionary coupling is a perfect complement to energy and dynamic coupling for the identification of allosteric pathways and sites. These concepts are described and discussed in the following paragraphs.

Given the probability distribution of a discrete random variable X, the Shannon or information entropy $H(X)$ measures the average minimum amount of information required to convey the identity of a random sample from that distribution [68]:

$$H(X) = -\sum_{x} p(x)\log p(x) \tag{7.27}$$

where x are the values or states that may be adopted by X, $p(x)$ is the marginal probability that X adopts state x, and the entropy units depend on the base of the logarithm—*bits* or *shannons* for base 2, *nats* for base e, and *hartleys* for base 10. Thus, one *nat* is the information content of an event of probability e^{-1}. Note that the information content of a state ($-\log p(x)$) is higher the lower its probability.

Importantly, there exists a relationship between information entropy and thermodynamic entropy [69], which becomes apparent from Gibbs' statistical mechanical definition of entropy, in its discrete form [70]:

$$S(X) = -k_B \sum_{x} p(x)\ln p(x) \tag{7.28}$$

where k_B is Boltzmann's constant. In this case, x are the microscopic states that may be adopted by the thermodynamic system X, as drawn from an equilibrium ensemble characterized by a probability distribution that is determined by the macroscopic (thermodynamic) state of the system (note that X in Eq. (7.28) implicitly incorporates the dependence on the thermodynamic state variables). In information-theoretic terms, the reduced Gibbs entropy (S/k_B) may thus be interpreted as the average minimum amount of information in *nats* required to fully specify a microstate, given that we know the system's macrostate [71].

Using the Shannon equation, one may also define the joint entropy $H(X,Y)$ of two random variables X and Y as:

$$H(X,Y) = -\sum_{x,y} p(x,y)\log p(x,y) \tag{7.29}$$

where $p(x, y)$ is the joint probability that X adopts state x and Y adopts state y.

One can also define the entropy $H(X|Y)$ of X conditioned to Y, or amount of information required to describe the state of X given that the state of Y is known, or amount of randomness in X given Y:

$$H(X|Y) = -\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(y)} \tag{7.30}$$

Finally, the mutual information $I(X, Y)$ measures the amount of information that one variable contains about the other, i.e., their mutual dependence:

$$I(X,Y) = \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{7.31}$$

where $I(X,Y)$ is nonnegative, $p(x, y) \geq p(x)p(y)$, the equality being valid for independent variables, and symmetric, $I(X,Y) = I(Y, X)$. Unlike the correlation given by Eq. (7.23), which reflects the degree of linear relationship between the two variables, the mutual information is also sensitive to nonlinear relationships.

The previous entropy measures can be then related to the mutual information through Eq. (7.32):

$$H(X,Y) = H(X) + H(Y) - I(X,Y) = H(X|Y) + H(Y|X) + I(X,Y) \tag{7.32}$$

which tells us that the information shared by the two variables (their dependence) reduces their joint entropy or amount of information required to encode a joint event.

To study the dynamics of shared information, we need to introduce directionality in Eq. (7.31) [72]. Let us first define the conditional mutual information for random variables X, Y given Z:

$$I(X,Y|Z) = H(X|Z) + H(Y|Z) - H(X,Y|Z) \tag{7.33}$$

Schreiber [73] defined transfer entropy $T_{X\to Y}$ as a measure of the causal information transfer between two systems that evolve in time, based on finite-order stationary Markov processes. Consider the simplest case of a random variable X approximated by a stationary Markov process of order 1—one where the probability to find X in state x at time $t + \tau$, τ being the coarse-graining lag time, depends only on the state of X at time t, and the probability of the transition is time invariant. Consider a second such variable Y. Then, the transfer entropy quantifies the deviation from the generalized Markov property $p(y_{t+\tau} \mid y_t) = p(y_{t+\tau} \mid y_t, x_t)$ that assumes that the two variables are independent and can be written as:

$$T_{X\to Y}(\tau) = \sum_{x_t}\sum_{y_t}\sum_{y_{t+\tau}} p(y_{t+\tau}, y_t, x_t)\log\frac{p(y_{t+\tau}|y_t, x_t)}{p(y_{t+\tau}|y_t)} \tag{7.34}$$

$$T_{X\to Y}(\tau) = \sum_{x_0}\sum_{y_0}\sum_{y_\tau} p(y_\tau, y_0, x_0)\log\frac{p(y_\tau|y_0, x_0)}{p(y_\tau|y_0)} \tag{7.35}$$

where, due to the condition of stationarity, the probabilities are independent of the value of t. For practical applications, one should first evaluate the dependence of $T_{X\to Y}$ on the value of τ or choose a value of τ consistent with the processes of interest—e.g., with the decay time of their time correlation function.

After simple manipulations and observing Eq. (7.33), it can be shown that:

$$T_{X\to Y}(\tau) = H(X_t|Y_t) + H(Y_{t+\tau}|Y_t) - H(Y_{t+\tau}, X_t|Y_t) = I(Y_{t+\tau}, X_t|Y_t) \quad (7.36)$$

Thus, the transfer entropy from a process X to a process Y is a measure of the amount of information that the current state of X provides on the future state of Y given the current state of Y, that is, the conditional mutual information.

These concepts have been recently used by Hacisuleyman and Erman [74] to evaluate the transfer entropy between the fluctuations of pairs of C_α atoms in ubiquitin (replacing X,Y by $\Delta\mathbf{r}_i$, $\Delta\mathbf{r}_j$ in the previous expressions) to determine net entropy acceptors and donors and entropy paths with potential roles in allostery. To this end, they histogrammed the values of $\Delta\mathbf{r}_i$ from a molecular dynamics simulation of the protein by dividing the distribution in an estimated optimum number of eight bins. This reduced number of states and a relatively long simulation of 600 ns enabled the evaluation of the single, double, and triple probabilities required for the calculation of the various entropy terms described above (with $\tau = 5$ ns). In addition, they complemented the transfer entropy calculations with results from time correlation analysis (see Sect. 7.2.2.1), which were also found to reveal causality relations.

The mutual information concept has been also used to derive a generalized correlation coefficient [75] that has been later used to study protein-cavity couplings from MD simulation data [59] or to identify the residues involved in allosteric signaling using MD simulation data and a graph-theoretical approach [76]. Mutual information has been also used to identify evolutionarily coupled residue pairs involved in the communication between ligand-binding pockets [77, 78].

In the latter context of coevolution of amino acid residues in a protein, Lockless and Ranganathan [79] introduced a successful metric to evaluate statistical energy couplings between positions in a multiple sequence alignment (MSA). The so-called statistical coupling analysis (SCA) views the evolution of a protein as the result of a random mutagenesis process constrained by the condition of function preservation. It assumes that, if there are no evolutionary constraints at one position in a MSA, the distribution of observed amino acids at that position should approach their mean abundance in proteins, and the degree of deviation from the mean value should be proportional to the degree of conservation of the position. It also assumes that the functional coupling of two positions, wherever they are in the structure, should mutually constrain evolution at those positions, and this should be reflected in the statistical coupling of the underlying amino acid distributions. The statistical coupling of two sites is defined as the degree to which amino acid frequencies at one site change—not necessarily changing the conservation of the site—in response to a perturbation of the frequencies observed at the other site.

The probability $p_i(x)$ of an amino acid x at site i in the MSA is given by the binomial probability of getting n_x instances of x at position i in the N sequences of the MSA, given a background probability of x in proteins of $p_p(x)$:

$$p_i(x) = \frac{N!}{n_x!(N-n_x)!} p_p^{n_x} \left(1-p_p\right)^{N-n_x} \tag{7.37}$$

One may treat the sequence as if described by a statistical ensemble, characterized by a given probability distribution over the microstates (exact amino acid sequences). Then, for a MSA representing the ensemble—with the right distribution of amino acids at the sites—the free energy change associated to the transition from a state with x at position i to a state with x at position j is given by:

$$\Delta G_{i\to j}^{x} = -k \ln \frac{p_j(x)}{p_i(x)} \tag{7.38}$$

where k is here an arbitrary constant with energy units (k_BT in the statistical-mechanical context). One may then take j as a hypothetical reference site for all amino acids, in which all amino acids are found at their mean frequencies in the MSA ($p_{\mathrm{msa}}(x)$), and define a vector $\mathbf{P}_i = (\Delta G_{i\to j}^{A}, \ldots, \Delta G_{i\to j}^{Y})$ of dimension 20. The empirical evolutionary conservation parameter $\Delta G_i^{\mathrm{stat}}$ for site i is then defined as the norm of vector $\mathbf{P}_i$:

$$\Delta G_i^{\mathrm{stat}} = k \left[\sum_x \left(\ln \frac{p_i(x)}{p_{\mathrm{msa}}(x)} \right)^2 \right]^{1/2} \tag{7.39}$$

To measure the functional coupling of sites, the authors measured the vector $\mathbf{P}_i$ for the full MSA and for a selected subset of the MSA representing a perturbation of the amino acid frequencies at another site j, $\mathbf{P}_{i|\delta j}$. The norm of the difference vector gives the statistical coupling energy between sites i and j:

$$\Delta\Delta G_{ij}^{\mathrm{stat}} = k \left[\sum_x \left(\ln \frac{p_{i|\delta j}(x)}{p_{\mathrm{msa}|\delta_j}(x)} - \ln \frac{p_i(x)}{p_{\mathrm{msa}}(x)} \right)^2 \right]^{1/2} \tag{7.40}$$

which represents the degree to which the probability of individual amino acids at i is dependent on the perturbation at j. If performed systematically for all sites i given a site j, it gives the full map of statistical couplings for position j over all proteins of the MSA.

Using this methodology, the same group later found that evolutionarily conserved sparse networks of coevolving amino acids—termed sectors—represent structural motifs with functional roles, including allosteric communication [80, 81].

7.2.2.3 Methods Based on Interaction Energies

The other quantity that can be used to evaluate the functional coupling between residues in a protein is energy. The method introduced by Colombo and collaborators [82] to analyze the distribution of the intramolecular nonbonded interaction energy in a protein represents an excellent example of this.

This method is based on the calculation of an $N \times N$ matrix describing residue-residue nonbonded interaction energies using configurations from a molecular dynamics simulation:

$$M_{ij} = \left\langle \sum_{\mathrm{k}\in\mathrm{i}} \sum_{l\in j} \left(V_{kl}^{\mathrm{vdW}} + V_{kl}^{\mathrm{ele}} \right) \right\rangle \tag{7.41}$$

where M_{ij} is the element of the matrix describing the interactions between residues i and j, V_{kl}^{vdW} is the van der Waals interaction energy between atom k of residue i and atom l of residue j, V_{kl}^{ele} is the corresponding electrostatic term, and the angle brackets denote an average over an equilibrium trajectory. Since the matrix $\mathbf{M}$ is symmetric, it can be decomposed into N real eigenvalues and corresponding orthogonal eigenvectors, which can be related to the elements M_{ij} through:

$$M_{ij} = \sum_{n} \lambda_n (u_n)_i (u_n)_j \tag{7.42}$$

where λ_n is an eigenvalue and $(u_n)_i$ and $(u_n)_j$ are the i and j components of the corresponding eigenvector $\mathbf{u}_n$ normalized to unit length. If the eigenvalues are labeled in increasing order, it is observed that $\lambda_1 \ll \lambda_2$, where λ_1 is negative. The interaction energy between residues i and j is then approximated by the first term in Eq. (7.42):

$$M_{ij} \approx \lambda_1 (u_1)_i (u_1)_j \tag{7.43}$$

This dimensionality reduction to a principal component allows for the description of the contribution of residue i to the intramolecular nonbonded interaction energy of the protein with a single parameter $(u_1)_i$. Thus, $\mathbf{u}_1$ defines a residue-based profile that is an approximated measure of the interaction energy distribution within the protein structure. The matrix given by Eq. (7.43) can be then used to map regions and pathways of high or low interaction energy density onto the three-dimensional structure of the protein. The scaling parameter λ_1 has been shown to be sensitive to global changes such as the introduction of destabilizing or stabilizing mutations [83], while the effect of conformational dynamics is sensed primarily by $\mathbf{u}_1$.

Note that the average in Eq. (7.41) may also be taken for each of a series of consecutive simulation intervals—eventually for every configuration, that is, with no averaging—to analyze the dependence of the energy distribution on changes in protein conformation or on binding of ligands. Furthermore, the average can be

performed for each of a number of sets of structures classified post-simulation with a structural clustering algorithm.

In recent applications of the method, the configurations from the molecular dynamics simulation are energy minimized before the calculation of the matrix. Although the simulations are performed with explicit solvation (and an ionic concentration when appropriate), by construction the residue-solvent interactions are not included in the matrix. This means that there is an energy imbalance between core and surface pair interactions that ignores solvent effects. To correct this, the protein's solvation free energy, decomposed into residue-pair contributions, is evaluated via the MM/PBSA method [84] and introduced in the calculations so that Eq. (7.41) becomes the potential of mean force:

$$M_{ij} = \left\langle \sum_{\mathrm{k}\in\mathrm{i}} \sum_{l\in j} \left(V_{kl}^{\mathrm{vdW}} + V_{kl}^{\mathrm{ele}} \right) \right\rangle + \Delta G_{ij}^{\mathrm{sol}} \tag{7.44}$$

where $\Delta G_{ij}^{\mathrm{sol}}$ is the solvation free energy contribution to the interaction between residues i and j, calculated over the same configurations as the interaction potential. Although the approximation describing the interaction potential of an amino acid residue with a single parameter might seem venturous, the authors have applied the method to different problems, including prediction of epitopes [85] and analysis of allosteric pathways [67], with significant success.

On a different basis, Yaliraki and collaborators [86] introduced a graph-theoretical model to predict allosteric sites and communication pathways through the derivation of so-called bond-to-bond propensities. In their graph model, nodes are atoms and edges represent covalent and noncovalent bonds (hydrogen bonds, salt bridges, hydrophobic tethers, and electrostatic interactions). A fundamental feature of this model is that edges receive weights derived from interatomic potentials. Earlier, Ribeiro and Ortiz [87] had devised a similar network model under the assumption that signal transmission and allostery could be more accurately described as efficient energy propagation than as correlations of atomic positions, an assumption that they validated by performing energy propagation MD simulations. The resulting all-atom graph is analyzed to calculate the effect that the fluctuations of an edge have on any other edge of the graph, giving a propensity score for each bond that measures the strength of its coupling to the active site through the graph. The significance of a high score propensity is evaluated by comparing each bond to the set of bonds at a similar geometric distance from the active site, as well as to a reference set of 100 representative proteins. Computation time scales almost linearly with the edges, making the method applicable to large systems.

7.2.2.4 Perturbation Methods

All methods discussed in the previous sections are susceptible of being used within a perturbation approach, most often consisting in the analysis of the response of the protein to the binding or unbinding of the ligand, by measuring a given metric or

group of metrics before and after the perturbation. In these experiments, the ligand may be a molecule known or expected to be an allosteric effector of the protein [88], or a simple volumetric probe aimed at efficiently scanning cavities with different characteristics in the protein [89]. When using a network model, the perturbation may be performed on a single structure [53] or on an ensemble of structures [90]. In some cases, as described in one example below, the perturbation may not be related to ligand binding but to an energy pulse, which propagation/dissipation can be tracked over time and space, indicating potential energy paths in the protein.

An interesting example of the ligand perturbation approach is given by recent work from Gohlke and coworkers [56]. They devised a model to study allosteric effects based solely on changes in flexibility: the perturbation introduced in the system does not consider, by actual design, the possibility of conformational changes between the bound and unbound states. By means of rigidity theory [91], proteins can be decomposed into flexible and rigid regions by determining the number and spatial distribution of independent internal degrees of freedom or so-called floppy modes, which are equivalent to zero-frequency vibrational modes of the system. In this model, the protein structure is represented as a network of atoms including covalent as well as noncovalent (hydrogen bonds, salt bridges, and hydrophobic tethers) constraints. A combinatorial algorithm [92] is then used to count the rotational degrees of freedom floppy modes of the network.

These authors had previously shown that changes in vibrational entropy upon binding of a ligand can be estimated from changes in the number of floppy modes with respect to changes in the number of noncovalent constraints of the network—resulting from the addition of the ligand [93]. This motivated the adoption of an ensemble-based perturbation approach. An ensemble of structures of the bound state, including the allosteric effector and eventually the orthosteric ligand, is first generated by MD simulation. The perturbed state is then obtained by removing the constraints associated with the allosteric effector and/or the orthosteric ligand in each of the networks generated from the chosen MD configurations. Finally, rigidity analysis of the two states is used to predict the residues involved in allosteric communication as well as the existence of cooperativity. To this end, the relative stability of the two states is estimated by calculation of the free energy change associated to the perturbation using the one-step perturbation formula [94], and the identification of residues potentially involved in allosteric signaling is performed by calculation of per-residue perturbation free energies using the linear response approximation [95].

As an example of the energy dissipation approach, Martínez and coworkers have recently combined the anisotropic thermal diffusion (ATD) protocol previously proposed by Ota and Agard [96] with a network model to study heat diffusion and predict residue-protein vibrational couplings [97]. An ATD MD simulation experiment consists of cooling a protein structure to a very low temperature (10 K, no temperature bath), then separately heating each residue by independent coupling to a heat reservoir, and measuring the temperature of the protein after a fixed interval. Thus, one can quantify the strength of the vibrational coupling of each residue to the protein as a whole. Residues which quickly dissipate excess vibrational energy have been previously found to be essential to protein activity in example cases [98]. The

authors investigated whether those residues correspond to nodes of higher than average centrality in the corresponding residue networks, by studying heat diffusion on the network as a function of the local temperatures and the connectivity between residues.

7.3 Outlook

Although a relatively young field, the computational study of allostery, including the prediction of allosteric sites, has already reached maturity. If we approach the problem from physics, the basic principles on which we can build, be it for the prediction of allosteric sites or to improve our understanding of allosteric communication, are known since a long time and come down to basic concepts such as atomic interactions and dynamics. These can be certainly approached from very different models and methods, but we should be worryingly surprised if they would give us fundamentally different answers. Significant progress in this front is therefore unlikely unless we see in the coming years a fundamental step forward in force field development or computational speed, or the irruption of new, unexpected experimental data making us revise how we see allostery from a physical point of view. If we approach the problem from biology, the concepts at stage seem also clear and may be based on sequence and/or structural conservation—not addressed here but also integrated in some strategies [99]—and coevolution of certain characters (residues or groups of them). Again, in the absence of orthogonal experimental data or the possibility to analyze much larger sets of allosteric proteins from a phylogenetics point of view, changing our views on how allostery emerged in evolution and evolved—we are probably more in the dark here than in the physics domain—significant progress in this front is also unlikely. The integration of different approaches and different reporters of allostery that have shown already some success seems, on the other hand, a good way to go from where we stand. Indeed, community efforts to benchmark [100] and integrate the different methods available will be at a certain point necessary if we want to progress.

The parallelisms between allostery and protein-folding research are remarkable. In both cases one would like to understand the problem (allosteric communication and associated conformational transitions in one case and the folding process in the other) from a physical and biological standpoint and be able to make predictions (the presence of allosteric sites or the native three-dimensional fold). Computational approaches used in these fields have also many similarities when they are not factually the same. If we look at the evolution of the older field of computational protein folding, we will see that there have been no major theoretical or methodological advances for several years now, with incremental steps being often enabled by corresponding advances in computational power or by the increasing information content of databases. Interestingly, DeepMind's AlphaFold has just entered the field of protein fold prediction with excellent results in CASP13 (http://predictioncenter.org/casp13). Comparatively, the irruption of artificial intelligence in the field of allostery—not covered in this revision—is still incipient. Nevertheless, there already

exist tools with prediction capacities at the level of the best methods [52, 101–103], often adding the machine learning layer on top of earlier models including information on dynamics. The impression is that radical improvement in this direction will require access to much larger amounts of data on allosteric proteins. This points out, once more, that continued efforts to identify and characterize allosteric proteins at the experimental level and to curate and classify these data to make it accessible through resources such as the AlloSteric Database [18] are key for the progress of the field.

Acknowledgements The Spanish Ministry for Science, Innovation and Universities is acknowledged for financial support under grant BIO2015-66674-R.

References

1. Yates RA, Pardee AB (1956) Control of pyrimidine biosynthesis in *Escherichia coli* by a feedback mechanism. J Biol Chem 221(2):757–770
2. Monod J, Jacob F (1961) General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol 26(0):389–401
3. Monod J, Changeux JP, Jacob F (1963) Allosteric proteins and cellular control systems. J Mol Biol 6(4):306–329
4. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12(1):88–118
5. Koshland DE, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5(1):365–385
6. McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267(5612):585–590
7. Cárdenas ML, Rabajille E, Niemeyer H (1978) Maintenance of the monomeric structure of glucokinase under reacting conditions. Arch Biochem Biophys 190(1):142–148
8. Guo J, Zhou HX (2016) Protein allostery and conformational dynamics. Chem Rev 116(11):6503–6515
9. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508(7496):331–339
10. Tsai CJ, Nussinov R (2014) A unified view of "how allostery works". PLoS Comput Biol 10(2):e1003394
11. Cooper A, Dryden DTF (1984) Allostery without conformational change. Eur Biophys J 11(2):103–109
12. Bowman GR, Geissler PL (2012) Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc Natl Acad Sci U S A 109(29):11681–11686
13. Cimermancic P, Weinkam P, Rettenmaier TJ, Bichmann L, Keedy DA, Woldeyes RA, Schneidman-Duhovny D, Demerdash ON, Mitchell JC, Wells JA, Fraser JS, Sali A (2016) Cryptosite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. J Mol Biol 428(4):709–719
14. Hardy JA, Lam J, Nguyen JT, O'Brien T, Wells JA (2004) Discovery of an allosteric site in the caspases. Proc Natl Acad Sci U S A 101(34):12461–12466
15. del Sol A, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 17(8):1042–1050
16. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57(3):433–443
17. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA (2013) Modulation of allostery by protein intrinsic disorder. Nature 498(7454):390–394

18. Huang Z, Mou L, Shen Q, Lu S, Li C, Liu X, Wang G, Li S, Geng L, Liu Y, Wu J, Chen G, Zhang J (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42(D1):D510–D516
19. Xu Q, Tang Q, Katsonis P, Lichtarge O, Jones D, Bovo S, Babbi G, Martelli PL, Casadio R, Lee GR, Seok C, Fenton AW, Dunbrack RL (2017) Benchmarking predictions of allostery in liver pyruvate kinase in CAGI4. Hum Mutat 38(9):1123–1131
20. Collier G, Ortiz V (2013) Emerging computational approaches for the study of protein allostery. Arch Biochem Biophys 538(1):6–15
21. Feher VA, Durrant JD, Van Wart AT, Amaro RE (2014) Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol 25:98–103
22. Greener JG, Sternberg MJ (2018) Structure-based prediction of protein allostery. Curr Opin Struct Biol 50:1–8
23. Lu S, Huang W, Zhang J (2014) Recent computational advances in the identification of allosteric sites in proteins. Drug Discov Today 19(10):1595–1600
24. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE (2016) Emerging computational methods for the rational discovery of allosteric drugs. Chem Rev 116(11): 6370–6390
25. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Ed 45(25):4064–4092
26. van Gunsteren WF, Berendsen HJC (1990) Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. Angew Chem Int Ed 29(9):992–1023
27. Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25(13):1584–1604
28. Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102(13):5451–5459
29. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. J Chem Phys 98(12):10089–10092
30. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 23(3):327–341
31. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (2016) Coarse-grained protein models and their applications. Chem Rev 116(14):7898–7936
32. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ (2008) The MARTINI coarse-grained force field: extension to proteins. J Chem Theory Comput 4(5):819–834
33. Periole X, Cavalli M, Marrink SJ, Ceruso MA (2009) Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. J Chem Theory Comput 5(9):2531–2543
34. Feig M, Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 14(2):217–224
35. Onufriev A (2008) Implicit solvent models in molecular dynamics simulations: a brief overview. Annu Rev Comput Chem 4:125–137
36. Hünenberger PH (2005) Thermostat algorithms for molecular dynamics simulations. Adv Polym Sci 173:105–149
37. Bahar I, Lezon TR, Bakan A, Shrivastava IH (2010) Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev 110(3):1463–1497
38. Sanejouand YH (2013) Elastic network models: theoretical and empirical foundations. Methods Mol Biol 924:601–616
39. Riccardi D, Cui Q, Phillips GN (2009) Application of elastic network models to proteins in the crystalline state. Biophys J 96(2):464–475

40. Zheng W, Brooks BR, Thirumalai D (2006) Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. Proc Natl Acad Sci U S A 103(20):7664–7669
41. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80(1):505–515
42. Tirion MM (1996) Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 77(9):1905–1908
43. Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2(3):173–181
44. Haliloglu T, Bahar I, Erman B (1997) Gaussian dynamics of folded proteins. Phys Rev Lett 79(16):3090–3093
45. Eyal E, Yang LW, Bahar I (2006) Anisotropic network model: systematic evaluation and a new web interface. Bioinformatics 22(21):2619–2627
46. Delarue M, Sanejouand YH (2002) Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 320(5): 1011–1024
47. Kundu S, Melton JS, Sorensen DC, Phillips GN (2002) Dynamics of proteins in crystals: comparison of experiment with simple models. Biophys J 83(2):723–732
48. Bahar I, Rader A (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15(5):586–592
49. Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, Gao J, Hong Y, Li Q, Ni D, Xu J, Chen G, Zhang J (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. Nucleic Acids Res 46(W1):W451–W458
50. Li H, Chang YY, Lee JY, Bahar I, Yang LW (2017) DynOmics: dynamics of structural proteome and beyond. Nucleic Acids Res 45(W1):W374–W380
51. Clarke D, Sethi A, Li S, Kumar S, Chang RWF, Chen J, Gerstein M (2016) Identifying allosteric hotspots with dynamics: application to inter and intra-species conservation. Structure 24(5):826–837
52. Greener JG, Sternberg MJ (2015) AlloPred: prediction of allosteric pockets on proteins using normal mode perturbation analysis. BMC Bioinf 16(1):335
53. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of protein allosteric and regulatory sites. Bioinformatics 30(9):1314–1315
54. Goncearenco A, Mitternacht S, Yong T, Eisenhaber B, Eisenhaber F, Berezovsky IN (2013) SPACER: server for predicting allosteric communication and effects of regulation. Nucleic Acids Res 41(W1):W266–W272
55. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13(9):831
56. Pfleger C, Minges A, Boehm M, McClendon CL, Torella R, Gohlke H (2017) Ensemble and rigidity theory-based perturbation approach to analyze dynamic allostery. J Chem Theory Comput 13(12):6343–6357
57. Nussinov R, Tsai CJ (2015) Allostery without a conformational change? Revisiting the paradigm. Curr Opin Struct Biol 30:17–24
58. Thayer KM, Galganov JC, Stein AJ (2017) Dependence of prevalence of contiguous pathways in proteins on structural complexity. PLoS One 12(12):e0188616
59. Barbany M, Meyer T, Hospital A, Faustino I, D'Abramo M, Morata J, Orozco M, de la Cruz X (2015) Molecular dynamics study of naturally existing cavity couplings in proteins. PLoS One 10(3):e0119978
60. Hardy JA, Wells JA (2004) Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 14(6):706–715
61. David CC, Jacobs DJ (2014) Principal component analysis: a method for determining the essential dynamics of proteins. Methods Mol Biol 1084:193–226
62. Karplus M, Kushick JN (1981) Method for estimating the configurational entropy of macromolecules. Macromolecules 14(2):325–332

63. Schlitter J (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chem Phys Lett 215(6):617–621
64. Allain A, Beauchêne ICd, Langenfeld F, Guarracino Y, Laine E, Tchertanov L (2014) Allosteric pathway identification through network analysis: from molecular dynamics simulations to interactive 2D and 3D graphs. Faraday Discuss 169(0):303–321
65. Missimer JH, Steinmetz MO, Baron R, Winkler FK, Kammerer RA, Daura X, van Gunsteren WF (2007) Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. Protein Sci 16(7):1349–1359
66. Kidera A, Go N (1992) Normal mode refinement: crystallographic refinement of protein dynamic structure: I. Theory and test by simulated diffraction data. J Mol Biol 225(2):457–475
67. Morra G, Genoni A, Colombo G (2014) Mechanisms of differential allosteric modulation in homologous proteins: insights from the analysis of internal dynamics and energetics of PDZ domains. J Chem Theory Comput 10(12):5677–5689
68. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3): 379–423
69. Bérut A, Arakelyan A, Petrosyan A, Ciliberto S, Dillenschneider R, Lutz E (2012) Experimental verification of Landauers principle linking information and thermodynamics. Nature 483(7388):187–189
70. Županović P, Kuić D (2018) Relation between Boltzmann and Gibbs entropy and example with multinomial distribution. J Phys Commun 2(4):045002
71. Kalinin MI, Kononogov SA (2005) Boltzmann's constant, the energy meaning of temperature, and thermodynamic irreversibility. Meas Tech 48(7):632–636
72. Hlavácková-Schindler K, Palus M, Vejmelka M, Bhattacharya J (2007) Causality detection based on information-theoretic approaches in time series analysis. Phys Rep 441(1):1–46
73. Schreiber T (2000) Measuring information transfer. Phys Rev Lett 85(2):461–464
74. Hacisuleyman A, Erman B (2017) Entropy transfer between residue pairs and allostery in proteins: quantifying allosteric communication in ubiquitin. PLoS Comput Biol 13(1): e1005319
75. Lange OF, Grubmüller H (2006) Generalized correlation for biomolecular dynamics. Proteins 62(4):1053–1061
76. Negre CFA, Morzan UN, Hendrickson HP, Pal R, Lisi GP, Loria JP, Rivalta I, Ho J, Batista VS (2018) Eigenvector centrality for characterization of protein allosteric pathways. Proc Natl Acad Sci U S A 115(52):E12201–E12208
77. Lopez T, Dalton K, Tomlinson A, Pande V, Frydman J (2017) An information theoretic framework reveals a tunable allosteric network in group II chaperonins. Nat Struct Mol Biol 24(9):726–733
78. Sung YM, Wilkins AD, Rodriguez GJ, Wensel TG, Lichtarge O (2016) Intramolecular allosteric communication in dopamine D2 receptor revealed by evolutionary amino acid covariation. Proc Natl Acad Sci U S A 113(13):3539–3544
79. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286(5438):295–299
80. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138(4):774–786
81. Reynolds K, McLaughlin R, Ranganathan R (2011) Hot spots for allosteric regulation on protein surfaces. Cell 147(7):1564–1575
82. Tiana G, Simona F, Mori GMSD, Broglia RA, Colombo G (2004) Understanding the determinants of stability and folding of small globular proteins from their energetics. Protein Sci 13(1):113–124
83. Morra G, Colombo G (2008) Relationship between energy distribution and fold stability: insights from molecular dynamics simulations of native and mutant proteins. Proteins 72(2):660–672
84. Wang W, Lim WA, Jakalian A, Wang J, Wang J, Luo R, Bayly CI, Kollman PA (2001) An analysis of the interactions between the Sem5 SH3 domain and its ligands using molecular

dynamics, free energy calculations, and sequence analysis. J Am Chem Soc 123(17): 3986–3994
85. Scarabelli G, Morra G, Colombo G (2010) Predicting interaction sites from the energetics of isolated proteins: a new approach to epitope mapping. Biophys J 98(9):1966–1975
86. Amor BRC, Schaub MT, Yaliraki SN, Barahona M (2016) Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. Nat Commun 7:12477
87. Ribeiro AAST, Ortiz V (2015) Energy propagation and network energetic coupling in proteins. J Phys Chem B 119(5):1835–1846
88. Morra G, Meli M, Colombo G (2018) How the ligand-induced reorganization of protein internal energies is coupled to conformational events. J Chem Theory Comput 14(11): 5992–6001
89. Panjkovich A, Daura X (2012) Exploiting protein flexibility to predict the location of allosteric sites. BMC Bioinf 13:273
90. Greener JG, Filippis I, Sternberg MJE (2017) Predicting protein dynamics and allostery using multi-protein atomic distance constraints. Structure 25(3):546–558
91. Hermans SMA, Pfleger C, Nutschel C, Hanke CA, Gohlke H (2017) Rigidity theory for biomolecules: concepts, software, and applications. WIREs Comput Mol Sci 7(4):e1311
92. Jacobs DJ, Thorpe MF (1995) Generic rigidity percolation: the pebble game. Phys Rev Lett 75(22):4051–4054
93. Gohlke H, Ben-Shalom IY, Kopitz H, Pfeiffer-Marek S, Baringhaus KH (2017) Rigidity theory-based approximation of vibrational entropy changes upon binding to biomolecules. J Chem Theory Comput 13(4):1495–1502
94. Zwanzig RW (1954) Hightemperature equation of state by a perturbation method. I. Nonpolar gases. J Chem Phys 22(8):1420–1426
95. Lee FS, Chu ZT, Bolger MB, Warshel A (1992) Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. Protein Eng Des Sel 5(3):215–228
96. Ota N, Agard DA (2005) Intramolecular signaling pathways revealed by modeling anisotropic thermal diffusion. J Mol Biol 351(2):345–354
97. Censoni L, dos Santos MH, Martnez L (2017) A network model predicts the intensity of residue-protein thermal coupling. Bioinformatics 33(14):2106–2113
98. Martínez L, Figueira ACM, Webb P, Polikarpov I, Skaf MS (2011) Mapping the intramolecular vibrational energy flow in proteins reveals functionally important residues. J Phys Chem Lett 2(16):2073–2078
99. Panjkovich A, Daura X (2010) Assessing the structural conservation of protein pockets to study functional and allosteric sites: implications for drug discovery. BMC Struct Biol 10:9
100. Huang W, Wang G, Shen Q, Liu X, Lu S, Geng L, Huang Z, Zhang J (2015) ASBench: benchmarking sets for allosteric discovery. Bioinformatics 31(15):2598–2600
101. Akbar R, Helms V (2018) ALLO: a tool to discriminate and prioritize allosteric pockets. Chem Biol Drug Des 91(4):845–853
102. Chen ASY, Westwood NJ, Brear P, Rogers GW, Mavridis L, Mitchell JBO (2016) A random forest model for predicting allosteric and functional sites on proteins. Mol Inf 35:125–135
103. Song K, Liu X, Huang W, Lu S, Shen Q, Zhang L, Zhang J (2017) Improved method for the identification and validation of allosteric sites. J Chem Inf Model 57(9):2358–2363

Chapter 8
Advances in NMR Methods to Identify Allosteric Sites and Allosteric Ligands

Hazem Abdelkarim, Ben Hitchinson, Avik Banerjee, and Vadim Gaponenko

Abstract NMR allows assessment of protein structure in solution. Unlike conventional X-ray crystallography that provides snapshots of protein conformations, all conformational states are simultaneously accessible to analysis by NMR. This is a significant advantage for discovery and characterization of allosteric effects. These effects are observed when binding at one site of the protein affects another distinct site through conformational transitions. Allosteric regulation of proteins has been observed in multiple physiological processes in health and disease, providing an opportunity for the development of allosteric inhibitors. These compounds do not directly interact with the orthosteric site of the protein but influence its structure and function. In this book chapter, we provide an overview on how NMR methods are utilized to identify allosteric sites and to discover novel inhibitors, highlighting examples from the field. We also describe how NMR has contributed to understanding of allosteric mechanisms and propose that it is likely to play an important role in clarification and further development of key concepts of allostery.

Keywords NMR · Allosteric ligands · Allosteric sites · NMR methods

Abbreviations

14–3-3	Conserved regulatory/adaptive proteins
^{19}F NMR	Fluorine NMR spectroscopy
Å	Angstroms
AMP	Adenosine-5-monophosphate

H. Abdelkarim · B. Hitchinson · V. Gaponenko (✉)
Department of Biochemistry and Molecular Genetics, College of Medicine, University of Illinois at Chicago, Chicago, IL, USA
e-mail: vadimg@uic.edu

A. Banerjee
Department of Chemistry, University of Illinois at Chicago, Chicago, IL, USA

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_8

CHESCA	Chemical shift covariance analysis
CORCEMA-ST	Conformational exchange matrix analysis of saturation transfer
CPMG	Carr-Purcell-Meiboom-Gill
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
DNA	Deoxyribonucleic acid
Eg5	Plus-end-directed kinesin-related protein
EPAC	Exchange protein directly activated by cAMP
F	Fluorine
FAXS	Fluorine chemical shift anisotropy and exchange for screening
GDP	Guanosine diphosphate
GMPPCP	β, γ-Methyleneguanosine triphosphate
GPCRs	G-protein-coupled receptors
GTP	Guanosine triphosphate
H	Hydrogen
HBGAs	Histo-blood antigens
HSQC/HMQC	Heteronuclear single/multiple quantum coherence/correlation
ITK	Interleukin-2-inducible T-cell kinase
KNF	Koshland, Némethy, and Filmer
MD	Molecular dynamics
MerR	The bacterial mercury resistance
MWC	Monod-Wyman-Changeux
N	Nitrogen
NMR	Nuclear magnetic resonance
NOEs	Nuclear Overhauser effects
pABA	p-Aminobenzoic acid
PCSs	Pseudocontact shifts
PDK1	Phosphoinositide-dependent kinase-1
PIF	PDK1 interacting fragment
PREs	Paramagnetic relaxation enhancements
R_1	^{15}N spin-lattice relaxation experiments
R_2	Spin-spin relaxation experiments
RDCs	Residual dipolar couplings
RNA	Ribonucleic acid
SOS	Son of sevenless
STD-NMR	Saturation transfer difference
TET	2,2,2-Trifluoroethanethiol
THF	Tetrahydrofolate
TROSY	Transverse relaxation optimized spectroscopy
VLPs	Norovirus virus-like particles
WaterLOGSY	Water ligand observed via gradient spectroscopy

8.1 Introduction

NMR is a technique that detects resonance frequencies of proteins and small molecules in a magnetic field. These resonance frequencies change in response to altering electromagnetic environment due to binding and conformational rearrangements. Proteins are inherently flexible macromolecules that constantly sample different conformational states. These dynamic properties are essential for protein functions as diverse as enzyme catalysis, signaling, building cellular structures, and driving cell motility. Understanding the basis of protein dynamics is crucial to understanding function and targeting proteins therapeutically.

One way to pursue proteins therapeutically is through allosteric targeting by designing ligands that bind at one site and induce conformational and functional changes at a different site. The mechanism of allostery is intrinsically linked to the dynamic nature of proteins, their ability to adopt an ensemble of conformations, and the ability of a binding event to shift the equilibrium of the conformational ensemble. Several biophysical and biochemical techniques have been employed to investigate allostery. These include, but are not limited to, ligand binding assays [48], fluorescence-based resonance energy transfer experiments [12], and X-ray crystallography [3]. While these methods have provided information about the affinity and binding sites of molecules, significant insight into conformational transitions has come from NMR experiments.

NMR has provided near-atomic resolution of dynamic processes on a timescale ranging from nano- to milliseconds. These dynamic processes underlie allosteric transitions in proteins and explain their thermodynamic and kinetic properties. Thus, NMR has become the preferred approach for investigating protein allostery.

In recent years, advances in NMR methods have made it possible to monitor dynamic processes that underlie allostery. In this chapter, we will focus on how NMR methods have been used to identify allosteric sites, effectors, and inhibitors and how NMR has advanced the concept of allostery.

8.2 NMR and the Concept of Allostery

NMR has contributed to the development of models that describe protein allostery. One important example is allosteric regulation of hemoglobin, a tetrameric oxygen transporting protein in red blood cells. The two widely accepted models of allostery in hemoglobin are the symmetry (concerted) model proposed by Monod-Wyman-Changeux (MWC) and the sequential model proposed by Koshland, Némethy, and Filmer (KNF). The MWC model postulates that the protein exists in two or more distinct symmetric states depending on their thermal equilibrium and stability. Effector binding shifts the equilibrium in favor of one of these states. On the other hand, the KNF model integrated the concept of cooperativity of protein subunits (positive and negative) and postulated that binding of an allosteric effector to one unit can cause a neighboring unit to change the conformation to a high/low affinity state.

Based on the MWC model, the sigmoidal binding of oxygen to hemoglobin can be explained as a shift in equilibrium from one state to another state as the oxygen concentration changes leading to a concerted cooperativity. However, NMR studies under high phosphate levels have revealed that the alpha subunit has higher affinity for oxygen than the beta subunit. This finding led to the conclusion that sequential cooperativity can exist in hemoglobin binding to oxygen in addition to the concerted mechanism that follows the MWC model of allostery [33].

The dynamic allostery model was introduced by Cooper et al. in 1984 to describe the existence of allosteric dynamic changes in proteins without a noticeable change or shift in the conformational ensembles. This allostery is dependent on the magnitude of changes in thermal amplitude or vibrational frequencies induced by allosteric ligand binding. The dynamic allostery model accounted for the entropic contribution to the mean conformational ensemble as a thermodynamic factor working together with the enthalpic contribution, considered in MWC and KNF models. NMR methods and approaches have been a cornerstone in providing experimental evidence in studying dynamic protein allostery structurally and quantitatively [28, 97]. NMR studies expanded the term to account for not only the thermodynamic nature of the changes, but also to include a perspective on protein kinetics and function [28, 97]. Several reviews have discussed the impact of NMR approaches on the concept of dynamic allostery, and in later sections of this book chapter we will highlight these approaches and how they facilitated the study of dynamic allostery [28, 97].

As it became apparent that the existing models of allostery explain some but not all characteristics of allosteric behavior in proteins, the need to unify these models arose. One response to this need is the development of the ensemble allosteric model, which interprets the energy landscape in terms of the probabilistic ensembles of conformations. In this model, the probability of sampling of specific conformations within the structural ensemble exhibited by the protein is defined thermodynamically, as a free energy balance [69, 100]. Significant support for the ensemble allosteric model came from NMR relaxation studies, as they described the entropic contribution to the allosteric process. These studies led to an idea that fast motions underlie conformational entropy. Importantly, the concept that conformational entropy constitutes a major part of the free energy protein-ligand interactions was developed [13, 101]. NMR investigation of the calcium-binding protein calmodulin provided evidence that conformational entropy, evidenced in local, long-range, or dynamic transitions, is directly involved in molecular recognition and, thus, protein function [49].

Interestingly, chemical shift analysis and ^{15}N Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR experiments, providing access to backbone dynamics on the millisecond to microsecond timescale, have revealed that the allosteric regulation of adenylate kinase can be tuned by altering dynamics of the spatially distinct mobile domain of the enzyme [81]. Dynamics of this domain were able to alter the affinity of adenylate kinase for its substrate. This study highlighted the importance of dynamic regions in the functional energy landscape of protein molecules and hinted at a possibility that unstructured regions within proteins can control thermodynamics and affinity for binding partners.

In line with this concept is the conformation selection and population shift model to describe protein folding, oligomerization, binding to effectors, and function. This model has significantly contributed to the concept of allostery by augmenting current models to consider conformational ensembles of several states instead of the simple two-state models [53, 96, 103]. Additionally, the Ranganathan research group, through statistical mapping of the amino acid networks, identified thermodynamic coupling between amino acids and their long-range effects on protein conformational ensembles [53, 55, 93]. Testing this hypothesis on important proteins will likely define the future of NMR investigations of allostery.

One example of the studies highlighting the role of protein dynamics in allosteric regulation is the Ras family of proto-oncogenes and the discovery of drugs that target these proteins [27, 60, 72, 74, 88]. Ras proteins (H-Ras, N-Ras and K-Ras4A and 4B) are small GTPases with a conserved catalytic domain and a highly flexible C-terminal hypervariable region. They regulate important signaling pathways by cycling between the active GTP-bound and the inactive GDP-bound states. Oncogenic mutations within the catalytic domain of Ras hinder the ability to hydrolyze GTP, rending the protein constitutively active. For a long period of time it was thought that Ras proteins existed only in on and off conformations.

Evidence that Ras proteins sample different conformational states came from NMR studies that discovered the presence of two states (state 1 and state 2) in H-Ras bound to the GTP analog GMPPNP [1, 26, 89, 90]. Only state 2 displayed high affinity for Raf kinase, an important effector of Ras. Further NMR studies revealed that the nucleotide sensing switch I and switch II regions exist in multiple conformations in GTP- and GDP-bound Ras [57]. Intriguingly, X-ray crystallography studies have found that allosteric regulation differs between wild type and oncogenic mutants of K-Ras and between Ras isoforms [7–9, 21, 70, 76]. Inhibitors of Ras are still lacking in the clinics and development of allosteric modulators of Ras as anticancer therapeutic agents is underway. For instance, Matsumoto et al. have reported the discovery of small molecules bearing a naphthalene group to be allosteric inhibitors of H-Ras in its GTP state [65]. Based on NMR studies, the compounds, via their naphthalene group, bind a hydrophobic pocket located between the switch I and II region of H-Ras. Upon binding, the inhibitor will allosterically influence the effector binding (c-Raf1) region by causing conformational changes in switch I and β2 strand region [65]. Also Maurer et al. have used saturation transfer difference (STD) NMR to identify fragments against K-Ras4B G12D [66]. The STD-NMR fragment-based screening yielded 240 primary hits that can target the protein in active (GMPPCP-loaded) and inactive (GDP-loaded) states [66]. Further NMR experiments supported by computational studies, crystallography, and biological approaches identified the new binding pocket with the ability to inhibit K-Ras activity at cellular level by interfering with son of sevenless (SOS)-mediated nucleotide exchange [66]. Because these ligands are not targeting the active pocket of K-Ras, they can potentially be considered allosteric inhibitors. Thus, NMR has allowed significant advancement in understanding allostery and discovery of allosteric inhibitors.

8.3 NMR Methods to Identify Allosteric Sites and Allosteric Ligands

Analysis of chemical shifts in two-dimensional heteronuclear single (or multiple) quantum coherence (HSQC or HMQC) experiments and their transverse relaxation optimized (TROSY) equivalents for large proteins can provide significant information about allostery because these shifts act as reporters of conformational changes. One example of how HSQC has enhanced our knowledge of protein targeting by allosteric molecules was provided by Oyen et al., who identified the allosteric binding site of a nanobody targeting the dihydrofolate reductase (DHFR) enzyme from *E. coli* using HSQC and ^{15}N-labeled protein [75]. DHFR catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate (THF) in an NADPH-dependent manner, with the rate-limiting step being THF release. This is a well-studied enzyme that has subsequently been used to establish the allosteric network paradigm using NMR relaxation experiments. These experiments commonly use the HSQC framework and other pulse sequence components to measure protein conformational dynamics [41]. Fast motions on the picosecond to nanosecond timescale are often measured in ^{1}H-^{15}N heteronuclear nuclear Overhauser effects (NOEs), ^{15}N spin-lattice (R_1), and spin-spin (R_2) relaxation experiments, while much slower conformational exchange can be characterized by the relaxation-compensated CPMG and off-resonance rotating frame relaxation experiments [52, 58, 59].

It is well established that within proteins, networks of interactions mediate the propagation of conformational changes caused by allosteric events. By plotting chemical shift patterns or by characterizing relaxation rates upon addition of ligands, we can follow transitions in protein conformation. If a library of ligands including allosteric binders is available, it is possible to interrogate the allosteric network in a protein. Selvaratnum et al. helped to establish this paradigm by studying the cAMP-binding domain of the exchange protein directly activated by cAMP (EPAC). EPAC is a guanine exchange factor that is regulated by cAMP. In these studies, they mapped allosteric networks activated by binding of different ligands based on covariance analysis of NMR chemical shifts [83]. This method, which they termed chemical shift covariance analysis (CHESCA), utilized five different EPAC ligands known to stabilize five different states of the protein in HSQC experiments. Once all resonances were assigned, the authors performed statistical analysis of chemical shifts for different residues via pairwise correlations, agglomerative clustering, and singular value decomposition.

Protein conformations can be analyzed with the aid of residual dipolar couplings (RDCs), dipolar interactions between neighboring nuclei, providing orientational information on internuclear vectors. Measurement of RDCs requires partial orientation of protein molecules in the magnetic field that is commonly accomplished by the addition of alignment media [94]. An intriguing example of the use of RDCs in understanding protein allostery was provided by Hansen et al. and Lukin et al. who studied hemoglobin. The authors found that the crystal structures of the original R hemoglobin poorly agreed with experimental RDCs measured in solution and

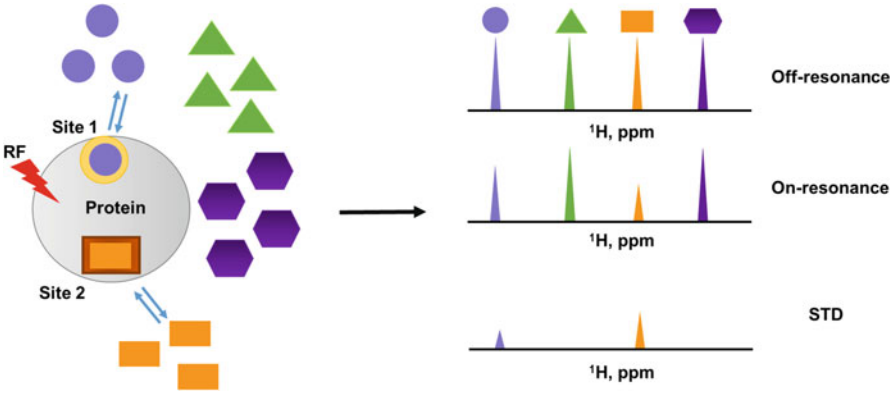


Fig. 8.1 Schematic representation of STD-NMR

proposed that hemoglobin exists in multiple conformational states [31, 61]. Thus, two-dimensional NMR methods can provide a wealth of information regarding protein allostery.

NMR has become a very useful technique for identification of allosteric ligands [6, 28, 52, 97]. The approach is to first identify binders using a ligand-based approach, such as saturation transfer difference NMR (STD-NMR) or water ligand observed via gradient spectroscopy (WaterLOGSY), and then to select ligands that do not directly interact with the orthosteric site.

STD-NMR provides information about ligand binding epitopes that are in direct contact with the protein of interest by showing a significantly stronger STD signal than those not in contact (Fig. 8.1). STD-NMR can be combined with computational and cell biology approaches to identify allosteric ligands. For instance, the complete relaxation and conformational exchange matrix analysis of saturation transfer (CORCEMA-ST) software has proved useful for the identification of allosteric ligand binding sites, where a known structure of a protein already exists [38, 39, 47, 67, 68]. CORCEMA-ST is a software that predicts STD-NMR intensities for a proposed interaction between a ligand and receptor using information about saturated protein protons and a variety of other parameters, such as correlation times, exchange rates, and spectrometer frequency [38, 78]. In 2013, Zhang et al. developed this approach to identify the allosteric binding site of several ligands of the plus-end-directed kinesin-5 subfamily protein, Eg5 [102].

More recently, Krimm [46] used orthosteric, fragment-like compounds termed "spy compounds" as probes for binding of allosteric ligands [46]. This technique found allosteric transitions induced by ligands via observing the changes in the binding mode of the orthosteric "spy" molecule in the absence and in the presence of an allosteric ligand. Importantly, the spy molecule binds at a site distinct from the allosteric ligand providing a method to assess the effects of allosteric binding and allosteric regulation of proteins. Additionally, Von Geldern et al. have used STD-NMR to support the identification of a set of novel allosteric inhibitors of

fructose-1,6-bisphosphatase [98]. Using N^6-methyl adenosine-5-monophosphate (AMP) as an allosteric probe molecule that binds to the AMP allosteric site, they ran competition STD-NMR experiments with benzoxazole benzenesulfonamide-based inhibitors [98]. The authors showed the ability of the inhibitors to suppress energy transfer to the probe, indicating binding to an allosteric site [98]. Further crystallography and biological data have confirmed the allosteric nature of these inhibitors [98]. Other examples of the use of STD-NMR to discover allosteric ligands include a screen for inhibitors of 3-phosphoinositide-dependent kinase-1 (PDK1) [92], determining the binding isotherms of L-fucose to virus-like particles [63], and identification of allosteric inhibitors for interleukin-2-inducible T-cell kinase [30].

In WaterLOGSY experiments, selective magnetization of bulk water molecules can be transferred to the target protein, to sequestered water molecules bound to the protein, and to water molecules in the interface of the protein-ligand complex [14, 15]. Hence, the ligand resonances belonging to the groups in contact with the protein will exhibit positive intensities, while the groups that do not interact with the protein will produce negative signals [14, 15] (Fig. 8.2).

As with STD-NMR, all binders are first identified in the ligand-based NMR screens and allosteric compounds are later selected in competition experiments and further validated using heteronuclear NMR experiments or other approaches. Examples of discovery of allosteric ligands using WaterLOGSY include fragment-based screening to target the myristoyl allosteric binding pocket of Abl kinase followed by NMR analysis of conformations of the enzyme labeled with ^{15}N valine [36], identification of allosteric inhibitors of farnesyl pyrophosphate synthase using WaterLOGSY competition assays [37], finding allosteric ligands for dihydropteroate synthase [29], and development of selective allosteric inhibitors of 14-3-3 adaptive protein [85]. Based on these examples, STD and WaterLOGSY-NMR approaches are powerful and fast approaches to identify allosteric ligands.

As the ^{19}F nucleus is exquisitely sensitive to its electromagnetic environment [10, 25, 34, 42–45, 50, 51, 56, 64, 77, 84, 99], ^{19}F NMR has become a powerful

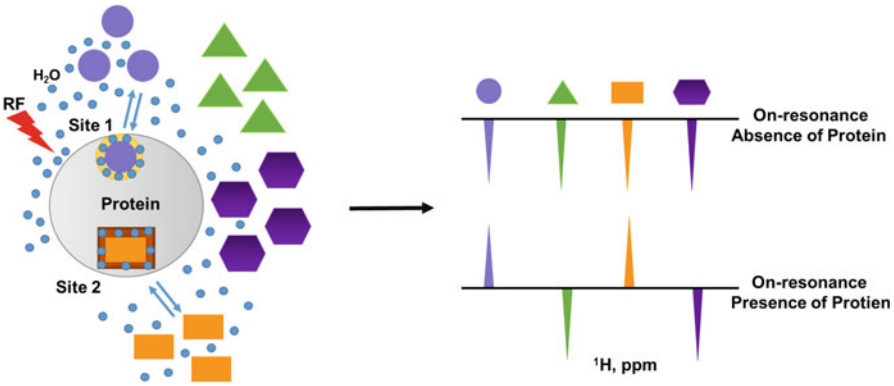


Fig. 8.2 Schematic representation of WaterLOGSY

approach to study allostery in protein folding/unfolding, dynamics, enzyme kinetics, conformational changes, binding of natural substrates and ligands, and even thermodynamic and physical properties of proteins [10, 25, 34, 42–45, 50, 51, 56, 64, 77, 84, 99]. Analysis of ^{19}F chemical shifts and signal line-widths in trifluoromethyl-containing chemical probes has helped to investigate allosteric behavior of insulin-metal hexamer complexes that play an important role in insulin assembly and storage in the pancreatic β-islets [5, 80]. Another ^{19}F NMR probe 1,1,1-trifluorobromoacetone was used to reveal the mechanism of allosteric regulation in citrate synthase [19]. ^{19}F can also be incorporated into proteins during synthesis via the use of 5-fluoro-tryptophane and 2-fluoro-tyrosine. This approach has helped to investigate allostery in human serum albumin in response to binding of a commonly prescribed blood thinner warfarin [40]; to study ligand-induced conformational changes in apical membrane antigen 1 (AMA1) [24], a malaria therapeutic target to several inhibitors [62]; and to discover a unique allostery network in bacterial mercury resistance (MerR) operon guided by metal and DNA binding [87].

An interesting approach to investigate allostery by ^{19}F NMR was employed by Assemat et al. who used fluorine chemical shift anisotropy and exchange for screening (FAXS, Fig. 8.3) [16] to identify glucokinase activators [2]. This enzyme can be modulated by activators or inhibitors that affect its transition between the active and inactive states. The ^{19}F NMR screening method was based on a spy molecule. By monitoring the R_2 relaxation time of the ^{19}F spin, eight known activators were used to validate the method. In a similar application, Skora and Jahnke developed a ^{19}F NMR dual-site screening assay to distinguish between active site and allosteric inhibitors of Abl [86]. This assay was designed to apply fluorinated reporter molecules known for their binding specificity to either the active site or the allosteric site.

Assessment of conformational rearrangements within G-protein-coupled receptors (GPCRs) is an important area of research that has benefitted from the use of ^{19}F NMR [18]. GPCRs convey signals from the extracellular environment to

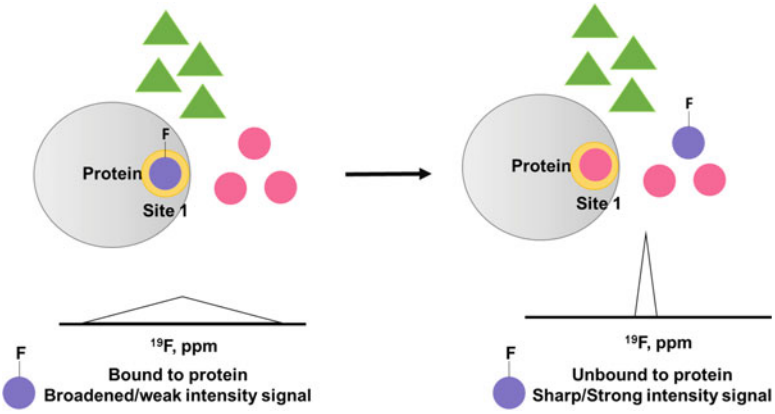


Fig. 8.3 Schematic representation of FAXS

intracellular structures and are highly dynamic proteins. Several methods have been used to incorporate specific labels into receptors to understand ligand binding and to gain insight into activation processes [18]. For instance, Liu et al. used 2,2,2-trifluoroethanethiol (TET) to label cytoplasmic cysteine of β_2-adrenergic receptor to study the binding of several ligands [54]. The ^{19}F NMR labels revealed that intracellular helices VI and VII can adapt two conformational states and ligands can differentially affect these states. Further ^{19}F NMR studies by Horst et al. determined the thermodynamic and kinetic changes associated with ligand binding and the conformational equilibria of β_2-adrenergic receptor [35]. Later, Staus et al., using ^{19}F NMR, showed that a negative allosteric antibody can stabilize one of the two conformational states [91]. In summary, ^{19}F NMR has been widely applied to investigate processes in proteins and to discover allosteric ligands.

A unique position among NMR methods is occupied by paramagnetic approaches. An unpaired electron creating paramagnetic effects can be incorporated into the protein structure via paramagnetic probes or metal-binding sites. Due to the very large magnetic moment of the unpaired electron, long-range distances can be measured from the electron to nuclei in the protein. In many cases, these distances surpass the 6 Å limit of NOEs [23, 32, 73]. Frequently, paramagnetic probes allow measurement of paramagnetic relaxation enhancements (PREs), pseudocontact shifts (PCSs), or RDCs [20, 22, 32, 95]. While PREs provide distance information, PCSs allow determination of orientation of electron-nuclear vectors with respect to the axes of the paramagnetic anisotropic susceptibility tensor. Paramagnetic centers can also induce partial orientation of macromolecules in the external magnetic field and allow measurement of RDCs, providing orientational information on bond vectors [22, 95]. Thus, the availability of paramagnetic probes in proteins allows collection of many types of structural information that can describe conformational rearrangements in protein molecules due to allosteric modulation.

Despite ample advantages offered by paramagnetic NMR approaches to the discovery of allosteric regulation in protein molecules, particularly with respect to long-range distance and angular information, there are relatively few reports describing their use. In most cases, PREs are utilized to measure distances between the protein and the ligand or within the protein itself. Some recent examples include discovery of allosteric regulatory mechanisms in intrinsically disordered proteins α-synuclein and osteopontin [4], detection of inhibitor-induced conformational changes in HIV-1 reverse transcriptase [82], and description of orientations of cyclic nucleotides in the binding site in the murine hyperpolarization-activated cyclic nucleotide-gated ion channel [17]. These studies involved incorporation of unique cysteine residues into proteins for attachment of paramagnetic probes. However, there may be cases where cysteine incorporation might create a problem with protein folding and/or aggregation. Additionally, calculations of paramagnetic anisotropic susceptibility tensor parameters for the use of PCSs and RDCs can be laborious and might hinder a wider use of these approaches. Despite the limitations, we expect that advances in paramagnetic NMR [11, 71, 79] will allow to more fully explore its use in studies of allosteric regulation in proteins.

8.4 Conclusion

In addition to other methods, NMR has guided the evolution of models of allostery, focusing on contributions of protein dynamics and conformational ensembles as drivers of allosteric regulation. Our review provides examples of the use of NMR in the development of models of allostery. These examples highlight the complexity and diversity of allosteric mechanisms and suggest that further refinement of the main concepts of allostery may be needed. We also propose that application of advanced concepts of allosteric regulation will likely lead to deeper understanding of important proteins, including Ras GTPases and GPCRs. Understanding how protein function can be allosterically modulated by ligand binding at a distant site is of paramount importance for drug discovery. NMR continues to play a key role in mechanistic studies of allostery and in identification of allosteric ligands. Multiple advanced NMR-based approaches are widely utilized to find protein binders and to discover the allosteric mechanisms used by these compounds to modulate protein function. Undoubtedly, further refinement of NMR methods will lead to new discoveries in the area of allosteric regulation of proteins.

Acknowledgments This work was funded by the National Institutes of Health (NIH)–National Cancer Institute (NCI) R01CA188427 grant to V.G., the American Cancer Society Postdoctoral Fellowship (132722-PF-18-196-01-DMC) to HA, and the Horizon award under the Congressionally Directed Medical Research Program (CDMPRP), Department of Defense (DoD) (W81XWH-17-10509) to A.B.

References

1. Araki M, Shima F, Yoshikawa Y, Muraoka S, Ijiri Y, Nagahara Y, Shirono T, Kataoka T, Tamura A (2011) Solution structure of the state 1 conformer of GTP-bound H-Ras protein and distinct dynamic properties between the state 1 and state 2 conformers. J Biol Chem 286:39644–39653
2. Assemat O, Antoine M, Fourquez JM, Wierzbicki M, Charton Y, Hennig P, Perron-Sierra F, Ferry G, Boutin JA, Delsuc MA (2015) ^{19}F nuclear magnetic resonance screening of glucokinase activators. Anal Biochem 477:62–68
3. Bauman JD, Patel D, Dharia C, Fromer MW, Ahmed S, Frenkel Y, Vijayan RS, Eck JT, Ho WC, Das K et al (2013) Detecting allosteric sites of HIV-1 reverse transcriptase by X-ray crystallographic fragment screening. J Med Chem 56:2738–2746
4. Beier A, Schwarz TC, Kurzbach D, Platzer G, Tribuzio F, Konrat R (2018) Modulation of correlated segment fluctuations in IDPs upon complex formation as an allosteric regulatory mechanism. J Mol Biol 430:2439–2452
5. Bonaccio M, Ghaderi N, Borchardt D, Dunn MF (2005) Insulin allosteric behavior: detection, identification, and quantification of allosteric states via ^{19}F NMR. Biochemistry 44:7656–7668
6. Boulton S, Melacini G (2016) Advances in NMR methods to map allosteric sites: from models to translation. Chem Rev 116:6267–6304
7. Buhrman G, Wink G, Mattos C (2007) Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. Structure 15:1618–1629

8. Buhrman G, Holzapfel G, Fetics S, Mattos C (2010) Allosteric modulation of Ras positions Q61 for a direct role in catalysis. Proc Natl Acad Sci USA 107:4931–4936
9. Buhrman G, Kumar VS, Cirit M, Haugh JM, Mattos C (2011) Allosteric modulation of Ras-GTP is linked to signal transduction through RAF kinase. J Biol Chem 286:3323–3331
10. Campos-Olivas R, Aziz R, Helms GL, Evans JN, Gronenborn AM (2002) Placement of ^{19}F into the center of GB1: effects on structure and stability. FEBS Lett 517:55–60
11. Clore GM (2015) Practical aspects of paramagnetic relaxation enhancement in biological macromolecules. Methods Enzymol 564:485–497
12. Cornea RL, Gruber SJ, Lockamy EL, Muretta JM, Jin D, Chen J, Dahl R, Bartfai T, Zsebo KM, Gillispie GD, Thomas DD (2013) High-throughput FRET assay yields allosteric SERCA activators. J Biomol Screen 18:97–107
13. Culik RM, Sekhar A, Nagesh J, Deol H, Rumfeldt JAO, Meiering EM, Kay LE (2018) Effects of maturation on the conformational free-energy landscape of SOD1. Proc Natl Acad Sci USA 115:E2546–E2555
14. Dalvit C, Pevarello P, Tato M, Veronesi M, Vulpetti A, Sundstrom M (2000) Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water. J Biomol NMR 18:65–68
15. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B (2001) Water LOGSY as a method for primary NMR screening: practical aspects and range of applicability. J Biomol NMR 21:349–359
16. Dalvit C, Ardini E, Flocco M, Fogliatto GP, Mongelli N, Veronesi M (2003) A general NMR method for rapid, efficient, and reliable biochemical screening. J Am Chem Soc 125:14620–14625
17. DeBerg HA, Brzovic PS, Flynn GE, Zagotta WN, Stoll S (2016) Structure and energetics of allosteric regulation of HCN2 Ion channels by cyclic nucleotides. J Biol Chem 291:371–381
18. Didenko T, Liu JJ, Horst R, Stevens RC, Wuthrich K (2013) Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. Curr Opin Struct Biol 23:740–747
19. Donald LJ, Crane BR, Anderson DH, Duckworth HW (1991) The role of cysteine 206 in allosteric inhibition of Escherichia coli citrate synthase. Studies by chemical modification, site-directed mutagenesis, and ^{19}F NMR. J Biol Chem 266:20709–20713
20. Dvoretsky A, Gaponenko V, Rosevear PR (2002) Derivation of structural restraints using a thiol-reactive chelator. FEBS Lett 528:189–192
21. Fetics SK, Guterres H, Kearney BM, Buhrman G, Ma B, Nussinov R, Mattos C (2015) Allosteric effects of the oncogenic RasQ61L mutant on Raf-RBD. Structure 23:505–516
22. Gaponenko V, Dvoretsky A, Walsby C, Hoffman BM, Rosevear PR (2000a) Calculation of z-coordinates and orientational restraints using a metal binding tag. Biochemistry 39:15217–15224
23. Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J, Hubbell WL, Rosevear PR (2000b) Protein global fold determination using site-directed spin and isotope labeling. Protein Sci 9:302–309
24. Ge X, MacRaild CA, Devine SM, Debono CO, Wang G, Scammells PJ, Scanlon MJ, Anders RF, Foley M, Norton RS (2014) Ligand-induced conformational change of Plasmodium falciparum AMA1 detected using ^{19}F NMR. J Med Chem 57:6419–6427
25. Gerig JT (1994) Fluorine NMR of proteins. Prog Nucl Magn Reson Spectrosc 26:293–370
26. Geyer M, Schweins T, Herrmann C, Prisner T, Wittinghofer A, Kalbitzer HR (1996) Conformational transitions in p21ras and in its complexes with the effector protein Raf-RBD and the GTPase activating protein GAP. Biochemistry 35:10308–10320
27. Gorfe AA (2010) Mechanisms of allostery and membrane attachment in Ras GTPases: implications for anti-cancer drug discovery. Curr Med Chem 17:1–9
28. Grutsch S, Bruschweiler S, Tollinger M (2016) NMR methods to study dynamic allostery. PLoS Comput Biol 12:e1004620

29. Hammoudeh DI, Date M, Yun MK, Zhang W, Boyd VA, Viacava Follis A, Griffith E, Lee RE, Bashford D, White SW (2014) Identification and characterization of an allosteric inhibitory site on dihydropteroate synthase. ACS Chem Biol 9:1294–1302
30. Han S, Czerwinski RM, Caspers NL, Limburg DC, Ding W, Wang H, Ohren JF, Rajamohan F, McLellan TJ, Unwalla R et al (2014) Selectively targeting an inactive conformation of interleukin-2-inducible T-cell kinase by allosteric inhibitors. Biochem J 460:211–222
31. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5:1065–1074
32. Hass MA, Ubbink M (2014) Structure determination of protein-protein complexes with long-range anisotropic paramagnetic NMR restraints. Curr Opin Struct Biol 24:45–53
33. Ho C, Lindstrom TR (1972) Functional non-equivalence of and hemes in human hemoglobins. Adv Exp Med Biol 28:65–76
34. Horng JC, Raleigh DP (2003) phi-Values beyond the ribosomally encoded amino acids: kinetic and thermodynamic consequences of incorporating trifluoromethyl amino acids in a globular protein. J Am Chem Soc 125:9286–9287
35. Horst R, Liu JJ, Stevens RC, Wuthrich K (2013) beta(2)-adrenergic receptor activation by agonists studied with (1)(9)F NMR spectroscopy. Angew Chem 52:10762–10765
36. Jahnke W, Grotzfeld RM, Pelle X, Strauss A, Fendrich G, Cowan-Jacob SW, Cotesta S, Fabbro D, Furet P, Mestan J, Marzinzik AL (2010a) Binding or bending: distinction of allosteric Abl kinase agonists from antagonists by an NMR-based conformational assay. J Am Chem Soc 132:7043–7048
37. Jahnke W, Rondeau JM, Cotesta S, Marzinzik A, Pelle X, Geiser M, Strauss A, Gotte M, Bitsch F, Hemmig R et al (2010b) Allosteric non-bisphosphonate FPPS inhibitors identified by fragment-based discovery. Nat Chem Biol 6:660–666
38. Jayalakshmi V, Krishna NR (2002) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of intermolecular saturation transfer effects in reversibly forming ligand-receptor complexes. J Magn Reson 155:106–118
39. Jayalakshmi V, Rama Krishna N (2004) CORCEMA refinement of the bound ligand conformation within the protein binding pocket in reversibly forming weak complexes using STD-NMR intensities. J Magn Reson 168:36–45
40. Jenkins BG, Lauffer RB (1990) Detection of site-specific binding and co-binding of ligands to human serum albumin using ^{19}F NMR. Mol Pharmacol 37:111–118
41. Kempf JG, Loria JP (2003) Protein dynamics from solution NMR: theory and applications. Cell Biochem Biophys 37:187–211
42. Kitevski-LeBlanc JL, Prosser RS (2012) Current applications of ^{19}F NMR to studies of protein structure and dynamics. Prog Nucl Magn Reson Spectrosc 62:1–33
43. Kitevski-LeBlanc JL, Evanics F, Prosser RS (2009) Approaches for the measurement of solvent exposure in proteins by ^{19}F NMR. J Biomol NMR 45:255–264
44. Kitevski-LeBlanc JL, Evanics F, Scott Prosser R (2010) Optimizing (1)(9)F NMR protein spectroscopy by fractional biosynthetic labeling. J Biomol NMR 48:113–121
45. Kitevski-Leblanc JL, Hoang J, Thach W, Larda ST, Prosser RS (2013) (1)(9)F NMR studies of a desolvated near-native protein folding intermediate. Biochemistry 52:5780–5789
46. Krimm I (2017) Identifying protein allosteric transitions for drug discovery with 1D NMR. ChemMedChem 12:901–904
47. Krishna NR, Moseley HNB (2002) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of NOESY spectra of reversibly forming ligand-receptor complexes application to transferred NOESY. In: Krishna NR, Berliner LJ (eds) Biological magnetic resonance: structure computation and dynamics in protein NMR. Boston, Springer US, pp 223–307
48. Leach K, Sexton PM, Christopoulos A (2011) Quantification of allosteric interactions at G protein-coupled receptors using radioligand binding assays. Curr Protocol Pharmacol Chapter 1(Unit 1):22

49. Lee AL, Kinnear SA, Wand AJ (2000) Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. Nat Struct Biol 7:72–77
50. Li H, Frieden C (2005) NMR studies of 4-^{19}F-phenylalanine-labeled intestinal fatty acid binding protein: evidence for conformational heterogeneity in the native state. Biochemistry 44:2369–2377
51. Li H, Frieden C (2006) Fluorine-19 NMR studies on the acid state of the intestinal fatty acid binding protein. Biochemistry 45:6272–6278
52. Lisi GP, Loria JP (2016) Solution NMR spectroscopy for the study of enzyme allostery. Chem Rev 116:6323–6369
53. Liu J, Nussinov R (2016) Allostery: an overview of its history, concepts, methods, and applications. PLoS Comput Biol 12:e1004966
54. Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012) Biased signaling pathways in beta2-adrenergic receptor characterized by ^{19}F-NMR. Science 335:1106–1110
55. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286:295–299
56. Loewen MC, Klein-Seetharaman J, Getmanova EV, Reeves PJ, Schwalbe H, Khorana HG (2001) Solution ^{19}F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. Proc Natl Acad Sci USA 98:4888–4892
57. Long D, Marshall CB, Bouvignies G, Mazhab-Jafari MT, Smith MJ, Ikura M, Kay LE (2013) A comparative CEST NMR study of slow conformational dynamics of small GTPases complexed with GTP and GTP analogues. Angew Chem 52:10771–10774
58. Loria JP, Rance M, Palmer AG (1999a) A relaxation-compensated Carr−Purcell−Meiboom−Gill sequence for characterizing chemical exchange by NMR spectroscopy. J Am Chem Soc 121:2331–2332
59. Loria JP, Rance M, Palmer AG 3rd (1999b) A TROSY CPMG sequence for characterizing chemical exchange in large proteins. J Biomol NMR 15:151–155
60. Lu S, Jang H, Muratcioglu S, Gursoy A, Keskin O, Nussinov R, Zhang J (2016) Ras conformational ensembles, allostery, and signaling. Chem Rev 116:6607–6665
61. Lukin JA, Kontaxis G, Simplaceanu V, Yuan Y, Bax A, Ho C (2003) Quaternary structure of hemoglobin in solution. Proc Natl Acad Sci USA 100:517–520
62. Macraild CA, Anders RF, Foley M, Norton RS (2011) Apical membrane antigen 1 as an antimalarial drug target. Curr Top Med Chem 11:2039–2047
63. Mallagaray A, Rademacher C, Parra F, Hansman G, Peters T (2017) Saturation transfer difference nuclear magnetic resonance titrations reveal complex multistep-binding of l-fucose to norovirus particles. Glycobiology 27:80–86
64. Martinez D, Gerig JT (2001) Intermolecular (1)H[(19)F] NOEs in studies of fluoroalcohol-induced conformations of peptides and proteins. J Magn Reson 152:269–275
65. Matsumoto S, Hiraga T, Hayashi Y, Yoshikawa Y, Tsuda C, Araki M, Neya M, Shima F, Kataoka T (2018) Molecular basis for allosteric inhibition of GTP-bound H-Ras protein by a small-molecule compound carrying a naphthalene ring. Biochemistry 57:5350–5358
66. Maurer T, Garrenton LS, Oh A, Pitts K, Anderson DJ, Skelton NJ, Fauber BP, Pan B, Malek S, Stokoe D et al (2012) Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. Proc Natl Acad Sci USA 109:5299–5304
67. Moseley HN, Curto EV, Krishna NR (1995) Complete relaxation and conformational exchange matrix (CORCEMA) analysis of NOESY spectra of interacting systems; two-dimensional transferred NOESY. J Magn Reson B 108:243–261
68. Moseley HN, Lee W, Arrowsmith CH, Krishna NR (1997) Quantitative determination of conformational, dynamic, and kinetic parameters of a ligand-protein/DNA complex from a complete relaxation and conformational exchange matrix analysis of intermolecular transferred NOESY. Biochemistry 36:5293–5299
69. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508:331–339

70. Nicely NI, Kosak J, de Serrano V, Mattos C (2004) Crystal structures of Ral-GppNHp and Ral-GDP reveal two binding sites that are also present in Ras and Rap. Structure 12:2025–2036
71. Nitsche C, Otting G (2018) NMR studies of ligand binding. Curr Opin Struct Biol 48:16–22
72. Nnadi CI, Jenkins ML, Gentile DR, Bateman LA, Zaidman D, Balius TE, Nomura DK, Burke JE, Shokat KM, London N (2018) Novel K-Ras G12C Switch-II covalent binders destabilize Ras and accelerate nucleotide exchange. J Chem Inf Model 58:464–471
73. Orton HW, Otting G (2018) Accurate electron-nucleus distances from paramagnetic relaxation enhancements. J Am Chem Soc 140:7688–7697
74. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM (2013) K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature 503:548–551
75. Oyen D, Wechselberger R, Srinivasan V, Steyaert J, Barlow JN (2013) Mechanistic analysis of allosteric and non-allosteric effects arising from nanobody binding to two epitopes of the dihydrofolate reductase of Escherichia coli. Biochim Biophys Acta 1834:2147–2157
76. Parker JA, Volmar AY, Pavlopoulos S, Mattos C (2018) K-Ras populates conformational states differently from its isoform H-Ras and oncogenic mutant K-RasG12D. Structure 26(810–820):e814
77. Peng JW (2001) Cross-correlated (19)F relaxation measurements for the study of fluorinated ligand-receptor interactions. J Magn Reson 153:32–47
78. Rama Krishna N, Jayalakshmi V (2006) Complete relaxation and conformational exchange matrix analysis of STD-NMR spectra of ligand–receptor complexes. Prog Nucl Magn Reson Spectrosc 49:1–25
79. Rogawski R, McDermott AE (2017) New NMR tools for protein structure and function: Spin tags for dynamic nuclear polarization solid state NMR. Arch Biochem Biophys 628:102–113
80. Roy M, Brader ML, Lee RW, Kaarsholm NC, Hansen JF, Dunn MF (1989) Spectroscopic signatures of the T to R conformational transition in the insulin hexamer. J Biol Chem 264:19081–19085
81. Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ (2018) Dynamic allostery can drive cold adaptation in enzymes. Nature 558:324–328
82. Seetaha S, Yagi-Utsumi M, Yamaguchi T, Ishii K, Hannongbua S, Choowongkomon K, Kato K (2016) Application of site-specific spin labeling for NMR detecting inhibitor-induced conformational change of HIV-1 reverse transcriptase. ChemMedChem 11:363–366
83. Selvaratnam R, Chowdhury S, VanSchouwen B, Melacini G (2011) Mapping allostery through the covariance analysis of NMR chemical shifts. Proc Natl Acad Sci USA 108:6133–6138
84. Shu Q, Frieden C (2005) Relation of enzyme activity to local/global stability of murine adenosine deaminase: ^{19}F NMR studies. J Mol Biol 345:599–610
85. Sijbesma E, Skora L, Leysen S, Brunsveld L, Koch U, Nussbaumer P, Jahnke W, Ottmann C (2017) Identification of two secondary ligand binding sites in 14-3-3 proteins using fragment screening. Biochemistry 56:3972–3982
86. Skora L, Jahnke W (2017) (19)F-NMR-based dual-site reporter assay for the discovery and distinction of catalytic and allosteric kinase inhibitors. ACS Med Chem Lett 8:632–635
87. Song L, Teng Q, Phillips RS, Brewer JM, Summers AO (2007) ^{19}F-NMR reveals metal and operator-induced allostery in MerR. J Mol Biol 371:79–92
88. Spencer-Smith R, Koide A, Zhou Y, Eguchi RR, Sha F, Gajwani P, Santana D, Gupta A, Jacobs M, Herrero-Garcia E et al (2017) Inhibition of RAS function through targeting an allosteric regulatory site. Nat Chem Biol 13:62–68
89. Spoerner M, Nuehs A, Herrmann C, Steiner G, Kalbitzer HR (2007) Slow conformational dynamics of the guanine nucleotide-binding protein Ras complexed with the GTP analogue GTPgammaS. FEBS J 274:1419–1433
90. Spoerner M, Hozsa C, Poetzl JA, Reiss K, Ganser P, Geyer M, Kalbitzer HR (2010) Conformational states of human rat sarcoma (Ras) protein complexed with its natural ligand GTP and their role for effector interaction and GTP hydrolysis. J Biol Chem 285:39768–39778

91. Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, Wingler LM, Ahn S, Chatterjee A, Masoudi A et al (2016) Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. Nature 535:448–452
92. Stockman BJ, Kothe M, Kohls D, Weibley L, Connolly BJ, Sheils AL, Cao Q, Cheng AC, Yang L, Kamath AV et al (2009) Identification of allosteric PIF-pocket ligands for PDK1 using NMR-based fragment screening and 1H-15N TROSY experiments. Chem Biol Drug Des 73:179–188
93. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10:59–69
94. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. Science 278:1111–1114
95. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1997) NMR evidence for slow collective motions in cyanometmyoglobin. Nat Struct Biol 4:292–297
96. Tsai CJ, Nussinov R (2014) A unified view of "how allostery works". PLoS Comput Biol 10: e1003394
97. Tzeng SR, Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. Curr Opin Struct Biol 21:62–67
98. von Geldern TW, Lai C, Gum RJ, Daly M, Sun C, Fry EH, Abad-Zapatero C (2006) Benzoxazole benzenesulfonamides are novel allosteric inhibitors of fructose-1,6-bisphosphatase with a distinct binding mode. Bioorg Med Chem Lett 16:1811–1815
99. Wang X, Mercier P, Letourneau PJ, Sykes BD (2005) Effects of Phe-to-Trp mutation and fluorotryptophan incorporation on the solution structure of cardiac troponin C, and analysis of its suitability as a potential probe for in situ NMR studies. Protein Sci 14:2447–2460
100. White JT, Li J, Grasso E, Wrabl JO, Hilser VJ (2018) Ensemble allosteric model: energetic frustration within the intrinsically disordered glucocorticoid receptor. Philos Trans R Soc Lond Ser B Biol Sci 373(1749):pii: 20170175
101. Yang D, Mok YK, Forman-Kay JD, Farrow NA, Kay LE (1997) Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation. J Mol Biol 272:790–804
102. Zhang W, Li R, Shin R, Wang Y, Padmalayam I, Zhai L, Krishna NR (2013) Identification of the binding site of an allosteric ligand using STD-NMR, docking, and CORCEMA-ST calculations. ChemMedChem 8:1629–1633
103. Zhuravlev PI, Papoian GA (2010) Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. Q Rev Biophys 43:295–332

Chapter 9
Interrogating Regulatory Mechanisms in Signaling Proteins by Allosteric Inhibitors and Activators: A Dynamic View Through the Lens of Residue Interaction Networks

Lindy Astl, Amanda Tse, and Gennady M. Verkhivker

Abstract Computational studies of allosteric interactions have witnessed a recent renaissance fueled by the growing interest in modeling of the complex molecular assemblies and biological networks. Allosteric interactions in protein structures allow for molecular communication in signal transduction networks. In this chapter, we discuss recent developments in understanding of allosteric mechanisms and interactions of protein systems, particularly in the context of structural, functional, and computational studies of allosteric inhibitors and activators. Computational and experimental approaches and advances in understanding allosteric regulatory mechanisms are reviewed to provide a systematic and critical view of the current progress in the development of allosteric modulators and highlight most challenging questions in the field. The abundance and diversity of genetic, structural, and biochemical data underlies the complexity of mechanisms by which targeted and personalized drugs can combat mutational profiles in protein kinases. Structural and computational studies of protein kinases have generated in recent decade significant insights that allowed leveraging knowledge about conformational diversity and allosteric regulation of protein kinases in the design and discovery of novel kinase drugs. We discuss recent developments in understanding multilayered allosteric regulatory machinery of protein kinases and provide a systematic view of the current state in

L. Astl · A. Tse
Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA

G. M. Verkhivker (✉)
Graduate Program in Computational and Data Sciences, Schmid College of Science and Technology, Chapman University, Orange, CA, USA

Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA, USA

Department of Pharmacology, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA
e-mail: verkhivk@chapman.edu

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_9

understanding molecular basis of allostery mediated by kinase inhibitors and activators. In conclusion, we highlight the current status and future prospects of computational biology approaches in bridging the basic science of protein kinases with the discovery of anticancer therapies.

Keywords Allosteric regulation · Siganling proteins · Allosteric modulators · Conformational dynamics · NMR Spectroscopy · Residue coevolution · Residue interaction networks · Protein kinases · Drug discovery

9.1 Introduction

Allosteric regulation is a common mechanism employed by proteins and complex biomolecular assemblies for regulating their activity and adaptability in the dynamic cellular environment during processes of signal transduction, catalysis, and gene regulation [31, 32, 109, 139, 158]. The initial framework and theoretical basis of allosteric regulation was proposed by Monod, Wyman, and Changeux (MWC model), which is often referred as the concerted model [139]. An alternative model was later developed by Koshland, Enmity, and Filmer and is known as the KNF model or sequential model [109]. These mechanistic qualitative models have been put forward to provide a plausible rationale of the allosteric regulation phenomena and explain cooperative binding by proteins, where binding event at the ligand site of one subunit can affect the protein response at another distant site, thereby altering ligand affinity to other protein subunits. The MWC model postulated that protein subunits undergo concerted conformational changes that occur by preserving symmetry as an important requirement of allosteric changes. According to this model, all subunits in each oligomer are in the R ("relaxed") or T ("tense") states where different conformational states of the allosteric protein coexist in a dynamic equilibrium and can be populated independently in the presence of the bound ligand. In the alternative KNF model, long-range allosteric coupling in proteins can occur through a mechanism of discrete and sequential conformational changes leading to global changes in the protein structure. In this model, ligand binding to one subunit triggers sequential conformational switching of the neighboring units to the higher affinity state. Computational and experimental studies of allosteric mechanisms have witnessed a recent renaissance that reshaped the conventional models of allostery, fueled by the growing interest in quantitative biology of complex biomolecular systems. The thermodynamics-based model of allosteric regulation has embraced the energy landscape perspective of protein structure and dynamics, according to which the statistical ensemble of preexisting conformational states and communication pathways can be explored and modulated by allosteric perturbations [46, 51, 126, 204, 205, 236]. New conceptual models such as conformational selection expanded our view of allostery beyond the induced fit by capitalizing on the free energy landscape framework where thermodynamic view of

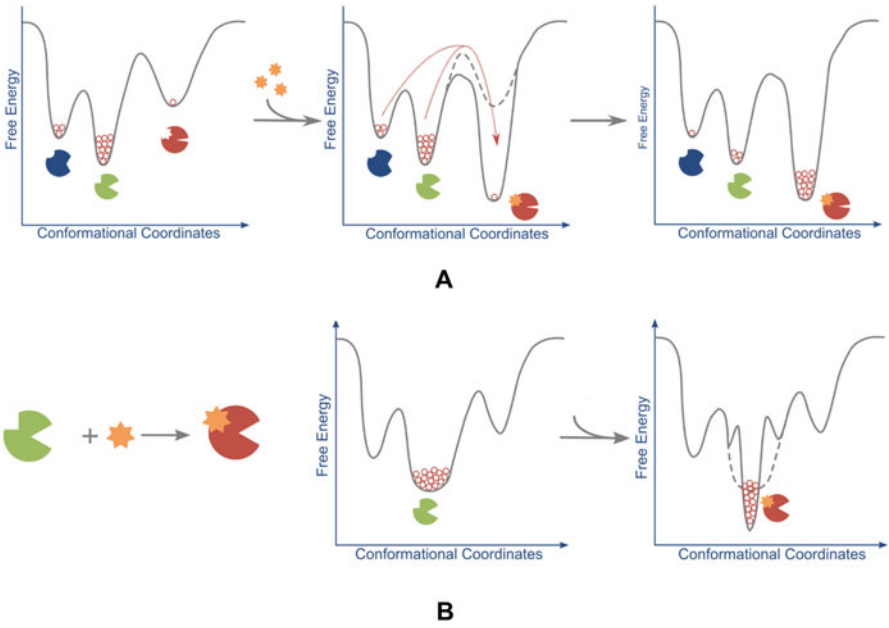


Fig. 9.1 (**a**) The conformational selection model. The conformations of the enzyme corresponding to local energy minimums are schematically shown. Red corresponds to the inactive enzyme, green to the active enzyme, and blue to another active-like conformation. The beads in the energy basins represent the relative population of each conformation. The energy surface prior to allosteric ligand binding is in a dashed line. (**b**) Dynamically driven allostery. The active enzyme (green) presents large fluctuations around the active site. These fluctuations are significantly reduced upon binding of an allosteric ligand rendering the inactive enzyme (red) without changing the mean structure of the enzyme. The surface prior to allosteric ligand binding is indicated by a dashed line

conformational selection is linked with the effector-induced equilibrium population shift [204, 206]. The conformational selection/population-shift model reformulated the original MWC model by assuming that proteins can sample ensembles of preexisting conformational states. In accordance with this model, allosteric ligand will bind weakly populated, high-energy conformations, altering the free energy landscape in this region and shifting the population toward this bound conformation (Fig. 9.1). The new view of protein allostery implies that ligand-induced redistribution of conformational fluctuations in regions that are linked by cooperative interactions may determine allosteric control and signaling. In this formulation, long-range interactions and allosteric changes are mediated not only by conformational transitions reflected in the enthalpy contribution but also by entropy-driven dynamic fluctuations around the average structure [77, 78, 158, 209, 210].

Protein kinases act as dynamic molecular switches in cellular signaling and their functional activity is essential for the integrity and viability of many signaling processes [4, 55, 106, 107, 135, 154, 178, 196, 198, 199]. The growing wealth of structural, genetic, and biochemical information in protein kinases has enhanced our

knowledge of the inhibitory mechanisms underlying drug–kinase binding and modulation of kinase activities. Recent advances in understanding allosteric regulation of protein kinases have facilitated efforts aiming in the discovery of allosteric kinase inhibitors that can provide target specificity and are at the forefront of the precision medicine initiative in oncology. We provide analysis of structural and network-based models of protein allostery and their applications to dissecting molecular determinants of allosteric inhibition and activation of human protein kinases.

9.2 Dynamic Allostery and Allosteric Communications in Proteins: Getting a Grip on Conformational Ensembles through Lens of NMR

While studies on allostery in protein systems are often focused on thermodynamic aspects of the mechanism, there is an increasing realization of the critical role of conformational dynamics, which is central to "entropy-driven" allostery (Fig. 9.1). This mechanism proposes that modulation of the thermal fluctuations around a mean structure rather than large structural shifts give rise to allostery in ligand binding [42]. Based on theoretical grounds, it was proposed that proteins may have evolved to take functional advantage not only of diverse conformational states but also of conformational dynamics about the mean structure, where effector-induced free energy changes are manifested exclusively in the modulation of the conformational distributions rather than structural transformations. Allosteric cooperative effects can be positive or negative, depending on whether the binding of the first ligand increases or decreases the affinity for subsequent ligands. While positive cooperativity is a fairly common mechanism for increasing the binding potential in allosteric systems [119], there are many examples of negative cooperativity in protein systems [48, 145, 188]. Negative cooperativity in proteins is mainly entropy-driven and occurs when successive binding to a regulatory site leads to a decrease in affinity and significant redistribution of conformational entropy. Dynamic allostery in proteins is often associated with negative cooperativity and significant entropy exchanges between local protein regions [6, 67, 134, 204, 206]. It was argued that fluctuation-induced allostery can arise in proteins that can be adequately described as inhomogeneous elastic entities in which domains are composed of rigid modules connected by more flexible interface regions [134]. The range of allosteric models employed to explain regulatory mechanisms in protein systems can vary from a sequential model, where binding of a molecule at one site causes a sequential propagation of conformational changes across the protein, to an intermediate, "block-based" model and a fully cooperative model, where structural changes are tightly coupled [104].

In the dynamics-driven allostery, effector ligands can induce redistribution of protein fluctuations in the regions linked by cooperative interactions and propagate signal through dynamic modulation of slow and fast functional motions located at

distant functional regions even in the absence of visible structural changes [99, 100, 158, 209, 210]. Recent time-resolved infrared spectroscopy experiments have indicated that the allosteric transitions occur on multiple timescales. By using non-equilibrium molecular dynamics (MD) simulations, a time-dependent picture of the allosteric communication revealed that allostery is manifested by hierarchical propagation of structural and dynamical changes suggesting a high degree of conformational heterogeneity of the ensemble of communication routes in proteins [26, 189]. The recent experimental breakthroughs in NMR technologies have enabled structural studies of large protein systems and conformational dynamic processes at atomic resolution that provide unique insights into allosteric mechanisms [89, 96, 101, 108, 165]. Relaxation dispersion NMR methods developed over the past decade have enabled detection and characterization of rare and energetically excited conformational states that play significant role in dynamic activation of protein function and allosteric mechanisms. The rapid conformational exchanges between ground and excited states occurring on the μs–ms timescale can now be detected by NMR and provide information about the kinetics and thermodynamics of the exchange process [101, 165]. Structural identification and characterization of very low populated states can allow for better understanding of allosteric activation mechanisms and enable atomic reconstruction of ligand-induced allosteric dynamic changes in large biomolecular systems. Characterization of low-lying excited states of proteins by high-pressure NMR under thermodynamic equilibrium conditions can allow for detection of reversible transitions that are functionally relevant, providing means for modulation of dynamic energy landscape sculpted to optimize functions via allosteric mechanisms [98, 219]. High-pressure NMR can help to identify these conformations, including low populated functional states, and characterize their energies and kinetics of conformational changes [219]. Pressure-dependent chemical shifts may also measure redistributions in conformational entropy and specify dynamic allosteric mechanisms, thereby providing the experimental platform for design of allosteric modulators that can sequester and stabilize otherwise low-populated functional states [98, 142, 219]. Allostery is a physical property of complex systems that can be characterized by dynamic networks of interactions between their components and integrated modules. The organization of these networks allows for formation of ensembles of pathways that transmit signals by propagating conformational fluctuations and functional motions between distant sites. The ensembles of allosteric communications and protein residues involved in signal transmission via population-shift or dynamics-based allostery can be experimentally examined by NMR spectroscopy [76, 184, 193]. NMR chemical shift responses to bound ligands are commonly employed as diagnostic tools for identifying coupled networks within allosteric proteins that could quantify potential communication pathways [122, 165, 180, 210]. The growing evidence that dynamics-driven allostery may be common to many protein systems has further expanded our view of diverse allosteric mechanisms that are not limited to population shift mediated exclusively through structural transitions that select and stabilize specific conformational states.

However, the main premise of the thermodynamics-based allosteric mechanism is shared by these models, postulating that allosteric proteins exist in different conformational states (often inactive and active type) and effector binding regulates protein responses and activity through modulation of the conformational equilibrium fluctuations. While allosteric ligand binding can alter distributions and energies of the different protein states, it typically has minor effect on conformations that are largely dictated by the intrinsic protein architecture. A rigorous thermodynamic model of allostery does not necessarily imply the existence of specific communication pathways as allosteric effects can propagate long distance through broad and diffusely distributed ensemble of heterogeneous conformational fluctuations [85, 103, 141, 220]. However, the ensemble-based model of allostery does not preclude thermodynamically weighted propagation of allosteric information through preferential preexisting pathways that can be modulated by ligand binding. This characterization can provide important insights into mechanisms and dynamics of signal transmission, identifying preferential mediators and carriers of dynamic fluctuations. Despite significant advances in computational and experimental studies of allosteric regulation mechanisms, molecular signatures of ensemble-based communication pathways and specific contributions of functional residues implicated in allosteric regulation are yet to be fully understood, rationalized, and reconciled with the population-shift and dynamics-driven allosteric mechanisms.

9.3 Molecular Mechanisms of Allosteric Inhibitors and Activators

Understanding molecular mechanisms underlying complexity of allosteric regulation in proteins has attracted growing attention partly due to considerable interest in the discovery of selective modulators of therapeutically important targets [149]. In the recent decade, drug discovery has been steadily shifting its focus toward targeting allosteric sites in order to improve compound selectivity [151, 194, 216]. An important factor behind these efforts is structural and evolutionary diversity of allosteric sites even among structurally similar proteins of the same family. Allosteric sites present a unique opportunity to selectively target one particular protein in a family of homologous proteins [53, 181]. Furthermore, allosteric binding sites tend to be under low evolutionary pressure and are often more dynamic than the active sites, facilitating and diversifying design of target-specific agents while addressing long-standing lingering problems of safety, toxicity, and side effects [149, 150]. Allosteric drugs also feature distinct physiochemical properties than their orthosteric counterparts, adding further freedom for the discovery of novel active compounds and can be often combined with orthosteric drugs into synergistic drug cocktails to modulate and improve enzyme activities, specificity, and pharmacological profiles. Analysis of the Ras cycle predicts at least eight drug-relevant functional states of the protein that can be used for an allosteric modulation of the

Ras activity, and some of these states were detected and localized in the activated Ras by high-pressure NMR techniques [98, 142]. These studies have suggested that allosteric inhibition based on targeting rare states revealed by NMR can open up new unexplored directions for the discovery of highly specific agents [98]. Targeted alterations of a dynamic conformational landscape may represent a feasible strategy for engineering allostery into protein scaffolds that lack intrinsic allosteric properties [38]. In particular, allostery can be dialed in by remodeling dynamic conformational ensembles through the introduction of short, flexible protein segments at the interface between two domains of non-switch fusion proteins. By exerting differential effects on the dynamic fluctuations in the presence and absence of effector, the flexible interdomain connector could confer allosteric switching between the inactive and active forms of the fusion protein [38]. Allostery can be also associated with the introduction of intrinsic disorder [11, 64, 68, 217]. While traditional orthosteric drugs usually bind to the active site and inhibit protein activity, allosteric modulators may not only inhibit but also increase protein activity. Chemical biology approaches have identified a number of allosteric activators for several natural enzyme targets, including glucokinase (GCK), AMP-activated protein kinase (AMPK), p300 histone acetyltransferase, RNase L, and the sirtuin family of NAD^+-dependent protein deacetylases [17].

Allosteric activators may exploit several general mechanisms: (a) binding to an allosteric site on the catalytic domain to promote an active conformation; (b) binding to an allosteric site to facilitate activating posttranslational modification; and (c) binding to a regulatory subunit to indirectly promote activity at the catalytic domain or to promote an activating oligomerization [237]. Direct binding to the catalytic domain of a dormant enzyme and induction of a population shift that stabilizes the active conformation is often a prevalent mechanism of allosteric activation [237]. This mechanism has been observed for GCK, SIRT1, PDK1, PP1, and p300 proteins. Human glucokinase (GCK) has emerged as a model system for understanding allostery in monomeric, single-site enzymes. Even in this simplest example, there exist two functionally distinct mechanisms of GCK activation that can elicit the allosteric response of GCK to glucose. While the first activation mechanism involves a population shift and a notable structural change to a closed active structure, limited proteolysis and NMR have also revealed that activation can be achieved by modulating the dynamic properties of an active site loop without perturbing the ensemble average structure, thereby presenting another example of dynamics-driven allostery [218]. Large protein assemblies such as 20S proteasome core particle (CP) can be controlled by activator complexes that bind 75 Å away from sites catalyzing proteolysis, revealing a population-shift-based allosteric mechanism by which proteasome interconverts between multiple conformations whose relative populations are shifted on binding of the activator or mutation of residues that contact activators [172]. Allosteric modulators that increase or decrease the activity of the orthosteric agonists by interacting with the residues that mediate allosteric communication between the orthosteric and the G-protein site emerge as new strategies in drug design for GPCRs [13, 19]. Analysis of differences in ensembles of communication pathways in the inactive and active states of GPCR

has shown several well-defined signaling routes, while a more dispersed pattern of delocalized pathway ensembles in the active form [13]. Allosteric nanobodies probed diverse mechanisms of GPCR activation where large ligand-specific effects are explained using an allosteric model of dynamic equilibrium among at least three receptor states [186]. While agonists exert efficacy by stabilizing the active antibody-stabilized receptor state, partial agonists incur both stabilization of the active state and destabilization of the inactive state to modulate receptor activation. These pioneering studies have concluded that allosteric GPCR agonists can elicit a range of cellular responses by differentially stabilizing multiple receptor states. NMR analysis of the activator complexes with the glutamine aminotransferase imidazole glycerol phosphate synthase (IGPS) has shown absence of ligand-induced structural changes in the IGPS enzyme, revealing clear signs of fluctuations-driven allosteric mechanism [124]. The allosteric network of IGPS appeared to be widely dispersed and illuminated presence of effector-specific pathways that may be essential for propagation of the allosteric signal. A broad and effector-dependent nature of allosteric network in the IGPS enzyme is manifested in the variable number and location of dynamically stimulated residues elicited by specific activators [124]. Interestingly, the strongly activating allosteric ligands perturb the chemical shifts of fewer residues than weak activators, and yet the allosteric pathways inferred from chemical shift data differ with each effector. These studies have also found that allosteric activator-induced changes in millisecond motions are responsible for enhancing the catalytic rate at a distant active site [121, 123, 124]. Despite several successful examples, the concept of allosteric activators has not yet obtained a widespread attention, which may reflect the relatively small number of structural and functional studies and lack of systematic validated strategy to rationally design enzyme activators. It has been noted that the repertoire of enzyme agonists and antagonists can often overlap and show how small changes may lead to switching from inhibition to activation [49]. The improved understanding of similarities and differences between these classes of allosteric modulators could offer new opportunities for probing and interrogating biological processes and pathways.

9.4 Structural and Network-Based Models of Allosteric Interactions in Proteins

The statistical ensemble-based models of protein allostery have been successfully used to model the effects of allosteric interactions in individual proteins, analysis of modulations on signaling pathways, cellular functions, and disease states [45, 74, 147–149, 152, 153]. Most recently, the modern statistical view of protein allostery has been further expanded to study disordered proteins [47], structure and dynamics of molecular networks [45], and the mechanisms of allosteric protein inhibition in signaling networks [194]. Elastic network models of protein dynamics [6, 7, 231] and the normal mode analysis [8, 127] have been integrated with the information-based theory of signal propagation [35, 36] in the development of a generalized

formalism of allosteric communication pathways in proteins [6, 37]. Among recently advocated approaches to explain communication of the allosteric signals is to consider protein representation as a network of residues interconnected via noncovalent interactions [24, 25, 52, 205, 213]. In this approach, allosteric effector can rewire part of the network to efficiently communicate allosteric information over long range through structural changes or dynamics redistributions. This model has led to the development of several methods that reproduced the atomistic mechanisms associated with transmission of ligand-induced allosteric signals [62]. The integration of network theory and evolution analysis with MD simulations and normal mode analysis (NMA) has emerged as a powerful and increasingly popular approach to study allosteric regulation and quantifying communication pathways in complex molecular machines [13, 15, 16, 18, 52, 53, 62, 69–71, 80, 176, 177, 191].

The central concept underlying the network approach in modeling of allosteric interactions and pathways is that effector binding can result in a dynamically propagating cascade of coupled residue fluctuations throughout the protein. By mapping these dynamic fluctuations and correlated motions onto a graph with nodes representing residues and edges representing weights of measured dynamic properties, one could determine global topological features and ligand-induced organization of the allosteric interaction network, identify key functional centers, and characterize ensembles of allosteric communication pathways in the system. The network-based approaches have offered a simple, robust, and conceptually attractive perspective, providing better understanding of protein structure and enabling local and global effects to be seamlessly incorporated into computationally transparent models. The network studies have also suggested that rapid transmission of allosteric interactions through small-world networks encoded in protein folds may be a universal requirement encoded across diverse protein families [18, 52, 205].

Functional sites that mediate communication pathways and determine organization of the residue interaction networks often correspond to dynamically coupled and coevolving residues. Statistical coupling analysis (SCA), mutual information (MI) model, and covariance-based approaches have employed sequence-based analysis of residue coevolution in homologous families to show that functional residues in residue networks are connected via strong coevolutionary relationships [2, 50, 79, 125, 128, 133, 183, 192]. Coevolution of protein residues can reflect a coordinated involvement of these sites in mediating residue–residue contacts [185], promoting protein folding [140], and facilitating allosteric signaling in multi-protein complexes [214]. Coevolving residues tend to be spatially coupled and coincide with functional positions exhibiting correlated and compensatory mutations in homologous proteins [232]. These residues could also form direct communication paths in the interaction networks with connections weighted according to dynamic couplings and coevolutionary interaction strengths between nodes [29, 30, 115, 146]. Furthermore, coevolving residues can assemble into structurally stable and quasi-independent modules of physically interacting residues termed "protein sectors" [79, 133]. Several computational methods have been developed to evaluate the extent of mutual information (MI) and coevolutionary dependencies between residue pairs [72, 128, 129, 183, 202]. In the MISTIC approach [128, 183], the residue-based mutual information score characterizes the extent of mutual information shared by a given

residue with all other protein residues across protein family. Based on this information, MISTIC tool defines cumulative mutual information (cMI) and proximity mutual information (pMI) residue scores. cMI score is a sequence-based parameter that measures the degree of shared mutual information of a given residue with the other protein residues. pMI score encapsulates both sequence and structural variations being defined as the average of cMI scores of all the residues within a predefined distance from a given residue in the protein structure [128, 183]. Coevolutionary analysis using MISTIC approach was integrated into dynamics-based network modeling of the residue interactions [187]. In this model, the network edges (interactions) are weighted based on both dynamic and coevolutionary residue correlations that determine the shortest communication paths between residue nodes. Residue centrality (residue betweenness) is a global network parameter that was computed to determine highly connected nodes in a global interaction network. A propensity of protein residues to serve as global mediating centers of allosteric interaction networks in this approach is evaluated by considering common peaks in the distributions of residue centrality and pMI profiles.

The important revelation of these studies was that dynamic and coevolutionary residue correlations may act as synchronizing forces to enable efficient and robust allosteric regulation, where key sites mediating allosteric signaling are often aligned with high centrality and high pMI residues. These specific characteristics may be necessary for the regulation of allosteric structural transitions and could distinguish regulatory sites from nonfunctional conserved residues. The observed confluence of dynamics correlations and coevolutionary residue couplings with global networking features may also determine modular organization of allosteric interaction networks [187]. Computational analysis of residue interaction networks has revealed that independent modules of protein residues anchored around functional sites can work cooperatively to mediate structural stability and conformational transitions required for diverse functions in the dynamic protein environment [88, 95, 128, 187, 201, 229, 235]. These results have motivated the development of novel community-hoping methods for modeling ensembles of allosteric communication pathways in protein structures [187, 208, 211]. In this model, cooperative transitions may occur between local modules (communities) of tightly coupled interacting residues that tend to switch their conformational states cooperatively and form a weakly coupled assembly acting as a communication pathway in signal transmission. The body of computational studies on allosteric mechanisms has indicated that integration of network-based approaches with NMR studies of protein dynamics and conformational ensembles may provide a robust platform for further exploration and atomistic characterization of allosteric states and regulatory mechanisms controlled by allostery. This has motivated the development and proliferation of various computational models and tools for predicting allosteric sites and pathways of the intramolecular signal propagation and communications ([22, 73, 102, 132, 136–138, 157, 187, 200]).

9.5 Data-Driven Modelling of Residue Interaction Networks: Network Signatures of Allosteric Inhibitors and Activators Inferred from Allosteric Database

A rapid growth of biochemical and structural information on allosteric proteins and allosteric modulators has benefited bioinformatics studies and development of various web-based database resources. Allosteric Database (ASD) initiated several years ago has received significant praise of computational community, providing a centralized resource for the display, search, and analysis of the structure and function of allosteric proteins [90, 91, 179]. Currently, ASD has catalogued and annotated allosteric proteins and modulators in three categories (activators, inhibitors, and regulators). ASDBench is an open-source subset of the ASD database that includes proteins from various species, primarily bacteria contributing 44% and human proteins amounting to 26% of the dataset. ASDBench is subdivided into Core set, which is composed of 235 unique allosteric protein entries and Core-Diversity set that contains 147 allosteric proteins with the additional filter of redundancy and the quality of the binding site. By performing large-scale analysis of residue interaction networks and coevolutionary residue couplings in allosteric proteins of the Core-Diversity set, we have determined the distributions of residue centralities and coevolutionary pMI parameters (Fig. 9.2). In this analysis, we specifically focused on a comparative analysis of complexes with allosteric inhibitors and allosteric activators, aiming to identify distinguishing molecular and network signatures of these two classes of allosteric modulators. Our analysis confirmed that protein structure topologies in allosteric proteins could produce small-world networks. Indeed, the residue centrality and pMI distributions revealed the characteristic small-world signature of a sharp decay and a longer tail corresponding to residues with very high centrality (high pMI) values (Fig. 9.2). Hence, allosteric inhibitors may often induce a significant network-bridging effect and produce large allosteric networks with the enhanced centrality of mediating residues that stabilize structural environment favored by the inhibitor-bound conformations. This may also reflect a potential enrichment of population-shift allosteric mechanism in complexes with the allosteric inhibitors. In this case, inhibitor binding can be often accompanied by the equilibrium shift to the specific state, with many spatially distributed high centrality centers in the allosteric network that links the allosteric and active sites. Our previous network studies of protein kinase complexes with ATP-competitive inhibitors related ligand-induced changes in the residue interaction networks with drug resistance effects, showing that network robustness may be compromised by targeted mutations of key mediating residues [207]. According to results of that study, the severity of drug resistance effects to inhibitor binding may be directly associated with the high network centrality and mediating function of mutation-targeted residues.

In some contrast, the centrality distribution of complexes with allosteric activators pointed to a noticeably shorter tail, indicating a shortage of high centrality sites and instead featuring many residues with moderate centrality (Fig. 9.2). These results may reflect the

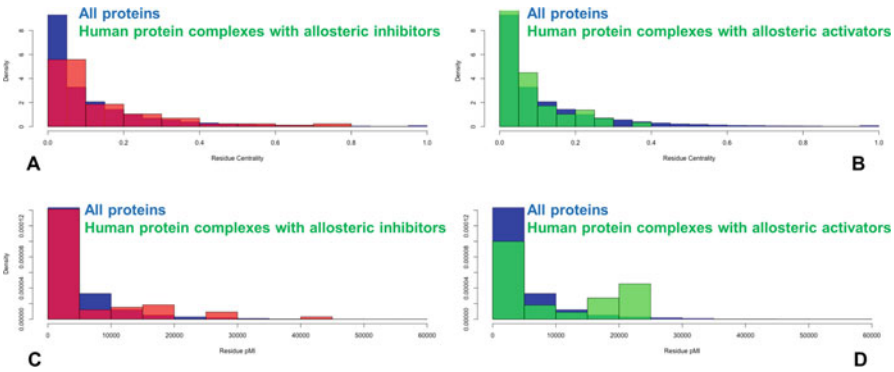


Fig. 9.2 (**a**) The residue centrality distribution for allosteric proteins from the Core-Diversity set. The distribution for all proteins is shown in blue and the distribution for allosteric inhibitors in human proteins is superimposed and shown in green. (**b**) The residue centrality distribution for all allosteric proteins is superimposed with the centrality profile for allosteric activators in human proteins (in green). (**c**) The pMI distribution for all allosteric proteins (in blue) and for allosteric inhibitors in human proteins (in green). (**d**) The pMI distribution for allosteric activators (in green) is superimposed on the overall pMI density

fact that an appreciable fraction of allosteric activators may operate under dynamically driven allosteric mechanisms when the protein does not significantly change in structure, giving rise to more dispersed and diffusely distributed ensembles of communication pathways with the fewer obligatory mediating hubs. In this scenario, the location and density of dynamically stimulated residues elicited by specific activators for a given protein may significantly differ [124]. It has been noted that the number of crystal structures of enzymes bound to allosteric activators is small as compared to the diversity of structures complexed with the allosteric inhibitors [237]. One of the potential reasons may be associated with the dominant allosteric mechanism as inhibitors usually stabilize and rigidify a specific inactive conformation, whereas activators could often modulate and stimulate protein dynamics making crystallographic determination of these complexes more challenging. Obviously, these observations do not imply that allosteric inhibitors and activators can be biased to any specific mechanism as allosteric modulators can exploit variations and synergistic combinations of different mechanisms including induced-fit, population-shift, and entropy-driven allostery. Our data argue that allosteric inhibitors may often leverage network efficiency by promoting signal transmission via several distinct "narrow tubes" of paths formed around specific routes of predominantly rigid high centrality residues. We have observed that allosteric activators may often promote adaptation of a moderate population-shift or dynamics-driven allosteric mechanism, where signaling routes are more diffusely distributed through many dynamic residues connecting distal sites. In this scenario, allosteric activators may exploit and functional

redundancy to ensure robust protein activation that would sustain random mutations. It was shown that dynamic allostery can better tolerate individual mutations due to many independent pathways as disabling one pathway by mutation could trigger compensation by another pathway [215].

Our previous studies of molecular chaperones [18, 187] and protein kinases [207, 208] have suggested that allosteric interactions may be regulated through a mechanism that combines efficient communication via specific pathway of predominantly rigid residues [160] and more robust signal transmission via an ensemble of multiple communication routes in which nonredundant rigid centers are coupled with more flexible residues [236]. The diversity of allosteric communication mechanisms could ensure a proper balance of the network efficiency and functional redundancy required to maintain resilience against random attacks in the fluctuating protein environment.

9.6 Allosteric Kinase Inhibitors

Among therapeutically important protein families, the human protein kinases play a special role by regulating functional processes in signal transduction networks and acting as dynamic molecular switches in cellular signaling [196, 198]. Recent studies of the structure and dynamic regulation of the SRC kinase [20, 57, 81, 166, 167, 168], ABL kinase [81, 156, 162], and ErbB kinases [21, 63, 98, 116, 117] have provided compelling evidence that protein kinase activity can be regulated via a dynamic equilibrium between the inactive and active states. Protein kinase activation and regulation are orchestrated by cooperative conformational changes of the conserved HRD motif in the catalytic loop and the DFG motif coupled with the αC-helix. These regulatory elements of the catalytic domain cooperate to form conserved intramolecular networks termed regulatory spine (R-spine) and catalytic spine (C-spine) whose assembly and stabilization are associated with the conformational transformations and kinase activation [196, 198]. A diverse repertoire of crystallographic conformations has also indicated that molecular mechanism of protein kinases may not necessarily imply an on–off binary switch between static inactive and active states but rather represents a dynamic multilayered regulatory sensor in which binding and external perturbations give rise to a continuum spectrum of inactive and active conformations exhibiting a range of activity levels. The wealth of structural knowledge about conformational states of the kinase catalytic domain, regulatory assemblies, and complexes with inhibitors has dramatically advanced out understanding of these allosterically regulated molecular switches, facilitating the development of selective and multi-targeted kinase inhibitors that play a dominant role in cancer research [58, 167–169, 221, 222].

On the basis of the molecular mechanism of action and crystallographic binding mode, several classes of kinase inhibitors have been determined [167–169, 234]. Type I inhibitors target the catalytically competent, active (DFG-in) conformation of the kinase domain with the αC-helix-in conformation and A-loop in the open ("out") active conformation. Type II inhibitors recognize the inactive DFG-out kinase conformation with the αC-helix-in conformation and A-loop in the closed ("in") inactive arrangement. Type III allosteric inhibitors do not compete with ATP and could be more selective than types I and II inhibitors by binding to the regulatory sites outside of the ATP binding site. An unintended consequence of kinase drug discovery has been the discovery of ATP-competitive inhibitors that behave as agonists of their kinase targets [49]. These studies have indicated a connection between ATP binding site occupancy and interaction networks of communications that can lead to the enhancement of catalytic activity. Allosteric kinase inhibitors can be further subdivided into two subclasses: inhibitors that bind in an adjacent allosteric site that does not overlap with the ATP binding pocket and inhibitors which bind to an allosteric site that is distant from the ATP binding pocket. Targeting allosteric pockets of kinases outside the highly conversed ATP pocket has been proposed as a promising alternative to overcome current barriers of kinase inhibitors, including poor selectivity and emergence of drug resistance. Structural diversity of allosteric regulatory mechanisms in protein kinases has been leveraged in the discovery of small-molecule allosteric inhibitors [43, 44, 49, 59, 212]. Allosteric compounds can often confer selectivity, allow access to novel and diverse chemical space beyond ATP-based scaffolds, and allow for better modulation of physicochemical properties. Although drug resistance may frequently arise, combination therapies of ATP-competitive and allosteric inhibitors can become a fruitful alternative to conventional design approaches.

The serine/threonine MEK kinases are among the most investigated targets for which allosteric inhibitors have been developed [61]. A number of compounds that inhibit MEK-1 and MEK-2 activity through an allosteric mechanism have been co-crystallized in complexes with the catalytic domain. These allosteric inhibitors can bind and stabilize a naturally occurring inactive form of MEK kinases and bind to the "back pocket" away from the ATP binding site [65, 163]. Structural study has identified that a potent and selective MEK allosteric inhibitor TAK-733 binds to the MEK1–ATP complex in the back cleft adjacent to the ATP-binding pocket [54]. The discovery of novel carboxamide-based allosteric inhibitor XL518 (GDC-0973) (pdb id 4LMN) (Fig. 9.3) and related analogs has produced further insights into mechanisms of allosteric MEK inhibitors [163]. Allosteric MEK1/2 inhibitors usually constrain the movement of the activation loop and thus decrease the rate of Raf-mediated MEK phosphorylation and lock the kinase in the catalytically inactive state. Hence, this class of allosteric inhibitors would likely operate under population-shift mechanism by biasing the dynamic equilibrium toward specific inactive conformation. AKT inhibitors can allosterically bind to both the active and inactive forms of the enzyme, and require presence of the pleckstrin-homology (PH) domain [10, 120]. The crystal structure of human AKT1 containing both the PH and kinase domains with a selective allosteric inhibitor bound in the interface was recently

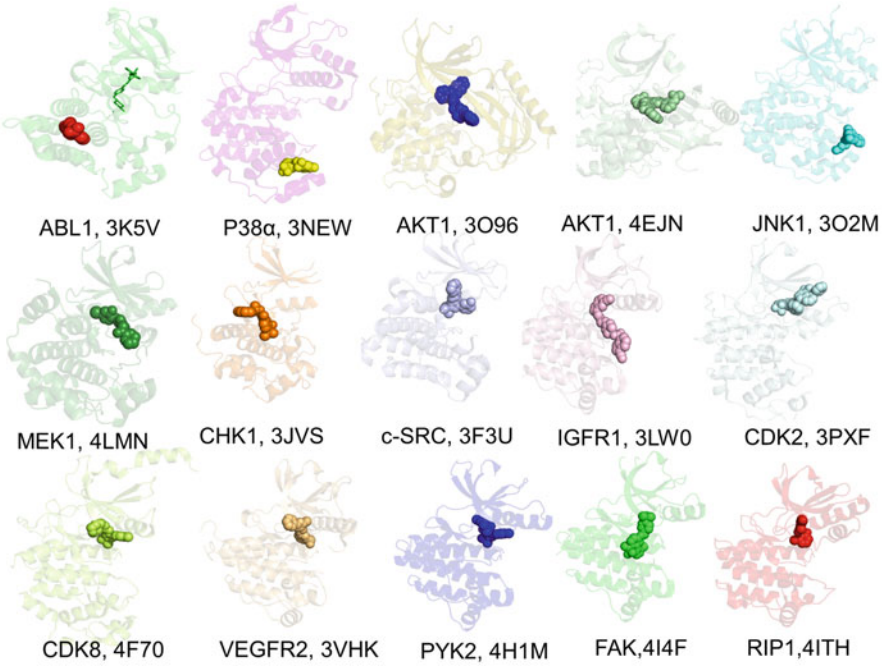


Fig. 9.3 Crystal structures of allosteric kinase inhibitors. The bound inhibitors are shown as spheres; the catalytic domain is shown in ribbons with a reduced transparency to highlight different positions of inhibitors. The crystal structures of kinases are annotated and corresponding PDB ID of the kinase complexes with allosteric inhibitors are also shown

determined (pdb id 3O96) (Fig. 9.3) [5]. This study has shown that the PH domain can lock the kinase in an inactive conformation, while the kinase domain disrupts the phospholipid binding site of the PH domain. The co-crystal structure of another AKT allosteric inhibitor occupies a cavity at the interface between the kinase domain and the PH domain (PDB entry 4EJN) (Fig. 9.3) [223]. While ATP-competitive AKT inhibitors bind to the activated state, 'PH-out' conformation, allosteric AKT inhibitors induce recruitment of the PH domain and stabilize the unique 'PH-in' conformation of the inactive AKT, thus inhibiting activation and phosphorylation of the enzyme [60, 131].

These studies have provided support to the conformational selection allosteric mechanism of AKT inhibitors that induce equilibrium shift toward the unique inactive form of the kinase. Allosteric kinase inhibitors can bind near the A-loop and MAPK insert region for JNK1 kinase (pdb id 3O2M) [40], PIF pocket of PD1 (pdb id 4AW1) [28], C-terminal lobe of CHK1 (pdb id 3JVS), DFG-out pocket of SRC kinase (pdb id 3F3U) ([182]), DFG-out pocket and A-loop of IGF1R (pdb id 3LW0) [84], DFG-out pocket of CDK8 (pdb id 4F70) [174] and PYK2 (pdb id 4H1M) [14], catalytic loop and DFG-out pocket of FAK (pdb id 4I4F) [203], and DFG-out pocket of RIP1 (pdb id 4ITJ) [227] (Fig. 9.3). The allosteric pockets

utilized by these inhibitors can involve structural rearrangements of the DFG motif and the activation loop, ordering or displacement of the αC-helix, and modulation of conformational dynamics in the functional regions.

To illustrate the nature of residue interaction networks and communication pathways, we selected several representative kinase complexes with allosteric inhibitors that feature a population-shift mechanism and are characterized by a number of high centrality and high pMI residues. By conducting atomistic simulations of the JNK kinase complex with the allosteric inhibitor (pdb id 3O2M), we constructed dynamic interaction network for this system, identified high centrality sites, and reconstructed ensembles of allosteric pathways connecting the allosteric and active sites (Fig. 9.4). The employed computational protocol is consistent with our previous network-based studies of allosteric regulation [187, 207, 208]. In this analysis, we focused on modeling of communication networks that link the peptide-binding with the catalytic site [113]. The allosteric compound binds on a ledge in JNK kinase provided by the MAP insertion on the surface of the protein, and this site is accessible only in the inactive form of the protein (Fig. 9.4). Allosteric inhibitor operates under population-shift mechanism by favoring the inactive JNK1 conformation characterized by rotational motions of N- and C-lobes and destabilization of the activation loop conformation [113]. The regulatory actions in JNK family kinases is exerted by the opposing actions of peptide and activation loop phosphorylation providing a control mechanism over the catalytic site, which lies between the lobes. Structural map of dominant communication pathways mediated by high centrality residues showed clusters near the peptide binding site and in the activation loop near the allosteric binding site that can be essential for carrying out the allosteric response. The communication map showed that allosteric binding in the lower lobe of the inactive JNK1 can affect signal transmission between the ATP site and the peptide binding site by rewiring preferential routes to link the allosteric and peptide sites instead. In this case, allosteric inhibitor may enhance network efficiency and connectivity of the inactive form by forming specific routes of high centrality residues linking the allosteric site with the rest of the protein. The inhibition mechanism via binding to the allosteric site in the inactive form can prevent phosphorylation of the activation loop and kinase activation. Our analysis also reveals a pocket crosstalk network, which passes through a set of small pockets that can be targeted by allosteric kinase inhibitors [110].

Crystal structures of the fluorophore 8-anilino-1-naphthalene sulfonate compound (ANS) bound with human CDK2 revealed two ANS molecules bound adjacent to one another in a large allosteric pocket that extends from the DFG region above the αC-helix [12] This pocket is formed by the αC-helix and the β-strands β3, β4, and β5 of the N-lobe. These small molecules act as allosteric inhibitors of CDK2, eliciting structural changes in the αC-helix and the β3, β4, and β5 strands thereby intervening with CDK2 association with cyclin A. The second ANS molecule binds adjacent to the primary site and also makes interactions with the αC-helix residues. However, the tight interaction of the CDK–cyclin complex requires allosteric inhibitors to be exceptionally potent as relatively moderate binding of ANS can be displaced from CDK2 upon cyclin A association [12, 130].

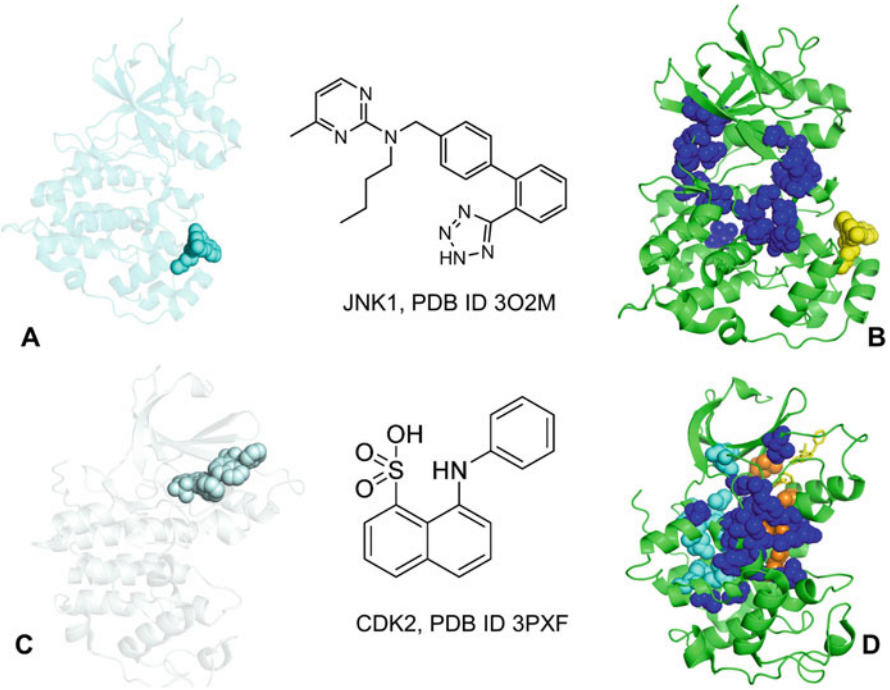


Fig. 9.4 The dynamic network analysis and structural mapping of the thermodynamically favorable communication routes from the ensemble of communication pathways. (**a**) The schematic view of JNK1 complex with the allosteric inhibitor (pdb id 3O2M). The inhibitor is shown in spheres; the protein structure is in ribbons with the reduced transparency. (**b**) The favorable communication paths are shown in blue spheres; the JNK structure is in green ribbons. The allosteric inhibitor is shown in yellow spheres. (**c**) The schematic view of CDK2 complex with the allosteric inhibitors (pdb id 3PXF). The inhibitors are shown in spheres; the protein structure is in ribbons with the reduced transparency. (**d**) The favorable communication paths are shown in blue spheres; the CDK2 structure is in green ribbons. The allosteric inhibitor is shown in yellow spheres. The R-spine residues (orange spheres) and C-spine residues (cyan spheres) highlight overlay and complementarity of these subnetworks with the preferential paths

Network-based modeling of communication pathways in the CDK2-inhibitor complex showed that these routes are mediated by high centrality residues that overlap and complement the R-spine and C-spine networks (Fig. 9.4). The localization of high centrality sites and the pathway map may reflect the underlying mechanism of allosteric inhibitor binding in CDK2 that is based on strengthening the unproductive spine arrangement that stabilizes the inactive CDK2 conformation. Our results suggested that structural mimicry of the R-spine residues by allosteric inhibitors can create opportunity for design of other compounds. Based on the importance of high centrality and high pMI residues, we suggest that further improvement of binding for ANS-derived compounds could be achieved by combination of complementary electrostatic (Lys33) and hydrophobic interactions with some important for allostery sites, including Leu55, Phe146, Tyr15, Phe80, Leu78,

Leu148, and Leu66. Using virtual screening campaign, the first-in-class type III allosteric ligands based on ANS scaffold was discovered [161]. These compounds have achieved the improved potency by utilizing interactions with key networking residues (Leu55, Phe146, Tyr15, Phe80, Leu78, Leu148, Leu66, Ile52, Leu76, Leu78, Leu37, Val69, and Ile35). The structural features of allosteric inhibitors binding with the different chemical scaffold were further explored using mutagenesis and biochemical studies [39] confirming binding to the allosteric pocket and structural mimicry of ANS interactions with the R-spine residues. Noteworthy, ligand-induced modulation of communication networks in JNK1 and CDK2 structures revealed presence of ligand-specific pathways that reflect the inhibitory mechanisms and stabilization of unique inactive kinase forms. In case of JNK1, inhibitor binding can modulate communications by blocking efficient signaling between the ATP and peptide binding sites (Fig. 9.4). Allosteric inhibitors of CDK2 kinase induce several communication lines that support and stabilize the inactive spine arrangements (Fig. 9.4).

Our previous studies of kinase interaction networks in the ABL kinase [207] and EGFR [94] presented a mechanistic model of allosteric coupling between the ATP-binding and substrate binding sites that may be conserved among kinases. In these studies, we have demonstrated that the θ-like shape of the interaction network can provide two lines of communication that connect the ATP binding site and the substrate binding regions. One side of the network could be traced from the active site to the αD-helix through the C-spine residues to the integral αF-helix and the substrate binding region. Another line of communication linked the ATP binding site with the αC-helix and via the R-spine residues to the substrate region. The θ-shaped residue network is rested on allosteric coupling between two lines of communications that are controlled by the regulatory DFG and HRD regions [207]. We observed that high centrality and highly coevolving kinase residues often coincide and correspond to the conserved functional sites mediating stability and allosteric communications in the kinase domain. These residues included the gatekeeper residue from the ATP site, HRD motif of the catalytic loop, and DFG motif and substrate binding motif in the C-lobe. This general topography of the residue interaction networks in kinases and conservation of mediating centers was also noted in a recent study of allosteric networks in SRC kinase [66], suggesting that the conserved allosteric network may define ensemble of preexisting communication pathways modulated by allosteric perturbations and ligand binding. Our data indicated that these allosteric networks and pathways can be dynamically remodeled and partly rewired based on the adopted allosteric mechanism of kinase inhibition (Fig. 9.5).

9.7 Allosteric Kinase Activators

The repertoire of kinase activators that can be utilized for interrogating and stimulation of biological pathways may be beneficial in some pathologies and could lead to new drug candidates as well as expanded tool kit for probing mechanisms of kinase actions. While the field of kinase inhibitors has enjoyed unprecedented

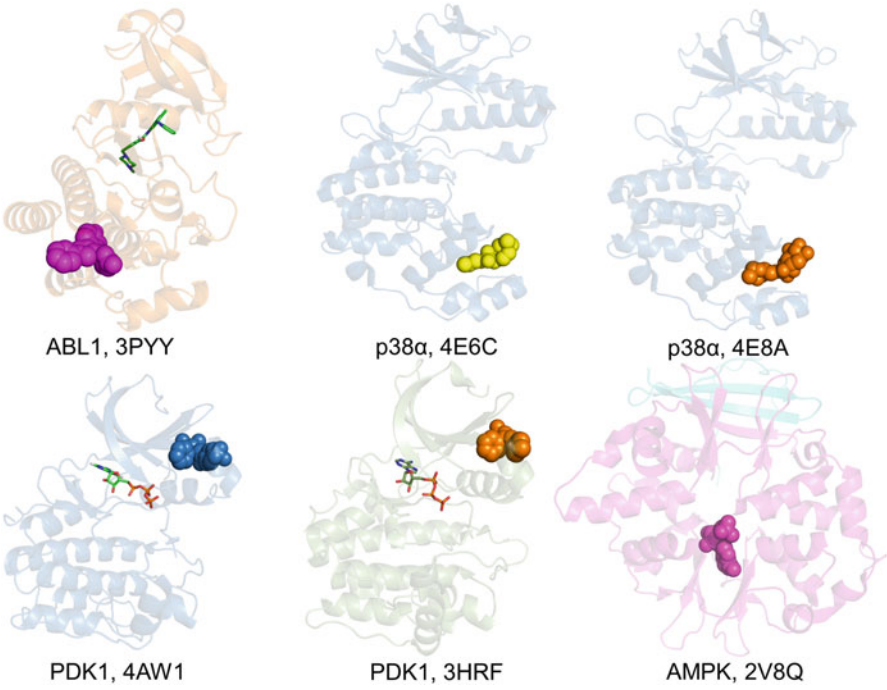


Fig. 9.5 Crystal structures of allosteric kinase activators. The bound inhibitors are shown as spheres; the catalytic domain is shown in ribbons with a reduced transparency to highlight different positions of inhibitors. The crystal structures of kinases are annotated and corresponding PDB ID of the kinase complexes with allosteric inhibitors are also shown

advances and clinical success manifested in multiple FDA-approved drugs, the number of studies focused on development of kinase activators is still very small. It has been noted that mechanisms underpinning allosteric action of kinase activators can proceed by destabilization of the inactive state, stabilization of the active state, facilitating of the active state, and dynamic responses to phosphorylation in regulatory sites [43, 49, 59]. Allosteric activation through binding to the catalytic domain of a dormant enzyme and induction of a conformational change is considered a prevalent mechanism of allosteric activation. Although binding of allosteric modulators of protein kinases can often involve significant rearrangements of the phosphorylated loop toward the active site and the rotation of the C-terminal domain, recent studies have indicated that switching functions of kinase structures and allosteric kinase activation can be achieved on different time scales by small, medium, and large conformational changes [225, 226]. NMR studies have shown that domain dynamics in the MAP kinase ERK2 is inherently constrained at the hinge region, distal from the site of phosphorylation. Phosphorylation can allosterically trigger a significant shift in conformational dynamic exchanges throughout the kinase core due to release of constraints in the hinge region [225]. The effect of

phosphorylation on ERK2 dynamics yields new models of kinase activation and inhibition, exposing cryptic sites that can be recognized by novel conformation selective inhibitors [170]. Interestingly, JNK1 phosphorylation can also increase conformational dynamics in the regions involved in catalysis and substrate binding that trigger kinase activation. It was shown that in the inactive form of JNK1, nucleotide binding may select conformation in which the domains are more open and the N-terminal domain is transiently destabilized, whereby phosphorylation shifts equilibrium toward a catalytically competent structure [155]. In this case, kinase activation may invoke a combination of dynamically driven allostery and population-shift mechanisms. Crystallographic and solution studies of allosteric inhibitor binding with PDK1 kinase have shown that small-molecule activators can allosterically activate the enzyme, leading to the ordering of the activation loop and controlling the assembly of an active kinase conformation [56, 190]. The crystal structure of the PDK1 complex with PS48 activator provided further insights into the allosteric mode of action of these modulators [86]. The allosteric effect of this activator is manifested in reducing conformational plasticity and ordering key functional elements of the kinase catalytic domain (the αC-helix, the glycine-rich loop, and the activation loop) that promotes stabilization of the active PDK1 conformation. Analogous to the discovery of activators that bind to the PIF pocket in PDK1, a small-molecule activator of PP1 to the allosteric site mimics binding of the endogenous protein regulator, increasing the enzymatic activity relative to the inhibited form [195]. Allosteric activation mechanisms can provide graded levels of regulatory control in protein kinases through phosphorylation-regulated protein dynamics in ERK2 [226] and AurA kinases [111, 118, 171]. In AurA kinase, phosphorylation of the activation loop can trigger a graded dynamic switch to a fully activated form and enhancement of the catalytic activity of the active form within a dynamic conformational ensemble [171]. Interestingly, allosteric activator of AurA kinase Tpx2 acts using a population-shift mechanism by switching from the inactive DFG-out to the active DFG-in state, while phosphorylation can induce activation through dynamics-driven allosteric modulation of the activation loop causing marginal stabilization of the active conformation [111, 118].

Allosteric activation by eliciting functional conformational changes in the catalytic domain through binding to the regulatory subunit underlies allosteric mechanism of AMPK [3, 41], PKA [105, 197], and PKC kinases [27]. Recent studies of the RAF regulatory mechanisms activation have revealed that RAF inhibitors can differentially affect the enzyme transactivation through allosteric modulation of the side-to-side dimerization [82, 83, 159]. These studies have discovered that most of the RAF inhibitors can induce dimerization and paradoxically stimulate RAS-dependent activation by conferring an active conformation of the BRAF kinase domain. As a result, the inhibitory potential of the existing RAF inhibitors can be compromised as they may paradoxically stimulate activation and increase ERK signaling, causing drug resistance [87]. The allosteric effect of these molecules is based on modulation of the dimerization interface and eliciting functional changes in the second protomer without altering structural topology of the dimer.

9.8 Allosteric Inhibitors and Activators of the ABL Kinase: Distinguishing Two Sides of Allostery

The activation and inhibition reactions in the ABL kinase are regulated allosterically through formation of multi-protein regulatory complexes involving the SH3 and SH2 domains and the kinase domain (Fig. 9.6). Recent structural and biochemical investigations of ABL kinase [81, 143, 144, 156] have reported molecular details of the regulatory interactions by which the SH3 and SH2 domains can act cooperatively with the catalytic core to suppress or promote kinase activation. In the downregulated inactive state of the ABL–SH2–SH3 complex, the SH3 domain binds to the linker that connects the SH2 domain and the kinase domain, while the

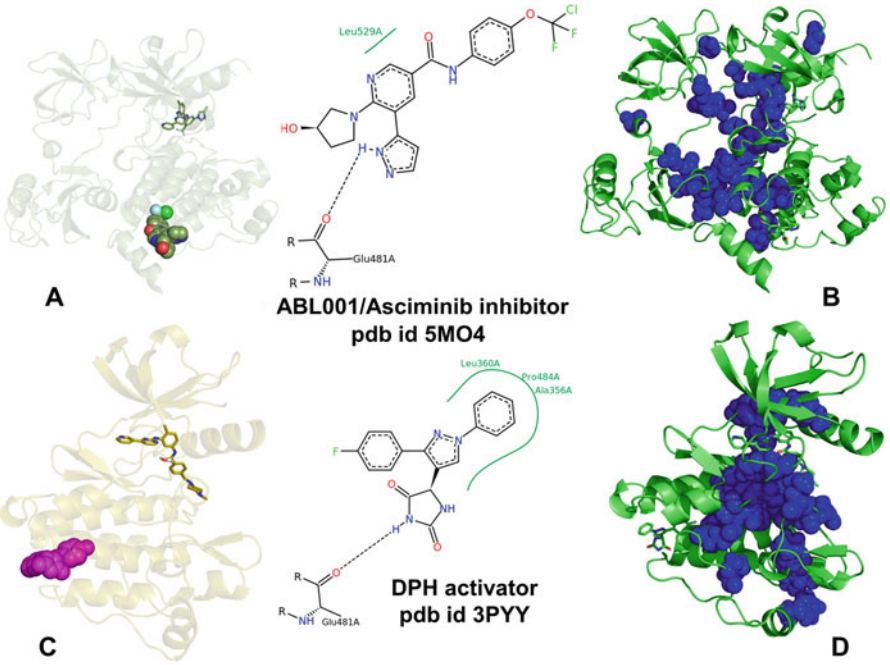


Fig. 9.6 The dynamic network analysis and structural mapping of the thermodynamically favorable communication routes in the ABL complexes with ABL001 inhibitor and DPH activator. (**a**) The schematic view of the ABL complex with the allosteric inhibitor ABL001 (asciminib) (pdb id 5MO4). The allosteric inhibitor is shown in spheres; the protein structure is in ribbons with the reduced transparency. The two-dimensional diagram of the allosteric inhibitor binding. Black dashed lines indicate hydrogen bonds, salt bridges, and metal interactions. Green solid line shows hydrophobic interactions. (**b**) The favorable communication paths are shown in blue spheres; the ABL structure is in green ribbons. The allosteric inhibitor asciminib and ATP-competitive inhibitor nilotinib are shown in atom-colored sticks. (**c**) The schematic view of ABL1 complex with the allosteric activator DPH (pdb id 3PYY). The activator is shown in spheres; the protein structure is in ribbons with the reduced transparency. (**d**) The favorable communication paths in the ABL complex with DPH are shown in blue spheres

SH2 domain interacts with the C-terminal kinase lobe acting as an autoinhibitory "clamp" that constrains catalytic activity of the kinase [92, 143]. The N-terminal "cap" (NCap) fragment can form direct interactions with the SH2 domain and the SH3–SH2 linker, thus strengthening the autoinhibitory "grip" over the catalytic core [33]. The disengagement of the SH2–SH3 domains that relieves the autoinhibitory constraints and yields an activated form can be promoted by phosphorylation of the SH3 tyrosine residues [9, 23, 34]. Phosphorylation and single mutations of functional residues in the SH3 and the linker regions can allosterically perturb the network of autoinhibitory interactions and induce a population shift between inactive and active kinase states. Recent biochemical studies have shown that the SH2–kinase binding can allosterically induce kinase activation, confirming a positive regulatory role of this interaction in kinase function [75, 112]. NMR studies of ABL complexes with inhibitors have established the existence of multiple conformations in which the SH2 and SH3 domains can adopt a range of different positions with respect to the kinase domain, providing an ensemble of preexisting conformational that can be accommodated by the kinase domain to produce states with different activity levels. The latest "tour de force" NMR study has revealed a detailed atomistic picture of allosteric regulation in the ABL kinase, showing how interacting signaling modules cooperate with the kinase domain to form a multilayered regulatory mechanism that exploits various allosteric switches powered by binding or phosphorylation at different sites of the regulatory assemblies [173].

Recently discovered allosteric inhibitor of the ABL kinase GNF-2 which binds to the myristic pocket near the C-terminus of the ABL kinase domain stabilizes the inactive conformation of the kinase [1, 233]. Using solution NMR, X-ray crystallography, mutagenesis, and hydrogen exchange mass spectrometry, it was shown that GNF-2 binding to the myristate binding site could induce changes in the dynamics of the ATP binding site. Although GNF-2 cannot inhibit several imatinib-resistant mutants, a combination of GNF-2 with another drug nilotinib was effective in inhibiting many imatinib-resistant mutants [233]. GNF-5 (a GNF-2 analog with improved pharmacokinetics), when used in combination with imatinib or nilotinib, could inhibit both wild-type and T315I mutant in biochemical and cellular assays [233]. According to the proposed inhibition mechanism of GNF-2/GNF-5 inhibitors to the myristate binding site, these compounds can induce a bent conformation of the αI-helix in the C-terminal lobe that serves as a switch promoting the SH2–KD interactions and stabilization of the inhibited conformation. NMR studies have confirmed that GNF2 and GNF5 binding to ABL can function via a population-shift mechanism that drives the equilibrium toward the assembled, inhibited state and decreases the activated-state population [233]. Simultaneous binding of GNF-5 and the ATP-competitive inhibitor dasatinib to the Abl-T315I mutant induced conformational changes similar to those observed with wild-type ABL in the presence of dasatinib alone [92]. Taken together, the results provided evidence for allosteric communication between the allosteric and ATP binding sites, revealing that dasatinib resistance can be overcome by simultaneous binding of GNF-5 that allosterically remodels the ATP binding site.

ABL001 (asciminib) is another allosteric ABL compound discovered through NMR-based conformational assay and structure-based design that inhibits ABL kinase activity by selectively binding to the myristoyl pocket and shifting equilibrium to the inactive kinase conformation [175, 224]. This allosteric inhibitor and second-generation ATP-competitive inhibitors have similar cellular potencies but distinct patterns of resistance mutations [115, 224]. While a number of drug-resistant mutations of allosteric ABL inhibitors are clustered near the myristoyl binding pocket, several drug-sensitive mutations in the kinase domain are located away from the allosteric site and confer resistance to allosteric inhibition by shifting the equilibrium and activating the kinase [115]. Asciminib is in phase I clinical trials as monotherapy and in combination with ABL drugs imatinib, nilotinib, and dasatinib for the treatment of patients with refractory chronic myeloid leukemia (CML). NMR-based conformational assays have shown that the conformational state of the αI-helix in the C-terminal lobe is a structural determinant for functional activity of allosteric myristate inhibitors. The allosteric inhibitors that mimic binding of myristate and induce bending of the αI-helix are functional antagonists, whereas allosteric ligands that bind to the same pocket but do not induce this local conformational change are kinase agonists [93].

A high-throughput screen has discovered another class of allosteric ABL compounds that targets the ABL myristoyl pocket and yet can further enhance the rate of kinase activation in the presence of ATP [230]. Unlike allosteric inhibitors, allosteric agonist DPH binds in the myristate pocket, forcing αI-helix into linear extended conformation (PDB code 3PYY) and promoting disengagement of the SH2 domain and subsequent conformational change to the active ABL form. Hence, although GNF-2/GNF-5, asciminib, and DPH ligands all bind to the myristate pocket and occupy the same structural positions, DPH activates ABL kinase, while other compounds suppress kinase activity. Quite surprisingly, recent NMR and biophysical methods have found that Gleevec is capable of binding to the myristate pocket with a sub-micromolar affinity and acts as an allosteric activator of ABL at higher doses of administering the drug [228]. Structural and biochemical studies of allosteric inhibitors and activators of ABL kinase have indicated that local structural environment near the myristate pocket can serve as a sensor of ligand binding, triggering either stabilization of the inactive state or large conformational shift and activation. Furthermore, synergistic actions of allosteric and ATP competitive inhibitors have provided strong evidence that binding can perturb dynamics at distal regions and elicit ligand-specific communication between binding sites. It remains unclear how this long-range effect is transmitted from the myristate to the ATP pocket.

We argued that reconstruction of residue interaction networks and ensembles of communication pathways in ABL complexes with inhibitors and activators may help to address these questions and detail allosteric signatures of two classes of allosteric modulators. In the presence of ABL001 inhibitor, we observed the emergence of several dominant allosteric communication pathways that are mediated by high centrality residues in the kinase domain, SH2 domain, and the SH2–SH3 linker (Fig. 9.6). The spatial organization of these pathways reflected a fairly large

allosteric network that favors distinct "narrow tubes" of paths connecting the allosteric site and ATP binding site (Fig. 9.6). The topography of dominant pathways in the ensemble is dictated by ligand-induced inhibitory conformation that locks the interactions between the SH2 and kinase domains. One of the dominant pathway tubes proceeds directly from the allosteric site navigating through several functionally important positions A337, A344, P465, V468, and F11 (corresponding to A356, A363, P484, V487, and F330, respectively, in the residue numbering of pdb id 5MO4). The existence of this direct path may exploit network efficiency in propagating this signal through cooperative interactions between these groups of rigid residues (Fig. 9.6). The emergence of this path in the statistical ensemble of network communications is supported by biochemical experiments showing that A344P, A337V, P465S, V468F, and Y469H substitutions near the allosteric site may cause drug resistance of ABL001 by relieving autoinhibition and promoting population shift to the activating form [115, 224]. The determined network of pathways specified a potential allosteric crosstalk between sites which goes through a set of functional residues located close to the αG- and αI-helices of the C-lobe, in a region that can be targeted by allosteric kinase inhibitors. Another dominant pathway in the ensemble is formed through couplings with the SH2 and SH2–SH3 linker residues (Fig. 9.6). Among key centrality residues serving as mediating sites in this path are SH2 residues (V151, A155, Y158) and linker residues (T243, P242, Y245, and P249). The importance of these centers is supported by NMR and biochemical experiments, demonstrating that mutations of Y158 in the SH2–KD interface can compromise the downregulated state and cause an increase in kinase activity [81]. Similarly, phosphorylation of Y245 in the linker can also disrupt the SH3–linker interactions and promote ABL kinase activity [34]. NMR data have revealed that the allosteric mutants in the linker region (P242E and P249E) can shift the population of the inhibited and activated state, giving rise to a substantially increased activation [173]. These allosteric drug-resistant mutants were shown to promote the activated state of ABL and compromise the inhibitory function of imatinib. Notably, preferential communication paths in the ensemble can weaken allosteric coupling between the ATP binding site and substrate binding site which is the intrinsic allosteric signature of the catalytic domain [66, 94, 207]. In contrast, in the ABL complex with DPH activator, we found a wider ensemble of communication pathways connecting the ATP binding sites and substrate binding site. Interestingly, through disengagement of the inhibitory lock with the SH2 domain, allosteric activator DPH switches the preferential pathways and leverages the C-spine network (V256, A269, L323, C369, L370, V317, L438, and I432) for efficient communication with the ATP site. In addition, DPH binding promotes a wide path connecting the ATP and substrate binding site, primarily by engaging the R-spine in ABL. This subnetwork consists of M290 from the C-terminal end of the αC-helix, L301 from the β4-strand, F382 of the DFG motif in the beginning of the A-loop, H361 of the HRD motif in the catalytic loop, and D421 of the αF-helix. Collectively, the broad ensemble of preferential pathways interlinks the allosteric binding site with the ATP and substrate binding pockets, ensuring robustness of the activating signal, which may be an important network signature of the active form of ABL kinase (Fig. 9.6).

These results indicated that allosteric modulators may selectively activate various communication routes from the ensemble that is intrinsically determined by the protein topology. The long-range communications between the allosteric and ATP binding sites in the inhibitory state can critically depend on structural integrity of these high centrality mediating sites. According to our findings, rapid communication in the ABL complex with allosteric inhibitor may come at the expense of high dependency and sensitivity on a relatively small number of coordinating modes, making allosteric drug binding vulnerable to targeted mutations of mediating centers. Exploiting and targeting this communication network may thus offer a way for modulating the kinase activity. In network terms, the emergence of several alternative pathways may provide better balance between the efficiency and resilience of the interaction networks. It is noteworthy that resilience introduces a certain degree of redundancy into the system, which may lead to a decreased efficiency but ensures stronger protection from random perturbations and targeted attacks.

9.9 Concluding Remarks

The recent years have witnessed significant advances in our understanding of the allosteric mechanisms and interactions in complex protein systems. Viewing proteins as dynamic regulatory machines that constantly fluctuate between different conformations in response to external perturbations and effector binding may be the key to understanding complexity and diversity of allosteric regulation. Although we have achieved a basic understanding of a variety of allosteric scenarios and have gained important insights into the function of allosteric proteins and modulators, the quantitate characterization of these highly dynamic and often elusive processes remains a great challenge. Fifty years after the KNF and MWC models were proposed, the processes involved in allosteric regulation are still not completely understood, and more work is still required before a consensus toward a general unified theory of allostery is reached. The lessons from our analysis suggest that allosteric regulation mechanisms can proceed via a subtle and nontrivial combination of the three classical models of allostery: conformational selection, induced-fit, and dynamic allostery. Although allostery takes place at the protein level, its effect has implications at the cellular level, as allostery enables to communicate environmental signals and alter specific cellular functions. Understanding of system-based relationships between protein robustness, disease, and allosteric drug binding at the structural level may prove to be useful in the development of theoretical and experimental approaches bridging structure-based network analysis of protein targets with modeling of protein interaction networks and pathways.

The emerging realization that allosteric inhibitors and activators can exploit distinct regulatory mechanisms and subtly modulate protein activities could open up new venues and opportunities for probing signaling processes, engineering allostery using synthetic biology principles, and development of drug combinations with improved therapeutic indices and broad range of activities. The next breakthrough in the discovery of allosteric

drugs may require integration of network science, systems biology approaches, and experiments to bridge microscopic analysis of genes with macroscopic understanding of cellular networks and signaling pathways. Data-driven and system-based modeling in combination with powerful machine learning and artificial intelligence tools could potentially provide a new robust platform to exploit advances in medicinal chemistry, chemical biology, and structural biology for the development of novel drugs and precision medicine therapies.

References

1. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y, Zhang G, Hur W, Ding S, Manley P, Mestan J, Fabbro D, Gray NS (2006) Allosteric inhibitors of Bcr-abl-dependent cell proliferation. Nat Chem Biol 2(2):95–102
2. Aguilar D, Oliva B, Marino Buslje C (2012) Mapping the mutual information network of enzymatic families in the protein structure to unveil functional features. PLoS One 7(7): e41430
3. Amodeo GA, Rudolph MJ, Tong L (2007) Crystal structure of the heterotrimer core of Saccharomyces cerevisiae AMPK homologue SNF1. Nature 449(7161):492–495
4. Artim SC, Mendrola JM, Lemmon MA (2012) Assessing the range of kinase autoinhibition mechanisms in the insulin receptor family. Biochem J 448(2):213–220
5. Ashwell MA, Lapierre JM, Brassard C, Bresciano K, Bull C, Cornell-Kennon S, Eathiraj S, France DS, Hall T, Hill J, Kelleher E, Khanapurkar S, Kizer D, Koerner S, Link J, Liu Y, Makhija S, Moussa M, Namdev N, Nguyen K, Nicewonger R, Palma R, Szwaya J, Tandon M, Uppalapati U, Vensel D, Volak LP, Volckova E, Westlund N, Wu H, Yang RY, Chan TC (2012) Discovery and optimization of a series of 3-(3-phenyl-3H-imidazo[4,5-b]pyridin-2-yl) pyridin-2-amines: orally bioavailable, selective, and potent ATP-independent Akt inhibitors. J Med Chem 55(11):5291–5310
6. Bahar I, Chennubhotla C, Tobi D (2007) Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Curr Opin Struct Biol 17(6):633–640
7. Bahar I, Lezon TR, Yang LW, Eyal E (2010) Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys 39:23–42
8. Bahar I, Rader AJ (2005) Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 15(5):586–592
9. Barila D, Superti-Furga G (1998) An intramolecular SH3-domain interaction regulates c-Abl activity. Nat Genet 18(3):280–282
10. Barnett SF, Defeo-Jones D, Fu S, Hancock PJ, Haskell KM, Jones RE, Kahana JA, Kral AM, Leander K, Lee LL, Malinowski J, McAvoy EM, Nahas DD, Robinson RG, Huber HE (2005) Identification and characterization of pleckstrin-homology-domain-dependent and isoenzyme-specific Akt inhibitors. Biochem J 385(Pt 2):399–408
11. Berlow RB, Dyson HJ, Wright PE (2018) Expanding the paradigm: intrinsically disordered proteins and allosteric regulation. J Mol Biol 430(16):2309–2320
12. Betzi S, Alam R, Martin M, Lubbers DJ, Han H, Jakkaraj SR, Georg GI, Schonbrunn E (2011) Discovery of a potential allosteric ligand binding site in CDK2. ACS Chem Biol 6(5):492–501
13. Bhattacharya S, Vaidehi N (2014) Differences in allosteric communication pipelines in the inactive and active states of a GPCR. Biophys J 107(2):422–434
14. Bhattacharya SK, Aspnes GE, Bagley SW, Boehm M, Brosius AD, Buckbinder L, Chang JS, Dibrino J, Eng H, Frederick KS, Griffith DA, Griffor MC, Guimaraes CR, Guzman-Perez A, Han S, Kalgutkar AS, Klug-McLeod J, Garcia-Irizarry C, Li J, Lippa B, Price DA, Southers JA, Walker DP, Wei L, Xiao J, Zawistoski MP, Zhao X (2012) Identification of novel series of

pyrazole and indole-urea based DFG-out PYK2 inhibitors. Bioorg Med Chem Lett 22(24):7523–7529
15. Bhattacharyya M, Vishveshwara S (2010) Elucidation of the conformational free energy landscape in H.pylori LuxS and its implications to catalysis. BMC Struct Biol 10:27
16. Bhattacharyya M, Vishveshwara S (2011) Probing the allosteric mechanism in pyrrolysyl-tRNA synthetase using energy-weighted network formalism. Biochemistry 50(28):6225–6236
17. Bishop AC, Chen VL (2009) Brought to life: targeted activation of enzyme function with small molecules. J Chem Biol 2(1):1–9
18. Blacklock K, Verkhivker GM (2014) Computational modeling of allosteric regulation in the Hsp90 chaperones: a statistical ensemble analysis of protein structure networks and allosteric communications. PLoS Comput Biol 10(6):21
19. Bock A, Bermudez M, Krebs F, Matera C, Chirinda B, Sydow D, Dallanoce C, Holzgrabe U, De Amici M, Lohse MJ, Wolber G, Mohr K (2016) Ligand binding ensembles determine graded agonist efficacies at a G protein-coupled receptor. J Biol Chem 291(31):16375–16389
20. Boggon TJ, Eck MJ (2004) Structure and regulation of Src family kinases. Oncogene 23(48):7918–7927
21. Bose R, Zhang X (2009) The ErbB kinase domain: structural perspectives into kinase activation and inhibition. Exp Cell Res 315(4):649–658
22. Bowman GR, Geissler PL (2012) Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc Natl Acad Sci U S A 109(29):11681–11686
23. Brasher BB, Roumiantsev S, Van Etten RA (2001) Mutational analysis of the regulatory function of the c-Abl Src homology 3 domain. Oncogene 20(53):7744–7752
24. Brinda KV, Vishveshwara S (2005) A network representation of protein structures: implications for protein stability. Biophys J 89(6):4159–4170
25. Brinda KV, Vishveshwara S (2010) Random network behaviour of protein structures. Mol BioSyst 6(2):391–398
26. Buchenberg S, Sittel F, Stock G (2017) Time-resolved observation of protein allosteric communication. Proc Natl Acad Sci U S A 114(33):E6804–e6811
27. Budas GR, Koyanagi T, Churchill EN, Mochly-Rosen D (2007) Competitive inhibitors and allosteric activators of protein kinase C isoenzymes: a personal account and progress report on transferring academic discoveries to the clinic. Biochem Soc Trans 35(Pt 5):1021–1026
28. Busschots K, Lopez-Garcia LA, Lammi C, Stroba A, Zeuzem S, Piiper A, Alzari PM, Neimanis S, Arencibia JM, Engel M, Schulze JO, Biondi RM (2012) Substrate-selective inhibition of protein kinase PDK1 by small compounds that bind to the PIF-pocket allosteric docking site. Chem Biol 19(9):1152–1163
29. Chakrabarti S, Panchenko AR (2009) Coevolution in defining the functional specificity. Proteins 75(1):231–240
30. Chakrabarti S, Panchenko AR (2010) Structural and functional roles of coevolved sites in proteins. PLoS One 5(1):e8591
31. Changeux JP (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 41:103–133
32. Changeux JP, Edelstein SJ (2005) Allosteric mechanisms of signal transduction. Science 308(5727):1424–1428
33. Chen S, Dumitrescu TP, Smithgall TE, Engen JR (2008a) Abl N-terminal cap stabilization of SH3 domain dynamics. Biochemistry 47(21):5795–5803
34. Chen S, O'Reilly LP, Smithgall TE, Engen JR (2008b) Tyrosine phosphorylation in the SH3 domain disrupts negative regulatory interactions within the c-Abl kinase core. J Mol Biol 383(2):414–423
35. Chennubhotla C, Bahar I (2006) Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. Mol Syst Biol 2:36
36. Chennubhotla C, Bahar I (2007) Signal propagation in proteins and relation to equilibrium fluctuations. PLoS Comput Biol 3(9):1716–1726
37. Chennubhotla C, Yang Z, Bahar I (2008) Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. Mol BioSyst 4(4):287–292

38. Choi JH, Laurent AH, Hilser VJ, Ostermeier M (2015) Design of protein switches based on an ensemble model of allostery. Nat Commun 6:6968
39. Christodoulou MS, Caporuscio F, Restelli V, Carlino L, Cannazza G, Costanzi E, Citti C, Lo Presti L, Pisani P, Battistutta R, Broggini M, Passarella D, Rastelli G (2017) Probing an allosteric pocket of CDK2 with small molecules. ChemMedChem 12(1):33–41
40. Comess KM, Sun C, Abad-Zapatero C, Goedken ER, Gum RJ, Borhani DW, Argiriadi M, Groebe DR, Jia Y, Clampit JE, Haasch DL, Smith HT, Wang S, Song D, Coen ML, Cloutier TE, Tang H, Cheng X, Quinn C, Liu B, Xin Z, Liu G, Fry EH, Stoll V, Ng TI, Banach D, Marcotte D, Burns DJ, Calderwood DJ, Hajduk PJ (2011) Discovery and characterization of non-ATP site inhibitors of the mitogen activated protein (MAP) kinases. ACS Chem Biol 6(3):234–244
41. Cool B, Zinker B, Chiou W, Kifle L, Cao N, Perham M, Dickinson R, Adler A, Gagne G, Iyengar R, Zhao G, Marsh K, Kym P, Jung P, Camp HS, Frevert E (2006) Identification and characterization of a small molecule AMPK activator that treats key components of type 2 diabetes and the metabolic syndrome. Cell Metab 3(6):403–416
42. Cooper A, Dryden DT (1984) Allostery without conformational change. A plausible model. Eur Biophys J 11(2):103–109
43. Cowan-Jacob SW, Jahnke W, Knapp S (2014) Novel approaches for targeting kinases: allosteric inhibition, allosteric activation and pseudokinases. Future Med Chem 6(5):541–561
44. Cox KJ, Shomin CD, Ghosh I (2011) Tinkering outside the kinase ATP box: allosteric (type IV) and bivalent (type V) inhibitors of protein kinases. Future Med Chem 3(1):29–43
45. Csermely P, Korcsmaros T, Kiss HJ, London G, Nussinov R (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. Pharmacol Ther 138(3):333–408
46. Csermely P, Palotai R, Nussinov R (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. Trends Biochem Sci 35(10):539–546
47. Csermely P, Sandhu KS, Hazai E, Hoksza Z, Kiss HJ, Miozzo F, Veres DV, Piazza F, Nussinov R (2012) Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function: hypotheses and a comprehensive review. Curr Protein Pept Sci 13(1):19–33
48. Dam TK, Roy R, Page D, Brewer CF (2002) Negative cooperativity associated with binding of multivalent carbohydrates to lectins. Thermodynamic analysis of the "multivalency effect". Biochemistry 41(4):1351–1358
49. Dar AC, Shokat KM (2011) The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. Annu Rev Biochem 80:769–795
50. de Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. Nat Rev Genet 14(4):249–261
51. del Sol A, Tsai CJ, Ma B, Nussinov R (2009) The origin of allosteric functional modulation: multiple pre-existing pathways. Structure 17(8):1042–1050
52. Di Paola L, Giuliani A (2015) Protein contact network topology: a natural language for allostery. Curr Opin Struct Biol 31:43–48
53. Dokholyan NV (2016) Controlling allosteric networks in proteins. Chem Rev 116(11):6463–6487
54. Dong Q, Dougan DR, Gong X, Halkowycz P, Jin B, Kanouni T, O'Connell SM, Scorah N, Shi L, Wallace MB, Zhou F (2011) Discovery of TAK-733, a potent and selective MEK allosteric site inhibitor for the treatment of cancer. Bioorg Med Chem Lett 21(5):1315–1319
55. Endicott JA, Noble ME, Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. Annu Rev Biochem 81:587–613
56. Engel M, Hindie V, Lopez-Garcia LA, Stroba A, Schaeffer F, Adrian I, Imig J, Idrissova L, Nastainczyk W, Zeuzem S, Alzari PM, Hartmann RW, Piiper A, Biondi RM (2006) Allosteric activation of the protein kinase PDK1 with low molecular weight compounds. EMBO J 25(23):5469–5480

57. Engen JR, Wales TE, Hochrein JM, Meyn MA 3rd, Banu Ozkan S, Bahar I, Smithgall TE (2008) Structure and dynamic regulation of Src-family kinases. Cell Mol Life Sci 65(19):3058–3073
58. Fabbro D (2015) 25 years of small molecular weight kinase inhibitors: potentials and limitations. Mol Pharmacol 87(5):766–775
59. Fang Z, Grutter C, Rauh D (2013) Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features. ACS Chem Biol 8(1):58–70
60. Fang Z, Simard JR, Plenker D, Nguyen HD, Phan T, Wolle P, Baumeister S, Rauh D (2015) Discovery of inter-domain stabilizers-a novel assay system for allosteric akt inhibitors. ACS Chem Biol 10(1):279–288
61. Fasano M, Della Corte CM, Califano R, Capuano A, Troiani T, Martinelli E, Ciardiello F, Morgillo F (2014) Type III or allosteric kinase inhibitors for the treatment of non-small cell lung cancer. Expert Opin Investig Drugs 23(6):809–821
62. Feher VA, Durrant JD, Van Wart AT, Amaro RE (2014) Computational approaches to mapping allosteric pathways. Curr Opin Struct Biol 25:98–103
63. Ferguson KM (2008) Structure-based view of epidermal growth factor receptor regulation. Annu Rev Biophys 37:353–373
64. Ferreon AC, Ferreon JC, Wright PE, Deniz AA (2013) Modulation of allostery by protein intrinsic disorder. Nature 498(7454):390–394
65. Fischmann TO, Smith CK, Mayhood TW, Myers JE, Reichert P, Mannarino A, Carr D, Zhu H, Wong J, Yang RS, Le HV, Madison VS (2009) Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. Biochemistry 48(12):2661–2674
66. Foda ZH, Shan Y, Kim ET, Shaw DE, Seeliger MA (2015) A dynamically coupled allosteric network underlies binding cooperativity in Src kinase. Nat Commun 6:5939
67. Frederick KK, Marlow MS, Valentine KG, Wand AJ (2007) Conformational entropy in molecular recognition by proteins. Nature 448(7151):325–329
68. Garcia-Pino A, Balasubramanian S, Wyns L, Gazit E, De Greve H, Magnuson RD, Charlier D, van Nuland NA, Loris R (2010) Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell 142(1):101–111
69. Ghosh A, Sakaguchi R, Liu C, Vishveshwara S, Hou YM (2011) Allosteric communication in cysteinyl tRNA synthetase: a network of direct and indirect readout. J Biol Chem 286(43):37721–37731
70. Ghosh A, Vishveshwara S (2007) A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. Proc Natl Acad Sci U S A 104(40):15711–15716
71. Ghosh A, Vishveshwara S (2008) Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. Biochemistry 47(44):11398–11407
72. Gloor GB, Martin LC, Wahl LM, Dunn SD (2005) Mutual information in protein multiple sequence alignments reveals two classes of coevolving positions. Biochemistry 44(19):7156–7165
73. Goncearenco A, Mitternacht S, Yong T, Eisenhaber B, Eisenhaber F, Berezovsky IN (2013) SPACER: server for predicting allosteric communication and effects of regulation. Nucleic Acids Res 41(Web Server issue):W266–W272
74. Good MC, Zalatan JG, Lim WA (2011) Scaffold proteins: hubs for controlling the flow of cellular information. Science 332(6030):680–686
75. Grebien F, Hantschel O, Wojcik J, Kaupe I, Kovacic B, Wyrzucki AM, Gish GD, Cerny-Reiterer S, Koide A, Beug H, Pawson T, Valent P, Koide S, Superti-Furga G (2011) Targeting the SH2-kinase interface in Bcr-Abl inhibits leukemogenesis. Cell 147(2):306–319
76. Grutsch S, Bruschweiler S, Tollinger M (2016) NMR methods to study dynamic allostery. PLoS Comput Biol 12(3):e1004620
77. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? Proteins 57(3):433–443

78. Guo J, Zhou HX (2016) Protein allostery and conformational dynamics. Chem Rev 116(11):6503–6515
79. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138(4):774–786
80. Hansia P, Ghosh A, Vishveshwara S (2009) Ligand dependent intra and inter subunit communication in human tryptophanyl tRNA synthetase as deduced from the dynamics of structure networks. Mol BioSyst 5(12):1860–1872
81. Hantschel O (2012) Structure, regulation, signaling, and targeting of abl kinases in cancer. Genes Cancer 3(5-6):436–446
82. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S, Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS, Malek S (2010) RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. Nature 464(7287):431–435
83. Heidorn SJ, Milagre C, Whittaker S, Nourry A, Niculescu-Duvas I, Dhomen N, Hussain J, Reis-Filho JS, Springer CJ, Pritchard C, Marais R (2010) Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF. Cell 140(2):209–221
84. Heinrich T, Gradler U, Bottcher H, Blaukat A, Shutes A (2010) Allosteric IGF-1R Inhibitors. ACS Med Chem Lett 1(5):199–203
85. Hilser VJ, Wrabl JO, Motlagh HN (2012) Structural and energetic basis of allostery. Annu Rev Biophys 41:585–609
86. Hindie V, Stroba A, Zhang H, Lopez-Garcia LA, Idrissova L, Zeuzem S, Hirschberg D, Schaeffer F, Jorgensen TJ, Engel M, Alzari PM, Biondi RM (2009) Structure and allosteric effects of low-molecular-weight activators on the protein kinase PDK1. Nat Chem Biol 5(10):758–764
87. Holderfield M, Nagel TE, Stuart DD (2014) Mechanism and consequences of RAF kinase activation by small-molecule inhibitors. Br J Cancer 111(4):640–645
88. Hsu YH, Traugh JA (2010) Reciprocally coupled residues crucial for protein kinase Pak2 activity calculated by statistical coupling analysis. PLoS One 5(3):e9455
89. Huang C, Kalodimos CG (2017) Structures of large protein complexes determined by nuclear magnetic resonance spectroscopy. Annu Rev Biophys 46:317–336
90. Huang Z, Mou L, Shen Q, Lu S, Li C, Liu X, Wang G, Li S, Geng L, Liu Y, Wu J, Chen G, Zhang J (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42(Database issue):D510–D516
91. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39(Database issue):D663–D669
92. Iacob RE, Zhang J, Gray NS, Engen JR (2011) Allosteric interactions between the myristate- and ATP-site of the Abl kinase. PLoS One 6(1):e15929
93. Jahnke W, Grotzfeld RM, Pelle X, Strauss A, Fendrich G, Cowan-Jacob SW, Cotesta S, Fabbro D, Furet P, Mestan J, Marzinzik AL (2010) Binding or bending: distinction of allosteric Abl kinase agonists from antagonists by an NMR-based conformational assay. J Am Chem Soc 132(20):7043–7048
94. James KA, Verkhivker GM (2014) Structure-based network analysis of activation mechanisms in the ErbB family of receptor tyrosine kinases: the regulatory spine residues are global mediators of structural stability and allosteric interactions. PLoS One 9(11):46
95. Jeon J, Nam HJ, Choi YS, Yang JS, Hwang J, Kim S (2011) Molecular evolution of protein conformational changes revealed by a network of evolutionarily coupled residues. Mol Biol Evol 28(9):2675–2685
96. Jiang Y, Kalodimos CG (2017) NMR Studies of Large Proteins. J Mol Biol 429(17):2667–2676

97. Jura N, Zhang X, Endres NF, Seeliger MA, Schindler T, Kuriyan J (2011) Catalytic control in the EGF receptor and its connection to general kinase regulatory mechanisms. Mol Cell 42(1):9–22
98. Kalbitzer HR, Rosnizeck IC, Munte CE, Narayanan SP, Kropf V, Spoerner M (2013) Intrinsic allosteric inhibition of signaling proteins by targeting rare interaction states detected by high-pressure NMR spectroscopy. Angew Chem Int Ed Eng 52(52):14242–14246
99. Kalodimos CG (2011) NMR reveals novel mechanisms of protein activity regulation. Protein Sci 20(5):773–782
100. Kalodimos CG (2012) Protein function and allostery: a dynamic relationship. Ann N Y Acad Sci 1260:81–86
101. Kay LE (2016) New views of functionally dynamic proteins by solution NMR spectroscopy. J Mol Biol 428(2 Pt A):323–331
102. Kaya C, Armutlulu A, Ekesan S, Haliloglu T (2013) MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. Nucleic Acids Res 41(Web Server issue):W249–W255
103. Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. Curr Opin Struct Biol 13(6):748–757
104. Kidd BA, Baker D, Thomas WE (2009) Computation of conformational coupling in allosteric proteins. PLoS Comput Biol 5(8):e1000484
105. Kim C, Cheng CY, Saldanha SA, Taylor SS (2007) PKA-I holoenzyme structure reveals a mechanism for cAMP-dependent activation. Cell 130(6):1032–1043
106. Kornev AP, Taylor SS (2015) Dynamics-driven allostery in protein kinases. Trends Biochem Sci 40(11):628–647
107. Kornev AP, Taylor SS, Ten Eyck LF (2008) A helix scaffold for the assembly of active protein kinases. Proc Natl Acad Sci U S A 105(38):14377–14382
108. Korzhnev DM, Kay LE (2008) Probing invisible, low-populated States of protein molecules by relaxation dispersion NMR spectroscopy: an application to protein folding. Acc Chem Res 41(3):442–451
109. Koshland DE Jr (1998) Conformational changes: how small is big enough? Nat Med 4(10):1112–1114
110. La Sala G, Decherchi S, De Vivo M, Rocchia W (2017) Allosteric communication networks in proteins revealed through pocket crosstalk analysis. ACS Cent Sci 3(9):949–960
111. Lake EW, Muretta JM, Thompson AR, Rasmussen DM, Majumdar A, Faber EB, Ruff EF, Thomas DD, Levinson NM (2018) Quantitative conformational profiling of kinase inhibitors reveals origins of selectivity for Aurora kinase activation states. Proc Natl Acad Sci U S A. https://doi.org/10.1073/pnas.1811158115
112. Lamontanara AJ, Georgeon S, Tria G, Svergun DI, Hantschel O (2014) The SH2 domain of Abl kinases regulates kinase autophosphorylation by controlling activation loop accessibility. Nat Commun 5:5470
113. Laughlin JD, Nwachukwu JC, Figuera-Losada M, Cherry L, Nettles KW, LoGrasso PV (2012) Structural mechanisms of allostery and autoinhibition in JNK family kinases. Structure 20(12):2174–2184
114. Lee BC, Park K, Kim D (2008) Analysis of the residue-residue coevolution network and the functionally important residues in proteins. Proteins 72(3):863–872
115. Lee BJ, Shah NP (2017) Identification and characterization of activating ABL1 1b kinase mutations: impact on sensitivity to ATP-competitive and allosteric ABL1 inhibitors. Leukemia 31(5):1096–1107
116. Lemmon MA (2009) Ligand-induced ErbB receptor dimerization. Exp Cell Res 315(4):638–648
117. Lemmon MA, Schlessinger J (2010) Cell signaling by receptor tyrosine kinases. Cell 141(7):1117–1134
118. Levinson NM (2018) The multifaceted allosteric regulation of Aurora kinase A. Biochem J 475(12):2025–2042

119. Li C, Ma N, Wang Y, Chen G (2014) Molecular dynamics simulation studies on the positive cooperativity of the Kemptide substrate with protein kinase A induced by the ATP ligand. J Phys Chem B 118(5):1273–1287
120. Lindsley CW, Zhao Z, Leister WH, Robinson RG, Barnett SF, Defeo-Jones D, Jones RE, Hartman GD, Huff JR, Huber HE, Duggan ME (2005) Allosteric Akt (PKB) inhibitors: discovery and SAR of isozyme selective inhibitors. Bioorg Med Chem Lett 15(3):761–764
121. Lisi GP, East KW, Batista VS, Loria JP (2017) Altering the allosteric pathway in IGPS suppresses millisecond motions and catalytic activity. Proc Natl Acad Sci U S A 114(17): E3414–e3423
122. Lisi GP, Loria JP (2016) Solution NMR spectroscopy for the study of enzyme allostery. Chem Rev 116(11):6323–6369
123. Lisi GP, Loria JP (2017) Allostery in enzyme catalysis. Curr Opin Struct Biol 47:123–130
124. Lisi GP, Manley GA, Hendrickson H, Rivalta I, Batista VS, Loria JP (2016) Dissecting dynamic allosteric pathways using chemically related small-molecule activators. Structure 24(7):1155–1166
125. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science 286(5438):295–299
126. Ma B, Tsai CJ, Haliloglu T, Nussinov R (2011) Dynamic allostery: linkers are not merely flexible. Structure 19(7):907–917
127. Ma J (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 13(3):373–380
128. Marino Buslje C, Teppa E, Di Domenico T, Delfino JM, Nielsen M (2010) Networks of high mutual information define the structural proximity of catalytic sites: implications for catalytic residue identification. PLoS Comput Biol 6(11):e1000978
129. Martin LC, Gloor GB, Dunn SD, Wahl LM (2005) Using information theory to search for co-evolving residues in proteins. Bioinformatics 21(22):4116–4124
130. Martin MP, Alam R, Betzi S, Ingles DJ, Zhu JY, Schonbrunn E (2012) A novel approach to the discovery of small-molecule ligands of CDK2. ChemBioChem 13(14):2128–2136
131. Mattmann ME, Stoops SL, Lindsley CW (2011) Inhibition of Akt with small molecules and biologics: historical perspective and current status of the patent landscape. Expert Opin Ther Pat 21(9):1309–1338
132. McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP (2009) Quantifying correlations between allosteric sites in thermodynamic ensembles. J Chem Theory Comput 5(9):2486–2502
133. McLaughlin RN Jr, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. Nature 491(7422):138–142
134. McLeish TC, Rodgers TL, Wilson MR (2013) Allostery without conformation change: modelling protein dynamics at multiple scales. Phys Biol 10(5):056004
135. Meharena HS, Chang P, Keshwani MM, Oruganty K, Nene AK, Kannan N, Taylor SS, Kornev AP (2013) Deciphering the structural basis of eukaryotic protein kinase regulation. PLoS Biol 11(10):e1001680
136. Ming D, Wall ME (2005) Quantifying allosteric effects in proteins. Proteins 59(4):697–707
137. Ming D, Wall ME (2006) Interactions in native binding sites cause a large change in protein dynamics. J Mol Biol 358(1):213–223
138. Mitternacht S, Berezovsky IN (2011) Binding leverage as a molecular basis for allosteric regulation. PLoS Comput Biol 7(9):e1002148
139. Monod J, Wyman J, Changeux JP (1965) ON the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
140. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci U S A 108(49):E1293–E1301
141. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508(7496):331–339

142. Munte CE, Beck Erlach M, Kremer W, Koehler J, Kalbitzer HR (2013) Distinct conformational states of the Alzheimer beta-amyloid peptide can be detected by high-pressure NMR spectroscopy. Angew Chem Int Ed Eng 52(34):8943–8947
143. Nagar B, Hantschel O, Seeliger M, Davies JM, Weis WI, Superti-Furga G, Kuriyan J (2006) Organization of the SH3-SH2 unit in active and inactive forms of the c-Abl tyrosine kinase. Mol Cell 21(6):787–798
144. Nagar B, Hantschel O, Young MA, Scheffzek K, Veach D, Bornmann W, Clarkson B, Superti-Furga G, Kuriyan J (2003) Structural basis for the autoinhibition of c-Abl tyrosine kinase. Cell 112(6):859–871
145. Nesmelova IV, Ermakova E, Daragan VA, Pang M, Menendez M, Lagartera L, Solis D, Baum LG, Mayo KH (2010) Lactose binding to galectin-1 modulates structural dynamics, increases conformational entropy, and occurs with apparent negative cooperativity. J Mol Biol 397(5):1209–1230
146. Nishi H, Hashimoto K, Panchenko AR (2011) Phosphorylation in protein-protein binding: effect on stability and function. Structure 19(12):1807–1815
147. Nussinov R (2012) How do dynamic cellular signals travel long distances? Mol BioSyst 8(1):22–26
148. Nussinov R, Ma B, Tsai CJ, Csermely P (2013a) Allosteric conformational barcodes direct signaling in the cell. Structure 21(9):1509–1521
149. Nussinov R, Tsai CJ (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305
150. Nussinov R, Tsai CJ (2015) Allostery without a conformational change? Revisiting the paradigm. Curr Opin Struct Biol 30:17–24
151. Nussinov R, Tsai CJ, Csermely P (2011) Allo-network drugs: harnessing allostery in cellular networks. Trends Pharmacol Sci 32(12):686–693
152. Nussinov R, Tsai CJ, Ma B (2013b) The underappreciated role of allostery in the cellular network. Annu Rev Biophys 42:169–189
153. Nussinov R, Tsai CJ, Xin F, Radivojac P (2012) Allosteric post-translational modification codes. Trends Biochem Sci 37(10):447–455
154. Oruganty K, Kannan N (2012) Design principles underpinning the regulatory diversity of protein kinases. Philos Trans R Soc Lond Ser B Biol Sci 367(1602):2529–2539
155. Owen GR, Stoychev S, Achilonu I, Dirr HW (2014) Phosphorylation- and nucleotide-binding-induced changes to the stability and hydrogen exchange patterns of JNK1beta1 provide insight into its mechanisms of activation. J Mol Biol 426(21):3569–3589
156. Panjarian S, Iacob RE, Chen S, Engen JR, Smithgall TE (2013) Structure and dynamic regulation of Abl kinases. J Biol Chem 288(8):5443–5450
157. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of Protein Allosteric and Regulatory Sites. Bioinformatics 30(9):1314–1315
158. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13(9):831–838
159. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N (2010) RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature 464(7287):427–430
160. Rader AJ, Brown SM (2011) Correlating allostery with rigidity. Mol BioSyst 7(2):464–471
161. Rastelli G, Anighoro A, Chripkova M, Carrassa L, Broggini M (2014) Structure-based discovery of the first allosteric inhibitors of cyclin-dependent kinase 2. Cell Cycle 13(14):2296–2305
162. Reddy EP, Aggarwal AK (2012) The ins and outs of bcr-abl inhibition. Genes Cancer 3(5-6):447–454
163. Rice KD, Aay N, Anand NK, Blazey CM, Bowles OJ, Bussenius J, Costanzo S, Curtis JK, Defina SC, Dubenko L, Engst S, Joshi AA, Kennedy AR, Kim AI, Koltun ES, Lougheed JC, Manalo JC, Martini JF, Nuss JM, Peto CJ, Tsang TH, Yu P, Johnston S (2012) Novel carboxamide-based allosteric MEK inhibitors: discovery and optimization efforts toward XL518 (GDC-0973). ACS Med Chem Lett 3(5):416–421

164. Rivalta I, Sultan MM, Lee NS, Manley GA, Loria JP, Batista VS (2012) Allosteric pathways in imidazole glycerol phosphate synthase. Proc Natl Acad Sci U S A 109(22):E1428–E1436
165. Rosenzweig R, Kay LE (2014) Bringing dynamic molecular machines into focus by methyl-TROSY NMR. Annu Rev Biochem 83:291–315
166. Roskoski R Jr (2004) Src protein-tyrosine kinase structure and regulation. Biochem Biophys Res Commun 324(4):1155–1164
167. Roskoski R Jr (2015a) A historical overview of protein kinases and their targeted small molecule inhibitors. Pharmacol Res 100:1–23
168. Roskoski R Jr (2015b) Src protein-tyrosine kinase structure, mechanism, and small molecule inhibitors. Pharmacol Res 94:9–25
169. Roskoski R Jr (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol Res 103:26–48
170. Rudolph J, Xiao Y, Pardi A, Ahn NG (2015) Slow inhibition and conformation selective properties of extracellular signal-regulated kinase 1 and 2 inhibitors. Biochemistry 54(1):22–31
171. Ruff EF, Muretta JM, Thompson AR, Lake EW, Cyphers S, Albanese SK, Hanson SM, Behr JM, Thomas DD, Chodera JD, Levinson NM (2018) A dynamic mechanism for allosteric activation of Aurora kinase A by activation loop phosphorylation. eLife 7. https://doi.org/10.7554/eLife.32766
172. Ruschak AM, Kay LE (2012) Proteasome allostery as a population shift between interchanging conformers. Proc Natl Acad Sci U S A 109(50):E3454–E3462
173. Saleh T, Rossi P, Kalodimos CG (2017) Atomic view of the energy landscape in the allosteric regulation of Abl kinase. Nat Struct Mol Biol 24(11):893–901
174. Schneider EV, Bottcher J, Huber R, Maskos K, Neumann L (2013) Structure-kinetic relationship study of CDK8/CycC specific compounds. Proc Natl Acad Sci U S A 110(20):8081–8086
175. Schoepfer J, Jahnke W, Berellini G, Buonamici S, Cotesta S, Cowan-Jacob SW, Dodd S, Drueckes P, Fabbro D, Gabriel T, Groell JM, Grotzfeld RM, Hassan AQ, Henry C, Iyer V, Jones D, Lombardo F, Loo A, Manley PW, Pelle X, Rummel G, Salem B, Warmuth M, Wylie AA, Zoller T, Marzinzik AL, Furet P (2018) Discovery of Asciminib (ABL001), an Allosteric Inhibitor of the Tyrosine Kinase Activity of BCR-ABL1. J Med Chem 61(18):8120–8135
176. Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA:protein complexes. Proc Natl Acad Sci U S A 106(16):6620–6625
177. Sethi A, Tian J, Derdeyn CA, Korber B, Gnanakaran S (2013) A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. PLoS Comput Biol 9(5):e1003046
178. Shaw AS, Kornev AP, Hu J, Ahuja LG, Taylor SS (2014) Kinases and pseudokinases: lessons from RAF. Mol Cell Biol 34(9):1538–1546
179. Shen Q, Wang G, Li S, Liu X, Lu S, Chen Z, Song K, Yan J, Geng L, Huang Z, Huang W, Chen G, Zhang J (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res 44(D1):D527–D535
180. Shi L, Kay LE (2014) Tracing an allosteric pathway regulating the activity of the HslV protease. Proc Natl Acad Sci U S A 111(6):2140–2145
181. Shukla D, Meng Y, Roux B, Pande VS (2014) Activation pathway of Src kinase reveals intermediate states as targets for drug design. Nat Commun 5:3397
182. Simard JR, Kluter S, Grutter C, Getlik M, Rabiller M, Rode HB, Rauh D (2009) A new screening assay for allosteric inhibitors of cSrc. Nat Chem Biol 5(6):394–396
183. Simonetti FL, Teppa E, Chernomoretz A, Nielsen M, Marino Buslje C (2013) MISTIC: mutual information server to infer coevolution. Nucleic Acids Res 41(Web Server issue):W8–W14
184. Smock RG, Gierasch LM (2009) Sending signals dynamically. Science 324(5924):198–203
185. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, Ranganathan R (2005) Evolutionary information for specifying a protein fold. Nature 437(7058):512–518
186. Staus DP, Strachan RT, Manglik A, Pani B, Kahsai AW, Kim TH, Wingler LM, Ahn S, Chatterjee A, Masoudi A, Kruse AC, Pardon E, Steyaert J, Weis WI, Prosser RS, Kobilka BK,

Costa T, Lefkowitz RJ (2016) Allosteric nanobodies reveal the dynamic range and diverse mechanisms of G-protein-coupled receptor activation. Nature 535(7612):448–452
187. Stetz G, Verkhivker GM (2017) Computational analysis of residue interaction networks and coevolutionary relationships in the Hsp70 chaperones: a community-hopping model of allosteric regulation and communication. PLoS Comput Biol 13(1):34
188. Stevens SY, Sanker S, Kent C, Zuiderweg ER (2001) Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. Nat Struct Biol 8(11):947–952
189. Stock G, Hamm P (2018) A non-equilibrium approach to allosteric communication. Philos Trans R Soc Lond Ser B Biol Sci 373(1749):20170187
190. Stockman BJ, Kothe M, Kohls D, Weibley L, Connolly BJ, Sheils AL, Cao Q, Cheng AC, Yang L, Kamath AV, Ding YH, Charlton ME (2009) Identification of allosteric PIF-pocket ligands for PDK1 using NMR-based fragment screening and 1H-15N TROSY experiments. Chem Biol Drug Des 73(2):179–188
191. Stolzenberg S, Michino M, LeVine MV, Weinstein H, Shi L (2016) Computational approaches to detect allosteric pathways in transmembrane molecular machines. Biochim Biophys Acta 1858(7 Pt B):1652–1662
192. Suel GM, Lockless SW, Wall MA, Ranganathan R (2003) Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 10(1):59–69
193. Swain JF, Gierasch LM (2006) The changing landscape of protein allostery. Curr Opin Struct Biol 16(1):102–108
194. Szilagyi A, Nussinov R, Csermely P (2013) Allo-network drugs: extension of the allosteric drug concept to protein- protein interaction and signaling networks. Curr Top Med Chem 13(1):64–77
195. Tappan E, Chamberlin AR (2008) Activation of protein phosphatase 1 by a small molecule designed to bind to the enzyme's regulatory site. Chem Biol 15(2):167–174
196. Taylor SS, Keshwani MM, Steichen JM, Kornev AP (2012) Evolution of the eukaryotic protein kinases as dynamic molecular switches. Philos Trans R Soc Lond Ser B Biol Sci 367(1602):2517–2528
197. Taylor SS, Kim C, Cheng CY, Brown SH, Wu J, Kannan N (2008) Signaling through cAMP and cAMP-dependent protein kinase: diverse strategies for drug design. Biochim Biophys Acta 1784(1):16–26
198. Taylor SS, Kornev AP (2011) Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci 36(2):65–77
199. Taylor SS, Zhang P, Steichen JM, Keshwani MM, Kornev AP (2013) PKA: lessons learned after twenty years. Biochim Biophys Acta 1834(7):1271–1278
200. Tehver R, Chen J, Thirumalai D (2009) Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. J Mol Biol 387(2):390–406
201. Teppa E, Wilkins AD, Nielsen M, Buslje CM (2012) Disentangling evolutionary signals: conservation, specificity determining positions and coevolution. Implication for catalytic residue prediction. BMC Bioinf 13:235
202. Tillier ER, Lui TW (2003) Using multiple interdependency to separate functional from phylogenetic correlations in protein alignments. Bioinformatics 19(6):750–755
203. Tomita N, Hayashi Y, Suzuki S, Oomori Y, Aramaki Y, Matsushita Y, Iwatani M, Iwata H, Okabe A, Awazu Y, Isono O, Skene RJ, Hosfield DJ, Miki H, Kawamoto T, Hori A, Baba A (2013) Structure-based discovery of cellular-active allosteric inhibitors of FAK. Bioorg Med Chem Lett 23(6):1779–1785
204. Tsai CJ, del Sol A, Nussinov R (2008) Allostery: absence of a change in shape does not imply that allostery is not at play. J Mol Biol 378(1):1–11
205. Tsai CJ, Del Sol A, Nussinov R (2009) Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. Mol BioSyst 5(3):207–216
206. Tsai CJ, Nussinov R (2014) A unified view of "how allostery works". PLoS Comput Biol 10(2):e1003394

207. Tse A, Verkhivker GM (2015) Molecular determinants underlying binding specificities of the ABL kinase inhibitors: combining alanine scanning of binding hot spots with network analysis of residue interactions and coevolution. PLoS One 10(6):43
208. Tse A, Verkhivker GM (2016) Exploring molecular mechanisms of paradoxical activation in the BRAF kinase dimers: atomistic simulations of conformational dynamics and modeling of allosteric communication networks and signaling pathways. PLoS One 11(11):e0166583
209. Tzeng SR, Kalodimos CG (2009) Dynamic activation of an allosteric regulatory protein. Nature 462(7271):368–372
210. Tzeng SR, Kalodimos CG (2011) Protein dynamics and allostery: an NMR view. Curr Opin Struct Biol 21(1):62–67
211. Verkhivker GM (2016) Molecular dynamics simulations and modelling of the residue interaction networks in the BRAF kinase complexes with small molecule inhibitors: probing the allosteric effects of ligand-induced kinase dimerization and paradoxical activation. Mol BioSyst 12(10):3146–3165
212. Verkhivker GM (2017) Leveraging structural diversity and allosteric regulatory mechanisms of protein kinases in the discovery of small molecule inhibitors. Curr Med Chem 24(42):4838–4872
213. Vijayabaskar MS, Vishveshwara S (2010) Interaction energy based protein structure networks. Biophys J 99(11):3704–3715
214. Wang J, Zhao Y, Wang Y, Huang J (2013) Molecular dynamics simulations and statistical coupling analysis reveal functional coevolution network of oncogenic mutations in the CDKN2A-CDK6 complex. FEBS Lett 587(2):136–141
215. Weinkam P, Pons J, Sali A (2012) Structure-based model of allostery predicts coupling between distant sites. Proc Natl Acad Sci U S A 109(13):4875–4880
216. Wenthur CJ, Gentry PR, Mathews TP, Lindsley CW (2014) Drugs for allosteric sites on receptors. Annu Rev Pharmacol Toxicol 54:165–184
217. White JT, Li J, Grasso E, Wrabl JO, Hilser VJ (2018) Ensemble allosteric model: energetic frustration within the intrinsically disordered glucocorticoid receptor. Philos Trans R Soc Lond Ser B Biol Sci 373(1749):20170175
218. Whittington AC, Larion M, Bowler JM, Ramsey KM, Bruschweiler R, Miller BG (2015) Dual allosteric activation mechanisms in monomeric human glucokinase. Proc Natl Acad Sci U S A 112(37):11553–11558
219. Williamson MP, Kitahara R (2018) Characterization of low-lying excited states of proteins by high-pressure NMR. Biochim Biophys Acta, Proteins Proteomics. https://doi.org/10.1016/j.bbapap.2018.10.014
220. Wrabl JO, Gu J, Liu T, Schrank TP, Whitten ST, Hilser VJ (2011) The role of protein conformational fluctuations in allostery, function, and evolution. Biophys Chem 159(1):129–141
221. Wu P, Clausen MH, Nielsen TE (2015a) Allosteric small-molecule kinase inhibitors. Pharmacol Ther 156:59–68
222. Wu P, Nielsen TE, Clausen MH (2015b) FDA-approved small-molecule kinase inhibitors. Trends Pharmacol Sci 36(7):422–439
223. Wu WI, Voegtli WC, Sturgis HL, Dizon FP, Vigers GP, Brandhuber BJ (2010) Crystal structure of human AKT1 with an allosteric inhibitor reveals a new mode of kinase inhibition. PLoS One 5(9):e12913
224. Wylie AA, Schoepfer J, Jahnke W, Cowan-Jacob SW, Loo A, Furet P, Marzinzik AL, Pelle X, Donovan J, Zhu W, Buonamici S, Hassan AQ, Lombardo F, Iyer V, Palmer M, Berellini G, Dodd S, Thohan S, Bitter H, Branford S, Ross DM, Hughes TP, Petruzzelli L, Vanasse KG, Warmuth M, Hofmann F, Keen NJ, Sellers WR (2017) The allosteric inhibitor ABL001 enables dual targeting of BCR-ABL1. Nature 543(7647):733–737
225. Xiao Y, Lee T, Latham MP, Warner LR, Tanimoto A, Pardi A, Ahn NG (2014) Phosphorylation releases constraints to domain motion in ERK2. Proc Natl Acad Sci U S A 111(7):2506–2511

226. Xiao Y, Liddle JC, Pardi A, Ahn NG (2015) Dynamics of protein kinases: insights from nuclear magnetic resonance. Acc Chem Res 48(4):1106–1114
227. Xie T, Peng W, Liu Y, Yan C, Maki J, Degterev A, Yuan J, Shi Y (2013) Structural basis of RIP1 inhibition by necrostatins. Structure 21(3):493–499
228. Xie T, Saleh T, Kalodimos CG (2018) Gleevec can act as an allosteric inhibitor of ABL. Biophys J 114(3):1150–PosB59
229. Xu F, Du P, Shen H, Hu H, Wu Q, Xie J, Yu L (2009) Correlated mutation analysis on the catalytic domains of serine/threonine protein kinases. PLoS One 4(6):e5913
230. Yang J, Campobasso N, Biju MP, Fisher K, Pan XQ, Cottom J, Galbraith S, Ho T, Zhang H, Hong X, Ward P, Hofmann G, Siegfried B, Zappacosta F, Washio Y, Cao P, Qu J, Bertrand S, Wang DY, Head MS, Li H, Moores S, Lai Z, Johanson K, Burton G, Erickson-Miller C, Simpson G, Tummino P, Copeland RA, Oliff A (2011) Discovery and characterization of a cell-permeable, small-molecule c-Abl kinase activator that binds to the myristoyl binding site. Chem Biol 18(2):177–186
231. Yang LW, Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA, Bahar I (2006) oGNM: online computation of structural dynamics using the Gaussian Network Model. Nucleic Acids Res 34(Web Server issue):W24–W31
232. Yeang CH, Haussler D (2007) Detecting coevolution in and among protein domains. PLoS Comput Biol 3(11):e211
233. Zhang J, Adrian FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F, Guo GR, Ding Q, Okram B, Choi Y, Wojciechowski A, Deng X, Liu G, Fendrich G, Strauss A, Vajpai N, Grzesiek S, Tuntland T, Liu Y, Bursulaya B, Azam M, Manley PW, Engen JR, Daley GQ, Warmuth M, Gray NS (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding-site inhibitors. Nature 463(7280):501–506
234. Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer 9(1):28–39
235. Zhao Y, Wang Y, Gao Y, Li G, Huang J (2015) Integrated analysis of residue coevolution and protein structures capture key protein sectors in HIV-1 proteins. PLoS One 10(2):e0117506
236. Zhuravlev PI, Papoian GA (2010) Protein functional landscapes, dynamics, allostery: a tortuous path towards a universal theoretical framework. Q Rev Biophys 43(3):295–332
237. Zorn JA, Wells JA (2010) Turning enzymes ON with small molecules. Nat Chem Biol 6(3):179–188

Chapter 10
GPCR Allosteric Modulator Discovery

Yiran Wu, Jiahui Tong, Kang Ding, Qingtong Zhou, and Suwen Zhao

Abstract G protein-coupled receptors (GPCRs) influence virtually every aspect of human physiology; about one-third of all marketed drugs target members of this family. GPCR allosteric ligands hold the promise of improved subtype selectivity, spatiotemporal sensitivity, and possible biased property over typical orthosteric ligands. However, only a small number of GPCR allosteric ligands have been approved as drugs or in clinical trials since the discovery process is very challenging. The rapid development of GPCR structural biology leads to the discovery of several allosteric sites and sheds light on understanding the mechanism of GPCR allosteric ligands, which is critical for discovering novel therapeutics. This book chapter summarized different GPCR allosteric modulating mechanisms and discussed validated mechanisms based on allosteric modulator-GPCR complex structures.

Keywords GPCR · allosteric modulator · drug discovery · membrane protein structures

10.1 GPCRs Are Allosteric Signaling Systems

G protein-coupled receptors (GPCRs), also known as seven-transmembrane domain receptors, represent the largest superfamily of proteins that regulate a wide range of fundamental physiological processes. They are one of the most important and successful classes of drug targets, and about one-third of marketed drugs function by targeting GPCRs [16, 17, 52].

Y. Wu · K. Ding · Q. Zhou
iHuman Institute, ShanghaiTech University, Shanghai, China

J. Tong · S. Zhao (✉)
iHuman Institute, ShanghaiTech University, Shanghai, China

School of Life Science and Technology, ShanghaiTech University, Shanghai, China
e-mail: zhaosw@shanghaitech.edu.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_10

All GPCRs present a common arrangement of the transmembrane domain (TMD), which consists of seven-transmembrane helices (TM1–7) linked by three extracellular loops (ECL1–3) and three intracellular loops (ICL1–3). TM3 is the most conserved helix that is surrounded by other helices in the 3D space, and it plays an axis role in GPCR activation. Upon agonist binding, residues in the cytoplasmic end of TM3 switch contact from residues in TM6 to TM7, the loosen contacts between TM3 and TM6 leads to the outward movement of the cytoplasmic end of TM6 (which is the hallmark of GPCR activation), and the cytoplasmic end of TM7 moves inward due to the engagement of TM3–TM7 residues.

GPCRs are typical allosteric signal transduction systems [32]. They receive various extracellular stimuli including peptides, neurotransmitters, odorants, lipids, and photons in the traditional ligand binding site and transduce the signal across the cell membrane through the conformational rearrangement of the receptors. The intracellular signaling pathway is subsequently triggered by recruiting effectors such as various heterotrimeric G proteins, β-arrestin, and GPCR kinases (GRKs) to the G protein binding pocket.

GPCRs are highly flexible, and they have many conformational states in the membrane, which comprise the free energy landscape of each receptor [14, 34]. Ligand(s) binding at any pocket(s) could perturb the free energy landscape and reshape the surface of intracellular binding pocket and thus affect the downstream signaling [35]. If the binding pocket of a ligand is different from the endogenous ligand binding site, the ligand is considered an allosteric modulator. Through many years' research, it has been more and more believed that one ligand leads to a unique free energy landscape that indicates a new biology. For the cases in which one orthosteric ligand and one allosteric ligand bound to the same GPCR, it can be expected that different ligands binding to same GPCR will lead to new free energy landscapes and there is possibility for stabilization of new functionally relevant conformational states and activation of new downstream cascades [14, 60].

The GPCR allosteric modulation is complex, where an allosteric ligand enhances an agonist-mediated receptor response and is, therefore, defined positive allosteric modulator (PAM). On the other hand, an allosteric ligand that attenuates an agonist-mediated receptor response is known as negative allosteric modulator (NAM). Despite the two classifications, an allosteric ligand is named as a neutral allosteric modulator, or neutral allosteric ligand (NAL), if it does not affect neither receptor nor orthosteric ligand activity. Moreover, a subclass of PAMs, called ago-PAM, can behave as agonist in the absence of orthosteric ligand [15, 40]. In an allosteric system like GPCRs, the ligands binding in spatially distinct sites can affect the affinity and efficacy of the other in a reciprocal manner. As shown in Fig. 10.1, the combination of affinity allosteric modulation and efficacy allosteric modulation is quite elaborate. For example, a PAM can potentiate the downstream signaling through four different ways: (1) promoting orthosteric agonist binding affinity but not directly affecting the signaling, (2) enhancing signaling directly without affecting the orthosteric agonist binding, (3) increasing the orthosteric ligand binding affinity and increasing the signaling by itself at the same time, and (4) decreasing the orthosteric ligand binding affinity but increasing the signaling by itself. A NAM can use similar combinations to diminish downstream signaling.

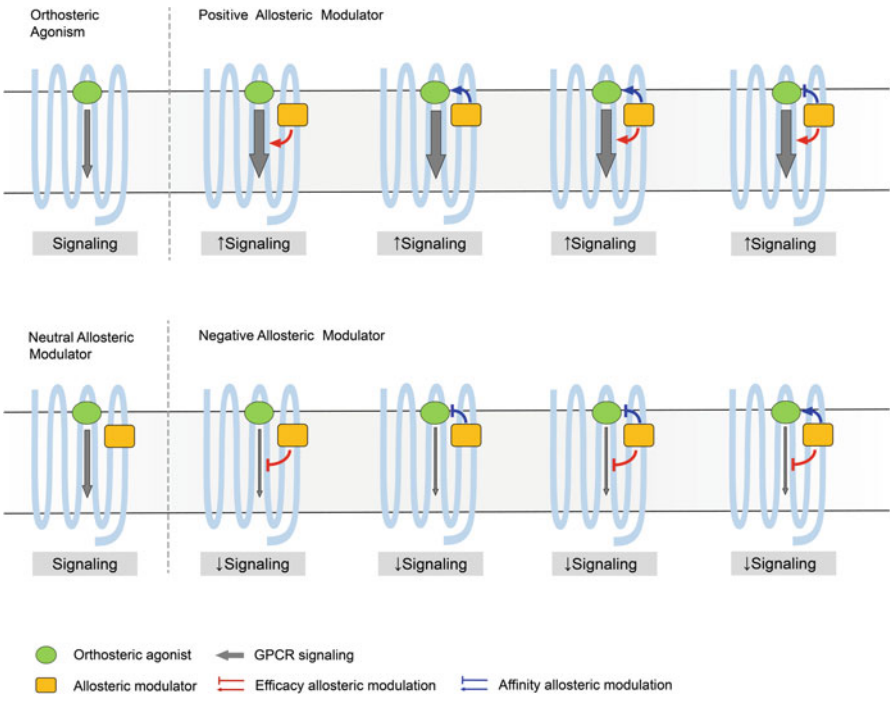


Fig. 10.1 Allosteric modulation mechanism of GPCRs. Orthosteric agonists bind to GPCRs and trigger downstream signaling (upper left). Allosteric ligands that have no effect on the binding affinity and signaling are neutral allosteric modulators (lower left). Positive allosteric modulators bind to GPCRs and enhance the signaling through either affinity allosteric modulation or efficacy allosteric modulation (top row). Negative allosteric modulators bind to GPCRs and decrease the signaling through either affinity allosteric modulation or efficacy allosteric modulation (bottom row)

10.2 Benefits of GPCR Allosteric Modulators

Due to their vital biological functions, GPCRs have been intensively researched during the past several decades. Numerous ligands, either endogenous or exogenous, natural or synthetic, small molecular or peptide, have been discovered. Some of these discoveries led to new achievements in medical field. A recent review [16] shows that 475 drugs (~34% of all drugs approved by the US Food and Drug Administration (FDA)) target at 108 unique GPCRs.

Most of these ligands bind to the traditional ligand-binding pocket which is formed extracellularly by the seven helix bundle. This pocket is referred as orthosteric pocket in most cases of class A and class B GPCRs. For class C and class F GPCRs, the endogenous ligands bind at their extracellular domain (ECD). Thus, many "allosteric" ligands for class C and class F interact with GPCRs in the traditional ligand binding pocket. Moreover, they present similar exposed versus buried surface area ratio with orthosteric ligands, while most class A and class B

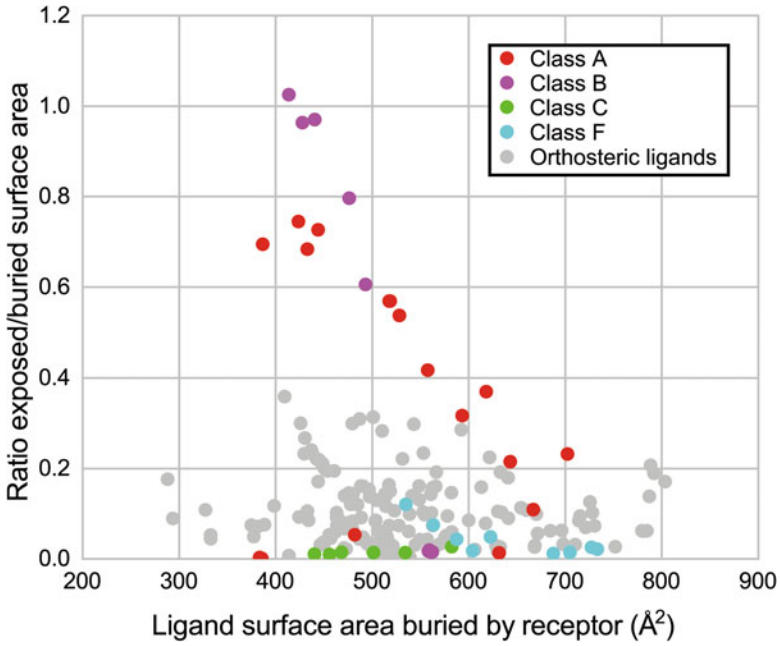


Fig. 10.2 Ratio of exposed/buried ligand surface area versus total ligand surface area buried by the receptor. Each ligand in the co-crystallized GPCR structure is a dot in the figure, with orthosteric ligand shown in gray and allosteric ligands shown in red, purple, green, and blue for class A, B, C, and F GPCRs

allosteric ligands show an elevated value for this ratio (Fig. 10.2). This book chapter is going to focus on discussing allosteric ligands of class A and class B receptors, with brief introduction of those involved in class C and class F GPCRs interaction.

Despite the remarkable success achieved in GPCR drug discovery, one of the major challenges remains the ligand selectivity. Indeed, the orthosteric binding sites within a subfamily show high sequence conservation, which allows the binding of the same/similar endogenous ligand(s). For this reason, an advantage in the use of allosteric modulators is a better receptor subtype selectivity, since the allosteric sites within a subfamily have theoretically less evolutionary pressure and are thus less conserved. Moreover, allosteric modulators may possess different subtype cooperativities due to the distinct allosteric pockets and pathways. Another advantage is that they maintain the natural spatiotemporal signaling rhythms of the endogenous orthosteric ligand when cofunctioning [15]. Besides, the cooperation between allosteric and orthosteric ligands may stabilize different conformation states of the receptor, which may lead to different function and result in biased signaling. Finally, the weak cooperativity of allosteric ligands leads to “effect ceiling” which can improve the safety of target in overdose situations [11].

10.3 Problems and Methods in Allosteric Modulator Discovery

The detection of allosteric behavior of a modulator from pharmacological experiments is still a challenging stage of the discovery procedure. Indeed, a PAM is not easily recognizable compared to an agonist, especially in high-throughput assay. The signal window of HTS assay decreases in the presence of the orthosteric ligand and increases the interference from noise. Meanwhile, the magnitude of the signal produced by PAMs is far less than that produced by a full agonist, in which case weaker PAMs may not show significant effect [3]. A NAM can be recognized if the signal decreases while orthosteric agonist's potency is unchanged. Otherwise, its behavior does not significantly differ from an orthosteric antagonist. Indeed, the pharmacological behavior of an allosteric modulator can be even more complicated. For example, ORG 27569, an allosteric modulator of cannabinoid receptor CB1, makes orthosteric agonist CP55,940 acting like an inverse agonist in cell signaling assay, but it actually increases the binding of CP55,940 [49].

Currently, triple-read assays have been developed as an effective HTS approach to detect allosteric modulators [26, 45, 51]. The first read tests ligand agonist activity, whereas the second read tests PAMs activity, and the final read measures the inhibition ability of tested compounds to an EC80 orthosteric agonist [3].

For many GPCRs, the binding site(s) determination of their allosteric modulator (s) is notoriously difficult. The design of point mutation is generally the most common method used to identify specific residues able to interact with the modulators. However, the consequences of mutations are often unpredictable, and false positives are often observed; therefore, further analysis should be included to interpret their impact on modulators. As an example, mutations of residues in different positions of the protein 3D structure have been reported to abolish the effect of ORG 27569 on cannabinoid receptor CB1 [53, 55]. A more reliable strategy would be the introduction of a covalent bond between the modulator and the receptor. For instance, a similar strategy has been used in PAMs of glucagon-like peptide 1 receptor [5, 46]. But modification of the modulator is limited by its chemical structure, and the rational design process is not always straightforward and could be very time consuming. Excitingly, solving the complex structure has been the most successful way to identify binding site and pose of allosteric modulators in GPCRs since 2013 [33]. So far, all the complex structures were obtained by crystallographic methods. Recently the fast developing cryo-electron microscopy demonstrated to be a powerful method to obtain complex structures at the atomic level. The study of allosteric modulators also helps drug discovery against GPCRs. Indeed, allosteric modulators of GPCRs in a number of classes (A, B1, C, and F) have been approved as drugs or tested in clinical trials in multiple cases.

With the development of structural biology, many computational methods based on structures spring up in allosteric site identification. For example, Allosite, a method for predicting allosteric sites developed by Zhang Jian et al., uses elegant algorithms such as pocket-based analysis and support vector machine (SVM)

classifier to predict the location of allosteric sites in proteins [23]. FTMAP, another computational method developed by Sandor Vajda et al., globally samples the surface of a target protein using small organic molecules as probes to discover favorable binding positions [44]. Furthermore, Miao Y et al. combined FTMAP and accelerated molecular dynamics (aMD) simulation, successfully mapping allosteric sites in activation-associated conformers of the M2 muscarinic receptor [41].

Visual screening also has successful examples in finding allosteric modulators in known allosteric sites. Miao Y et al. performed long-time scale aMD simulation on M2 muscarinic acetylcholine receptor to construct structural ensembles. During virtual screening, they combined induced-fit docking and ensemble docking. Eventually, they successfully identified four new NAMs and one PAM for M2 mAChR [42].

Given the fact that noteworthy reviews about GPCRs allosteric regulation are already available in literature, this book chapter aims to summarize different GPCR allosteric modulating mechanisms and discuss validated mechanisms based on allosteric modulator-GPCR complex structures.

10.4 Allosteric Modulating Issues for GPCRs in Different Orthosteric Binding Mechanisms

Allosteric modulating issues are distinct issues when the orthosteric sites on GPCRs are at different positions.

10.4.1 The Classic Orthosteric Binding Mechanism

The most common mechanism indicates that the orthosteric site locates among the transmembrane helices close to the extracellular end. This happens in most class A (the largest class) GPCRs. Although class B1 GPCR has an extracellular domain at the N-terminal (called hormone receptor domain) critical for endogenous peptide ligand binding, it also shares orthosteric site mechanism, due to the fact that the peptide ligand binding site extends into the pocket among the transmembrane helices.

For receptors following this mechanism, the orthosteric site must be elsewhere than the pocket among the transmembrane helices close to the extracellular end.

10.4.2 Special Orthosteric Binding Mechanism: Extracellular Domain

Several groups of GPCRs bind to endogenous ligands only with extracellular domains. (1) Class C GPCR has two extracellular domains, a Venus flytrap domain and nine-cysteine domain, and it uses the former to recognize endogenous small-molecule ligand. (2) Class F receptor has an N-terminal Fz domain (or called cysteine-rich domain) responsible for binding to WNT protein binding in frizzled receptors (PDB ID: 4F0A [27]) or for the interaction with steroids in smoothened receptor (both functional [65] and structural evidences, PDB ID: 5L7D [6]). (3) There are eight evolutionarily related receptors in class A GPCRs, each having a leucine-rich repeat-containing domain. All the five members with known functions, including three protein receptors and two peptide receptors, have been validated to bind to the endogenous ligands only with the extracellular leucine-rich repeats-containing domains. The other three members (leucine-rich repeat-containing G-protein coupled receptors 4, 5, and 6) remain orphan (for their ligands inducing G protein coupling/pathway are not found). Nevertheless, they have been reported to function in WNT signaling pathway by binding to R-spondin solely with their extracellular domains.

For GPCRs with this extracellular domain binding mechanism, the site among the transmembrane helices, proximal to the extracellular end, is possibly involved in allosteric modulating. During long evolution, the diverse ligand types of GPCRs have proved this site to be ideal for ligand binding. Therefore, discovering allosteric modulators at this site should be relatively easy. This is in accordance with the published statistic data of allosteric modulators: class C GPCRs contribute to more than half of the modulators; smoothened and the three leucine-rich repeats-containing protein receptors (follicle-stimulating hormone receptor, lutropin-choriogonadotropic hormone receptor, thyrotropin receptor, and glycoprotein hormone receptor) in class A GPCRs have multiple allosteric modulators, as reported in Fig. 10.3.

10.4.3 Special Orthosteric Mechanism: Proton Sensing

Four receptors (GPR4, GPR65, GPR68, and GPR132) in class A GPCRs sense pH and are involved in cancer cell regulation [30]. The recognition of acidic pH is achieved by protonation of histidine residues in the extracellular region [43, 59]; thus, the pocket among transmembrane helices is not the orthosteric site and is potentially involved in allosteric modulating. Allosteric modulators of GPR4, GPR65, and GPR68 have been reported [24, 59].

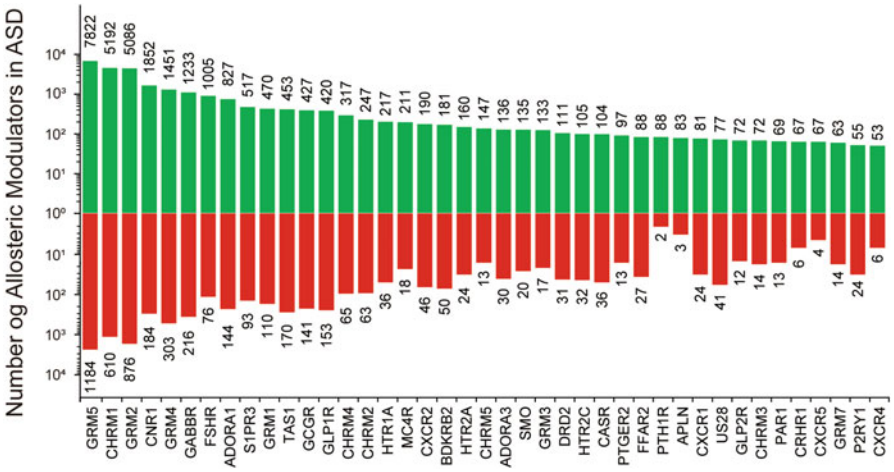


Fig. 10.3 Allosteric ligands of GPCRs in Allosteric Database (ASD). Green bars indicate the amount of allosteric modulators of GPCRs in ASD; red bars indicate the clusters of GPCRs when all ligands are clustered with a similarity cutoff of 0.60

10.5 Many GPCR Allosteric Modulators Have Been Reported

10.5.1 Basic Statistics of Reported GPCR Allosteric Modulators in The Allosteric Database

The Allosteric Database 16-18 (ASD, http://mdl.shsmu.edu.cn/ASD/) [22] records 30712 allosteric modulating on 92 different GPCRs. Most modulators are organic compounds (29694), and the rest are polypeptides (409) and ions (7). The data was manually extracted from 729 citations, including 545 journal articles and 184 patents. However, patents contribute more than half of the modulating data (18472/30712, 60%).

Among all 30712 modulating records, 18861 are recorded as "activator," 6010 as "inhibitor," and 5841 as "regulator" for the impact on activity is not clear.

10.5.2 Target Analysis of Reported Allosteric Modulating

For targets that have more than 50 allosteric modulators reported in ASD, the distribution of allosteric modulators is shown in Fig. 10.3 and representative allosteric ligands for top 27 GPCRs with most allosteric ligands are shown in Fig. 10.4. The number of modulators for each GPCR is largely different: each of the top three receptors (metabotropic glutamate receptor 5, muscarinic acetylcholine receptor M1,

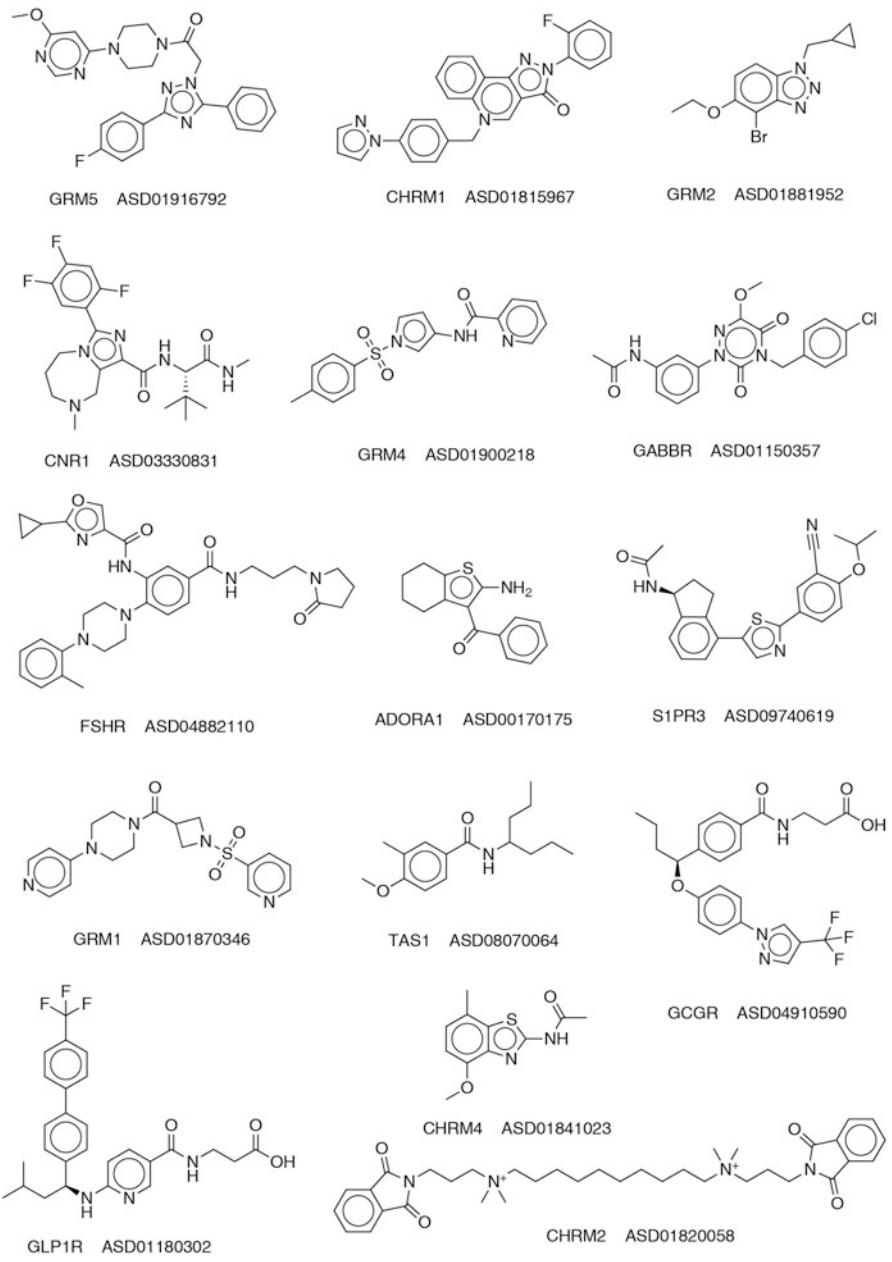


Fig. 10.4 Representative allosteric ligands for top 27 GPCRs with most allosteric ligands

and metabotropic glutamate receptor 2) has more than 5000 modulators, and they contribute to 59% of all the modulating records in total; the top ten receptors have over 500 modulators in average and contribute to 83% of all the modulating records. This high inequality of the number of reported modulators for different GPCR is

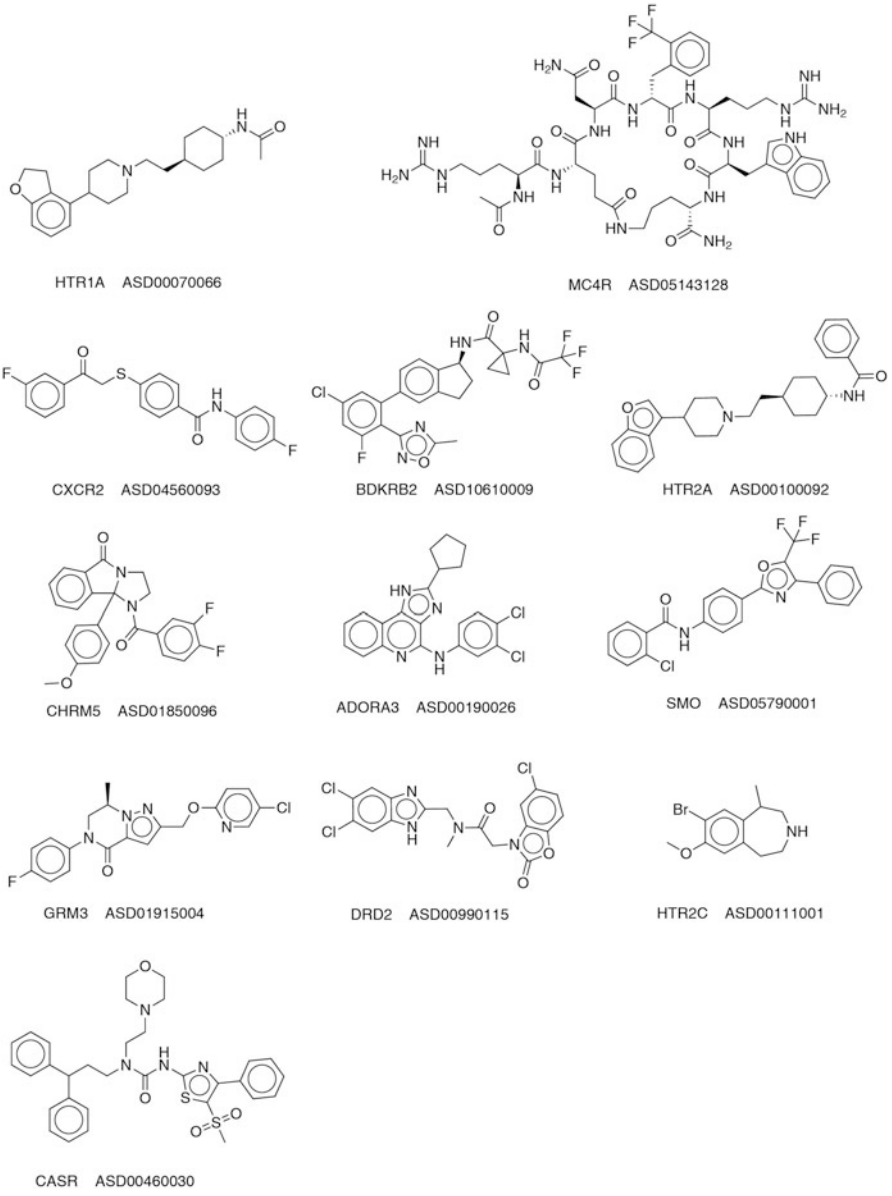


Fig. 10.4 (continued)

caused by three major reasons: (1) some GPCR members can easily target in allosteric sites; (2) some GPCR members are highly focused in research, and thus their allosteric modulators are more likely to be discovered; (3) after an allosteric

modulator is identified, derivatives from this parent compound would be soon developed and tested on the same target.

Among the 92 GPCR targets registered in ASD, the majority (71) belong to the class A receptors (the rhodopsin family), while class C (the metabotropic glutamate family, 11) and class B1 (the secretin family, 6) have fewer members. However, the class C receptors contribute to more than half of the modulating molecules (16858/30712, 55%). Binding to endogenous ligand using the extracellular domain, and keeping the pocket among transmembrane helices, an allosteric site easy to target is the main reason for having a considerable amount of allosteric modulators from class C GPCRs reported in literature. Besides, these receptors show a role related to the specific functions in the nervous system, a fact that makes them widely studied. In class A GPCRs, the most notable group is the muscarinic acetylcholine receptors (CHRMs) 1–5: all five members are on the list, and CHRM1 has the largest number of modulators (5200) in class A receptors.

10.6 GPCR Allosteric Binding Sites Identified by Crystallography

The breakthrough of GPCR structural biology in the past decade has resulted ~300 structure for more than 50 receptors. Although most of these structures are in complex with orthosteric ligands, more and more allosteric-receptor complexes become available in PDB. In total, the structures of 32 unique allosteric ligand-receptors complexes have been solved (Table. 10.1). Detailed analyses for binding sites of these structures are in the following.

10.6.1 Lipidic Interface

As shown by many structures, GPCRs can bind allosteric ligands using the extra-helical region, between the interface of GPCR and membrane lipids. So far, four different binding sites at GPCR lipidic interface have been identified by crystal structures, and they are referred as AS3, AS4, AS5, and AS6 (Table. 10.1, Fig. 10.5 and 10.6). Although the predominant interactions are hydrophobic, at least one hydrogen bond can be identified in each case. Therefore, hydrogen bond donor/acceptor atoms exposed at the lipidic interface are likely to be hot spots for such allosteric ligand binding. However, these binding sites are usually quite shallow compared to typical sites in soluble proteins and the orthosteric binding pockets in GPCRs, which make allosteric drug discovery targeting pockets in protein-lipid interface extremely difficult. Even after the validation of the binding site, the discovery of a ligand using molecular docking remains challenging. This is because many scoring functions were developed for pockets exposed to water environment,

Table 10.1 Crystal structures of GPCRs in complex with allosteric ligands

Class	GPCR	Ligand name	Site	Ligand function	PDB ID
Class A	CCR2	CCR2-RA-[R]	AS2	NAM	5T1A [70]
	CCR5	Maraviroc	AS1	NAM	4MBS [57]
		Compound 21	AS1	NAM	6AKX [48]
		Compound 34	AS1	NAM	6AKY [48]
	CCR9	Vercirnon	AS2	NAM	5LWE [47]
	ACM2	LY2119620	AS1	PAM	4MQT [33]
	P2RY1	BPTU	AS3	NAM	4XNV [67]
	ADRB2	Cmpd-15	AS2	NAM	5X7D [36]
	C5AR1	NDT9513727	AS5	NAM	6C1Q [37], 5O9H [50]
		Avacopan	AS5	NAM	6C1R [37]
	PAR2	AZ3451	AS4	NAM	5NDZ [8]
	FFAR1	Ap8	AS5	PAM	5TZY [38]
		Compound 1	AS5	PAM	5KW2 [18]
Class B	CRFR1	CP-376395	AS7	NAM	4K5Y [19], 4Z9G [13]
	GCGR	MK-0893	AS6	NAM	5EE7 [28]
		NNC0640	AS6	NAM	5XF1,5XEZ [68]
	GLP1R	PF-06372222	AS6	NAM	5VEW [54]
		NNC0640	AS6	NAM	5VEX [54]
Class C	GRM1	FITM	AS8	NAM	4OR2 [64]
	GRM5	Mavoglurant	AS8	NAM	4OO9 [12]
		Compound 14	AS8	NAM	5CGC [9]
		HTL14242	AS8	NAM	5CGD [9]
		Fenobam	AS8	NAM	6FFH [10]
		MMPEP	AS8	NAM	6FFI [10]
Class F	SMO	LY2940680	AS8	NAM	4JKV [61]
		SANT-1	AS8	NAM	4N4W [61]
		Cyclopamine	AS8	NAM	4O9R [63]
		SAG1.5	AS8	PAM	4QIN [61]
		ANTA XV	AS8	NAM	4QIM [61]
		Vismodegib	AS8	NAM	5L7I [6]
		TC114	AS8	NAM	5V56, 5V57 [69]
	xSMO	Cyclopamine	AS8	NAM	6D32 [25]

so the empirical and/or statistical terms are not suitable for pockets embedded in membrane.

By analyzing these cases, it is possible to obtain some valuable and comprehensive understanding of the mechanisms of these modulators.

Three out of four lipidic interface allosteric sites are for NAM binding (AS3, AS4, and AS6 in Fig. 10.6), whereas the other one is for both PAM and NAM binding (AS5 in Fig. 10.6). Interestingly, all the NAMs were not identified as allosteric modulators but as antagonists/inverse agonists, until their binding sites were determined by crystallography (except for AZ3451 in protease-activated receptor 2, firstly

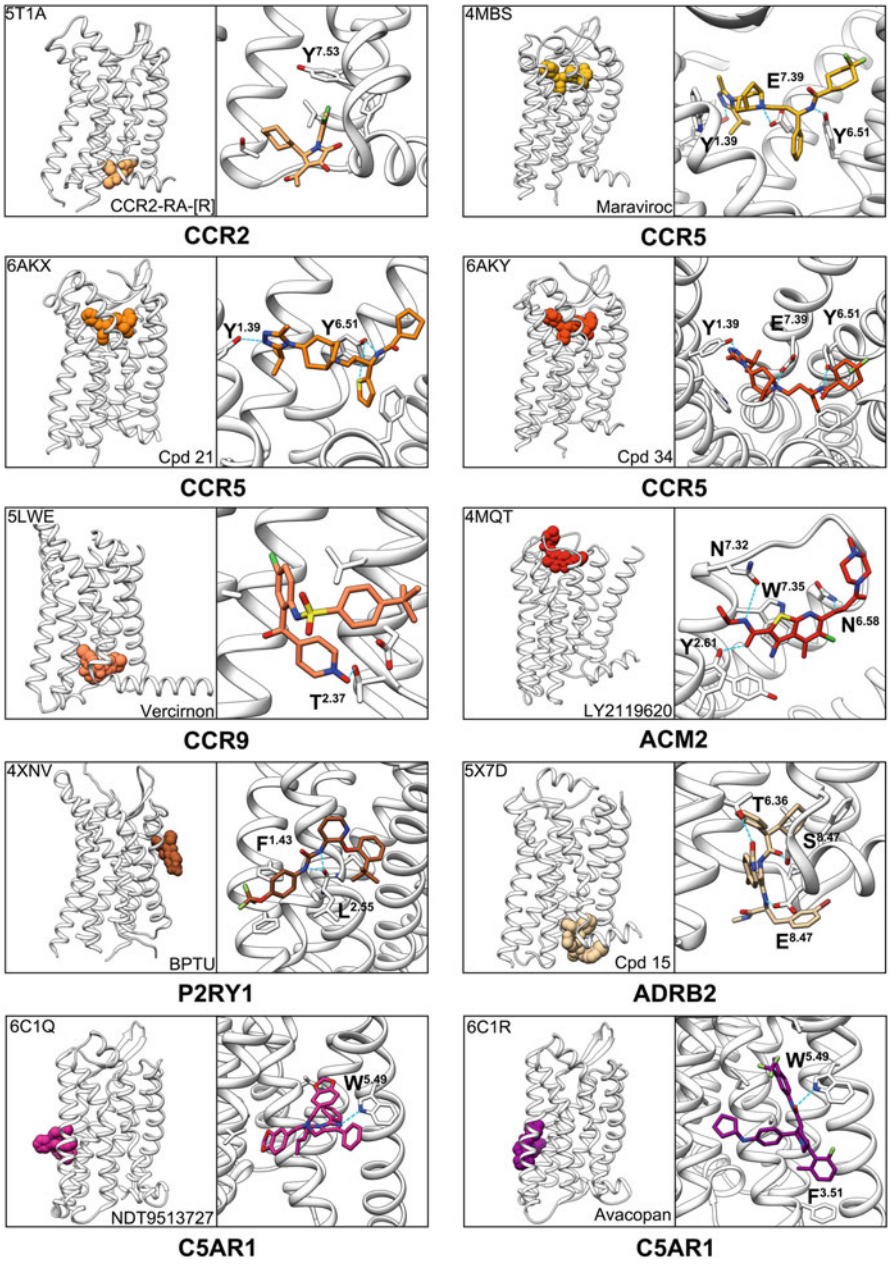


Fig. 10.5 Binding sites and interactions with receptors for all co-crystallized allosteric ligands and bitopic ligands of GPCR. For each ligand, left figure shows binding pocket. Upper left of left figure shows PDB ID and lower right shows ligand name. Right figure shows interactions between receptor and ligand. Blue dotted lines indicate hydrogen bonds. GPCRs in different classes are shown in different color: class A in black, class B in purple, class C in green, and class F in blue. Receptors of bitopic ligands are shown in gray

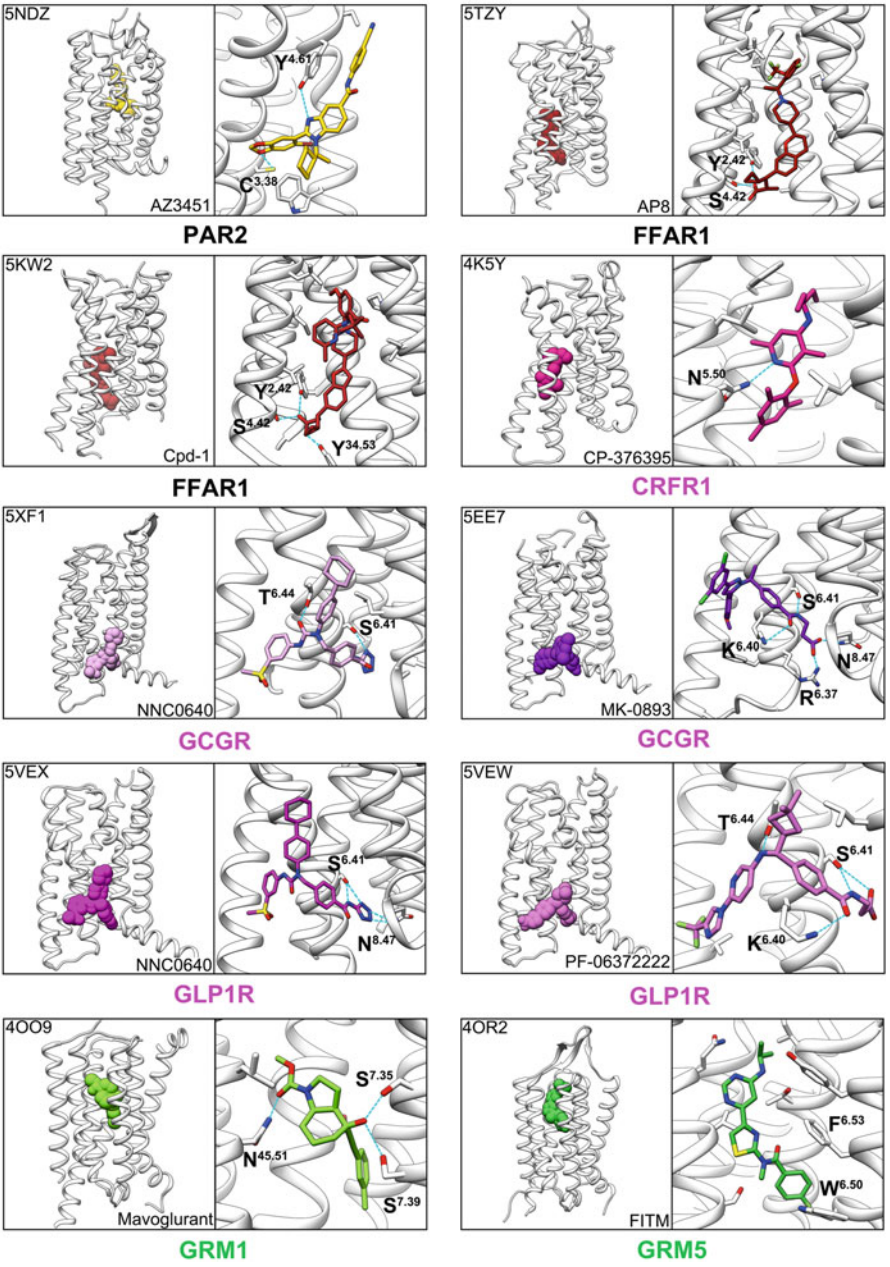


Fig. 10.5 (continued)

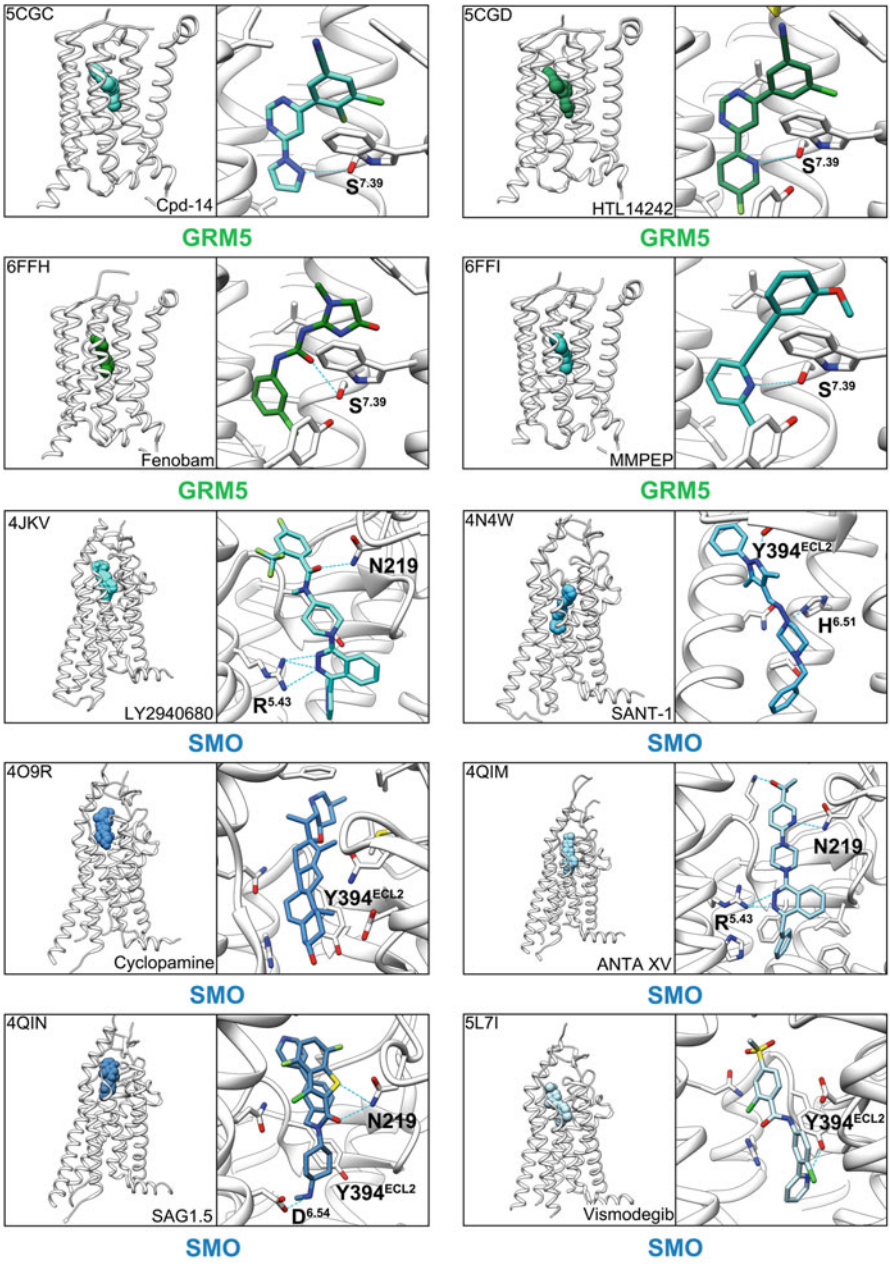


Fig. 10.5 (continued)

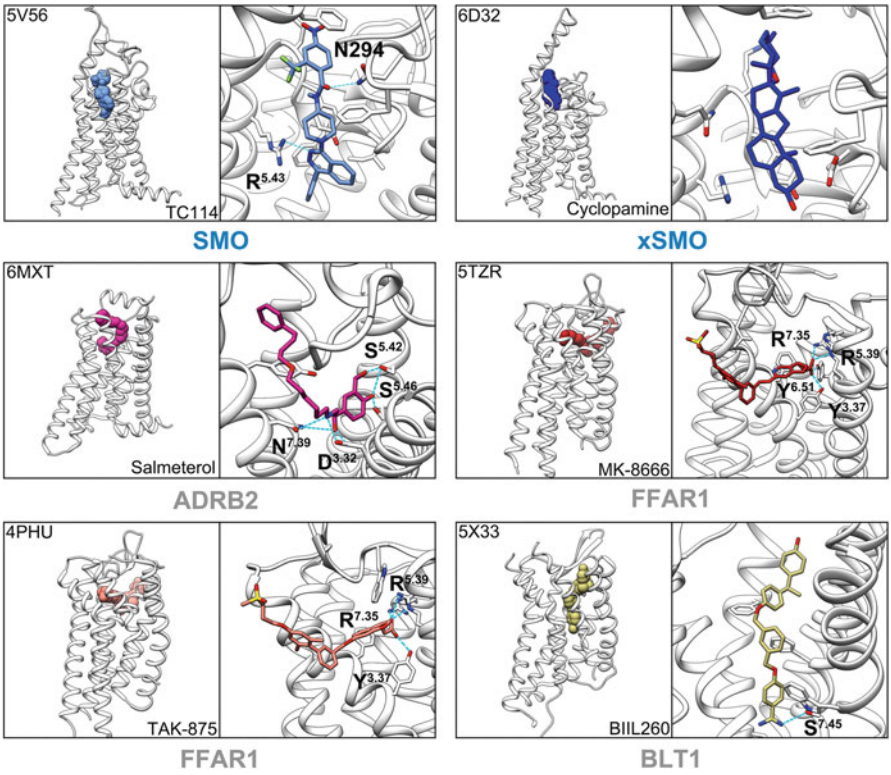


Fig. 10.5 (continued)

reported with the crystal structure [8]). Among these NAMs, NDT9513727 in C5a anaphylatoxin chemotactic receptor 1 [50] was even tested in radioligand binding experiments in previous work and was found to have the same effect as competitor to the endogenous ligand [4]. This phenomenon firmly reveals that the determination of NAMs by pharmacological experiments is not trivial.

The positions of the allosteric sites are quite dispersed: two in four sites stay close to the extracellular end and the other two close to the cytoplasmic end. Although all transmembrane helices are likely involved in allosteric modulator binding, the distribution of allosteric sites around the receptor is unequal: from the side view, it's possible to conclude that AS3–5 (Fig. 10.6a) is mainly concentrated on one side (TM1-2-3-4-5), while only AS6 (Fig. 10.6b) stands at the other side of the structure (TM5-6-7-1). There are two possible reasons for this inequality: (1) TM2–4 generally remain stable during activation, while TM5–7 are much more flexible; thus, it is easier for modulators to bind to TM2–4. Another reason lays in the fact that modulators, interacting with TM2–4, stabilize the helices and are easier to be crystalized. (2) Only a few GPCR allosteric sites have been found, and these new sites interacting with TM6–7 may remain unrevealed.

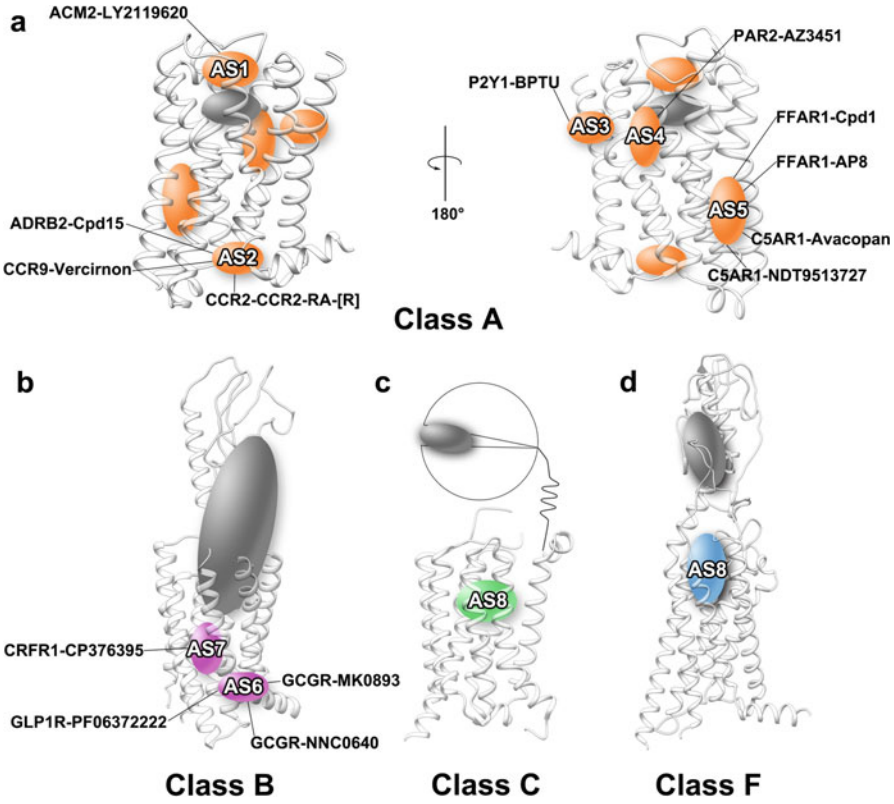


Fig. 10.6 Summary of allosteric sites for all co-crystallized allosteric ligands of GPCRs. Orthosteric ligands are shown in gray. Allosteric site (AS) of each class are numbered and shown in different color, respectively: class A in orange, class B in purple, class C in green, and class F in blue

Unlike AS3–5 that were identified from receptors in the rhodopsin family, AS6 keeps unique for its specific distribution in secretin family (Fig. 10.6b). Two ago-PAMs of glucagon-like peptide 1 receptor, compound 2 and BTEP, covalently link to Cys347 close to this site [5, 46], the ago-PAM effect of which was investigated by performing MD simulations. PF-06372222 (NAM) stabilizes the ionic interactions among TM2-3-6-7 and blocking TM6–7 separation which is required for receptor activation. Compound 2 (PAM) induces a conformational change in the intracellular regions of TM5 and 6 that results in disruption of the intracellular ionic lock, leading to a cavity at a similar location for G protein binding in class A GPCRs [54].

AS5 is an interesting site since both NAM and PAM can bind in this pocket, even though the interacting segments are different. AS5 resides at a position TM4-3-5 packing more tightly in active state than in inactive state; therefore, horizontally extending modulators such as NDT9513727 and avacopan of C5a anaphylatoxin

chemotactic receptor 1 (Table. 10.1, Fig. 10.5 and 10.6b) stabilizing the loosely packing conformation are NAMs, while vertically extending modulators such as AP8 and compound 1 in FFAR1 (Table. 10.1, Fig. 10.5 and 10.6b) stabilizing the tightly packing conformation are PAMs. In addition, the ternary complex of FFAR1/MK-8666/AP8 [38] provides a very special case of a receptor simultaneously binding to two allosteric modulators.

BPTU in AS3 (Fig. 10.5 and 10.6b) of P2Y purinoceptor 1 (P2RY1) [67] is special since it forms two hydrogen bonds to the backbone carbonyl of Leu102$^{2.55}$ (Fig. 10.5). The unpaired backbone carbonyl locates in the bulge on helix II, which is caused by Pro105$^{2.58}$. This case suggests a possible universal mechanism for lipidic allosteric modulators: using bulges on transmembrane helices as an anchor for binding. Pro$^{2.58}$ is conserved in 74% (nonolfactory, the same for the below in this section) class A GPCRs. There are only three other proline residues conserved in more than 50% class A GPCRs, 4.60, 6.50, and 7.50, but none of them faces to the lipidic interface. Therefore, Pro$^{2.58}$ and other nonconserved proline residues that face the lipidic interface are possible anchors for allosteric modulators.

10.6.2 Extracellular Interface

The allosteric modulator binding close to the orthosteric ligand, at the extracellular vestibule of a GPCR, is a straightforward mechanism. LY2119620 in AS1 (Fig. 10.6a) of muscarinic acetylcholine receptor 2 (CHRM2) [33] is the only example for this case. Many analogs of LY2119620 reported as PAMs to CHRM 1, 2, and 4 have proved these sites possessing potential universal distribution in muscarinic acetylcholine receptors.

The muscarinic acetylcholine receptors are highly concentrated in allosteric modulator study. The endogenous ligand, acetylcholine, is a tiny quaternary ammonium and small in size. The dominant interaction is cation-π between the quaternary ammonium cation of acetylcholine and conserved aromatic residues in the receptors. The binding site of acetylcholine has three aromatic residues as ceiling: Tyr$^{3.33}$, Tyr$^{6.51}$, and Tyr$^{7.39}$. The three tyrosine residues isolate the orthosteric binding site from the allosteric pocket above it, which locates in the extracellular vestibule and exposes to water. It is highly possible that this site is responsible for the thousands of reported allosteric modulators (both NAMs and PAMs) for muscarinic acetylcholine receptors.

Most GPCRs do not have separate pockets above the orthosteric site, but allosteric binding is still possible. Unidentified density other than orthosteric ligand at this position was observed in crystal structure of orexin receptor type 2 (PDB ID: 5WQC [56]) and assigned as polyethylene glycol, just like such unidentified density in CHRM4 without allosteric modulator (PDB ID: 5DSG [58]).

10.6.3 Cytoplasmic Interface

GPCRs use their cytoplasmic interface to bind to downstream partners including G proteins and arrestins. This interface can also be targeted by allosteric modulators and is referred as AS2, which have been proved by three cases (Table. 10.1, Fig. 10.5 and Fig. 10.6a). In the three cases, the small-molecule modulators are all NAMs stabilizing inactive states of the receptors and blocking G protein recruitment. The NAMs all interact with TM2, 7, and helix 8. They show π-π interaction (parallel or T-shaped) with $Y^{7.53}$, and this interaction blocks the $NP^{7.50}$xxY motif in TM7 to move which is necessary for activation.

As this cytoplasmic NAM mechanism is found in receptors with distant evolutionarily relationship (aminergic and chemokine receptors), it is possible that the mechanism is universal to class A GPCRs. However, this strategy has not yet been considered in drug design due to lower selectivity for these sites among different GPCRs.

High-affinity nanobodies were also developed to bind in this pocket, and they showed ability to stabilize multiple states of the receptors (Fig. 10.7), including active state (PDB ID: 3P0G), inactive state (PDB ID: 5JQH), and partial active state (PDB ID: 6MXT).

10.6.4 Classic Orthosteric Site

For a GPCR using other positions to recognize the endogenous ligand, the traditional binding site of GPCRs (referred as AS8 in Fig. 10.6) is potentially a good pocket for allosteric modulators. As stated in Sects. 10.2 and 10.3, this allosteric modulating mechanism has been found in class C receptors, smoothened receptor, leucine-rich repeat-containing protein receptors, and GPR65/68, with class C and smoothened receptors having available crystal structures.

For class C GPCRs, complex structures of metabotropic glutamate receptor 1 (GRM1) with one allosteric modulator and metabotropic glutamate receptor 5 (GRM5) with five different modulators have been obtained. All the modulators are NAMs. All the NAMs of GRM5 are highly overlapped and form similar interactions to TM2, 3, 5, 6, and 7. The NAM of GRM1, FITM (Table. 10.1), binds at a shallower position and interacts with ECL2.

For smoothened receptor, seven TM modulators (7 NAMs and 1 PAM) are observed by crystal structures (Table. 10.1). The depth of the NAMs largely varies, with cyclopamine and SANT-1 representing too extreme cases. The position of the only PAM, SAG1.5 (Table. 10.1), occupies about the same depth as that of the shallowest NAMs. However, the structures fail to display active/inactive conformational transition. Vismodegib (a crystalized NAM) is an FDA-approved drug to treat basal cell carcinoma. The frizzled receptors share the same class with smoothened receptor and some of them have been proved to be involved in cancer onset.

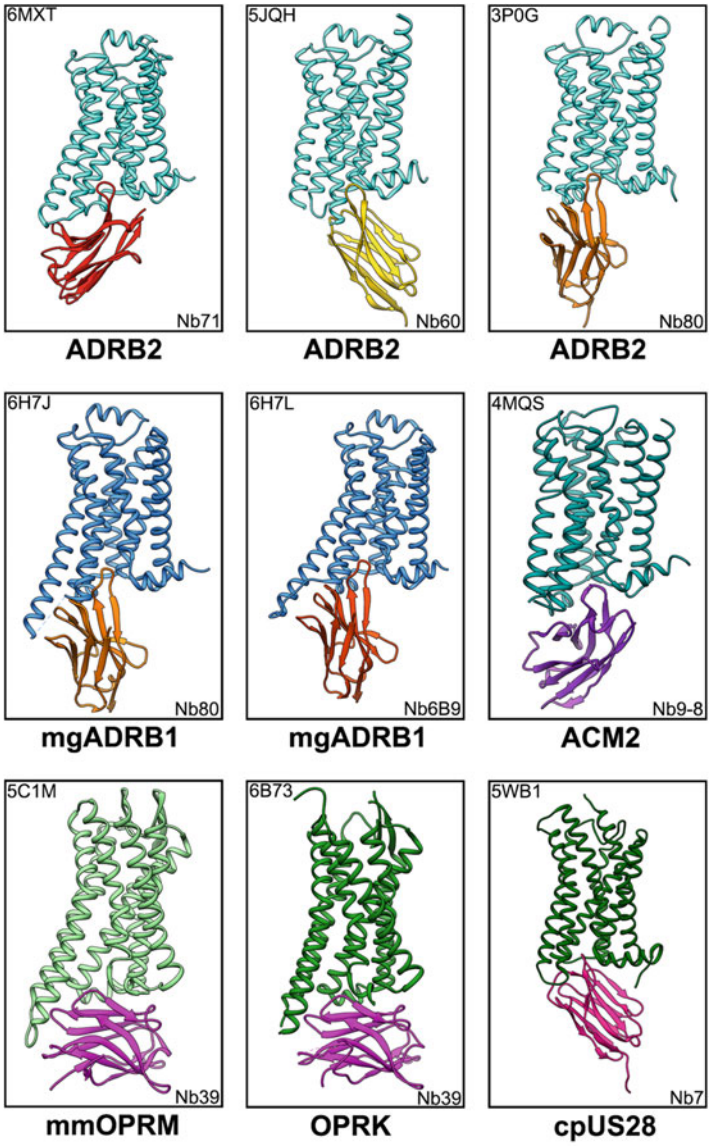


Fig. 10.7 Nanobody-stabilized structures of GPCR. For each structure, upper left shows the PDB ID, and lower right shows name of nanobody. The same receptors and nanobody are in the same color

Unfortunately, according to the recent discovery provided by their crystal structure of the first frizzled receptor TMD domain, it has been found that the hypothetical binding pocket is very narrow and hard to accommodate small molecules (PDB ID: 6BD4 [66]).

10.6.5 Na^+ and Na^+ Binding Pocket

In addition to the allosteric sites occupied by G proteins and other GPCR effectors, Na^+ binding site formed by D2.50, S3.39, N7.45, and N7.49 is highly conserved in class A GPCRs. The pocket has been reported in dozens of high-resolution crystal structures including adenosine A2a (PDB ID: 4EIY [62]) and protease-activated receptor 1 (PDB ID: 3VW7 [7]). Na^+ stabilizes the inactive state of GPCRs, and agonist initiates GPCR activation by triggering the collapse of Na^+ binding pocket [31] and the inward movement of TM7 toward TM3. The possibility to design bitopic ligands which occupy both Na^+ pocket and traditional orthosteric binding pocket via connection of two pharmacophores is possible due to the proximity of the two pockets. The pharmacophores of orthosteric sites in many GPCRs are quite well known; in contrast, there are limited choices for pharmacophores in the small and polar Na^+ pocket. As a result, it is convenient to generate possible bitopic ligands for many class A GPCRs by linking orthosteric pharmacophore and several Na^+ mimic pharmacophores. The bitopic ligands occupying both orthosteric pocket and Na^+ pocket are likely to be antagonist since they prevent the pocket to collapse which is essential for GPCR activation.

10.6.6 Bitopic Ligands

Bitopic ligands have been reported to target both orthosteric and allosteric sites, so that they can provide higher selectivity and binding affinity by taking advantages from both the types of ligands. Three receptors, free fatty acid receptor 1, leukotriene B4 (LTB4) receptor, and beta2AR, have been co-crystalized with their bitopic ligands TAK-875 (PDB ID: 4PHU [1]) or MK-8666 (PDB ID: 5TZY [38]), BIIL260 (PDB ID: 5X33 [21]), and salmeterol (PDB ID:6MXT [39]), respectively (Fig. 10.5). The structure of BLT1 shows that the amidine group of BIIL260 occupies the Na^+ pocket and interacts with $D^{2.50}$, stabilizing the inactive state of BLT1 (Fig. 10.5), while salmeterol in B2AR extends to the exosite where B1AR and B2AR are distinct. CP-376395 in corticotropin-releasing factor receptor 1, a class B receptor (PDB ID: 4K5Y in Fig. 10.5), shows a unique allosteric binding site (referred as AS7 in Fig. 10.6c) formed by TM3, TM5, and TM6.

10.7 Concluding Remarks

Allosteric regulation of GPCRs has substantial implications in both fundamental and translational biological research. Over the past decade, studies regarding the allosteric regulation of GPCRs have been flourishing. Many allosteric modulators have been identified for different classes of GPCRs, with a large number of modulators

having unknown binding sites. As a result, many questions regarding functions and mechanisms of these allosteric modulators remain less well understood due to lack of structural information. Till today, only 32 novel allosteric modulator-GPCR complex structures with 8 different allosteric sites have been solved, and the best method to identify novel binding sites of allosteric ligands is still X-ray crystallography. Effective and efficient approaches, either experimental or computational, for accurately locating allosteric sites of GPCRs with/without structures are needed in the future. Furthermore, identifying binding pockets in lipidic environment requires extra attention, since 4 out of 8 discovered allosteric ligand binding sites in GPCRs are exposed to lipidic environment. With more and more allosteric sites discovered, bitopic ligands that target two sites (either orthosteric-allosteric or two allosteric) through linking the pharmacophores of the sites can be readily developed as tool compounds and therapeutic agents. While there are obstacles to identify allosteric binding pockets, to screen allosteric modulators from high-throughput functional assays, to design allosteric modulators through computer-aided drug discovery, etc., progresses have been made for each of these obstacles during the past 10 years. GPCR allosteric ligands still hold the promise to reach better subtype selectivity, spatiotemporal sensitivity, and possible biased property.

Glossary

Transmembrane domains (TMD) the core domain that is common to all GPCRs, consisted by seven-transmembrane helices (TM1–7) that are linked by three extracellular loops (ECL1–3) and three intracellular loops (ICL1–3).

Agonist a molecule that binds to and activates a receptor to increase the receptor signaling over basal levels.

Antagonist a molecule that binds to a receptor and inhibits agonist signaling.

Positive allosteric modulator (PAM) an allosteric ligand that potentiates an agonist-mediated receptor response is referred to as a positive allosteric modulator.

Negative allosteric modulator (NAM) an allosteric ligand that attenuates an agonist-mediated receptor response is referred as a negative allosteric modulator.

Neutral allosteric ligand (NAL) an allosteric ligand that does not affect receptor or orthosteric ligand activity is referred as a neutral allosteric ligand.

Ago-PAM an allosteric ligand that is capable of directly activating the receptor from an allosteric site even in the absence of an orthosteric agonist is referred as an ago-PAM.

Bitopic ligand a hybrid molecule that occupies both orthosteric and allosteric sites to mediate a novel pharmacology.

GPCR generic numbering systems sequence-based generic GPCR residue numbering systems exist for class A (Ballesteros-Weinstein numbering) [2], class B

(Wootten numbering) [20], class C (Pin numbering) [29], and class F (Wang numbering) [61]. A numbering is two numbers separated by a dot. The first number denotes the transmembrane helix (1–7). For the second number, the most conserved position in the helix is assigned as 50. That is, the most conserved residue in a helix x is denoted x 50, while all other residues on the same helix are numbered relative to it. For example, 2.50 denotes the most conserved residue in transmembrane helix 2, while one position after it is 2.51.

Orthosteric binding site where endogenous ligands bind. For GPCR there are different orthosteric binding site in different classes. In class A and B, the orthosteric site locates among the transmembrane helices close to the extracellular end. In class C and D, the orthosteric site locates in extracellular domain.

Allosteric binding site sites for ligand binding to a receptor that are remote from the orthosteric binding site.

References

1. Ankita S, Jason Y, Yoshihiko H, Georgia K, Franz G, Gyorgy S, Weston L, Anthony I, Kathleen A, Jasmine N (2014) High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. Nature 513(7516):124
2. Ballesteros JA, Weinstein H (1995) ^{19}F Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. Meth Neurosci 25(5):366–428
3. Bertekap RL Jr, Burford NT, Li Z, Alt A (2015) High-throughput screening for allosteric modulators of GPCRs. Meth Mol Biol (Clifton, NJ) 1335:223–240
4. Brodbeck RM, Cortright DN, Kieltyka AP, Yu J, Baltazar CO, Buck ME, Meade R, Maynard GD, Thurkauf A, Chien DS, Hutchison AJ, Krause JE (2008) Identification and characterization of NDT 9513727 [N,N-bis(1,3-benzodioxol-5-ylmethyl)-1-butyl-2,4-diphenyl-1H-imidazole-5-methanam ine], a novel, orally bioavailable C5a receptor inverse agonist. J Pharmacol Exp Ther 327(3):898–909
5. Bueno AB, Showalter AD, Wainscott DB, Stutsman C, Marin A, Ficorilli J, Cabrera O, Willard FS, Sloop KW (2016) Positive allosteric modulation of the glucagon-like peptide-1 receptor by diverse electrophiles. J Biol Chem 291(20):10700–10715
6. Byrne EF, Sircar R, Miller PS, Hedger G, Luchetti G, Nachtergaele S, Tully MD, Mydock-Mcgrane L, Covey DF, Rambo RP (2016) Structural basis of Smoothened regulation by its extracellular domains. Nature 535(7613):517–522
7. Cheng Z, Yoga S, Arlow DH, Juan Jose F, Daniel P, Yaowu Z, Green HF, Anjali P, Dror RO, Shaw DE (2012) High-resolution crystal structure of human protease-activated receptor 1. Nature 492(7429):387–392
8. Cheng RKY, Fiez-Vandal C, Schlenker O, Edman K, Aggeler B, Brown DG, Brown GA, Cooke RM, Dumelin CE, Dore AS, Geschwindner S, Grebner C, Hermansson NO, Jazayeri A, Johansson P, Leong L, Prihandoko R, Rappas M, Soutter H, Snijder A, Sundstrom L, Tehan B, Thornton P, Troast D, Wiggin G, Zhukov A, Marshall FH, Dekker N (2017) Structural insight into allosteric modulation of protease-activated receptor 2. Nature 545(7652):112–115
9. Christopher JA, Aves SJ, Bennett KA, Doré AS, Errey JC, Ali J, Marshall FH, Krzysztof O, Serrano-Vega MJ, Tehan BG (2015) Fragment and structure-based drug discovery for a class C GPCR: discovery of the mGlu5 negative allosteric modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzonitrile). J Med Chem 58(16):6653–6664

10. Christopher JA, Orgován Z, Congreve M, Doré AS, Errey JC, Marshall FH, Mason JS, Okrasa K, Rucktooa P, Serrano-Vega MJ (2018) Structure-based optimization strategies for G Protein-Coupled Receptor (GPCR) allosteric modulators: a case study from analyses of new metabotropic glutamate receptor 5 (mGlu5) X-ray structures. J Med Chem 62(1):207–222
11. Christopoulos A (2002) Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat Rev Drug Discov 1(3):198–210
12. Dore AS, Okrasa K, Patel JC, Serrano-Vega M, Bennett K, Cooke RM, Errey JC, Jazayeri A, Khan S, Tehan B, Weir M, Wiggin GR, Marshall FH (2014) Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. Nature 511(7511):557–562
13. Dore AS, Bortolato A, Hollenstein K, Cheng RKY, Read RJ, Marshall FH (2017) Decoding corticotropin-releasing factor receptor type 1 crystal structures. Curr Mol Pharmacol 10(4):334–344
14. Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, Borhani DW, Shaw DE (2011) Activation mechanism of the beta2-adrenergic receptor. Proc Natl Acad Sci U S A 108(46):18684–18689
15. Gentry PR, Sexton PM, Christopoulos A (2015) Novel allosteric modulators of G protein-coupled receptors. J Biol Chem 290(32):19478–19488
16. Hauser AS, Attwood MM, Rask-Andersen M, Schioth HB, Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov 16(12):829–842
17. Hauser AS, Chavali S, Masuho I, Jahn LJ, Martemyanov KA, Gloriam DE, Babu MM (2018) Pharmacogenomics of GPCR drug targets. Cell 172(1–2):41–54. e19
18. Ho JD, Chau B, Rodgers L, Lu F, Wilbur KL, Otto KA, Chen Y, Song M, Riley JP, Yang HC, Reynolds NA, Kahl SD, Lewis AP, Groshong C, Madsen RE, Conners K, Lineswala JP, Gheyi T, Saflor MD, Lee MR, Benach J, Baker KA, Montrose-Rafizadeh C, Genin MJ, Miller AR, Hamdouchi C (2018) Structural basis for GPR40 allosteric agonism and incretin stimulation. Nat Commun 9(1):1645
19. Hollenstein K, Kean J, Bortolato A, Cheng RK, Dore AS, Jazayeri A, Cooke RM, Weir M, Marshall FH (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. Nature 499(7459):438–443
20. Hollenstein K, Graaf CD, Bortolato A, Wang MW, Marshall FH, Stevens RC (2014) Insights into the structure of class B GPCRs. Trends Pharmacol Sci 35(1):12–22
21. Hori T, Okuno T, Hirata K, Yamashita K, Kawano Y, Yamamoto M, Hato M, Nakamura M, Shimizu T, Yokomizo T, Miyano M, Yokoyama S (2018) Na(+)-mimicking ligands stabilize the inactive state of leukotriene B4 receptor BLT1. Nat Chem Biol 14(3):262–269
22. Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39(Database issue):D663–D669
23. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T, Zhang J (2013) Allosite: a method for predicting allosteric sites. Bioinformatics (Oxford, England) 29(18):2357–2359
24. Huang XP, Karpiak J, Kroeze WK, Zhu H, Chen X, Moy SS, Saddoris KA, Nikolova VD, Farrell MS, Wang S, Mangano TJ, Deshpande DA, Jiang A, Penn RB, Jin J, Koller BH, Kenakin T, Shoichet BK, Roth BL (2015) Allosteric ligands for the pharmacologically dark receptors GPR68 and GPR65. Nature 527(7579):477–483
25. Huang P, Zheng S, Wierbowski BM, Kim Y, Nedelcu D, Aravena L, Liu J, Kruse AC, Salic A (2018) Structural basis of smoothened activation in hedgehog signaling. Cell 174(2):312–324
26. Jacoby E, Bouhelal R, Gerspacher M, Seuwen K (2006) The 7 TM G-protein-coupled receptor target family. Chem Med Chem 1(8):761–782
27. Janda CY, Waghray D, Levin AM, Thomas C, Garcia KC (2012) Structural basis of Wnt recognition by Frizzled. Science (New York, NY) 337(6090):59–64

28. Jazayeri A, Doré AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, Bortolato A, Koglin M, Robertson NJ, Errey JC (2016) Extra-helical binding site of a glucagon receptor antagonist. Nature 533(7602):274
29. Jean-Philippe P, Thierry G, Laurent PJ (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. Pharmacol Ther 98(3):325–354
30. Justus CR, Dong L, Yang LV (2013) Acidic tumor microenvironment and pH-sensing G protein-coupled receptors. Front Physiol 4:354
31. Katritch V, Fenalti G, Abola EE, Roth BL, Cherezov V, Stevens RC (2014) Allosteric sodium in class A GPCR signaling. Trends Biochem Sci 39(5):233–244
32. Kobilka BK (2007) G protein coupled receptor structure and activation. Biochim Biophys Acta 1768(4):794–807
33. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, Hubner H, Pardon E, Valant C, Sexton PM, Christopoulos A, Felder CC, Gmeiner P, Steyaert J, Weis WI, Garcia KC, Wess J, Kobilka BK (2013) Activation and allosteric modulation of a muscarinic acetylcholine receptor. Nature 504(7478):101–106
34. Latorraca NR, Venkatakrishnan AJ, Dror RO (2017) GPCR Dynamics: Structures in Motion. Chem Rev 117(1):139–155
35. Liu JJ, Horst R, Katritch V, Stevens RC, Wuthrich K (2012) Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR. Science (New York, NY) 335(6072):1106–1110
36. Liu X, Ahn S, Kahsai AW, Meng KC, Latorraca NR, Pani B, Venkatakrishnan AJ, Masoudi A, Weis WI, Dror RO (2017) Mechanism of intracellular allosteric β2AR antagonist revealed by X-ray crystal structure. Nature 548(7668):480–484
37. Liu H, Kim HR, Deepak R, Wang L, Chung KY, Fan H, Wei Z, Zhang C (2018) Orthosteric and allosteric action of the C5a receptor antagonists. Nat Struct Mol Biol 25(6):472–481
38. Lu J, Byrne N, Wang J, Bricogne G, Brown FK, Chobanian HR, Colletti SL, Di Salvo J, Thomas-Fowlkes B, Guo Y, Hall DL, Hadix J, Hastings NB, Hermes JD, Ho T, Howard AD, Josien H, Kornienko M, Lumb KJ, Miller MW, Patel SB, Pio B, Plummer CW, Sherborne BS, Sheth P, Souza S, Tummala S, Vonrhein C, Webb M, Allen SJ, Johnston JM, Weinglass AB, Sharma S, Soisson SM (2017) Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. Nat Struct Mol Biol 24(7):570–577
39. Masureel M, Zou Y, Picard LP, van der Westhuizen E, Mahoney JP, Rodrigues J, Mildorf TJ, Dror RO, Shaw DE, Bouvier M, Pardon E, Steyaert J, Sunahara RK, Weis WI, Zhang C, Kobilka BK (2018) Structural insights into binding specificity, efficacy and bias of a beta2AR partial agonist. Nat Chem Biol 14(11):1059–1066
40. May LT, Leach K, Sexton PM, Christopoulos A (2007) Allosteric modulation of G protein-coupled receptors. Annu Rev Pharmacol Toxicol 47:1–51
41. Miao Y, Nichols SE, McCammon JA (2014) Mapping of allosteric druggable sites in activation-associated conformers of the M2 muscarinic receptor. Chem Biol Drug Des 83(2):237–246
42. Miao Y, Goldfeld DA, Moo EV, Sexton PM, Christopoulos A, McCammon JA, Valant C (2016) Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor. Proc Natl Acad Sci U S A 113(38):E5675–E5684
43. Mohebbi N, Benabbas C, Vidal S, Daryadel A, Bourgeois S, Velic A, Ludwig MG, Seuwen K, Wagner CA (2012) The proton-activated G protein coupled receptor OGR1 acutely regulates the activity of epithelial proton transport proteins. Cell Physiol Biochem 29(3-4):313–324
44. Ngan CH, Bohnuud T, Mottarella SE, Beglov D, Villar EA, Hall DR, Kozakov D, Vajda S (2012) FTMAP: extended protein mapping with user-selected probe molecules. Nucleic acids research. Nucleic Acids Res 40(Web Server issue):W271–W275
45. Noblin DJ, Bertekap RL Jr, Burford NT, Hendricson A, Zhang L, Knox R, Banks M, O'Connell J, Alt A (2012) Development of a high-throughput calcium flux assay for identification of all ligand types including positive, negative, and silent allosteric modulators for G protein-coupled receptors. Assay Drug Dev Technol 10(5):457–467

46. Nolte WM, Fortin JP, Stevens BD, Aspnes GE, Griffith DA, Hoth LR, Ruggeri RB, Mathiowetz AM, Limberakis C, Hepworth D, Carpino PA (2014) A potentiator of orthosteric ligand activity at GLP-1R acts via covalent modification. Nat Chem Biol 10(8):629–631
47. Oswald C, Rappas M, Kean J, Doré AS, Errey JC, Bennett K, Deflorian F, Christopher JA, Jazayeri A, Mason JS (2016) Intracellular allosteric antagonism of the CCR9 receptor. Nature 540(7633):462
48. Peng P, Chen H, Zhu Y, Wang Z, Li J, Luo RH, Wang J, Chen L, Yang LM, Jiang H, Xie X, Wu B, Zheng YT, Liu H (2018) Structure-based design of 1-heteroaryl-1,3-propanediamine derivatives as a novel series of CC-chemokine receptor 5 antagonists. J Med Chem 61(21):9621–9636
49. Price MR, Baillie GL, Thomas A, Stevenson LA, Easson M, Goodwin R, McLean A, McIntosh L, Goodwin G, Walker G, Westwood P, Marrs J, Thomson F, Cowley P, Christopoulos A, Pertwee RG, Ross RA (2005) Allosteric modulation of the cannabinoid CB1 receptor. Mol Pharmacol 68(5):1484–1495
50. Robertson N, Rappas M, Dore AS, Brown J, Bottegoni G, Koglin M, Cansfield J, Jazayeri A, Cooke RM, Marshall FH (2018) Structure of the complement C5a receptor bound to the extra-helical antagonist NDT9513727. Nature 553(7686):111–114
51. Rodriguez AL, Grier MD, Jones CK, Herman EJ, Kane AS, Smith RL, Williams R, Zhou Y, Marlo JE, Days EL, Blatt TN, Jadhav S, Menon UN, Vinson PN, Rook JM, Stauffer SR, Niswender CM, Lindsley CW, Weaver CD, Conn PJ (2010) Discovery of novel allosteric modulators of metabotropic glutamate receptor subtype 5 reveals chemical and functional diversity and in vivo activity in rat behavioral models of anxiolytic and antipsychotic activity. Mol Pharmacol 78(6):1105–1123
52. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI, Overington JP (2017) A comprehensive map of molecular drug targets. Nat Rev Drug Discov 16(1):19–34
53. Shore DM, Baillie GL, Hurst DH, Navas F 3rd, Seltzman HH, Marcu JP, Abood ME, Ross RA, Reggio PH (2014) Allosteric modulation of a cannabinoid G protein-coupled receptor: binding site elucidation and relationship to G protein signaling. J Biol Chem 289(9):5828–5845
54. Song G, Yang D, Wang Y, de Graaf C, Zhou Q, Jiang S, Liu K, Cai X, Dai A, Lin G, Liu D, Wu F, Wu Y, Zhao S, Ye L, Han GW, Lau J, Wu B, Hanson MA, Liu ZJ, Wang MW, Stevens RC (2017) Human GLP-1 receptor transmembrane domain structure in complex with allosteric modulators. Nature 546(7657):312–315
55. Stornaiuolo M, Bruno A, Botta L, La Regina G, Cosconati S, Silvestri R, Marinelli L, Novellino E (2015) Endogenous vs Exogenous Allosteric Modulators in GPCRs: a dispute for shuttling CB1 among different membrane microenvironments. Sci Rep 5:15453
56. Suno R, Kimura KT, Nakane T, Yamashita K, Wang J, Fujiwara T, Yamanaka Y, Im D, Horita S, Tsujimoto H, Tawaramoto MS, Hirokawa T, Nango E, Tono K, Kameshima T, Hatsui T, Joti Y, Yabashi M, Shimamoto K, Yamamoto M, Rosenbaum DM, Iwata S, Shimamura T, Kobayashi T (2018) Crystal structures of human Orexin 2 receptor bound to the subtype-selective antagonist EMPA. Structure 26(1):7–19. e15
57. Tan QX, Zhu Y, Li J, Chen ZX, Han GW, Kufareva I, Li TT, Ma LM, Fenalti G, Li J, Zhang WR, Xie X, Yang HY, Jiang HL, Cherezov V, Liu H, Stevens RC, Zhao Q, Wu BL (2013) Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. Science (New York, NY) 341(6152):1387–1390
58. Thal DM, Sun B, Feng D, Nawaratne V, Leach K, Felder CC, Bures MG, Evans DA, Weis WI, Bachhawat P, Kobilka TS, Sexton PM, Kobilka BK, Christopoulos A (2016) Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. Nature 531(7594):335–340
59. Tobo A, Tobo M, Nakakura T, Ebara M, Tomura H, Mogi C, Im DS, Murata N, Kuwabara A, Ito S, Fukuda H, Arisawa M, Shuto S, Nakaya M, Kurose H, Sato K, Okajima F (2015) Characterization of imidazopyridine compounds as negative allosteric modulators of proton-sensing GPR4 in extracellular acidification-induced responses. PLoS One 10(6):e0129334

60. Venkatakrishnan AJ, Deupi X, Lebon G, Heydenreich FM, Flock T, Miljus T, Balaji S, Bouvier M, Veprintsev DB, Tate CG, Schertler GF, Babu MM (2016) Diverse activation pathways in class A GPCRs converge near the G-protein-coupling region. Nature 536(7617):484–487
61. Wang C, Wu H, Evron T, Vardy E, Han GW, Huang XP, Hufeisen SJ, Mangano TJ, Dan JU, Katritch V (2014) Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. Nat Commun 5(5):4355
62. Wei L, Eugene C, Thompson AA, Pavel C, Fei X, Vsevolod K, Gye Won H, Roth CB, Heitman LH, Ijzerman AP (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. Science 337(6091):232
63. Weierstall U, James D, Wang C, White TA, Wang D, Liu W, Spence JC, Bruce DR, Nelson G, Fromme P (2014) Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. Nat Commun 5(1):3309
64. Wu H, Wang C, Gregory KJ, Han GW, Cho HP, Xia Y, Niswender CM, Katritch V, Meiler J, Cherezov V, Conn PJ, Stevens RC (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. Science (New York, NY) 344(6179):58–64
65. Xiao X, Tang JJ, Peng C, Wang Y, Fu L, Qiu ZP, Xiong Y, Yang LF, Cui HW, He XL, Yin L, Qi W, Wong CC, Zhao Y, Li BL, Qiu WW, Song BL (2017) Cholesterol modification of smoothened is required for hedgehog signaling. Mol Cell 66(1):154–162.e110
66. Yang S, Wu Y, Xu TH, de Waal PW, He Y, Pu M, Chen Y, DeBruine ZJ, Zhang B, Zaidi SA, Popov P, Guo Y, Han GW, Lu Y, Suino-Powell K, Dong S, Harikumar KG, Miller LJ, Katritch V, Xu HE, Shui W, Stevens RC, Melcher K, Zhao S, Xu F (2018) Crystal structure of the Frizzled 4 receptor in a ligand-free state. Nature 560(7720):666–670
67. Zhang D, Gao ZG, Zhang K, Kiselev E, Crane S, Wang J, Paoletta S, Yi C, Ma L, Zhang W, Han GW, Liu H, Cherezov V, Katritch V, Jiang H, Stevens RC, Jacobson KA, Zhao Q, Wu B (2015) Two disparate ligand-binding sites in the human P2Y1 receptor. Nature 520(7547):317–321
68. Zhang H, Qiao A, Yang D, Yang L, Dai A, Graaf CD, Reedtzrunge S, Dharmarajan V, Zhang H, Han GW (2017a) Structure of the full-length glucagon class B G-protein-coupled receptor. Nature 546(7657):259
69. Zhang X, Zhao F, Wu Y, Yang J, Han GW, Zhao S, Ishchenko A, Ye L, Lin X, Ding K (2017b) Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. Nat Commun 8:15383
70. Zheng Y, Qin L, Zacarías NV, De VH, Han GW, Gustavsson M, Dabros M, Zhao C, Cherney RJ, Carter PJN (2016) Structure of CC chemokine receptor 2 with orthosteric and allosteric antagonists. Nature 540(7633):458–461

Chapter 11
Allosteric Small-Molecule Serine/Threonine Kinase Inhibitors

Resmi C. Panicker, Souvik Chattopadhaya, Anthony G. Coyne, and Rajavel Srinivasan

Abstract Deregulation of protein kinase activity has been linked to many diseases ranging from cancer to AIDS and neurodegenerative diseases. Not surprisingly, drugging the human kinome – the complete set of kinases encoded by the human genome – has been one of the major drug discovery pipelines. Majority of the approved clinical kinase inhibitors target the ATP binding site of kinases. However, the remarkable sequence and structural similarity of ATP binding pockets of kinases make selective inhibition of kinases a daunting task. To circumvent these issues, allosteric inhibitors that target sites other than the orthosteric ATP binding pocket have been developed. The structural diversity of the allosteric sites allows these inhibitors to have higher selectivity, lower toxicity and improved physiochemical properties and overcome drug resistance associated with the use of conventional kinase inhibitors. In this chapter, we will focus on the allosteric inhibitors of selected serine/threonine kinases, outline the benefits of using these inhibitors and discuss the challenges and future opportunities.

Keywords Kinase allostery · Allosteric inhibitors · SMKI · FBDD · MEK allosteric inhibitors · Aurora A kinase · Aurora A-TPX2 · CK2α · CDK · Akt

Authors Resmi C. Panicker and Souvik Chattopadhaya have equally contributed to this chapter.

R. C. Panicker · R. Srinivasan (✉)
School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, People's Republic of China
e-mail: rajavels@tju.edu.cn

S. Chattopadhaya
School of Applied Sciences, Republic Polytechnic, Singapore, Singapore

A. G. Coyne
University Chemical Laboratory, University of Cambridge, Cambridge, UK

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_11

11.1 Introduction

11.1.1 Kinases and Small-Molecule Kinase Inhibitors

Protein kinases catalyse the transfer of the gamma phosphoryl group of ATP to specific hydroxyl (-OH) group of their substrates, including proteins, lipids and sugars, and provide key mechanisms for control of cellular signalling processes [30, 60]. The human "kinome" has been known to encompass 518 kinases and phosphorylate up to one-third of the proteome [36, 80]. Phosphorylation occurs on serine, threonine or tyrosine residues, and kinases can therefore be broadly classified as serine/threonine kinases or tyrosine kinases and, in some instances, dual-specificity kinases wherein they phosphorylate both serine/threonine and tyrosine residues. Dysregulation of kinase signalling pathways and/or overexpression of kinases is associated with a variety of pathophysiological conditions ranging from cancer, inflammatory diseases to metabolic disorders and neurodegenerative diseases [42]. Kinases have been intensively pursued as important therapeutic targets by pharmaceutical industries given the pivotal role of enzymes in diseases as well as the established druggability and clinical safety profile of approved kinase inhibitors. Small-molecule kinase inhibitors (SMKIs) entered clinical stage in the 1990s, and discovery efforts were bolstered in 2001 with the FDA approval of imatinib (also known as Gleevec®) – the first oral kinase inhibitor used in the treatment of patients with chronic myeloid leukaemia (CML) [59]. Culmination of the work done in the past three decades has resulted in 38 kinase inhibitors being approved till date [35].

SMKIs can be classified as Types I–IV [96, 115]. Type I inhibitors bind to ATP pocket of active kinases when the conserved aspartate–phenylalanine–glycine (DFG) motif at the N-terminus of the activation lobe is "in" conformation. In contrast, Type II inhibitors like imatinib, nilotinib and sorafenib stabilize inactive kinases by binding to an extended pocket that includes the DGF motif in the "out" conformation and an adjacent, less-conserved allosteric site [38]. Both Types I and II are ATP-competitive inhibitors. Type III and IV inhibitors are allosteric inhibitors. While Type III inhibitors like trametinib and cobimetinib bind within the cleft between the small and large lobes adjacent to the ATP binding pocket [96, 114], Type IV inhibitors bind out of the cleft and phosphoacceptor region [40]. Besides these main classes of kinase inhibitors, there are other types of non-covalent kinase inhibitors. Lamba and Ghosh [69] labelled bivalent or bisubstrate molecules that span two regions of the protein kinase domain as Type V inhibitors. Zuccotto et al. introduced Type 1½ inhibitors as compounds that bind kinases in the DFG-Asp in and C-helix out conformation [116]. Due to the remarkable structural and sequence similarity of ATP binding site, finding selective SMKIs with good physiochemical and pharmacokinetic properties is a daunting task. In some cases, the multikinase inhibition can be advantageous, as exemplified by the MAPK inhibitor sorafenib, which also inhibits kinases involved in tumour signalling and vasculature, making it effective in targeting tumour growth [108]. However, in general, the clinical applications of Type I and II SMKIs have been undermined from undesired target

selectivity and specificity profiles resulting in off-target side effects [37]. Besides attaining selectivity required for pharmacological target validation, several other challenges need to be overcome in order to realize the full potential of kinases as therapeutic targets in cancer and beyond. These include validation of novel kinases (such as atypical and pseudokinases) as targets, targeting non-cysteine residues like lysine [3] or methionine [74], developing specific modulators that can inhibit kinase signalling cascades or promote kinase degradation, overcoming drug resistance due to mutations in ATP binding pocket and developing new technologies for efficient screening and profiling of kinase inhibitors [19, 35].

11.1.2 Allosteric Kinase Inhibitors

To circumvent the issues associated with the use of ATP-competitive inhibitors, an alternate approach relies on the development of allosteric inhibitors. The idea of allosteric inhibition stems from the observation that non-catalytic functions of kinases involve scaffolding, protein-protein interaction, allosteric effector on other enzymes and DNA binding functions. Unlike phosphorylation, these functions are not as conserved among the different kinases and hence blocking or enhancing these functions would allow for modulation of kinase activities [92]. Allosteric inhibitors bind to an effector site distinct from the orthosteric ATP binding pocket in kinases or to other interacting proteins [76]. Even though allosteric inhibitors are non-ATP-competitive, some of these inhibitors stabilize the inactive kinase conformation and appear “ATP-competitive”, while others show “non-ATP-competitive” kinetics of inhibition [22, 81]. Since trametinib was granted FDA approval as the first allosteric SMKI in 2013, the field of allosteric inhibitors has rapidly burgeoned with more than ten other allosteric inhibitors of MEK (CI-1040, GDC-0973, pimasertib) [61] and Akt (MK- 2206) [109] in clinical trials. Inhibitors for LIMK2, PAK, IRE1 and RIP1 have also been now reported [110]. Identification of allosteric sites and modulators has become a central theme in the kinase drug discovery programs.

11.1.3 Advantages of Allosteric Inhibitors

Allosteric inhibitors have the advantage of specific inhibition and minimal off-target pharmacology, given the low sequence homology of allosteric sites [34]. Even isoform selectivity is possible as demonstrated by the cases of AKT1/2/3 [4] and PAK1/2 [64] inhibitors. Such inhibitors expand the chemical space for the identification of novel kinase inhibitors. The ATP pocket is deep and hydrophobic – so ATP-competitive pharmacophores tend to be poorly soluble in water and can be of higher molecular weight. On the contrary, allosteric modulators have a pronounced shape dictated by the allosteric site itself and it is possible to develop compounds with improved physiochemical properties [22]. Since allosteric inhibitors need not

compete with high levels of intracellular ATP, weak binders can be easily identified in compound screens. Resistance to therapy is a particular problem with cancer and anti-infectives. For example, leukaemia patients become resistant to the kinase inhibitor imatinib, due to mutation in the gatekeeper residue (T315I) at the ATP binding pocket of BCR-ABL1 [43]. Identification of allosteric sites provides a means to "rescue" treatment for such patients. In addition, one can envisage that resistance can be avoided by using combination therapies in which both the catalytic and non-catalytic sites are targeted simultaneously [113]. As allosteric modulators do not interfere with native ligands, their binding can cause activation or inhibition of kinase activity, and therefore these compounds can be used as mechanistic probes to delineate kinase pathways [41] and better understand their catalytic and non-catalytic activities (scaffolding, homo- and heterodimerization) and regulation [34, 110].

11.2 Experimental Methods to Identify Allosteric Ligands

One of the prerequisites for identification, screening and optimization of allosteric modulators is a knowledge of the allosteric site. But the detection of such sites on proteins is not straightforward. Serendipity has played an important role in the discovery of most allosteric inhibitors. These inhibitors have been found generally through high-throughput screenings (HTS) while measuring the binding affinities or kinetics of protein-ligand interactions such as the identification of compound binding to myristate pocket of ABL kinase using a cytotoxicity screen of combinatorial kinase-directed heterocycle library [1]. Confirmation of binding sites and mechanism of action of allosteric modulators tends to be even more laborious relative to ATP-mimetic inhibitors. In recent years, guided by the knowledge of allosteric proteins and modulators, several experimental and computational approaches have been developed to identify allosteric sites. These approaches have been discussed in detail in several excellent reviews [22, 76]. For the sake of brevity, this chapter will only highlight those approaches which have been used to find the allosteric sites on kinases.

11.2.1 High-Throughput Screening (HTS)

HTS, as the name suggests, attempt to identify potent chemical leads ($IC_{50} < 10$ μM) by screening millions of rule-of-five compliant small (< 500 Da) molecules. HTS assays require rigorous kinetic experiments to ascertain activation/inactivation of kinases prior to X-ray crystallographic studies to determine binding sites. The earliest allosteric inhibitors identified, like those targeting MEK, AKT and Bcr-Abl, were mostly discovered starting with hits derived from HTS assays. As the initial hits from HTS assays of allosteric inhibitors are usually of low or moderate potency, they require optimizations guided by NMR or co-crystal structures to yield

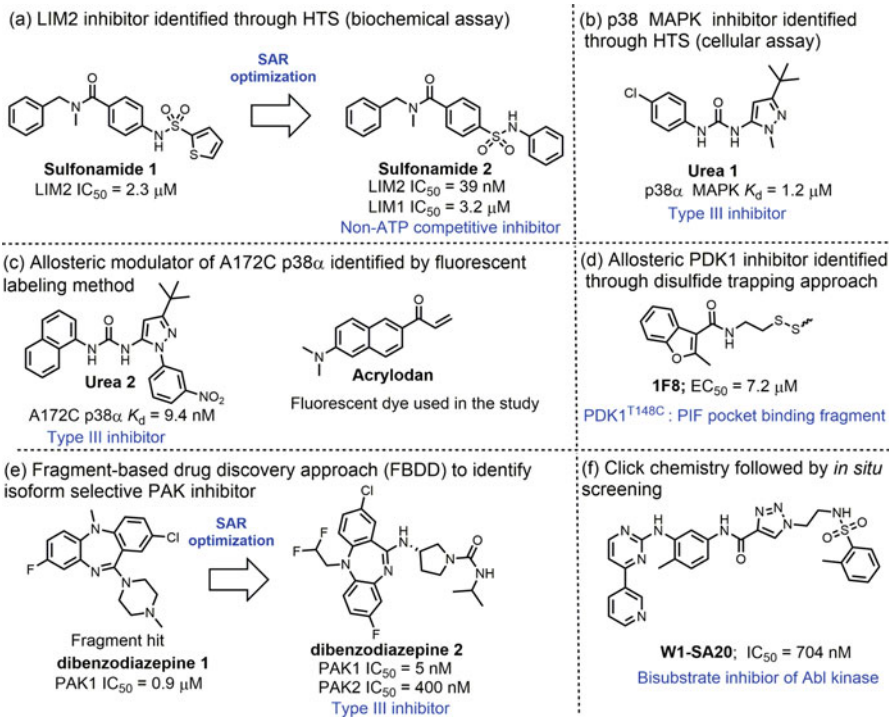


Fig. 11.1 Allosteric ligands identified through different experimental methods (**a**–**f**)

potent compounds with improved physicochemical properties. HTS are done using biochemical or cellular assays in high-throughput format. The following paragraphs give a snippet of some of the allosteric inhibitors identified from HTS screens. Care has been taken to include diverse inhibitors identified using biochemical (LIMK2) and cellular assays (MEK) to give the readers an overview of HTS in the realm of allosteric inhibitor discovery [106].

HTS of LIMK2 identified a sulphonamide **1** with an IC_{50} of 2.3 μM (Fig. 11.1a). Subsequent SAR optimizations led to sulphonamide **2** that showed 800-fold selectivity for LIMK2 (IC_{50} 39 nM) over LIMK1 (IC_{50} 3.2 μM). Even in the presence of 300 μM ATP, sulphonamide **2** showed no significant activity (IC_{50} > 100 μM) against a plethora of kinases – PKA, PKC, CDKs, Abl, etc. A further optimized analogue when co-crystallized with LIMK2 indeed confirms that the inhibitor was a Type III binding mode [47].

Cellular-based HTS followed by deconvolution of the target using biochemical assays have been used to identify MEK inhibitors (Fig. 11.1b). In one such reporter assay, employing the transcription factor activator protein 1 (AP-1) in transfected COS-7 cells led to the identification of urea **1** with modest binding and cellular activity against p38α MAPK (K_d = 1.2 μM). Resolution of co-crystal complex of urea **1** with p38α confirmed an allosteric binding mode [89].

11.2.2 Fluorescent Labelling

Fluorescent labelling entails selective, covalent labelling of a macromolecule like proteins using a reactive derivative of an environmentally sensitive fluorophore [82]. Choice of amino acid is crucial for the success of this approach as the amino acid should be solvent exposed (for fluorophore attachment) and must undergo significant movement upon ligand binding. This method is particularly amenable to detecting allosteric modulators since allosteric regulation frequently involves large conformational changes in proteins like kinases. Kinase activation involves the movement of the conserved DFG motif from the "out" to the "in" configuration. Based on this structural transition and using an acrylodan dye to label A172C mutant of p38α, Simard et al. were able to deduce the SAR of pyrazolourea derivatives that bind to allosteric pockets of p38α [102]. Urea **2** had a K_d of 9.4 nm against the A172C mutant of p38α (Fig. 11.1c).

11.2.3 Disulfide Trapping

Disulfide trapping (or tethering) developed by Wells [31, 32] is another useful approach to detect and characterize known, suspected or orphan allosteric sites on proteins. The method entails screening of disulfide-containing compounds, under partially reducing conditions, for their ability to form a mixed disulfide with a natural or an engineered cysteine residue near a site of interest of protein target. Mass spectrometry is used to identify the small molecule, and binding mode is further confirmed by X-ray crystallographic studies. Allosteric ligand (**1F8**) (Fig. 11.1d) binding at the PIF pocket of 3-phosphoinositide-dependent kinase 1 (PDK1) have been identified by this approach [97].

11.2.4 Fragment-Based Drug Discovery (FBDD)

Although HTS is a powerful method to identify lead molecules, there are several limitations – limited chemical diversity, lower "hit" rates and difficulties in lead optimization to drug-like molecules. Fragment-based drug discovery (FBDD) has emerged as a powerful, complementary method to HTS for the generation of new chemical hits. FBDD involves the detection and subsequent elaboration of weak binding chemical fragments into high-binding lead compounds [24]. Fragments are usually defined as having less than 20 non-hydrogen (or "heavy") atoms [32]. Most fragment libraries adhere to the "rule-of-three" molecular weight < 300 Da, fewer than three hydrogen bond donors and acceptors, fewer than three rotatable bonds and

cLogP of three or below [20]. The binding mode of these fragments is determined using both "classical" methods of fragment screening like X-ray crystallography, NMR spectroscopy, surface plasmon resonance (SPR) and more recent approaches such as thermal shift assays (TSAs) and microscale thermophoresis. In contrast to HTS methods which identify potent chemical leads (<10 μM), FBDD fragments have lower affinities (μM to mM) but possess higher ligand efficiency (LE) [9]. FBDD also allows a chemically unbiased screen of the available protein surface, thereby maximizing the probability of finding alternate binding sites as well as uncovering cooperativity between the various binding sites [67]. In several cases, identification of clinical candidates has been accelerated by FBDD. The first fragment-derived FDA-approved drug on the market – vemurafenib – took only 6 years from project initiation to FDA approval. Vemurafenib is an inhibitor for mutant BRAF kinase and has been approved for treatment of melanoma [12]. Fragment screens allow sampling of a greater chemical space using smaller chemical libraries [83] and provide opportunities for selectivity and intellectual property.

Fragment-based screening led to the identification of dibenzodiazepine **1** (Fig. 11.1e) PAK1 inhibitor that bound to the allosteric pocket next to the ATP binding pocket with the DFG motif in the "out" conformation. Structure-assisted optimizations delivered dibenzodiazepine **2** with improved physiochemical properties. The dibenzodiazepine 2 compound showed a high selectivity when profiled against 442 kinases. It was also able to preferentially bind PAK1 isoform ($K_d = 7$ nM) over PAK2 ($K_d = 400$ nM) even though both isoforms have a 93% sequence similarity [64]. However, the in vitro studies were not promising due to the poor solubility of the inhibitor in the rat liver microsomes.

11.2.5 In Situ Click Chemistry

In situ click chemistry is another fragment-based approach that helps identify ligands that bind at secondary sites that are adjacent to the active site [66]. This method relies on near-perfect and robust chemical transformations that are modular in nature. The most widely used reaction for this approach is the copper (I)-catalysed 1,3-dipolar cycloaddition between azide and alkyne fragments which is generally referred to as "click" chemistry. A prerequisite for this approach is the a priori knowledge about the nature of the allosteric site and its distance from the active site. Following assembly, the compound library can be directly screened against the biological target as the efficiency of the "click" reaction obviates the need for further purification of the assembled molecules. This method has been used to identify secondary site binding fragments on several different protein targets including tyrosine phosphatases [104, 105] and Abelson (Abl) tyrosine kinase [63] (Fig. 11.1f).

11.2.6 Computational Approaches to Identify Allosteric Modulators

Structure-based virtual screening techniques are some of the most successful methods for augmenting the drug discovery process. Recent advances in computing power, MD simulation and bioinformatic analyses have allowed the development of a suite of computational methods to identify allosteric sites. These approaches seek to complement experimental approaches highlighted in the previous section. Several different prediction approaches based on sequence, dynamics, structure, topology and normal mode analysis are available. For a detailed review on all these approaches, readers are referred to recent publication by Jian Zhang [84]. One such web server for the accurate prediction tool of allosteric sites is AlloSite (http://mdl.shsmu.edu.cn/AST). AlloSite was developed using pocket-based analysis and support vector machine (SVM) classifier. From a total of 218 allosteric sites deposited in ASDv2.0, Huang and co-workers selected 90 nonredundant allosteric sites and 1360 non-allosteric sites predicted by Fpocket to carry out SVM learning. A set of 21 descriptors was selected to characterize these allosteric sites and these descriptors were further optimized using the SVM classifier. This resulted in the structure-based model AlloSite [55, 56] which has a 96% accuracy rate to predict allosteric sites.

Other web servers that predict allosteric sites based on normal mode analysis (NMA) are AllositePro [103], PARS (http://bioinf.uab.cat/pars) [88] and more recently CavityPlus (http://repharma.pku.edu.cn/cavityplus) – a webserver that allows pharmacophore modelling and protein cavity detection in addition to allosteric site detection [111]. CavityPlus detects potential binding sites on the surface of target protein and ranks these sites with ligandability and druggability scores.

The next part of this chapter will focus on selected allosteric Ser/Thr kinase inhibitors of therapeutic relevance.

11.3 Allosteric Serine/Kinase Inhibitors

11.3.1 Allosteric MEK Inhibitors

Mitogen-activated protein kinases (MAPK) are serine/threonine kinases that mediate intracellular signalling via protein phosphorylation. About 70 genes are now known to encode close to 200 distinct components of the MAPK system (Table 11.1). MAPK signalling system consists of three separate pathways that transduce extracellular signals including growth factors, cytokines, mitogens and chemical and physical stresses to control a plethora of intracellular processes such as gene transcription, metabolism, cell proliferation, synaptic plasticity, apoptosis, etc. The three main signalling pathways are (A) c-Jun N-terminal kinase (JNK) pathway, (B) extracellular signal-regulated kinase (ERK) pathway and (C) p38 pathways. Signalling starts by binding of an extracellular signal/stimuli binding to cell surface

Table 11.1 Components of the MAPK signalling pathway

	Erk	JNK	p38
Transducer	GCK, GLK, HPK1, SOS, Rac, cdc42, Ras		
MAPKKKs	Raf-1, B-Raf, PAK	MEKK1-4, MLK1-3, ASK1, ASK3, TAK1,Tpl2	TAK
MAPKKs	MEK1, MEK2, MEK5	MEK4, MEK7	MEK3
MAPK	Erk1, Erk2, Erk5, Erk7, Erk8	JNK1, JNK2, JNK3	p38α, p38α, p38δ, p38γ
Target proteins	ATF2, ETS, Elk1, c-Jun, Jun-B, Jun-D, MEF2, MSK1,2, SAP-1		
Phosphatases	MKP-1, MKP-2, MKP-3, MKP-4, MKP-5, VHR, PAC1, Pyst2, STEP, hVH3/B23		

receptors. Transmission of these signals is then mediated by activation of G-protein (e.g. Ras) or by interaction of the upstream components with adaptor proteins. Signals are transduced further by cytosolic kinases that are organized generally in three tiers (MAPKKK, MAPKK, MAPK). The kinases in each tier phosphorylate and activate downstream kinases to allow rapid and regulated transmission to signals to the cellular targets. The MAPK cascade is a critical pathway for cancer cell survival, dissemination and resistance to drug therapy, and its understanding is crucial to designing drugs that can alter/regulate tumour signalling in this complex network of co-dependent pathways

The Ras-Raf-MEK1/2-ERK1/2 pathway is a highly conserved pathway in eukaryotic cells that transduces extracellular signals into cellular responses. The canonical MAPK signalling cascade is activated by binding of a mitogen to transmembrane receptors – GPCRs, cytokine receptors and receptor tyrosine kinases (RTKs). Activated receptors undergo dimerization and transphosphorylation at tyrosine residues, leading to their activation. Thereafter, small adapter proteins, such as SHC, SOS and GRB, associate with activated receptors and recruit guanine nucleotide exchange factors (GEFs). GEFs mediate conversion of inactive Ras-GDP (H-, K-, N-Ras) to active Ras-GTP at the inner leaflet of the plasma member. Ras-GTP promotes the homo- or heterodimerization of A-Raf, B-Raf and C-Raf through a multistage process. Phosphorylated Raf, a serine/threonine kinase, catalyses the phosphorylation and activation of MAPK/ERK kinase 1 (MEK1) and MEK2. MEK1/2 are dual specificity kinases and in turn phosphorylate ERK1/2 on the conserved Thr-Glu-Tyr (TEY) sequence that occurs on the activation loop of ERK (Thr202/Tyr204 for human ERK1 and Thr 185/Tyr187 for human ERK2). Significant ERK activation requires phosphorylation at both sites, with tyrosine phosphorylation preceding that of threonine. Unlike Raf and MEK1/2 which are fastidious in terms of the substrate, ERK1/2 has broad substrate specificity and can catalyse the phosphorylation of hundreds of downstream proteins such as kinases, phosphatases, transcription factors such as MYC and FOS, cell cycle proteins like cyclin D as well as regulatory proteins SPRY and DUSP which mediate negative feedback on MAPK pathway.

One of the most frequently deregulated pathways in human cancer is the signalling cascades mediated by Ras-Raf-MEK-ERK [17, 29].

Since the downstream effectors of this pathway have important roles in regulation of cell fate, genomic integrity and cellular survival, mutations in genes can lead to increase protein amplification and alter the tumour microenvironment, thus overactivating the pathways. These mutations can occur upstream in membrane receptor genes, such as epithelial growth factor receptor *(EGFR*), signal transducers (*RAS*), regulatory partners and downstream kinases belonging to MAPK/ERK pathway itself (*BRAF*). Genetic profiling of tumours has revealed that BRAF is the most frequently mutated oncogene in human cancer [95]. BRAF somatic missense mutation has been reported in 66% of malignant melanoma and at a lower frequency in a wide range of human cancer [26]. Two FDA-approved compounds (vemurafenib and dabrafenib) were used for the treatment of metastatic and unresectable BRAF-mutated melanomas. However, it has since become apparent that BRAF mutational status alone does not predict therapeutic responses in all cancer types. Preclinical and early clinical data strongly suggest that combination therapy of BRAF inhibitor and either MEK or conceivably an ERK inhibitor will have the greatest efficacy [53]. This has prompted the search for novel Akt and MEK inhibitors – both ATP-competitive and allosteric binders.

The first-generation MEK1/2 inhibitors such as CI-1040 demonstrated poor exposure in human subjects, while second-generation inhibitors like PD325901 had toxicity issues. Subsequent generation of MEK inhibitors are now being investigated and appear to be more potent and better tolerated. The subsequent paragraphs discuss only those allosteric MEK1/2 inhibitors that are approved or in preclinical stages (Fig. 11.2)

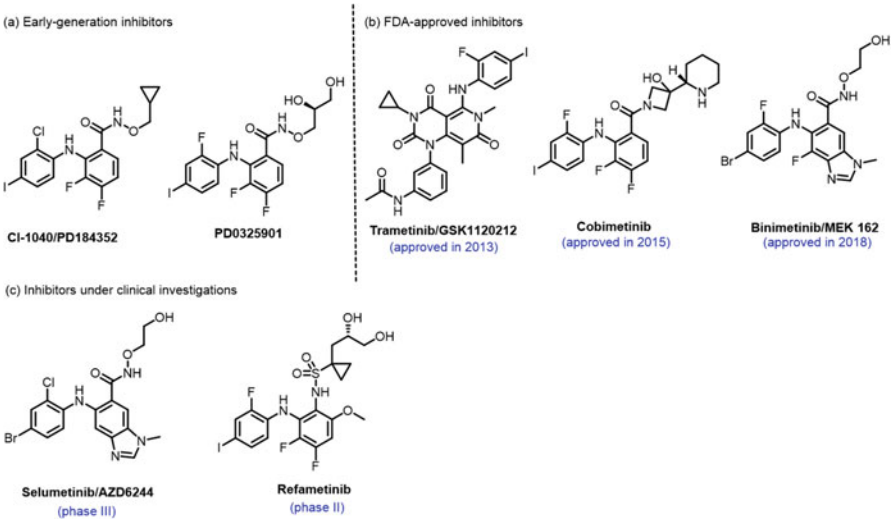


Fig. 11.2 Selected MEK1/2 inhibitors of clinical significance

Trametinib (Mekinist®) is a pyridone-based non-ATP-competitive MEK1/2 inhibitor that was developed by GSK [45]. This drug is the first FDA-approved allosteric inhibitor for MEK1/2. It was approved in 2013 as a single agent for treating patients with advanced melanoma with either B-Raf V600E or V600K mutation. Trametinib inhibits phosphorylation by RAF on serine 217 of MEK1 and supresses phosphorylated ERK [77]. Besides its use as a single agent, trametinib has also been used in combination therapies with BRAF inhibitor dabrafenib (combination therapy was given FDA approval in 2014).

Another allosteric MEK inhibitor approved for combination therapy is cobimetinib (GDC-0973, XL518). Cobimetinib was developed by Roche in collaboration with Exelixis. It is a diarylamino compound tethered to an azetidine derivative via an amide linkage. X-ray crystal structure analysis indicated the binding of cobimetinib to a pocket adjacent to the ATP binding site of MEK1, picking up several interactions with the activation segment, catalytic, C-helix and G-rich loops of the kinase [94]. The drug stabilizes the inactive conformation of MEK1/2 and promotes the dislocation of Raf-MEK1/2 complexes similar to that of trametinib. But, unlike trametinib, it does not block the Ras-catalysed phosphorylation of MEK1/2. The FDA has approved the use of cobimetinib in 2015 along with the FDA-approved B-Raf inhibitor vemurafenib (Zelboraf®) to treat metastasized melanoma with B-Raf V600E/K mutation [49]. It is also being investigated for the treatment of a number of other cancers.

Binimetinib (MEK162, ARRY-438162) is a diarylamine-based MEK1/2 inhibitor developed by Array Biopharma. Binimetinib has been documented to inhibit tumour volume in KRAS, NRAS and BRAG mutant in vitro cell line and in vivo tumour models. FDA-approved binimetinib is to be used along with a B-Raf inhibitor, encorafenib, for the treatment of metastasized melanoma with B-Raf V600E/K mutation [65]. The drug combination therapy received FDA approval in 2018.

Selumetinib (AZD6244, ARRY-142886) is a selective allosteric inhibitor for MEK 1 and 2 with an IC_{50} of 10–14 nM and currently undergoing clinical trials [33]. Selectivity of selumetinib for MEK has been documented using a panel of kinases at concentration of 10 μM [112]. The compound has shown antitumour activity in BRAF and NRAF mutant preclinical models including cell lines and xenografts from a variety of carcinomas like pancreas, lung, colon, breast, etc. In Phase I–III trials, selumetinib has been evaluated in a variety of tumours, either alone or in combination with other agents, such as cytotoxic chemotherapy. When used in conjunction with docetaxel, AZD6244 showed enhanced antitumour efficacy. Clinical trials have also been conducted using AZD6244 and Akt inhibitor (MK-2202) as it has been speculated that activation of the PI3K and AKT pathway could render melanomas refractory to MEK inhibitors.

Refametinib (BAY 869766, RDEA119) developed by Bayer is a sulphonamide-based diarylamino compound. It is an orally active exquisitely selective MEK1 inhibitor (IC_{50} = 19 nM) with potential antineoplastic activity. X-ray crystallographic analysis indicated that the molecule binds to an allosteric pocket adjacent to the Mg-ATP binding site and makes hydrogen bond and hydrophobic interaction

with the ATP, activation loops and adjacent residues. In vivo, it has shown significant activity in xenograft models of colon cancer, melanoma and epidermal carcinoma. In Phase I clinical trials, the molecule was well tolerated in patients with solid tumours. Refametinib has also been used in combination therapies with sorafenib for treatment with Ras mutant hepatocellular carcinoma (HCC) and is currently being evaluated in Phase II trials [73].

For more information on other types of MEK inhibitors, readers are referred to recent reviews by Luke [77] and Zhang [100].

11.3.2 Aurora A (AurA) Kinase

Aurora kinases are Ser/Thr kinases that play vital roles during cell division [46]. Aurora A, B and C are three highly conserved kinases in this family but have distinct functions [18]. Aurora B is involved in the late anaphase and cytokinesis, while Aurora A regulates the initial stages of the mitosis, which include centrosome maturation and separation, mitotic entry, the formation of the bipolar spindle and recruitment and regulations of other proteins at centrosomes during mitosis [5, 10, 23, 44, 48]. Overall AurA is responsible for safeguarding the genomic integrity of the daughter cells [46]. AurA kinase is overexpressed in a wide range of human cancers including ovarian, prostatic, pancreatic, breast, gastric and colon [37]. The overexpression of AurA is known to induce resistance to taxol, a first-line chemotherapeutic agent that targets the spindle checkpoint [2]. Apart from this, AurA has also been shown to phosphorylate and deregulate tumour suppressor proteins such as p53 [75] and BRCA1. The inhibition of AurA in tumour cells results in the premature exit from mitosis and ultimately to cell death. All of these factors suggest that AurA is a potential drug target for treatments against various types of cancers. A number of companies are working to deliver drugs targeting AurA for cancer treatments, and there are already several AurA inhibitors in various phases of clinical trials [6]. For example, alisertib (Phase II), danusertib (Phase II) and VX-689 (completed Phase I) are at various stages [13]; however, all these compounds bind to the ATP binding pocket of the kinase. This poses possible serious selectivity problems as the ATP binding site is highly conserved among different kinases [27]. Hence, there is a pressing need to develop inhibitors that could selectively target AurA with fewer off-target effects. These inhibitors can be considered as the next-generation therapies for treating cancers and various other diseases.

Kinases interact with their protein partners using their allosteric sites. The selective modulation of a kinase could be achieved by targeting those sites as they are not conserved among the members of kinases [76, 85, 110]. AurA interacts with various protein partners such as targeting protein for Xenopus kinesin-like protein 2 (TPX2) [68], Ajuba [52], Bora [57] and N-Myc [86]. The interaction with TPX2 is well characterized both structurally and functionally. TPX2 is a much larger protein than AurA, but the residues 1–43 of TPX2 are sufficient for binding and activating AurA. The phosphorylation of Thr288 residue located within the activation segment of AurA is crucial for the kinase activity. The exposed phosphorylated Thr288

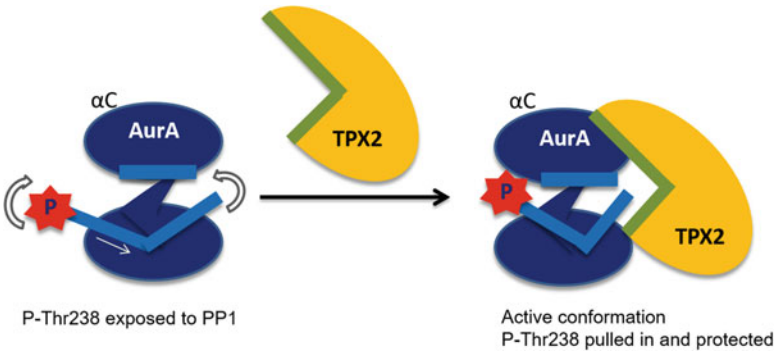


Fig. 11.3 Diagrammatic representation of AurA activation by TPX2

(P-Thr288) residue is susceptible to hydrolysis by phosphatase PP1. TPX2 binding locks AurA into an active conformation and pulls in P-Thr288 residue and keeps it away from PP1 (Fig. 11.3). TPX2 is also responsible for localizing AurA onto spindle microtubules. Mutation (Y8, Y10 and D11) on TPX2 peptide results in spindle defects [11]. Hence, TPX2 is essential for the activation and function of AurA. The X-ray crystal structure of AurA-TPX2(1–43) complex has been well studied. TPX2(1–43) binds to AurA at two sites. The upstream site of TPX2(8–19) binds at N-terminus lobe of the kinase and is responsible for fastening TPX2 on AurA, while the helical downstream site (residues 30–38) interacts with the activation segment of AurA and thus keeping P-Thr288 hidden from PP1 [7]. Both AurA and TPX2 work in concert with each other and this has been identified as an oncogenic holoenzyme. The targeting of this interaction using small molecules would deliver selective inhibitors without targeting the ATP binding site [87]. A number of research groups are working on targeting this protein-protein interaction using single-domain antibody [8], peptides [93] and small molecules [87]. The research is still at an early stage, and no clinical candidates are known as of now for this target. Some of the published works on small-molecule-based AurA-TPX inhibitors are summarized below.

Conti, Bayliss and co-workers were among the pioneers in elucidating the activation mechanism of AurA by TPX2(1–43) peptide using the X-ray crystal structures of phosphorylated AurA and AurA-TPX2(1–43) complex [7]. Based on the structural information and computational studies, they came up with a series of indole and indene derivatives as allosteric inhibitors of AurA [21]. They found that some of these compounds bind to the TPX2 binding site on AurA and decrease the activity of the kinase. However, no X-ray crystal structures of these compounds bound to AurA were reported or have been deposited in the PDB. In the cell-based assay, compounds **1** and **2** showed selectivity of AurA over protein kinase A (PKA) and compound **2** displayed an IC_{50} of 37 μM against AurA-TPX2 interaction (Fig. 11.4). Apart from the work by Conti and co-workers, a handful of other groups recently identified allosteric inhibitors of AurA by using in silico methods. However, no X-ray crystal structures of the inhibitor-AurA complexes were reported to verify the binding site.

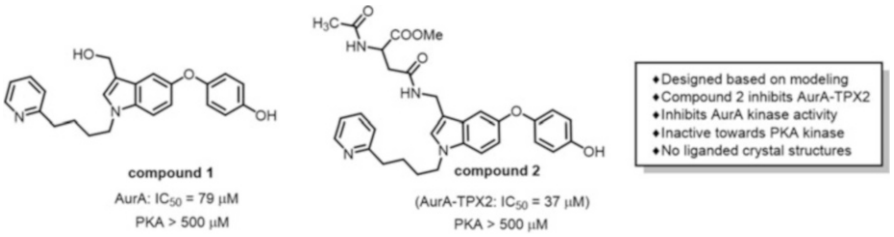


Fig. 11.4 AurA-TPX2 inhibitors developed using computational modelling

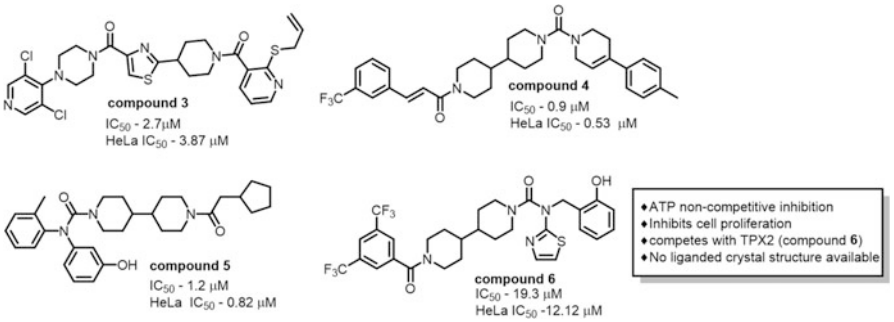


Fig. 11.5 AurA-TPX2 inhibitors identified through HTS

Lewis and co-workers at EMBL, Heidelberg, were among the first to identify non-ATP-competitive small-molecule AurA inhibitors through their high-throughput screening campaign [72]. A commercial library of 53,000 compounds was screened for the AurA kinase activity using standard kinase assay. The active compounds identified were subjected to a fluorescence resonance energy transfer (FRET)-based biochemical assay (Z'Lyte Kinase™ assay) to ascertain if they were ATP-competitive. Four compounds (compounds **3**–**6**, Hela IC_{50} of 0.53–12.12 μM) were identified and shown to be non-ATP-competitive inhibitors (Fig. 11.5). Three of these four compounds are based on 4,4'-bipiperidine scaffold containing urea linkages. The fourth hit was a simple nicotinoyl derivative with two different amides linkages. Further, to ascertain if the compounds bind at the TPX2 binding site on AurA, a TPX2 competition assay was performed using recombinant TPX2 protein. The IC_{50} value of compound **6** increased from 4.8 μM in the absence of TPX2 to 78.8 μM in the presence of TPX2 suggesting that the compound is competing with TPX2 and possibly it is binding on the TPX2 binding site on AurA. However, no X-ray crystal structure of compound **6** bound to AurA was reported.

A team of researchers at the University of Cambridge discovered an allosteric inhibitor against AurA-TPX2 interaction and characterized the binding by X-ray crystal structure in vivo and in vitro studies [58]. At the onset of the study, the team screened a library of 170,000 compounds using an in-house developed FP assay and identified 15 hit compounds. In order to ascertain if these hits are non-ATP-competitive, further screening was performed by blocking the APT binding site on AurA

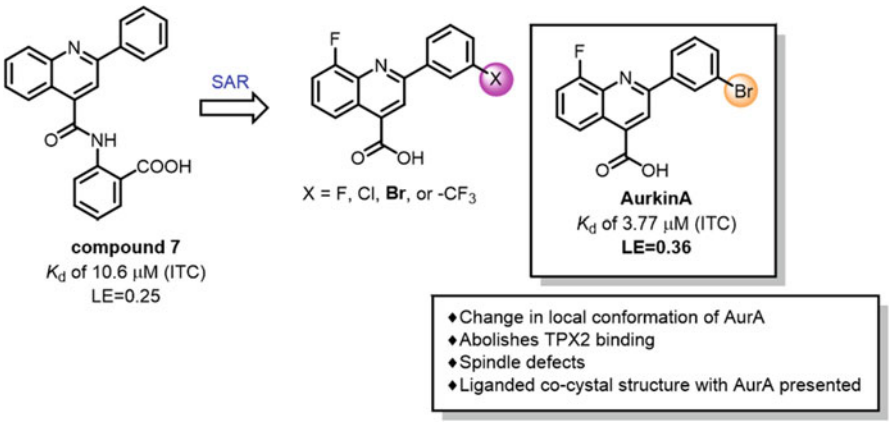


Fig. 11.6 Structurally characterized AurA-TPX2 inhibitor

using a potent Aurora kinase inhibitor JNJ-7706621 (JNJ). The binding of JNJ was also shown to have no significant observed effect on binding of the TPX2(1–43) peptide to AurA. A quinoline-based compound **7** was identified to be the most promising hit with a K_d of 10.6 μM (ITC) and an ligand efficiency (LE) of 0.25 (Fig. 11.6). A small amount of SAR was explored around compound **7**. The Introduction of fluorine and bromine atoms on compound **7** delivered a better affinity (AurkinA) with a K_d of 3.77 μM and LE of 0.36. Upon examination of the X-ray crystal structures, it was found that AurkinA fits into a hydrophobic pocket called the "Y-pocket" on AurA where a part of the upstream stretch of TPX2 containing the YSY motif binds. The binding of AurkinA to the Y-pocket prevents TPX2 from binding to AurA and also induces a structural chance making the kinase inactive but at the same time does not affect the ATP binding, demonstrating a new allosteric inhibition mechanism. In the cell-based assay, it was found that AurkinA was capable of mislocalizing AurA in mitotic cells and also arresting the kinase activity suggesting the druggability of the Y-pocket. This work has laid the foundation for the development of small-molecule-based allosteric inhibitors of AurA using structural studies.

11.3.3 Cyclin-Dependent Kinases (CDKs)

Cyclin-dependent kinases (CDKs) are serine/threonine kinases that are important targets for cancer treatment [70, 98]. The human CDK family consists of 21 members, and some of them (such as CDKs 1–4, 6) regulate the cell cycle processes, while others are involved in cell cycle-independent processes such as regulating transcription, DNA repair and neuronal functions [50, 79]. To date, three inhibitors targeting CDK4/6 have been approved as drugs and all these three drugs,

palbociclib, ribociclib and abemaciclib, are ATP-competitive inhibitors for the treatment of advanced breast cancer [78]. However, over 100 different ATP-competitive inhibitors of CDKs are known, but a majority of them lack selectivity and are associated with toxicity problems [62]. There is a burgeoning interest in the development of allosteric inhibitors of CDK kinases.

Following the discovery of an open allosteric pocket adjacent to the αC-helix of the CDK2, Rastelli and co-workers used computational and docking methods to identify ligands capable of binding into this allosteric pocket [91]. Screening identified a small series of hit compounds, and competition experiments were performed with the ATP binding site ligand staurosporine to confirm that the hit compounds were not binding at the ATP binding site. The best hit was a tetrahydroquinolinone (compound **8**) (Fig. 11.7) which inhibited the proliferation of breast cancer cells with an IC_{50} value of 4 μM. More recently, mutagenesis studies also confirmed the binding of compound **8** to the allosteric pocket on CDK2.

Non-ATP binding site ligands have been identified for CDK8/cyclin C (CycC) complex [99]. This complex is characterized by a deep DMG-out pocket (equivalent to the DGF-out pocket on other kinases) that is amenable for targeting by small molecules. A designed fragment library was screened and this led to the identification of a pyrazolyl urea (compound **9**) (Fig. 11.7) with a long residence time and a K_d value of 3.24 μM against this complex. It was shown by X-ray crystallography that compound **9** binds deep into the DMG-out pocket of CDK8, picking up hydrophobic and H-bonding interactions with the DMG motif and the surrounding residues but without any contact with the hinge region of the kinase.

Zhang and co-workers have discovered non-peptide allosteric inhibitors of CDK2 that could disrupt the CDK2-cyclin A3 interaction [54]. The hit compounds were identified via virtual high-throughput screening and were further evaluated using kinase assay. The imine-based lead compound **10** (Fig. 11.7) displayed an IC_{50} of 52 μM against CDK2-cyclin A3 interaction and was active against an array of human cancer cell lines. Computational and docking studies suggested that the compound **10** fits well into an allosteric pocket close to the interface of the CDK2-cyclin A3 complex.

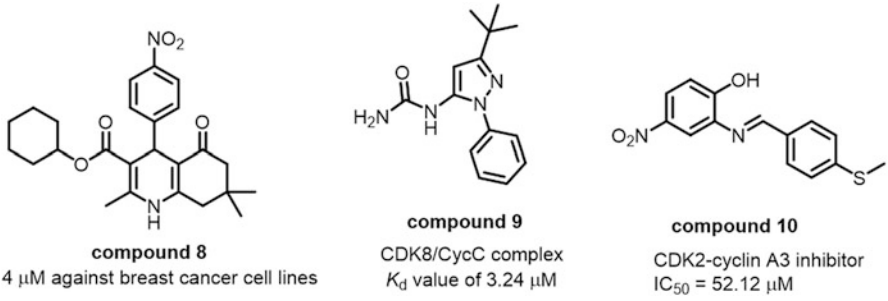


Fig. 11.7 Allosteric CDK inhibitors

11.3.4 CK2α (Casein Kinase 2α)

Casein kinase 2 (CK2) is a serine/threonine selective kinase and is capable of phosphorylating over 300 substrates. These substrates have an acidic consensus sequence flanking the phosphoacceptor residue which uses either ATP or GTP as the phosphate source. In humans, CK2 is a heterotetrameric holoenzyme (α_2/β_2). This is composed of two catalytic subunits (CK2α) attached to a dimer of two regulatory subunits (CK2β). CK2 is involved in a wide range of cellular process and elevated levels of CK2 are associated with a number of different cancer types due to dysregulated CK2 activity. It has been shown that cancer cells are susceptible to CK2 inhibition as they rely on high levels of CK2 to survive. Because of this, the development of inhibitors of CK2 has become a major area for the development of anticancer therapies [14, 25, 90].

A diverse range of compounds have been developed against the ATP binding site of CK2. The most advanced inhibitor CX-4945 (silmitasertib) is a potent, selective and orally bioavailable inhibitor that targets the ATP binding site. Currently, this drug is in clinical trials and has been granted orphan drug status by the FDA for treatment of cholangiocarcinoma, a cancer of the bile cell ducts. While ATP-competitive inhibitors have been shown to be hugely successful in targeting kinases, they do suffer from drawbacks such as selectivity. In the case of CX-4945, this has been shown to inhibit at least 12 other kinases with nanomolar affinities [101]. There have been many elegant strategies to targeting kinase selectivity including covalent inhibition [35]. However, one of the more interesting, all be it challenging, methods is allosteric inhibition.

Hyvonen, Spring and co-workers have developed a small-molecule inhibitor that targets a selective pocket adjacent to the ATP binding site of CK2. This site was identified using a fragment-based approach where X-ray crystallography was used as the primary screening method [15, 16, 28]. Fragment-based drug discovery is a methodology that uses small molecules or fragments as start points for elaboration. These fragments typically have a molecular weight of approximately 180 Da, and while they are weakly binding, they can make optimal interactions with the protein. Once a fragment has been identified, this is then elaborated using a number of different strategies, fragment growing, fragment merging and fragment linking. This methodology has been widely adapted in both academia and industry, and this is now used routinely in conjunction with high-throughput screening (HTS).

A fragment screen was carried out against CK2α and this led to the identification of 3,4-dichlorophenethylamine **11** as a fragment hit by X-ray crystallography. This fragment was observed to bind to CK2α at multiple binding sites. One of the sites identified was the opening of a pocket adjacent to the ATP binding site of the kinase. This new binding site (αD pocket) is located behind the αD-helix and is hydrophobic in character. Upon binding, this fragment displaces Tyr125 from its normal position and releases the αD-helix from the C-lobe opening up this allosteric pocket.

The initial fragment hit **11** was optimized to **12** which was found to have K_D on 270 μM. Upon examination of the X-ray crystal structure of this elaborated

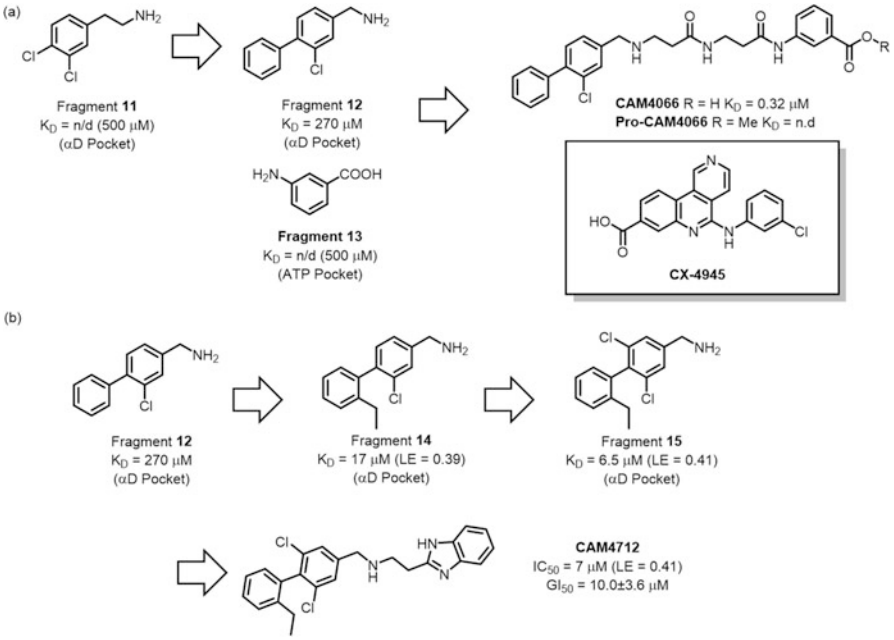


Fig. 11.8 Fragment-based drug discovery of allosteric inhibitors of CK2α (**a**). Fragment **11** was identified to bind to multiple sites on CK2α and one molecule was shown to bind to the αD pocket. This fragment hit was elaborated to fragment **12** which had a measurable K_D of 0.27mM. Fragment **13** was identified by X-ray crystallography to bind to the ATP site and a fragment-linking strategy led to the development of **CAM4066** which was shown to have a K_D of 0.32 μM. The **pro-CAM4066** was shown to have good cell activity, whereas **CAM4066** was shown to have none. The compound CX-4945, which is currently in clinical trials, is shown in the box. (**b**) The fragment **12** was optimized on both of the phenyl rings of the biaryl which led to the development of compound **15** which had an affinity of 7 μM. Subsequent elaboration of the fragment led to the development of CAM4712

fragment, it was hypothesized that this could be linked to fragments bound into the ATP site. In the screening of the fragment library, fragment **13** was identified to bind to the ATP binding site. Interestingly for both the fragment that bound to the αD site and the fragment that bound to the ATP binding site, no measurable K_D (ITC) was obtained for these. When these fragments were linked, this led to the development of compound CAM4066 which was shown to have a K_D of 0.32 μM which has an almost 1000-fold improvement over the fragment **12** (Fig. 11.8a). When this compound was screened in a CK2α kinase assay, this compound was shown to have an IC_{50} of 370 nM. In order to explore the selectivity, CAM4066 was screened across a panel of 52 kinases at a concentration of 2 μM; no other kinases in the panel showed any significant inhibition. The greatest inhibition was shown to be the kinase IGF-1R with an %inhibition of 20% (SD 9%). A number of closely related kinases (DYRK1A, HIPK2, HIPK3 and SRPK1) were shown not to be inhibited by CAM4066. The compound CAM4066 was screened against three different cell lines, HCT116, Jurkat and A549 cells; however they showed no effect, possibly

due to poor cell penetration. When the methyl ester derivative, pro-CAM4066, was screened against the cell lines, this showed good dose response with GI_{50}'s 9 μM (HCT116), 6 μM (Jurkat) and 20 μM (A549). Interestingly when compared to the compound CX4945 (GI_{50} of 5 μM (HCT116), 15 μM (Jurkat) and 17 μM (A549)), this compared favourably although the IC_{50} of CX-4945 is considerably better.

Hyvonen, Spring and co-workers further optimized CAM4066 by improving the physicochemical properties of the molecule [71]. This optimization first focused on the αD site fragment. It was observed that the phenyl ring of the biaryl bound into a hydrophobic pocket; however, a side channel was observed filled by several water molecules. A number of different substituents were introduced onto the *ortho*-position of the ring, and it was found that an ethyl substituent proved beneficial for the affinity. The subsequent introduction of a chlorine atom onto the second phenyl ring in the 6-position led to the development of fragment **15** which had an affinity of 6.5 μM (Fig. 11.8b).

In the development of CAM4066, a number of linkers were explored to build to the ATP site, and it was observed that these compounds opened a small channel between the αD and ATP sites, and the aim was to see whether this small channel could be targeted. This led to the introduction of a benzimidazole ring linked to the fragment at the αD pocket. This compound, CAM4712, was shown to have an IC_{50} of 7 μM; however, due to solubility issues, direct ITC measurements could not be carried out. A number of competition experiments were carried out and a K_D of approximately 4 μM was measured. This compound was screened against the HCT116 cell line and a GI_{50} of 10.0 ± 3.6 μM was obtained which is comparable to that of pro-CAM4066. Upon screening of the compound CAM4712 against a panel of kinases, the selectivity was shown to be reduced where it was observed that inhibition of CAMK1, SmMLCK, EF2K and SGK1 was at over 50%. This study showed a good correlation between the measured IC_{50} and the GI_{50} where no observable drop-off was seen.

11.3.5 Akt (Protein Kinase B)

The serine/threonine kinase Akt, also known as protein kinase B, is involved in the cellular survival pathways. There have been a number of ATP-competitive inhibitors of Akt developed; however, these have shown poor selectivity especially with the AGC class of kinases. Because of this, the development of inhibitors of Atk is still a challenge, and only a handful of small molecules have been developed as lead candidates. One of the best compounds which has been identified is MK-2206 [51]. This compound targets a unique allosteric pocket at the interface of the catalytic kinase domain and the regulatory pleckstrin homology (PH) domain of the inactive conformation of Atk. Currently, this is in Phase I and II clinical trials, and it has been shown to be highly selective to the three Atk isoforms, Atk1, Atk2 and Atk3.

The allosteric inhibitor **MK-2206** stabilizes an inactive conformation where it binds at the interface of the PH and kinase domains. The IC_{50} for this compound is 6.5 ± 0.8 nM, against the wtAkt1 isoform. Rauh and co-workers developed covalent

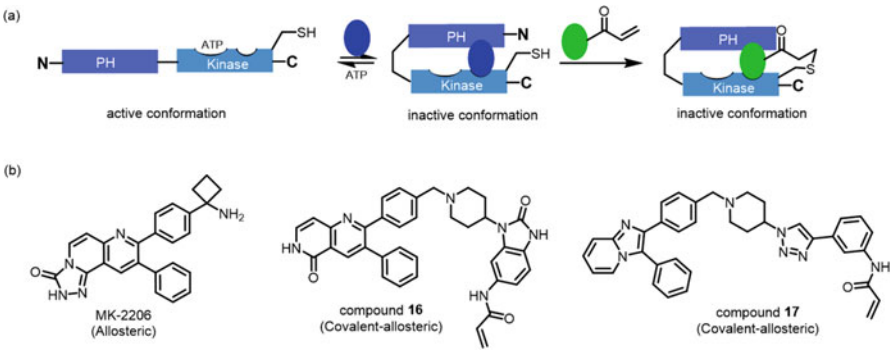


Fig. 11.9 (**a**) Structure of Akt showing the active and inactive conformations. The allosteric inhibitors bind to the interface of the PH and kinase domains. The covalent allosteric inhibitors irreversibly stabilize the inactive conformation through covalent attachment to a cysteine residue in the kinase domain. (**b**) Chemical structures of **MK-2206** and two covalent allosteric inhibitors **16** and **17**

allosteric inhibitors of Akt1 [107]. Two inhibitors were developed based on the 1,6-naphthyridine **16** and imidazo-1,2-pyridine **17** scaffolds. Compound **16** was shown to have an affinity of 0.2 ± 0.1 nM for wtAkt1. The alternative covalent allosteric scaffold was shown to have an IC_{50} of 372 ± 48 nM, which is significantly less than that of MK2206 and the covalent allosteric compound **16** (Fig. 11.9b). The covalent interaction of compounds **16** and **17** was confirmed by ESI-MS when these compounds incubated with wtAkt1. This resulted in the mass increases equivalent to the corresponding single labelled Akt1 (587 and 572 Da, respectively), compared to the control when treated with DMSO. Upon tryptic digest, it was found that compound **16** was bound to Cys296 and compound **17** was covalently bound to Cys301, both of which are in close proximity to the allosteric interface. The covalent allosteric inhibitor **16** shows selectivity for Akt1 when screened against a panel of over 100 kinases. The Akt isoforms 1–3 showed >80% inhibition at a concentration of 1 μM. The kinase MAP 4K5 was shown to be moderately inhibited at 37%. Compound **16** was screened against a range of cancer cell lines and induced a sensitive dose-dependent decrease of pAkt1 at both Thr308 and Ser473 in PC3 and BT474 cancer cell lines.

11.4 Conclusion

Allosteric inhibition of kinases using small molecules is of paramount importance in the kinase drug discovery arena. These small molecules are considered to be the next-generation medicines with high selectivity, fewer off-target effects and less prone to drug resistance problems when compared to that of the ATP site targeting compounds. Apart from potential clinical use, allosteric inhibitors could also act as valuable tools in deciphering new scientific understanding concerning the biological

roles of kinases and their protein partners. However, identification of allosteric site binding ligands is a significant challenge. Several computational and experimental approaches such as fragment-based drug discovery techniques are evolving to identify these ligands. Here, we outline these approaches, with a focus on examples of the successful use of such techniques. On a closing note, research towards kinase allostery is advancing at a faster pace and expected to deliver fruitful outcome in terms of novel therapeutics and provide insights into the complex cellular mechanism of kinases.

Acknowledgements RS acknowledges the School of Pharmaceutical Science and Technology (SPST), Tianjin University for the funding support and Tianjin municipal government for the Tianjin 1000-Talents (Youth) award. RCP acknowledges the Tianjin University for the Post-Doctoral Fellowship support.

References

1. Adrian FJ, Ding Q, Sim T, Velentza A, Sloan C, Liu Y, Zhang G, Hur W, Ding S, Manley P, Mestan J, Fabbro D, Gray NS (2006) Allosteric inhibitors of Bcr-abl-dependent cell proliferation. Nature Chem Biol 2(2):95–102
2. Anand S, Penrhyn-Lowe S, Venkitaraman AR (2003) AURORA-A amplification overrides the mitotic spindle assembly checkpoint, inducing resistance to Taxol. Cancer Cell 3(1):51–62
3. Anscombe E, Meschini E, Mora-Vidal R, Martin MP, Staunton D, Geitmann M, Danielson U, Stanley WA, Wang LZ, Reuillon T, Golding BT, Cano C, Newell DR, Noble ME, Wedge SR, Endicott JA, Griffin RJ (2015) Identification and characterization of an irreversible inhibitor of CDK2. Chem Biol 22:1159–1164
4. Ashwell MA, Lapierre JM, Brassard C, Bresciano K, Bull C, Cornell-Kennon S, Eathiraj S, France DS, Hall T, Hill J, Kelleher E, Khanapurkar S, Kizer D, Koerner S, Link J, Liu Y, Makhija S, Moussa M, Namdev N, Nguyen K, Nicewonger R, Palma R, Szwaya J, Tandon M, Uppalapati U, Vensel D, Volak LP, Volckova E, Westlund N, Wu H, Yang RY, Chan TC (2012) Discovery and optimization of a series of 3-(3-phenyl-3H-imidazo[4,5-b]pyridin-2-yl) pyridin-2-amines: orally bioavailable, selective, and potent ATP-independent Akt inhibitors. J Med Chem 55:5291–5310
5. Barr AR, Gergely F (2007) Aurora-A: the maker and breaker of spindle poles. J Cell Sci 120(Pt 17):2987–2996
6. Bavetsias V, Linardopoulos S (2015) Aurora kinase inhibitors: current status and outlook. Front Oncol 5:278
7. Bayliss R, Sardon T, Vernos I, Conti E (2003) Structural basis of Aurora-A activation by TPX2 at the mitotic spindle. Mol Cell 12(4):851–862
8. Bayliss R, Burgess SG, McIntyre PJ (2017) Switching Aurora-A kinase on and off at an allosteric site. FEBS J 284(18):2947–2954
9. Bembenek SD, Tounge BA, Reynolds CH (2009) Ligand efficiency and fragment-based drug discovery. Drug Discov Today 14:278–283
10. Berdnik D, Knoblich JA (2002) Drosophila Aurora-A is required for centrosome maturation and actin-dependent asymmetric protein localization during mitosis. Curr Biol 12(8):640–647
11. Bird AW, Hyman AA (2008) Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A. J Cell Biol 182(2):289–300
12. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat Rev Drug Discov 11(11):873–886
13. Borisa AC, Bhatt HG (2017) A comprehensive review on Aurora kinase: small molecule inhibitors and clinical trial studies. Eur J Med Chem 140:1–19

14. Bortolato A, Cozza G, Moro S (2008) Protein kinase CK2 inhibitors: emerging anticancer therapeutic agents? Anticancer Agents Med Chem 7(8):789–806
15. Brear P, De Fusco C, Georgiou KH, Francis-Newton NJ, Stubbs CJ, Sore HF, Venkitaraman AR, Abell C, Spring DR, Hyvönen M (2016) Specific inhibition of CK2α from an anchor outside the active site. Chem Sci 7(11):6839–6845
16. Brear P, North A, legre J, Georgiou KH, Lubin A, Carro L, Green W, Sore HF, Hyvönen M, Spring DR (2018) Novel non-ATP competitive small molecules targeting the CK2α/β interface. Bioorg Med Chem 26:3016–3020
17. Burotto M, Chiou VL, Lee JM, Kohn EC (2014) The MAPK pathway across different malignancies: a new perspective. Cancer 120(22):3446–3456
18. Carmena M, Earnshaw WC (2003) The cellular geography of aurora kinases. Nat Rev Mol Cell Biol 4(11):842–854
19. Cohen P, Alessi DR (2013) Kinase drug discovery – what's next in the field? ACS Chem Biol 8:96–104
20. Congreve M, Carr R, Murray C, Jhoti H (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877
21. Conti E, Bayliss R, Schultz C, Vernos I, Sardon T (2006) Crystals of an aurora-A TPX2 complex, TPX2 binding site of aurora-A, aurora-A ligands and their use. WO2005040368
22. Cowan-Jacob SW, Jahnke W, Knapp S (2014) Novel approaches for targeting kinases: allosteric inhibition, allosteric activation and pseudokinases. Future Med Chem 6(5):541–561
23. Cowley DO, Rivera-Pérez JA, Schliekelman M, He YJ, Oliver TG, Lu L, O'Quinn R, Salmon ED, Magnuson T, Van Dyke T (2009) Aurora-A kinase is essential for bipolar spindle formation and early development. Mol Cell Biol 29(4):1059–1071
24. Coyne AG, Scott DE, Abell C (2010) Drugging challenging targets using fragment-based approaches. Curr Opin Chem Biol. 14(3):299–307
25. Cozza G, Pinna LA (2016) Caesin kinases as potential therapeutic targets. Expert Opin Ther Targets 20(3):319–340
26. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR, Futreal PA (2002) Mutations of the BRAF gene in human cancer. Nature 417(6892):949–954
27. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, Hocker M, Treiber DK, Zarrinkar PP (2011) Comprehensive analysis of kinase inhibitor selectivity. Nat Biotechnol 29(11):1046–1051
28. De Fusco C, Brear P, legre J, Georgiou KH, Sore HF, Hyvönen M, Spring DR (2017) A fragment-based approach leading to the discovery of novel binding site and the selective CK2 inhibitor CAM4066. Biorg Med Chem 25(13):3471–3482
29. De Luca A, Maiello MR, D'Alessio A, Pergameno M, Normanno N (2012) The RAS/RAF/MEK/ERK and the PI3K/AKT signalling pathways: role in cancer pathogenesis and implications for therapeutic approaches. Expert Opin Ther Targets 16(Suppl 2):S17–S27
30. Endicott JA, Noble ME, Johnson LN (2012) The structural basis for control of eukaryotic protein kinases. Annu Rev Biochem 81:587–613
31. Erlanson DA, Wells JA, Braisted AC (2004) Tethering: Fragment-based drug discovery. Annu Rev Biophys Biomol Struct 33:199–223
32. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H (2016) Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov 15(9):605–619
33. Faghfuri E, Nikfar S, Niaz K, Faramarzi MA, Abdollahi M (2018) Mitogen-activated protein kinase (MEK) inhibitors to treat melanoma alone or in combination with other kinase inhibitors. Expert Opin Drug Metab Toxicol 14(3):317–330
34. Fang Z, Grütter C, Rauh D (2013) Strategies for the selective regulation of kinases with allosteric modulators: exploiting exclusive structural features. ACS Chem Biol 8(1):58–70

35. Ferguson FM, Gray NS (2018) Kinase inhibitors: the road ahead. Nat Rev Drug Discov 17(5):353–377
36. Ficarro SB, McCleland ML, Stukenberg PT, Burke DJ, Ross MM, Shabanowitz J, Hunt DF, White FM (2002) Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae. Nat Biotechnol 20(3):301–305
37. Fischer PM (2004) The design of drug candidate molecules as selective inhibitors of therapeutically relevant protein kinases. Curr Med Chem 11(12):1563–1583
38. Garuti L, Roberti M, Bottegoni G (2010) Non-ATP competitive protein kinase inhibitors. Curr Med Chem 17(25):2804–2821
39. Gautschi O, Heighway J, Mack PC, Purnell PR, Lara PN, Jr Gandara DR (2008) Aurora kinases as anticancer drug targets. Clin Cancer Res 14(6):1639–1648
40. Gavrin LK, Saiah E (2013) Approaches to discover non-ATP site inhibitors. Med Chem Commun 4:41–51
41. Getlik M, Simard JR, Termathe M, Grütter C, Rabiller M, van Otterlo WA, Rauh D (2012) Fluorophore labelled kinase detects ligands that bind within the MAPK insert of p38α kinase. PLoS One 7(7):e39713
42. Gharwan H, Groninger H (2016) Kinase inhibitors and monoclonal antibodies in oncology: clinical implications. Nat Rev Clin Oncol 13(4):209–227
43. Gibbons DL, Pricl S, Kantarjian H, Cortes J, Quintás-Cardama A (2012) The rise and fall of gatekeeper mutations? The BCR-ABL1 T315I paradigm. Cancer 118(2):293–299
44. Giet R, McLean D, Descamps S, Lee MJ, Raff JW, Prigent C, Glover DM (2002) Drosophila Aurora A kinase is required to localize D-TACC to centrosomes and to regulate astral microtubules. J Cell Biol 156(3):437–451
45. Gilmartin AG, Bleam MR, Groy A, Moss KG, Minthorn EA, Kulkarni SG, Rominger CM, Erskine S, Fisher KE, Yang J, Zappacosta F, Annan R, Sutton D, Laquerre SG (2011) GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. Clin Cancer Res 17(5):989–1000
46. Goldenson B, Crispino JD (2015) The aurora kinases in cell cycle and leukemia. Oncogene 34:537–545
47. Goodwin NC, Cianchetta G, Burgoon HA, Healy J, Mabon R, Strobel ED, Allen J, Wang S, Hamman BD, Rawlins DB (2014) Discovery of a type III inhibitor of LIM knase 2 that binds in a DFG-Out conformation. ACS Med Chem Lett 6(1):53–57
48. Hannak E, Kirkham M, Hyman AA, Oegema K (2001) Aurora-A kinase is required for centrosome maturation in Caenorhabditis elegans. J Cell Biol 155(7):1109–1116
49. Hatzivassiliou G, Haling JR, Chen H, Song K, Price S, Heald R, Hewitt JF, Zak M, Peck A, Orr C, Merchant M, Hoeflich KP, Chan J, Luoh SM, Anderson DJ, Ludlam MJ, Wiesmann C, Ultsch M, Friedman LS, Malek S, Belvin M (2013) Mechanism of MEK inhibition determines efficacy in mutant KRAS- versus BRAF-driven cancers. Nature 501(7466):232–236
50. Heptinstall AB, Adiyasa I, Cano C, Hardcastle IR (2018) Recent advances in CDK inhibitors for cancer therapy. Future Med Chem 10(11):1369–1388
51. Hirai H, Sootome H, Nakatsuru Y, Miyama K, Taguichi S, Tsujikoa K, Ueno Y, Hatch H, Majumder PK, Pan B-S, Kotani H (2010) MM-2206, an allosteric Atk inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs *in vitro* and *in vivo*. Mol Cancer Ther 9(7):1956–1967
52. Hirota T, Naoko K, Saya H (2003) Aurora-A and an interacting activator, the LIM protein Ajuba, are requiredfor mitotic commitment in human cells. Cell 114(5):585–598
53. Holderfield M, Deuker MM, McCormick F, McMahon M (2014) Targeting RAF kinases for cancer therapy: BRAF-mutated melanoma and beyond. Nat Rev Cancer 14(7):455–467
54. Hu Y, Li S, Liu F, Geng L, Shu X, Zhang J (2015) Discovery of novel nonpeptide allosteric inhibitors interrupting the interaction of CDK2/cyclin A3 by virtual screening and bioassays. Bioorg Med Chem Lett 25(19):4069–4073
55. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T, Zhang J (2013) Allosite: a method for predicting allosteric sites. Bioinformatics 29:2357–2359

56. Huang Z, Mou L, Shen Q, Lu S, Li C, Liu X, Wang G, Li S, Geng L, Chen G, Zhang J (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42:D510–D516
57. Hutterer A, Berdnik D, Wirtz-Peitz F, Zigman M, Schleiffer A, Knoblich JA (2006) Mitotic activation of the kinase Aurora-A requires its binding partner Bora. Dev Cell 11(2):147–157
58. Janeček M, Rossmann M, Sharma P, Emery A, Huggins DJ, Stockwell SR, Stokes JE, Tan YS, Almeida EG, Hardwick B, Narvaez AJ, Hyvönen M, Spring DR, McKenzie GJ, Venkitaraman AR (2016) Allosteric modulation of AURKA kinase activity by a small-molecule inhibitor of its protein-protein interaction with TPX2. Sci Rep 6:28528
59. Jeong W, Doroshow JH, Kummar S (2013) US FDA approved oral kinase inhibitors for the treatment of malignancies. Curr Probl Cancer 37(3):110–144
60. Johnson LN, Lewis RJ (2001) Structural basis for control by phosphorylation. Chem Rev 101(8):2209–2242
61. Jokinen E, Koivunen JP (2015) MEK and PI3K inhibition in solid tumours: rationale and evidence to date. Ther Adv Med Oncol 7:170–180
62. Jorda R, Hendrychová D, Voller J, Řezníčková E, Gucký T, Kryštof V (2018) How selective are pharmacological inhibitors of cell-cycle-regulating cyclin-dependent kinases? J Med Chem 61(20):9105–9120
63. Kalesh KA, Liu K, Yao SQ (2009) Rapid synthesis of Abelson tyrosine kinase inhibitors using click chemistry. Org Biomol Chem 7(24):5129–5136
64. Karpov AS, Amiri P, Bellamacina C, Bellance MH, Breitenstein W, Daniel D, Denay R, Fabbro D, Fernandez C, Galuba I, Guerro-Lagasse S, Gutmann S, Hinh L, Jahnke W, Klopp J, Lai A, Lindvall MK, Ma S, Mobitz H, Pecchi S, Rummel G, Shoemaker K, Trappe J, Voliva C, Cowan-Jacob SW, Marzinzik AL (2015) Optimization of a dibenzodiazepine hit to a potent and selective allosteric PAK1 inhibitor. ACS Med Chem Lett 6(7):776–781
65. Koelblinger P, Dornbierer J, Dummer R (2017) A review of binimetinib for the treatment of mutant cutaneous melanoma. Future Oncol 13(20):1755–1766
66. Kolb HC, Sharpless KB (2003) The growing impact of click chemistry on drug discovery. Drug Discov Today 8(24):1128–1137
67. Krimm I, Lancelin JM, Praly JP (2012) Binding evaluation of fragment-based scaffolds for probing allosteric enzymes. J Med Chem 55(3):1287–1295
68. Kufer TA, Silljé HH, Körner R, Gruss OJ, Meraldi P, Nigg EA (2002) Human TPX2 is required for targeting Aurora-A kinase to the spindle. J Cell Biol 158(4):617–623
69. Lamba V, Ghosh I (2012) New directions in targeting protein kinases: focusing upon true allosteric and bivalent inhibitors. Curr Pharm Des 18:2936–2945
70. Law ME, Corsino PE, Narayan S, Law BK (2015) Cyclin-dependent kinase inhibitors as anticancer therapeutics. Mol Pharmacol 88(5):846–852
71. legre J, Braer P, De Fusco C, Yoshida M, Mitchell SL, Rossmann M, Carro L, Sore HF, Hyvönen M, Spring DR (2018) Second-generation CK2α inhibitors targeting the αD pocket. Chem Sci 9(11):3041–3049
72. Lewis J (2007) Aurora Kinase Inhibitors. WO2007115805
73. Lim HY, Merle P, Weiss KH, Yau T, Ross P, Mazzaferro V, Blanc JF, Ma YT, Yen CJ, Kocsis J, Choo SP, Sukeepaisarnjaroen W, Gérolami R, Dufour JF, Gane EJ, Ryoo BY, Peck-Radosavljevic M, Dao T, Yeo W, Lamlertthon W, Thongsawat S, Teufel M, Roth K, Reis D, Childs BH, Krissel H, Llovet JM (2018) Phase II studies with refametinib or refametinib plus sorafenib in patients with RAS-mutated hepatocellular carcinoma. Clin Cancer Res 24(19):4650–4661
74. Lin S, Yang X, Jia S, Weeks AM, Hornsby M, Lee PS, Nichiporuk RV, Iavarone AT, Wells JA, Toste FD, Chang CJ (2017) Redox-based reagents for chemoselective methionine bioconjugation. Science 355:597–602
75. Liu Q, Kaneko S, Yang L, Feldman RI, Nicosia SV, Chen J, Cheng JQ (2004) Aurora-A abrogation of p53 DNA binding and transactivation activity by phosphorylation of serine 215. J Biol Chem 279(50):52175–52182
76. Lu S, Li S, Zhang J (2014) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34(6):1242–1285

77. Luke JL, Ott PA, Shapiro GI (2014) The biology and clinical development of MEK inhibitors for cancer. Drugs 74:2111–2128
78. Lynce F, Shajahan-Haq AN, Swain SM (2018) CDK4/6 inhibitors in breast cancer therapy: Current practice and future opportunities. Pharmacol Ther 191:65–73
79. Malumbres M (2014) Cyclin-dependent kinases. Genome Biol 15(6):122
80. Manning G, Whyte DB, Martinez R, Hunter T, Sundarsanam S (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934
81. Mobitz H, Jahnke W, Cowan-Jacob SW (2017) Expanding the opportunities for modulating kinase targets with allosteric approaches. Curr Top Med Chem 17(1):59–70
82. Modesti M (2018) Fluorescent Labelling of Proteins. Methods Mol Biol 1665:115–134
83. Murray CW, Rees DC (2009) The rise of fragment-based drug discovery. Nat Chem 1(3):187–192
84. Ni D, Lu S, Zhang J (2018) Methods applied for the allosteric site revelation. Encyclopedia of Analytical Chemistry. Wiley.
85. Nussinov R, Tsai CJ, Liu J (2014) Principles of allosteric interactions in cell signaling. J Am Chem Soc 136(51):17692–17701
86. Otto T, Horn S, Brockmann M, Eilers U, Schüttrumpf L, Popov N, Kenney AM, Schulte JH, Beijersbergen R, Christiansen H, Berwanger B, Eilers M (2009) Stabilization of N-Myc is a critical function of Aurora A in human neuroblastoma. Cancer Cell 15(1):67–78
87. Panicker RC, Coyne AG, Srinivasan R (2017) Allosteric targeting of Aurora A kinase using small molecules: a step forward towards next generation medicines? Curr Med Chem doi. https://doi.org/10.2174/0929867324666170727120315
88. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of protein allosteric and regulatory sites. Bioinformatics 30(9):1314–1315
89. Pargellis C, Tong L, Churchill L, Cirillo PF, Gilmore T, Graham AG, Grob PM, Hickey ER, Moss N, Pav S, Regan J (2002) Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. Nat Struct Biol 9(4):268–272
90. Prudent R, Cochet C (2009) New protein kinase CK2 inhibitors: jumping out of the catalytic box. Chem Biol 16(2):112–120
91. Rastelli G, Anighoro A, Chripkova M, Carrassa L, Broggini M (2014) Structure-based discovery of the first allosteric inhibitors of cyclin-dependent kinase 2. Cell Cycle 13(14):2296–2305
92. Rauch J, Volinsky N, Romano D, Kolch W (2011) The secret life of kinases: functions beyond catalysis. Cell Commun Signal. 9(1):23
93. Rennie YK, McIntyre PJ, Akindele T, Bayliss R, Jamieson AG (2016) A TPX2 Proteomimetic has enhanced affinity for Aurora-A due to hydrocarbon stapling of a helix. ACS Chem Biol 11(12):3383–3390
94. Rice KD, Aay N, Anand NK, Blazey CM, Bowles OJ, Bussenius J, Costanzo S, Curtis JK, Defina SC, Dubenko L, Engst S, Joshi AA, Kennedy AR, Kim AI, Koltun ES, Lougheed JC, Manalo JC, Martini JF, Nuss JM, Peto CJ, Tsang TH, Yu P, Johnston S (2013) Novel Carboxamide-Based Allosteric MEK Inhibitors: Discovery and Optimization Efforts toward XL518 (GDC-0973). ACS Med Chem Lett 3(5):416–421
95. Roberts PJ, Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. Oncogene 26(22):3291–3310
96. Roskoski R Jr (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol Res 103:26–48
97. Sadowsky JD, Burlingame MA, Wolan DW, McClendon CL, Jacobson MP, Wells JA (2011) Turning a protein kinase on or off from a single allosteric site via disulfide trapping. Proc Natl Acad Sci USA 108(15):6056–6061
98. Sánchez-Martínez C, Gelbert LM, Lallena MJ, de Dios A (2015) Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. Bioorg Med Chem Lett 25(17):3420–3435
99. Schneider R, Becker C, Simard JR, Cetlik M, Bohlke N, Janning P, Rauh D (2012) Direct binding assay for the detection of type IV allosteric inhibitors of Abl. J Am Chem Soc 134:9138–9141

100. Shang J, Lu S, Jiang Y, Zhang J (2016) Allosteric modulators of MEK1: drug design and discovery. Allosteric modulators of MEK1: drug design and discovery. Chem Biol Drug Des 88(4):485–497
101. Siddiqui-Jain A, Drygin D, Streiner N, Chua P, Pierre F, O'Brien SE, Bilesath J, Omori M, Huser M, Huser N, Ho C, Proffitt C, Schwaebe MK, Ryckman DM, Rice WG, Anderes K (2010) CX-4945, an orally bioavailable selective inhibitor of protein kinase CK2, inhibits prosurvival and angiogenic signalling and exhibits antitumor efficacy. Cancer Res 70(24):10288–10298
102. Simard JR, Getlik M, Grutter C, Pawar V, Wulfert S, Rabiller M, Rauth D (2009) Development of a fluorescent-tagged kinase assay system for the detection and characterization of allosteric kinase inhibitors. J Am Chem Soc 131:13286–13296
103. Song K, Liu X, Huang W, Lu S, Shen Q, Zhang L, Zhang J (2017) Improved method for the identification and validation of allosteric site. J Chem Inf Model 57(9):2358–2363
104. Srinivasan R, Uttamchandani M, Yao SQ (2006) Rapid assembly and in situ screening of bidentate inhibitors of protein tyrosine phosphatases. Org Lett 8(4):713–716
105. Srinivasan R, Li J, Ng SL, Kalesh KA, Yao SQ (2007) Methods of using click chemistry in the discovery of enzyme inhibitors. Nat Protoc 2(11):2655–2664
106. Ward RA, Goldberg FW (2018) Kinase drug discovery: modern approaches. Royal Society of Chemistry, Cambridge
107. Weisner J, Gontla R, van den Westhuizen L, Oeck S, Ketzer J, Janning P, Richters A, Mühlenberg T, Fang Z, Taher A, Jendrossek V, Pelly SC, Bauer S, van Otterlo WAL, Rauh D (2015) Covalent-allosteric kinase inhibitors. Angew Chem Int Ed 54(35):10313–10316
108. Wilhelm S, Carter C, Lynch M, Lowinger T, Dumas J, Smith RA, Schwartz B, Simantov R, Kelley S (2016) Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. Nat Rev Drug Discov 5:835–844
109. Wisinski KB, Tevaarwerk AJ, Burkard ME, Rampurwala M, Eickhoff J, Bell M, Kolesar JM, Flynn C, Liu G (2016) Phase 1 study of an AKT-inhibitor (MK-2206) combined with lapatinib in adult solid tumours followed by dose-expansion in advanced HER2+ breast cancer. Clin Cancer Res 22:2659–2667
110. Wu P, Clausen MH, Nielsen TE (2015) Allosteric small-molecule kinase inhibitors. Pharmacol Ther 156:59–68
111. Xu Y, Wang S, Hu Q, Gao S, Ma X, Zhang W, Shen Y, Chen F, Lai L, Pei J (2018) CavityPlus: a web server for protein cavity detection with pharmacophore modelling, allosteric site identification and covalent ligand binding ability prediction. Nucleic Acids Res 46(W1): W374–W379
112. Yeh TC, Marsh V, Bernat BA, Ballard J, Colwell H, Evans RJ, Parry J, Smith D, Brandhuber BJ, Gross S, Marlow A, Hurley B, Lyssikatos J, Lee PA, Winkler JD, Koch K, Wallace E (2007) Biological characterization of ARRY-142886 (AZD6244), a potent, highly selective mitogen-activated protein kinase kinase 1/2 inhibitor. Clin Cancer Res 13(5):1576–1583
113. Zhang J, Adrian FJ, Jahnke W, Cowan-Jacob SW, Li AG, Iacob RE, Sim T, Powers J, Dierks C, Sun F, Guo GR, Ding Q, Okram B, Choi Y, Wojciechowski A, Deng X, Liu G, Fendrich G, Strauss A, Vajpai N, Grzesiek S, Tuntland T, Liu Y, Bursulaya B, Azam M, Manley PW, Engen JR, Daley GQ, Warmuth M, Gray NS (2010) Targeting Bcr-Abl by combining allosteric with ATP-binding site inhibitors. Nature 463(7280):501–506
114. Zhao Z, Xie L, Bourne PE (2017) Insights into the binding mode of MEK type-III inhibitors. A step towards discovering and designing allosteric kinase inhibitors across the human kinome. PLoS ONE 12(6):e0179936
115. Zheng Z, Bourne PE (2018) Progress with covalent small molecule kinase inhibitors. Drug Discov Today 23(3):727–735
116. Zuccotto F, Ardini E, Casale E, Angiolini M (2010) Through the gatekeeper door: exploiting the active kinase conformation. J Med Chem 53:2681–2694

Chapter 12
Allosteric Regulation of Protein Kinases Downstream of PI3-Kinase Signalling

Alejandro E. Leroux, Lissy Z. F. Gross, Mariana Sacerdoti, and Ricardo M. Biondi

Abstract Allostery is a basic principle that enables proteins to process and transmit cellular information. Protein kinases evolved allosteric mechanisms to transduce cellular signals to downstream signalling components or effector molecules. Protein kinases catalyse the transfer of the terminal phosphate from ATP to protein substrates upon specific stimuli. Protein kinases are targets for the development of small molecule inhibitors for the treatment of human diseases. Drug development has focussed on ATP-binding site, while there is increase interest in the development of drugs targeting alternative sites, i.e. allosteric sites. Here, we review the mechanism of regulation of protein kinases, which often involve the allosteric modulation of the ATP-binding site, enhancing or inhibiting activity. We exemplify the molecular mechanism of allostery in protein kinases downstream of PI3-kinase signalling with a focus on phosphoinositide-dependent protein kinase 1 (PDK1), a model kinase where small compounds can allosterically modulate the conformation of the kinase bidirectionally.

Keywords Allostery · Kinase · PDK1 · PIF pocket · Small molecules · AGC kinase · Allosteric drug · Bidirectional allostery · ATP binding site

A. E. Leroux · L. Z. F. Gross · M. Sacerdoti
Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA) – CONICET – Partner Institute of the Max Planck Society, Buenos Aires, Argentina

R. M. Biondi (✉)
Instituto de Investigación en Biomedicina de Buenos Aires (IBioBA) – CONICET – Partner Institute of the Max Planck Society, Buenos Aires, Argentina

Department of Internal Medicine I, Universitätsklinikum Frankfurt, Frankfurt, Germany

DKTK German Cancer Consortium (DKTK), Frankfurt, Germany

German Cancer Research Center (DKFZ), Heidelberg, Germany
e-mail: biondi@med.uni-frankfurt.de

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_12

12.1 Allostery in Cellular Signalling

Allostery is a mechanism of intramolecular signal transmission whereby local perturbations on a protein affect the structure and dynamics of specific distal regions [1–3]. Today, the definition of allostery also includes cases where dimerisation or oligomerisation is not known to occur [4, 5].

Allostery is central for the passage of the information from the cellular receptors to the intracellular effectors, in a process termed cellular signalling or signal transduction [6]. Protein kinases are enzymes that, once active, have the ability to transfer a terminal phosphate from ATP to protein substrates in a process termed protein phosphorylation. The phosphorylation of proteins catalysed by protein kinases is one key mechanism used by cells to transduce the intracellular signals [7].

There are more than 500 genes coding for protein kinases in humans [8], while most of the proteins in human cells are phosphorylated. In this complex context, protein kinases have evolved the ability to phosphorylate specific substrate proteins at a particular location and time. Protein phosphatases will dephosphorylate the target proteins and stop the protein kinase signalling, stimulating or inhibiting the function of the effector molecules. Thus, protein kinases and protein phosphatases need to act in concert to transmit the appropriate signals and produce the physiological effects.

The phosphates placed at specific locations in substrate proteins can change the properties of the targeted proteins, in particular, enabling them to form hydrogen bonds or salt bridges either intra- or intermolecularly [9]. In turn, the change produced by the phosphorylation may be important to relay the information downstream or to produce the final effect. For example, the binding of insulin to a cellular membrane receptor will produce a cascade of protein phosphorylations catalysed by protein kinases, which increases the activity of PI3-kinase, generating the second messenger Ptdins(3,4,5)P_3; in the presence of Ptdins(3,4,5)P_3, the phosphoinositide-dependent protein kinase-1 (PDK1) will phosphorylate its substrate protein kinase B (Akt/PKB). Once active, Akt/PKB phosphorylates glycogen synthase kinase 3 (GSK3), which reduces its activity. Inhibition of GSK3 eventually leads to the dephosphorylation and activation of glycogen synthase, the enzyme that catalyses the synthesis of glycogen in response to insulin. With a different timing after PI3-kinase activation, PDK1 phosphorylates p70-S6-kinase (S6K) leading to insulin-stimulated protein synthesis. S6K, in turn, participates in a negative feedback loop, inhibiting the pathway. This simplified linear path is then part of a broad network of cellular phosphorylation cascades that are stimulated by insulin and growth factors, which are dysfunctional in cancer, insulin resistance, type II diabetes, neurological disorders, etc. [10]. Allosteric processes coupled to conformational sensors (i.e. PDK1 [11]) are central to the transmission of information within the cell, being arguably the core of the intracellular signal transduction language [6, 12].

12.2 Allostery in Protein Kinases

The sequencing of DNA from cancers that are resistant to protein kinase inhibitors unveiled that numerous mutations that constitutively activate the kinase or promote drug resistance, are located far from the active site and drug-binding site, as observed for Bcr-Abl [13]. Mutations can allosterically affect the active site of the kinase by a static effect, stabilising *on* or *off* states or, alternatively, by affecting the propagation of the allosteric effects [3]. Molecular dynamic simulations verified that the effect of mutations found in cancer can be explained from their allosteric effects [14, 15]. On the other hand, perhaps the most abundant source of information on the allosteric systems in protein kinases arose from information obtained from crystal structures of protein kinases stabilised either in active conformations or in different inactive conformations [16, 17]. Although these are static pictures that only poorly provide a whole vision of the dynamics, they also provide strong support to the existence of allosteric communications that must participate in the conformational changes from inactive to active forms of the kinase.

12.3 Allostery: The Traditional Use and Misuse of the Terminology in the Field of Protein Kinases

The concept of allostery requires the definition of at least two sites. The interaction of a metabolite, an interacting protein, a substrate or a small molecule binding to a first site is said to produce an allosteric effect if it affects the function or interactions at a second site, which is not in direct contact with the first one.

The protein kinase research field named “type I inhibitors” to compounds binding at the ATP-binding site, “type II inhibitors” to those binding at the ATP-binding site and at a neighbouring site and “type III inhibitors” to those targeting a pocket different to the ATP-binding site [18]. However, in the common language for those in the field, the “type III” compounds were often considered equivalent to “truly allosteric” drugs. Indeed, the nomenclature on the mode of action of compounds on protein kinases has been rather confusing regarding the “allosteric” terminology because the wording has been often used in cases where a second site is not known to be affected. The recommendations of the International Union of Basic and Clinical Pharmacology for the nomenclature of receptor allosterism and allosteric ligands is that the term “allosteric” “be reserved for instances where the properties of one ligand (small molecule or protein) are altered upon binding of a second ligand at a nonoverlapping, topographically distinct site and where, ideally, reciprocity in this interaction can be demonstrated” [19]. To avoid confusion, the suggestion is to employ the “allostery” terminology only when two sites are being mentioned, a binding site and a second site that is being affected. In addition, the classical protein kinase inhibitor nomenclature (types I, II and III) should be reserved to describe the location of the compound binding site.

12.4 The Protein Kinase Catalytic Domain

The current chapter focusses on the catalytic domain of protein kinases as an allosteric system. The catalytic domain of the protein kinase PKA was the first one to be crystallised and it is still today the protein kinase prototype [20]; the state of activation of protein kinases are often evaluated in comparison to it (PDB: 1ATP). PKA is an AGC kinase regulated by allosteric effects on the regulatory subunit (PKA-R). PKA-R has two cAMP binding sites and binds to the catalytic domain, occupying the peptide substrate binding site and inhibiting the ability of the catalytic subunit to phosphorylate its substrates. The interaction of cAMP with the regulatory subunit produces conformational changes on PKA-R that free the catalytic subunit from the inhibition, enabling PKA to phosphorylate its substrates. In this case, PKA-R is thus the central regulatory allosteric unit, while the PKA kinase domain is locked in an active conformation by N- and C-terminal extensions to the catalytic domain [21].

The conserved features of protein kinase catalytic domain include the general structure that resembles a "bean", with a small lobe and a large lobe and the ATP-binding site sitting in between the two lobes [22] (Fig. 12.1). Thus, the bottom of the ATP-binding site corresponds to regions on the large lobe, while the ceiling of the ATP-binding site is the small lobe, including the β1 and β2 strands and the Gly-rich loop that connects both β-strands. In all structures of protein kinase in complex with substrates, the region of the substrate to be phosphorylated is extended, sitting at the edge of the small lobe-large lobe interface as first depicted by the interaction of PKA with the peptide pseudosubstrate PKI [22], and also found in substrate peptides bound to insulin receptor kinase [23], Akt/PKB [24], etc.

One important feature from protein kinases is the dynamic open-close conformational changes that open and close relative positions of the small and large lobes. The catalytic domain of the prototype protein kinase PKA was crystalised in "open", "intermediate" and "closed" conformations [21]. The open-close conformational changes correlate with catalysis in PKA and is expected to help in the release of ADP, which is the rate-limiting step in the activity of PKA [25].

The open-close structures reveal the most widely observed dynamic feature conserved in protein kinases, the hinge motion. In the protein kinases that have been studied, the close structure provides the snapshot of the catalytically competent structure, which would enable the transfer of the terminal phosphate of ATP to the peptide substrate.

12.4.1 The Activation Loop

The activation loop is varied in amino acid length but is always present. An important feature is the ability to participate in the activation of many kinases by phosphorylation. In PKA and other AGC kinases, the phosphate at the activation

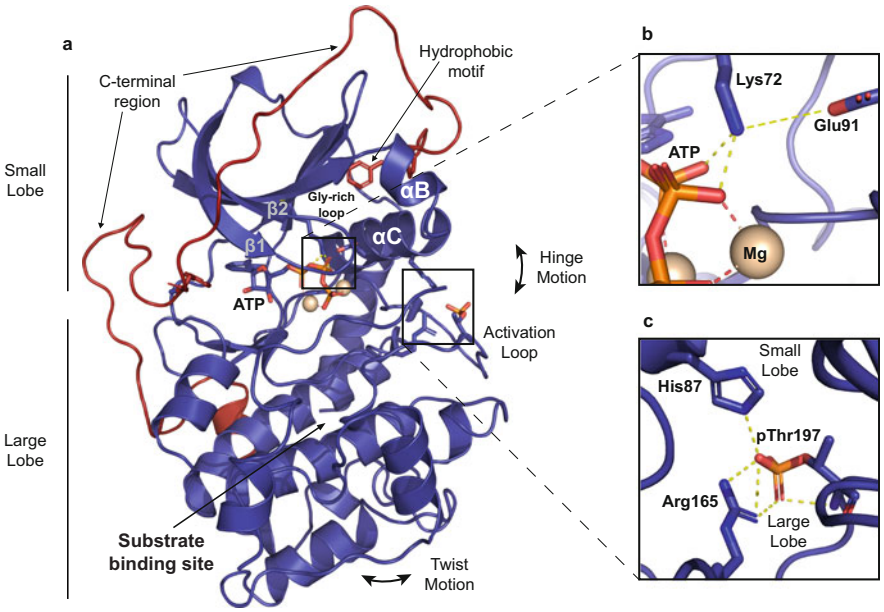


Fig. 12.1 The overall structure of AGC protein kinases. (**a**) The crystal structure of PKA (PDB code 1ATP) is shown in cartoon representation. The C-terminal extension is shown in red, with the Phe residues of the truncated hydrophobic motif (shown as sticks) binding to the PIF-pocket (flanked by the β-sheet, αB and αC helices). Indicated are common features of AGC protein kinases. (**b**) Close-up representation of the Glu-Lys salt bridge that enables the correct position of the ATP phosphates at the active site. (**c**) Close-up representation of the interactions between the phosphorylated residue of the activation loop and residues of the small (His87) and large lobe (Arg165) that contribute to position the activation loop and stabilise the "closed" active structure

loop makes specific interactions with His87 in PKA (Arg129 in PDK1), a residue from the αC helix and Arg165 in PKA (Arg204 in PDK1), located in the large lobe. Thus, the phosphorylation fixes the loop at a particular position in between the small and large lobes. Furthermore, by this fixing, the activation loop phosphate also stabilises a conformation of the kinase and the αC helix. Although the activation of protein kinases by phosphorylation of the activation loop is widely conserved, the mechanism of activation may differ between different kinases.

12.4.2 The ATP-Binding Site

The ATP-binding site has some features that are worth mentioning. The adenine ring makes specific interactions with the so-called hinge region (the hinge between the small and large lobe). This site is almost invariably participating in the interaction with compounds that bind at the ATP-binding site. The phosphates of ATP are positioned for catalysis by their interaction with Mg^{2+} and Lys72 in PKA (Lys111 in

PDK1). In turn, Lys72 in PKA is positioned by a salt bridge interaction with Glu91 (Glu130 in PDK1) from the αC helix, an interaction that turned out to be a hallmark of active protein kinases and is considered one of the most important features in eukaryotic protein kinases.

12.4.3 The αB and αC Helices

In the detailed studies to be described subsequently, we will name the αB helix, which is present in the active forms of AGC kinases. This helix is most often not present in other groups of protein kinases, where it is generally a loop that connects β3 strand with the αC helix. In AGC kinases, the region between αB helix, αC helix and the β-sheet is the region termed "PIF-binding pocket" or PIF-pocket (see Sect. 12.6.2). Since the β-sheet is very stable, the regulatory properties of the PIF-pocket are thus given by the αB and αC helices. One key structural element is the so-called regulatory spine [26], where the αC helix also takes part. It consists of a stack of four hydrophobic residues (RS1/RS2 from the large lobe and RS3/RS4 from the small lobe) that are aligned in active kinases but broken in inactive kinases. RS1 (Tyr164 in PKA) comes from the HRD sequence, RS2 (Phe185 in PKA) from the DFG sequence that marks the beginning of the activation loop, RS3 (Leu95 in PKA) from the αC helix and RS4 (Leu106 in PKA) from the β4 strand.

In addition to the dynamic equilibrium between "open" and "closed" structures, the αC helix (and αB helix) can also be very dynamic and is a key regulatory element in protein kinases from different groups. In PKA, where the catalytic domain is constitutively in the active conformation, the αC helix is stabilised by a "truncated" hydrophobic motif (HM, Phe-Xaa-Xaa-Phe_{350}-COOH) located at the C-terminus of the protein sequence. The Phe residues of the truncated HM dock into the hydrophobic cavity of the PIF-pocket. On the other side of the αC helix, the phosphate from the activation loop brings together the αC helix and the large lobe. Together, the truncated HM and the activation loop stabilise the αC helix from both sides allowing the generation of the Glu91-Lys72 salt bridge. This interaction builds the active regulatory spine, enabling Lys72 to fruitfully position the terminal phosphates from ATP, thus connecting the stabilisation of the helix αC with the catalytic site.

In contrast to the constitutive stability observed in PKA, the αC helix is a central node for the regulation of protein kinases from diverse groups. The regulatory properties of the αC helix were first obtained from crystal structures of CDKs. In CDK2, the inactive structure has the αC helix bended outwards, lacking the Glu-Lys salt bridge. The binding of the cyclin stabilises the active form, bringing the αC helix closer to the catalytic core and establishing the Glu51-Lys33 salt bridge [27, 28]. This inwards/outwards bending of the αC helix and the formation/breakage of the Glu-Lys salt bridge was first described in CDKs and now appears as a feature broadly disseminated in the kinome [29]. Cyclins bind to CDKs occupying a broad surface that encompasses the region in between the αC helix and the β-sheet and also extends to the other side of the αC helix, interacting with the activation loop. While

the overall interaction stabilised the active conformation of the catalytic domain, it was not originally known if any specific interacting region was the key to the activation of CDKs. A considerably smaller viral cyclin (cyclin T1) was later crystallised in complex with CDK9 [30]. The viral cyclin T1 produces the activation of CDK9 by occupying only the region between the αC helix and the β-sheet. Thus, the stabilisation of the αC helix and the activation of CDKs can be achieved by interacting molecules binding at a localised hydrophobic site that is equivalent to the PIF-pocket regulatory site in AGC kinases.

The activation of transmembrane tyrosine kinases like the epidermal growth factor receptor (EGFR) requires the formation of a head-to-tail asymmetric dimer in which an activator kinase domain allosterically activates a receiver kinase [31]. This interaction stabilises the αC helix and is analogous to the activation of AGC kinases by interactions with the HM and the activation of CDKs by cyclin. EGFR allosteric activation by dimerisation is also regulated by the intracellular juxtamembrane segment (for a review, see [32]). More recently, the juxtamembrane region has also been described to play an important role on RET kinase activation. Interactions between the juxtamembrane segment and the αC helix regulate RET kinase activity with striking similarities to the PIF-pocket/hydrophobic motif interaction of AGC kinases [33]. The αC helix also has a regulatory role in non-receptor tyrosine kinases from the Src family [34, 35], in Bruton's tyrosine kinase (BTK) [36], in the zeta chain-associated protein of 70 kDa (ZAP-70) [37, 38], in c-Fes [39] and in protein kinase Csk [40].

In conclusion, a wide variety of mechanisms exist in different kinase families that ultimately affect the conformation of the αC helix, which in turn transfers the information to the active site. The αC helix can be observed as a central hub to relay the effects of inter- and intramolecular interactions to the catalytic pocket. The regulation of AGC kinases, described below in more detail, represents a particular allosteric regulation which resembles the allosteric regulation mediated by the helix αC in many other protein kinases.

12.4.4 The Recognition of Protein Kinase Substrates by Docking Interactions

Most substrates of protein kinases are known to be recognised by means of specific recognition motifs around the actual phosphorylation site [41–45]. However, even early research has found that many protein substrates of kinases had different kinetics from the isolated polypeptide [46]. An alternative mechanism for the specificity of kinases for their substrates is referred to as "docking sites". These are very common in protein kinases from the MAPK family, in CDKs and also in the recognition of substrates by PDK1 [47].

Mitogen-activated protein kinases (MAPKs) are divided in 3 subgroups, extracellular signal-regulated kinases (ERKs), p38 MAP kinases and c-Jun amino-terminal kinases (JNKs). The substrates of the different MAPKs have a very loose recognition motif around the phosphorylation site. To achieve substrate recognition, the substrates possess a separate sequence that docks at a site behind the hinge region, at the back of the kinase domain. Interestingly, a related sequence that binds at the same docking pocket is present in MAPK substrates, in MAPK upstream kinases and in phosphatases that dephosphorylate the MAPK [47–50]. Notably, the fact that the same site is used by different regulators suggests that these docking interactions must be regulated.

Thus, even if the concept of recognition of linear substrate epitopes for phosphorylation has enormously helped the field, it is still not clear if most protein kinase substrates will follow such simple model, or, if alternatively, recognition may be more often mediated by docking interactions or combined models. As will be detailed subsequently in the PDK1 example (Sect. 12.6.2), docking interactions are a possible source of allosteric regulatory mechanisms in protein kinases.

12.5 The Active and Inactive Structures of Protein Kinases

12.5.1 The Active Structures of Protein Kinases

The crystal structure of the active form of PKA in complex with ATP-Mg and PKI unveiled the overall common aspects of the catalytic domain of all protein kinases in active conformations. The crystal structures of all protein kinases in overall active conformations reveal a similar disposition of residues around the ATP-binding site, in the relative position of the two lobes and in the Glu91-Lys72 salt bridge (PKA numbering). As stated earlier, the case of PKA is special in that the catalytic domain structure of the active and inactive conformations are identical. Other protein kinases that are also inhibited with a pseudosubstrate mechanism do not share this property.

The structures of active protein kinases are very well conserved, and the salt bridges equivalent to Glu91-Lys72 are almost invariably present in the protein kinase superfamily. One known exception is the eukaryotic-like Ser/Thr protein kinase PknG from *Mycobacterium tuberculosis*, which is active in the absence of this salt bridge [51].

12.5.2 The Snapshots of Inactive Structures of Protein Kinases

While there is a consensus that the active structures of protein kinases are very conserved, the inactive structures of the protein kinases, on the other hand, are

represented by a broad range of different conformations of the catalytic domain. However, there is a set of structural sites in the catalytic domain that are almost always distorted when a kinase is in an inactive state which include activation loop, Gly-rich loop and αC helix. In addition, the blocking of the ATP-binding pocket by intra- or intermolecular interactions (i.e. pseudosubstrate sequences) can be observed in some inactive structures [52].

For example, in Aurora kinase, c-Abl, B-RAF and insulin receptor, the inactive structure is stabilised by a flipped DFG motif and a non-phosphorylated activation loop sticking into the inactive structure of the catalytic core. This conformation yields an active site that is not aligned for ATP and peptide recognition [53–55].

In Src family protein kinases, phosphorylation of a highly conserved tyrosine residue in the tail region (Tyr-527 in c-Src) induces an intramolecular interaction with the SH2 domain in addition to an interaction of the SH3 domain with the linker connecting the SH2 and catalytic domain. The interactions cause the αC helix to swing outwards into a conformation incompatible with catalytic activity. Therefore, the SH2 and SH3 are essential to achieve an inhibited conformation (for a review, see [35]). The inactive form of Abl shares a lot of similarities with autoinhibited Src, where SH2 and SH3 domains interact with the catalytic domain and negatively regulate the activity. The autoinhibition is stabilised by the binding of an N-terminal myristic acid group to a deep hydrophobic pocket in the large lobe of the catalytic domain which induces a bend in αI helix. The interaction of the myristoylated N-terminus with the large lobe is essential to maintain the autoinhibited state as mutations affecting the interaction yield a highly active kinase. The myristoyl binding pocket is allosterically linked to the active site, as small molecules targeting the regulatory pocket affect the kinase activity (for a review, see [56]).

The earlier snapshots of inactive conformation of kinases reveal a large number of potential mechanisms to keep different protein kinases inactive. To some degree, those inactive structures also show the different regions of the kinase that must have important dynamics, in order to enable the switch between inactive and active structures.

12.6 The AGC Group of Protein Kinases

The AGC kinases comprise group of 63 protein kinases that include protein kinase A, protein kinase G and protein kinase C isoforms that give the name to the group. The group also comprises three pseudokinases, RSKL1, RSKL2 and RSKLK (previously termed Sgk494) [17].

PKAs, PKGs and PKCs acquire their active conformations soon after synthesis and are afterwards regulated by second messengers that bind to the regulatory domains, releasing pseudosubstrate interactions. PRK/PKNs, MRCK and ROCK are important downstream effectors of the GTP-binding protein Rho. NDR and LATS are implicated in the Hippo signalling pathway. RSK and MSK are important

downstream mediators of the Ras-ERK and p38-MAPK signal transduction pathways. Akt/PKB is a central node for PI3K signalling and its deregulation has been linked to cancer, diabetes and cardiovascular and neurological diseases. S6K is also a downstream effector of PI3-kinase, but its activation also requires phosphorylation by mTOR. It is involved on metabolism and protein synthesis control and is part of a negative feedback loop on insulin/growth factor signalling. SGK is involved in cell proliferation, aldosterone and insulin release and the regulation of ion transporters (see reviews [17, 57]).

12.6.1 The αC Helix, the PIF-Pocket and Their Dynamics in AGC Kinases

Experiments have shown that the αC and αB helices in AGC kinases are very dynamic in solution, as evidenced by hydrogen-deuterium exchange (HDX) experiments in PDK1 [58], PKCzeta [59] and PKCiota [60]. In these examples, the αC helix is always well resolved in the crystal structure of the active conformations of these protein kinases; however, the HDX studies of the isolated catalytic domains revealed that both αB and αC helices are prominently exposed to HDX in solution, revealing that the helices are dynamically formed and unformed. The original data from the HDX experiments on the catalytic domain of PDK1 are shown in Fig. 12.2.

Some regions of the protein kinase catalytic domain of PDK1 have a very low exchange throughout experiments lasting several hours. For example, polypeptides 49–69 and 146–155 do not show any HDX throughout the 2-h experiment (see black lines in Fig. 12.2). On the opposite spectrum, polypeptides 218–247 and 225–247 comprising the activation loop show 80% H exchanged for D already at the first time point. This is not surprising since loops are expected to have a maximal exchange in the time frame investigated. In addition to the activation loop, an almost complete exchange is also observed in overlapping polypeptides encompassing the αB and the αC helices (see Fig. 12.2b, c and e). Polypeptides showing time-dependent HDX between 2 min and 2 h correspond to the glycine-rich loop (polypeptides 49–93 and 52–93; Fig. 12.2f, g).

12.6.2 The PIF-Pocket Regulatory Site

PKA has a C-terminal extension to the catalytic core, with a truncated HM (Phe-Xaa-Xaa-Phe$_{350}$-COOH) that docks into a hydrophobic pocket on the small lobe of the kinase domain. The C-terminal extension to the catalytic core is quite divergent in different AGC kinases. However, it was found that most AGC kinases also have an extended HM (Phe-Xaa-Xaa-Phe-**Ser/Thr**-Tyr), often comprising a phosphorylation at a Ser or Thr site, indicated in bold.

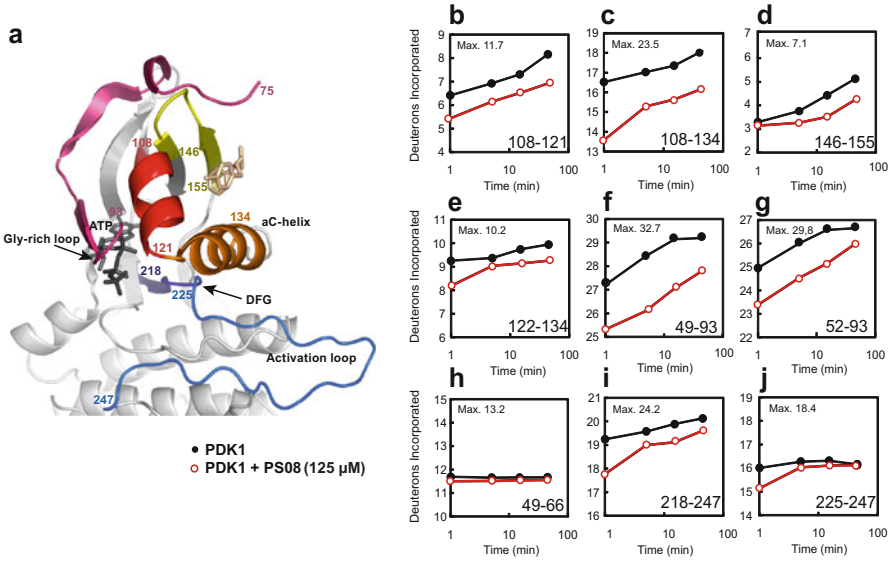


Fig. 12.2 Hydrogen/deuterium exchange coupled with mass spectrometry to measure PDK1 dynamics and the effects of an allosteric compound. (**a**) Structure of the catalytic domain of PDK1 (PBD code 3HRF), indicating the analysed peptides. (**b–j**) Deuterium incorporation at 1, 5, 15 and 45 min in the indicated peptides from PDK1 samples incubated in the presence or absence of the allosteric activator PS08 that binds to the PIF-pocket. Filled black circles correspond to the peptides in the absence of added compound; empty red circles correspond to peptides in the presence of PS08. (Modified from [58])

Similarly to the active structures of PKA, the active structures of different AGC kinases have the HM bound to the PIF-pocket. In contrast to PKA, the binding of the HM to the PIF-pocket in other AGC kinases is the essential regulatory mechanism participating in the concerted activation by the phosphorylation of the activation loop and HM. The phosphorylation of the HM site acts in concert with the activation loop phosphate to stabilise the helices from both sides of the αC helix. This allosteric communication between the PIF-pocket and the ATP-binding site is central to the intramolecular mechanism of phosphorylation-dependent activation of AGC kinases. Also, the PIF-pocket on PDK1 is key to the docking interaction with its AGC kinase substrates.

12.6.2.1 AGC Kinases Constitutively Phosphorylated by PDK1 and Phosphorylated by PDK1 upon Insulin/Growth Factor Stimulation of PI3-Kinase Pathway

The PIF-pocket regulatory site in PDK1 is used for the docking interaction with substrates. A subset of substrates of PDK1, such as PKA, PKG and PKC isoforms, are constitutively phosphorylated by PDK1. Those phosphorylations do not activate

the kinases but keep them competent to be activated by different second messengers (i.e. Ca^{2+} and DAG will activate classical PKCs). On the other hand, a number of other substrates are phosphorylated by PDK1 in response to growth factors and the subsequent stimulation of PI3-kinase signalling (i.e. PKB/Akt, S6K, SGK, RSK, PRK/PKNs).

How is PDK1 regulated so that it can phosphorylate some substrates constitutively and other substrates at different times after growth factor stimulation? How are those AGC kinases activated by phosphorylation?

12.6.2.2 The PIF-Pocket of PDK1 Provides Substrate Recognition, a Regulated Docking Interaction and Differential Timing for the Phosphorylation of Substrates

The C-terminal region of the protein kinase PRK2 was identified by 2-hybrid screening to bind to PDK1 and termed "PDK1 interacting fragment", PIF [61]. In PRK2, the HM has a phosphate-mimicking Asp residue (Phe-Xaa-Xaa-Phe-Asp-Tyr). The binding site for PIF on PDK1 was suggested by the homology of the HM of PRK2 with the truncated HM of PKA. The binding site for the HM of PIF on PDK1 was termed PIF-binding pocket, or PIF-pocket [62].

Most of the substrates of PDK1 require the docking interaction between the PIF-pocket of PDK1 and the hydrophobic motif from substrates. An exception is the phosphorylation of PKB/Akt upon PI3-kinase activation, which is independent of the PIF-pocket docking interaction [63]. The requirement of the docking interaction between the HM from substrates and the PIF-pocket on PDK1 was first described in vitro [62], then validated in knock-in cells that express a PDK1 mutant affected in the PIF-pocket and further validated by small compounds that bind to the PIF-pocket of PDK1 and inhibit S6K phosphorylation and activation but do not inhibit the PI3-kinase-stimulated phosphorylation of PKB/Akt by PDK1 [64–66].

Even then, these protein kinase substrates are phosphorylated at different timing. What defines that a substrate will become phosphorylated by PDK1? The available data supports a model where the substrates acquire the regulated ability to interact with PDK1. In this model, the activity of PDK1 does not need to vary, but instead the substrates are modified to acquire the ability to dock to PDK1 (see Fig. 12.3). In SGK, the phosphorylation of the HM is considered to trigger the interaction with PDK1. In S6K, phosphorylations within an autoinhibitory segment (about 100 C-terminal residues) are required to enable the HM phosphorylation, which enhances the interaction with PDK1 [63]. RSK has two catalytic domains: the C-terminal kinase domain is activated by ERK1/2 and then phosphorylates the HM, triggering the interaction with PDK1 and the activation of the N-terminal kinase domain that phosphorylates transcription factor effectors of the signal [67]. As a note of caution, we should comment that some of the studies that lead us to reach the aforementioned model arise from experiments where the non-phosphorylated forms of substrates also interact with PDK1 under overexpression conditions and the phosphorylations are often mimicked with Asp or Glu.

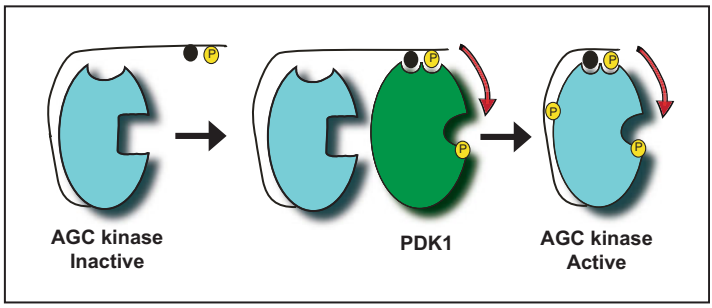


Fig. 12.3 Schematic representation of the regulated docking interaction between PDK1 and its substrates. PDK1 is a conformational sensor that detects the exposed HMs of the inactive forms of AGC kinase substrates. Interaction of phosphorylated HMs of substrates with PDK1 PIF-pocket promotes allosteric effects on the active site of PDK1, increasing its specific activity and enabling the phosphorylation of substrates at the activation loop. In turn, the phosphorylation by PDK1 prompts conformational changes in the substrate, where the binding of the phosphorylated HM to their own PIF-pockets allosterically activates the substrate kinase by stabilising its helix αC and affecting the ATP-binding site

In PRK2, the HM comprised within PIF does not have a phosphorylation site but has an Asp instead. In this case, the binding to PDK1 is regulated by phosphorylation of PRK2 at the turn-motif/zipper site (Thr958), 15 amino acids preceding the HM [68] (see Sect. 12.6.3). The phosphorylation does not directly inhibit the interaction of PRK2 with PDK1 but it indirectly disrupts the interaction by enhancing the intramolecular interaction of the turn-motif/zipper phosphate with a phosphate-binding site on the small lobe of PRK2.

12.6.2.3 Activation of PDK1 and Other AGC Kinases by HM Phosphorylation

Interestingly, the docking interaction with the PIF-pocket of PDK1 not only provides specificity for substrate recognition but the binding also allosterically "activates" PDK1. Phosphorylated HM polypeptides and the particular HM corresponding to PRK2 (REPRILSEEEQEM**F**RD**F**D**Y**IADWC; PIFtide; HM in bold, Asp978 replacing the phosphorylation site underlined) bind to the PIF-pocket of PDK1 and allosterically activate PDK1 [62, 63, 69].

The intrinsic activity of the protein kinase PDK1 can be measured using a peptide substrate derived from the PDK1 phosphorylation site in substrates, i.e. a peptide derived from the activation loop of Akt/PKB, termed T308tide [62, 64]. The isolated peptide is a weak in vitro substrate for the kinase; physiologically, the activation loop polypeptide substrate sites are phosphorylated because PDK1 possesses other means of interaction with substrates, as described earlier. PDKtide was designed by joining PIFtide with T308tide. PDKtide is a vastly improved substrate because it provides PIFtide-mediated docking interaction to the T308tide substrate [62].

The isolated PDK1 has a basal specific activity, which is increased three- to fourfold by the interaction with peptides derived from the HM of its substrates [62]. Similarly to the case with PDK1, other AGC kinases (i.e. Akt/PKB, PKCzeta, RSK, MSK, SGK, S6K) are also activated in vitro by phospho-HM polypeptide but not by the non-phosphorylated peptides (see Fig. 12.4) [70]. Thus, it is considered that the activation of most AGC kinases involves the intramolecular docking of the phosphorylated HM to the PIF-pocket and the consequent allosteric effects on the ATP-binding site.

The HM of PRK2, PIF, has Asp in place of the HM phosphorylation site. The role of Asp978 within the HM of PIF on the activation of PDK1 was investigated by testing the effect of PIFtide or the PIFtide [Asp978Ala] mutated polypeptide on the activation of PDK1. The PIFtide [Asp978Ala] mutated polypeptide had 10 times decreased affinity and fully activated PDK1 at 10 times the concentration [62]. The same happens when the activation of Akt/PKB by PIFtide is studied in comparison to the activation by PIFtide [Asp978Ala] [71]. Together, the data indicates that the negative charge provides increased affinity of the HM to the PIF-pocket and thus participates indirectly in the mechanism of activation of PDK1 and Akt/PKB.

12.6.3 Active Structures of AGC Kinases

The active structures of the catalytic domain of protein kinases are highly conserved. Active structures of AGC kinases PKA (i.e. PDB code 1ATP), PDK1 (i.e. PDB code 1H1W), ROCK (PDB code 2F2U), PRK1/2 (PDB codes 4CRS, 4OTD), Akt/PKB (PDB codes 1O6K, 1O6L, 4EKK, 6BUU)[63, 72, 73], MRCK/DMPK (PDB code 3TKU), PKCiota (PDB code 3A8W), PKCbeta2 (PDB code 2I0E) [74], GSKs (i.e. PDB code 3NYN) [75] as well as the active structure of the related kinase Aurora kinase (PDB code 1OL5) [76] have been published. Except for PDK1, all other active structures have the Phe residues from the HM (Akt/PKB, **F**-P-Q-**F**-S-**Y**; SGK, **F**-L/V-G-**F**-S-**Y**; ROCK, **F**-V/I-G-**F**-T-**Y**; LATS, **F**-Y-E-**F**-T-**F**; PKCiota, **F**-E-G-**F**-E-**Y**) bound to their corresponding PIF-pockets as originally identified for the truncated hydrophobic motif of PKA (**F**-T-E-**F**). The first crystal structure of PDK1 was solved with a sulphate making interactions with Arg131, Thr148, Lys76 and Gln150 and suggested to form the phosphate binding site that interacts with the phosphate from the hydrophobic motif. Indeed, Gln150 and Lys76 provide interactions with phosphate from the HM of Akt/PKB and with the Asp from the HM of PIF in the crystal structures of PDK1 in complex with the HM of Akt/PKB (PDB code 5LVP) [77] and the HM of PRK2 (PIF) (PDB code 4RRV) [66].

An Akt/PKB-PRK2 chimera was produced comprising the PH domain and the catalytic domain of Akt/PKB as a fusion with the C-terminal region of PRK2 (Akt/PKB-CT-PIF), the resulting Akt/PKB-CT-PIF chimera was constitutively active in vitro [63]. The crystal of Akt/PKB-CT-PIF shows a kinase domain stabilised by the C-terminal region of PRK2 (PIF) in the active conformation

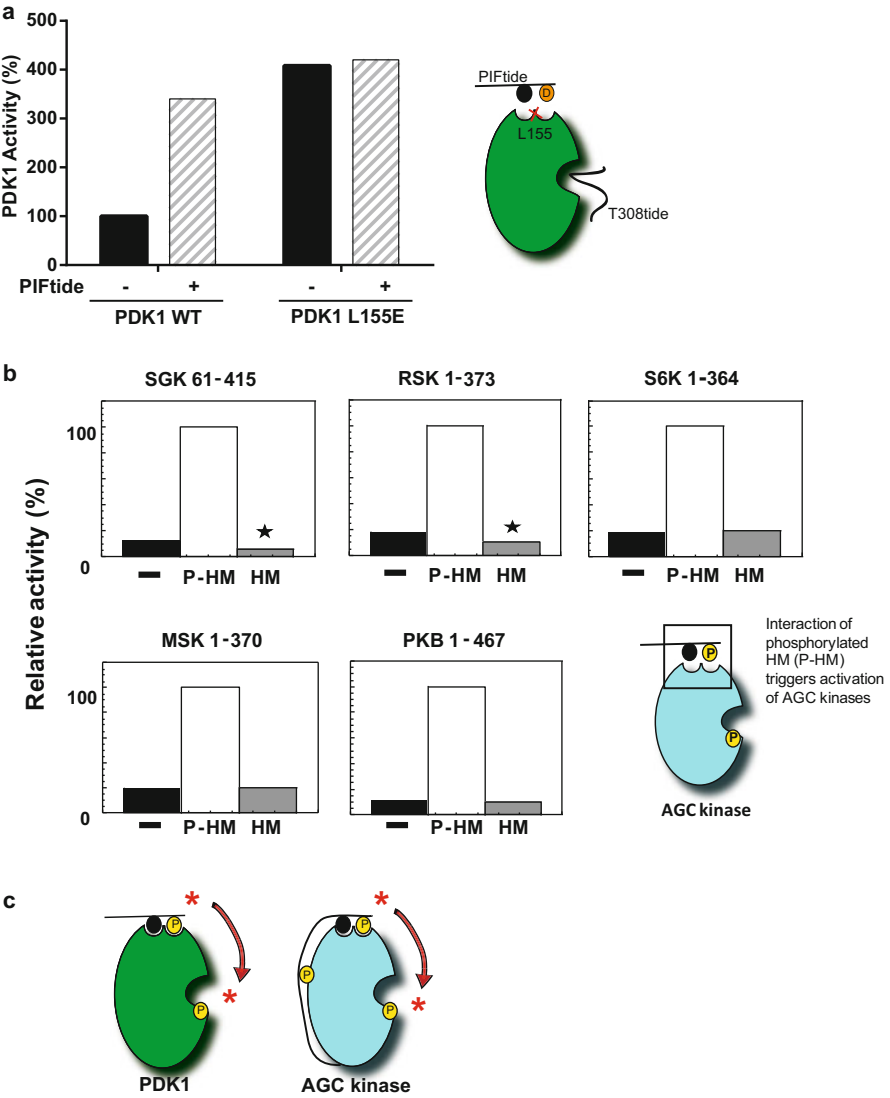


Fig. 12.4 Conserved mechanism of activation of AGC kinases by the binding of HM to the PIF-pocket. (**a**) Wild-type (WT) PDK1 is activated by the HM polypeptide PIFtide, while PDK1 mutant Leu155Glu (L155E) has higher specific activity, which is not modified by PIFtide. For the original data, see [62]. (**b**) The catalytic domains of RSK2, S6K1, MSK1, PKBα/Akt1 and SGK1 are also activated by phosphorylated HMs (pHM) and not affected by the non-phosphorylated HM peptides. The kinase activity is expressed as a percentage of maximal activity. For the original data, see [70]. (**c**) Schematic representation of how the interacting phosphopeptides allosterically affect the ATP-binding site and increase activity. A similar allosteric mechanism of activation by interactions at the site equivalent to the PIF-pocket is expected in other distant protein kinases, like EGFR [31], CDKs [27, 28, 30] and RET [33]

[24]. Notably, the generation of LATS-CT-PIF and NDR-CT-PIF chimeras also produces a constitutively active kinase [78–80]. The binding of the hydrophobic motif to the PIF-pocket of Akt/PKB was important to achieve the active structure. The construct lacking the HM or where the HM was not phosphorylated crystallised inactive [71] (see Sect. 12.6.4).

Besides the activation loop phosphorylation site and the hydrophobic motif phosphorylation site, many AGC kinases also possess a third phosphorylation site, located about 15 amino acids preceding the hydrophobic motif. It was originally termed "turn motif" phosphorylation site, because in PKA, the only crystallised AGC kinase at the time, the phosphorylation is located next to a turn in the C-terminal segment. In PKA, the phosphate points outwards and does not make interactions with the catalytic domain. In contrast, the phosphate from other AGC kinases makes specific interactions with residues from the small lobe, i.e. in PRK2, the turn-motif/zipper phosphates interact with Lys689 and with Arg665 and Lys670 from the Gly-rich loop; therefore, it does not seem equivalent in function to the turn-motif phosphate in PKA. This rather conserved phosphorylation site participates in the activation of AGC kinases by providing an additional interaction of the C-terminal segment to the catalytic domain, enhancing the docking of the HM to the PIF-pocket. Because of this function, the site was also termed "zipper" phosphorylation site [59]. Two of the three residues forming the zipper phosphate binding site (i.e. Arg665 and Lys670 in PRK2) are located in the Gly-rich loop, which corresponds to the "ceiling" of the ATP-binding site, thus suggesting that the phosphorylation at this site could also affect the ATP-binding site more directly. In PRK2, the turn-motif/zipper phosphate plays two known roles: it decreases the binding of PRK2 to PDK1 [68] and promotes the interaction of the C-terminal HM to it's own PIF-pocket, enhancing activity by stabilising the αC helix [59]. The phosphate at the turn-motif/zipper site could in addition support the stabilisation of the active conformation of the ATP-binding site by its interaction with Arg665 and Lys670 located in the Gly-rich loop. More recently, studies on Akt/PKB estimated very high Km values for ATP in comparison to previous work [72]. Interestingly, the Akt/PKB protein used lacked the "zipper" phosphorylation site, while Akt/PKB structure suggested the existence of the conserved binding site for the zipper phosphate (Arg184 from the β-strand preceding the helix αB, Lys160 and Lys165 located in the Gly-rich loop). It is tempting to speculate that the increase Km was due to the lack of the turn motif/zipper phosphorylation. If this were the case, the "zipper" phosphorylation site would have a second function, directly on the conformation of the ATP-binding site, controlling the Km towards ATP.

In summary, the full activity of AGC kinases relies on the stabilisation of the small lobe and the closing of the small lobe into the large lobe. On the one hand, the αC helix is stabilised by the phosphorylated HM binding to the PIF-pocket on one side of the helix and the phosphate of the activation loop at the other side. In turn, the activation loop phosphate interacts with Arg/Lys residue on the large lobe. Together, the chain of interactions produces the stabilisation of the αC helix and the closure of the catalytic domain. On the other hand, the "zipper" phosphorylation stabilises the HM into the PIF-pocket and also stabilises the Gly-rich loop, possibly also affecting

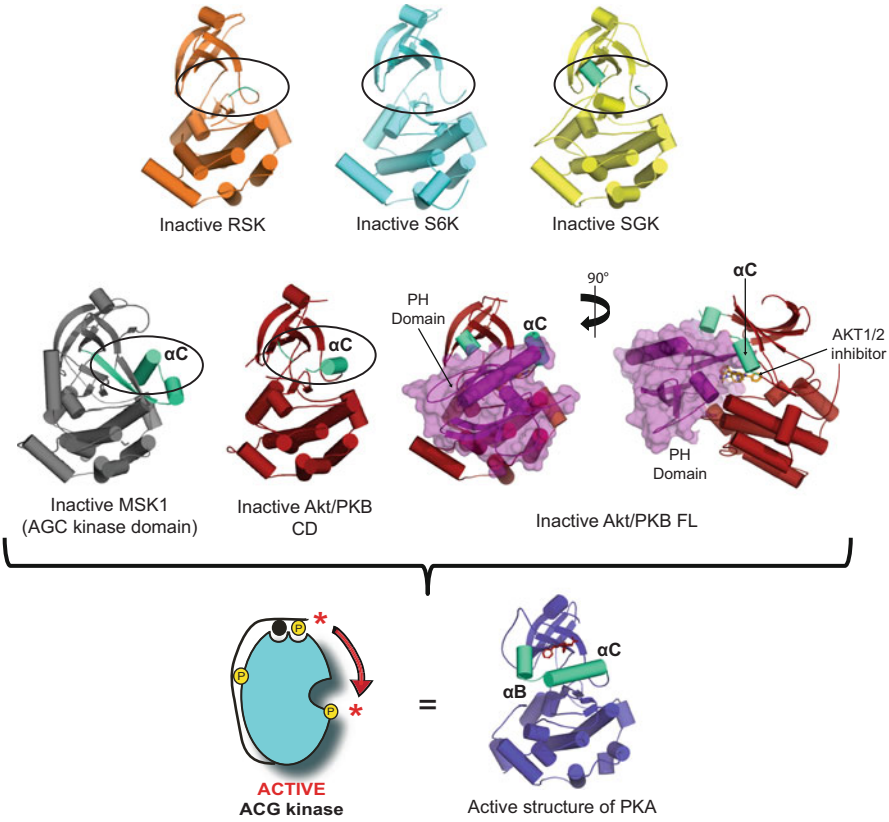


Fig. 12.5 Comparison of different inactive structures of AGC kinases. While the active forms of kinases are all equivalent (represented here by the structure of PKA; PDB code 1ATP), the inhibited forms differ from each other with their own characteristics. Helices αB and αC (marked in green) are affected in all inactive states. RSK (PDB code 2Z7Q), S6K (PDB code 3A60), SGK (PDB code 2R5T), MSK1 (PDB code 1VZO), Akt/PKB catalytic domain (CD; PDB code 1GZN) and Akt/PKB full length (FL; PDB code 3O96)

the conformation of the ATP-binding site. Finally, besides these general aspects, each AGC kinase has its own particular modifications on the general theme.

12.6.4 The Inactive and Inhibited Forms of Selected AGC Kinases: Akt/PKB and PRKs

While the active form of the kinases are all equivalent, the inhibited forms have each their own characteristics (see Fig. 12.5). The catalytic subunit of PKA is inhibited by the polypeptide inhibitor PKI, or the regulatory subunit, like a pseudosubstrate. In this case, the inhibited kinase keeps the active conformation of the catalytic domain.

This is not a generality in protein kinases inhibited by a pseudosubstrate mechanism. The isoforms of PKC are inhibited by an intramolecular pseudosubstrate sequence; however, in these cases, there is no crystal structure of the inhibited kinase domain, and several lines of evidence [60, 81] suggest that the kinase domain is not stabilised by the pseudosubstrate in the active conformation as in PKA.

Akt/PKB is possibly the most complete example of an inactive structure of an AGC kinase. A crystal structure of the inactive conformation of the catalytic domain of Akt/PKB was published in 2002 (i.e. PDB codes 1GZO, 1GZN) [71]. The catalytic domain not phosphorylated at the HM or lacking the C-terminal segment crystallised in an inactive conformation, with disordered regions in the area corresponding to the helices αB and αC [71]. For comparison, the active structure of the catalytic domain was obtained with a constitutively active construct comprising the fusion of the catalytic domain of Akt/PKB with the C-terminal region of the protein kinase PRK2 (Akt/PKB-CT-PIF) [24, 63]. The active structure of the catalytic domain was therefore modulated by the C-terminal regulation and the HM binding to the PIF-pocket on Akt/PKB and stabilising the αB and αC helices.

In the case of Akt/PKB, we know the mechanism of the allosteric inhibition from the crystal structure of the full-length protein in complex with the allosteric inhibitor Akti1-2 (PDB code 3O96) [82]. The structure shows that the PH domain folds onto the small lobe at the site where normally the αB and αC helices form the active PIF-pocket in the active structures (Fig. 12.5). The small compound Akti1-2 sits as the ham of the sandwich between the small lobe and the PH domain of Akt/PKB. Follow-up compounds from Merck (i.e. MK2206 [83]) and Bayer (i.e. BAY1125976 [84]) are also expected to interact with Akt/PKB using a similar mode of action. Interestingly, the HM of Akt/PKB is not present in the inactive structure. One possibility is that in the inactive form the HM interacts with heat shock proteins or alternatively that it could sit at a position overlapping with the Akti1-2 site. The activation of Akt/PKB is considered to require two steps: first, the binding of the second messenger Ptdins(3,4,5)P_3 to the PH domain would open the structure, and only after this step happens, PDK1 would phosphorylate the activation loop. Recent studies show that phosphorylation of Ser473 by mTORC2 can facilitate Akt/PKB activation via the interaction of pSer473 with Arg144 within the PH-kinase linker which would help relieving the autoinhibition by the PH domain [72]. It was shown that the phosphorylation of the activation loop of Akt/PKB by PDK1 upon PI3-kinase activation does not require the docking interaction of its HM with the PIF-pocket of PDK1 [63]. Also it was recently observed that the phosphorylated form of Akt/PKB can be activated directly by Ptdins(3,4,5)P_3 [85]. The dual phosphorylation of Ser477/Thr479 has also been proposed as an alternative activation mechanism [86]. The phosphorylation of those residues within the C-terminal tail can trigger an interaction with the activation loop and relieve Akt/PKB autoinhibition [72]. In addition, recent data shows that Akt/PKB can become phosphorylated and activated independently of Ptdins(3,4,5)P_3, with the requirement of E3 ubiquitin ligase Skp2 and ubiquitination of Akt/PKB [87]. The mechanism of phosphorylation by PDK1 may also be different if the process happens in the presence of Ptdins(3,4,5)P_3 or in its absence. In the absence of Ptdins(3,4,5)P_3, the

activation of Akt/PKB may be dependent on the docking interaction with the PIF-pocket of PDK1, just as the other substrates of PDK1.

12.7 Allostery in AGC Kinases: From the Regulatory PIF-Pocket Site to the ATP-Binding Site

The allosteric communication between the PIF-pocket and the activation loop has been studied in PDK1 using a polypeptide comprising the HM of PRK2, PIFtide, and using small compounds that specifically bind to the PIF-pocket. While PIFtide has 23 amino acids and may exert its effects by interaction at the PIF-pocket and additionally at different sites, the small compounds precisely unveil the effect of the binding solely to the PIF-pocket.

In vitro, PIFtide increases the ability of PDK1 to phosphorylate a peptide substrate derived from the activation loop of PKB/Akt. Small compounds PS46 [64], PS48 (PDB code 3HRF) [58, 88], PS114 (PDB code 4A06) [81, 89], PS189 (PDB code 4AW0), PS210 (PDB code 4AW1) [65, 90] and RS1 (PDB code 4RQK) [66] are reversible allosteric activators that bind to the PIF-pocket as revealed in the crystal structure of the complexes and, like PIFtide (PDB code 4RRV) and phospho-HM of Akt/PKB (PDB code 5LVP), also activate PDK1 in vitro.

The polypeptides or small compound activators do not modify the Km of ATP but rather increase the Vmax, thus providing evidence that more molecules are stabilised in active conformation by the compound activators [62, 64]. The crystal structure of the PDK1-ATP-PS48 complex was very similar to the protein in complex with ATP, with the structure in an overall active conformation—but not fully active conformation because the structure of PDK1 in complex with ATP and PS48 was not in a closed conformation but an "intermediate" conformation—and the beta and gamma phosphates of ATP were not positioned as in the closed-active structure of PKA (see Fig. 12.6a, b) [58]. The structure of PDK1 in complex with ATP and PS48 showed small but significant changes upon the binding of PS48. The deeper binding of PS48 into the PIF-pocket induced local changes (i.e. Phe159 changed conformation). In addition, the structure revealed small changes in distant regions, i.e. the Gly-rich loop region and stabilisation of the activation loop. The allosteric effect of compounds binding at the PIF-pocket site was studied in solution using a fluorescent ATP analogue, TNP-ATP, to probe the conformation of the ATP-binding site (see Fig. 12.6c). Binding of PS48 or PS08 to PDK1 decreases the fluorescence of TNP-ATP [58]. The specificity of the signal was validated using compounds PS47 and PS133, which are stereoisomers of PS48 and PS08 where the carboxylate points to the opposite direction. These control compounds do not activate PDK1 and do not affect the fluorescence of TNP-ATP. The allosteric effect in solution in the presence of PS48 was also studied using HDX experiments (see Fig. 12.2, red lines in comparison to black lines) [58]. Addition of PS48 produced local protection of the polypeptides that are part of the PIF-pocket, i.e. polypeptide from the αB and αC

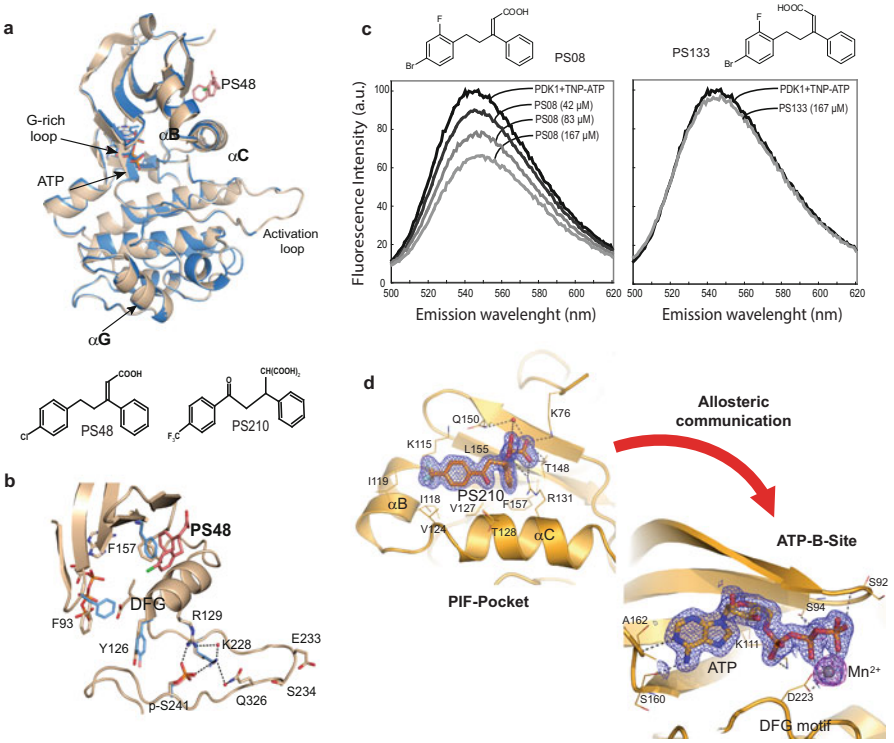


Fig. 12.6 Allosteric effects from the PIF-pocket to the ATP-binding site. (**a** and **b**) Comparison of the structure of the catalytic domain of PDK1 in the presence or absence of the allosteric activator PS48 (PDB code 3HRF). The compound binds to the PIF-pocket, inducing local and allosteric changes. (**c**) The effects of the small compounds on the conformation of the ATP-binding site can be studied using a fluorescent ATP analogue (TNT-ATP). Binding of PS08 (an analogue of PS48) decreases fluorescence, while the stereoisomer PS133, used as control, has no effect. For the original data, see [58]. (**d**) PS210 is a potent PDK1 activator that binds to the PIF-pocket and closes the ATP-binding site as a consequence of the allosteric communication between both sites (PDB code 4AW1 [65])

helices. In addition, PS48 induced protection of overlapping polypeptides that are not in direct contact with the compound, comprising the Gly-rich loop and encompassing a large polypeptide that corresponds to the activation loop and the DFG motif. Interestingly, the binding of PS48 to the PIF-pocket of PDK1 also produced a significant protection of a polypeptide that forms the αG helix within the large lobe of PDK1, at about 30 Å distance from the PIF-pocket [58]. The meaning of this protection remains unclear but may be related to the interaction of PDK1 with full-length substrates. The conformational change induced by the binding to the PIF-pocket could potentially support the binding or the release of substrates from PDK1.

The crystallisation of PDK1 in complex with the potent compound PS210 unveiled the completely closed PDK1 structure (Fig. 12.6d). The PS210-related

compound PS182, which induces half of the maximal activation produced by PS210, did not show the closure of the structure, indicating that the stronger allosteric effect is due to the larger displacement of the Gly-rich loop which makes specific interactions with the uncoordinated β-phosphate of ATP. In the structure of PDK1 in complex with ATP and PS210 (PDB code 4AW1), the nucleotide is positioned exactly as in the structure of the active-closed form of PKA (PDB code 1ATP).

The allosteric communication between the PIF-pocket and the ATP-binding site in PDK1 was also investigated using molecular dynamics simulations. As a first step, non-biased simulations of the catalytic domain of PDK1 in the PDK1-ATP complex identified two dynamic features, the open-close motion and the extension of the helix αB. To fully sample the conformational changes from μs to ms, the detailed analysis was then performed using long multiple-replica parallel tempering (PT; [91]) simulations in the well-tempered ensemble (WTE; [92]). The effects of the compounds on the open-close populations and on the conformation of the helix αB are graphically presented in Fig. 12.7. The distance between residues Asp205 and the Gly-rich loop measures the open-close conformation, while the end-to-end distance of helix αB represents the effects on the conformation of that helix (the smaller distance represents the well-folded helix αB). PDK1 in complex with ATP shows two populations that represent an equilibrium between open and closed conformations and two populations in relation to the conformation of the helix αB. PS210 restricts the kinase open-close dynamics, enforcing a more closed (active) catalytic domain and further stabilises the better folded helix αB.

The aforementioned studies show an allosteric communication between the PIF-pocket and the ATP-binding site of PDK1. A similar allosteric communication is expected to happen in other AGC kinases that are activated by HM-polypeptides binding to the PIF-pocket. In an equivalent manner, AGC kinases also inhibit the kinase catalytic domain by interactions with the αC helix (or what is left from the αC helix), as demonstrated for Akt/PKB and also aPKCs [81]. Small compound variants of PDK1 activators resulted in allosteric inhibitors of aPKCs, i.e. PS171 that is a weak inhibitor of PKCζ was obtained as a variant from PS114 that is an activator of PDK1 [81, 89]. Follow-up small compounds directed to the PIF-pocket of aPKCs have been developed [93–95]. In particular, HDX experiments indicate that the compound PS315 binds to the PIF-pocket with minor allosteric effects, while the crystal structure of PS315 in complex with a chimera comprising the catalytic domain of PDK1 with PIF-pocket residues corresponding to PKCζ shows that PS315 occupies a deep tunnel at the bottom of the PIF-pocket (PDB code 4CT1). PS315 binding disturbs the localisation of Lys111 [60], which positions the terminal phosphates of ATP for catalysis in active structures by participating in the salt bridge with Glu130 (from helix αC). Thus, the compounds binding to the PIF-pocket can also inhibit the kinase activity of aPKCs by subtly affecting the active site of the kinase, allosterically.

A similar phenomenon and similar possibilities for allosteric modulation with small compounds are expected to take place in more evolutionary distant protein kinases where the site equivalent to the PIF-pocket or the stabilisation of the αC helix participates in the activation of kinases. Aurora A is activated by TPX2 [76, 96] and

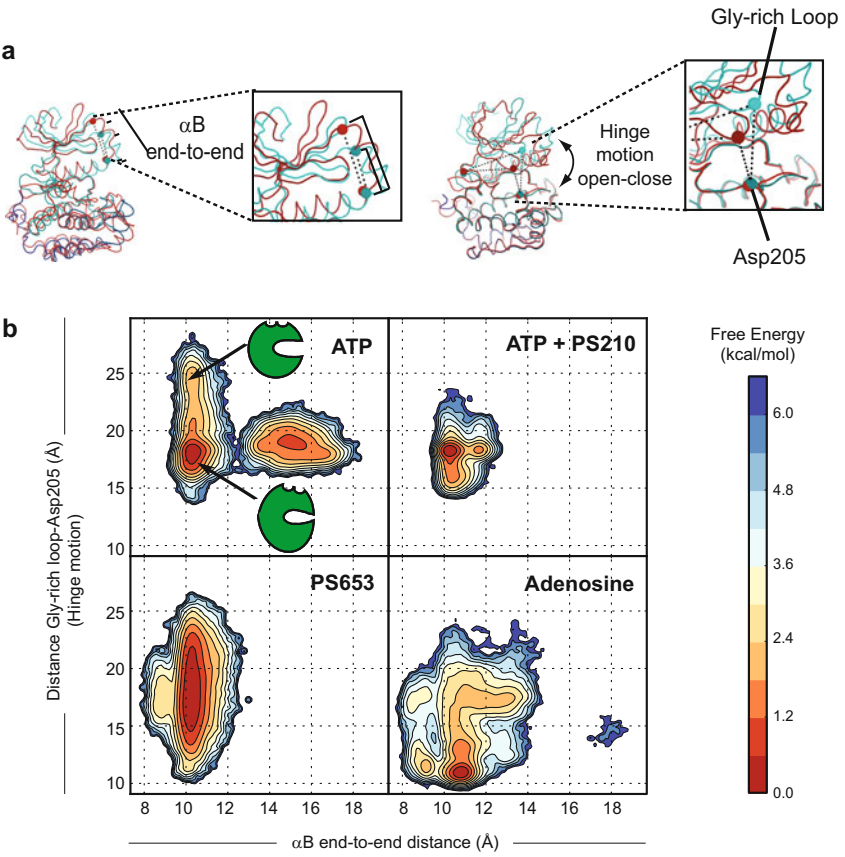


Fig. 12.7 Molecular dynamics simulation to assess kinase dynamics and allosteric effects. (**a**) Schematic representation of the parameters evaluated during the PDK1 simulations: conformation of the αB helix (end-to-end distance) and open-close motion (hinge motion; distance Gly-rich loop and Asp205). (**b**) Free energy calculation of PDK1 in the presence of ATP, PS210, PS653 and adenosine. (For the original data, see [77])

was recently found to be allosterically inhibited by small compounds binding to the TPX2-binding site, which is equivalent to the PIF-pocket [97–100]. CDK7 is activated by viral cyclin T1, which precisely binds into the site equivalent to the PIF-pocket [30]. EGFR is activated by dimerisation at a site equivalent to the PIF-pocket [31]. Interestingly, it is claimed that small allosteric inhibitors of CDKs and EGFR also bind at the site equivalent to the PIF-pocket [101].

Notably, as there is an allosteric communication from the PIF-pocket to the ATP-binding site, then due to thermodynamics, the reverse should also be possible; thus compounds, proteins or intramolecular interactions at the active site could also modulate the conformation or dynamics of the PIF-pocket.

12.8 Allostery in AGC Kinases: From the ATP-Binding Site to the PIF-Pocket Regulatory Site

A screening of libraries of compounds, using a PDK1 homogeneous assay that measures the interaction between PDK1 and the polypeptide PIFtide, identified hits that displaced the interaction by two different mechanisms. Further characterisations proved that some of the compounds that were activators of PDK1 in vitro did not affect the activity of PDK1 [Leu155Glu] (a PIF-pocket mutant that does not bind small compounds) and crystalised at the PIF-pocket, i.e. PSE10 (PDB code 5LVO) [77]. Such kind of compound inhibited the binding of PDK1 to PIFtide by competing with the interaction site with PIFtide (see Fig. 12.8).

More interestingly, another hit compound termed PS653 was identified to displace the interaction of PDK1 with PIFtide (see Fig. 12.8a) but in secondary assays proved to inhibit the activity of PDK1 in vitro, where its potency was maintained against the PDK1 mutated at the PIF-pocket [Leu155Glu]. PS653 crystallised in complex with PDK1 occupying the ATP-binding site (PDB code 5LVL) [77]. Thus, PS653 displaced the interaction of PDK1 with PIFtide allosterically but in the opposite direction to the allosteric activation by compounds such as PS48 and PS210 and by the known physiological mechanism of regulation of the kinase by HM polypeptides from substrates that bind to the PIF-pocket of PDK1.

PS653 is a very small compound that, in the crystal, made specific interactions with the hinge region, similarly to the interactions made by de adenine ring of ATP. Adenine, AMP, ADP and ATP do not produce any significant allosteric effect on the interactions of PDK1 with PIFtide. However, adenosine produces an effect that is the opposite of the allosteric effect produced by PS653. Adenosine binds to the ATP-binding site and enhances the binding of PIFtide at the PIF-pocket [77] (see Fig. 12.8b).The crystal structure of PDK1 in complex with PS653, adenine (PDB code 5LVM) and adenosine (PDB code 5LVN) did not provide clues about the mechanism of the negative or positive allosteric modulation of PIFtide binding.

Molecular dynamics simulations were performed to investigate the mechanism by which PS653 and adenosine modulated the interaction of PDK1 with PIFtide (Fig. 12.7) [77]. PT-WTE simulations show that in the presence of PS653 the interactions between the large lobe and the small lobe are lost. Without such interactions, the PIF-pocket is stabilised with well-structured αB and αC helices but with a more open conformation of the catalytic domain. In contrast, the binding of adenosine stabilises a more wobbly structure of the PIF-pocket, where the overall populations are in a more closed conformation than in the presence of PS653 but with less structured αB and αC helices. These findings suggest that the increased binding of PIFtide is a product from a less rigid structure of the PIF-pocket to enable the binding.

Altogether, the PDK1 model here presented represents a conclusive example of allostery, where the bidirectionality of the system has been depicted with a series of complementing approaches, including biochemical and biophysical studies, synthetic chemistry, crystallography of PDK1 with the allosteric compounds and molecular dynamic simulations.

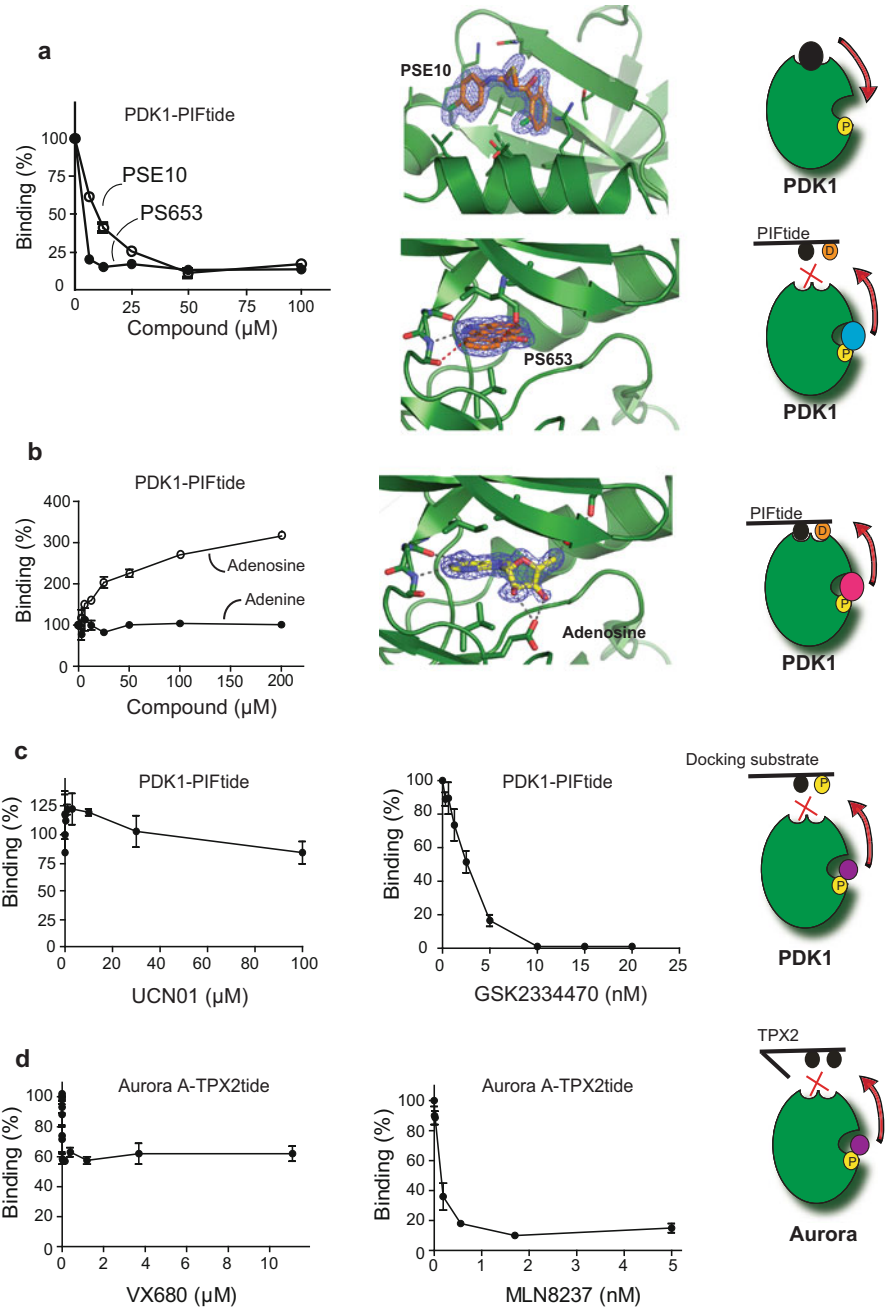


Fig. 12.8 Allosteric effects from the ATP-binding site to the PIF-pocket regulatory site. (**a**) PSE10 and PS653 displace the interaction between PDK1 and PIFtide using an AlphaScreen assay. Crystallographic structures show that PSE10 binds at the PIF-pocket, while PS653 occupies the ATP-binding site. PS653 therefore allosterically *inhibits* the interaction with PIFtide. (**b**) Adenosine binds at the ATP-binding site and allosterically *enhances* the interaction between PDK1 and PIFtide, while adenine has no effect. (**c**) Effect of UCN01 and GSK2334470, two

12.9 Allosteric Modulation of Protein Kinases by Inhibitors (Drugs) Binding to the ATP-Binding Site

Having reviewed the data on bidirectional allosteric communication between the PIF-pocket regulatory site and the ATP-binding site, it becomes obvious that drugs developed by the pharmaceutical industry occupying the ATP-binding site could produce allosteric effects at regulatory regions of the kinase.

UCN01 and GSK2334470 are small compounds developed by the pharmaceutical industry that potently inhibit the protein kinase PDK1 by occupying the ATP-binding site [102, 103]. Interestingly, in cellular assays, GSK2334470 compound was found to inhibit more potently the activation of one of its substrates S6K but not Akt/PKB [104]. We found that the binding of UCN01 on PDK1 does not affect the interaction of PDK1 with PIFtide (see Fig. 12.8c). However, in sharp contrast, GSK2334470, also binding at the ATP-binding site, produced a potent inhibition of the interaction with PIFtide [77]. The findings indicate that GSK2334470 binds to the ATP-binding site and allosterically affects the conformation or dynamics of the PIF-pocket, inhibiting the docking interaction with HM polypeptide. Interestingly, the "reverse allosteric effect" by GSK2334470 could explain the stronger inhibition of S6K activation comparing to Akt/PKB, since S6K requires docking interaction of its HM to the PIF-pocket of PDK1, while PI3-kinase-mediated activation of Akt/PKB does not require the docking to the PIF-pocket of PDK1. The finding suggests that different ATP-binding site drugs in development by the pharmaceutical industry can have important allosteric effects on the regulatory sites, affecting docking interactions with substrates, localisation, etc.

Aurora kinases are closely related to the AGC group of protein kinases. Drugs targeting the ATP-binding site that inhibit the Aurora kinase activity are of interest for the treatment of cancers. Aurora A is activated by TPX2 [76, 96] in a manner that resembles the activation of PDK1 by HM polypeptides [97–100]. TPX2 provides activation and also the proper localisation of the kinase. In addition, Aurora A stabilises the oncogenes N-Myc and C-Myc [105]. The ideal drug against Aurora A would be one that binds to the ATP-binding site and allosterically displaces the interaction with TPX2, affecting the localisation of Aurora and, in addition, destabilising C-Myc and N-Myc to promote their degradation. VX680 and MLN8237 are both potent inhibitors of Aurora A. But they differ in their ability to allosterically inhibit the interaction of Aurora A with TPX2. While VX680 is virtually ineffective in allosterically displacing the interaction with TPX2,

Fig. 12.8 (continued) ATP-competitive inhibitors of PDK1, on the interaction between the catalytic domain of PDK1 and PIFtide. While UCN01 does not show any effect, GSK2334470 strongly displaces the interaction. (**d**) Effect of VX680 and MLN8237, two ATP- competitive inhibitors of Aurora kinase, on the interaction between Aurora A and TPX2tide, a peptide derived from the Aurora kinase regulator TPX2. MLN8237 is a potent inhibitor of the interaction, while VX680 does not affect the interaction. (For the original data, see [77])

MLN8237 binds to the ATP-binding site and allosterically displaces the interaction with TPX2 (Fig. 12.8d) [77]. Gustafson et al. further identified a series of compounds that also target the ATP-binding site on Aurora A and are far more potent than MLN8237 in destabilising N-Myc [106].

Compounds binding at the ATP-binding site can also allosterically enhance the dimer interactions of B-Raf [107, 108] and RNASE activity of IRE1 [109]. Also, compounds targeting the ATP-binding site of the JNK can differentially affect the interaction of peptides that bind at a docking site that is located at the back of the kinase catalytic domain [110].

12.10 Perspectives

The PI3-kinase signalling pathway is important for proper response to insulin. Deregulation of the pathway is responsible for the lack of response to insulin in type II diabetes, while mutations that activate PI3-kinase or inactivate the Ptdins $(3,4,5)P_3$ phosphatase PTEN are present in almost 40% of all cancers. The understanding of the allosteric mechanisms by which the PDK1 and its AGC kinase substrates are regulated can set the bases for innovative drugs to inhibit the signalling in cancers or to specifically activate certain phosphorylations and mimic essential insulin signalling for the treatment of type II diabetes.

The perspectives on the discovery and development of protein kinase inhibitors have evolved through the years. Two decades ago, the perspective was the potential to identify and develop sufficiently selective inhibitors of protein kinases for treatment of human disorders. Later, the perspective was shifted to the possibility of developing allosteric drugs, binding to regulatory sites or docking sites and allosterically affecting the active site and the consequent kinase activity. Today, those perspectives have been achieved for many protein kinases and new developments are constantly being disclosed. The understanding of the allosteric process in protein kinases lags behind those developments. The description of the allosteric mechanism in PDK1 is perhaps the best understood allosteric system within protein kinases, where for the first time, the bidirectionality using small compounds has been described. Those studies reinforce potentially important features both for drug discovery and physiological regulation of protein kinases.

Recent work shows that drugs developed by the pharmaceutical industry targeting the ATP-binding site can have strong allosteric effects and positively or negatively modulate protein kinase interactions. This realisation should encourage the revision of the programs where drugs failed clinical trials and should stimulate checking the allosteric effects of those compounds as well as alternative leads and clinical candidates, which, by having different allosteric effects, could have different effects on patients. On the other hand, the recent detailed studies on the allosteric system on PDK1 identified that adenosine binds to the active site of PDK1 and enhances the interaction of PDK1 with the peptide PIFtide, which binds to the PIF-pocket regulatory site on PDK1. The identification of adenosine suggests that

protein kinase docking interactions could be modulated physiologically by cellular metabolites. Although it seems early to emphasise the importance of the finding, the discovery opens a potentially important regulatory mechanism used by nature to modulate protein kinases signalling by modulating the interaction with substrates and the formation of multiprotein complexes. Such mechanisms of regulation of protein kinase conformations by metabolites targeting the ATP-binding site could also regulate the catalytically dead pseudokinases.

Acknowledgements The work was supported by DFG BI 1044/12-1, CONICET (subsidio P. UE.); ANPCyT (subsidio PICT PRH-2016-4835), ANPCyT (subsidio PICT -2016-3525), FOCEM-Mercosur (COF 03/11) and CONICET.

References

1. Changeux JP (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. Annu Rev Biophys 41:103–133. https://doi.org/10.1146/annurev-biophys-050511-102222
2. Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. Nat Chem Biol 4(8):474–482. https://doi.org/10.1038/nchembio.98
3. Nussinov R, Tsai CJ (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305. https://doi.org/10.1016/j.cell.2013.03.034
4. Changeux JP (2013) 50 years of allosteric interactions: the twists and turns of the models. Nat Rev Mol Cell Biol 14(12):819–829. https://doi.org/10.1038/nrm3695
5. Whittington AC, Larion M, Bowler JM, Ramsey KM, Bruschweiler R, Miller BG (2015) Dual allosteric activation mechanisms in monomeric human glucokinase. Proc Natl Acad Sci U S A 112(37):11553–11558. https://doi.org/10.1073/pnas.1506664112
6. Nussinov R, Tsai CJ, Ma B (2013) The underappreciated role of allostery in the cellular network. Annu Rev Biophys 42:169–189. https://doi.org/10.1146/annurev-biophys-083012-130257
7. Pawson T, Scott JD (2005) Protein phosphorylation in signaling--50 years and counting. Trends Biochem Sci 30(6):286–290
8. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934
9. Hunter T (2012) Why nature chose phosphate to modify proteins. Philos Trans R Soc Lond Ser B Biol Sci 367(1602):2513–2516. https://doi.org/10.1098/rstb.2012.0013
10. Manning BD, Toker A (2017) AKT/PKB signaling: navigating the network. Cell 169(3):381–405. https://doi.org/10.1016/j.cell.2017.04.001
11. Biondi RM (2004) Phosphoinositide-dependent protein kinase 1, a sensor of protein conformation. Trends Biochem Sci 29(3):136–142
12. Tompa P (2016) The principle of conformational signaling. Chem Soc Rev 45(15):4252–4284. https://doi.org/10.1039/c6cs00011h
13. Cowan-Jacob SW, Guez V, Fendrich G, Griffin JD, Fabbro D, Furet P, Liebetanz J, Mestan J, Manley PW (2004) Imatinib (STI571) resistance in chronic myelogenous leukemia: molecular basis of the underlying mechanisms and potential strategies for treatment. Mini-Rev Med Chem 4(3):285–299
14. Saladino G, Gervasio FL (2016) Modeling the effect of pathogenic mutations on the conformational landscape of protein kinases. Curr Opin Struct Biol 37:108–114. https://doi.org/10.1016/j.sbi.2016.01.005

15. Marino KA, Sutto L, Gervasio FL (2015) The effect of a widespread cancer-causing mutation on the inactive to active dynamics of the B-Raf kinase. J Am Chem Soc 137(16):5280–5283. https://doi.org/10.1021/jacs.5b01421
16. Shi Z, Resing KA, Ahn NG (2006) Networks for the allosteric control of protein kinases. Curr Opin Struct Biol 16(6):686–692. https://doi.org/10.1016/j.sbi.2006.10.011
17. Leroux AE, Schulze JO, Biondi RM (2017) AGC kinases, mechanisms of regulation and innovative drug development. Semin Cancer Biol. https://doi.org/10.1016/j.semcancer.2017.05.011
18. Dar AC, Shokat KM (2011) The evolution of protein kinase inhibitors from antagonists to agonists of cellular signaling. Annu Rev Biochem 80:769–795. https://doi.org/10.1146/annurev-biochem-090308-173656
19. Christopoulos A, Changeux JP, Catterall WA, Fabbro D, Burris TP, Cidlowski JA, Olsen RW, Peters JA, Neubig RR, Pin JP, Sexton PM, Kenakin TP, Ehlert FJ, Spedding M, Langmead CJ (2014) International Union of Basic and Clinical Pharmacology. XC. Multisite pharmacology: Recommendations for the nomenclature of receptor allosterism and allosteric ligands. Pharmacol Rev 66(4):918–947. https://doi.org/10.1124/pr.114.008862
20. Taylor SS, Zhang P, Steichen JM, Keshwani MM, Kornev AP (2013) PKA: lessons learned after twenty years. Biochim Biophys Acta 1834(7):1271–1278. https://doi.org/10.1016/j.bbapap.2013.03.007
21. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan M, Taylor SS (2001) Dynamics of cAMP-dependent protein kinase. Chem Rev 101(8):2243–2270
22. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science 253(5018):407–414
23. Hubbard SR (1997) Crystal structure of the activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog. EMBO J 16(18):5572–5581. https://doi.org/10.1093/emboj/16.18.5572
24. Yang J, Cron P, Good VM, Thompson V, Hemmings BA, Barford D (2002) Crystal structure of an activated Akt/protein kinase B ternary complex with GSK3-peptide and AMP-PNP. Nat Struct Biol 9(12):940–944
25. Srivastava AK, McDonald LR, Cembran A, Kim J, Masterson LR, McClendon CL, Taylor SS, Veglia G (2014) Synchronous opening and closing motions are essential for cAMP-dependent protein kinase A signaling. Structure 22(12):1735–1743. https://doi.org/10.1016/j.str.2014.09.010
26. Kornev AP, Haste NM, Taylor SS, Eyck LF (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc Natl Acad Sci U S A 103(47):17783–17788. https://doi.org/10.1073/pnas.0607656103
27. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. Cell 109(3):275–282
28. Jeffrey PD, Russo AA, Polyak K, Gibbs E, Hurwitz J, Massague J, Pavletich NP (1995) Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. Nature 376(6538):313–320. https://doi.org/10.1038/376313a0
29. Huang H, Zhao R, Dickson BM, Skeel RD, Post CB (2012) alphaC helix as a switch in the conformational transition of Src/CDK-like kinase domains. J Phys Chem B 116(15):4465–4475. https://doi.org/10.1021/jp301628r
30. Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, Debreczeni JE, Knapp S, Johnson LN (2008) The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. EMBO J 27(13):1907–1918. https://doi.org/10.1038/emboj.2008.121
31. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125(6):1137–1149

32. Lemmon MA, Schlessinger J, Ferguson KM (2014) The EGFR family: not so prototypical receptor tyrosine kinases. Cold Spring Harb Perspect Biol 6(4):a020768. https://doi.org/10.1101/cshperspect.a020768
33. Plaza-Menacho I, Barnouin K, Barry R, Borg A, Orme M, Chauhan R, Mouilleron S, Martinez-Torres RJ, Meier P, McDonald NQ (2016) RET functions as a dual-specificity kinase that requires allosteric inputs from juxtamembrane elements. Cell Rep 17(12):3319–3332. https://doi.org/10.1016/j.celrep.2016.11.061
34. Register AC, Leonard SE, Maly DJ (2014) SH2-catalytic domain linker heterogeneity influences allosteric coupling across the SFK family. Biochemistry 53(44):6910–6923. https://doi.org/10.1021/bi5008194
35. Engen JR, Wales TE, Hochrein JM, Meyn MA 3rd, Banu Ozkan S, Bahar I, Smithgall TE (2008) Structure and dynamic regulation of Src-family kinases. Cell Mol Life Sci 65(19):3058–3073. https://doi.org/10.1007/s00018-008-8122-2
36. Marcotte DJ, Liu YT, Arduini RM, Hession CA, Miatkowski K, Wildes CP, Cullen PF, Hong V, Hopkins BT, Mertsching E, Jenkins TJ, Romanowski MJ, Baker DP, Silvian LF (2010) Structures of human Bruton's tyrosine kinase in active and inactive conformations suggest a mechanism of activation for TEC family kinases. Protein Sci 19(3):429–439. https://doi.org/10.1002/pro.321
37. Deindl S, Kadlecek TA, Brdicka T, Cao X, Weiss A, Kuriyan J (2007) Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. Cell 129(4):735–746. https://doi.org/10.1016/j.cell.2007.03.039
38. Kannan N, Neuwald AF, Taylor SS (2008) Analogous regulatory sites within the alphaC-beta4 loop regions of ZAP-70 tyrosine kinase and AGC kinases. Biochim Biophys Acta 1784(1):27–32. https://doi.org/10.1016/j.bbapap.2007.09.007
39. Filippakopoulos P, Kofler M, Hantschel O, Gish GD, Grebien F, Salah E, Neudecker P, Kay LE, Turk BE, Superti-Furga G, Pawson T, Knapp S (2008) Structural coupling of SH2-kinase domains links Fes and Abl substrate recognition and kinase activation. Cell 134(5):793–803
40. Ogawa A, Takayama Y, Sakai H, Chong KT, Takeuchi S, Nakagawa A, Nada S, Okada M, Tsukihara T (2002) Structure of the carboxyl-terminal Src kinase, Csk. J Biol Chem 277(17):14351–14354. https://doi.org/10.1074/jbc.C200086200
41. Kemp BE, Bylund DB, Huang TS, Krebs EG (1975) Substrate specificity of the cyclic AMP-dependent protein kinase. Proc Natl Acad Sci U S A 72(9):3448–3452
42. Zetterqvist O, Ragnarsson U, Humble E, Berglund L, Engstrom L (1976) The minimum substrate of cyclic AMP-stimulated protein kinase, as studied by synthetic peptides representing the phosphorylatable site of pyruvate kinase (type L) of rat liver. Biochem Biophys Res Commun 70(3):696–703
43. Yaffe MB, Leparc GG, Lai J, Obata T, Volinia S, Cantley LC (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat Biotechnol 19(4):348–353. https://doi.org/10.1038/86737
44. Pinna LA, Ruzzene M (1996) How do protein kinases recognize their substrates? Biochim Biophys Acta 1314(3):191–225
45. Kettenbach AN, Wang T, Faherty BK, Madden DR, Knapp S, Bailey-Kellogg C, Gerber SA (2012) Rapid determination of multiple linear kinase substrate motifs by mass spectrometry. Chem Biol 19(5):608–618. https://doi.org/10.1016/j.chembiol.2012.04.011
46. Kemp BE, Pearson RB (1991) Design and use of peptide substrates for protein kinases. Methods Enzymol 200:121–134
47. Biondi RM, Nebreda AR (2003) Signalling specificity of Ser/Thr protein kinases through docking-site-mediated interactions. Biochem J 372(Pt 1):1–13
48. Gavin AC, Nebreda AR (1999) A MAP kinase docking site is required for phosphorylation and activation of p90(rsk)/MAPKAP kinase-1. Curr Biol 9(5):281–284
49. Kallunki T, Deng T, Hibi M, Karin M (1996) c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. Cell 87(5):929–939

50. Xu B, Stippec S, Robinson FL, Cobb MH (2001) Hydrophobic as well as charged residues in both MEK1 and ERK2 are important for their proper docking. J Biol Chem 276(28):26509–26515
51. Lisa MN, Gil M, Andre-Leroux G, Barilone N, Duran R, Biondi RM, Alzari PM (2015) Molecular basis of the activity and the regulation of the eukaryotic-like S/T protein kinase PknG from Mycobacterium tuberculosis. Structure 23(6):1039–1048. https://doi.org/10.1016/j.str.2015.04.001
52. Engh RA, Bossemeyer D (2002) Structural aspects of protein kinase control-role of conformational flexibility. Pharmacol Ther 93(2-3):99–111
53. Cheetham GM, Knegtel RM, Coll JT, Renwick SB, Swenson L, Weber P, Lippke JA, Austen DA (2002) Crystal structure of aurora-2, an oncogenic serine/threonine kinase. J Biol Chem 277(45):42419–42422. https://doi.org/10.1074/jbc.C200426200
54. Hubbard SR, Wei L, Ellis L, Hendrickson WA (1994) Crystal structure of the tyrosine kinase domain of the human insulin receptor. Nature 372(6508):746–754. https://doi.org/10.1038/372746a0
55. Wan PT, Garnett MJ, Roe SM, Lee S, Niculescu-Duvaz D, Good VM, Jones CM, Marshall CJ, Springer CJ, Barford D, Marais R (2004) Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell 116(6):855–867
56. Panjarian S, Iacob RE, Chen S, Engen JR, Smithgall TE (2013) Structure and dynamic regulation of Abl kinases. J Biol Chem 288(8):5443–5450. https://doi.org/10.1074/jbc.R112.438382
57. Arencibia JM, Pastor-Flores D, Bauer AF, Schulze JO, Biondi RM (2013) AGC protein kinases: from structural mechanism of regulation to allosteric drug development for the treatment of human diseases. Biochim Biophys Acta 1834(7):1302–1321. https://doi.org/10.1016/j.bbapap.2013.03.010
58. Hindie V, Stroba A, Zhang H, Lopez-Garcia LA, Idrissova L, Zeuzem S, Hirschberg D, Schaeffer F, Jorgensen TJD, Engel M, Alzari PM, Biondi RM (2009) Structure and allosteric effects of low molecular weight activators on the protein kinase PDK1. Nat Chem Biol 5(10):758–764. https://doi.org/10.1038/nchembio.208
59. Hauge C, Antal TL, Hirschberg D, Doehn U, Thorup K, Idrissova L, Hansen K, Jensen ON, Jorgensen TJ, Biondi RM, Frodin M (2007) Mechanism for activation of the growth factor-activated AGC kinases by turn motif phosphorylation. EMBO J 26(9):2251–2261
60. Zhang H, Neimanis S, Lopez-Garcia LA, Arencibia JM, Amon S, Stroba A, Zeuzem S, Proschak E, Stark H, Bauer AF, Busschots K, Jorgensen TJ, Engel M, Schulze JO, Biondi RM (2014) Molecular mechanism of regulation of the atypical protein kinase C by N-terminal domains and an allosteric small compound. Chem Biol 21(6):754–765. https://doi.org/10.1016/j.chembiol.2014.04.007
61. Balendran A, Casamayor A, Deak M, Paterson A, Gaffney P, Currie R, Downes CP, Alessi DR (1999) PDK1 acquires PDK2 activity in the presence of a synthetic peptide derived from the carboxyl terminus of PRK2. Curr Biol 9(8):393–404
62. Biondi RM, Cheung PC, Casamayor A, Deak M, Currie RA, Alessi DR (2000) Identification of a pocket in the PDK1 kinase domain that interacts with PIF and the C-terminal residues of PKA. EMBO J 19(5):979–988
63. Biondi RM, Kieloch A, Currie RA, Deak M, Alessi DR (2001) The PIF-binding pocket in PDK1 is essential for activation of S6K and SGK, but not PKB. EMBO J 20(16):4380–4390
64. Engel M, Hindie V, Lopez-Garcia LA, Stroba A, Schaeffer F, Adrian I, Imig J, Idrissova L, Nastainczyk W, Zeuzem S, Alzari PM, Hartmann RW, Piiper A, Biondi RM (2006) Allosteric activation of the protein kinase PDK1 with low molecular weight compounds. EMBO J 25(23):5469–5480
65. Busschots K, Lopez-Garcia LA, Lammi C, Stroba A, Zeuzem S, Piiper A, Alzari PM, Neimanis S, Arencibia JM, Engel M, Schulze JO, Biondi RM (2012) Substrate-selective inhibition of protein kinase PDK1 by small compounds that bind to the PIF-pocket allosteric docking site. Chem Biol 19(9):1152–1163. https://doi.org/10.1016/j.chembiol.2012.07.017

66. Rettenmaier TJ, Sadowsky JD, Thomsen ND, Chen SC, Doak AK, Arkin MR, Wells JA (2014) A small-molecule mimic of a peptide docking motif inhibits the protein kinase PDK1. Proc Natl Acad Sci U S A 111(52):18590–18595. https://doi.org/10.1073/pnas.1415365112
67. Frodin M, Jensen CJ, Merienne K, Gammeltoft S (2000) A phosphoserine-regulated docking site in the protein kinase RSK2 that recruits and activates PDK1. EMBO J 19(12):2924–2934
68. Dettori R, Sonzogni S, Meyer L, Lopez-Garcia LA, Morrice NA, Zeuzem S, Engel M, Piiper A, Neimanis S, Frodin M, Biondi RM (2009) Regulation of the interaction between protein kinase C-related protein kinase 2 (PRK2) and its upstream kinase, 3-phosphoinositide-dependent protein kinase 1 (PDK1). J Biol Chem 284(44):30318–30327. https://doi.org/10.1074/jbc.M109.051151
69. Biondi RM, Komander D, Thomas CC, Lizcano JM, Deak M, Alessi DR, van Aalten DM (2002) High resolution crystal structure of the human PDK1 catalytic domain defines the regulatory phosphopeptide docking site. EMBO J 21(16):4219–4228
70. Frodin M, Antal TL, Dummler BA, Jensen CJ, Deak M, Gammeltoft S, Biondi RM (2002) A phosphoserine/threonine-binding pocket in AGC kinases and PDK1 mediates activation by hydrophobic motif phosphorylation. EMBO J 21(20):5396–5407
71. Yang J, Cron P, Thompson V, Good VM, Hess D, Hemmings BA, Barford D (2002) Molecular mechanism for the regulation of protein kinase B/Akt by hydrophobic motif phosphorylation. Mol Cell 9(6):1227–1240
72. Chu N, Salguero AL, Liu AZ, Chen Z, Dempsey DR, Ficarro SB, Alexander WM, Marto JA, Li Y, Amzel LM, Gabelli SB, Cole PA (2018) Akt Kinase Activation Mechanisms Revealed Using Protein Semisynthesis. Cell 174(4):897–907.e814. https://doi.org/10.1016/j.cell.2018.07.003
73. Lin K, Lin J, Wu WI, Ballard J, Lee BB, Gloor SL, Vigers GP, Morales TH, Friedman LS, Skelton N, Brandhuber BJ (2012) An ATP-site on-off switch that restricts phosphatase accessibility of Akt. Sci Signal 5(223):ra37. https://doi.org/10.1126/scisignal.2002618
74. Grodsky N, Li Y, Bouzida D, Love R, Jensen J, Nodes B, Nonomiya J, Grant S (2006) Structure of the catalytic domain of human protein kinase C beta II complexed with a bisindolylmaleimide inhibitor. Biochemistry 45(47):13970–13981. https://doi.org/10.1021/bi061128h
75. Boguth CA, Singh P, Huang CC, Tesmer JJG (2010) Molecular basis for activation of G protein-coupled receptor kinases. EMBO J 29(19):3249–3259. https://doi.org/10.1038/emboj.2010.206
76. Bayliss R, Sardon T, Vernos I, Conti E (2003) Structural basis of Aurora-A activation by TPX2 at the mitotic spindle. Mol Cell 12(4):851–862
77. Schulze JO, Saladino G, Busschots K, Neimanis S, Suss E, Odadzic D, Zeuzem S, Hindie V, Herbrand AK, Lisa MN, Alzari PM, Gervasio FL, Biondi RM (2016) Bidirectional allosteric communication between the ATP-binding site and the regulatory PIF pocket in PDK1 protein kinase. Cell Chem Biol 23(10):1193–1205. https://doi.org/10.1016/j.chembiol.2016.06.017
78. Stegert MR, Tamaskovic R, Bichsel SJ, Hergovich A, Hemmings BA (2004) Regulation of NDR2 protein kinase by multi-site phosphorylation and the S100B calcium-binding protein. J Biol Chem 279(22):23806–23812. https://doi.org/10.1074/jbc.M402472200
79. Ultanir SK, Hertz NT, Li G, Ge WP, Burlingame AL, Pleasure SJ, Shokat KM, Jan LY, Jan YN (2012) Chemical genetic identification of NDR1/2 kinase substrates AAK1 and Rabin8 Uncovers their roles in dendrite arborization and spine development. Neuron 73(6):1127–1142. https://doi.org/10.1016/j.neuron.2012.01.019
80. Cook D, Hoa LY, Gomez V, Gomez M, Hergovich A (2014) Constitutively active NDR1-PIF kinase functions independent of MST1 and hMOB1 signalling. Cell Signal 26(8):1657–1667. https://doi.org/10.1016/j.cellsig.2014.04.011
81. Lopez-Garcia LA, Schulze JO, Frohner W, Zhang H, Suss E, Weber N, Navratil J, Amon S, Hindie V, Zeuzem S, Jorgensen TJ, Alzari PM, Neimanis S, Engel M, Biondi RM (2011) Allosteric regulation of protein kinase PKCzeta by the N-terminal C1 domain and small compounds to the PIF-pocket. Chem Biol 18(11):1463–1473. https://doi.org/10.1016/j.chembiol.2011.08.010

82. Wu WI, Voegtli WC, Sturgis HL, Dizon FP, Vigers GP, Brandhuber BJ (2010) Crystal structure of human AKT1 with an allosteric inhibitor reveals a new mode of kinase inhibition. PLoS One 5(9):e12913. https://doi.org/10.1371/journal.pone.0012913
83. Hirai H, Sootome H, Nakatsuru Y, Miyama K, Taguchi S, Tsujioka K, Ueno Y, Hatch H, Majumder PK, Pan BS, Kotani H (2010) MK-2206, an allosteric Akt inhibitor, enhances antitumor efficacy by standard chemotherapeutic agents or molecular targeted drugs in vitro and in vivo. Mol Cancer Ther 9(7):1956–1967. https://doi.org/10.1158/1535-7163.MCT-09-1012
84. Politz O, Siegel F, Barfacker L, Bomer U, Hagebarth A, Scott WJ, Michels M, Ince S, Neuhaus R, Meyer K, Fernandez-Montalvan AE, Liu N, von Nussbaum F, Mumberg D, Ziegelbauer K (2017) BAY 1125976, a selective allosteric AKT1/2 inhibitor, exhibits high efficacy on AKT signaling-dependent tumor growth in mouse models. Int J Cancer 140(2):449–459. https://doi.org/10.1002/ijc.30457
85. Ebner M, Lucic I, Leonard TA, Yudushkin I (2017) PI(3,4,5)P3 engagement restricts Akt activity to cellular membranes. Mol Cell 65(3):416–431.e416. https://doi.org/10.1016/j.molcel.2016.12.028
86. Liu P, Wang Z, Wei W (2014) Phosphorylation of Akt at the C-terminal tail triggers Akt activation. Cell Cycle 13(14):2162–2164. https://doi.org/10.4161/cc.29584
87. Clement E, Inuzuka H, Nihira NT, Wei W, Toker A (2018) Skp2-dependent reactivation of AKT drives resistance to PI3K inhibitors. Sci Signal 11(521). https://doi.org/10.1126/scisignal.aao3810
88. Stroba A, Schaeffer F, Hindie V, Lopez-Garcia L, Adrian I, Frohner W, Hartmann RW, Biondi RM, Engel M (2009) 3,5-Diphenylpent-2-enoic acids as allosteric activators of the protein kinase PDK1: structure-activity relationships and thermodynamic characterization of binding as paradigms for PIF-binding pocket-targeting compounds. J Med Chem 52(15):4683–4693. https://doi.org/10.1021/jm9001499
89. Frohner W, Lopez-Garcia LA, Neimanis S, Weber N, Navratil J, Maurer F, Stroba A, Zhang H, Biondi RM, Engel M (2011) 4-benzimidazolyl-3-phenylbutanoic acids as novel Pif-Pocket-targeting allosteric inhibitors of protein kinase PKCzeta. J Med Chem 54(19):6714–6723. https://doi.org/10.1021/jm2005892
90. Wilhelm A, Lopez-Garcia LA, Busschots K, Frohner W, Maurer F, Boettcher S, Zhang H, Schulze JO, Biondi RM, Engel M (2012) 2-(3-Oxo-1,3-diphenylpropyl)malonic acids as potent allosteric ligands of the PIF pocket of phosphoinositide-dependent kinase-1: development and prodrug concept. J Med Chem 55(22):9817–9830. https://doi.org/10.1021/jm3010477
91. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314(1–2):141–151. https://doi.org/10.1016/S0009-2614(99)01123-9
92. Bonomi M, Parrinello M (2010) Enhanced Sampling in the well-tempered ensemble. Phys Rev Lett 104(19). https://doi.org/10.1103/Physrevlett.104.190601
93. Abdel-Halim M, Abadi AH, Engel M (2018) Design and synthesis of novel 1,3,5-triphenyl pyrazolines as potential anti-inflammatory agents through allosteric inhibition of protein kinase Czeta (PKCzeta). MedChemComm 9(6):1076–1082. https://doi.org/10.1039/c8md00100f
94. Abdel-Halim M, Diesel B, Kiemer AK, Abadi AH, Hartmann RW, Engel M (2014) Discovery and optimization of 1,3,5-trisubstituted pyrazolines as potent and highly selective allosteric inhibitors of protein kinase C-zeta. J Med Chem 57(15):6513–6530. https://doi.org/10.1021/jm500521n
95. Arencibia JM, Frohner W, Krupa M, Pastor-Flores D, Merker P, Oellerich T, Neimanis S, Schmithals C, Koberle V, Suss E, Zeuzem S, Stark H, Piiper A, Odadzic D, Schulze JO, Biondi RM (2017) An allosteric inhibitor scaffold targeting the PIF-pocket of atypical protein kinase C isoforms. ACS Chem Biol 12(2):564–573. https://doi.org/10.1021/acschembio.6b00827

96. Eyers PA, Erikson E, Chen LG, Maller JL (2003) A novel mechanism for activation of the protein kinase Aurora A. Curr Biol 13(8):691–697. https://doi.org/10.1016/S0960-9822(03)00166-0
97. Janecek M, Rossmann M, Sharma P, Emery A, Huggins DJ, Stockwell SR, Stokes JE, Tan YS, Almeida EG, Hardwick B, Narvaez AJ, Hyvonen M, Spring DR, McKenzie GJ, Venkitaraman AR (2016) Allosteric modulation of AURKA kinase activity by a small-molecule inhibitor of its protein-protein interaction with TPX2. Sci Rep 6:28528. https://doi.org/10.1038/srep28528
98. Asteriti IA, Daidone F, Colotti G, Rinaldo S, Lavia P, Guarguaglini G, Paiardini A (2017) Identification of small molecule inhibitors of the Aurora-A/TPX2 complex. Oncotarget 8(19):32117–32133. https://doi.org/10.18632/oncotarget.16738
99. Bayliss R, Burgess SG, McIntyre PJ (2017) Switching Aurora-A kinase on and off at an allosteric site. FEBS J 284(18):2947–2954. https://doi.org/10.1111/febs.14069
100. Panicker RC, Coyne AG, Srinivasan R (2017) Allosteric targeting of Aurora A kinase using small molecules: a step forward towards next generation medicines? Curr Med Chem. https://doi.org/10.2174/0929867324666170727120315
101. Carlino L, Christodoulou MS, Restelli V, Caporuscio F, Foschi F, Semrau MS, Costanzi E, Tinivella A, Pinzi L, Lo Presti L, Battistutta R, Storici P, Broggini M, Passarella D, Rastelli G (2018) Structure-activity relationships of hexahydrocyclopenta[c]quinoline derivatives as allosteric inhibitors of CDK2 and EGFR. ChemMedChem 13(24):2627–2634. https://doi.org/10.1002/cmdc.201800687
102. Komander D, Kular GS, Bain J, Elliott M, Alessi DR, Van Aalten DM (2003) Structural basis for UCN-01 (7-hydroxystaurosporine) specificity and PDK1 (3-phosphoinositide-dependent protein kinase-1) inhibition. Biochem J 375(Pt 2):255–262
103. Axten JM, Blackledge CW, Brady GP, Feng Y, Grant SW, Medina JR, Millller WH, Romeril SP (2010) Preparation of 6-(4-pyrimidinyl)-1H-indazole derivatives as PDK1 inhibitors. PCT Int Appl WO 2010059658
104. Najafov A, Sommer EM, Axten JM, Deyoung MP, Alessi DR (2011) Characterization of GSK2334470, a novel and highly specific inhibitor of PDK1. Biochem J 433(2):357–369. https://doi.org/10.1042/BJ20101732
105. Richards MW, Burgess SG, Poon E, Carstensen A, Eilers M, Chesler L, Bayliss R (2016) Structural basis of N-Myc binding by Aurora-A and its destabilization by kinase inhibitors. Proc Natl Acad Sci U S A 113(48):13726–13731. https://doi.org/10.1073/pnas.1610626113
106. Gustafson WC, Meyerowitz JG, Nekritz EA, Chen J, Benes C, Charron E, Simonds EF, Seeger R, Matthay KK, Hertz NT, Eilers M, Shokat KM, Weiss WA (2014) Drugging MYCN through an allosteric transition in Aurora kinase A. Cancer Cell 26(3):414–427. https://doi.org/10.1016/j.ccr.2014.07.015
107. Poulikakos PI, Zhang C, Bollag G, Shokat KM, Rosen N (2010) RAF inhibitors transactivate RAF dimers and ERK signalling in cells with wild-type BRAF. Nature 464(7287):427–430. https://doi.org/10.1038/nature08902
108. Hatzivassiliou G, Song K, Yen I, Brandhuber BJ, Anderson DJ, Alvarado R, Ludlam MJ, Stokoe D, Gloor SL, Vigers G, Morales T, Aliagas I, Liu B, Sideris S, Hoeflich KP, Jaiswal BS, Seshagiri S, Koeppen H, Belvin M, Friedman LS et al (2010) RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth. Nature 464(7287):431–435. https://doi.org/10.1038/nature08833
109. Wang L, Perera BG, Hari SB, Bhhatarai B, Backes BJ, Seeliger MA, Schurer SC, Oakes SA, Papa FR, Maly DJ (2012) Divergent allosteric control of the IRE1alpha endoribonuclease using kinase inhibitors. Nat Chem Biol 8(12):982–989. https://doi.org/10.1038/nchembio.1094
110. Lombard CK, Davis AL, Inukai T, Maly DJ (2018) Allosteric Modulation of JNK Docking Site Interactions with ATP-Competitive Inhibitors. Biochemistry 57(40):5897–5909. https://doi.org/10.1021/acs.biochem.8b00776

Chapter 13
Allosteric Modulators of Protein–Protein Interactions (PPIs)

Duan Ni, Na Liu, and Chunquan Sheng

Abstract Protein–protein interactions (PPIs) represent promising drug targets of broad-spectrum therapeutic interests due to their critical implications in both health and disease circumstances. Hence, they are widely accepted as the Holy Grail of drug development. Historically, PPIs were rendered "undruggable" for their large, flat, and pocket-less structures. Current attempts to drug these "intractable" targets include orthosteric and allosteric methodologies. Previous efforts employing orthosteric approaches like protein therapeutics and orthosteric small molecules frequently suffered from poor performance caused by the difficulties in directly targeting PPI interfaces. As structural biology progresses rapidly, allosteric modulators, which direct to the allosteric regulatory sites remote to the PPI surfaces, have gradually established as a potential solution. Allosteric pockets are topologically distal from the PPI orthosteric sites, and their ligands do not need to compete with the PPI partners, which helps to improve the physiochemical and pharmacological properties of allosteric PPI modulators. Thus, exploiting allostery to tailor PPIs is regarded as a tempting strategy in future PPI drug discovery. Here, we provide a comprehensive review of our representative achievements along the way we utilize allosteric effects to tame the difficult PPI systems into druggable targets. Importantly, we provide an in-depth mechanistic analysis of this success, which will be instructive to future related lead optimizations and drug design. Finally, we discuss the current challenges in allosteric PPI drug discovery. Their solutions as well as future perspectives are also presented.

Keywords Protein–protein interaction · Allostery · Allosteric modulator · Drug design

D. Ni
Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis of Chinese Ministry of Education, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

N. Liu · C. Sheng (✉)
School of Pharmacy, Second Military Medical University, Shanghai, China
e-mail: shengcq@smmu.edu.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_13

13.1 Introduction

A number of important biological processes, such as cell division, adaption, and extracellular signal response, are modulated by protein-protein interactions (PPIs) [5, 10, 83, 122]. The complex network of human PPIs (interactome) is estimated to comprise about 130,000 to 650,000 types of PPI [98, 110, 111, 116]. Due to the pivotal role of PPIs both in normal and disease states, PPIs represent a rich source of drug targets for the development of new generation of therapeutics [1, 5, 62, 83, 104, 113, 122, 128]. PPI complexes can be modulated by peptides or small molecules through orthosteric binding (inhibition or stabilization) and allosteric binding (inhibition or stabilization). An orthosteric PPI modulator binds to the PPI interaction surface, thus sterically preventing or stabilizing the binding of one protein to its partner protein. Orthosteric PPI inhibition represents a major strategy in current PPI-based drug discovery because most PPI modulators marketed or in clinical development are small molecule inhibitors [50, 63, 79, 83, 117, 129]. However, the discovery of drug-like small-molecule PPI inhibitors is still considered to be highly challenging [39, 104, 123]. As compared to traditional drug targets (e.g., enzymes, G-protein-coupled receptors) with well-defined pockets for binding small molecules, the PPI interface is generally large (about 1500–6000 $Å^2$) [6, 56, 109], flat [10], and dominated with hydrophobic or charged residues [6, 52, 56, 64]. Thus, these features of PPI interfaces make it difficult to design selective and potent small-molecule inhibitors [16]. Stabilization of PPI complexes is an alternative way to interfere with the PPI-associated biological functions [113, 125]. Instead of competing with any of the protein partners, PPI stabilizers target the regions at or near the PPI interfaces. Thus, PPI stabilizers are able to exert desired biological response with relatively low binding affinity (micromolar range) and are more likely to achieve specificity [74, 113, 125]. Recently, there is an increase of research interests in investigating the mode of action, molecular design, and therapeutic effects of PPI stabilizers (e.g., stabilizers of 14-3-3 PPI) [74, 113, 125].

In contrast to orthosteric modulators, allosteric PPI modulators bind at a remote site outside the PPI interface and interfere with the protein binding by triggering conformational change [14, 23, 41, 83, 104, 106]. Due to the unique binding mechanism that targeting regions spatially distinct from intractable PPI interfaces, allosteric modulators have several advantages over orthosteric PPI ligands and could partly address the difficulties in designing orthosteric modulators. First, PPI targets that are fundamentally intractable to orthosteric small molecule binders can be modulated by allosteric binding toward remote pockets. The structures of allosteric sites are more suitable to bind drug-like small molecules relative to large and flat PPI interfaces. Also, allosteric ligands do not need to compete with the bulky, and relatively high-affinity PPI partners, which is the major obstacle for orthosteric binding. Hence, allosteric modulators generally have lower molecular weight and less hydrophobicity than orthosteric PPI ligands, and thus are considered to be more drug-like [61, 115, 119]. Second, allosteric mechanism of modulation allows for the possibility to achieve better functional control of a drug target in which binding and

functional modulations can be separated [14, 41, 66, 85]. Currently, various allosteric PPI modulators were identified, and particularly modulators for tubulin dimerization [60, 76] have been intensively studied and even reached clinic.

Although allosteric regulation has shown promising future in PPI-based drug discovery, it is still highly challenging to discover small molecule allosteric PPI modulators with desired functions. This is probably due to limited knowledge about the allosteric mechanisms of PPI [4]. Instead of rational design, most allosteric PPI modulators were discovered serendipitously, but still, rapid progress has been made in theories and drug design approaches for allosteric PPI modulators [14, 65, 66, 84, 118]. The improvement of chemical and structural biology techniques facilitated better understanding of allosteric binding mechanisms and rational drug design.

In this chapter, we aim to provide an overview of allosteric PPI modulators. Design strategies and successful examples for the discovery of allosteric modulators will be highlighted. Current challenges and future perspectives will also be provided to inspire future research of allosteric PPI-based drug discovery.

13.2 CDK2–Cyclin A3

Cell cycle or cell division is an important cellular event strictly controlled throughout the process of development and growth. In addition to its critical role in normal physiological conditions, malfunctions of division frequently lead to the pathogenesis of series of diseases such as cancers [69, 70, 77, 80, 96, 105]. The accurate process of cell cycle is exactly regulated by a family of serine/threonine kinases, the cyclin-dependent protein kinases (CDKs), and their PPI partner cyclins [9, 68]. During cell division, cyclins form heterodimers with CDKs [68, 96], which further phosphorylate downstream targets such as Whi5 and Ndd1 [24, 27, 28, 94] and drive cell cycle in a timely manner. Malfunction of the CDK–cyclin PPIs results in the loss of correct timing and order of cell division processes, leading to pathophysiological conditions such as cancers [29, 43, 77, 80, 105]. Hence, the interactions between CDKs and cyclins have been intensively investigated as attractive antitumor targets and the CDK2–cyclin A3 PPI is among one of them [31, 33, 36, 68]. Previous small peptides or peptidomimetic inhibitors for this PPI system are frequently hampered by poor metabolic stability, low oral bioavailability, and easy degradation. Therefore, allosteric modulators supply a new opportunity for targeting CDK2–cyclin A3 PPI with improved performances and represent a novel direction for cancer therapeutics development [31, 33, 36, 69].

Recently, a novel allosteric pocket on CDK2 was uncovered [55] and it was quickly translated into the development of a set of allosteric inhibitors for the CDK2–cyclin A3 PPI [46] (Fig. 13.1). In the research carried out by Hu et al., virtual screening is combined with bioassays to yield the promising lead compounds. SPECS chemical library (http://www.specs.net) and Chemdiv chemical library (http://www.chemdiv.com) were employed for in silico screening using Grid-based ligand docking from GLIDE software [38] (Schrödinger, Portland, OR) on

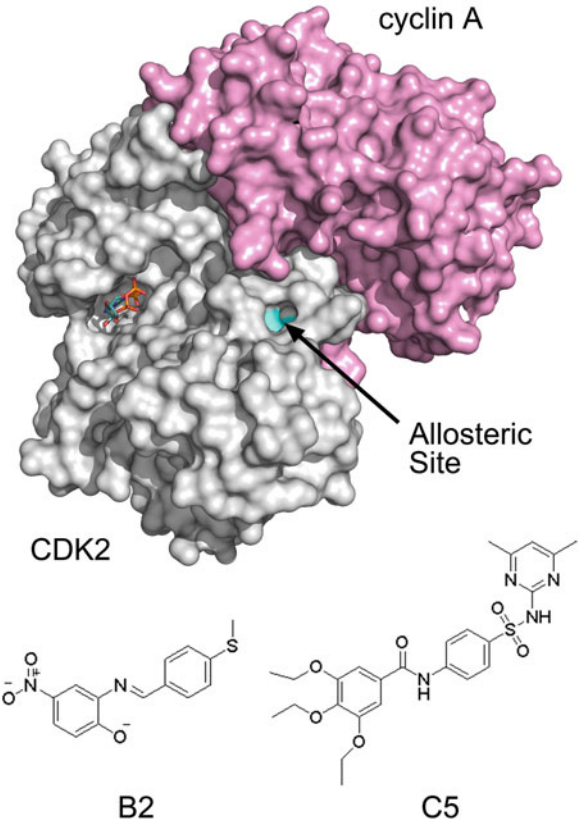


Fig. 13.1 Structural overview of the CDK2–cyclin A PPI complex system with the allosteric site highlighted (PDB ID, 2CCI; CDK2, grey; cyclin A, pink; ATP, blue; and allosteric site, cyan) and chemical structure of two allosteric PPI inhibitors for CDK2–cyclin A system, B2 and C5

the co-crystal structure of CDK2 and cyclin A [17] (PDB ID: 2CCI). Two rounds of screening and docking with standard-precision (SP) and extra-precision (XP) docking were carried out and the screening hits and their potential binding poses were retrieved. Final choices for the possible candidates were based on the Glide scoring function (G-Score), docking analysis, and manual selection, during which docking free energy, binding pose, and molecular interactions were major criteria for decisions. Fifteen compounds were obtained from virtual screening, and they were tested using a commercially available luminescent-based approach (ADP-Glo™ Max Assay, Promega). Since the phosphorylation activity of CDK2 depends on its PPI with cyclin A3, the ability of CDK2 in the complex system to phosphorylate synthetic peptide substrate was applied to the measurement of the capacity of the compound candidates to disrupt the CDK2–cyclin A3 interaction. Two compounds, B2 and C5, were found to be effective PPI inhibitors with IC_{50} of 52.12 ± 1.350 μM and 105.1 ± 1.428 μM (Fig. 13.1). Their inhibitory effects were also confirmed through GST pull-down assays. Further mutagenesis experiments verified the allosteric binding of the compounds into a pocket distinct from the PPI interface, composed of R126, R150, and Y180 (Fig. 13.1).

The lead compound, B2, was tested in human cancer cell lines A549 (non-small cell lung carcinoma), HepG2 (liver hepatocellular carcinoma), and MDA-MB-231 (breast carcinoma) by MTS assays. It showed that B2 exhibits a concentration-dependent anti-proliferative activity. Flow cytometry showed that B2 arrested cell cycle at S phage and therefore prevents cell growth. Thus, as allosteric PPI inhibitors for CDK2–cyclin A3 system, B2 and its derivatives shed light on drug discovery relevant to cell cycle and may serve as the starting point for future optimization and development of related therapeutic agents.

13.3 CD40 – CD40L

PPI between the TNF family cytokine CD40 ligand (CD40L, aka CD154) and CD40 has long been regarded as a compelling target for drug development due to its critical role in immune system and involvement in diseases such as atherothrombosis and cancers [3, 32, 93, 126]. Previously, a number of antibodies had been developed to block the CD40 – CD40L interaction, and some of them like hu5c8 had even advanced into clinical trials for autoimmune diseases such as alloislet graft rejection [18, 107] and lupus nephritis [26]. However, most of these protein therapeutics for the CD40 – CD40L system failed due to adverse events such as thromboembolic complications [11], which mainly resulted from the cross-linking or effector functions of the antibodies [34]. Hence, small molecule inhibitors may be a new direction for drug design targeting the PPI between CD40 and CD40L.

One of the recent advances in the development of CD40 – CD40L inhibitors is BIO8898 (Fig. 13.2a), reported by Biogen Idec. Based on compound library screening, BIO8898 was identified as a CD40L binder, and it was further characterized as an inhibitor for the interaction between CD40L and CD40 by Silvian et al. [108]. With a dissociation-enhanced lanthanide fluorescence immunoassay (DELFIA) [54], researchers first demonstrated that BIO8898 could inhibit CD40L binding toward CD40 with an IC_{50} of ~25 μM. The inhibitory effect of BIO8898 was also tested in cellular context utilizing a baby hamster kidney (BHK) cell line with a chimeric receptor fusing CD40 to TNFRp75, which could trigger apoptosis upon CD40 – CD40L interaction [91, 103]. The PPI-induced cell death in BHK cell could be inhibited by BIO8898, thereby confirming its PPI disruption activity in vivo.

To investigate the binding mode and structural mechanisms underlying the action of BIO8898, the crystal structure of CD40L in complex with the compound was solved. *Apo* CD40L was constitutively homotrimeric comprising of three identical subunits Chain A, B, and C. Surprisingly, BIO8898 was found to intercalate between subunits, distal to the CD40 – CD40L interface, therefore indicating its allosteric regulatory role on the PPI system (Fig. 13.2b). This compound positioned into an induced cleft made up of L168, Y170, V228, E230, L259, and L261 from Chain A; Y170, Y172, A173, Q174, H224, and L225 from Chain B; and A123, H125, Y145, Y146, Y170, Y172, H224, S255, L259, and L261 from Chain C. The binding pocket is 81% hydrophobic, 11% neutral, and 6% polar. The bipyridyl core of BIO8898

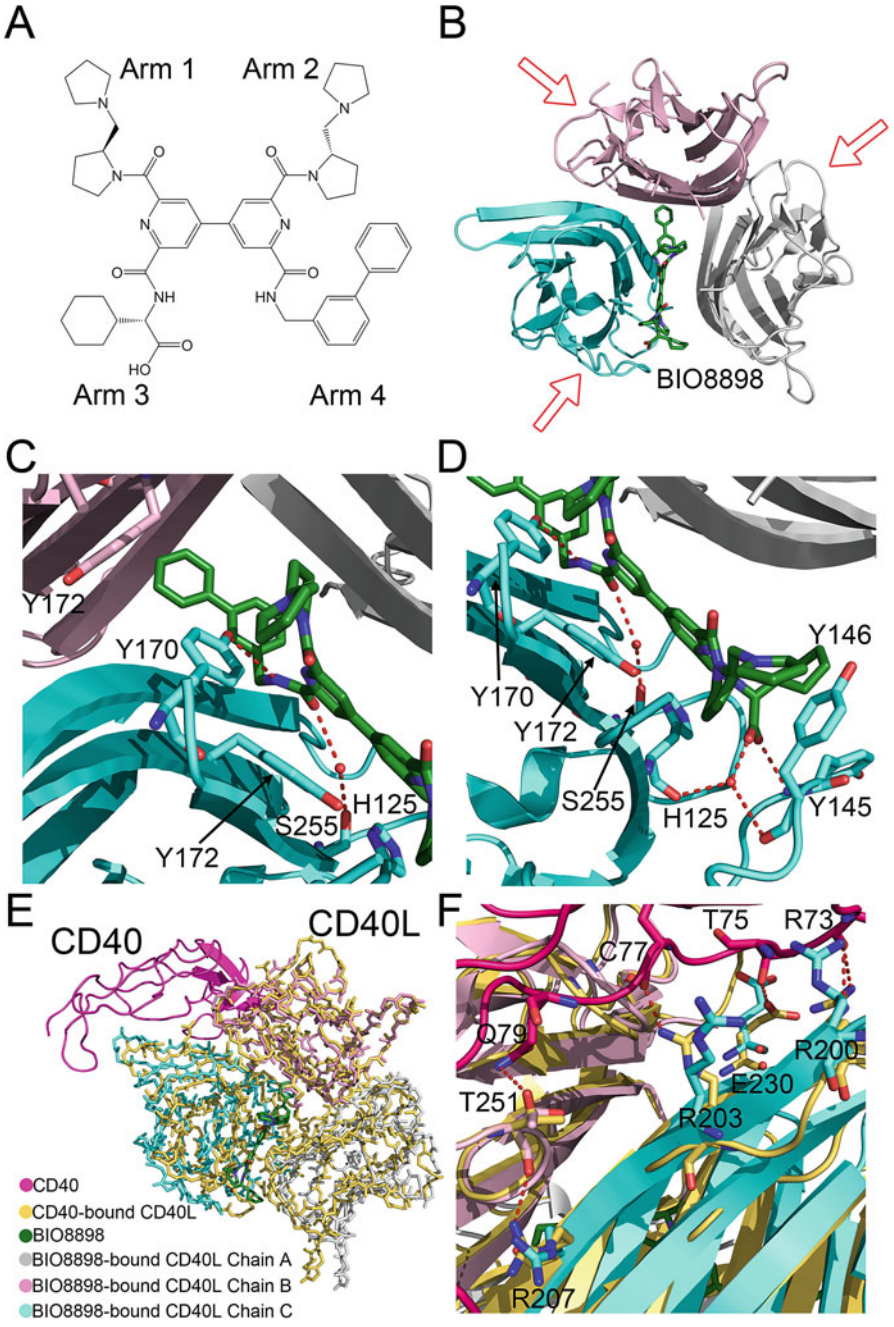


Fig. 13.2 (**a**) Chemical structure of the allosteric CD40 − CD40L PPI inhibitor BIO8898. (**b**) BIO8898 intercalates between 3 subunits of CD40L (PDB ID: 3LKJ, 3 subunits are colored differently, and the PPI binding sites for CD40 are labeled with red arrows). (**c** and **d**) Interaction details between BIO8898 and CD40L. (**e**) BIO8898 intercalates into the CD40L trimer, disrupting the trimeric symmetry, affecting the CD40 binding surface, and therefore allosterically inhibit the CD40 − CD40L PPI. (**f**) Contacts between CD40 and CD40L are disrupted upon BIO8898 binding and thus result in the allosteric PPI inhibition (PDB ID: BIO8898-bound CD40L: 3LKJ, CD40-bound CD40L: 3QD6, color representations and labeling are shown in Panel **E**)

made π-stack with Y172 in Chain C, and the biphenyl moiety of Arm 4 in the inhibitor formed a perpendicular π–π interaction with Y172 from Chain B. Several hydrogen bonds were found between the compound and Chain C. The carboxylate group in BIO8898 was hydrogen bonded to Y145 and made water-mediated hydrogen bonds with H125 and Y146. The amide NH of the biphenyl arm in the ligand made a hydrogen bond with Y170 and the amide carbonyl in this arm formed a water-mediated hydrogen bond with S255 (Fig. 13.2c and d). These inter-molecular interactions stabilized the ligand binding, and affected the overall conformations of the trimeric CD40L as well as the CD40 − CD40L interface.

Structural superposition revealed that BIO8898 would not cause dissociation of the CD40L homotrimer, but rather, it would interrupt the original symmetry of the protein trimer, and allosterically distort the binding site for CD40 (Fig. 13.2e). The CD40 − CD40L interface became relatively disordered upon inhibitor binding, especially the elastic β sheets involved. Also, the conformations of the interfacial hot spot residues such as R200 and R203 in Chain C from CD40L were significantly altered, which thereby interfered the CD40 − CD40L contacts, such as hydrogen bonds (Fig. 13.2f). Hence, BIO8898 represented a novel mode for the inhibition of the critical CD40 − CD40L PPI target through allostery and supplied a starting point for future PPI drug development related to autoimmune diseases and cancer immune therapy [18, 107] with great prospects.

13.4 α-Tubulin−β-Tubulin

In addition to the majority of allosteric PPI inhibitors, allosteric modulators to stabilize PPIs also possess great potential for drug discovery, since in many biological contexts, both physiological and pathological, prolonged lifespans of PPIs will be more desirable [113, 125]. Previous research on this topic reveals that nature has already offered fruitful impressive examples of allosteric PPI stabilizers. The most representative one of them is Paclitaxel (Taxol®) from *Taxus brevifolia* stabilizing the inter-tubulin PPI, which has achieved great success as an anti-cancer agent [25, 53, 88] (Fig. 13.3a).

Cytoskeleton is a major cellular mechanical structure, and microtubules (MTs) are its key components. MTs are built up of α- and β-tubulin subunits [2, 12, 113], and the PPIs between them are involved in the growth and shrinkage of MTs and the cytoskeleton dynamics. Hence, inter-tubulin PPIs are central to various cytoskeleton-mediated cellular activities like cell cycle and subcellular transport [90] and malfunctions of them frequently lead to oncogenesis, during which cells feature over-activated self-replication and proliferation [19, 73, 86]. Microtubule-stabilizing agents (MSAs), which stabilize these PPIs and thereby halt abnormal cell cycle and induce apoptosis [30, 86, 100], have been established as a promising direction for anti-mitotic agents development as cancer therapies [53], and Paclitaxel (Taxol®) is among the best example.

Paclitaxel (Taxol®) is extracted from the bark of the tree *Taxus brevifolia*. It stabilizes the inter-tubulins PPIs both in vitro and in vivo [100, 101] and displays

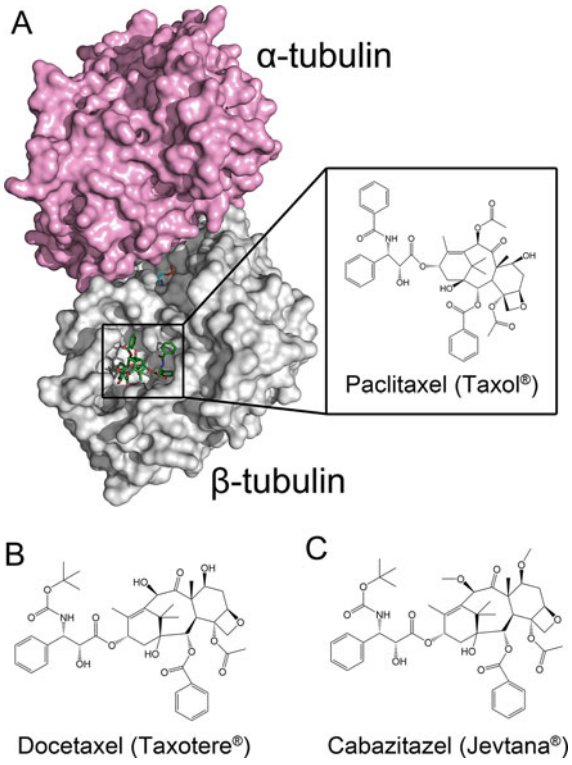


Fig. 13.3 (**a**) Structural overview of the α- and β-tubulins with the allosteric PPI stabilizer Paclitaxel (Taxol®) (PDB ID: 3J6G). (**b**) Chemical structure of allosteric PPI stabilizer docetaxel (Taxotere®). (**c**) Chemical structure of allosteric PPI stabilizer cabazitazel (Jevtana®)

antitumor capacity in several experimental systems [120]. In 1992, Paclitaxel (Taxol®) was clinically approved as the first MSA chemotherapeutic drug and administered to treat breast cancer, non-small-cell lung cancer, and ovarian cancer [71].

Despite the great success of Paclitaxel (Taxol®), its underlying mechanism has long been under debate. Only until very recently, with the help of cryo-EM, researchers revealed the binding of Paclitaxel (Taxol®) in a pocket distal to the tubulins PPI interface and therefore unraveled its allosteric PPI stabilization effect [2] (Fig. 13.3a). Binding of the compound restores the microtubule lattice and remodels the interfaces between α- and β-tubulins, and therefore, the tubulin PPIs are stabilized allosterically [2]. More in-depth mechanistic details for Paclitaxel's action remain elusive due to the vulnerabilities of microtubule and tubulin structures during crystallography study, but Paclitaxel (Taxol®) still supplies great opportunities for related allosteric PPI drug development. Based on its ligand-binding pocket, more allosteric inter-tubulin PPI stabilizers have been reported, especially the derivatives of Paclitaxel (Taxol®) itself. Some of these compounds have also reached clinic [12, 30, 35, 40]. For example, docetaxel (Taxotere®) (Fig. 13.3b) was approved in 2006 and is now used for treatment of prostate, head, and neck cancers [30]; and cabazitazel (Jevtana®) (Fig. 13.3c) was approved in 2010 and is effective

for treating metastatic hormone-refractory prostate cancer [35, 40]. As allosteric PPI stabilizers, Paclitaxel series have achieved great success in both pharmaceutical research and clinical application; and they still represent a promising direction for future rational design of MSAs and anti-mitotic cancer drugs.

13.5 Cdc34A–Ubiquitin

Another representative example of a PPI amenable to allosteric stabilization is that occurring between Cdc34A, an E2 ubiquitin-conjugating enzyme, and ubiquitin. Ubiquitination is one of the most important protein post-translational modifications [57, 124], and its functions are closely related to the ubiquitin-proteasome system (UPS), which comprises a conserved cycle of E1 activating enzymes, E2 conjugating enzymes, and E3 ubiquitin ligases [57, 82]. Perturbation of UPS is commonly associated with neurological and immunological disorders [7, 20, 82] as well as cancers. Thus, regulation of the UPS enzymatic cascade has attracted intense interest for drug discovery [20, 82].

One of the latest advances in UPS drug discovery is a first-in-class Cdc34A small molecule inhibitor, CC0651 (Fig. 13.4a), that allosterically stabilizes the Cdc34A – ubiquitin PPI [13], thereby interfering the normal functions of Cdc34A. In search of small molecule inhibitors for ubiquitination catalyzed by the multisubunit Skp1-Cdc53/Cullin-F box protein (SCF) ubiquitin ligases, Ceccarelli et al. performed a high-throughput-compatible assay based on the ubiquitination reaction of human CDK inhibitor $p27^{Kip1}$ by the SCF^{Skp2} E3 complex [13, 15]. CC0651 was identified as a potent hit (IC_{50} = 1.72 μM), and its target was revealed to be Cdc34A after numerous tests with different enzymes involved in that process [13]. Further biochemical analysis unveiled that the PPI between Cdc34A and ubiquitin was enhanced upon CC0651 treatment and the time-resolved Förster resonance energy transfer (TR-FRET) assay also confirmed that CC0651 could potentiate the binding of ubiquitin to Cdc34A with a half-maximum effective concentration (EC_{50}) of 14 ± 2 μM. Hence, CC0651 was proposed to inhibit Cdc34A through stabilizing the Cdc34A – ubiquitin PPI and trapping this E2 enzyme in an intermediate state along the ubiquitination process.

To investigate into the underlying mechanism, crystal structures of *apo* and *holo* forms of Cdc34A were solved. CC0651 was found to sit in a partially induced pocket distinct from the ubiquitin binding surface, which indicated its allosteric mechanism for the PPI stabilization. This cavity was formed by the α1-β1 linker, C-terminal end of the α2 helix, β2-β3 linker, C-terminal end of the α3 helix, and α3-α4 linker (Fig. 13.4b). Comparison of the *apo* and *holo* structures showed that the most significant conformational changes induced by CC0651 were located at the α2 helix (Fig. 13.4c), which in turn leads to conformational changes on the Cdc34A – ubiquitin interface. Researchers inferred that the interfacial structural alterations might impair the reactivity of the labile thioester intermediate between Cdc34A – ubiquitin and further strengthen the Cdc34A – ubiquitin linkage, thereby

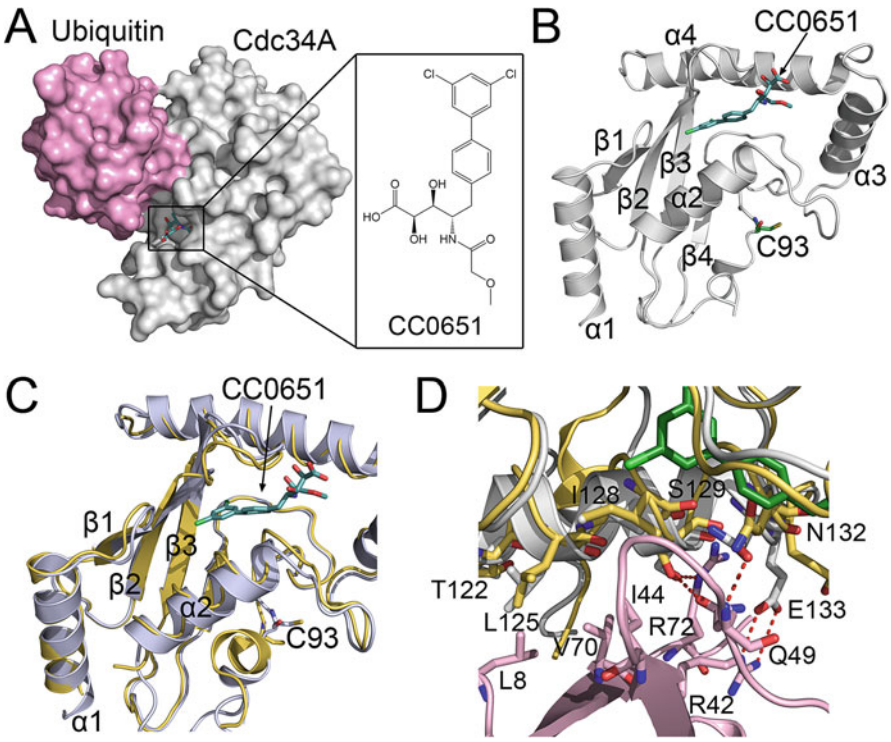


Fig. 13.4 (**a**) Structural overview of the Cdc34A – ubiquitin PPI system in complex with the allosteric PPI stabilizer CC0651 (PDB ID: 4MDK). (**b**) Structural details into the binding pocket of CC0651. (**c**) Conformational changes induced by CC0651 binding. (**d**) CC0651 binding can enhance interactions between Cdc34A and ubiquitin and thereby stabilize the PPI

arresting this PPI. Also, NMR analysis uncovered that more pronounced interactions were induced by CC0651 between Cdc34A and ubiquitin and the low-affinity interactions between them were stabilized by the compound binding. In the liganded form, notable hydrogen bonds and electrostatic interactions were found between S129, N132, and E133 in Cdc34A and R42, Q49, and R72 in ubiquitin. T122, L125, S126, I128, and S129 in Cdc34A also formed hydrophobic contacts with L8, I44, and V70 in ubiquitin (Fig. 13.4d). Hence, through strengthening both the covalent and non-covalent interactions between Cdc34A and ubiquitin, CC0651 allosterically stabilize this PPI system.

Further cellular experiments unveiled that CC0651 and its analogs can inhibit the proliferation of cancer cell lines. They increased the intracellular level of the ubiquitination substrate of Cdc34A and prevented it from proteasomal degradation [13]. Given its central role in UPS, the E2 ubiquitin-conjugating enzyme Cdc34A was established as a novel viable therapeutic target through the success of CC0651, which provides a feasible strategy for drugging the critical UPS in cell.

13.6 AmtB–GlnK

Most allosteric PPI modulators exert a sole effect, for example, stabilizing or inhibiting the PPI activity. However, recent studies have revealed that under specific circumstances, endogenous allosteric molecules such as lipids can confer different effects by fine-tuning PPIs in different ways [21, 59, 97]. Allosteric PPI modulators with multiple functions have been exemplified by the various regulatory roles of lipid molecules with different modifications [21, 59]. Recently, Laganowsky et al. determined the effects of individual binding events for a series of lipid molecules on the PPI between two membrane proteins, ammonium channel, AmtB, and GlnK [21, 59]. To explore the allosteric effects of lipid molecules on the AmtB−GlnK PPI, the protein complex system was titrated with lipid molecules with different headgroups, tails or stereochemistries. The results showed that the addition of phosphatidylglycerol (PG) results in dramatic enhancement of the AmtB−GlnK PPI via an allosteric mechanism (Fig. 13.5). Then, the effects of different headgroups of lipid molecules on the AmtB−GlnK PPI were analyzed. Phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), PG, phosphatidylserine (PS) containing 1-palmitoyl-2-oleoyl (PO, 16:0–18:1) tails and cardiolipin (TOCDL, 1,1′2,2′-tetraoleyl-cardiolipin) were chosen as representatives of lipid molecules. Analysis of the coupling factors suggested that individual binding of POPE and POPG exerts an activation effect of stabilizing PPI, while adding POPS and TOCDL alone has no effect on the AmtB−GlnK interaction. Intriguingly, adding POPA resulted in a slight inhibitory effect. Investigation of the effects of the lipid tails was further carried out using PG harboring tails with different lengths. However, upon modification with different types of tails, the function of PG was subtly altered. Tails

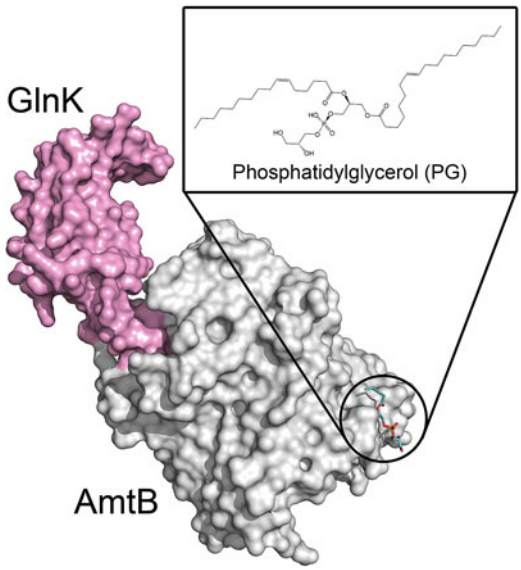


Fig. 13.5 Structural overview of AmtB−GlnK PPI system in complex with allosteric PPI modulator phosphatidylglycerol (PG) (PDB ID: GlnK: 2NS1, AmtB: 4NH2, and PG: 4NH2)

with lengths of 12 and 16 carbons were found to exhibit activation allosteric effects, whereas a 14-carbon tail resulted in neither stabilizing nor disrupting performance for PG. In addition to the modification of different headgroups and tails, the impact of stereochemistry was also studied, and PE lipids with dioleoyl (DO, 18:1) tails in both *cis* and *trans* configurations and with 1-stearoyl-2-oleoyl (SO, 18:0 saturated, 18:1 *cis*-monounsaturated) tails were investigated. The results showed that both the *cis* and *trans* configurations of DOPE display activation effects, but that of the *cis* form is less pronounced. As for SOPE, it conveys an effect similar to that of DOPE.

Taken together, this is the first report that individual lipid-binding events can exert allosteric modulation on PPI and it also revealed that even though with similar parent cores, the regulatory events of allosteric compounds on PPIs may have large dependence on their structural modifications and stereochemistry.

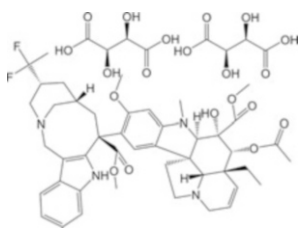
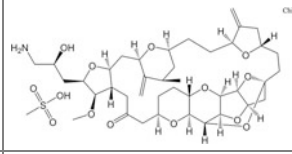
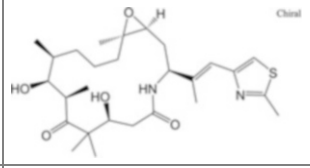
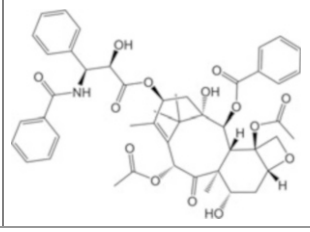
13.7 Conclusions

Underlying various kinds of cellular processes as well as both normal and medical conditions, PPIs are regarded as the Holy Grail of therapeutic intervention and drug development [4, 51, 83, 104, 122]. They provide a plethora of unique structural landscapes as promising drug targets [22, 39, 62, 83, 104, 113, 122, 128], and have gathered pace in pharmaceutical field with the rapid advances in chemical biology and structural biology.

Traditional PPI drug design mainly focuses on direct binding toward PPI interfaces [51, 83, 106, 122]. However, as a result of the relatively large, flat, and featureless interaction surfaces, limited success has been achieved using orthosteric methodology. Also, orthosteric approaches targeting the PPI interfaces such as antibodies frequently suffer from problems such as poor bioavailability [83, 104, 109, 122]. Hence, as one of the cutting-edge breakthrough in structural biology and drug discovery, harnessing allostery to tailor PPIs has emerged as a promising alternative and complement to orthosteric modulators [4, 37, 41, 66, 84, 109].

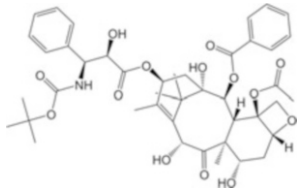
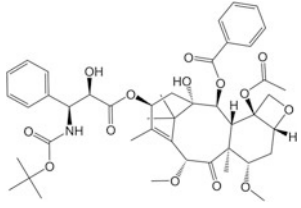
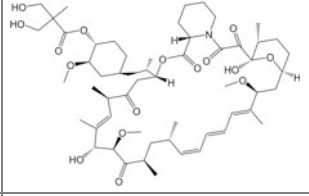
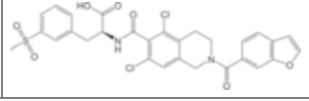
Compared with the difficult topology and conformations of the orthosteric PPI interfaces, allosteric pockets are more suitable for drug design and the following optimization. Additionally, targeting the allosteric sites spatially and topologically remote from the orthosteric interfaces, allosteric ligands do not need to compete with the bulky PPI partners, which can contribute to their improved physiochemical properties and pharmacological performances [14, 84, 115, 119]. In the recent decades, even though we have already witnessed an upsurge in the success of allosteric PPI modulators in drug development, suggesting the promising future in this aspect (Table 13.1), there still exist series of problems and challenges. Obstacles like suboptimal pharmacokinetic properties frequently emerge, and our understanding toward PPI allostery is still relatively limited. In the field of PPI drug discovery, most of the allosteric success so far are achieved completely by serendipity, which largely hampers the further optimizations and structure-based rational design. However, we still believe that more momentum will be fueled into this field with the

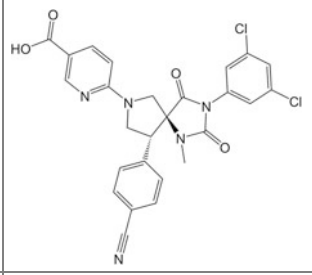
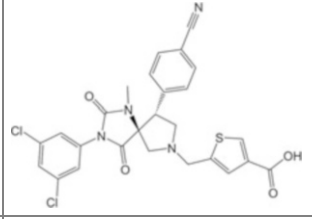
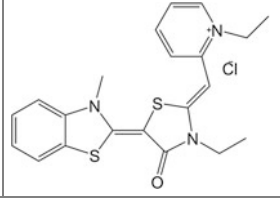
Table 13.1 Recent examples of allosteric PPI modulators in different clinical phages

Generic name	Chemical structure	Target PPI	Regulatory effect	Clinical phage	Refs
Vinflunine ditartrate or PM391		α-Tubulin −β-tubulin	Inhibition	In the clinic	Vu et al. [117]
Eribulin mesylate or E7389		α-Tubulin −β-tubulin	Inhibition	In the clinic	Towle et al. [114] and Morris [78]
Ixabepilone		α-Tubulin −β-tubulin	Stabilization	In the clinic	Osborne et al. [87] and Lee et al. [60]
Paclitaxel or Taxol®		α-Tubulin −β-tubulin	Stabilization	In the clinic	Richardson et al. [95] and Saloustros et al. [99]

(continued)

Table 13.1 (continued)

Generic name	Chemical structure	Target PPI	Regulatory effect	Clinical phage	Refs
Docetaxel or Taxotere®		α-Tubulin −β-tubulin	Stabilization	In the clinic	James et al. [49] and Heery et al. [44]
Cabazitaxel or Jevtana®		α-Tubulin −β-tubulin	Stabilization	In the clinic	Mita et al. [75] and Nabholtz and Gligorov [81]
Temsirolimus or CCI-779		mTOR−FKBP12	Inhibition	In the clinic	Schulze et al. [102] and Bellmunt et al. [8]
Lifitegrast or SAR1118		LFA-1 − ICAM-1	Inhibition	In the clinic	Zhong et al. [127] and Holland et al. [45]

BMS-688521		LFA-1 – ICAM-1	Inhibition	Preclinic	Watterson et al. [121]
BMS-587101		LFA-1 – ICAM-1	Inhibition	Preclinic	Suchard et al. [112]
MKT-077		HSP70-BAG3	Inhibition	Preclinical but abandoned	Propper et al. [92] and Koya et al. [58]

advances in bioinformatics allosteric site prediction tools like AlloFinder [47, 48], AlloSigMa [42], and PARS [89] and breakthroughs in structural biology such as cryo-EM method for atomistic resolution of the previously inaccessible PPI systems [2, 72]. With these state-of-the-art technologies, allosteric strategy for PPI modulation undoubtedly has great potential to enter into the mainstream of pharmaceutical industry and related research in the near future.

Previously, allosteric effect has been well-known for taming the historically challenging targets such as GPCRs and protein kinases [66, 67, 84, 85]. Given their prominent advantageous characteristics in PPI modulations, we believe allosteric modulators can still follow their path of success and tackle the obstacles posed by the diversified and fascinating PPI targets and supply a novel methodology for future drug discovery with great expectations.

References

1. Aeluri M, Chamakuri S, Dasari B, Guduru SK, Jimmidi R, Jogula S, Arya P (2014) Small molecule modulators of protein-protein interactions: selected case studies. Chem Rev 114(9):4640–4694
2. Alushin GM, Lander GC, Kellogg EH, Zhang R, Baker D, Nogales E (2014) High-resolution microtubule structures reveal the structural transitions in alphabeta-tubulin upon GTP hydrolysis. Cell 157(5):1117–1129
3. Antoniades C, Bakogiannis C, Tousoulis D, Antonopoulos AS, Stefanadis C (2009) The CD40/CD40 ligand system: linking inflammation with atherothrombosis. J Am Coll Cardiol 54(8):669–677
4. Arkin MR, Tang Y, Wells JA (2014) Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. Chem Biol 21(9):1102–1114
5. Arkin MR, Wells JA (2004) Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. Nat Rev Drug Discov 3(4):301–317
6. Bahadur RP, Zacharias M (2008) The interface of protein-protein complexes: analysis of contacts and prediction of interactions. Cell Mol Life Sci 65(7–8):1059–1072
7. Bedford L, Lowe J, Dick LR, Mayer RJ, Brownell JE (2011) Ubiquitin-like protein conjugation and the ubiquitin-proteasome system as drug targets. Nat Rev Drug Discov 10(1):29–46
8. Bellmunt J, Szczylik C, Feingold J, Strahs A, Berkenblit A (2008) Temsirolimus safety profile and management of toxic effects in patients with advanced renal cell carcinoma and poor prognostic features. Ann Oncol 19(8):1387–1392
9. Bloom J, Cross FR (2007) Multiple levels of cyclin specificity in cell-cycle control. Nat Rev Mol Cell Biol 8(2):149–160
10. Blundell TL, Burke DF, Chirgadze D, Dhanaraj V, Hyvonen M, Innis CA, Parisini E, Pellegrini L, Sayed M, Sibanda BL (2000) Protein-protein interactions in receptor activation and intracellular signalling. Biol Chem 381(9–10):955–959
11. Boumpas DT, Furie R, Manzi S, Illei GG, Wallace DJ, Balow JE, Vaishnaw A, Group BGLNT (2003) A short course of BG9588 (anti-CD40 ligand antibody) improves serologic activity and decreases hematuria in patients with proliferative lupus glomerulonephritis. Arthritis Rheum 48(3):719–727
12. Cao YN, Zheng LL, Wang D, Liang XX, Gao F, Zhou XL (2018) Recent advances in microtubule-stabilizing agents. Eur J Med Chem 143:806–828
13. Ceccarelli DF, Tang X, Pelletier B, Orlicky S, Xie W, Plantevin V, Neculai D, Chou YC, Ogunjimi A, Al-Hakim A, Varelas X, Koszela J, Wasney GA, Vedadi M, Dhe-Paganon S,

Cox S, Xu S, Lopez-Girona A, Mercurio F, Wrana J, Durocher D, Meloche S, Webb DR, Tyers M, Sicheri F (2011) An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. Cell 145(7):1075–1087
14. Changeux JP (2013) The concept of allosteric modulation: an overview. Drug Discov Today Technol 10(2):e223–e228
15. Chen Q, Xie W, Kuhn DJ, Voorhees PM, Lopez-Girona A, Mendy D, Corral LG, Krenitsky VP, Xu W, Moutouh-de Parseval L, Webb DR, Mercurio F, Nakayama KI, Nakayama K, Orlowski RZ (2008) Targeting the p27 E3 ligase SCF(Skp2) results in p27- and Skp2-mediated cell-cycle arrest and activation of autophagy. Blood 111(9):4690–4699
16. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, Salzberg AC, Huang ES (2007) Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotechnol 25(1):71–75
17. Cheng KY, Noble ME, Skamnaki V, Brown NR, Lowe ED, Kontogiannis L, Shen K, Cole PA, Siligardi G, Johnson LN (2006) The role of the phospho-CDK2/cyclin a recruitment site in substrate recognition. J Biol Chem 281(32):23167–23179
18. Cho CS, Burkly LC, Fechner JH Jr, Kirk AD, Oberley TD, Dong Y, Brunner KG, Peters D, Tenhoor CN, Nadeau K, Yagci G, Ishido N, Schultz JM, Tsuchida M, Hamawy MM, Knechtle SJ (2001) Successful conversion from conventional immunosuppression to anti-CD154 monoclonal antibody costimulatory molecule blockade in rhesus renal allograft recipients. Transplantation 72(4):587–597
19. Cirillo L, Gotta M, Meraldi P (2017) The elephant in the room: the role of microtubules in Cancer. Adv Exp Med Biol 1002:93–124
20. Cohen P, Tcherpakov M (2010) Will the ubiquitin system furnish as many drug targets as protein kinases? Cell 143(5):686–693
21. Cong X, Liu Y, Liu W, Liang X, Laganowsky A (2017) Allosteric modulation of protein-protein interactions by individual lipid binding events. Nat Commun 8(1):2203
22. Corbi-Verge C, Garton M, Nim S, Kim PM (2017) Strategies to develop inhibitors of motif-mediated protein-protein interactions as drug leads. Annu Rev Pharmacol Toxicol 57:39–60
23. Cossins BP, Lawson AD (2015) Small molecule targeting of protein-protein interactions through allosteric modulation of dynamics. Molecules 20(9):16435–16445
24. Costanzo M, Nishikawa JL, Tang X, Millman JS, Schub O, Breitkreuz K, Dewar D, Rupes I, Andrews B, Tyers M (2004) CDK activity antagonizes Whi5, an inhibitor of G1/S transcription in yeast. Cell 117(7):899–913
25. Crown J, O'Leary M (2000) The taxanes: an update. Lancet 355(9210):1176–1178
26. Daikh DI, Finck BK, Linsley PS, Hollenbaugh D, Wofsy D (1997) Long-term inhibition of murine lupus by brief simultaneous blockade of the B7/CD28 and CD40/gp39 costimulation pathways. J Immunol 159(7):3104–3108
27. Darieva Z, Pic-Taylor A, Boros J, Spanos A, Geymonat M, Reece RJ, Sedgwick SG, Sharrocks AD, Morgan BA (2003) Cell cycle-regulated transcription through the FHA domain of Fkh2p and the coactivator Ndd1p. Curr Biol 13(19):1740–1745
28. de Bruin RA, McDonald WH, Kalashnikova TI, Yates J 3rd, Wittenberg C (2004) Cln3 activates G1-specific transcription via phosphorylation of the SBF bound repressor Whi5. Cell 117(7):887–898
29. Dobashi Y (2005) Cell cycle regulation and its aberrations in human lung carcinoma. Pathol Int 55(3):95–105
30. Dumontet C, Jordan MA (2010) Microtubule-binding agents: a dynamic field of cancer therapeutics. Nat Rev Drug Discov 9(10):790–803
31. Echalier A, Endicott JA, Noble ME (2010) Recent developments in cyclin-dependent kinase biochemical and structural studies. Biochim Biophys Acta 1804(3):511–519
32. Elgueta R, Benson MJ, de Vries VC, Wasiuk A, Guo Y, Noelle RJ (2009) Molecular mechanism and function of CD40/CD40L engagement in the immune system. Immunol Rev 229(1):152–172
33. Emanuel S, Rugg CA, Gruninger RH, Lin R, Fuentes-Pesquera A, Connolly PJ, Wetter SK, Hollister B, Kruger WW, Napier C, Jolliffe L, Middleton SA (2005) The in vitro and in vivo

effects of JNJ-7706621: a dual inhibitor of cyclin-dependent kinases and aurora kinases. Cancer Res 65(19):9038–9046
34. Ferrant JL, Benjamin CD, Cutler AH, Kalled SL, Hsu YM, Garber EA, Hess DM, Shapiro RI, Kenyon NS, Harlan DM, Kirk AD, Burkly LC, Taylor FR (2004) The contribution of Fc effector mechanisms in the efficacy of anti-CD154 immunotherapy depends on the nature of the immune challenge. Int Immunol 16(11):1583–1594
35. Field JJ, Diaz JF, Miller JH (2013) The binding sites of microtubule-stabilizing agents. Chem Biol 20(3):301–315
36. Fischer PM (2004) The use of CDK inhibitors in oncology: a pharmaceutical perspective. Cell Cycle 3(6):742–746
37. Fischer G, Rossmann M, Hyvonen M (2015) Alternative modulation of protein-protein interactions by small molecules. Curr Opin Biotechnol 35:78–85
38. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 47(7):1739–1749
39. Fry DC (2008) Drug-like inhibitors of protein-protein interactions: a structural examination of effective protein mimicry. Curr Protein Pept Sci 9(3):240–247
40. Galsky MD, Dritselis A, Kirkpatrick P, Oh WK (2010) Cabazitaxel. Nat Rev Drug Discov 9(9):677–678
41. Guarnera E, Berezovsky IN (2016) Allosteric sites: remote control in regulation of protein activity. Curr Opin Struct Biol 37:1–8
42. Guarnera E, Tan ZW, Zheng Z, Berezovsky IN (2017) AlloSigMA: allosteric signaling and mutation analysis server. Bioinformatics 33(24):3996–3998
43. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674
44. Heery CR, Ibrahim NK, Arlen PM, Mohebtash M, Murray JL, Koenig K, Madan RA, McMahon S, Marte JL, Steinberg SM, Donahue RN, Grenga I, Jochems C, Farsaci B, Folio LR, Schlom J, Gulley JL (2015) Docetaxel alone or in combination with a therapeutic cancer vaccine (PANVAC) in patients with metastatic breast cancer: a randomized clinical trial. JAMA Oncol 1(8):1087–1095
45. Holland EJ, Luchs J, Karpecki PM, Nichols KK, Jackson MA, Sall K, Tauber J, Roy M, Raychaudhuri A, Shojaei A (2017) Lifitegrast for the treatment of dry eye disease: results of a phase III, randomized, double-masked, placebo-controlled trial (OPUS-3). Ophthalmology 124(1):53–60
46. Hu Y, Li S, Liu F, Geng L, Shu X, Zhang J (2015) Discovery of novel nonpeptide allosteric inhibitors interrupting the interaction of CDK2/cyclin A3 by virtual screening and bioassays. Bioorg Med Chem Lett 25(19):4069–4073
47. Huang M, Song K, Liu X, Lu S, Shen Q, Wang R, Gao J, Hong Y, Li Q, Ni D, Xu J, Chen G, Zhang J (2018) AlloFinder: a strategy for allosteric modulator discovery and allosterome analyses. Nucleic Acids Res 46(W1):W451–W458
48. Huang W, Lu S, Huang Z, Liu X, Mou L, Luo Y, Zhao Y, Liu Y, Chen Z, Hou T, Zhang J (2013) Allosite: a method for predicting allosteric sites. Bioinformatics 29(18):2357–2359
49. James ND, Pirrie SJ, Pope AM, Barton D, Andronis L, Goranitis I, Collins S, Daunton A, McLaren D, O'Sullivan J, Parker C, Porfiri E, Staffurth J, Stanley A, Wylie J, Beesley S, Birtle A, Brown J, Chakraborti P, Hussain S, Russell M, Billingham LJ (2016) Clinical outcomes and survival following treatment of metastatic castrate-refractory prostate cancer with docetaxel alone or with strontium-89, zoledronic acid, or both: the TRAPEZE randomized clinical trial. JAMA Oncol 2(4):493–499
50. Jiang Y, Zhuang C, Chen L, Lu J, Dong G, Miao Z, Zhang W, Li J, Sheng C (2017) Structural biology-inspired discovery of novel KRAS–PDEδ inhibitors. J Med Chem 60:9400–9406
51. Jin L, Wang W, Fang G (2014) Targeting protein-protein interaction by small molecules. Annu Rev Pharmacol Toxicol 54:435–456

52. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci U S A 93(1):13–20
53. Jordan MA, Wilson L (2004) Microtubules as a target for anticancer drugs. Nat Rev Cancer 4(4):253–265
54. Jurgens G, Hermann A, Aktuna D, Petek W (1992) Dissociation-enhanced lanthanide fluorescence immunoassay of lipoprotein(a) in serum. Clin Chem 38(6):853–859
55. Karthiga A, Tripathi SK, Shanmugam R, Suryanarayanan V, Singh SK (2015) Targeting the cyclin-binding groove site to inhibit the catalytic activity of CDK2/cyclin a complex using p27 (KIP1)-derived peptidomimetic inhibitors. J Chem Biol 8(1):11–24
56. Kim WK, Henschel A, Winter C, Schroeder M (2006) The many faces of protein-protein interactions: a compendium of interface geometry. PLoS Comput Biol 2(9):e124
57. Komander D, Rape M (2012) The ubiquitin code. Annu Rev Biochem 81:203–229
58. Koya K, Li Y, Wang H, Ukai T, Tatsuta N, Kawakami M, Shishido CLB (1996) MKT-077, a novel rhodacyanine dye in clinical trials, exhibits anticarcinoma activity in preclinical studies based on selective mitochondrial accumulation. Cancer Res 56(3):538–543
59. Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV (2014) Membrane proteins bind lipids selectively to modulate their structure and function. Nature 510(7503):172–175
60. Lee FY, Borzilleri R, Fairchild CR, Kim SH, Long BH, Reventos-Suarez C, Vite GD, Rose WC, Kramer RA (2001) BMS-247550: a novel epothilone analog with a mode of action similar to paclitaxel but possessing superior antitumor efficacy. Clin Cancer Res 7(5):1429–1437
61. Li X, Chen Y, Lu S, Huang Z, Liu X, Wang Q, Shi T, Zhang J (2013) Toward an understanding of the sequence and structural basis of allosteric proteins. J Mol Graph Model 40:30–39
62. Liu N, Tu J, Dong G, Wang Y, Sheng C (2018) Emerging new targets for the treatment of resistant fungal infections. J Med Chem 61:5484–5511
63. Liu N, Zhu S, Zhang X, Yin X, Dong G, Yao J, Miao Z, Zhang W, Zhang X, Sheng C (2016) The discovery and characterization of a novel scaffold as a potent hepatitis C virus inhibitor. Chem Commun 52:3340–3343
64. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein-protein recognition sites. J Mol Biol 285(5):2177–2198
65. Lu S, Huang W, Zhang J (2014a) Recent computational advances in the identification of allosteric sites in proteins. Drug Discov Today 19(10):1595–1600
66. Lu S, Li S, Zhang J (2014b) Harnessing allostery: a novel approach to drug discovery. Med Res Rev 34(6):1242–1285
67. Lu S, Zhang J (2018) Small molecule allosteric modulators of G-protein-coupled receptors: drug-target interactions. J Med Chem. https://doi.org/10.1021/acs.jmedchem.7b01844
68. Malumbres M (2014) Cyclin-dependent kinases. Genome Biol 15(6):122
69. Malumbres M, Barbacid M (2001) To cycle or not to cycle: a critical decision in cancer. Nat Rev Cancer 1(3):222–231
70. Malumbres M, Barbacid M (2005) Mammalian cyclin-dependent kinases. Trends Biochem Sci 30(11):630–641
71. Mekhail TM, Markman M (2002) Paclitaxel in cancer therapy. Expert Opin Pharmacother 3(6):755–766
72. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, Subramaniam S (2016) Breaking Cryo-EM resolution barriers to facilitate drug discovery. Cell 165(7):1698–1707
73. Miao Z, Ali A, Hu L, Zhao F, Yin C, Chen C, Yang T, Qian A (2017) Microtubule actin cross-linking factor 1, a novel potential target in cancer. Cancer Sci 108(10):1953–1958
74. Milroy LG, Bartel M, Henen MA, Leysen S, Adriaans JM, Brunsveld L, Landrieu I, Ottmann C (2015) Stabilizer-guided inhibition of protein-protein interactions. Angew Chem Int Ed Eng 54(52):15720–15724

75. Mita AC, Denis LJ, Rowinsky EK, Debono JS, Goetz AD, Ochoa L, Forouzesh B, Beeram M, Patnaik A, Molpus K, Semiond D, Besenval M, Tolcher AW (2009) Phase I and pharmacokinetic study of XRP6258 (RPR 116258A), a novel taxane, administered as a 1-hour infusion every 3 weeks in patients with advanced solid tumors. Clin Cancer Res 15(2):723–730
76. Mitra A, Sept D (2008) Taxol allosterically alters the dynamics of the tubulin dimer and increases the flexibility of microtubules. Biophys J 95(7):3252–3258
77. Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. Annu Rev Cell Dev Biol 13:261–291
78. Morris PG (2010) Advances in therapy: eribulin improves survival for metastatic breast cancer. Anti-Cancer Drugs 21(10):885–889
79. Mullard A (2012) Protein-protein interaction inhibitors get into the groove. Nat Rev Drug Discov 11(3):173–175
80. Musgrove EA, Caldon CE, Barraclough J, Stone A, Sutherland RL (2011) Cyclin D as a therapeutic target in cancer. Nat Rev Cancer 11(8):558–572
81. Nabholtz JM, Gligorov J (2005) The role of taxanes in the treatment of breast cancer. Expert Opin Pharmacother 6(7):1073–1094
82. Nalepa G, Rolfe M, Harper JW (2006) Drug discovery in the ubiquitin-proteasome system. Nat Rev Drug Discov 5(7):596–613
83. Nero TL, Morton CJ, Holien JK, Wielens J, Parker MW (2014) Oncogenic protein interfaces: small molecules, big challenges. Nat Rev Cancer 14(4):248–262
84. Nussinov R, Tsai CJ (2013) Allostery in disease and in drug discovery. Cell 153(2):293–305
85. Nussinov R, Tsai CJ, Csermely P (2011) Allo-network drugs: harnessing allostery in cellular networks. Trends Pharmacol Sci 32(12):686–693
86. Olziersky AM, Labidi-Galy SI (2017) Clinical development of anti-mitotic drugs in cancer. Adv Exp Med Biol 1002:125–152
87. Osborne C, Challagalla JD, Eisenbeis CF, Holmes FA, Neubauer MA, Koutrelakos NW, Taboada CA, Vukelja SJ, Wilks ST, Allison MA, Reddy P, Sedlacek S, Wang Y, Asmar L, O'Shaughnessy J (2018) Ixabepilone and carboplatin for hormone receptor positive/HER2-neu negative and triple negative metastatic breast cancer. Clin Breast Cancer 18(1):e89–e95
88. Ottaggio L, Bestoso F, Armirotti A, Balbi A, Damonte G, Mazzei M, Sancandi M, Miele M (2008) Taxanes from shells and leaves of Corylus avellana. J Nat Prod 71(1):58–60
89. Panjkovich A, Daura X (2014) PARS: a web server for the prediction of protein allosteric and regulatory sites. Bioinformatics 30(9):1314–1315
90. Pegoraro AF, Janmey P, Weitz DA (2017) Mechanical properties of the cytoskeleton and cells. Cold Spring Harb Perspect Biol 9(11):a022038
91. Plumb JA (2004) Cell sensitivity assays: the MTT assay. Methods Mol Med 88:165–169
92. Propper DJ, Braybrooke JP, Taylor DJ, Lodi R, Styles P, Cramer JA, Collins WC, Levitt NC, Talbot DC, Ganesan TS, Harris AL (1999) Phase I trial of the selective mitochondrial toxin MKT077 in chemo-resistant solid tumours. Ann Oncol 10(8):923–927
93. Quezada SA, Jarvinen LZ, Lind EF, Noelle RJ (2004) CD40/CD154 interactions at the interface of tolerance and immunity. Annu Rev Immunol 22:307–328
94. Reynolds D, Shi BJ, McLean C, Katsis F, Kemp B, Dalton S (2003) Recruitment of Thr 319-phosphorylated Ndd1p to the FHA domain of Fkh2p requires Clb kinase activity: a mechanism for CLB cluster gene activation. Genes Dev 17(14):1789–1802
95. Richardson DL, Sill MW, Coleman RL, Sood AK, Pearl ML, Kehoe SM, Carney ME, Hanjani P, Van Le L, Zhou XC, Alvarez Secord A, Gray HJ, Landrum LM, Lankes HA, Hu W, Aghajanian C (2018) Paclitaxel with and without Pazopanib for persistent or recurrent ovarian cancer: a randomized clinical trial. JAMA Oncol 4(2):196–202
96. Roskoski R Jr (2016) Cyclin-dependent protein kinase inhibitors including palbociclib as anticancer drugs. Pharmacol Res 107:249–275
97. Rouhana J, Padilla A, Estaran S, Bakari S, Delbecq S, Boublik Y, Chopineau J, Pugniere M, Chavanieu A (2013) Kinetics of interaction between ADP-ribosylation factor-1 (Arf1) and the Sec7 domain of Arno guanine nucleotide exchange factor, modulation by allosteric factors, and the uncompetitive inhibitor brefeldin A. J Biol Chem 288(7):4659–4672

98. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437(7062):1173–1178
99. Saloustros E, Mavroudis D, Georgoulias V (2008) Paclitaxel and docetaxel in the treatment of breast cancer. Expert Opin Pharmacother 9(15):2603–2616
100. Schiff PB, Fant J, Horwitz SB (1979) Promotion of microtubule assembly in vitro by taxol. Nature 277(5698):665–667
101. Schiff PB, Horwitz SB (1980) Taxol stabilizes microtubules in mouse fibroblast cells. Proc Natl Acad Sci U S A 77(3):1561–1565
102. Schulze M, Stock C, Zaccagnini M, Teber D, Rassweiler JJ (2014) Temsirolimus. Recent Results Cancer Res 201:393–403
103. Schwabe RF, Hess S, Johnson JP, Engelmann H (1997) Modulation of soluble CD40 ligand bioactivity with anti-CD40 antibodies. Hybridoma 16(3):217–226
104. Scott DE, Bayly AR, Abell C, Skidmore J (2016) Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat Rev Drug Discov 15(8):533–550
105. Shen Q, Wang G, Li S, Liu X, Lu S, Chen Z, Song K, Yan J, Geng L, Huang Z, Huang W, Chen G, Zhang J (2016) ASD v3.0: unraveling allosteric regulation with structural mechanisms and biological networks. Nucleic Acids Res 44(D1):D527–D535
106. Sheng C, Dong G, Miao Z, Zhang W, Wang W (2015) State-of-the-art strategies for targeting protein-protein interactions by small-molecule inhibitors. Chem Soc Rev 44(22):8238–8259
107. Sho M, Sandner SE, Najafian N, Salama AD, Dong V, Yamada A, Kishimoto K, Harada H, Schmitt I, Sayegh MH (2002) New insights into the interactions between T-cell costimulatory blockade and conventional immunosuppressive drugs. Ann Surg 236(5):667–675
108. Silvian LF, Friedman JE, Strauch K, Cachero TG, Day ES, Qian F, Cunningham B, Fung A, Sun L, Shipps GW, Su L, Zheng Z, Kumaravel G, Whitty A (2011) Small molecule inhibition of the TNF family cytokine CD40 ligand through a subunit fracture mechanism. ACS Chem Biol 6(6):636–647
109. Smith MC, Gestwicki JE (2012) Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. Expert Rev Mol Med 14:e16
110. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122(6):957–968
111. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C (2008) Estimating the size of the human interactome. Proc Natl Acad Sci U S A 105(19):6959–6964
112. Suchard SJ, Stetsko DK, Davis PM, Skala S, Potin D, Launay M, Dhar TG, Barrish JC, Susulic V, Shuster DJ, McIntyre KW, McKinnon M, Salter-Cid L (2010) An LFA-1 (alphaLbeta2) small-molecule antagonist reduces inflammation and joint destruction in murine models of arthritis. J Immunol 184(7):3917–3926
113. Thiel P, Kaiser M, Ottmann C (2012) Small-molecule stabilization of protein-protein interactions: an underestimated concept in drug discovery? Angew Chem Int Ed Eng 51 (9):2012–2018
114. Towle MJ, Salvato KA, Budrow J, Wels BF, Kuznetsov G, Aalfs KK, Welsh S, Zheng W, Seletsky BM, Palme MH, Habgood GJ, Singer LA, Dipietro LV, Wang Y, Chen JJ, Quincy DA, Davis A, Yoshimatsu K, Kishi Y, Yu MJ, Littlefield BA (2001) In vitro and in vivo anticancer activities of synthetic macrocyclic ketone analogues of halichondrin B. Cancer Res 61(3):1013–1021
115. van Westen GJ, Gaulton A, Overington JP (2014) Chemical, target, and bioactive properties of allosteric modulation. PLoS Comput Biol 10(4):e1003559

116. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M (2009) An empirical framework for binary interactome mapping. Nat Methods 6(1):83–90
117. Vu B, Wovkulich P, Pizzolato G, Lovey A, Ding Q, Jiang N, Liu JJ, Zhao C, Glenn K, Wen Y, Tovar C, Packman K, Vassilev L, Graves B (2013) Discovery of RG7112: a small-molecule MDM2 inhibitor in clinical development. ACS Med Chem Lett 4(5):466–469
118. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE (2016) Emerging computational methods for the rational discovery of allosteric drugs. Chem Rev 116(11):6370–6390
119. Wang Q, Zheng M, Huang Z, Liu X, Zhou H, Chen Y, Shi T, Zhang J (2012) Toward understanding the molecular basis for chemical allosteric modulator design. J Mol Graph Model 38:324–333
120. Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT (1971) Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 93(9):2325–2327
121. Watterson SH, Xiao Z, Dodd DS, Tortolani DR, Vaccaro W, Potin D, Launay M, Stetsko DK, Skala S, Davis PM, Lee D, Yang X, KW MI, Balimane P, Patel K, Yang Z, Marathe P, Kadiyala P, Tebben AJ, Sheriff S, Chang CY, Ziemba T, Zhang H, Chen BC, Del Monte AJ, Aranibar N, McKinnon M, Barrish JC, Suchard SJ, Murali Dhar TG (2010) Small molecule antagonist of leukocyte function associated antigen-1 (LFA-1): structure-activity relationships leading to the identification of 6-((5S,9R)-9-(4-cyanophenyl)-3-(3,5-dichlorophenyl)-1-methyl-2,4-dioxo-1,3,7-tria zaspiro[4.4]nonan-7-yl)nicotinic acid (BMS-688521). J Med Chem 53(9):3814–3830
122. Wells JA, McClendon CL (2007) Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature 450(7172):1001–1009
123. Whitty A, Kumaravel G (2006) Between a rock and a hard place? Nat Chem Biol 2(3):112–118
124. Yau R, Rape M (2016) The increasing complexity of the ubiquitin code. Nat Cell Biol 18(6):579–586
125. Zarzycka B, Kuenemann MA, Miteva MA, Nicolaes GA, Vriend G, Sperandio O (2016) Stabilization of protein-protein interaction complexes through small molecules. Drug Discov Today 21(1):48–57
126. Zhang B, Wu T, Chen M, Zhou Y, Yi D, Guo R (2013) The CD40/CD40L system: a new therapeutic target for disease. Immunol Lett 153(1–2):58–61
127. Zhong M, Gadek TR, Bui M, Shen W, Burnier J, Barr KJ, Hanan EJ, Oslob JD, Yu CH, Zhu J, Arkin MR, Evanchik MJ, Flanagan WM, Hoch U, Hyde J, Prabhu S, Silverman JA, Wright J (2012) Discovery and development of potent LFA-1/ICAM-1 antagonist SAR 1118 as an ophthalmic solution for treating dry eye. ACS Med Chem Lett 3(3):203–206
128. Zhuang C, Miao Z, Sheng C, Zhang W (2014a) Updated research and applications of small molecule inhibitors of Keap1-Nrf2 protein-protein interaction: a review. Curr Med Chem 21:1861–1870
129. Zhuang C, Miao Z, Wu Y, Guo Z, Li J, Yao J, Xing C, Sheng C, Zhang W (2014b) Double-edged swords as cancer therapeutics: novel, orally active, small molecules simultaneously inhibit p53–MDM2 interaction and the NF-κB pathway. J Med Chem 57:567–577

Chapter 14
Allosteric Modulation of Intrinsically Disordered Proteins

Ashfaq Ur Rehman, Mueed Ur Rahman, Taaha Arshad, and Hai-Feng Chen

Abstract The allosteric property of globular proteins is applauded as their intrinsic ability to regulate distant sites, and this property further plays a critical role in a wide variety of cellular regulatory mechanisms. Recent advancements and studies have revealed the manifestation of allostery in intrinsically disordered proteins or regions as allosteric sites present within or mediated by IDP/IDRs facilitates the signaling interactions for various biological mechanisms which would otherwise be impossible for globular proteins to regulate. This thematic review has highlighted the biological outcomes that can be achieved by the mechanism of allosteric regulation of intrinsically disordered proteins or regions. The similar mechanism has been implemented on Adenovirus 5 early region 1A and tumor apoptosis protein p53 in correspondence with other partners in binary and ternary complexes, which are the subject of the current review. Both these proteins regulate once they bind to their partners, consequently, forming either a binary or a ternary complex. Allosteric regulation by IDPs is currently a subject undergoing intense study, and the ongoing

Ashfaq Ur Rehman and Mueed Ur Rahman contributed equally to this chapter.

A. U. Rehman
State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Department of Biochemistry, Abdul Wali Khan University Mardan, Mardan, Pakistan

M. U. Rahman · T. Arshad
State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

H.-F. Chen (✉)
State Key Laboratory of Microbial Metabolism, Department of Bioinformatics and Biostatistics, National Experimental Teaching Center for Life Sciences and Biotechnology, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

Shanghai Center for Bioinformation Technology, Shanghai, China
e-mail: haifengchen@sjtu.edu.cn

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_14

research work will ensure a better understanding of precision and efficiency of cellular regulation by them. Allosteric regulation mechanism can also be researched by intrinsically disordered protein-specific force field.

Keywords Allosteric regulation · IDPs · E1A · p53 · Partners

Abbreviations

ATM	Ataxia-telangiectasia mutated (gene ATM)
ATR	Ataxia-telangiectasia and Rad3-related (ATR)
CAK	CDK-activating kinase
CBP	CREB-binding protein (CREBBP)
CDC2	Cell division cycle 2 kinase
CDKs	Cyclin-dependent kinases (multiple members)
CHK1	Cell cycle checkpoint kinase 1 (CHEK1)
CHK2	Cell cycle checkpoint kinase 2 (CHEK2)
CK1	Casein kinase 1 (multiple isoforms)
CK2	Casein kinase 2 (multiple isoforms)
CSN	COP9 signalosome (protein complex)
DNA-PK	DNA-dependent protein kinase (PRKDC)
ERK2	p42 mitogen-activated protein kinase (MAPK1)
FACT	Facilitating chromatin-mediated transcription
HIPK2	Homeodomain-interacting protein kinase 2
JNK	Jun N-terminal kinase (MAPK8)
MDM2	Mouse double-minute 2 homologue
P38	p38 mitogen-activated protein kinase (MAPK14)
P300	E1A-binding protein, 300-kDa (EP300)
PCAF	P300/CBP-associated factor
PKC	Protein kinase C (multiple isoforms)
PKR	Double-stranded RNA-dependent protein kinase (PRKR)
PIAS	Protein inhibitor of activated STAT (multiple isoforms)
PIN1	Peptidyl-prolyl-cis-trans isomerase 1
RSK2	Ribosomal S6 kinase 2 (RPS6KA3)
SET9	SET domain-containing protein 9 (SET9)
STK15	Serine/threonine protein kinase 15
TAF II250	TATA-binding protein-associated factor 250-kD (TAF1)

14.1 Introduction

This thematic issue focuses on allosteric modulation of intrinsically disordered proteins and regions (IDPs and IDRs). The word allostery is initially articulated over 50 years before [1–3] and had endured a central focus in the field of biology.

To understand most of the biological processes beyond the molecular level, such as cellular disease and signaling, the quantitative description of allostery plays a critically important role and is a fundamental tool [4, 5]. Allostery has even been stated as the "second secret of life" [4, 5]. The classical characterization of the word allostery [6] was ground on the three-dimensional (3D) protein subunit, structural annotations of their cooperative behaviors. The concept has since been expanded to accommodate a wide variety of cellular regulatory proteins to reveal the regulatory mechanisms concerning an isoform of proteins, where a single modulator (either protein or small molecule) binds at the active or catalytic site and ultimately can modulate functions at a distant site. Extensive research about the allosteric regulation from theoretical and experimental data focused on the presence of distinct energy conformational states of an allosteric effector (in the presence and absence). Subsequently from these concepts regarding the allosteric regulation mechanism from distinct energy states has been broadly advanced our understanding of how small molecule or macromolecule can facilitate a wide range of cellular signals to reveal the specific regulation. It became clear since the discovery of word allostery that conformational changes in the protein structure alone could not interpret for all observations of allosteric processes in the biological system [7, 8]. The modulatory properties are interdependent with the ternary complex consisting of the target protein, the primary ligand, and the modulator. There is a different type of allosteric modulator, which can be described as positive, negative, and silent allosteric modulators, respectively. The positive allosteric modulators (PAMs) strengthen the effect of the primary ligand (PL) [9]. Most benzodiazepines act as PAMs at the $GABA_A$ receptor, while the negative allosteric modulators (NAMs) ultimately reduce the impact of the PL-(Ro15-4513) at the $\alpha1\beta2\gamma2$-$GABA_A$ receptor. The silent allosteric modulators (SAMs) occupy the allosteric binding site and behave functionally neutral.

Over the past decade, extensive research work has driven a reevaluation of the descriptions of allosteric phenomena of IDPs and IDRs. Nearly 1/3 of human proteins contain disordered regions around =>30 amino acid in length [10]. IDPs and IDRs are flexible in behavior and proficient of quickly sampling an ensemble of energy states protein conformational [11]. Generally, the IDPs often function as a molecular hub in molecular interaction networks and are highly enriched in sites for posttranslational modification [12–14]. These fundamental characteristics of IDPs make them ideally appropriate for cellular function (regulation and signaling) [14, 15]. Recent studies have shown that the tremendous capability of these IDPs to rapidly participate with a wide variety of cellular proteins and convey a wide range of cellular signals could be linked to a mechanistic process called allostery [11, 16, 17]. Though, in some cases, the phenomena of allostery were observed within a single disordered polypeptide chain, in which a coupled folding and binding event in one site influences subsequent interactions or modifications at a distant site, in other cases, these IDPs act as extremely particular allosteric effectors, cooperatively controlling various cellular protein binding networks on protein surfaces and sophisticatedly orchestrating cellular responses to stimuli and decisions about cell fortune.

Despite its importance, allosteric mechanisms in most instances remain a biophysical enigma, eluding a general, quantifiable, and predictive atomic description. Here in this review, we focused on the binary and ternary complexes of the intrinsically disordered proteins (IDPs), for example, the adenovirus early region 1A (E1A) oncoprotein that de/regulates cell growth by interacting with multiple cellular proteins and playing a role as a hub in such allosteric regulation/deregulation. In this review, we highlight the multiplicity of an allosteric mechanism for IDP-associated molecular process and distinctive regulatory sensitivity that can be accomplished over IDP-facilitated allostery. Each example is going to be cited with a spotlight on the biological and purposeful significance of the IDPs concerned, further as inside the context of presently offered theoretical descriptions and mechanistic models of allosteric regulation mediated by IDPs. Finally, we will provide our perspective on the longer-term growth of the thought of allostery to incorporate the processes delineated here and on the experimental strategies and theoretical framework which will allow for further characterization of these complex yet highly essential systems.

14.2 Adenovirus 5 Early Region 1A Is an Intrinsically Disordered Protein

The previous literature study regarding virus and their genome ensure the researcher to let propose that virus is the obligate intracellular parasites and their genomes do not seem to be massive enough to encode all the functions mandatory to produce progeny independently; thus, all the viruses are entirely dependent on host cell functions. Mechanistically, in eukaryotes, these processes (host cell) are finely established on a complex series of molecular interactions. Specifically, the implementation of these complex biological processes let based on the interaction of thousands of different partner proteins and hence called complex molecular networks [18, 19]. Generally, the protein interaction network among the partner proteins is associated with the cellular proteins either two or one in number. However, the minority of proteins form protein network hub via interaction with tens, or even hundreds, of other proteins.

Consequently, this protein network hub plays essential roles in regulating the activity of the partner protein and efficiently forming useful modules within the cellular interactome [20, 21]. In the time of or during viral infection, the cellular hub protein served a central role in regulating cell functions, more ever making them an ideal target. Targeting single or multiple cellular hub effect differently; by targeting single, a virus-related regulatory protein can effectively gain control over an entire module, potentially contained plenty of proteins (>100s), while, on the other hand, targeting multiple, a virally encoded hub can transform the configuration of the entire cellular protein network, and virtually reprogrammed the entire physiology and related function. The virus-related oncogenes of the small DNA tumor viruses

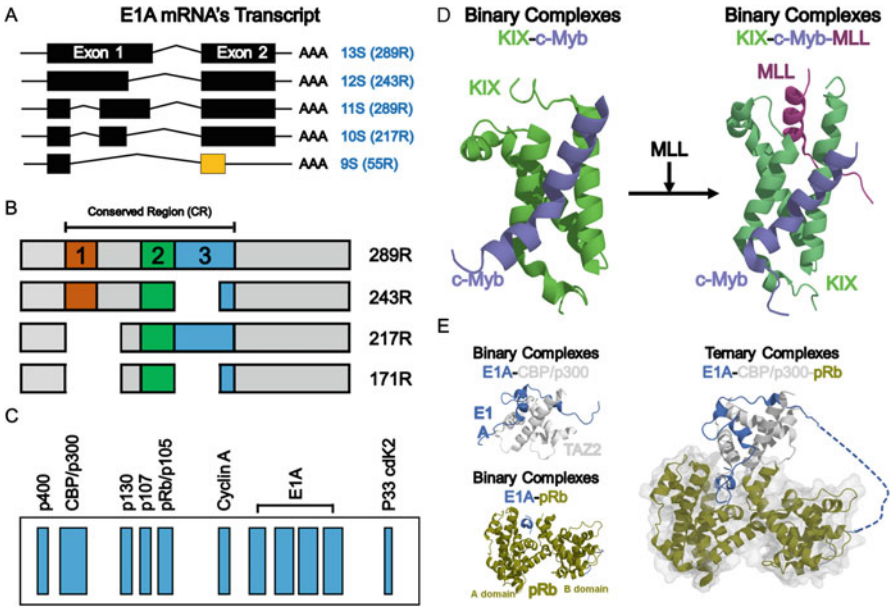


Fig. 14.1 The graphical representation of the E1A mRNAs and their proteins containing conserved regions. (**a**) The viral protein E1A coding regions represented as a black color-filled box all read in the same frame except for the second exon of the 9S mRNA (light brown color-filled box), and (**b**) different regions of the conserved residue sequence in the E1A proteins were labeled as CR1, CR2, and CR3. (**c**) Representation of an idealized fluorograph of the proteins co-immunoprecipitated with E1A from lysates of radiolabeled Ad5-infected cells. (**d**) The structural basis of positive cooperativity for ligand binding by the KIX domain of CBP/p300; the structures of the binary KIX:c-Myb complex (left; PDB ID, 1SB0) and the ternary MLL:KIX:c-Myb complex (right; PDB ID, 2AGH). (**e**) Folding of the intrinsically disordered protein E1A induced by binding to pRb and the TAZ2 domain of CBP/p300. An E1A binding and folding equilibrium, showing the formation of the ternary complex from the unbound intrinsically disordered protein state by way of two binary intermediate complexes

encode potent viral hub proteins. Among them, the intrinsically disordered Adenovirus 5 early region 1A (E1A) is one of the best and well-characterized viral hub proteins. The E1A oncogene performs many genes and growth regulatory activities, making E1A a useful tool with which to study important cellular allosteric processes.

The transcription from the intrinsically disordered Adenovirus type 5 and highly related type 2 becomes detectable after 1 hour of infection. Subsequently, the first transcribed viral gene is E1A [22]. Next, this primary E1A transcript is further processed by differential splicing to produce five distinct messages (13S-9S) as shown in Fig. 14.1a [23, 24]. At the early time of infection, these transcripts (13S and 12S) are the most abundant one among all the five transcripts. The rest of the transcript become more abundant at late times after infection, i.e., the mRNAs 11S and 10S which are less abundant at an early stage and are minor species [23, 24]. The mRNAs of E1A (13S, 12S, 11S, 10S, and 9S) encode for distinct protein like 289 residues (R), 243R, 217R, 171R, and 55R as shown in Fig. 14.1a. All the product

is detectable in vivo except the 9S product, which has only been detected in vitro [23]. In the infected cell, mainly, the major E1A proteins (289R and 243R) regulate transcription of both cellular and viral genes.

The major products encoded by the viral E1A proteins (289R and 243R) possess similar N- and C-terminal amino acid sequence. The E1A protein 289R possess additionally 46 amino acids which makes them bigger and different from the other isoforms of the same protein, and these additional 46 residues are the result of the differential splicing of the primary E1A transcript as shown in Fig. 14.1a. This extra internal stretch is often stated as the "unique region." The extra stretch comprised of 46 amino acids required for efficient transactivation of early viral promoters [25].

Comparing E1A sequences of various human and chimpanzee, gorilla, monkey (all simian) adenovirus serotypes has similarities in the CR1-3 regions have conserved amino acid [26, 27]. In Ad5, the amino acid sequences (from res 40 to res 80) of the conserved region 1 (CR1) maps, CR2 between amino acids res 121 to res 139, and CR3 between res 140 to res 188 which roughly overlaps with the 13S unique region as shown in Fig. 14.1b. The evolutionary conservation of these sequences further suggests that these sequences have a critically important role (critical) in E1A function but by no means limit the possibility that other regions are at least correspondingly important.

The protein sequence of E1A have plenty of proline contents (P), acidic, and localized in the nucleus. The rapid nuclear localization is facilitated by a small peptide comprised of highly basic amino acid sequence (L-R-P-R-P) at the extreme C-terminus of the polypeptides [28]. The presence of the rich proline contents in the E1A protein executes a conformational constraint which likely limits the formation of substantial secondary structure in the E1A proteins. The extreme heat stability of bacterially produced E1A protein, which recalls strong transcriptional activation activity even after boiling for 5 minutes [29], suggests that E1A either can readily refold to an active conformation or can function as a random coil. The viral E1A protein possesses a remarkable ability to tolerate insertions of new sequences or deletions without interruption of its biological function, and by such ability, the researchers emphasized to propose the concept that E1A could be a series of small allosteric modular domains that are comparatively independent of surrounding sequences.

The extra internal stretch among the encoded conserved regions (in 289R) has been identified that this extra stretch residue is the potential target for the metal binding with a consensus zinc finger motif (cysteine-Xaa2-cysteine-Xaa13-cysteine-Xaa2-cysteine) as in Fig. 14.2a [25]. In the consensus sequence, the cysteine (Cys) residues serve as postulated ligands [25]. The E1A isoform 289R does bind a single zinc ion, and as expected the other E1A isoform (smaller E1A) that lacks this unique region does not bind. The binding of Zn^{2+} to four Cys residues plays an essential role in transcriptional activation [25]. While replacing a single Cys residue of the finger with other amino acids destroying the transactivating capability of the 289R of E1A, even though replaced amino acid belongs to structurally or functionally conserved [25]. To date, it has remained unclear whether this Zn^{2+} would facilitate or not the interaction among the E1A and DNA or with other cellular proteins.

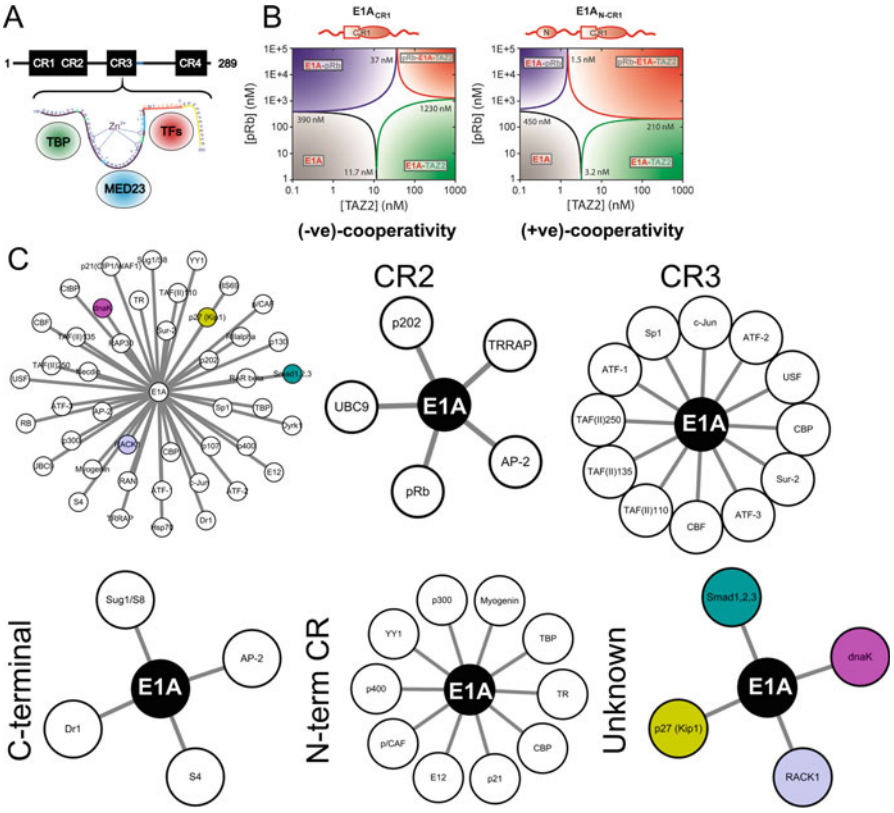


Fig. 14.2 (**a**) The current model of E1A CR3-dependent activation of transcription. Top, linear representation of 289R E1A. CRs are labeled, and AR1 is denoted in yellow. Bottom, residues of E1A CR3 from 139 to 200 are indicated using a one-letter code. The coordinating cysteines are circled, and the critical targets of CR3 are shown in green, blue, and red, respectively. The corresponding colors indicate the key residues interacting with each target. The boundaries of the residues of CR3 known to be required for interaction with APIS (residues 169–188) and the 20S proteasome (residues 161–177) are marked by # and ∗ Published in J Virol, vol. 82, p. 7252–7263. (**b**) A cooperativity switch in the adenoviral oncoprotein E1A. Phase diagrams for pRb and TAZ2 binding illustrate how the formation of pRb:E1A:TAZ2 ternary complexes can be described by either negative (left) or positive cooperativity (right), depending on the availability of binding sites. Kd values for the formation of various complexes with the E1A constructs shown above the phase diagrams are indicated at the phase boundaries. Published in Nature, vol. 498, p. 392, 2013 by Springer Nature, and (**c**) Interactions of the N-terminal, CR1 and CR2 motifs of E1A with cellular proteins

14.3 Intrinsically Disordered Proteins Act as Allosteric Hub

It is generally expected that the E1A protein indirectly affects the activity of the associated cellular protein, subsequently either enhancing or reducing their activities. Due to that reason, the E1A protein does not possess the specific DNA binding activity nor an intrinsic enzymatic activity. The immunoprecipitation of E1A proteins, extracted from [35S]-Met-labeled infected cells, shows that the E1A complexes specifically with numerous cellular regulatory proteins. Moreover, the major species of coprecipitating proteins migrate at rates corresponding to 105 kDa, 107 kDa, and 300kDa, with minor species migrating at 33 kDa, 60 kDa, and 130 kDa, respectively, as shown in Fig. 14.1c [30, 31]. From the previous literature study, it is identified that the E1A protein interacts with numerous cellular proteins as shown in Fig. 14.2c. For example, CREB-binding protein (CBP), a known transcriptional regulatory protein, and its associated protein p300 both function as principal regulators of cellular machinery by mediating interactions between the basal transcription machinery and different transcriptional factors [32, 33]. The CBP is bulky, contains multiple functional protein domain features, and even possesses stably folded domains, single molten globular domain, and many disordered regions; among them, some have known function, and many of them have unknown function [34, 35]. These both disordered and fully structured regions of the CBP protein interacting with a wide variety of other disordered protein partners, for example recent studies revealed the allosteric regulation mechanism of the CBP and its related protein p300. Additionally, it was assumed that the allosteric mechanism plays an essential role in modulating the transcriptional co-activation function of CBP/p300 as well as interact with the retinoblastoma protein (pRb) [36, 37], subsequently provoke the acetylation of RB1 through engaging EP300 and RB1 into a multiple interactive protein complexes [38]. The interaction of E1A with host CTBP1-2 and associated protein seems to be involved in viral replication [39, 40], and likely E1A interacts with a wide variety of cellular proteins like BS69 [41], p60/cyclin A [42], p107 [43], p33 cdk2 [44], and p130 [45, 46]. Similarly, interaction of E1A with host KPNA4 allows E1A import into the host nucleus [47]. While EIA-TBP interaction probably disrupts the TBP-TATA complex [48], the TATA-binding protein (TBP) [48–50], and various other components of the TFIID complex [51].

In biochemistry, the KIX domain (kinase-inducible domain (KID) interacting domain) or CREB-binding domain is a protein domain of the eukaryotic transcriptional co-activators CBP and P300. It serves as a docking site for the formation of heterodimers between the coactivator and specific transcription factors. Structurally, the KIX domain is a globular domain consisting of three α-helices and two short 310 helices. This KIX domain is an essential binding site for several disordered transcriptional factors involved in cell differentiation, leukemogenesis, and viral transformation [52, 53].

The activation domains of both transcriptional factors (pKID and c-Myb) bind to a common site on this KIX, whereas the mixed lineage leukemia (MLL) protein binds to a distinct surface on the opposite face of KIX as shown in Fig. 14.1d

[54, 55]. The structural and biochemical studies show the cooperative binding behavior of the polypeptide complexes (pKID and MLL or c-Myb and MLL) to KIX to make ternary complexes; the binding of MLL in its related site increases the binding of both c-Myb and pKID and vice versa [54, 56, 57]. The mentioned studies indicated that KIX functions not only as a structured scaffold for binding of disordered ligands but also as an allosteric modulator of transcription. Moreover, the researchers mapped the regions of the E1A protein required for interaction with other proteins in a variety of cells using a multitude of mutants. However, it seems likely that the targeted protein E1A functions indirectly through its association with these and other yet unidentified cellular proteins.

To the best of our knowledge, the deep literature search regarding the interaction of E1A with other associated proteins lets the researcher conclude and assume that E1A binds to these associated proteins and either sequesters them, preventing their action or modifying their enzymatic activities. The critically important property of IDPs is always appreciated that they possess multiple interaction motifs for their interactive cellular partners, hence, they can potentially bind to multiple cellular partners in a variety of combinations, and allowing them to function as molecular hubs in interaction complex protein networks [13, 58]. As previously mentioned, the one of the tremendous characteristic of the IDPs is used as a hub protein in a wide variety of protein networks, for example adenovirus oncoprotein early region (E1A), which interacts with the TAZ2 domain of CBP/p300 and the retinoblastoma protein (pRb) to further induce epigenetically reprograming cellular transcriptional processes in the host cell as reported previously [59, 60]. Interestingly, the protein complex of E1A with its partner can be modulated by both (+ve) and (−ve) cooperativity, based on the availability of various binding sites for its partner [61]. For example, the construct of unbound E1A contains the only CR1, which is capable of binding to both pRb and TAZ2 and displays (−ve) cooperativity in the formation of a ternary complex of an E1A-TAZ2-pRb, once the E1A CR1 region is already bound to other cellular partners as shown in Fig. 14.2b^{left}. In case of displays (+ve) cooperativity, once extension in the E1A construct conduct is to include the amino-terminal domain further and results in complex ternary formation (E1A-TAZ2-pRb) as shown in Fig. 14.2b^{right}. The switching pattern of cooperativity from (−ve) to (+ve) for both ternary complexes is determined by the incorporation of additional interaction site of the TAZ2 domain, hence suggesting an allosteric regulatory mechanism in which occupancy of the amino-terminal domain controls the interaction of pRb in the CR1 region. Remarkably, binding promiscuity of E1A with partner protein complex has significant consequences for the functional output. The binary complex of E1A and pRb regulates the cell cycle, whereas the binary complex between E1A and TAZ2 eases acetylation of E1A by CBP/p300 and subsequently induces epigenetic regulation of cellular transcription. The allosterically enhanced ternary complex formed between E1A, CBP/p300, and pRb permits for efficient interruption of the cell cycle by endorsing pRb acetylation through CBP/p300 cellular protein, thus targeting this pRb protein for degradation and further permitting the virus to take over the whole machinery of cellular transcription, compelling cell cycle progression, and unrestrained cellular proliferation [62].

14.4 The p53 Protein as an IDP

The pattern of disorder within molecular hub proteins can modulate functional output. Intrinsically disordered proteins (IDPs) can modulate their multiple interaction motifs (fully or partially disordered binding sites) according to the structural propensity of their molecular partners. By this way, they function as molecular hubs in interaction networks where they can potentially bind to multiple cellular partners in a variety of combinations [13, 58]. Molecular hub proteins show varying degrees of binding propensity, i.e., some binding modules prefer interactions with high affinity and specificity, while other interactions can favor low affinity and low specificity and all possible combinations in between. Among the other IDPs, p53 protein is considered as a major intrinsically disordered hub protein. To improve understanding of the use of disorder for binding diversity, p53 was studied as a prototypical example of hub proteins involved in crucial biological functions.

p53 is a key player in a large signaling network involving the expression of genes carrying out such processes as cell cycle progression, apoptosis induction, DNA repair, and response to cellular stress [63]. Loss of p53 function, either directly through mutation or indirectly through several other mechanisms, is often accompanied by cancerous transformation [64]. The p53 protein interacts with many other proteins in order to carry out its signal transduction function. The disordered regions enable p53 to bind with its multiple partners at the same time, as reported previously [65–81]. A number of these binding partners are downstream targets, such as transcription factors, and others are activators or inhibitors of p53's transactivation function. From the disorder point of view, four regions or domains in p53 are characterized as the N-terminal transcriptional activation domain, the central DNA binding domain, the C-terminal tetramerization domain, and the C-terminal regulatory domain. The last two could be a single C-terminal domain with two subregions. The tetramerization domain is intrinsically disordered, and the structure is acquired upon the formation of the complex, while the DNA binding domain is intrinsically structured, and the terminal domains are intrinsically disordered [82, 83].

14.4.1 The Associated Binding Partners for the Formation of a Ternary Complex with p53

The IDPs/IDRs could be implemented in protein-protein interactions in at least two different modes: One disordered region could bind to many partners (one-to-many signaling), and multiple disordered regions could bind to one site on one partner (many-to-one signaling) [84–86]. The p53 adopts the one-to-many mode of interaction due to the intrinsically structured and disordered nature of the proteins. The interacting partners are characterized into three main categories due to the binding site they interacted with and shown in Table 14.1. For example, experimentally characterized intrinsically disordered C-terminal domain is making a complex with

Table 14.1 p53 associated with protein partner

N-terminal domain	PR	DBD	C-terminal domain	
TFIID	Sin3	SV40 Tag	TBP	Brca1-exon11
TAFII31	WOX1	hRAD51	YB1	MUC1
Polβ	IkBa	Sp1	DNMT1	ATF3
Pin1		Tip60	NF-Y	hRAD51
CREB		JNK Axin	E2F1	hRAD54
COP9		Brca1-BRCT2	Sp3	NIR
Seladin-1		HIF-1α	ik3-1/cables	GRP75
ING1b		Bcl-2	hCdc14A/B	hSir2/SIRT1
FAK		Plk1	TFIIH	Cdk9
PRMT1		MAML1	WRN	CSB
CHC		53BP1	BLM	14-3-3
PID/MTA2		Pirh2	topoI	IFI16
Smad2/3		Bak	topoII	Nucleolin
Rpa70		Chk2	hRAD51	Securin
MDM2		ASPP2	hRAD54	parc
Tfb1		Bcl-X	NIR	CUL7
		DMC1	GRP75	CK2b
		PML-IV	hSir2/SIRT1	Cyclin A
		ZBP-89	Cdk9	PTEN
		NIR		ERα
		Brn-3a	ZBP-89	CARM1
		E4F1	YY1	mot-2
		TR-3	STK15	Sp1
		DNA	AP2α	C/EBPβ
		53BP1	HAUSP/USP7	p18
		53BP2	HSF3	
		sv40	TR-3	set9
			CBP	tGcn5

cyclin A (PDB ID: 1H26) [75], sirtuin (PDB ID: 1MA3) [65], CREB-binding protein (CBP; PDB ID: 1JSP) [76], S100ββ (PDB ID: 1DT7) [81], set9 (PDB ID: 1XQH) [68], and tGcn5 (PDB ID: 1Q2D) [79], while intrinsically disordered N-terminal domain interacts with Rpa70 (PDB ID: 2B3G) [66], Mdm2 (PDB ID: 1YCR) [72], Tfb1 (PDB ID:2GS0) [69], and itself [73]. The interactions mediated by the structured region (DBD) of p53 and its partners are DNA (PDB ID: 1TSR) [67], the BRCT domain of BP1 (PDB ID: 1L0B) [71], the SH3 domain of BP2 (PDB ID: 1YCS) [70] and large T-antigen (LTag) from simian virus 40 (PDB ID: 2H1L) [74]. More interestingly, the disordered C-terminal regulatory region (from residues 374 to 388) forms all three major secondary structure types in the bound state: a helix when associating with S100ββ, a sheet with sirtuin, and distinct irregular structures with CBP and cyclin A2 (Fig. 14.3).

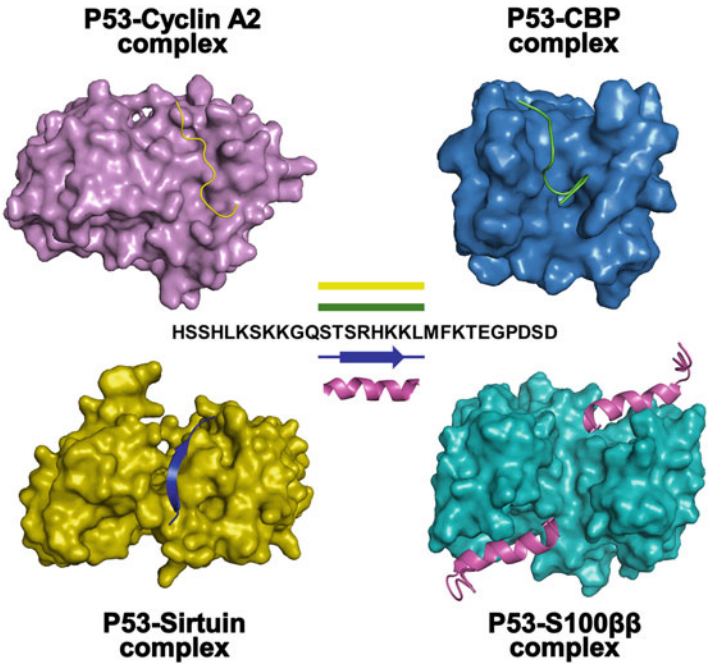


Fig. 14.3 **S**tructural comparison of p53 as an IDP hub protein**.** Primary, secondary, and quaternary structure of p53 complexes. (Published in BMC Genomics 2008, doi:10.1186/1471-2164-9-S1-S1)

Similarly, the structural study reveals that intrinsically disordered N-terminal transactivation domain (TAD) utilizes two molecular recognition elements, the amphipathic AD1 and AD2, to interact with a variety of cellular targets [87, 88], including complexes with multiple domains of the transcriptional coactivators CBP/p300 [89, 90], MDMX [91], and MDM2 [92]. In its free form, AD1 of the N-terminal TAD exists in equilibrium between disordered and partially helical conformations [83, 91–93], whereas residues 19–25 form a stable amphipathic α-helix in the Mdm2 complex [72]. Recently, it has been reported that the mutation of proline in residue C-terminal of MDM2-binding motif substantially increases both the residual helicity and the binding affinity for MDM2. Conserved proline residues outside the Mdm2 binding site preserve these defined levels of helicity, which is ultimately required for productive p53 signaling [94]. In p53, drastic conformational changes enable distinct disordered regions to modulate into conformations suitable for binding with different binding partners, in which the structured partner alters the propensity of the disordered region via a binding-induced manner (Fig. 14.4).

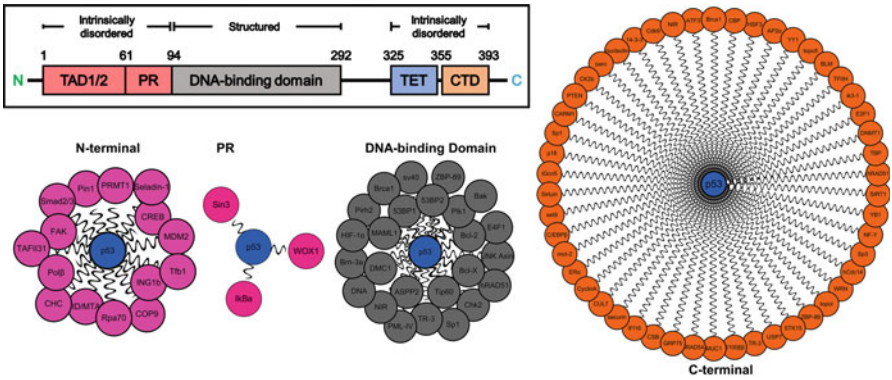


Fig. 14.4 Protein interaction network for p53 protein

14.4.2 Posttranslational Modifications (PTMs)

Posttranslational modification (PTM) of proteins refers to the chemical changes that occur after mRNA is translated into protein products, which can play important roles in controlling the affinity and the order of binding events in IDPs. Multiple different posttranslational modifications (PTMs) have been identified in p53 (Fig. 14.5), thereby altering protein interactions, namely, phosphorylation, cis-trans isomerization, acetylation, ubiquitination, methylation, SUMOylating, and neddylation at more than one site. The conformational flexibility of the intrinsically disordered sites/regions facilitates the access and the sequential regulation by multiple PTMs [95]. Multiple studies have shown that the PTMs occur most often in disordered regions [96, 97]. For example, the C-terminal domain of p53 contains residues modified by different PTMs in order to tune its specific interaction properties of p53 [98].

14.4.2.1 Phosphorylation

p53 phosphorylation has been widely investigated and associated with protein stabilization. For example, the interaction between p53 and its significant negative regulator, MDM2, is diminished when three N-terminal sites, Ser15, Thr18, and Ser20, are phosphorylated, while P300 (acetyltransferase) binding is increasing the level and stability of p53. Interestingly, increase in the number of phosphorylation events on the N-terminal TAD directly increases the binding affinity of the TAD for

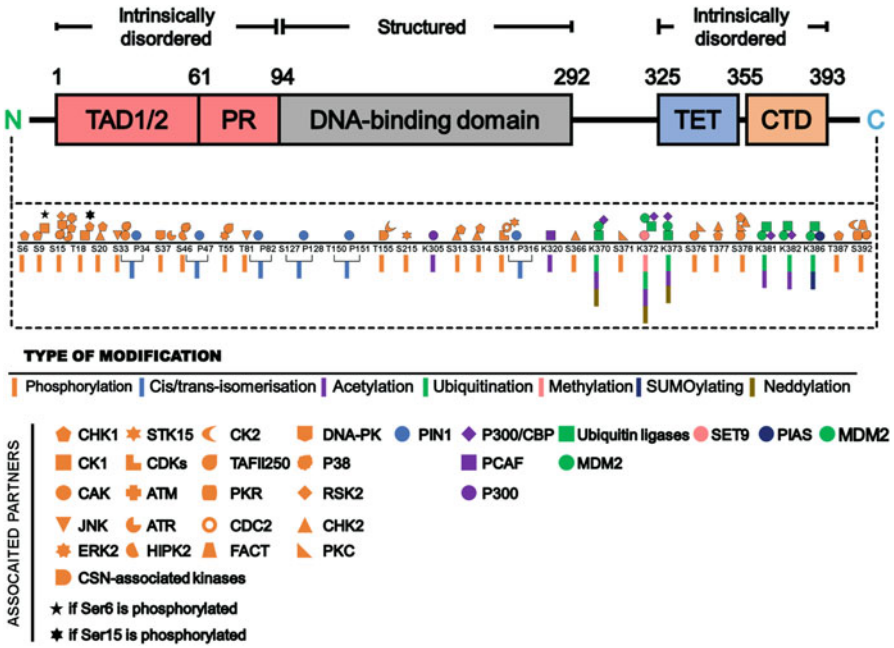


Fig. 14.5 **p53** posttranslational modifications (PTMs) at different sites via various interacting partners

the TAZ1, TAZ2, and KIX domains of CBP/p300, with each successive phosphorylation event having an additive effect on the measured binding affinity [99]. Similarly, phosphorylation at Ser46 prefer to bind at PTEN (a tumor suppressor) promoter for MDM2, thus abolish the autoregulatory loop that contributes to control the p53 level, while phosphorylation of a single threonine residue (Thr18) as part of a signaling cascade impairs binding to the E3 ubiquitin ligase HDM2 [99, 100]. In both of these cases, a single phosphorylation event is sufficient for the regulation of critical cellular processes, highlighting the exquisite sensitivity and responsiveness of IDPs to posttranslational modifications. The disordered N-terminal transactivation domain (TAD) contains nine known phosphorylation sites which targeted serine, threonine, and proline residues (Fig. 14.5).

14.4.2.2 Cis-Trans Isomerization

Another PTM is the cis-trans isomerization which alters the affinity of p53 for specific partners. p53-peptidyl-prolyl-cis-trans isomerase 1 (PIN1) interaction involves the pro-activation conformational change of p53, thus generating cis-trans isomerization of specific proline residues. PIN1 interacting with Ser33–Pro34,

Ser46–Pro47, Ser127–Pro128, Ser315–Pro316, Thr81–Pro82, and Thr150–Pro151 residues consisting of a phosphorylated ser-pro or the-pro motif, followed by proline isomerization, which changes the conformation of p53. MDM2 pre-binding inhibition or detachment is induced by the modulation of p53 via PIN1 brought about by p53 stabilization. Moreover, p53 acetylation regulated by P300 at C-terminal lysine can be promoted due to the possible conformational change, and consequently it can enhance the binding affinity of p53 core domain for specific promoter-associated sites, particularly apoptosis-promoting sites [101]. PIN1 also regulated the CHK2-Pro82 binding as well as the consequent Ser20 phosphorylation in response to DNA damage. In another sequence of events, the Thr81 phosphorylation allows PIN1 p53 isomerization, which further promotes the Ser20 phosphorylation and p53-CHK2 interaction [102].

14.4.2.3 Acetylation

The role of acetylation has been reported as the regulation of p53-DNA binding at DNA binding site, stabilization by impairing ubiquitination, and p53-mediated transactivation of target genes through the recruitment of coactivators [103, 104]. P300, CBP, and PCAF proteins are involved in p53 acetylation. According to the previous study, phosphorylation of N-terminal sites (specifically Ser15) may promote P300/CBP binding for MDM2 to N-terminal transactivation domain, so that a reduction in MDM2-p53 interaction is related with acetylation of Lys373 and Lys382 [105]. Other p53 residues acetylated by P300/CBP are Lys370, Lys372, and Lys381. P300/CBP-associated factor (PCAF) along with homeodomain-interacting protein kinase 2 (HIPK2) may upregulate the p53 acetylation at site Lys320 upon non-apoptotic DNA damage [106].

14.4.2.4 Ubiquitination

Ubiquitination involves the protein degradation into peptides by ubiquitin-26S proteasome system (or a highly conserved independent ubiquitin protein) in normal cells and shows a high level of target specificity. MDM2 is considered as the significant ubiquitin ligase that regulates the amount of p53 by binding to the N-terminal region and represses p53 activity via two mechanisms: by promoting the export of p53 to the cytoplasm and its consequent degradation and by blocking its transcriptional activation. Another protein which interacts and promotes the export of p53 is a nuclear export signal (NES). Lys370, Lys372, Lys373, Lys381, Lys382, and Lys386 residues located at the C-terminus of p53 are involved in MDM2 ubiquitination [107], resulting in exposure of NES. Moreover, a p53-MDM2 autoregulatory loop is generated in case of MDM2 upregulation.

Other proteins involved in p53 interaction during ubiquitination are the COP1 (constitutive photomorphogenesis protein 1), a RING domain ubiquitin ligase that

negatively regulates p53-dependent transcription [108], the cytosolic chaperone-associated U-box domain ubiquitin ligase CHIP (C-terminus of hsc70-interacting protein) that is involved in degradation of p53 using the proteasome machinery [109], and the cullin-domain ubiquitin ligase CUL4A (cullin 4a) associated with MDM2-p53; PARC (p53-associated parkin-like cytoplasmic protein) is a RING domain ubiquitin ligase considered as a critical regulator in p53 subcellular localization that directly interacts with p53 in the cytoplasm and subsequently restrains its function [110]. The PIRH2 (p53-induced protein, RING-H2-domain-containing) is a RING domain ubiquitin ligase that participates in an autoregulatory feedback loop that controls p53 function as well as promotes MDM2-independent p53 ubiquitination [111].

14.4.2.5 SUMOylating, Methylation, and Neddylation

The p53 residue Lys386 may be sumoylated. SUMO (small ubiquitin-related modifier) proteins covalently bind to their partners via a mechanism similar to ubiquitination. For example, protein inhibitor of activated STAT (PIAS)-1, PIASxa, PIASxb, and PIASy function as SUMO ligases, interact with p53, and affect protein function either by promoting specific alteration (i.e., ubiquitination at the same acceptor site) or by altering subcellular localization [112]. Methylation of p53 is taking place by binding of SET9 (SET domain-containing protein 9) and target Lys372. As a result, it enhances the protein stability in the nucleus [68]. Neddylation is catalyzed by NEDD8 (neuronal precursor cell-expressed developmentally downregulated protein 8), a small ubiquitin-like protein which has been reported as a p53 modulator in an MDM2-dependent manner and inhibits its transcriptional activity [113].

14.5 Future Perspectives

Intrinsically disordered proteins (IDPs) have unique structural and functional peculiarities among the other proteins in cell regulation and recognition and signal transduction. Their functional behavior is modulated via interaction with its protein partner and various PTMs; by this way they can change their binding specificity and affinity in signaling interactions. Among many IDPs, three proteins KID, E1A, and p53 are discussed here, which have a vital role as major hub proteins in cellular protein network. These proteins carry a wide variety of functional regulators of transcription and cell cycle, thereby act as a molecular hub and physically interact with various cellular partner proteins, and harmonize their activities, while many of them act as a molecular hub themselves. Intrinsically disordered proteins participate in one-to-many signaling cascade which confers the induced folding mechanism for the target protein and most probably forms the root for the multifunctional properties of the complex, which are obligatory for defining a cell fate. The IDPs discussed in

this review revealed their role in the allosteric regulation of various cellular signaling processes. The role of IDPs in cellular control mechanism has been explained at various levels. However, the work that has been done so far in this field indisputably signifies only the tip of the iceberg. With the emergent progress about the flexible nature in biological functions of IDPs, we can expect prompt advances toward an understanding of the allosteric mechanisms associated with IDPs which can exert and consequently drive the potential allosteric control throughout the crucial cellar progression. The human proteomes contain plenty of proteins that can be categorized into various sizes (from large complexes to small peptides) and structural types (from fully ordered to partially and fully disordered), which function synergistically to regulate various cellular activities; this might be a valuable strategy to develop novel allosteric inhibitors again challenging drug targets. Extensive research regarding IDP-specific force field had opened new windows in order to further study the allosteric mechanism between IDPs and their partners [114–118]. Further study regarding IDP dynamics and molecular mechanism of allostery will undoubtedly boost our thought about the function of a substantial class of related proteins, driving their critically important role in cellular protein regulation and offering new opportunities for targeted therapeutic intervention.

Acknowledgment This work was supported by the Center for HPC at Shanghai Jiao Tong University, the National Key Research and Development Program of China (2018YFC0310803 and 2017YFE0103300), the National Natural Science Foundation of China (31770771 and 31620103901), and the Medical Engineering Cross Fund of Shanghai Jiao Tong University (YG2015MS56 and YG2017MS08).

References

1. Changeux JP (1961) The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. Cold Spring Harb Symp Quant Biol 26:313–318
2. Monod J, Jacob F (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harb Symp Quant Biol 26:389–401
3. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
4. Freiburger LA, Baettig OM, Sprules T, Berghuis AM, Auclair K, Mittermaier AK (2011) Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme. Nat Struct Mol Biol 18(3):288–294
5. Nussinov R, Tsai CJ, Ma B (2013) The underappreciated role of allostery in the cellular network. Annu Rev Biophys 42:169–189
6. Koshland DE Jr, Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5(1):365–385
7. Hilser VJ, Wrabl JO, Motlagh HN (2012) Structural and energetic basis of allostery. Annu Rev Biophys 41:585–609
8. Nussinov R, Tsai CJ (2015) Allostery without a conformational change? Revisiting the paradigm. Curr Opin Struct Biol 30:17–24
9. May LT, Leach K, Sexton PM, Christopoulos A (2007) Allosteric modulation of G protein-coupled receptors. Annu Rev Pharmacol Toxicol 47:1–51

10. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM (2014) Classification of intrinsically disordered regions and proteins. Chem Rev 114(13):6589–6631
11. Csizmok V, Follis AV, Kriwacki RW, Forman-Kay JD (2016) Dynamic protein interaction networks and new structural paradigms in signaling. Chem Rev 116(11):6424–6462
12. Bah A, Forman-Kay JD (2016) Modulation of intrinsically disordered protein function by post-translational modifications. J Biol Chem R115:695056
13. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. PLoS Comput Biol 2(8):e100
14. Wright PE, Dyson HJ (2014) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18
15. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK (2002) Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323(3):573–584
16. Flock T, Weatheritt RJ, Latysheva NS, Babu MM (2014) Controlling entropy to tune the functions of intrinsically disordered regions. Curr Opin Struct Biol 26:62–72
17. Tompa P (2013) Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. Chem Rev 114(13):6715–6732
18. Gandhi TK, Zhong J, Mathivanan S, Karthick L, Chandrika KN, Mohan SS, Sharma S, Pinkert S, Nagaraju S, Periaswamy B, Mishra G, Nandakumar K, Shen B, Deshpande N, Nayak R, Sarker M, Boeke JD, Parmigiani G, Schultz J, Bader JS, Pandey A (2006) Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. Nat Genet 38(3):285–293
19. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksoz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE (2005) A human protein-protein interaction network: a resource for annotating the proteome. Cell 122(6):957–968
20. Fraser HB, Hirsh AE, Steinmetz LM, Scharfe C, Feldman MW (2002) Evolutionary rate in the protein interaction network. Science (New York, NY) 296(5568):750–752
21. Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411(6833):41–42
22. Nevins JR, Ginsberg HS, Blanchard JM, Wilson MC, Darnell JE Jr (1979) Regulation of the primary expression of the early adenovirus transcription units. J Virol 32(3):727–733
23. Stephens C, Harlow E (1987) Differential splicing yields novel adenovirus 5 E1A mRNAs that encode 30 kd and 35 kd proteins. EMBO J 6(7):2027–2035
24. Ulfendahl PJ, Linder S, Kreivi JP, Nordqvist K, Sevensson C, Hultberg H, Akusjarvi G (1987) A novel adenovirus-2 E1A mRNA encoding a protein with transcription activation properties. EMBO J 6(7):2037–2044
25. Culp JS, Webster LC, Friedman DJ, Smith CL, Huang WJ, Wu FY, Rosenberg M, Ricciardi RP (1988) The 289-amino acid E1A protein of adenovirus binds zinc in a region that is important for trans-activation. Proc Natl Acad Sci U S A 85(17):6450–6454
26. Kimelman D, Miller JS, Porter D, Roberts BE (1985) E1a regions of the human adenoviruses and of the highly oncogenic simian adenovirus 7 are closely related. J Virol 53(2):399–409
27. van Ormondt H, Maat J, Dijkema R (1980) Comparison of nucleotide sequences of the early E1a regions for subgroups A, B and C of human adenoviruses. Gene 12(1–2):63–76
28. Lyons RH, Ferguson BQ, Rosenberg M (1987) Pentapeptide nuclear localization signal in adenovirus E1a. Mol Cell Biol 7(7):2451–2456
29. Krippl B, Ferguson B, Rosenberg M, Westphal H (1984) Functions of purified E1A protein microinjected into mammalian cells. Proc Natl Acad Sci 81(22):6988–6992
30. Harlow E, Whyte P, Franza BR Jr, Schley C (1986) Association of adenovirus early-region 1A proteins with cellular polypeptides. Mol Cell Biol 6(5):1579–1589

31. Yee SP, Branton PE (1985) Detection of cellular proteins associated with human adenovirus type 5 early region 1A polypeptides. Virology 147(1):142–153
32. Blobel GA (2000) CREB-binding protein and p300: molecular integrators of hematopoietic transcription. Blood 95(3):745–755
33. Giordano A, Avantaggiati ML (1999) p300 and CBP: partners for life and death. J Cell Physiol 181(2):218–230
34. Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6(3):197–208
35. Dyson HJ, Wright PE (2016) Role of intrinsic protein disorder in the function and interactions of the transcriptional coactivators CREB-binding protein (CBP) and p300. J Biol Chem 291(13):6714–6722
36. Egan C, Bayley ST, Branton PE (1989) Binding of the Rb1 protein to E1A products is required for adenovirus transformation. Oncogene 4(3):383–388
37. Whyte P, Buchkovich KJ, Horowitz JM, Friend SH, Raybuck M, Weinberg RA, Harlow E (1988) Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. Nature 334(6178):124–129
38. Chan HM, Krstic-Demonacos M, Smith L, Demonacos C, La Thangue NB (2001) Acetylation control of the retinoblastoma tumour-suppressor protein. Nat Cell Biol 3(7):667–674
39. Subramanian T, Zhao LJ, Chinnadurai G (2013) Interaction of CtBP with adenovirus E1A suppresses immortalization of primary epithelial cells and enhances virus replication during productive infection. Virology 443(2):313–320
40. Tremblay ML, Dumont DJ, Branton PE (1989) Analysis of phosphorylation sites in the exon 1 region of E1A proteins of human adenovirus type 5. Virology 169(2):397–407
41. Hateboer G, Gennissen A, Ramos YF, Kerkhoven RM, Sonntag-Buck V, Stunnenberg HG, Bernards R (1995) BS69, a novel adenovirus E1A-associated protein that inhibits E1A transactivation. EMBO J 14(13):3159–3169
42. Pines J, Hunter T (1990) Human cyclin A is adenovirus E1A-associated protein p60 and behaves differently from cyclin B. Nature 346(6286):760–763
43. Ewen ME, Xing YG, Lawrence JB, Livingston DM (1991) Molecular cloning, chromosomal mapping, and expression of the cDNA for p107, a retinoblastoma gene product-related protein. Cell 66(6):1155–1164
44. Tsai LH, Harlow E, Meyerson M (1991) Isolation of the human cdk2 gene that encodes the cyclin A- and adenovirus E1A-associated p33 kinase. Nature 353(6340):174–177
45. Hannon GJ, Demetrick D, Beach D (1993) Isolation of the Rb-related p130 through its interaction with CDK2 and cyclins. Genes Dev 7(12a):2378–2391
46. Li Y, Graham C, Lacy S, Duncan AM, Whyte P (1993) The adenovirus E1A-associated 130-kD protein is encoded by a member of the retinoblastoma gene family and physically interacts with cyclins A and E. Genes Dev 7(12a):2366–2377
47. Cohen MJ, Yousef AF, Massimi P, Fonseca GJ, Todorovic B, Pelka P, Turnell AS, Banks L, Mymryk JS (2013) Dissection of the C-terminal region of E1A redefines the roles of CtBP and other cellular targets in oncogenic transformation. J Virol 87(18):10348–10355
48. Geisberg JV, Lee WS, Berk AJ, Ricciardi RP (1994) The zinc finger region of the adenovirus E1A transactivating domain complexes with the TATA box binding protein. Proc Natl Acad Sci U S A 91(7):2488–2492
49. Hateboer G, Timmers HT, Rustgi AK, Billaud M, van't Veer LJ, Bernards R (1993) TATA-binding protein and the retinoblastoma gene product bind to overlapping epitopes on c-Myc and adenovirus E1A protein. Proc Natl Acad Sci U S A 90(18):8489–8493
50. Song CZ, Loewenstein PM, Toth K, Green M (1995) Transcription factor TFIID is a direct functional target of the adenovirus E1A transcription-repression domain. Proc Natl Acad Sci U S A 92(22):10330–10333
51. Boyer TG, Berk AJ (1993) Functional interaction of adenovirus E1A with holo-TFIID. Genes Dev 7(9):1810–1823

52. Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. Genes Dev 14(13):1553–1577
53. Thakur JK, Yadav A, Yadav G (2013) Molecular recognition by the KIX domain and its role in gene regulation. Nucleic Acids Res 42(4):2112–2125
54. Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE (2002) Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP) The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. J Biol Chem 277(45):43168–43174
55. Radhakrishnan I, Pérez-Alvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator: coactivator interactions. Cell 91(6):741–752
56. De Guzman RN, Goto NK, Dyson HJ, Wright PE (2006) Structural basis for cooperative transcription factor binding to the CBP coactivator. J Mol Biol 355(5):1005–1013
57. Ernst P, Wang J, Huang M, Goodman RH, Korsmeyer SJ (2001) MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein. Mol Cell Biol 21(7):2249–2258
58. Kim PM, Sboner A, Xia Y, Gerstein M (2008) The role of disorder in interaction networks: a structural analysis. Mol Syst Biol 4:179
59. Ferrari R, Pellegrini M, Horwitz GA, Xie W, Berk AJ, Kurdistani SK (2008) Epigenetic reprogramming by adenovirus e1a. Science 321(5892):1086–1088
60. White E (2001) Regulation of the cell cycle and apoptosis by the oncogenes of adenovirus. Oncogene 20(54):7836
61. Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE (2009b) Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. Proc Natl Acad Sci U S A 106(32):13260–13265
62. Ferreon ACM, Ferreon JC, Wright PE, Deniz AA (2013) Modulation of allostery by protein intrinsic disorder. Nature 498(7454):390
63. Anderson CW, Appella E (2009) Signaling to the p53 tumor suppressor through pathways activated by genotoxic and non-genotoxic stresses. In: Handbook of cell signaling, 2nd edn. Elsevier, pp 2185–2204
64. Hollstein M, Sidransky D, Vogelstein B, Harris CC (1991) p53 mutations in human cancers. Science 253(5015):49–53
65. Avalos JL, Celic I, Muhammad S, Cosgrove MS, Boeke JD, Wolberger C (2002) Structure of a Sir2 enzyme bound to an acetylated p53 peptide. Mol Cell 10(3):523–535
66. Bochkareva E, Kaustov L, Ayed A, Yi G-S, Lu Y, Pineda-Lucena A, Liao JC, Okorokov AL, Milner J, Arrowsmith CH (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. Proc Natl Acad Sci U S A 102(43):15412–15417
67. Cho Y, Gorina S, Jeffrey PD, Pavletich NP (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. Science 265(5170):346–355
68. Chuikov S, Kurash JK, Wilson JR, Xiao B, Justin N, Ivanov GS, McKinney K, Tempst P, Prives C, Gamblin SJ (2004) Regulation of p53 activity through lysine methylation. Nature 432(7015):353
69. Di Lello P, Jenkins LMM, Jones TN, Nguyen BD, Hara T, Yamaguchi H, Dikeakos JD, Appella E, Legault P, Omichinski JG (2006) Structure of the Tfb1/p53 complex: Insights into the interaction between the p62/Tfb1 subunit of TFIIH and the activation domain of p53. Mol Cell 22(6):731–740
70. Gorina S, Pavletich NP (1996) Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. Science 274(5289):1001–1005
71. Joo WS, Jeffrey PD, Cantor SB, Finnin MS, Livingston DM, Pavletich NP (2002) Structure of the 53BP1 BRCT region bound to p53 and its comparison to the Brca1 BRCT structure. Genes Dev 16(5):583–593

72. Kussie PH, Gorina S, Marechal V, Elenbaas B, Moreau J, Levine AJ, Pavletich NP (1996) Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 274(5289):948–953
73. Kuszewski J, Gronenborn AM, Clore GM (1999) Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. J Am Chem Soc 121(10):2337–2338
74. Lilyestrom W, Klein MG, Zhang R, Joachimiak A, Chen XS (2006) Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. Genes Dev 20(17):2373–2382
75. Lowe ED, Tews I, Cheng KY, Brown NR, Gul S, Noble ME, Gamblin SJ, Johnson LN (2002) Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. Biochemistry 41(52):15625–15634
76. Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, Sanchez R, Zeleznik-Le NJ, Ze R, Zhou M-M (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. Mol Cell 13(2):251–263
77. Obsil T, Ghirlando R, Klein DC, Ganguly S, Dyda F (2001) Crystal structure of the 14-3-3ζ: serotonin N-acetyltransferase complex: a role for scaffolding in enzyme regulation. Cell 105(2):257–267
78. Petosa C, Masters SC, Bankston LA, Pohl J, Wang B, Fu H, Liddington RC (1998) 14-3-3ζ binds a phosphorylated Raf peptide and an unphosphorylated peptide via its conserved amphipathic groove. J Biol Chem 273(26):16305–16310
79. Poux AN, Marmorstein R (2003) Molecular basis for Gcn5/PCAF histone acetyltransferase selectivity for histone and nonhistone substrates. Biochemistry 42(49):14366–14374
80. Rittinger K, Budman J, Xu J, Volinia S, Cantley LC, Smerdon SJ, Gamblin SJ, Yaffe MB (1999) Structural analysis of 14-3-3 phosphopeptide complexes identifies a dual role for the nuclear export signal of 14-3-3 in ligand binding. Mol Cell 4(2):153–166
81. Wu H, Maciejewski MW, Marintchev A, Benashski SE, Mullen GP, King SM (2000) Solution structure of a dynein motor domain associated light chain. Nat Struct Mol Biol 7(7):575
82. Dawson R, Müller L, Dehner A, Klein C, Kessler H, Buchner J (2003) The N-terminal domain of p53 is natively unfolded. J Mol Biol 332(5):1131–1141
83. Lee H, Mok KH, Muhandiram R, Park K-H, Suk J-E, Kim D-H, Chang J, Sung YC, Choi KY, Han K-H (2000) Local structural elements in the mostly unstructured transcriptional activation domain of human p53. J Biol Chem 275(38):29426–29432
84. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN (2005) Flexible nets. FEBS J 272(20):5129–5148
85. Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE (1998) Protein disorder and the evolution of molecular recognition: theory, predictions and observations. In: Pacific symposium on biocomputing, pp 473–484
86. Uversky VN, Oldfield CJ, Dunker AK (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 18(5):343–384
87. Ferreon JC, Lee CW, Arai M, Martinez-Yamout MA, Dyson HJ, Wright PE (2009a) Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2. Proc Natl Acad Sci U S A 106(16):6591–6596
88. Teufel DP, Freund SM, Bycroft M, Fersht AR (2007) Four domains of p300 each bind tightly to a sequence spanning both transactivation subdomains of p53. Proc Natl Acad Sci U S A 104(17):7009–7014
89. Feng H, Jenkins LMM, Durell SR, Hayashi R, Mazur SJ, Cherry S, Tropea JE, Miller M, Wlodawer A, Appella E (2009) Structural basis for p300 Taz2-p53 TAD1 binding and modulation by phosphorylation. Structure 17(2):202–210
90. Miller Jenkins LM, Feng H, Durell SR, Tagad HD, Mazur SJ, Tropea JE, Bai Y, Appella E (2015) Characterization of the p300 Taz2–p53 TAD2 complex and comparison with the p300 Taz2–p53 TAD1 complex. Biochemistry 54(11):2001–2010

91. Popowicz G, Czarna A, Holak T (2008) Structure of the human Mdmx protein bound to the p53 tumor suppressor transactivation domain. Cell Cycle 7(15):2441–2443
92. Vise PD, Baral B, Latos AJ, Daughdrill GW (2005) NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. Nucleic Acids Res 33(7):2061–2077
93. Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. Proc Natl Acad Sci U S A 105(15):5762–5767
94. Borcherds W, Theillet F-X, Katzer A, Finzel A, Mishall KM, Powell AT, Wu H, Manieri W, Dieterich C, Selenko P (2014) Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. Nat Chem Biol 10(12):1000
95. Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, Radivojac P (2014) The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci 23(8):1077–1093
96. Gao J, Xu D (2012) Correlation between posttranslational modification and intrinsic disorder in protein. In: Biocomputing 2012. World Scientific, pp 94–103
97. Theillet F-X, Smet-Nocca C, Liokatis S, Thongwichian R, Kosten J, Yoon M-K, Kriwacki RW, Landrieu I, Lippens G, Selenko P (2012) Cell signaling, post-translational protein modifications and NMR spectroscopy. J Biomol NMR 54(3):217–236
98. Tsai C-J, Ma B, Nussinov R (2009) Protein–protein interaction networks: how can a hub protein bind so many different partners? Trends Biochem Sci 34(12):594–600
99. Lee CW, Ferreon JC, Ferreon ACM, Arai M, Wright PE (2010) Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. Proc Natl Acad Sci U S A 107(45):19290–19295
100. Schon O, Friedler A, Bycroft M, Freund SM, Fersht AR (2002) Molecular mechanism of the interaction between MDM2 and p53. J Mol Biol 323(3):491–501
101. Kohn KW, Pommier Y (2005) Molecular interaction map of the p53 and Mdm2 logic elements, which control the Off–On switch of p53 in response to DNA damage. Biochem Biophys Res Commun 331(3):816–827
102. Berger M, Stahl N, Del Sal G, Haupt Y (2005) Mutations in proline 82 of p53 impair its activation by Pin1 and Chk2 in response to DNA damage. Mol Cell Biol 25(13):5380–5388
103. Brooks CL, Gu W (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol 15(2):164–171
104. Kawaguchi Y, Ito A, Appella E, Yao T-P (2006) Charge modification at multiple C-terminal lysine residues regulates p53 oligomerization and its nucleus-cytoplasm trafficking. J Biol Chem 281(3):1394–1400
105. Appella E, Anderson CW (2001) Post-translational modifications and activation of p53 by genotoxic stresses. Eur J Biochem 268(10):2764–2772
106. Di Stefano V, Soddu S, Sacchi A, D'orazi G (2005) HIPK2 contributes to PCAF-mediated p53 acetylation and selective transactivation of p21 Waf1 after nonapoptotic DNA damage. Oncogene 24(35):5431
107. Rodriguez MS, Desterro JM, Lain S, Lane DP, Hay RT (2000) Multiple C-terminal lysine residues target p53 for ubiquitin-proteasome-mediated degradation. Mol Cell Biol 20(22):8458–8467
108. Dornan D, Bheddah S, Newton K, Ince W, Frantz GD, Dowd P, Koeppen H, Dixit VM, French DM (2004) COP1, the negative regulator of p53, is overexpressed in breast and ovarian adenocarcinomas. Cancer Res 64(20):7226–7230
109. Esser C, Scheffner M, Höhfeld J (2005) The chaperone associated ubiquitin ligase CHIP is able to target p53 for proteasomal degradation. J Biol Chem
110. Nikolaev AY, Li M, Puskas N, Qin J, Gu W (2003) Parc: a cytoplasmic anchor for p53. Cell 112(1):29–40

111. Leng RP, Lin Y, Ma W, Wu H, Lemmers B, Chung S, Parant JM, Lozano G, Hakem R, Benchimol S (2003) Pirh2, a p53-induced ubiquitin-protein ligase, promotes p53 degradation. Cell 112(6):779–791
112. Schmidt D, Müller S (2002) Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity. Proc Natl Acad Sci U S A 99(5):2872–2877
113. Xirodimas DP, Saville MK, Bourdon J-C, Hay RT, Lane DP (2004) Mdm2-mediated NEDD8 conjugation of p53 inhibits its transcriptional activity. Cell 118(1):83–97
114. Liu H, Song D, Lu H, Luo R, Chen HF (2018) Intrinsically disordered protein specific force field CHARMM 36 IDPSFF. Chem Biol Drug Des 92(4):1722–1735
115. Song D, Luo R, Chen H-F (2017a) The IDP-specific force field ff14IDPSFF improves the conformer sampling of intrinsically disordered proteins. J Chem Inf Model 57(5):1166–1178
116. Song D, Wang W, Ye W, Ji D, Luo R, Chen HF (2017b) ff14IDPs force field improving the conformation sampling of intrinsically disordered proteins. Chem Biol Drug Des 89(1):5–15
117. Wang W, Ye W, Jiang C, Luo R, Chen HF (2014) New force field on modeling intrinsically disordered proteins. Chem Biol Drug Des 84(3):253–269
118. Ye W, Ji D, Wang W, Luo R, Chen H-F (2015) Test and evaluation of ff99IDPs force field for intrinsically disordered proteins. J Chem Inf Model 55(5):1021–1029

Chapter 15
Engineering Allostery into Proteins

Scott D. Gorman, Rebecca N. D'Amico, Dennis S. Winston, and David D. Boehr

Abstract Our ability to engineer protein structure and function has grown dramatically over recent years. Perhaps the next level in protein design is to develop proteins whose function can be regulated in response to various stimuli, including ligand binding, pH changes, and light. Endeavors toward these goals have tested and expanded on our understanding of protein function and allosteric regulation. In this chapter, we provide examples from different methods for developing new allosterically regulated proteins. These methods range from whole insertion of regulatory domains into new host proteins, to covalent attachment of photoswitches to generate light-responsive proteins, and to targeted changes to specific amino acid residues, especially to residues identified to be important for relaying allosteric information across the protein framework. Many of the examples we discuss have already found practical use in medical and biotechnology applications.

Keywords Allostery · Protein regulation · Protein engineering · Energy landscape · Amino acid network · Domain insertion · Covalent modification

15.1 Introduction

Allostery in biomacromolecules can be broadly defined as "action at a distance," in which a perturbation at one site (e.g., ligand binding, amino acid change) has an effect on the structure or function of some other distant site. The ability to allosterically regulate biomacromolecules is essential for life itself, with examples in DNA replication [57, 116], gene regulation [21, 56, 66, 119, 129], and biosynthetic pathways [2, 38, 58]. There is renewed interest in the design of therapeutics that target allosteric sites in proteins involved in disease [18, 60, 64, 100, 106, 121, 132]. There have also been significant advances in the engineering of novel allosteric

S. D. Gorman · R. N. D'Amico · D. S. Winston · D. D. Boehr (✉)
Department of Chemistry, The Pennsylvania State University, University Park, PA, USA
e-mail: ddb12@psu.edu

J. Zhang, R. Nussinov (eds.), *Protein Allostery in Drug Discovery*,
Advances in Experimental Medicine and Biology 1163,
https://doi.org/10.1007/978-981-13-8719-7_15

regulation mechanisms into proteins. Such proteins are of substantial interest in biotechnology and medical fields, including in the production of biofuels and pharmaceuticals [80, 102, 103].

We have only just begun to understand allostery, this "second secret of life" [53, 136]. While some details differ among allosteric proteins, it is often useful to develop more general models to both understand and then even design new allosteric proteins. Towards this end, we first briefly outline two of the most widely accepted modern frameworks for allostery, the free energy landscape and network models, and then present examples that leverage these models toward engineering allostery into proteins. It should be noted that our examples are representative of recent work, but are not exhaustive.

15.1.1 The Energy Landscape Model

The best-known early models for describing allostery are the Monod-Wyman-Changeux (MWC) and Koshland-Nemethy-Filmer (KNF) models, which focused on conformational changes in multimeric complexes [73]. While these models have proven extremely useful for quantifying conformational states for large complexes such as hemoglobin, they do not take into account non-rigid body dynamics. For example, there are now well-known examples of allosteric proteins that do not undergo substantial structural changes upon ligand binding [88, 94, 97], which could thus not be described by either the MWC or KNF type models. The free energy landscape model [25] can be viewed as a broader generalization of these early models, which can then also describe allostery in proteins that lack large conformational changes, and also expand our definition of allostery to include the effects of amino acid changes or changes to other physiochemical parameters (e.g., temperature, pressure). Within the free energy landscape model, proteins can fluctuate among many unique substates populated according to their relative free energies. Low-amplitude, high-frequency motions such as a methyl rotation or a change in a side chain dihedral angle are represented as small energy barriers between substates, while larger barriers represent lower-frequency, higher-amplitude motions such as exchange between major and minor conformational states. Molecular dynamics (MD) techniques such as Markov state models [93] and biophysical techniques such as nuclear magnetic resonance (NMR) spin relaxation [19] and relaxation dispersion [3], fluorescence resonance energy transfer (FRET) [78], and isothermal titration calorimetry (ITC) [92] are commonly used together to better understand these energy landscapes.

The free energy landscape of a protein can be highly malleable in that it changes in response to stimuli [76] such as ligand binding, mutagenesis, or even pressure [83, 139]. Efforts to engineer allostery in the context of the energy landscape model therefore focus on ways to restrict or increase conformational exchange under certain conditions, or to stabilize certain minor states correlated with protein function (Fig. 15.1a).

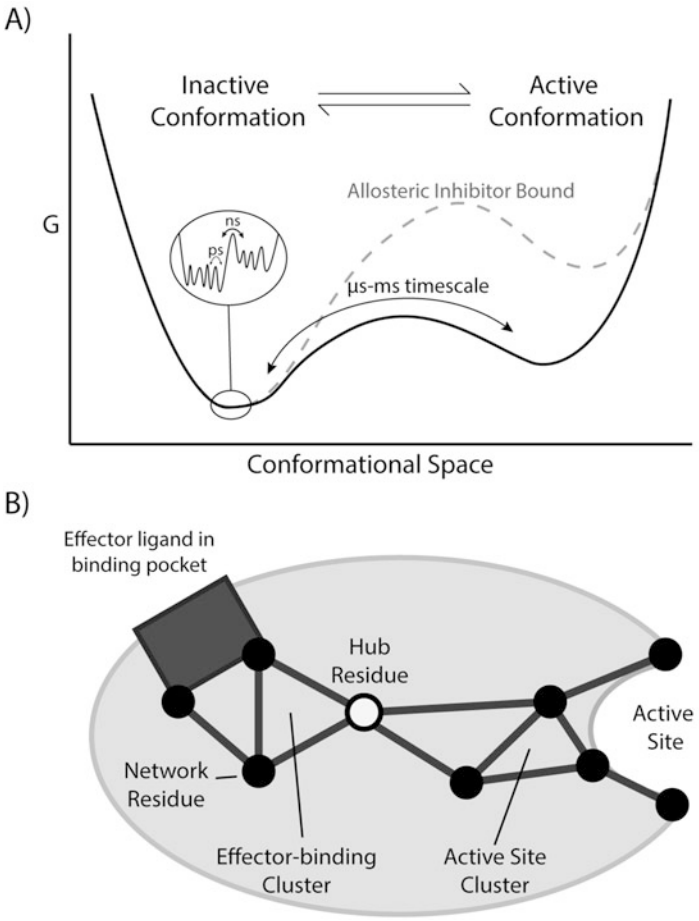


Fig. 15.1 Frameworks for understanding allostery. (**a**) In the free energy landscape model, proteins are considered to be ensembles of different conformations. Local minima represent different conformations, with their free energies determining relative populations, and free energy barriers determining exchange rates between conformations. Allosteric inhibitors and activators can change the topology of the energy landscape by altering populations or the kinetics of exchange between conformations. (**b**) The network model places emphasis on correlated structural dynamics of amino acid residues being responsible for allosteric effects. A perturbation at one site can have an effect on a surrounding cluster of amino acids, which then can propagate through the amino acid residue interaction network to effect changes at a distant active site or other binding site. The network signals travel through highly connected hub residues, which may be important engineering points to change protein function and regulation

15.1.2 The Network Model

Another way to view a protein is to focus on the atomic-level noncovalent interactions that hold together its three-dimensional structure [89]. Within an amino acid

interaction network, the nodes are usually full amino acid residues, and the edges contain information on which residues are interacting and the relative strength of those interactions (Fig. 15.1b). Residues with a large number of interactions are often considered to be important "hub" residues, and groups of interacting residues are sorted into cliques, communities, or clusters depending on the connectivity within the group [123]. Networks can be identified through a variety of techniques including bioinformatics analyses of multiple sequence alignments to identify evolutionarily coupled residues as performed by the statistical coupling analysis [63] (SCA) and direct coupling analysis [75] (DCA) algorithms. MD cross-correlation analysis [29] can identify energetically and dynamically coupled residues. NMR spectroscopy also offers a number of physics-based methods to identify networks, including the Chemical Shift Covariance Analysis (CHESCA) [11] and RASSMM [40] algorithms.

Ligand binding to an allosteric protein at one site may transmit a series of structural changes, and changes to the noncovalent interactions lead to structural and functional changes at a distant site. Efforts to engineer allostery in the context of the network model focus on changing network residues by mutagenesis or covalent modification to allosterically "tune" an enzyme's activity or even give it new functionality by modifying these network connections [1, 6, 7].

In our view, the network model is compatible with the free energy landscape model; the network model focuses on paths and webs of physical interactions within the protein, while the free energy landscape attempts to assign these interactions to various conformational substates. Traversing over the free energy landscape from one conformational substate to another necessitates the breakage and formation of noncovalent interactions that might be identified as part of an important amino acid interaction network.

15.2 Engineering Protein Allostery Through Domain Insertion

Under favorable circumstances, new allosteric proteins can be created by inserting a regulatory domain from a known allosteric protein into a new host protein (Fig. 15.2 and 15.3). Upon allosteric effector binding, the regulatory domain undergoes a large conformational change, leading to substantial structural and functional changes in the host protein. The regulatory domain often changes the free energy landscape of the host protein by introducing new steric interactions between the regulatory domain and host protein. The goals of this section are to present select examples that highlight the different approaches for creating these chimeric proteins.

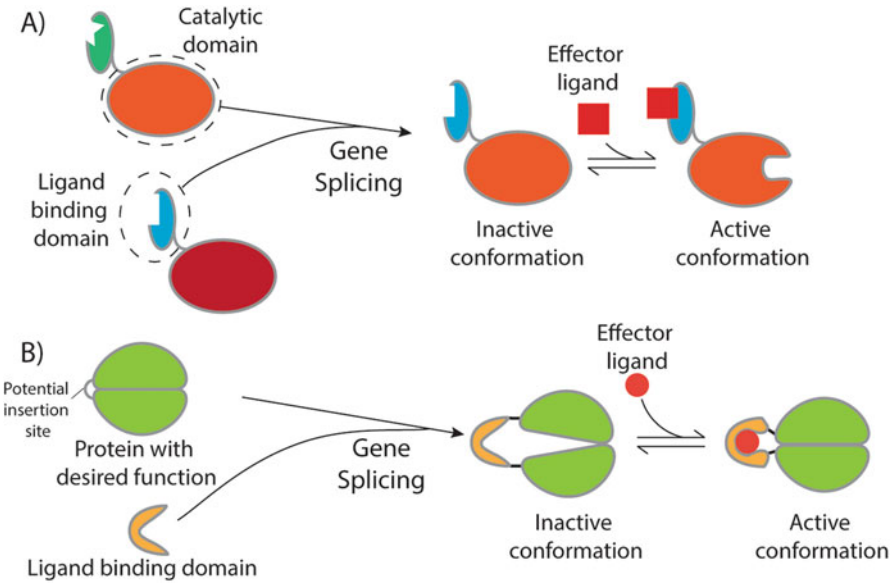


Fig. 15.2 Methods of rational domain insertion. (**a**) Homologous proteins with catalytic domains are often structurally similar enough that regulatory domains sensitive to different effectors can be swapped or added to homologs without a regulatory domain, as in the above example where the blue regulatory domain is spliced onto the orange catalytic domain. (**b**) In cases where the protein with the desired function does not have any allosteric homologs, a conditionally disordered linker that becomes ordered in response to environmental changes or a regulatory domain can be inserted into the middle of a loop in the host protein sequence. In this example, the green enzymatic domain is inactive when the yellow regulatory domain is in its open apo conformation. Once the regulatory domain binds its effector ligand, it adopts a conformation that allows the enzymatic domain to adopt its active conformation

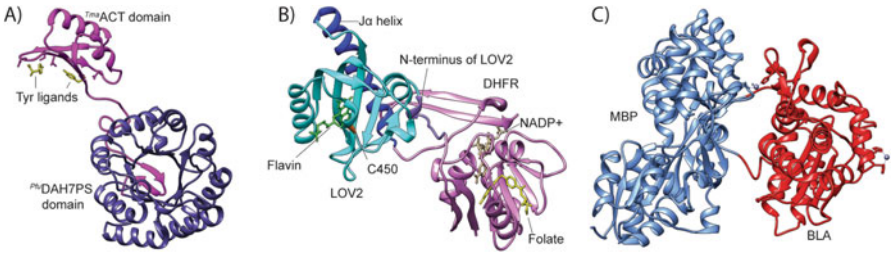


Fig. 15.3 Examples of allosteric enzymes engineered through domain insertion. (**A**) *Pfu*ACT*Tma*DAH7PS with Tyr (yellow) bound to the ACT domain (magenta) with the DAH7PS domain in slate (PDB: 4GRS) [16]. *Pfu*ACT*Tma*DAH7PS is typically depicted as a homotetramer. It should be noted that only one subunit is shown here for clarity. (**B**) DHFR-LOV2 with LOV2 domain inserted at residue 120 of DHFR (PDBs: 1RX2, 2V0U) [59]. Upon irradiation with light, C450 covalently bonds to the flavin cofactor in the LOV2 domain. The changes in conformation and dynamics propagate through both domains. (**C**) The MBP-BLA construct RG13 generated by random domain insertion of BLA (red) into MBP (blue) (PDB: 4DXB) [48]

15.2.1 Rational Insertion of Regulatory Domains

The Parker lab has explored how allostery has evolved in nature, specifically in the amino acid biosynthesis pathways of various bacteria [4, 5, 16, 23, 43, 47, 58, 126]. Their method has primarily consisted of splicing ACT or CM regulatory domains from one $(\beta/\alpha)_8$ barrel enzyme onto another, similar to the schematic in Fig. 15.2a. The ACT domain is named for aspartate kinase, (AroH) chorismate mutase, and TyrA (i.e., TyrA is the bifunctional chorismate mutase/prephenate dehydratase in the Tyr biosynthetic pathway), all of which contain this domain. ACT domains consist of a βαββαβ fold, and typically occur when in contact with a second ACT domain to form a binding pocket for one or more small molecule effector ligands [58]. Ligand binding triggers a conformational change or a change in the oligomeric state of the enzyme that in turn modulates function of the catalytic domain. The AroQ chorismate mutase effector binding (CM) domain, in contrast, is an all-α-helical bundle that can turnover chorismic acid to prephenic acid, and binds prephenic acid when used as a regulatory binding domain [23]. Upon binding prephenic acid, the CM domain also undergoes a large conformational change. Cross et al. [16] spliced an ACT domain from *Thermotoga maritima* 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAH7PS) onto the catalytic $(\beta/\alpha)_8$ barrel domain of the unregulated DAH7PS from *Pyrococcus furiosus* via a flexible linker on the N-terminus of *P. furiosus* DAH7PS (Fig. 15.3a). Surprisingly, the new chimeric protein was both catalytically active and allosterically inhibited by Tyr in a manner similar to that of *T. maritima* DAH7PS without any further modification. In a later publication, Fan et al. [23] swapped the regulatory domains of DAH7PS from *T. maritima* and *Geobacillus* sp. (^{Tma}DAH7PS and ^{Gsp}DAH7PS, respectively). ^{Tma}DAH7PS and ^{Gsp}DAH7PS differ structurally, primarily in that ^{Tma}DAH7PS has an ACT regulatory domain and ^{Gsp}DAH7PS has an AroQ CM regulatory domain. In ^{Gsp}DAH7PS, the CM reduces DAH7PS activity in response to binding prephenic acid, and when transplanted onto the ^{Tma}DAH7PS $(\beta/\alpha)_8$ catalytic domain conferred that regulation. Similarly, the fusion of the ^{Tma}DAH7PS Tyr-binding ACT domain onto the $(\beta/\alpha)_8$ catalytic domain of ^{Gsp}DAH7PS conferred inhibition upon Tyr binding. These studies (and others[48]) were remarkable in how readily new allosteric mechanisms could be conferred onto these proteins, suggesting that the ability to allosterically regulate the catalytic domains may be intrinsic to these proteins, and perhaps to these types of protein folds.

For the creation of new allosteric chimeric proteins, it is common to insert the regulatory domain into the middle of the catalytic domain (Fig. 15.2b). Some of these new allosteric proteins have inserted an intrinsically disordered protein (IDP) domain that undergoes a transition between a high entropy disordered state to a low entropy ordered state, depending on the ligand-bound state, temperature, and/or pH. For example, Choi et al. [14] designed a maltose-dependent β-lactamase-maltose binding protein (BLA-MBP) chimeric protein in which the BLA domain was inserted in place of residues 317 and 318 of the MBP domain, and the amino acid sequence Asp-Lys-Thr was placed between the N-terminus of BLA and residue

319 of MBP. The construct, known as C4, was not allosterically regulated but did have BLA activity. Nonetheless, by changing the linker, they were able to develop allosterically regulated variants. For example, they added varying amounts of Gly residues to the linker resulting in temperature-dependent allosteric regulation, which was attributed to differences to entropic contributions with and without maltose. In another variant, a varying repeat Glu-Ala-Gln-Ala linker was substituted for Asp-Lys-Thr. Since Glu-Ala-Gln-Ala is more α-helical at low pH and more random coil at high pH, the proposal was to create variants that had a pH-dependent sensitivity to maltose. Indeed, experimental evidence showed that higher linker flexibility at high pH or high temperature resulted in greater conformational heterogeneity in the BLA domain, resulting in a higher population of less-active states until maltose was bound. Maltose binding under high pH or high temperature conditions reduced conformational entropy but increased the population of the active conformations, resulting in greater β-lactamase activity. The constructs lacked allosteric regulation when their respective linkers were α-helical at low pH or under low temperature conditions. These studies indicated that proteins can be engineered to respond simultaneously to ligand binding and changes to physiochemical parameters (i.e., pH and temperature), providing additional levels of control over protein function.

In a similar example, Meister et al. [71] took advantage of a previously identified insertion site in BLA [31–33] and inserted calmodulin (CaM) in such a way that BLA was split into two halves, BLA 24–194 and BLA 196–286, in a manner analogous to Fig. 15.2b. The resulting chimeric protein had β-lactamase activity in the absence of Ca^{2+} because the central helix of CaM was disordered under those conditions, allowing the two halves of the BLA domain to come together. Upon addition of Ca^{2+}, the helix became rigid, which separated the two half BLA domains and abrogated catalytic activity. In this case, the high-entropy state was more active because it is likely thermodynamically favorable for the hydrophobic core of BLA to come back together and exclude solvent from those nonpolar residues. These sets of studies indicated that domain insertion and thoughtful linker engineering can lead to rational enthalpy and entropy changes to the underlying free energy landscapes of these proteins.

In a rather unique example, Guo et al. [34] created an allosteric Ca^{2+} electrochemical sensor. A calmodulin domain was inserted into the redox enzyme PQQ-glucose dehydrogenase with the goal of creating a biosensor that could quantify Ca^{2+} concentrations in biological fluids. PQQ-glucose dehydrogenase had been used as an electrochemical biosensor for glucose, but Guo et al. attempted to further engineer this enzyme through domain insertion of a calmodulin domain. The loop connecting strands A and B of β-sheet 3 of PQQ-glucose dehydrogenase was chosen as the insertion site since strand A is part of the catalytic domain involved in abstracting a proton from the glucose O1 atom. The thought was that the large conformational change in the CaM domain upon binding Ca^{2+} would result in a structural change that would activate the redox reaction catalyzed by PQQ-glucose dehydrogenase, resulting in an electrical impulse. Indeed, the chimeric protein could detect dilute Ca^{2+} in saliva. In a later refinement, the chimeric protein was placed in

direct contact with conducting electrodes and a semiconductor by conjugating it to a graphene nanosheet associating ligand, PBSE [54].

In the above examples, new allosteric proteins were created that responded to ligand binding. There are also several examples of newly designed light-controlled proteins. The LOV2 domain from *Avena sativa* has been used to control the function of a variety of proteins via photoswitchable allosteric regulation [15, 26, 27, 39, 59, 65, 72, 74, 77, 84, 86, 105, 110, 114, 117, 125, 127, 130, 133, 137, 138]. A conformational change occurs upon absorption of blue light due to a conserved cysteine in the LOV2 domain forming a covalent bond with the flavin cofactor, altering the conformation of the Jα helix at the C-terminus and a helix at the N-terminus (Fig. 15.3b). The differences in structure and conformational dynamics between the "dark" and "light" states allow the LOV2 domain to be used as a photoswitch [35, 36]. Another strategy for rationally engineering light-controllable allostery is exemplified in the design of a photoswitchable mammalian pyruvate kinase by Gehrig et al. [27]. Through MD simulations, they found sites on the kinase that are highly surface-exposed and undergo conformational changes when an allosteric effector is bound. They inserted the LOV2 domain into these locations and demonstrated reversible control of the enzyme by light in mammalian cells.

Most of the aforementioned examples can be framed within the free energy landscape model, but in one notable example, Lee et al. fused a LOV2 domain and *Escherichia coli* dihydrofolate reductase (DHFR) at surface sites in each protein that are connected to allosteric amino acid interaction networks (Fig. 15.3b). By connecting the allosteric networks of the two proteins, they were able to control the catalytic activity of the DHFR enzyme with light [59].

15.2.2 Random Insertion of Regulatory Domains

While there are many elegant examples of well-designed chimeric allosteric proteins, rational domain insertion often requires prior knowledge of the regulatory domain and host proteins' mechanisms and structures, which may not be generally available for proteins of interest. In these cases, random domain insertion may still generate new allosteric proteins, regardless of the depth of protein structure and function knowledge. For random domain insertion, an appropriate library of insertion variants must be created, and then a method for screening functional allosteric enzymes must be available. The Ostermeier lab's methods for the creation of insertion libraries include random double-stranded breaks introduced by DNase I [17, 31, 32] or S1 nuclease [132], or by inverse PCR [91] into the insertion target within an expression vector (Fig. 15.4). The Ostermeier methods utilize random deletions and insertions of amino acids near the gene insertion site and random circular permutations of the insertion DNA to increase the diversity of the library. The resulting plasmid prior to insertion is blunt-ended linear DNA, with productive linear plasmids having the insert target split on either end. The blunt-ended DNA insert is then ligated into the plasmid, resulting in a circular expression vector for the chimeric protein. In contrast, the method of the Kim lab [95, 112, 113] involves the

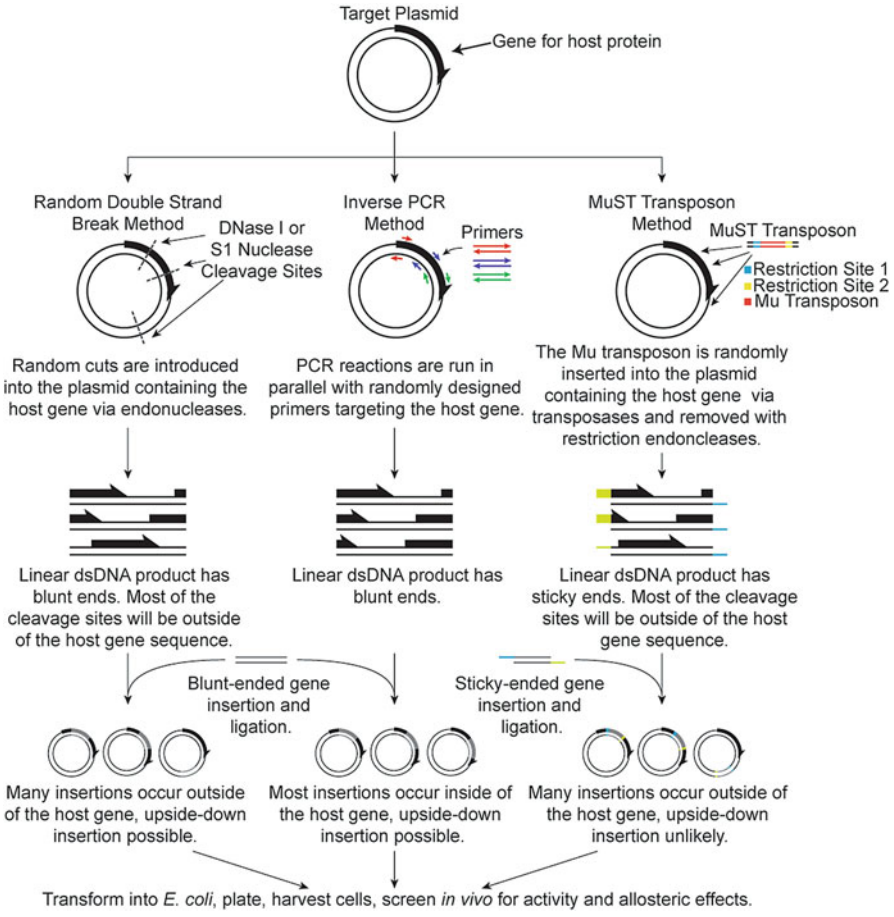


Fig. 15.4 Methods of random domain insertion. Details on the major steps of creating gene insertion libraries are shown in brief

use of a Mu transposon for gene insertion (Fig. 15.4). A transposon is a double-stranded DNA sequence that can be randomly inserted into a target DNA sequence through the use of a transposase. This method also allows for random circular permutations. In both methods, the resulting libraries contain many variations on the insert sequence. The need for screening limits which chimeric proteins can be generated through random domain insertion. There must be an observable indication that either the insertion target or the insertion domain is active, where cell- or lysate-based assays are more amenable to high throughput screening.

The work of Guntas et al. [31–33] provided an early template for random domain insertion for chimeric protein production with circular permutations and tandem deletions and repeats. The initial screening process for the target maltose-dependent β-lactamase consisted of transforming the vector library into *E. coli* and plating the cells on Luria-Bertani agar with ampicillin and maltose. Only cells that contained library members with

functional β-lactamase activity were further tested for maltose-dependent β-lactamase activity. In the first iterations of this study [32], 70% of library members had β-lactamase successfully inserted into MBP, and 4% could also bind maltose. One member, RG13 (Fig. 15.3c), was shown to have a great amount of switching behavior, with very little β-lactamase activity in the absence of maltose and a 25-fold increase in the β-lactamase activity in the presence of maltose. Wright et al. [131] determined that the BLA domain of the chimeric protein primarily samples conformations where it is less active, and addition of maltose allows BLA to sample conformations closer to its native state in the full BLA protein.

Wright et al. [132] later applied the blunt-end random domain insertion method (Fig. 15.4) to create a prodrug-activating protein that selectively activates in the presence of the hypoxia-inducible factor 1α (HIF-1α), a marker for some types of cancer. This protein, Haps59, was a fusion of the HIF-1α-binding CH1 domain of the p300 protein and yeast cytosine deaminase (yCD), an enzyme that converts the prodrug 5-fluorocytosine (5FC) to the toxic 5-fluorouracil (5FU), a chemotherapeutic agent. The library was first transformed into GIA39 *E. coli*, a uracil auxotroph lacking cytosine deaminase. Since the target chimeric protein would convert 5FC into the toxin 5FU in the presence of HIF-1α, the first stage of screening consisted of negative selection where 5FC and uracil were present and cells possessing cytosine deaminase activity in the absence of HIF-1α did not grow due to 5FU toxicity. Surviving cells then had their plasmids isolated and transformed into GIA39 cells containing a glutathione-s-transferase-HIF-1a fusion (gstHIF-1α) plasmid with an arabinose promoter. The cells were plated on agar supplemented with cytosine and arabinose so that members that had cytosine deaminase activity in the presence of HIF-1α could grow but members without cytosine deaminase activity could not. The most promising chimeric protein, Haps59, was demonstrated to significantly decrease cell survival in human breast cancer and colorectal cancer cells in the presence of 5FC under hypoxic conditions similar to those in tumors, but not under normal oxygen conditions present outside of tumors.

Flow cytometry analysis fluorescent-assisted cell sorting (i.e., FACS) has become a popular method for screening libraries for desired functions. Ribeiro et al. [102, 103] used this selection method to create a more efficient endo-β-1,4-xylanase. Xylanase is an important enzyme for the conversion of plant biomass to fermentable sugars for biofuel production. Unfortunately, product inhibition from xylose is a significant limiting factor for xylanase function. The xylose binding domain from xylose binding protein (XBP) was selected as an insert domain since it undergoes a significant conformational change upon binding xylose. The resulting protein possessed increased xylanase activity even after large amounts of xylose product were formed. Because the goal was to create an improved version of an already active enzyme where the effector molecule was also the product, initial screening consisted of determining xylose-binding activity. *E. coli* cells lacking wild-type XBP were used since XBP is required for the efficient uptake of xylose. Cells were transformed with a pT7T3GFP_XBP plasmid, which contained a xylose promoter and would express GFP and the chimeric protein containing the XBP domain. FACS was used to sort for cells that displayed xylose binding, and these cells were further screened

for xylanase activity. Cells with xylanase activity were determined by the formation of halos on agar xylan plates after staining with Congo Red, indicating xylan breakdown to xylose.

Nadler et al. [81] utilized the Mu transposon to insert GFP (green fluorescent protein) into a ligand-binding domain to demonstrate the creation of allosteric metabolite sensors. To demonstrate their technique, they created a single-fluorescent protein biosensor (SFPB) to quantify in-cell levels of trehalose. A binding event in the ligand binding domain was predicted to trigger an allosteric change in fluorescence in the inserted circularly permuted GFP domain. Oakes et al. [90] also utilized this technique with the goal of introducing allosteric regulation into the CRISPR-associated protein Cas9 by inserting the alpha estrogen receptor, which crystallographic evidence had shown undergoes a significant conformational change upon ligand binding. To identify sites where the estrogen receptor domain could be inserted, mouse PDZ-domain insertion sites were screened for several rounds in *E. coli*, with round one selecting for insertions that did not cause a frameshift and were not reversed. Round two screened for Cas9's ability to suppress RFP (red fluorescent protein) expression with the insertion present at a random site to weed out non-functional Cas9 activity. Sites on Cas9 that best tolerated insertion were clustered around flexible loops, the ends of helices, and on the surface of the protein. Once functional insertion sites were identified, the protein variants were then screened for estrogen receptor insertion and assayed for Cas9 activity in a similar way to the PDZ screening in the presence and absence of a 4-hydroxytamoxifen ligand.

The rational design methods of regulatory domain insertion benefit from computation and structural knowledge to tune the thermodynamic and steric parameters of the chimeric protein to achieve the desired functional modulation. The random domain insertion methods take advantage of clever library development and screening assays, benefitting from but not necessitating structural knowledge of the regulatory domains and host proteins. These methods have brought new insights into the free energy landscapes underlying protein regulation, and have already resulted in new allosteric proteins finding practical use in biotechnology and medical applications.

15.3 Engineering Protein Allostery Through Covalent Modification

Allostery and long-range communication may be intrinsic to all protein folds. Instead of whole domain insertions, more subtle changes may likewise result in new allosteric regulation. Covalent modification of protein scaffolds represents another method for altering or creating allostery in proteins. Rather than adding a new domain to perturb a structural ensemble, covalent modification often either restricts the accessible conformations with a linker or perturbs an amino acid interaction network by conjugating a bulky molecule to a labile, surface-exposed amino acid.

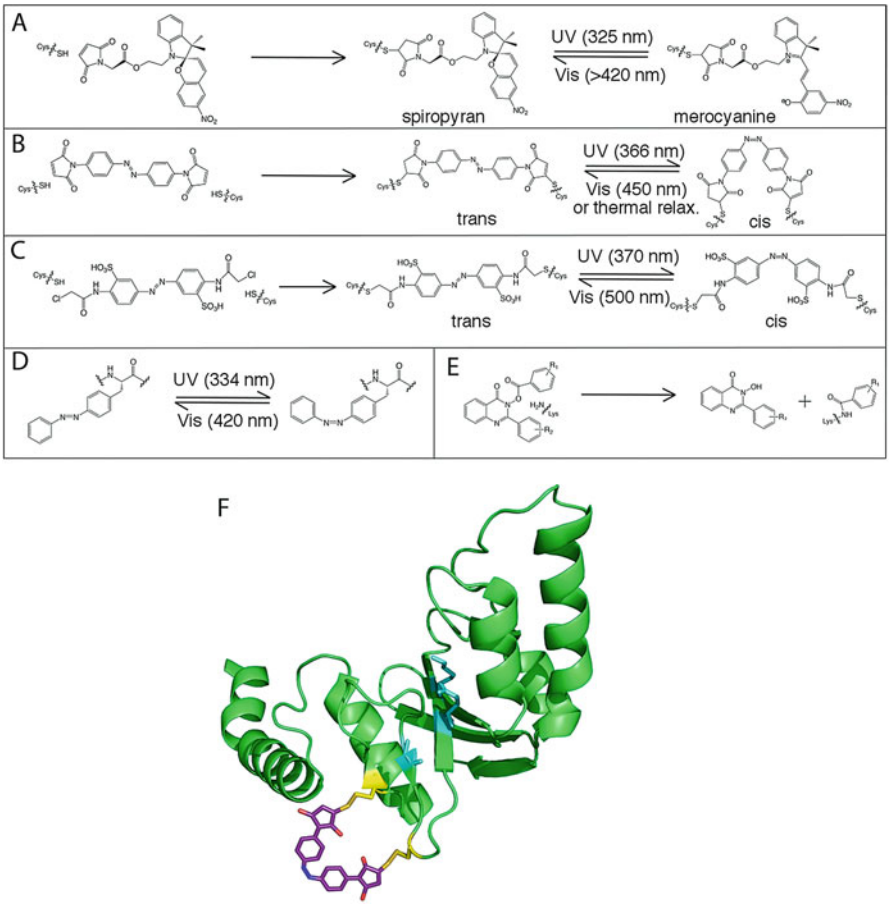


Fig. 15.5 Covalent modification of proteins to control allostery. (**a**) Conjugation of spiropyran to a Cys residue and isomerization to the merocyanine form. (**b**) Conjugation of the bifunctional azobenzene derivative 4,4′-bis (maleimido)azobenzene to two cysteine residues and isomerization between *trans* and *cis* forms. (**c**) Conjugation of the bifunctional azobenzene derivative 3,3′-bis (sulfonato)-4,4′-bis (chloroacetamido)azobenzene (BSBCA) to two Cys residues and isomerization between *trans* and *cis* forms. (**d**) Isomerization of the genetically encoded azobenzene derivative AzoPhe. (**e**) Covalent modification of a Lys residue by quinazolin-4 (3H)-one hydroxamic ester. (**f**) Covalent modification of PvuII with azomal (PDB: 1NI0), adapted from Schierling et al. [108]. Active site residues (cyan) are disrupted upon isomerization of azomal to the *trans* form

Putri et al. used a spiropyrin photoswitch linker (Fig. 15.5a) to engineer a light-sensitive human serum albumin (HSA) [98]. The linker was conjugated to a surface-accessible cysteine via Michael addition in subdomain IA, which resulted in the regulation of ligand release at subdomain IB greater than 130 Å away. For example, the HSA ligands methylene orange and bromocresol green were released upon UV irradiation, and methyl orange binding affinity decreased by three-fold after UV

irradiation. MD simulations confirmed that the subdomains were not directly interactions, indicating that changes to ligand affinities were due to longer-range network interactions.

One of the most common and useful covalent attachments is the azobenzene photoswitch, which reversibly isomerizes between *cis* and *trans* conformations [67, 98, 101]. Several strategies have been employed to integrate the azobenzene photoswitch: it can be directly incorporated into a peptide during peptide synthesis reactions [61, 122], incorporated into a protein via non-natural amino acids in vitro *[79, 82]* or in vivo *[9]*, crosslinked to the protein to one or two solvent-exposed cysteines [79, 128, 135], and/or incorporated into a protein-binding ligand [96, 109, 124]. Generally, the isomerization of an azobenzene derivative between *trans* and *cis* conformations changes the relative free energies of the active and inactive protein conformations. In one example, Schierling et al. crosslinked an azobenzene derivative to the restriction enzyme PvuII, allowing allosteric inhibition to be reversibly controlled by light. Here, two cysteine residues were conjugated to the bifunctional azobenzene derivative 4,4′-bis (maleimido)azobenzene (Fig. 15.5b, f). The switch from *cis* to *trans* conformation disrupted the PvuII secondary structure, altered the active site structure, and reduced enzyme activity by 16-fold [108]. In another example, Ritterson et al. crosslinked two cysteine residues in cadherin, a cell-cell adhesion protein, using the bifunctional azobenzene derivative BSBCA (Fig. 15.5c). Upon irradiation with near-UV light, the apparent Ca^{2+} binding affinity decreased 18-fold. Because Ca^{2+} binding is linked to cadherin dimerization, the function of cadherin was controlled by light. The photoswitch was reversible up to at least three light/dark cycles. A genetically encoded azobenzene derivative has also been developed for use in *E. coli* allowing for the generation of azobenzene photoswitches in live cells [79] (Fig. 15.5d).

Key surface exposed residues might also be covalently labeled to control protein function. For example, Bongard et al. [8] discovered a class of small molecules, quinazolin-4 (3H)-one hydroxamic esters, that covalently modify Lys residues (Fig. 15.5e). They found that these molecules allosterically activate DegS, a bacterial serine protease, by modifying the conformation of a loop in the protease domain. By changing the Lys residues that were covalently modified, they were able to determine which modifications were responsible for allosteric activation. The Lys residue responsible for the activation was found to be part of an allosteric amino acid interaction network. This strategy of covalent modification can be applied to other proteins for allosteric regulation or to find noncatalytic residues that are part of allosteric amino acid interaction networks (see Fig. 15.1b).

15.4 Modifying Protein Allostery Through Mutagenesis

The ability to quickly create point mutations in a target protein has been a boon for molecular biology and biochemistry. With some prior knowledge about a protein's structure and amino acid sequence, small modifications to the amino acid sequence

can be made to create pH switchable proteins or identify and modify the amino acid interaction networks that allow a protein to function. These studies generally represent more subtle changes than whole domain insertions or covalent modification. This section will summarize select examples from the literature where mutagenesis was used as a tool to modify, study, or create new allostery.

15.4.1 Generating pH-Dependent Proteins Through Mutagenesis

Since a protein's conformation is so dependent on its charge, there are many examples of pH-switchable proteins [44, 51, 68, 70, 87, 107, 115, 120]. Examples found in nature include the *E. coli* chaperone HdeA [41], β-ureidopropionase in the pyrimidine catabolic pathway [69], bovine β-lactoglobulin [99], human prolactin [49, 55], and nitrophorin 4 (NP4) [20]. The ability to engineer pH-dependent allosteric proteins has great medical potential. For example, it is well known that tumors often create an acidic environment, which may allow for pH-sensitive proteins to functionally target cancer cells. There are also stark pH differences within cells, potentially allowing for even more specific targeting in diseased states that affect membrane-bound organelles such as mitochondria or lysosomes.

NP4 is a protein used by the kissing bug *Rhodnius prolixus* to selectively release nitric oxide into its victim's tissues via pH-dependent conformational changes. When deprotonated at the higher pH in the tissue, a hydrogen bond with Asp30 is broken, causing a conformational change that results in the release of bound nitric oxide (Fig. 15.6). Di Russo et al. used pH replica exchange molecular dynamics (pH-REMD) [46, 118] to describe the free energy landscape of NP4's pH-dependent conformational change. By creating a variant that destabilized the open conformation but did not affect the pK_a of Asp30 in either conformation, they were able to show that the observed pK_a depends on both the equilibrium constant for the conformations and the pK_a at each conformation as predicted by their model. In addition to changing the equilibrium constant between conformations, the pK_a and local dynamics of buried ionizable residues in proteins can be altered by changing the distribution of surface charges, as demonstrated by Pey et al. using *E. coli* thioredoxin [104].

In these studies, the coupling between protonation state equilibrium and conformational equilibrium must be considered in order to understand the thermodynamics of the pH-dependent conformational change. Exploring this further, Liu et al. studied the coupling of the protonation state to the conformation of the inner residues of staphylococcal nuclease by substituting ionizable residues into the interior of the protein and measuring the pK_a values of these residues [62]. By taking into account the conformational equilibrium and the pK_a of the residue of interest in each conformation there was excellent agreement with the experimentally observed pK_a

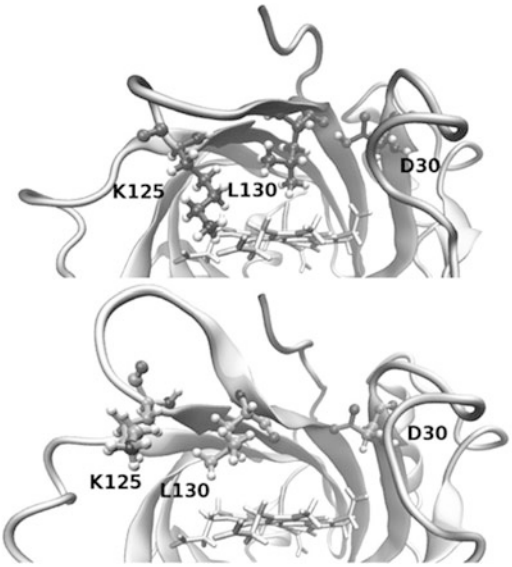


Fig. 15.6 pH-dependent modulation of the NP4 protein. Closed (top) and open (bottom) conformations of NP4. Protonation of Asp30 stabilizes the closed conformation, while deprotonation stabilizes the open conformation. (Reprinted with permission from Di Russo, N. V.; Martí, M. A.; Roitberg, A. E. Underlying Thermodynamics of PH-Dependent Allostery. J. Phys. Chem. B 2014, 118 (45), 12,818–12,826. Copyright 2014 American Chemical Society)

values and structural information for buried ionizable residues in the same protein [45].

There are now various examples of engineering proteins to be responsive to pH changes by substituting buried residues for ionizable residues. Zimenkov et al. [140] designed peptides that reversibly self-assemble into fibrils in response to pH changes. The self-assembly results from helix-coil transitions in the peptides when His residues in the helix are protonated. There are also biomedical applications to engineering pH-sensitive proteins. The fibronectin (Fn3) domain is a useful scaffold for engineering proteins that bind to proteins associated with cardiovascular disease and cancer. By engineering His residues into the hydrophobic core, Heinzelman et al. [37] engineered Fn3 domains that bind epidermal growth factor receptor (EGFR) with much lower affinity at low pH, and the effect was shown to be reversible. A similar strategy was used to engineer pH-responsive binding proteins for the Fc portion of human immunoglobulin (HIgG) derived from the hyperthermophilic Sso7d protein by Gera et al. [28]

Although most examples of engineered pH-dependent allostery involve substituting ionizable residues into the hydrophobic core of a protein, other residues can also be substituted to create pH-dependent allostery. Lysteriolysin O, a pore-forming protein secreted by *Listeria cytogenes* that destroys cholesterol-rich cell membranes, was engineered to be activated at slightly acidic pH via an allosteric mechanism. Kisovec et al. [50] substituted an Ala residue for a Tyr residue in the vicinity of ionizable residues previously identified to affect pH-dependent stability of the protein, changing the pK_a values of the buried ionizable residues.

15.4.2 Perturbing Allosteric Amino Acid Interaction Networks with Mutagenesis

In the past decade, there have been several examples of mutational studies that identified and perturbed amino acid interaction networks. In several cases described here, changing network residues distal from the binding or active sites of proteins have wide implications for protein function and allostery. These studies have been of recent interest due to the rise of allosteric drugs and the hunt for druggable "cryptic" allosteric sites.

A long-standing problem in cellular signaling is tuning substrate and protein interaction specificity. For example, several PDZ domains have intersecting substrate specificity but play discrete roles in signaling pathways [24]. Gianni et al. investigated this problem using double mutant cycle analysis [42] on two specific PDZ domains that have computational evidence supporting distinct allosteric networks [52]. Double mutant cycle analysis helps to determine if two sites are allosterically and energetically coupled. Through these studies, they were able to determine the binding kinetics for 31 protein variants of each binding partner. These energetic constraints determined that the two binding partners had distinct allosteric mechanisms and that these allosteric networks were not conserved across the protein family. The authors emphasized the importance of amino acid composition to allosteric networks [30].

Introducing point mutations can also provide experimental evidence for the presence of amino acid interaction networks that play a critical role in allostery. Holliday and coworkers identified two dynamic amino acid clusters within cyclophilin A using RASSMM (relaxation techniques and single-site multiple mutations) [40]. RASSMM utilizes the high sensitivity of NMR methods to identify hot spots that are related to active site structural dynamics. Within the two coupled networks in cyclophilin A, two key "hot spot" residues were identified (Fig. 15.7). When these residues were perturbed, the active site dynamics and enzyme turnover were negatively impacted, despite both residues being greater than 15 Å from the active site. Similar methods can be applied to discover dynamically-coupled networks in other proteins.

There are several other NMR methods that can be used to identify and perturb allosteric networks in proteins. Among these is chemical shift covariance analysis (CHESCA) [111], which is similar to RASSMM in that it helps to elucidate allosteric mechanisms in proteins. CHESCA has been utilized to determine amino acid interaction networks in several proteins, including the alpha subunit of tryptophan synthase [2], a DNA repair protein (Rad50) [10], and phosophohexomutase [134]. In a different NMR approach to allostery, Cui and coworkers described a method that solely relies on total protein ^{1}H, ^{15}N chemical shift perturbations, or CSPs [17]. In order to probe the allosteric sites in tyrosine phosphatases, they introduced several Ala substitutions in an acidic loop region and then monitored chemical shift changes in residues distal from the active site. Using this method, they

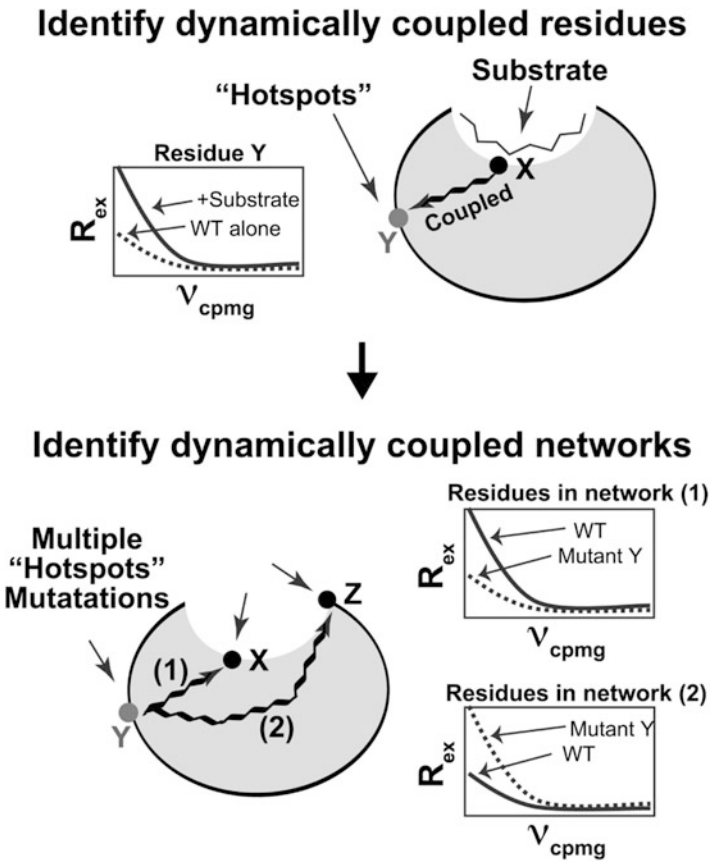


Fig. 15.7 The RASMM method for identifying dynamically coupled amino acid interaction networks in proteins. Amino acids distal from the active site are identified as "hot spots" based on their R_2 relaxation dispersion profiles in the apo and substrate-bound states. Residues that are linked allosterically to the active site are expected to show a measurable change in R_2 relaxation dispersion behavior in the presence of substrate. To identify the global communications networks, single site mutations of "hot spot" residues are generated and global R_2 relaxation dispersion profiles are examined. Reprinted from Structure, Volume 25 Issue 2, MJ Holliday, C Camilloni, GS Armstrong, M Vendruscolo, EZ Eisenmesser. Networks of Dynamic Allostery Regulate Enzyme Function, 276–286, Copyright (2017), with permission from Elsevier

identified two separate amino acid interaction networks. When network residues were changed, a marked decrease in enzyme activity was noted, despite the distance of these network residues from the active site. Tyrosine phosphatases have been proposed to be potential drug targets for the treatment of diabetes, cancer, and obesity, and so the identification of these network residues helps to identify protein surfaces that might be targetable by allosteric drugs.

15.4.3 *Engineering Allostery "Out Of" Proteins*

All of the examples above have explored ways to generate new allosteric proteins. However, it is sometimes advantageous to render a protein unresponsive to allosteric regulation while maintaining other functions. For example, metabolic enzymes are often allosterically regulated, but such regulation might interfere in a particular synthetic scheme [22]. One great recent example is the engineering of tryptophan synthase (TS) for the synthesis of novel indole analogs. TS is composed of two subunits, the alpha subunit (αTS) that generates indole, which is then directly channeled through a hydrophobic tunnel to the beta subunit (βTS) that completes tryptophan biosynthesis. Only the catalytic activity of βTS was desired for the synthesis of indole analogs, but the large decrease in βTS activity in the absence of αTS made synthesis difficult [85]. The Arnold lab [13] sought to remedy this problem by using directed evolution to generate βTS that recapitulated the allosteric activation in the absence of αTS. The resulting βTS had greater catalytic activity than the native complex and also a more diversified substrate portfolio.

In further studies, Buller and coworkers sought to identify the mechanism by which their evolved βTS was able to fully function in the absence of αTS [12]. By monitoring the steady-state concentrations of βTS intermediates, they determined that their directed evolution variants progressively changed the rate-limiting step of the reaction. In addition, their variants shifted the steady-state distribution, favoring the tryptophan bound products (Fig. 15.8). They concluded that the amino acid

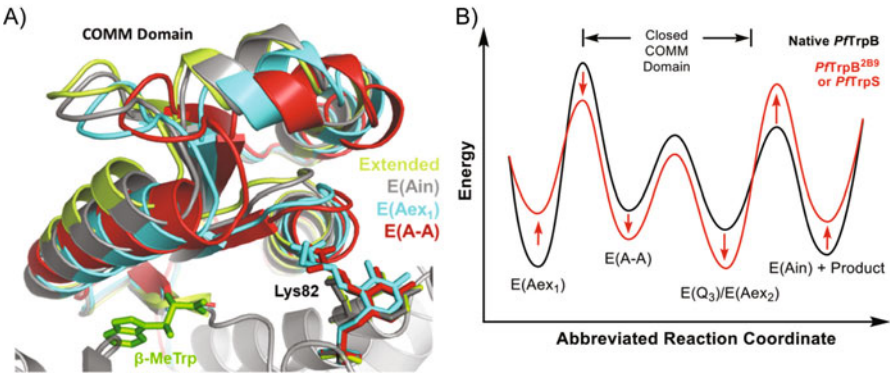


Fig. 15.8 Engineering allostery "out of" the beta subunit of tryptophan synthase. (**a**) A crystal structure of the COMM domain of βTS representing the differences in the open/closed conformations through the catalytic cycle. The "extended" state represents the open conformation. As the reaction proceeds, the COMM domain enters the fully closed conformation (E (A-A)). (**b**) A theoretical reaction coordinate illustrating the energy levels of the βTS intermediates. The engineered beta subunit ($PfTrp^{B2B9}$) creates a decrease in energy of the product bound state. This results in a more efficient enzyme without the presence of the allosteric effector (αTS). Reprinted with permission from Buller, A. R. et al. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. J. Am. Chem. Soc. 140, 7256–7266 (2018). Copyright 2018 American Chemical Society

substitutions stabilized a population in βTS that is typically only favored when bound to αTS – therefore recapitulating the complex activity in just the βTS subunit. The TS example provides an important counterexample for allostery engineering – instead of engineering in allostery, the goal was to engineer allostery out of βTS so the enzyme was fully capable of functioning alone.

15.5 Conclusion

The ability to control engineered proteins allosterically would revolutionize medicine and industry. Here, we have provided examples that illuminated the fundamentals of allostery, and have shown that there are newly designed allosteric proteins already finding practical use. Rational and random domain insertion methods appear to be the most prolific in creating novel allosteric proteins. The modular nature of certain types of catalytic and regulatory domains may be why these approaches are so successful. The ability to engineer photoswitchable proteins is especially exciting, which has been accomplished mainly through LOV2 domain insertion and/or insertion of the azobenzene photoswitch through covalent modification. Likewise, pH-switchable proteins could offer means to specifically target different cellular locations or activate only under certain disease conditions. The identification of amino acid interaction networks through a variety of biophysical and computational methods [1] offers not only a means to potentially engineer allostery and function, but these methods can also identify protein surfaces that may be amenable to small molecule targeting and allosteric regulation.

References

1. Axe JM, O'Rourke KF, Kerstetter NE, Yezdimer EM, Chan YM, Chasin A, Boehr DD (2015) Severing of a hydrogen bond disrupts amino acid networks in the catalytically active state of the alpha subunit of tryptophan synthase. Protein Sci 24(4):484–494
2. Axe JM, Yezdimer EM, O'Rourke KF, Kerstetter NE, You W, Chang CE, Boehr DD (2014) Amino acid networks in a (beta/alpha) (8) barrel enzyme change during catalytic turnover. J Am Chem Soc 136(19):6818–6821
3. Berliner L (2015) Protein NMR: modern techniques and biomedical applications
4. Blackmore NJ, Nazmi AR, Hutton RD, Webby MN, Baker EN, Jameson GB, Parker EJ (2015) Complex formation between two biosynthetic enzymes modifies the allosteric regulatory properties of both: AN EXAMPLE OF MOLECULAR SYMBIOSIS. J Biol Chem 290(29):18187–18198
5. Blackmore NJ, Reichau S, Jiao W, Hutton RD, Baker EN, Jameson GB, Parker EJ (2013) Three sites and you are out: ternary synergistic allostery controls aromatic amino acid biosynthesis in Mycobacterium tuberculosis. J Mol Biol 425(9):1582–1592
6. Boehr DD, D'Amico RN, O'Rourke KF (2018) Engineered control of enzyme structural dynamics and function. Protein Sci 27(4):825–838

7. Boehr DD, Schnell JR, McElheny D, Bae SH, Duggan BM, Benkovic SJ, Dyson HJ, Wright PE. A distal mutation perturbs dynamic amino acid networks in dihydrofolate reductase. pp 1520–4995 (Electronic)
8. Bongard J, Lorenz M, Vetter IR, Stege P, Porfetye AT, Schmitz AL, Kaschani F, Wolf A, Koch U, Nussbaumer P, Klebl B, Kaiser MA, Ehrmann MA. Identification of Noncatalytic Lysine Residues from Allosteric Circuits via Covalent Probes. pp 1554–8937 (Electronic)
9. Bose M, Groff D, Xie J, Brustad E, Schultz PG (2006) The incorporation of a photoisomerizable amino acid into proteins in E. coli. J Am Chem Soc 128(2):388–389
10. Boswell ZK, Rahman S, Canny MD, Latham MP (2018) A dynamic allosteric pathway underlies Rad50 ABC ATPase function in DNA repair. Sci Rep 8(1):1639
11. Boulton S, Akimoto M, Selvaratnam R, Bashiri A, Melacini G (2014) A tool set to map allosteric networks through the NMR chemical shift covariance analysis. Sci Rep 4:7306
12. Buller AA-O, van Roye P, Cahn JKB, Scheele RA, Herger M, Arnold FA-OX. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. pp 1520–5126 (Electronic))
13. Buller AR, Brinkmann-Chen S, Romney DK, Herger M, Murciano-Calles J, Arnold FH (2015) Directed evolution of the tryptophan synthase beta-subunit for stand-alone function recapitulates allosteric activation. Proc Natl Acad Sci U S A 112(47):14599–14604
14. Choi JH, Laurent AH, Hilser VJ, Ostermeier M (2015) Design of protein switches based on an ensemble model of allostery. Nat Commun 6:6968
15. Cristian C, Laura A, Sabrina G, Marco A, Edoardo R, Solei C, Paolo Z, Jan P, Monica B, van Etten JL (2015) Optogenetics. Engineering of a light-gated potassium channel. Science 348(6235):707
16. Cross PJ, Allison TM, Dobson RC, Jameson GB, Parker EJ. Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. pp 1091–6490 (Electronic)
17. Cui DS, Beaumont V, Ginther PS, Lipchock JM, Loria JP (2017) Leveraging reciprocity to identify and characterize unknown allosteric sites in protein tyrosine phosphatases. J Mol Biol 429(15):2360–2372
18. Davies BR, Guan N, Logie A, Crafter C, Hanson L, Jacobs V, James N, Dudley P, Jacques K, Ladd B, D'Cruz CM, Zinda M, Lindemann J, Kodaira M, Tamura K, Jenkins EL (2015) Tumors with AKT1E17K mutations are rational targets for single agent or combination therapy with AKT inhibitors. Mol Cancer Ther 14(11):2441–2451
19. Dayie KT, Wagner G, Lefevre JF (1996) Theory and practice of nuclear spin relaxation in proteins. Annu Rev Phys Chem 47:243–282
20. Di Russo NV, Marti MA, Roitberg AE (2014) Underlying thermodynamics of pH-dependent allostery. J Phys ChemB 118(45):12818–12826
21. Djuranovic S, Nahvi A, Green R (2011) A parsimonious model for gene regulation by miRNAs. Science (New York, NY) 331(6017):550–553
22. Du L, Lou L (2010) PKS and NRPS release mechanisms. Nat Prod Rep 27(2):255–278
23. Fan Y, Cross PJ, Jameson GB, Parker EJ (2018) Exploring modular allostery via interchangeable regulatory domains. Proc Natl Acad Sci U S A 115(12):3006–3011
24. Feng W, Zhang M (2009) Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. Nat Rev Neurosci 10(2):87–99
25. Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science (New York, NY) 254(5038):1598–1603
26. Gautier A, Gauron C, Volovitch M, Bensimon D, Jullien L, Vriz S (2014) How to control proteins with light in living systems. Nat Chem Biol 10(7):533–541
27. Gehrig S, Macpherson JA, Driscoll PC, Symon A, Martin SR, MacRae JI, Kleinjung J, Fraternali F, Anastasiou D (2017) An engineered photoswitchable mammalian pyruvate kinase. FEBS J 284(18):2955–2980
28. Gera N, Hill AB, White DP, Carbonell RG, Rao BM. Design of pH sensitive binding proteins from the hyperthermophilic Sso7d scaffold. pp 1932–6203 (Electronic)

29. Ghosh A, Vishveshwara S (2007) A study of communication pathways in methionyl- tRNA synthetase by molecular dynamics simulations and structure network analysis. Proc Natl Acad Sci U S A 104(40):15711–15716
30. Gianni S, Haq SR, Montemiglio LC, Jurgens MC, Engstrom A, Chi CN, Brunori M, Jemth P. Sequence-specific long range networks in PSD-95/discs large/ZO-1 (PDZ) domains tune their binding selectivity. pp 1083–351X (Electronic)
31. Guntas G, Mansell TJ, Kim JR, Ostermeier M (2005) Directed evolution of protein switches and their application to the creation of ligand-binding proteins. Proc Natl Acad Sci U S A 102(32):11224–11229
32. Guntas G, Mitchell SF, Ostermeier M (2004) A molecular switch created by in vitro recombination of nonhomologous genes. Chem Biol 11(11):1483–1487
33. Guntas G, Ostermeier M (2004) Creation of an allosteric enzyme by domain insertion. J Mol Biol 336(1):263–273
34. Guo Z, Johnston WA, Stein V, Kalimuthu P, Perez-Alcala S, Bernhardt PV, Alexandrov K. Engineering PQQ-glucose dehydrogenase into an allosteric electrochemical Ca (2+) sensor. pp 1364–548X (Electronic)
35. Halavaty AS, Moffat K (2007) N- and C-terminal flanking regions modulate light-induced signal transduction in the LOV2 domain of the blue light sensor phototropin 1 from Avena sativa. Biochemistry 46(49):14001–14009
36. Harper SM, Neil LC, Gardner KH (2003) Structural basis of a phototropin light switch. Science (New York, NY) 301(5639):1541–1544
37. Heinzelman P, Krais J, Ruben E, Pantazes R (2015) Engineering pH responsive fibronectin domains for biomedical applications. J Biol Eng 9:6
38. Helmstaedt K, Krappmann S, Braus GH (2001) Allosteric regulation of catalytic activity: Escherichia coli aspartate transcarbamoylase versus yeast chorismate mutase. Microbiol Mol Biol R 65(3):404–421, table of contents
39. Hisatomi O, Furuya K (2015) A light-regulated bZIP module, photozipper, induces the binding of fused proteins to the target DNA sequence in a blue light-dependent manner. Photochem Photobiol Sci: Off J Eur Photochem Assoc Eur Soc Photobiol 14(11):1998–2006
40. Holliday MJ, Camilloni C, Armstrong GS, Vendruscolo M, Eisenmesser EZ (2017) Networks of dynamic Allostery regulate enzyme function. Structure (London, England: 1993) 25(2):276–286
41. Hong W, Jiao W, Hu J, Zhang J, Liu C, Fu X, Shen D, Xia B, Chang Z (2005) Periplasmic protein HdeA exhibits chaperone-like activity exclusively within stomach pH range by transforming into disordered conformation. The J Biol Chem 280(29):27029–27034
42. Horovitz A, Fersht AR (1990) Strategy for analysing the co-operativity of intramolecular interactions in peptides and proteins. J Mol Biol 214(3):613–617
43. Huisman FH, Koon N, Bulloch EM, Baker HM, Baker EN, Squire CJ, Parker EJ (2012) Removal of the C-terminal regulatory domain of alpha-isopropylmalate synthase disrupts functional substrate binding. Biochemistry 51(11):2289–2297
44. Idili A, Vallee-Belisle A, Ricci F (2014) Programmable pH-triggered DNA nanoswitches. J Am Chem Soc 136(16):5836–5839
45. Isom DG, Castaneda CA, Cannon BR, Garcia-Moreno B. Large shifts in pKa values of lysine residues buried inside a protein. pp 1091–6490 (Electronic)
46. Itoh SG, Damjanovic A, Brooks BR (2011) pH replica-exchange method based on discrete protonation states. Proteins 79(12):3420–3436
47. Jiao W, Hutton RD, Cross PJ, Jameson GB, Parker EJ. Dynamic cross-talk among remote binding sites: the molecular basis for unusual synergistic allostery. pp 1089–8638 (Electronic)
48. Ke W, Laurent AH, Armstrong MD, Chen Y, Smith WE, Liang J, Wright CM, Ostermeier M, van den Akker F (2012) Structure of an engineered beta-lactamase maltose binding protein fusion protein: insights into heterotropic allosteric regulation. PLoS One 7(6):e39168

49. Keeler C, Jablonski EM, Albert YB, Taylor BD, Myszka DG, Clevenger CV, Hodsdon ME (2007) The kinetics of binding human prolactin, but not growth hormone, to the prolactin receptor vary over a physiologic pH range. Biochemistry 46(9):2398–2410
50. Kisovec M, Rezelj S, Knap P, Cajnko MM, Caserman S, Flasker A, Znidarsic N, Repic M, Mavri J, Ruan Y, Scheuring S, Podobnik M, Anderluh G. Engineering a pH responsive pore forming protein. pp 2045–2322 (Electronic)
51. Kohse S, Neubauer A, Pazidis A, Lochbrunner S, Kragl U. Photoswitching of enzyme activity by laser-induced pH-jump. pp 1520–5126 (Electronic)
52. Kong Y, Karplus M (2009) Signaling pathways of PDZ2 domain: a molecular dynamics interaction correlation analysis. Proteins 74(1):145–154
53. Koshland DE Jr, Nemethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry 5(1):365–385
54. Koushanpour A, Gamella M, Guo Z, Honarvarfard E, Poghossian A, Schoning MJ, Alexandrov K, Katz EA. Ca (2+)-Switchable Glucose Dehydrogenase Associated with Electrochemical/Electronic Interfaces: Applications to Signal-Controlled Power Production and Biomolecular Release. pp 1520–5207 (Electronic)
55. Kulkarni MV, Tettamanzi MC, Murphy JW, Keeler C, Myszka DG, Chayen NE, Lolis EJ, Hodsdon ME. Two independent histidines, one in human prolactin and one in its receptor, are critical for pH-dependent receptor recognition and activation. pp 1083–351X (Electronic)
56. Kumar R, McEwan IJ (2012) Allosteric modulators of steroid hormone receptors: structural dynamics and gene regulation. Endocr Rev 33(2):271–299
57. Lafrance-Vanasse J, Williams GJ, Tainer JA (2015) Envisioning the dynamics and flexibility of Mre11-Rad50-Nbs1 complex to decipher its roles in DNA replication and repair. Prog Biophys Mol Bio 117(2–3):182–193
58. Lang EJ, Cross PJ, Mittelstadt G, Jameson GB, Parker EJ (2014) Allosteric ACTion: the varied ACT domains regulating enzymes of amino-acid metabolism. Curr Opin Struct Biol 29:102–111
59. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R (2008) Surface sites for engineering allosteric control in proteins. Science (New York, NY) 322(5900):438–442
60. Ling Y, Jing M, Wang XD (2015) Allosteric therapies for lung cancer. Cancer Metast Rev 34(2):303–312
61. Liu D, Karanicolas J, Yu C, Zhang Z, Woolley GAJB, Letters MC (1997) Site-specific incorporation of photoisomerizable azobenzene groups into ribonuclease S. Bioorganic Med Chem Lett 7(20):2677–2680
62. Liu J, Swails J, Zhang JZH, He X, Roitberg AE (2018) A coupled ionization-conformational equilibrium is required to understand the properties of Ionizable residues in the hydrophobic interior of staphylococcal nuclease. J Am Chem Soc 140(5):1639–1648
63. Lockless SW, Ranganathan R (1999) Evolutionarily conserved pathways of energetic connectivity in protein families. Science (New York, NY) 286(5438):295–299
64. Lu S, Li S, Zhang J. Harnessing allostery: a novel approach to drug discovery. pp 1098–1128 (Electronic)
65. Lungu OI, Hallett RA, Choi EJ, Aiken MJ, Hahn KM, Kuhlman B (2012) Designing photoswitchable peptides using the AsLOV2 domain. Chem Biol 19(4):507–517
66. Mandal M, Breaker RR (2004) Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5(6):451–463
67. Mart RJ, Allemann RK (2016) Azobenzene photocontrol of peptides and proteins. Chem Commun 52(83):12262–12277
68. Marttila AT, Hytonen VP, Laitinen OH, Bayer EA, Wilchek M, Kulomaa MS. Mutation of the important Tyr-33 residue of chicken avidin: functional and structural consequences. pp 0264–6021 (Print)
69. Maurer D, Lohkamp B, Krumpel M, Widersten M, Dobritzsch DA. Crystal structure and pH-dependent allosteric regulation of human beta-ureidopropionase, an enzyme involved in anticancer drug metabolism. pp 1470–8728 (Electronic)

70. McConnell EM, Bolzon R, Mezin P, Frahm G, Johnston M, DeRosa MC (2016) pHAST (pH-driven aptamer switch for thrombin) catch-and-release of target protein. Bioconjug Chem 27(6):1493–1499
71. Meister GE, Joshi NS (2013) An engineered calmodulin-based allosteric switch for peptide biosensing. Chembiochem 14(12):1460–1467
72. Mills E, Chen X, Pham E, Wong S, Truong K (2012) Engineering a photoactivated caspase-7 for rapid induction of apoptosis. ACS Synth Biol 1(3):75–82
73. Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: a plausible model. J Mol Biol 12:88–118
74. Mootz HD (2017) Split Inteins
75. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M (2011) Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc Natl Acad Sci U S A 108(49):E1293–E1301
76. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. Nature 508(7496):331–339
77. Motta-Mena LB, Reade A, Mallory MJ, Glantz S, Weiner OD, Lynch KW, Gardner KH (2014) An optogenetic gene expression system with rapid activation and deactivation kinetics. Nat Chem Biol 10(3):196–202
78. Munro JB, Vaiana A, Sanbonmatsu KY, Blanchard SC (2008) A new view of protein synthesis: mapping the free energy landscape of the ribosome using single-molecule FRET. Biopolymers 89(7):565–577
79. Muranaka N, Hohsaka T, Sisido M (2002) Photoswitching of peroxidase activity by position-specific incorporation of a photoisomerizable non-natural amino acid into horseradish peroxidase. FEBS Lett 510(1–2):10–12
80. Murciano-Calles J, Romney DK, Brinkmann-Chen S, Buller AR, Arnold FH (2016) A panel of TrpB biocatalysts derived from tryptophan synthase through the transfer of mutations that mimic allosteric activation. Angew Chem Int Ed Engl 55(38):11577–11581
81. Nadler DC, Morgan SA, Flamholz A, Kortright KE, Savage DF (2016) Rapid construction of metabolite biosensors using domain-insertion profiling. Nat Commun 7:12266
82. Nakayama K, Endo M, Majima T (2004) Photochemical regulation of the activity of an endonuclease BamHI using an azobenzene moiety incorporated site-selectively into the dimer interface. Chem Commun 21:2386–2387
83. Nguyen LM, Roche J (2017) High-pressure NMR techniques for the study of protein dynamics, folding and aggregation. J Magnet Reson (San Diego, Calif: 1997) 277:179–185
84. Nihongaki Y, Kawano F, Nakajima T, Sato M (2015) Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nat Biotechnol 33(7):755–760
85. Niks D, Hilario E, Dierkers A, Ngo H, Borchardt D, Neubauer TJ, Fan L, Mueller LJ, Dunn MF. Allostery and substrate channeling in the tryptophan synthase bienzyme complex: evidence for two subunit conformations and four quaternary states. pp 1520–4995 (Electronic)
86. Niopek D, Benzinger D, Roensch J, Draebing T, Wehler P, Eils R, Di Ventura B (2014) Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. Nat Commun 5:4404
87. Nordlund HR, Hytonen VP, Laitinen OH, Uotila ST, Niskanen EA, Savolainen J, Porkka E, Kulomaa MS (2003) Introduction of histidine residues into avidin subunit interfaces allows pH-dependent regulation of quaternary structure and biotin binding. FEBS Lett 555(3):449–454
88. Nussinov R, Tsai CJ (2015) Allostery without a conformational change? Revisiting the paradigm. Curr Opin Struct Biol 30:17–24
89. O'Rourke KF, Gorman SD, Boehr DD. Biophysical and computational methods to analyze amino acid interaction networks in proteins. pp 2001–0370 (Print)
90. Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, Savage DF (2016) Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. Nat Biotechnol 34(6):646–651

91. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. Genetics 120(3):621–623
92. Olsson U, Wolf-Watz M (2010) Overlap between folding and functional energy landscapes for adenylate kinase conformational change. Nat Commun 1:111
93. Pande VS, Beauchamp K, Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask. Methods (San Diego, Calif) 52(1):99–105
94. Petit CM, Zhang J, Sapienza PJ, Fuentes EJ, Lee AL (2009) Hidden dynamic allostery in a PDZ domain. Proc Natl Acad Sci U S A 106(43):18249–18254
95. Pierre B, Shah V, Xiao J, Kim JR. Construction of a random circular permutation library using an engineered transposon. pp 1096–0309 (Electronic)
96. Pittolo S, Gomez-Santacana X, Eckelt K, Rovira X, Dalton J, Goudet C, Pin JP, Llobet A, Giraldo J, Llebaria A, Gorostiza P (2014) An allosteric modulator to control endogenous G protein-coupled receptors with light. Nat Chem Biol 10(10):813–815
97. Popovych N, Sun S, Ebright RH, Kalodimos CG (2006) Dynamically driven protein allostery. Nat Struct Mol Biol 13(9):831–838
98. Putri RM, Zulfikri H, Fredy JW, Juan A, Tananchayakul P, Cornelissen J, Koay MST, Filippi C, Katsonis N (2018) Photoprogramming Allostery in human serum albumin. Bioconjug Chem 29(7):2215–2224
99. Qin BY, Bewley MC, Creamer LK, Baker HM, Baker EN, Jameson GB (1998) Structural basis of the Tanford transition of bovine beta-lactoglobulin. Biochemistry 37(40):14014–14023
100. Ramanathan RK, McDonough SL, Kennecke HF, Iqbal S, Baranda JC, Seery TE, Lim HJ, Hezel AF, Vaccaro GM, Blanke CD (2015) Phase 2 study of MK-2206, an allosteric inhibitor of AKT, as second-line therapy for advanced gastric and gastroesophageal junction cancer: a SWOG cooperative group trial (S1005). Cancer 121(13):2193–2197
101. Renner C, Moroder L (2006) Azobenzene as conformational switch in model peptides. Chembiochem 7(6):868–878
102. Ribeiro LF, Nicholes N, Tullman J, Ribeiro LF, Fuzo CA, Vieira DS, Furtado GP, Ostermeier M, Ward RJ (2015) Insertion of a xylanase in xylose binding protein results in a xylose-stimulated xylanase. Biotechnol Biofuels 8:118
103. Ribeiro LF, Tullman J, Nicholes N, Silva SR, Vieira DS, Ostermeier M, Ward RJ (2016) A xylose-stimulated xylanase-xylose binding protein chimera created by random nonhomologous recombination. Biotechnol Biofuels 9:119
104. Richman DE, Majumdar A, Garcia-Moreno EB (2015) Conformational reorganization coupled to the ionization of internal Lys residues in proteins. Biochemistry 54(38):5888–5897
105. Richter F, Fonfara I, Bouazza B, Schumacher CH, Bratovic M, Charpentier E, Moglich A (2016) Engineering of temperature- and light-switchable Cas9 variants. Nucleic Acids Res 44(20):10003–10014
106. Rockwell KLA, (24 Nov 2016) Allosterism in drug discovery
107. Sarkar CA, Lowenhaupt K, Horan T, Boone TC, Tidor B, Lauffenburger DA (2002) Rational cytokine design for increased lifetime and enhanced potency using pH-activated "histidine switching". Nat Biotechnol 20(9):908–913
108. Schierling B, Noel AJ, Wende W, Hien LT, Volkov E, Kubareva E, Oretskaya T, Kokkinidis M, Rompp A, Spengler B, Pingoud A. Controlling the enzymatic activity of a restriction enzyme by light. pp 1091–6490 (Electronic)
109. Schonberger M, Trauner D (2014) A photochromic agonist for mu-opioid receptors. Angew Chem Int Ed Engl 53(12):3264–3267
110. Seifert S, Brakmann S (2018) LOV Domains in the design of photoresponsive enzymes. ACS Chem Biol 13(8):1914–1920
111. Selvaratnam R, Chowdhury S Schouwen B, Melacini G. Mapping allostery through the covariance analysis of NMR chemical shifts. pp 1091–6490 (Electronic)
112. Shah V, Kim JR (2016) Transposon for protein engineering. Mob Genet Elem 6(6):e1239601

113. Shah V, Pierre B. Fau-Kim JR, Kim JR. Facile construction of a random protein domain insertion library using an engineered transposon. pp 1096–0309 (Electronic)
114. Spiltoir JI, Strickland D, Glotzer M, Tucker CL (2016) Optical control of Peroxisomal trafficking. ACS Synth Biol 5(7):554–560
115. Srivastava J, Barber DL, Jacobson MP (2007) Intracellular pH sensors: design principles and functional significance. Physiology (Bethesda, Md) 22:30–39
116. Stauffer ME, Chazin WJ (2004) Structural mechanisms of DNA replication, repair, and recombination. J Biol Chem 279(30):30915–30918
117. Strickland D, Moffat K, Sosnick TR (2008) Light-activated DNA binding in a designed allosteric protein. Proc Natl Acad Sci U S A 105(31):10709–10714
118. Swails JM, Roitberg AE. Enhancing conformation and protonation state sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. pp 1549–9618 (Print)
119. Swint-Kruse L, Matthews KS (2009) Allostery in the LacI/GalR family: variations on a theme. Curr Opin Microbiol 12(2):129–137
120. Tillotson BJ, Goulatis LI, Parenti I, Duxbury E, Shusta EV (2015) Engineering an anti-transferrin receptor ScFv for pH-sensitive binding leads to increased intracellular accumulation. PLoS One 10(12):e0145820
121. Tsai YT, Chuang MJ, Tang SH, Wu ST, Chen YC, Sun GH, Hsiao PW, Huang SM, Lee HJ, Yu CP, Ho JY, Lin HK, Chen MR, Lin CC, Chang SY, Lin VC, Yu DS, Cha TL (2015) Novel Cancer therapeutics with allosteric modulation of the mitochondrial C-Raf-DAPK complex by Raf inhibitor combination therapy. Cancer Res 75(17):3568–3582
122. Ueda T, Murayama K, Yamamoto T, Kimura S, Imanishi Y (1994) Photo-regulation of hydrolysis activity of semisynthetic mutant phospholipases A2 replaced by non-natural aromatic amino acids. J Chem Soc 2(2):225–230
123. Vishveshwara S, Brinda K, Nkjjo T, Chemistry C (2002) Protein structure: insights from graph theory. J Theoret Comput Chem 1(01):187–211
124. Volgraf M, Gorostiza P, Szobota S, Helix MR, Isacoff EY, Trauner D (2007) Reversibly caged glutamate: a photochromic agonist of ionotropic glutamate receptors. J Am Chem Soc 129(2):260–261
125. Wang H, Vilela M, Winkler A, Tarnawski M, Schlichting I, Yumerefendi H, Kuhlman B, Liu R, Danuser G, Hahn KM (2016) LOVTRAP: an optogenetic system for photoinduced protein dissociation. Nat Methods 13(9):755–758
126. Webby CJ, Jiao W, Hutton RD, Blackmore NJ, Baker HM, Baker EN, Jameson GB, Parker EJ (2010) Synergistic allostery, a sophisticated regulatory network for the control of aromatic amino acid biosynthesis in Mycobacterium tuberculosis. J Biol Chem 285(40):30567–30576
127. Wehler P, Niopek D, Eils R, Di Ventura B (2016) Optogenetic control of nuclear protein import in living cells using light-inducible nuclear localization signals (LINuS). Curr Protoc Chem Biol 8(2):131–145
128. Willner I, Rubin S, Riklin A (1991) Photoregulation of papain activity through anchoring photochromic azo groups to the enzyme backbone, vol 113
129. Wilson CJ, Zhan H, Swint-Kruse L, Matthews KS (2007) The lactose repressor system: paradigms for regulation, allosteric behavior and protein folding. Cell Mol Life Sci 64(1):3–16
130. Wong S, Mosabbir AA, Truong K (2015) An engineered Split Intein for Photoactivated protein trans-splicing. PLoS One 10(8):e0135965
131. Wright CM, Majumdar A, Tolman JR, Ostermeier M (2010) NMR characterization of an engineered domain fusion between maltose binding protein and TEM1 beta-lactamase provides insight into its structure and allosteric mechanism. Proteins 78(6):1423–1430
132. Wright CM, Wright RC, Eshleman JR, Ostermeier M (2011) A protein therapeutic modality founded on molecular regulation. Proc Natl Acad Sci U S A 108(39):16206–16211
133. Wu YI, Frey D, Lungu OI, Jaehrig A, Schlichting I, Kuhlman B, Hahn KM (2009) A genetically encoded photoactivatable Rac controls the motility of living cells. Nature 461(7260):104–108

134. Xu J, Sarma AVS, Wei Y, Beamer LJ, Van Doren SR (2017) Multiple ligand-bound states of a Phosphohexomutase revealed by principal component analysis of NMR peak shifts. Sci Rep 7(1):5343
135. Yamada MD, Nakajima Y, Maeda H, Maruta S (2007) Photocontrol of kinesin ATPase activity using an azobenzene derivative. J Biochem 142(6):691–698
136. Yates FE, Iberall ASJAOBE (1973) Chance and necessity: an essay on the natural philosophy of modern biology by Jacques Monod. Ann Biomed Eng 1(3):381–384
137. Yi JJ, Wang H, Vilela M, Danuser G, Hahn KM (2014) Manipulation of endogenous kinase activity in living cells using photoswitchable inhibitory peptides. ACS Synth Biol 3(11):788–795
138. Yumerefendi H, Lerner AM, Zimmerman SP, Hahn K, Bear JE, Strahl BD, Kuhlman B. Light-induced nuclear export reveals rapid dynamics of epigenetic modifications. pp 1552–4469 (Electronic)
139. Zhang Y, Kitazawa S, Peran I, Stenzoski N, McCallum SA, Raleigh DP, Royer CA (2016) High pressure ZZ-exchange NMR reveals key features of protein folding transition states. J Am Chem Soc 138(46):15260–15266
140. Zimenkov Y, Dublin SN, Ni R, Tu RS, Breedveld V, Apkarian RP, Conticello VP (2006) Rational design of a reversible pH-responsive switch for peptide self-assembly. J Am Chem Soc 128(21):6770–6771