Eur. Phys. J. B (2021) 94:203
https://doi.org/10.1140/epjb/s10051-021-00207-7

THE EUROPEAN PHYSICAL JOURNAL B

Topical Review - Computational Methods

Tyrosine kinases: complex molecular systems challenging computational methodologies

Trayder Thomas[1,2] and Benoît Roux[1,2,a]

[1] Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA
[2] Gordon Center for Integrative Science, 929 E 57th Street, Room W323B, Chicago, IL 60637, USA

Received 22 April 2021 / Accepted 14 September 2021 / Published online 11 October 2021

Abstract. Classical molecular dynamics (MD) simulations based on atomic models play an increasingly important role in a wide range of applications in physics, biology, and chemistry. Nonetheless, generating genuine knowledge about biological systems using MD simulations remains challenging. Protein tyrosine kinases are important cellular signaling enzymes that regulate cell growth, proliferation, metabolism, differentiation, and migration. Due to the large conformational changes and long timescales involved in their function, these kinases present particularly challenging problems to modern computational and theoretical frameworks aimed at elucidating the dynamics of complex biomolecular systems. Markov state models have achieved limited success in tackling the broader conformational ensemble and biased methods are often employed to examine specific long timescale events. Recent advances in machine learning continue to push the limitations of current methodologies and provide notable improvements when integrated with the existing frameworks. A broad perspective is drawn from a critical review of recent studies.

Introduction

Classical molecular dynamics (MD) simulations based on atomic models play an increasingly important role in a wide range of applications in physics, biology, and chemistry. The approach consists of constructing detailed atomic models of the macromolecular system and, having described the microscopic forces with a potential function, using Newton's classical equation to literally "simulate" the dynamical motions of all the atoms as a function of time. The calculated trajectory, though an approximation to the real world, provides detailed information about the time course of the atomic motions, which is impossible to access experimentally. While great progress has been made, producing genuine knowledge about biological systems using MD simulations remains challenging. Among the most difficult problems are the characterization of large conformational changes occurring over long timescales that impact biological function.

In the last decade, a powerful theoretical and computational paradigm for studying complex biomolecular systems has emerged. Building on classical statistical mechanical techniques such as alchemical free energy perturbation (FEP) [1,2,4] and umbrella sampling (US) in multiple dimensions [5,6], it consists in combining various strategies including, transition pathway determination [7–11] and stochastic Markov modeling [12–14], together with time-independent component analysis (TICA) [15,16] and kinetic mapping [15]. Theoretical frameworks relying on transition path theory (TPT) [17–20], and on a variational formulation of the transfer operator [21–24], serve as foundations to expand many of these novel ideas toward various data-driven machine learning (ML) techniques [24–30].

Protein tyrosine kinases offer particularly challenging systems to advance and put to the test this emerging computational framework. After providing some background on the structure and function of kinases, we will broadly review previous computational studies of these important systems and highlight the main findings. We will conclude with an outlook on the considerable challenges that remain and the promise of the most recent theoretical advances in the treatment of biomolecular dynamical systems.

Some background on protein tyrosine kinases

Protein kinases are enzymes that catalyze the covalent transfer of the γ-phosphate of an adenosine triphosphate (ATP) molecule onto a tyrosine, serine, threonine, or histidine residue on targeted and specific substrate proteins (including kinases), thereby sending a "downstream" chemical signal throughout a network of proteins. Phosphorylation of a substrate protein can result in a functional change and modification of cellular localization, with a potential impact on numerous cellular processes. The human genome encodes more

[a] e-mail: roux@uchicago.edu (corresponding author)

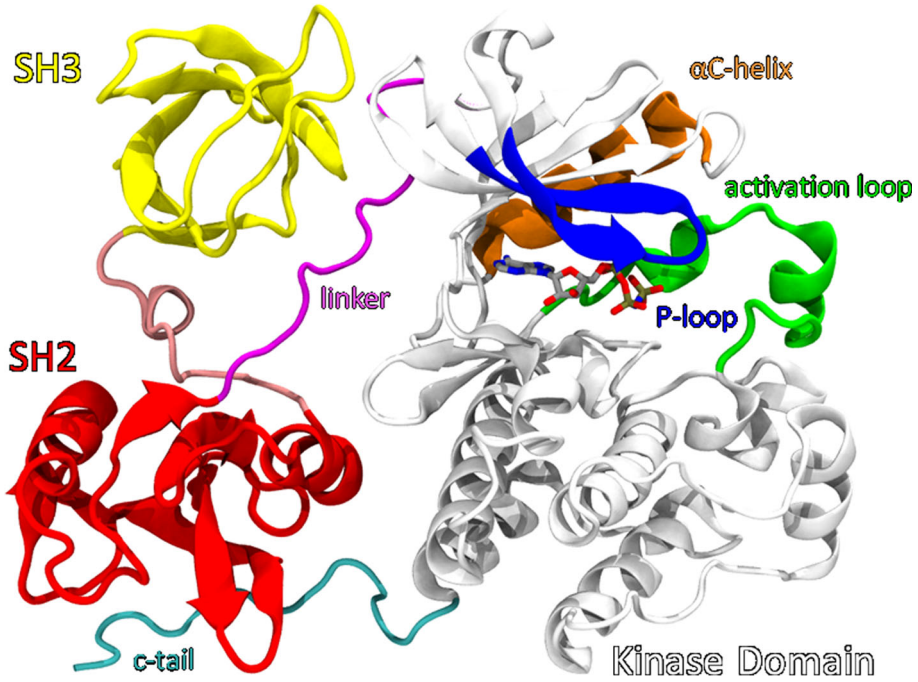

Fig. 1 c-Src tyrosine kinase in the autoinhibitory assembled form (PDB 2SRC). The regulatory domains SH3 (yellow) and SH2 (red) are shown along with the linker (purple) to the kinase domain (white). Important structural features of the kinase domain have also been highlighted: P-loop (blue), activation loop (green), αC-helix (orange), and c-tail (cyan). An analogue of ATP (grey sticks) is also shown bound to the kinase domain

than 500 protein kinases, making the kinome one of the largest gene families [31]. Kinases are important cellular signaling enzymes that regulate cell growth, proliferation, metabolism, differentiation, and migration. Unregulated kinase activity and loss of function are often involved in a wide range of human diseases. For this reason, these enzymes are major therapeutic targets, and the discovery of kinase-specific inhibitors has been intensely pursued by the pharmaceutical industry for many years [32–37]. Kinases are highly validated drug targets with a vast amount of available structural data.

The members of the Src family of protein tyrosine kinases represent prototypical model systems of great interest to study the regulation mechanisms. All nine members of the Src family (Src, Yes, Fyn, Lyn, Lck, Blk, Hck, Fgr, and Yrk) are highly homologous proteins with a similar regulatory mechanism sharing a multidomain architecture illustrated in Fig. 1 that comprises a catalytic tyrosine kinase domain (grey), preceded by two peptide-binding regulatory modules, the Src-homology domains SH2 (red) and SH3 (yellow) [38]. The kinase members of the Src family are allosteric molecular switches. Their catalytic activity can be modulated in response to cellular signals. Phosphorylation of the protein kinase is especially crucial to the regulation of the catalytic activity.

Molecular dynamics studies

Molecular dynamics (MD) simulations provide one important avenue to better understand protein kinases. Many computational studies focused on the conformational transition taking place within the catalytic domain, which is from an inactive to its catalytically competent active state [39,40]. Characterizing the properties of the isolated catalytic domain without the SH2 and SH3 regulatory domains has also been of considerable interest [41], as it is known to be constitutively active in solution according to experiment [39]. Smaller conformational transitions have also attracted interest, in particular, a 3-residue motif comprised of Asp-Phe-Gly (DFG) near the N-terminus of the activation (A–) loop covering the catalytic site is involved in the activation process and its conformation is known to be critical for binding some kinase inhibitors [42–44]. An understanding of how the conformational ensemble of kinases is influenced by ligand binding, phosphorylation, autoinhibition, and drug resistant mutations is of keen interest, and there is a long history of MD studies investigating the effects of these phenomena on the conformational landscape of kinases. In one of the earliest studies, Young et al. examined the effect of mutations on the structural fluctuations within the multidomain c-Src kinase [45]. Suenaga et al. observed the effect of phosphorylation on the structural flexibility of Shc kinase [46]. Kuriyan and co-workers observed a Src-like inactive conformation in the Abl tyrosine kinase domain [47]. Dixit et al. examined the activation mechanisms in the catalytic domain of Abl and EGFR kinases [48]. In time, computational studies became ever more ambitious. Cambran et al. studied the effect of myristoylation, phosphorylation, and ligand binding on the conformational ensemble of protein kinase A

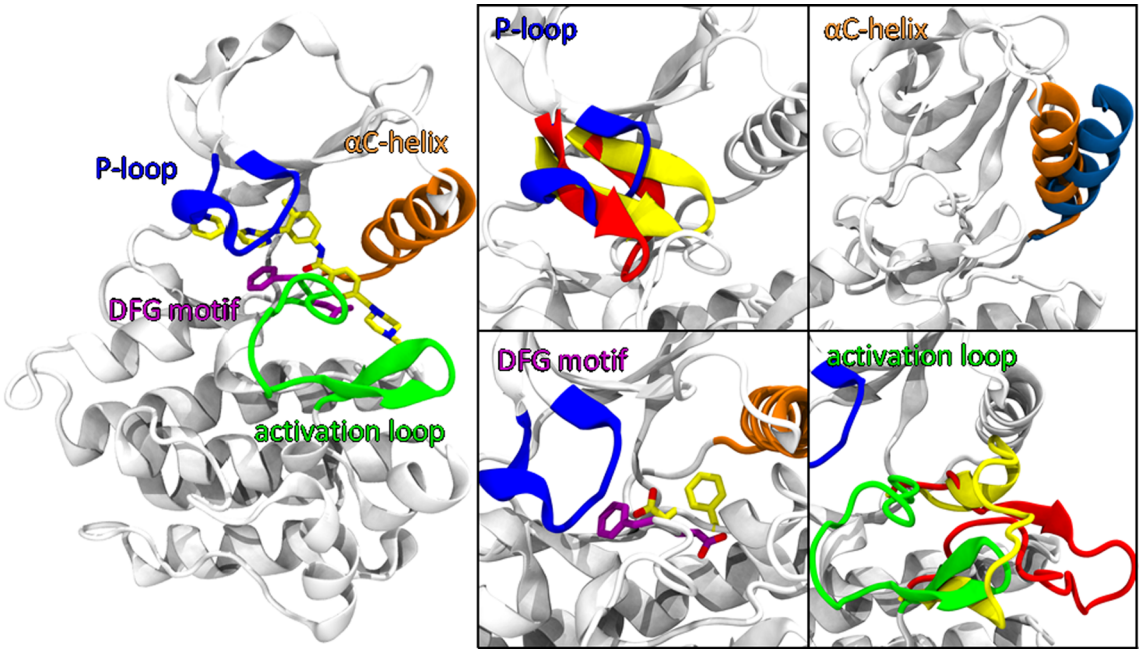

Fig. 2 Four main structural elements in Abl kinase. (Left) Crystal structure of Abl kinase (PDB 2HYY) highlighting the P-loop (blue), DFG motif (purple), activation loop (green), αC-helix (orange), and the bound ligand imatinib (yellow). (Right) Alternative conformations of structural elements from three Abl kinase experimental structures (PDB 2HYY, 2G1T, 6XR6), illustrating the broad conformational space of the binding site

(PKA) [49,50], whilst Boras et al. conducted a similar study focusing on the autoinhibition by the regulatory domains [51]. Lopez et al. used an array of methods to assess the impact of phosphorylation on the conformation of ERK2 [52]. In a study of the Janus kinase 2 (JAK2) tyrosine kinase [53], the DESRES team led by David Shaw used unbiased MD simulations to investigate autoinhibition. Similar characterizations have also been conducted on the conformation of the epidermal growth factor receptor (EGFR) and its changes due to dimerization and mutations [54–59]. Yan et al. studied the hepatocyte growth factor receptor (HGFR, also known as c-Met) [60], and found that the allosteric inhibitor tivantinib caused a transition from the DFG-in to DFG-out state in a sub-microsecond timescale allowing them to observe the DFG-flip during several 250 ns simulations.

Conformation of the DFG motif

As mentioned above, the orientation of the small DFG motif plays a role in the inhibition mechanism of the Abelson tyrosine kinase (Abl) by the anticancer drug imatinib [43]. Imatinib displays a much lower inhibitory effect on c-Src, even though these two kinases display a high level of sequence identity (47%) and similar structural features [43,61,62]. The similarity of binding site residues between the two kinases, combined with the idea that imatinib binds exclusively to a DFG-out conformation, led to the view that there was a "conformational selection" mechanism of binding specificity. But the crystal structure of imatinib in complex with the c-Src kinase domain subsequently showed that it also adopted the same inactive DFG-out conformation [61,63,64]. This led to the view that there must be some "thermodynamic penalty" associated with the DFG-out conformation [65]. Because of its great importance, the conformational transition of the DFG motif and its impact on the binding specificity of inhibitors has been the subject of many computational studies [2,60,62,66–70]. Researchers from DESRES characterized the conformation of Abl kinase [62], identifying multiple factors that contributed to the stability of the DFG-in or DFG-out poses, including a pH dependence in the conformation of the DFG motif that they attributed to electrostatic changes during the catalytic cycle. An analysis of the kinetics of imatinib binding in this same study reinforced that imatinib selectively bound to the DFG-out conformation and estimated the timescale of the DFG-flip to be in the DFG-flip to be in the order of 10 s of milliseconds.

The DFG-flip conformational transition and its thermodynamic impact specifically regarding the selectivity of imatinib has also been the subject of several computational studies [2–4,48,62,66,68–72]. Lovera et al. calculated a free energy landscape of the DFG-flip transition in c-Abl and c-Src using meta-dynamics MD simulations with the AMBER force field [68], yielding free energy differences 4.0 and 6.0 kcal/mol in Abl and c-Src, respectively. Roux and co-workers estimated the same quantity using umbrella sampling with the CHARMM force field and obtained 1.4 and 5.4 kcal/mol in Abl and c-Src, respectively [2,69]. Furthermore, computations based on alchemical free energy

perturbation demonstrate that different hidden factors (e.g. DFG-flip, intrinsic affinity, steric clashes with the P-loop) that are not readily detectable by experiments partly cancel one another to explain the observed K_d of imatinib and of a related inhibitor G6G [4]. Lovera et al. investigated the free energy landscape of several widespread drug-resistant mutations of Abl kinase, finding a category of mutations that reduced imatinib affinity by favoring the DFG-in conformations of Abl [67]. All these studies indicate that the DFG-flip can happen for both Abl and c-Src, but the latter simply incurs a larger "thermodynamic penalty". Such findings are not unique to Abl, and similar studies have been conducted across the kinome [73,74]. Nonetheless, the issue is not settled, as an alternative model to explain these observations was recently proposed by Kern and co-workers, who found from their kinetic experiments that imatinib bound in a two-step process which differed significantly in kinetics between Abl and Src [75].

Activation pathways and string method

Because the timescales for the conformational changes leading to activation and inactivation are very long, many of the early studies relied on some type of non-equilibrium perturbation to accelerate transitions of interest, such as targeted-MD driven by the root-mean-square deviations [45,76–79], the half quadratic biasing for positive selection of spontaneous fluctuations [80–82], or chemical perturbation (protonation of titratable residues) [62]. While these efforts provided some useful mechanistic insight, the path generated by these biased methods must be considered with caution because it generally depends on the choice of progress variable used to move the system between the initial and final states [83].

While such driven (biased) simulations based on non-equilibrium trahectories can generate useful information to characterize conformational transition pathways, they are often insufficient to tackle slow conformational transitions occurring on long timescales. These issues are resolved with the more rigorous computational framework provided by the string method [7,8]. The string method represents the pathway as a chain of M "images" (copies of the full system) in the multidimensional space of the set of collective variables $\mathbf{Z}$, e.g., the pathway is given as $\{Z^{(1)}, Z^{(2)}, \ldots, Z^{(M)}\}$. In the string method with swarms-of-trajectories [9], the transition pathway is iteratively refined starting from an initial guess by using the mean drift of short unbiased simulations launched from each of the M images. Because all these simulations are uncorrelated, the procedure can be efficiently scaled up to a very large number of CPUs using the general multiple-copy algorithm framework implemented in NAMD [84]. It is also possible to accelerate the convergence of the iterative refinement using a multi-scale pre-conditioner [85].

The first application of the string method with swarms-of-trajectories was to determine the activation pathway in the isolated catalytic domain of Src kinase [86]. A subsequent study reported a very ambitious characterization of the activation pathway of the catalytic domain in the presence of the SH2 and SH3 regulatory domains [87]. Alternative approaches to the string method have also been used to study conformational transition in kinases, including the adaptively biased path optimization (ABPO) method to determine the principal curve defining a conformational transition between two known end states [10]. Post and co-workers used ABPO to examine the transitions within the catalytic domain of c-Src and CDK2 kinase [88–90]. This analysis highlighted the importance of a previously identified switched electrostatic network [81,82] with respect to the αC helix and the activation loop motions [88–90]. The string method was also used to determine the transition pathway of the DFG-flip [69]. Another pathway algorithm developed by Levy and Elber, the "Rock-Climbing" method, was to characterize the DFG-flip transition in Abl kinase [11,91].

Umbrella sampling and free energy landscape

Umbrella sampling simulations were used to calculate the potential of mean force (PMF) along selected collective coordinates to examine the free energy landscape associated with various processes [10,69,77,87,92–95]. Studies included the conformational equilibrium of the N-terminus of the catalytic domain [77], the DFG-flip [69], and the complete transition from the inactive to the active state in the catalytic domain [10,87,92,93]. In a number of cases, the transition pathway was exploited as a basis to map the free energy landscape for these conformational changes. In one of the earliest efforts, Yang et al. constructed a low resolution free energy landscape for the activation of Hck kinase from unbiased equilibrium MD simulations initiated along a TMD transition pathway, revealing the existence of intermediate conformations along the activation pathway [92]. Sampling strategies relying on the curvilinear transition pathways also include the Milestoning method [96], with a recent application to the DFG-flip transition in Abl kinase [91]. To enhance the sampling, many of these PMF calculations were carried out using a multiple-copies replica-exchange algorithm [84]. The ABPO method was used compute the PMF along the curvilinear activation transition pathway of Src [90]. Metadynamics is also a possible avenue that has been used to enhance the sampling capability of MD simulations in studies of kinases [54,68,97].

Alchemical FEP

Alchemical free energy perturbation (FEP) MD simulations (FEP/MD) is a method that can be used to calculate the equilibrium binding free energy of ligands to their target receptor [98]. The method is generally considered to be accurate to approximately 1 kcal/mol, and the cancellation of errors between similar binding sites has been shown to result in increased accuracy when predicting selectivity for kinases [99]. Alchemical

FEP/MD simulations were used to examine the binding free energy of kinase inhibitors [2–4]. While these calculations are mainly focused on the binding affinity of a ligand to a given protein conformation, the critical issue of conformational diversity cannot be avoided, even if it does not appear directly in alchemical FEP/MD simulations. For example, the "in" or "out" conformation of the DFG motif is critical to ascertain the binding specificity of imatinib to Abl and c-Src kinases [2]. The general take-home message of alchemical FEP/MD is that the method is useful, but the underlying issue of conformational diversity must be resolved, at least partially, in order to draw meaningful conclusions.

Coarse-grained models

In part to circumvent the nearly insurmountable sampling problems associated with a large, complex protein, a number of studies have resorted to a simpler coarse-grained (CG) representation of the kinase. Yang and Roux used a multistate coarse-grained Gō-like model to simulate the activation of the catalytic domain of Hck kinase [100]. A similar multistate Gō-like model of Lyn kinase was exploited to search the optimal path for the same transition using the maximum flux transition path (MFTP) method [88]. Simplified CG models have also been used to test algorithms to determine conformational pathways [10], and examine the computational feasibility of Markov state models (MSMs) [100]. While it is clear that detailed all-atom models are more realistic, simplified CG models have played an important role to test and illustrate new algorithms in the context of kinase structures. This type of CG model also played a useful role in the interpretation of experimental data from small angle X-ray scattering (SAXS) of Hck kinase in solution [92,101].

Markov state models

Markov state models (MSMs) provide a powerful computational strategy combining the information accumulated from a large number of MD simulations [12]. Building an MSM involves splitting up a system's conformational space into a set of microstates, running independent trajectories visiting each of those regions, and then tabulating the observed state-to-state transitions into a Markov matrix of transition probabilities [12–14,102–104]. Traditionally, MSMs have been constructed from aggregate MD simulation data from multiple trajectories. With the tremendous progress in accelerating classical MD simulations on inexpensive graphics processing units (GPUs) [105–108] the simulation of tens of thousands of short, independent trajectories can be executed with relative ease using OpenMM [105–107] or AMBER-GPU [108]. Sophisticated analysis software packages MSMBuilder (http://msmbuilder.org) [13,109] and PyEMMA [110] play an important role for a sustained progress in the field. The approach was used to characterize the activating transition in the catalytic domain of c-Src kinase [94]. The qualitative conclusions confirmed the existence of a putative intermediate state along the activation pathway, as previously revealed [92]. The MSMs thus constructed, enabled a transition path theory (TPT) [19] analysis of the nature of the "reaction tube" supporting the conformational change for the activation of c-Src [111]. Sultan et al. used MSMs to analyze 1.7 ms of unbiased simulations of BTK kinase to construct a MSM that captured the DFG-flip [112]. This methodology was later expanded with a further aggregate 4 ms of MD data across 6 Src-family kinases that was combined into single MSM, establishing a common conformational ensemble [113].

It is clear that the large number of conformations accessible to kinases poses a special challenge to efforts aimed at designing inhibitors binding to a given kinase with high specificity and affinity. Conventional docking algorithms can produce native like complex structures when a number of conditions are satisfied (e.g., if a deep binding pocket has a single dominant conformation) [114]. However, they often fall short when a protein is flexible and can adopt multiple conformations. Analysis based on a single non-dominant structure, whether it is from X-ray crystallography or from homology modeling, can lead to an erroneous SBDD strategy. For kinases structural elements near the binding pocket (A-loop, P-loop, DFG motif, catalytic spine, etc.) are able to adopt multiple conformations (see Fig. 2). These structural elements affect the shape and properties of the binding pocket. As a consequence, it becomes highly difficult to ascertain which of those conformations should be considered in a drug discovery project. For example, the DFG motif can point inward (DFG-in) or outward (DFG-out), which strongly affects the type of inhibitors that can bind to the pocket; the A-loop and P-loop can adopt a variety of conformations, creating a large number of combinatorial possibilities. In an effort to try to consider multiple conformations, one can try to harvest the possible conformations from similar proteins in the PDB, or by computational sampling (such as Rosetta) to generate alternative structures. Nevertheless, such knowledge-based heuristic approaches are not guaranteed to discover all accessible conformations of the various structural elements for the protein of interest. More importantly, it does not provide information about the relative importance of multiple different conformations of the target protein (the frequency of conformations in the PDB is not a measure of energetic stability). In large part to explore this issue, the MSM framework was also used to examine the conformational diversity of Abl kinase in the absence of any ligand [115]. The efforts showed that it is feasible to discover and identify most of the most relevant conformations of the catalytic domain of a kinase, although ranking them in terms of their equilibrium population is harder and more uncertain. While these states are likely to be metastable, a characterization of the kinetic timescales is extremely difficult [116]. Even with a little over 100 μs of aggregate MD data, the eigenvalues of the transition matrix remain highly uncertain. Similar issues are also present when trying

to characterize the binding of a specific ligand such as imatinib to Abl kinase [117].

Relating computational studies to functional assays

Relating the results of computational studies to experimental observation about kinases also presents some special challenges. The activity of signaling enzymes is often measured indirectly via cellular assays that reveal small but biologically meaningful trends on macroscopic timescales [45]. For this reason, validating the results of computational studies with respect to kinase function is not always straightforward. Generally, it is clear that microscopic factors that stabilize the down-regulated inactive conformation of the kinase should reduce activity [40]. Correspondingly, factors that destabilize the inactive state should appear as increased kinase activity in experimental assays. But this naïve understanding rapidly encounters difficulties. For example, the isolated catalytic domain (without its regulatory domains) is known to be constitutively active according to experiments [39]. Presumably, the active catalytic domain is trans-phosphorylated via a bimolecular encounter with another kinase in order to stabilize the fully active state. Indeed, the calculated free energy landscape shows that phosphorylation of Y416 in the A-loop essentially "locks" the kinase into its catalytically competent conformation [93]. Nonetheless, while this conceptual framework is reasonable, it is unclear whether it can be reconciled with experimental observations: the timescale of the interconversion between the inactive and active state from the MSM is on the order of 100 μs whereas the experimentally observed timescale for Src-family tyrosine kinase autophosphorylation at residue Tyr416 is on the order of minutes. It was possible to relate these vastly disparate timescales by constructing a simple kinetic model using the data extracted from atomistic simulations as well as a reasonable estimate for the bimolecular rate of kinase phosphorylation [94]. A similar conceptual framework was used to explain the functional behavior of the activating W260A mutant [118]. At first glance, the computational result seems to be inconsistent with experiments: while the W260A mutation in c-Src is markedly up-regulated [119], the free energy cost to reach the active-like unphosphorylated state A is only slightly smaller in the W260A mutant compared with WT [118]. Nonetheless, such a small difference in stability accounts for the large increase in the activity of the mutant that is observed experimentally [118]. Furthermore, umbrella sampling computations indicate that the configuration of the SH2 and SH3 regulatory domains, greatly affects the ability of the catalytic domain to adopt the active or inactive conformation [87]. Conversely, individual inhibitors binding to the catalytic domain can have differing allosteric effects on whether the regulatory domains adopt a closed or open conformation [120]. Computational methods based on simplified CG models have played a useful role to interpret SAXS data on Hck kinase in solution in terms of the configuration of the multi-domain kinase [92,101]. For instance, MSM one can observe the organization of the SH2, SH3, and catalytic domain in response to different external factors [92], or mutations [101]. This type of strategy has also been useful to interpret data about the membrane anchoring tail of Hck kinase from neutron reflectometry [121]. Xie et al. used NMR to identify conformational states occupied by Abl kinase and were then able to fit their chemical exchange data to a model in order to predict populations and kinetics of those states for the wildtype and various mutants [122]. MD simulations have also been important to develop a perspective on substrate recognition integrating information from NMR [123].

Binding of inhibitors

There is also considerable interest in using computational methods to understand experimental data about the binding of specific kinase inhibitors. One of the most studied prototypical systems in this case is the binding of imatinib and how it is affected by the conformation of the DFG motif [43,75,124–126]. The tryptophan fluorescence quenching experiments of Agafonov et al. [75] revealed that the ligand imatinib (L) binds to the Abl kinase protein (P) via a two-step process: $P + L \rightarrow P \sim L \rightarrow PL$. While the first step is a concentration-dependent bimolecular association leading to the formation of a protein-ligand intermediate, the second step corresponds to a slow concentration-independent unimolecular rearrangement that ultimately leads this intermediate to the strongly bound protein-ligand complex. The authors interpreted this as evidence for an induced-fit conformational change of the protein kinase that is directly responsible for the binding specificity. However, an extensive MSM study indicates that there exists a large number of metastable states corresponding to non-specific association of imatinib with Abl kinase along two main pathways leading imatinib to the binding pocket [117]. The MSM analysis based on more than 500 μs of aggregate MD data strongly supports the notion that the interconversion from the metastable binding modes of the ligand to the fully bound state is extremely slow, suggesting that the set of metastable binding modes of imatinib could reasonably give rise to the long-lived intermediate state revealed by the tryptophan fluorescence data [75]. In that sense, the picture emerging from the MSM analysis is consistent with the actual experimental observation of a slow concentration-independent unimolecular rearrangement leading to the final complex. This picture departs from the traditional view of the induced-fit mechanism, where an initial weakly bound protein-ligand complex induces some conformational changes in the protein that leads to a strongly bound protein-ligand complex, while the computational results rather point to the slow rearrangement of the ligand toward its correct binding pocket from a multitude of long-lived metastable states that underlies the slow step detected in the fluorescence experiment. More generally, kinases

are viewed as highly flexible allosteric proteins with multiple moving parts that can adopt different conformations, which clearly impacts on our ability to design small molecule inhibitors binding with high specificity. In Fig. 2, four structural elements in Abl kinase displaying multiple conformational variations—the P-loop (blue), the DFG motif (purple), the A-loop (green), and the αC-helix (orange)—are illustrated. Those conformational variants affect the binding pocket. While a large amount of attention has been devoted to the features of the isolated catalytic domain by itself, understanding the influence of the regulatory SH2 and SH3 domains is tremendously important for the biological cellular signaling function [87,101,127], and membrane co-localization [121].

Overall assessment of computational methods

At the present time, Markov state models (MSM) perhaps represent the richest and most powerful computational framework to characterize the dynamics of complex biomolecular systems [12–14]. Yet, in spite of a huge amount of aggregate MD data, our own efforts with kinase proteins were only partial successes [94,115–117]. One of the fundamental difficulties and an important requirement for the construction of a reliable and accurate MSM is the definition of "states", dynamically meaningful regions of conformational space that must correctly map out the behavior of the system. Several structural elements of kinases display great flexibility. Ultimately, it is important to realize that MSMs only provide a model of the explored conformational space. There are several slow processes that may not be well sampled, even with a millisecond of data. In our MSM analysis, dimensionality reduction of the full configurational space was carried out via the time-lagged independent component analysis (TICA) [15,16], applied to pre-identified structural regions displaying high conformational variability, e.g., the A-loop, the αC-helix, the DFG motif, and the P-loop [115,116]. The ideas of Taylor and co-workers to subdivide a kinase structure into contiguous "communities" that exhibit internally correlated motions provide some interesting directions worth exploring to improve the identification of relevant structural protein regions and select appropriate "features" for the MSM analysis [128]. These limitations notwithstanding, the MSM framework was undoubtedly invaluable in exploring the conformational landscape of apo-Abl kinase [115], allowing us to identify and rank the metastability of novel states. Further analysis of these states by comparing them to kinase crystal structures available in the PDB. Paul et al. [116] revealed high similarity between our identified apo-Abl conformations and ligand-bound structures across other kinase families. This analysis provided remarkable insight into both the similarities in conformational behavior across kinase families, and the druggability of our identified conformations. Nevertheless, despite the 800 μs of MD simulation used in these studies, we still fell short of reproducing the correct kinetics or thermodynamics of the molecular system. MSMs are prone to discretization error arising during the clustering steps that causes underestimation of energy, and error due to under-sampling of rare transitions that can cause overestimation of the same barriers. Shütte and collaborators have developed approaches for estimating the discretization error in MSM and address this issue [129–131].

Understanding the binding of specific kinase inhibitors is likely to remain an important objective for many years to come. In some cases, when the binding process is fairly fast, it has been possible to construct an accurate MSM from the aggregate data of unbiased MD simulations [132,133]. For example, to study the binding of trypsin and its competitive inhibitor benzamidine, Buch et al. [132]. used 495 unbiased MD of 100 ns each for a total of 49 μs. For slower binding processes where slow induced-fit conformational changes occur concertedly during the binding process, this direct strategy would not be practical. It is unlikely that brute-force MD could simulate the complete unbinding (k_{off}) of a slow kinase inhibitor such as imatinib, which takes place over $\sim$ 40 ms. From a more mundane point of view, improving the accuracy of the atomic force field will also be needed. For reliable computational studies, the force field for the ligand inhibitors must be obtained in an objective manner, using a protocol with no direct manipulation by the users, e.g. GAFF [134], CGENFF [135], or built from the general automated atomic model parameterization (GAAMP) protocol [136]. Another new frontier to tackle is the binding of covalent inhibitors, which are the focus of increased interest following the recent clinical success of drugs such as ibrutinib [137]. Properly optimized covalent inhibition offers potential for longer on target residency time and selectivity. The design of accurate computational treatments is very challenging because the action of covalent inhibitors involves a variety of microscopic processes that take place on timescales that are far beyond the reach of standard MD simulations. Furthermore, because the formation of a covalent bond is a quantum mechanical (QM) process it is necessary to go beyond classical force fields. While simulating the formation of the covalent bond requires a quantum mechanical/molecular mechanical (QM/MM) representation of the system, the rate for the process itself can naturally be incorporated into the MSM framework. In fact, MSM is arguably the only framework by which the covalent bond formation can be incorporated into the kinetics of the slower diffusion processes.

Future outlook

Many of the theoretical frameworks used to extract kinetics from MD simulations find their underpinning in transition path theory (TPT) [17–20], which considers a reaction across a rugged energy landscape via the pathways of highest flux of probability between two states. This method has been expanded upon through a vari-

ational approach to molecular kinetics [21–24], which has been more suitable to machine learning approaches and led to methods such as VAMPnets [26]. Klus et al. used an alternative kernel-based approach and achieved a competitive result [138]. Boltzmann generators are a different strategy that bypasses the required sampling of rare events by drawing statistically independent samples from a prior distribution and reversibly mapping them to a "latent space" where pathways can be linearly interpolated between minima and then mapped back into configuration space with an associated probability [29]. The length of most simulations, which is orders of magnitudes smaller than that of the processes of interest, is likely to remain a key limitation. Previous applications show that efficient sampling of rare events remain the greatest challenge in extracting useful kinetic information from MD simulations; these events are the most relevant to the kinetics but also the most likely to be undersampled. The information from unbiased and biased MD trajectories may be combined to improve the statistical robustness and convergence of the MSM, either using the dynamical weighted histogram analysis (DHAM) [139], or the discrete transition-based reweighting analysis method (dTRAM) [140,141].

Some promising advancements in this area may be provided by machine learning (ML). As an example, Wang et al. approached the sampling problem by utilizing machine learning to design a coarse-grain force field for their system [142]. The recent success of ML techniques suggests that they may be capable of approximating high-dimensional functions with controllably small errors [27,28,30,143–146]. They may assist in solving the sampling problem by combining these approaches with some aspects of importance sampling methods for data acquisition to focus the computational resources on the critical part of a problem. One approach has been to leverage information from the string method to construct models focused on individual pathways [96,147].

The problem of accurately defining a state space for the MSM also has promising solutions in machine learning. VAMPnets [26] combine the optimization of hyperparameters with the featurization, dimensionality reduction, and clustering of the system into a single machine learning pipeline, thereby eliminating many sources of error. We have applied this methodology to the imatinib-Abl system [117], and demonstrated its ability to identify metastable states of the protein, although it does not solve undersampling problems. Ung et al. used machine learning to classify the conformation of all kinase structures in the PDB and a similar approach could be applied to simulation data [148].

While these methods, particularly in the area of ML, represent a significant advancement in the amounts of simulation data we can effectively utilize, and the accuracy of the analysis we can perform, they are not yet closing the gap between the timescales of what we can access, and the events we are interested in. This is largely because the methods endeavor to be unbiased and undirected, if we wish to close the gap faster than waiting for the gradual improvement of hardware, we will need methods that are more integrated with data generation. Fortunately, there is every indication that this is an active field with a huge potential for innovation.

Acknowledgements This work is supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) through grant R01-CAO93577.

Author contributions

TT and BR wrote the paper.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This is a review, therefore no data was deposited.]

References

1. Y. Deng, B. Roux, Computations of standard binding free energies with molecular dynamics simulations. J. Phys. Chem. **113**, 2234–2246 (2009)
2. Y.L. Lin, Y. Meng, W. Jiang, B. Roux, Explaining why Gleevec is a specific and potent inhibitor of Abl kinase. Proc. Natl. Acad. Sci. U.S.A. **110**, 1664–1669 (2013)
3. Y.L. Lin, B. Roux, Computational analysis of the binding specificity of Gleevec to Abl, c-Kit, Lck, and c-Src tyrosine kinases. J. Am. Chem. Soc. **135**, 14741–14753 (2013)
4. Y.L. Lin, Y. Meng, L. Huang, B. Roux, Computational study of Gleevec and G6G reveals molecular determinants of kinase inhibitor selectivity. J. Am. Chem. Soc. **136**, 14753–14762 (2014)
5. W. Jiang, Y. Luo, L. Maragliano, B. Roux, Calculation of free energy landscape in multi-dimensions with Hamiltonian-exchange umbrella sampling on petascale supercomputer. J. Chem. Theory Comput. **8**, 4672–4680 (2012)
6. W. Wojtas-Niziurski, Y. Meng, B. Roux, S. Berneche, Self-learning adaptive umbrella sampling method for the determination of free energy landscapes in multiple dimensions. J. Chem. Theory Comput. **9**, 1885–1895 (2013)
7. E. Weinan, W. Ren, E. Eijnden, String method for the study of rare events. Phys. Rev. B **66**, 052301 (2002)
8. L. Maragliano, A. Fischer, E. Vanden-Eijnden, G. Ciccotti, String method in collective variables: minimum free energy paths and isocommittor surfaces. J. Chem. Phys. **125**, 24106 (2006)
9. A.C. Pan, D. Sezer, B. Roux, Finding transition pathways using the string method with swarms of trajectories. J. Phys. Chem. **112**, 3432–3440 (2008)
10. B.M. Dickson, H. Huang, C.B. Post, Unrestrained computation of free energy along a path. J. Phys. Chem. B **116**, 11046–11055 (2012)
11. C. Templeton, S.H. Chen, A. Fathizadeh, R. Elber, Rock climbing: a local-global algorithm to compute

minimum energy and minimum free energy pathways. J. Chem. Phys. **147**, 152718 (2017)
12. G.R. Bowman, V.S. Pande, F. Noé, An introduction to Markov state models and their application to long timescale molecular simulation. In: Advances in Experimental Medicine and Biology, vol. 797. Springer, Netherlands (2014)
13. V.S. Pande, K. Beauchamp, G.R. Bowman, Everything you wanted to know about Markov state models but were afraid to ask. Methods **52**, 99–105 (2010)
14. J.H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schutte, F. Noe, Markov models of molecular kinetics: generation and validation. J. Chem. Phys. **134**, 174105 (2011)
15. F. Noe, C. Clementi, Kinetic distance and kinetic maps from molecular dynamics simulation. J. Chem. Theory Comput. **11**, 5002–5011 (2015)
16. G. Perez-Hernandez, F. Paul, T. Giorgino, G. De Fabritiis, F. Noe, Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. **139**, 015102 (2013)
17. P. Metzner, C. Schutte, E. Vanden-Eijnden, Illustration of transition path theory on a collection of simple examples. J. Chem. Phys. 125, 084110 (2006). https://doi.org/10.1063/1.2335447
18. P. Metzner, C. Schutte, E. Vanden-Eijnden, Transition path theory for Markov jump processes. Multiscale Model. Simul. **7**, 1192–1219 (2009)
19. E. Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events. Annu. Rev. Phys. Chem. **61**, 391–420 (2010)
20. E. Vanden-Eijnden, Transition path theory. Adv. Exp. Med. Biol. **797**, 91–100 (2014)
21. F. Nuske, B.G. Keller, G. Perez-Hernandez, A.S. Mey, F. Noe, Variational approach to molecular kinetics. J. Chem. Theory Comput. **10**, 1739–1752 (2014)
22. E.H. Thiede, D. Giannakis, A.R. Dinner, J. Weare, Galerkin approximation of dynamical quantities using trajectory data. J. Chem. Phys. **150**, 244111 (2019)
23. C. Lorpaiboon, E.H. Thiede, R.J. Webber, J. Weare, A.R. Dinner, Integrated variational approach to conformational dynamics: a robust strategy for identifying eigenfunctions of dynamical operators. J. Phys. Chem. B **124**, 9354–9364 (2020)
24. A. Bittracher, R. Banisch, C. Schutte, Data-driven computation of molecular reaction coordinates. J. Chem. Phys. **149**, 154103 (2018). https://doi.org/10.1063/1.5035183
25. N.D. Conrad, M. Weber, C. Schutte, Finding dominant structures of nonreversible Markov processes. Multiscale Model. Simul. **14**, 1319–1340 (2016)
26. A. Mardt, L. Pasquali, H. Wu, F. Noe, VAMPnets for deep learning of molecular kinetics. Nat. Commun. **9**, 5 (2018)
27. F. Noe, G. De Fabritiis, C. Clementi, Machine learning for protein folding and dynamics. Curr. Opin. Struct. Biol. **60**, 77–84 (2020)
28. F. Noe, A. Tkatchenko, K.R. Muller, C. Clementi, Machine learning for molecular simulation. Annu. Rev. Phys. Chem. **71**, 361–390 (2020)
29. F. Noe, S. Olsson, J. Kohler, H. Wu, Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. Science **365**, 1001 (2019). https://doi.org/10.1126/science.aaw1147
30. A. Bittracher, C. Schutte, A probabilistic algorithm for aggregating vastly undersampled large Markov chains. Phys. D **416**, 132799 (2021). https://doi.org/10.1016/j.physd.2020.132799
31. G. Manning, D.B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. Science **298**, 1912–1934 (2002)
32. D. Fabbro, C. Garcia-Echeverria, Targeting protein kinases in cancer therapy. Curr. Opin. Drug Discov. Dev. **5**, 701–712 (2002)
33. P. Cohen, Protein kinases-the major drug targets of the twenty-first century? Nat. Rev. Drug Discov. **1**, 309–315 (2002)
34. M.E.M. Noble, J.A. Endicott, L.N. Johnson, Protein kinase inhibitors: insights into drug design from structure. Science **303**, 1800–1805 (2004)
35. J.M. Zhang, P.L. Yang, N.S. Gray, Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer **9**, 28–39 (2009)
36. S.Y. Zhang, D.H. Yu, Targeting Src family kinases in anti-cancer therapies: turning promise into triumph. Trends Pharmacol. Sci. **33**, 122–128 (2012)
37. F.M. Ferguson, N.S. Gray, Kinase inhibitors: the road ahead. Nat. Rev. Drug Discov. **17**, 353–377 (2018)
38. T.J. Boggon, M.J. Eck, Structure and regulation of Src family kinases. Oncogene **23**, 7918–7927 (2004)
39. H. Yamaguchi, W.A. Hendrickson, Structural basis for activation of human lymphocyte kinase Lck upon tyrosine phosphorylation. Nature **384**, 484–489 (1996)
40. F. Sicheri, I. Moarefi, J. Kuriyan, Crystal structure of the Src family tyrosine kinase Hck. Nature **385**, 602–609 (1997)
41. A.P. Kornev, N.M. Haste, S.S. Taylor, L.F. Ten Eyck, Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc. Natl. Acad. Sci. U.S.A. **103**, 17783–17788 (2006)
42. B. Nagar, W.G. Bornmann, P. Pellicena, T. Schindler, D.R. Veach, W.T. Miller, B. Clarkson, J. Kuriyan, Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). Cancer Res. **62**, 4236–4243 (2002)
43. T. Schindler, W. Bornmann, P. Pellicena, W.T. Miller, B. Clarkson, J. Kuriyan, Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. Science **289**, 1938–1942 (2000)
44. N. Vajpai, A. Strauss, G. Fendrich, S.W. Cowan-Jacob, P.W. Manley, S. Grzesiek, W. Jahnke, Solution conformations and dynamics of ABL kinase-inhibitor complexes determined by NMR substantiate the different binding modes of imatinib/nilotinib and dasatinib. J. Biol. Chem. **283**, 18292–18302 (2008)
45. M.A. Young, S. Gonfloni, G. Superti-Furga, B. Roux, J. Kuriyan, Dynamic coupling between the SH2 and SH3 domains of c-Src and hck underlies their inactivation by C-terminal tyrosine phosphorylation. Cell **105**, 115–126 (2001)
46. A. Suenaga, A.B. Kiyatkin, M. Hatakeyama, N. Futatsugi, N. Okimoto, Y. Hirano, T. Narumi, A. Kawai, R. Susukita, T. Koishi, H. Furusawa, K. Yasuoka, N. Takada, Y. Ohno, M. Taiji, T. Ebisuzaki, J.B. Hoek, A.

Konagaya, B.N. Kholodenko, Tyr-317 phosphorylation increases Shc structural rigidity and reduces coupling of domain motions remote from the phosphorylation site as revealed by molecular dynamics simulations. J. Biol. Chem. **279**, 4657–4662 (2004)
47. N.M. Levinson, O. Kuchment, K. Shen, M.A. Young, M. Koldobskiy, M. Karplus, P.A. Cole, J. Kuriyan, A Src-like inactive conformation in the Abl tyrosine kinase domain. PLoS Biol. **4**, 753–767 (2006)
48. A. Dixit, G.M. Verkhivker, Hierarchical modeling of activation mechanisms in the ABL and EGFR kinase domains: thermodynamic and mechanistic catalysts of kinase activation by cancer mutations. PLoS Comput. Biol. **5**, e10004487 (2009)
49. A. Cembran, L.R. Masterson, C.L. McClendon, S.S. Taylor, J.L. Gao, G. Veglia, Conformational equilibrium of N-myristoylated cAMP-dependent protein kinase A by molecular dynamics simulations. Biochemistry **51**, 10186–10196 (2012)
50. L.R. Masterson, A. Cembran, L. Shi, G. Veglia, in *Adv. Protein Chem. Struct. Biol.*, vol. 87, ed. by C. Christov, T. Karabencheva-Christova (Academic Press, 2012), Ch. 12, pp. 363–389
51. B.W. Boras, A. Kornev, S.S. Taylor, A.D. McCulloch, Using Markov state models to develop a mechanistic understanding of protein kinase A regulatory subunit RI alpha activation in response to cAMP binding. J. Biol. Chem. **289**, 30040–30051 (2014)
52. E.D. Lopez, O. Burastero, J.P. Arcon, L.A. Defelipe, N.G. Ahn, M.A. Marti, A.G. Turjanski, Kinase activation by small conformational changes. J. Chem. Inf. Model. **60**, 821–832 (2020)
53. Y.B. Shan, K. Gnanasambandan, D. Ungureanu, E.T. Kim, H. Hammaren, K. Yamashita, O. Silvennoinen, D.E. Shaw, S.R. Hubbard, Molecular basis for pseudokinase-dependent autoinhibition of JAK2 tyrosine kinase. Nat. Struct. Mol. Biol. **21**, 579–584 (2014)
54. L. Sutto, F.L. Gervasio, Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. Proc. Natl. Acad. Sci. U.S.A. **110**, 10616–10621 (2013)
55. Y.B. Shan, M.P. Eastwood, X.W. Zhang, E.T. Kim, A. Arkhipov, R.O. Dror, J. Jumper, J. Kuriyan, D.E. Shaw, Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. Cell **149**, 860–870 (2012)
56. A. Arkhipov, Y.B. Shan, R. Das, N.F. Endres, M.P. Eastwood, D.E. Wemmer, J. Kuriyan, D.E. Shaw, Architecture and membrane interactions of the EGF receptor. Cell **152**, 557–569 (2013)
57. A. Arkhipov, Y.B. Shan, E.T. Kim, R.O. Dror, D.E. Shaw, Her2 activation mechanism reflects evolutionary preservation of asymmetric ectodomain dimers in the human EGFR family. Elife **2**, e00708 (2013). https://doi.org/10.7554/eLife.00708
58. N.F. Endres, R. Das, A.W. Smith, A. Arkhipov, E. Kovacs, Y.J. Huang, J.G. Pelton, Y.B. Shan, D.E. Shaw, D.E. Wemmer, J.T. Groves, J. Kuriyan, Conformational coupling across the plasma membrane in activation of the EGF receptor. Cell **152**, 543–556 (2013)
59. Y.B. Shan, A. Arkhipov, E.T. Kim, A.C. Pan, D.E. Shaw, Transitions to catalytically inactive conformations in EGFR kinase. Proc. Natl. Acad. Sci. U.S.A. **110**, 7270–7275 (2013)
60. M. Yan, H. Wang, Q. Wang, Z. Zhang, C. Zhang, Allosteric inhibition of c-Met kinase in submicrosecond molecular dynamics simulations induced by its inhibitor, tivantinib. Phys. Chem. Chem. Phys. **18**, 10367–10374 (2016)
61. M.A. Seeliger, P. Ranjitkar, C. Kasap, Y.B. Shan, D.E. Shaw, N.P. Shah, J. Kuriyan, D.J. Maly, Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. Cancer Res. **69**, 2384–2392 (2009)
62. Y.B. Shan, M.A. Seeliger, M.P. Eastwood, F. Frank, H.F. Xu, M.O. Jensen, R.O. Dror, J. Kuriyan, D.E. Shaw, A conserved protonation-dependent switch controls drug binding in the Abl kinase. Proc. Natl. Acad. Sci. U.S.A. **106**, 139–144 (2009)
63. A.C. Dar, M.S. Lopez, K.M. Shokat, Small molecule recognition of c-Src via the lmatinib-binding conformation. Chem. Biol. **15**, 1015–1022 (2008)
64. M.A. Seeliger, B. Nagar, F. Frank, X. Cao, M.N. Henderson, J. Kuriyan, c-Src binds to the cancer drug imatinib with an inactive Abl/c-Kit conformation and a distributed thermodynamic penalty. Structure **15**, 299–311 (2007)
65. S.W. Cowan-Jacob, H. Mobitz, D. Fabbro, Structural biology contributions to tyrosine kinase drug discovery. Curr. Opin. Cell Biol. **21**, 280–287 (2009)
66. A. Aleksandrov, T. Simonson, Molecular dynamics simulations show that conformational selection governs the binding preferences of imatinib for several tyrosine kinases. J. Biol. Chem. **285**, 13807–13815 (2010)
67. S. Lovera, M. Morando, E. Pucheta-Martinez, J.L. Martinez-Torrecuadrada, G. Saladino, F.L. Gervasio, Towards a molecular understanding of the link between imatinib resistance and kinase conformational dynamics. PLoS Comput. Biol. **11**, e1004578 (2015)
68. S. Lovera, L. Sutto, R. Boubeva, L. Scapozza, N. Dolker, F.L. Gervasio, The different flexibility of c-Src and c-Abl kinases regulates the accessibility of a druggable inactive conformation. J. Am. Chem. Soc. **134**, 2496–2499 (2012)
69. Y. Meng, Y.L. Lin, B. Roux, Computational study of the "DFG-flip" conformational transition in c-Abl and c-Src tyrosine kinases. J. Phys. Chem. B **119**, 1443–1456 (2015)
70. H. Vashisth, L. Maragliano, C.F. Abrams, "DFG-Flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. Biophys. J. **102**, 1979–1987 (2012)
71. F. Filomia, F. De Rienzo, M.C. Menziani, Insights into MAPK p38 alpha DFG flip mechanism by accelerated molecular dynamics. Bioorgan. Med. Chem. **18**, 6805–6812 (2010)
72. A. Dixit, G.M. Verkhivker, Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. PLoS Comput. Biol. **7**, e1002179 (2011)
73. R.S.K. Vijayan, P. He, V. Modi, K.C. Duong-Ly, H.C. Ma, J.R. Peterson, R.L. Dunbrack, R.M. Levy, Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. J. Med. Chem. **58**, 466–479 (2015)

74. A. Haldane, W.F. Flynn, P. He, R.S.K. Vijayan, R.M. Levy, Structural propensities of kinase family proteins from a Potts model of residue co-variation. Protein Sci. **25**, 1378–1384 (2016)
75. R.V. Agafonov, C. Wilson, R. Otten, V. Buosi, D. Kern, Energetic dissection of Gleevec's selectivity toward human tyrosine kinases. Nat. Struct. Mol. Biol. **21**, 848–853 (2014)
76. J. Mendieta, F. Gago, In silico activation of Src tyrosine kinase reveals the molecular basis for intramolecular autophosphorylation. J. Mol. Graph. Model. **23**, 189–198 (2004)
77. N.K. Banavali, B. Roux, The N-terminal end of the catalytic domain of Src kinase Hck is a conformational switch implicated in long-range allosteric regulation. Structure **13**, 1715–1723 (2005)
78. N.K. Banavali, B. Roux, Anatomy of a structural pathway for activation of the catalytic domain of Src kinase Hck. Proteins Struct. Funct. Bioinform. **67**, 1096–1112 (2007)
79. N.K. Banavali, B. Roux, Flexibility and charge asymmetry in the activation loop of Src tyrosine kinases. Proteins **74**, 378–389 (2009)
80. E. Paci, M. Karplus, Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J. Mol. Biol. **288**, 441–459 (1999)
81. E. Ozkirimli, C.B. Post, Src kinase activation: a switched electrostatic network. Protein Sci. **15**, 1051–1062 (2006)
82. E. Ozkirimli, S.S. Yadav, W.T. Miller, C.B. Post, An electrostatic network and long-range regulation of Src kinases. Protein Sci. **17**, 1871–1880 (2008)
83. X. Huang, Y. Yao, G.R. Bowman, J. Sun, L.J. Guibas, G. Carlsson, V.S. Pande, Constructing multi-resolution Markov State Models (MSMs) to elucidate RNA hairpin folding mechanisms, in Pac Symp Biocomput, pp. 228–239 (2010)
84. W. Jiang, J.C. Phillips, L. Huang, M. Fajer, Y. Meng, J.C. Gumbart, Y. Luo, K. Schulten, B. Roux, Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. Comput. Phys. Commun. **185**, 908–916 (2014)
85. J.O. Tempkin, B. Qi, M.G. Saunders, B. Roux, A.R. Dinner, J. Weare, Using multiscale preconditioning to accelerate the convergence of iterative molecular calculations. J. Chem. Phys. **140**, 184114 (2014)
86. W. Gan, S. Yang, B. Roux, Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. Biophys. J. **97**, L8–L10 (2009)
87. M. Fajer, Y. Meng, B. Roux, The activation of c-Src tyrosine kinase: conformational transition pathway and free energy landscape. J. Phys. Chem. B **121**, 3352–3363 (2017)
88. H. Huang, R.J. Zhao, B.M. Dickson, R.D. Skeel, C.B. Post, alpha C helix as a switch in the conformational transition of Src/CDK-like kinase domains. J. Phys. Chem. B **116**, 4465–4475 (2012)
89. H. Wu, C.B. Post, Protein conformational transitions from all-atom adaptively biased path optimization. J. Chem. Theory Comput. **14**, 5372–5382 (2018)
90. H. Wu, H. Huang, C.B. Post, All-atom adaptively biased path optimization of Src kinase conformational inactivation: switched electrostatic network in the concerted motion of alphaC helix and the activation loop. J. Chem. Phys. **153**, 175101 (2020)
91. B. Narayan, A. Fathizadeh, C. Templeton, P. He, S. Arasteh, R. Elber, N.V. Buchete, R.M. Levy, The transition between active and inactive conformations of Abl kinase studied by rock climbing and milestoning. Biochim. Biophys. Acta Gen. Subj. **1864**, 129508 (2020)
92. S. Yang, N.K. Banavali, B. Roux, Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. Proc. Natl. Acad. Sci. U.S.A. **106**, 3776–3781 (2009)
93. Y. Meng, B. Roux, Locking the active conformation of c-Src kinase through the phosphorylation of the activation loop. J. Mol. Biol. **426**, 423–435 (2014)
94. D. Shukla, Y. Meng, B. Roux, V.S. Pande, Activation pathway of Src kinase reveals intermediate states as targets for drug design. Nat. Commun. **5**, 3397 (2014)
95. Y. Meng, L.G. Ahuja, A.P. Kornev, S.S. Taylor, B. Roux, A catalytically disabled double mutant of Src tyrosine kinase can be stabilized into an active-like conformation. J. Mol. Biol. **430**, 881–889 (2018)
96. L. Maragliano, E. Vanden-Eijnden, B. Roux, Free energy and kinetics of conformational transitions from Voronoi tessellated milestoning with restraining potentials. J. Chem. Theory Comput. **5**, 2589–2594 (2009)
97. M.A. Morando, G. Saladino, N. D'Amelio, E. Pucheta-Martinez, S. Lovera, M. Lelli, B. Lopez-Mendez, M. Marenchino, R. Campos-Olivas, F.L. Gervasio, Conformational selection and induced fit mechanisms in the binding of an anticancer drug to the c-Src kinase. Sci. Rep. **6**, 24439 (2016)
98. Y. Deng, B. Roux, Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. J. Chem. Phys. **128**, 115103 (2008)
99. S.K. Albanese, J.D. Chodera, A. Volkamer, S. Keng, R. Abel, L. Wang, Is structure-based drug design ready for selectivity optimization? J. Chem. Inf. Model. **60**, 6211–6227 (2020)
100. S. Yang, B. Roux, Src kinase conformational activation: thermodynamics, pathways, and mechanisms. PLoS Comput. Biol. **4**, e1000047 (2008)
101. L. Huang, M. Wright, S. Yang, L. Blachowicz, L. Makowski, B. Roux, Glycine substitution in SH3-SH2 connector of Hck tyrosine kinase causes population shift from assembled to disassembled state. Biochim. Biophys. Acta Gen. Subj. **1864**, 129604 (2020)
102. G.R. Bowman, K.A. Beauchamp, G. Boxer, V.S. Pande, Progress and challenges in the automated construction of Markov state models for full protein systems. J. Chem. Phys. **131**, 124101 (2009)
103. G.R. Bowman, X. Huang, V.S. Pande, Using generalized ensemble simulations and Markov state models to identify conformational states. Methods **49**, 197–201 (2009)
104. G.R. Bowman, D.L. Ensign, V.S. Pande, Enhanced modeling via network theory: adaptive sampling of Markov state models. J. Chem. Theory Comput. **6**, 787–794 (2010)

105. M.S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A.L. Beberg, D.L. Ensign, C.M. Bruns, V.S. Pande, Accelerating molecular dynamic simulation on graphics processing units. J. Comput. Chem. **30**, 864–872 (2009)
106. E. Luttmann, D.L. Ensign, V. Vaidyanathan, M. Houston, N. Rimon, J. Oland, G. Jayachandran, M. Friedrichs, V.S. Pande, Accelerating molecular dynamic simulation on the cell processor and Playstation 3. J. Comput. Chem. **30**, 268–274 (2009)
107. P. Eastman, M.S. Friedrichs, J.D. Chodera, R.J. Radmer, C.M. Bruns, J.P. Ku, K.A. Beauchamp, T.J. Lane, L.P. Wang, D. Shukla, T. Tye, M. Houston, T. Stich, C. Klein, M.R. Shirts, V.S. Pande, OpenMM 4: a reusable, extensible, hardware independent library for high performance molecular simulation. J. Chem. Theory Comput. **9**, 461–469 (2013)
108. R. Salomon-Ferrer, A.W. Gotz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. J. Chem. Theory Comput. **9**, 3878–3888 (2013)
109. K.A. Beauchamp, G.R. Bowman, T.J. Lane, L. Maibaum, I.S. Haque, V.S. Pande, MSMBuilder2: modeling conformational dynamics at the picosecond to millisecond scale. J. Chem. Theory Comput. **7**, 3412–3419 (2011)
110. M.K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Perez-Hernandez, M. Hoffmann, N. Plattner, C. Wehmeyer, J.H. Prinz, F. Noe, PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. J. Chem. Theory Comput. **11**, 5525–5542 (2015)
111. Y. Meng, D. Shukla, V.S. Pande, B. Roux, Transition path theory analysis of c-Src kinase activation. Proc. Natl. Acad. Sci. U.S.A. **113**, 9193–9198 (2016)
112. M.P. Harrigan, M.M. Sultan, C.X. Hernandez, B.E. Husic, P. Eastman, C.R. Schwantes, K.A. Beauchamp, R.T. McGibbon, V.S. Pande, MSMBuilder: statistical models for biomolecular dynamics. Biophys. J. **112**, 10–15 (2017)
113. M.M. Sultan, G. Kiss, V.S. Pande, Towards simple kinetic models of functional dynamics for a kinase subfamily. Nat. Chem. **10**, 903–909 (2018)
114. M.J. Keiser, J.J. Irwin, B.K. Shoichet, The chemical basis of pharmacology. Biochemistry **49**, 10267–10276 (2010)
115. Y. Meng, C. Gao, D.K. Clawson, S. Atwell, M. Russell, M. Vieth, B. Roux, Predicting the conformational variability of Abl tyrosine kinase using molecular dynamics simulations and Markov state models. J. Chem. Theory Comput. **14**, 2721–2732 (2018)
116. F. Paul, Y. Meng, B. Roux, Identification of druggable kinase target conformations using Markov model metastable states analysis of apo-Abl. J. Chem. Theory Comput. **16**, 1896–1912 (2020)
117. F. Paul, T. Thomas, B. Roux, Diversity of long-lived intermediates along the binding pathway of imatinib to Abl kinase revealed by MD simulations. J. Chem. Theory Comput. **16**, 7852–7865 (2020)
118. Y. Meng, B. Roux, Computational study of the W260A activating mutant of Src tyrosine kinase. Protein Sci. **25**, 219–230 (2016)
119. M. LaFevre-Bernt, F. Sicheri, A. Pico, M. Porter, J. Kuriyano, W.T. Miller, Intramolecular regulatory interactions in the Src family kinase Hck probed by mutagenesis of a conserved tryptophan residue. J. Biol. Chem. **273**, 32129–32134 (1998)
120. L. Fang, J. Vilas-Boas, S. Chakraborty, Z.E. Potter, A.C. Register, M.A. Seeliger, D.J. Maly, How ATP-competitive inhibitors allosterically modulate tyrosine kinases that contain a Src-like regulatory architecture. ACS Chem. Biol. **15**, 2005–2016 (2020)
121. M.P. Pond, R. Eells, B.W. Treece, F. Heinrich, M. Losche, B. Roux, Membrane anchoring of Hck kinase via the intrinsically disordered SH4-U and length scale associated with subcellular localization. J. Mol. Biol. **432**, 2985–2997 (2020)
122. T. Xie, T. Saleh, P. Rossi, C.G. Kalodimos, Conformational states dynamically populated by a kinase determine its function. Science **370**, 189 (2020). https://doi.org/10.1126/science.abc2754
123. M.K. Joshi, R.A. Burton, H. Wu, A.M. Lipchik, B.P. Craddock, H. Mo, L.L. Parker, W.T. Miller, C.B. Post, Substrate binding to Src: a new perspective on tyrosine kinase substrate recognition from NMR and molecular dynamics. Protein Sci. **29**, 350–359 (2020)
124. S. Swendeman, B. Nagar, D. Wisniewski, A. Strife, C. Lambek, C. Liu, D. Veach, W. Bornmann, J. Kuriyan, B. Clarkson, Crystal structures of the c-Abl tyrosine kinase domain in complex with STI-571 and a novel Bcr-Abl inhibitor, PD1173955. Clin. Cancer Res. **7**, 3768s–3768s (2001)
125. F. Pontiggia, D.V. Pachov, M.W. Clarkson, J. Villali, M.F. Hagan, V.S. Pande, D. Kern, Free energy landscape of activation in a signalling protein at atomic resolution. Nat. Commun. **6**, 7284 (2015)
126. C. Wilson, R.V. Agafonov, M. Hoemberger, S. Kutter, A. Zorba, J. Halpin, V. Buosi, R. Otten, D. Waterman, D.L. Theobald, D. Kern, Kinase dynamics. Using ancient protein kinases to unravel a modern cancer drug's mechanism. Science **347**, 882–886 (2015)
127. S. Yang, L. Blachowicz, L. Makowski, B. Roux, Multidomain assembled states of Hck tyrosine kinase in solution. Proc. Natl. Acad. Sci. U.S.A. **107**, 15757–15762 (2010)
128. C.L. McClendon, A.P. Kornev, M.K. Gilson, S.S. Taylor, Dynamic architecture of a protein kinase. Proc. Natl. Acad. Sci. U.S.A. **111**, E4623–E4631 (2014)
129. J.H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J.D. Chodera, C. Schutte, F. Noe, Markov models of molecular kinetics: generation and validation. J. Chem. Phys. **134**, 174105 (2011)
130. F. Nuske, H. Wu, J.H. Prinz, C. Wehmeyer, C. Clementi, F. Noe, Markov state models from short non-equilibrium simulations-analysis and correction of estimation bias. J. Chem. Phys. **146** (2017)
131. M. Weber, K. Fackeldey, C. Schutte, Set-free Markov state model building. J. Chem. Phys. **146**, 124133 (2017). https://doi.org/10.1063/1.4978501
132. I. Buch, T. Giorgino, G. De Fabritiis, Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc. Natl. Acad. Sci. U.S.A. **108**, 10184–10189 (2011)
133. N. Plattner, F. Noe, Protein conformational plasticity and complex ligand-binding kinetics explored by atom-

istic simulations and Markov models. Nat. Commun. **6**, 7653 (2015)
134. J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field. J. Comput. Chem. **25**, 1157–1174 (2004)
135. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell Jr., CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. **31**, 671–690 (2010)
136. L. Huang, B. Roux, Automated force field parameterization for non-polarizable and polarizable atomic models based on target data. J. Chem. Theory Comput. **9**, 3543–3556 (2013). https://doi.org/10.1021/ct4003477
137. J. Singh, R.C. Petter, T.A. Baillie, A. Whitty, The resurgence of covalent drugs. Nat. Rev. Drug Discov. **10**, 307–317 (2011)
138. S. Klus, A. Bittracher, I. Schuster, C. Schutte. A kernel-based approach to molecular conformation analysis. J. Chem. Phys. **149**, 244109 (2018). https://doi.org/10.1063/1.5063533
139. E. Rosta, G. Hummer, Free energies from dynamic weighted histogram analysis using unbiased Markov state model. J. Chem. Theory Comput. **11**, 276–285 (2015)
140. A.S.J.S. Mey, H. Wu, F. Noe, xTRAM: estimating equilibrium expectations from time-correlated simulation data at multiple thermodynamic states. Phys. Rev. X **4**, 041018 (2014)
141. H. Wu, A.S. Mey, E. Rosta, F. Noe, Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states. J. Chem. Phys. **141**, 214106 (2014)
142. J. Wang, S. Chmiela, K.R. Muller, F. Noe, C. Clementi, Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. J. Chem. Phys. **152**, 194106 (2020)
143. T.O.F. Conrad, M. Genzel, N. Cvetkovic, N. Wulkow, 1286 A. Leichtle, J. Vybiral, G. Kutyniok, C. Schutte. Sparse Proteomics Analysis – a compressed sensing-based approach for feature selection and classification of high-dimensional proteomics mass spectrometry data. BMC Bioinform. **18**, 160 (2017). https://doi.org/10.1186/s12859-017-1565-4
144. G.M. Rotskoff, E. Vanden-Eijnden, Dynamical computation of the density of states and Bayes factors using nonequilibrium importance sampling. Phys. Rev. Lett. **122**, 150602 (2019)
145. J. Wang, S. Olsson, C. Wehmeyer, A. Perez, N.E. Charron, G. de Fabritiis, F. Noe, C. Clementi, Machine learning of coarse-grained molecular dynamics force fields. ACS Cent. Sci. **5**, 755–767 (2019)
146. Bittracher, A., Klus, S., Hamzi, B. et al. Dimensionality Reduction of Complex Metastable Systems via Kernel Embeddings of Transition Manifolds. J. Nonlinear Sci. **31**, 3 (2021). https://doi.org/10.1007/s00332-020-09668-z
147. A.C. Pan, B. Roux, Building Markov state models along pathways to determine free energies and rates of transitions. J. Chem. Phys. **129**, 064107 (2008)
148. P.M. Ung, R. Rahman, A. Schlessinger, Redefining the protein kinase conformational space with machine learning. Cell Chem. Biol. **25**, 916-924 e912 (2018)