Molecular Modeling of Protein Kinases: Current Status and Challenges

Antti Poso

Contents

Abstract Molecular modeling and virtual screening are currently among the main tools in kinase inhibitor design. In addition, molecular dynamics is actively used to study structural features and function of kinases. During the last 20 years, computational power has dramatically increased, and at the same time, algorithms have become more user-friendly. This has resulted in a situation where these powerful methods are more easily available for even larger groups of scientists. To effectively use computational methods, one should understand in great detail how protein kinases are functioning and how current protein kinase inhibitors interact with the kinase domain. This short review presents some of the main topics of kinase modeling, concentrating especially on proper selection and preparation of protein structure and general usage of MD simulations.

Keywords Docking, Molecular dynamics, Structure-based drug design, Virtual screening

A. Poso (✉)
School of Pharmacy, University of Eastern Finland, Kuopio, Finland

Department of Internal Medicine VIII, University Hospital Tübingen, Tübingen, Germany
e-mail: antti.poso@uef.fi

1 Introduction

One of the first molecular modeling studies dealing with protein kinases was published by Fry, Kuby, and Mildvan [1] in the mid-1980s. The authors used NMR NOE's and molecular modeling to understand how MgATP interacts with rabbit muscle adenylate kinase. This study can be understood as a starting point for a great deal of and increasingly more active research on small molecule-kinase interactions. Surprisingly, we are still facing partially the same obstacles as Fry et al. some 35 years ago. In a highly simplified way, we are trying to understand structural properties of kinases and how kinase inhibitors are modifying these properties. The current main question is how kinase function and conformation effect upon the inhibitor binding are related to each other. Although the current paradigm in kinase drug design is to interfere the biological activity of kinase with small molecules, we do now understand that this inhibition cannot be modeled only by a simple docking experiment between a small drug-like molecule and the ATP-binding site of the target kinase. Instead, it is mandatory to study the whole kinase domain with solvent and, in many cases, with additional domains and interacting proteins.

This chapter will deal with the molecular modeling of kinases. Although some structural biology data is also presented, I would warmly recommend the reader to study the excellent text by Röhm, Krämer, and Knapp in this book (Chap. XX) to begin with. Modeling is, after all, based on our knowledge of structural biology, and very little can be achieved without high-quality protein structures. In addition, protein kinases share several unique structural features, like hydrophobic spines [2], which one should know prior to looking at the details of molecular modeling around kinases. This chapter is not to be taken as a guide on how to model kinases, neither is it a complete review of the topic. The emphasis is more on indicating those critical factors which one must consider when and if protein kinases are modeled. At the same time, this chapter concentrates mainly on structure-based drug design aspects, and detailed analysis of quantum mechanical studies or QSAR/machine learning, for example, is not included. One reason why QSAR and related methods are not analyzed is that high-quality QSAR studies of kinase inhibitors are rare and most of the time only explanatory in nature. One can even argue that since the invention of 3D-QSAR studies in the late 1980s [3], the development of QSAR methods in drug discovery has been quite negligible, and structure-based methods are now the mainstream in drug design.

So, what are the modeling issues we are currently struggling with, and what are the main approaches computational medicinal chemists and molecular modelers are utilizing? A simple answer to this question is "molecular motion" and "molecular dynamics." In other words, the aim is to go beyond simple virtual screening and docking and look at how topics like solvent effects, local and global molecular motions, and protein-protein interactions are modeled.

And yet, there is still one preliminary question to be answered: what is molecular modeling? Maybe the best response is offered by Ander Leach: "Molecular

modelling encompasses all methods, theoretical and computational, used to model or mimic the behavior of molecules." The art of modeling is to include all those factors which are needed to gain correct results and not those which are not affecting the outcome. In early 2000 many scientists used to think that all that was needed to model kinase inhibitor binding and biological activity was a proper knowledge of the kinase binding cavity structure and a good scoring function. Unfortunately, kinase life (like protein life in general) is more complicated, and several findings have forced us to rethink what is important. As an example, the first-generation Raf inhibitors turned out to be both kinase inhibitors and activators at the same time [4]. This paradoxical finding cannot be explained simply by binding interactions or structural data based on protein kinase X-ray structures but requires considering kinase dimerization and allosteric effects between kinase domains [5]. Another classical example, shown by Wood et al. [6], demonstrates how three kinase inhibitors, lapatinib (Tykerb, GlaxoSmithKline: GW572916), gefitinib (Iressa, AstraZeneca: ZD-1839), and erlotinib (Tarceva, OSI: OSI-774) all bind the EGFR kinase domain with almost equal IC_{50} values but yet a dramatically different effect on cell cultures. As it turned out, these compounds have big differences in target residence times. So, simple IC_{50} or binding affinity in the form of pK_i is not the dictating factor for biological activity, but, instead, dynamic properties are critical. A third example demonstrates how solvent effects do explain kinase inhibitors' structure-activity relationships. Direct interactions between cyclin G-associated kinase (GAK) and a library of 4-anilinoquin(az)olines were not able to explain structure-activity relationships (SAR). Instead, desolvation energies, reflecting enthalpy and entropy of individual water molecules within the GAK binding site, were critical for building a systematic SAR model. This example clearly indicates that water should not be neglected during kinase modeling [7]. Although these examples may seem to be quite unique, there is one common factor combining all the cases. To gain useful modeling results, we must consider the protein including solvent and dynamic aspects of the whole molecular system.

So, should we model protein kinases alone or inhibitor-kinase systems in general? A simple, fundamental answer originates from thermodynamics. Equation (1) shows the very basic relationship between binding affinity and Gibbs energy of binding:

$$\Delta G = -RT \ln K_a = -RT \ln \left(\frac{1}{K_d} \right) = \mu_{PL} - \mu_L - \mu_P \qquad (1)$$

Equation (1), Gibbs energy of binding (ΔG^0), R = Gas constant, T = temperature (K), K_a = drug-binding association constant, K_d = drug-binding dissociation constant, μ_{PL} = chemical potential of protein/ligand complex in solution, μ_L = chemical potential of Ligand in solution and μ_P = chemical potential of protein in solution

As binding energy (and affinity) is related to the chemical potential of the protein-ligand system in solution, one must consider all the factors which are involved in the chemical potential. Thus, one must study how the solvent interacts with the protein/ligand and the protein-ligand complex. In addition, one must consider all the configurations (protein-ligand poses with different conformations) of the protein-ligand complex which are relevant for binding. This is not done by two very popular methods, namely, docking and QSAR, since in docking every individual pose is scored independently, and QSAR does usually not consider anything else than ligand 2D or 3D structural descriptors. Both of these methods have been successfully used for quite a long time, QSAR since early 1960 [8] and docking from the early 1990s [9]. As one can easily understand, those methods were developed to be fast and easily available, thus not requiring substantial computational power. This was only possible by making those major simplifications which, at the time, were acceptable but should be reconsidered in the current world.

Thanks to the current massive GPU and classical supercomputer environments, it is now possible to study a full protein-solvent-ligand ensemble in a dynamical fashion. Without going into details, it can be stated that molecular dynamics (MD) approaches are the natural answer to the problem presented in Eq. (1). Unfortunately, usage of MD simulations means that the computational burden is much higher than with classical molecular docking or QSAR. This is not the only issue, since results from MD simulations are quite complicated. Both, docking and QSAR, are popular methods, partially, because they deliver simple numerical results (scoring or predictions), easy to understand, and be compared. Even the most "complicated" QSAR method, CoMFA [3], returns a clear (and often misleading) 3D image indicating those regions around the ligand structure which should be modified to gain better binding interactions. The results from MD simulations are in the form of molecular trajectories, describing atomistic movement and corresponding kinetic and potential energies. One must use a substantial amount of time and, paradoxically, computing power to analyze large MD trajectories before results can be used to guide medicinal chemistry work. At the same time, there is no easy and general procedure how to analyze MD trajectories quantitatively. Analytical procedure strongly depends on the research question. Thus it may be very time-consuming just to find what to search for from the trajectory data.

Besides understanding atomic motion, one must use an appropriate protein conformation for kinase modeling. Kinase inhibitors are classified as types I, 1½, and II–VI [10]. The consensus is that type I inhibitors target catalytically active, DFG-in conformation, and thus compete with ATP, while type II inhibitors target inactive DFG-out conformation which lacks the ATP. Type 1½ inhibitors have high affinity toward both DFG-in-like and DFG-out conformations, while types III and IV are used for allosteric inhibitors. The last two types, V (bivalent inhibitor) and VI (covalent) are not commonly used. Since this classification is based on the kinase conformation, as seen in the corresponding inhibitor-kinase complex, one can easily understand that protein kinase conformation does actually matter. Modeling must be based on the protein structure matching the requirements of an inhibitor. Thus, if one is modeling a classical type II inhibitor but the target protein conformation is a

catalytically active DFG-in (type I), all the structure-based modeling methods will ultimately yield false results. The true issue is the fact that the kinase inhibitor structure alone is not enough to predict if the inhibitor is type I, II, or something else. Also very minor changes to the inhibitor structure might change the conformation of the target kinase [11].

2   Virtual Screening and Docking

Docking is the most commonly used tool in virtual screening. In the case of kinase inhibitors, one can easily find tens if not hundreds of publications showing different types of docking approaches used. There is indeed a large number of different software packages and scoring functions to choose from (for a recent review about docking in general and especially about the pitfalls, see Pantsar and Poso [12]), but one cannot claim that any specific method would be clearly better than another. This does not mean that all the approaches are working or that it doesn't matter how virtual screening is carried out [13]. Maybe one of the most critical aspects in molecular modeling of kinases is the selection of protein kinase conformations to be used in virtual screening. Kinases are well-known enzymes, and thus conformational variation has been extensively studied [14]. The main way to classify kinase structures is to use DFG-in and DFG-out families, which refer to the DFG domain orientation [15]. Although DFG-in and DFG-out are also well explained elsewhere in this book, it is good to look at the definition on a general level.

The activation loop of the kinase protein controls the enzymatic activity by relocating itself onto the surface of the protein, resulting in kinase inactivation. Additional activity control is reached by the DFG motif conformational shifts, so that the phenylalanine of DFG occupies the ATP binding pocket, and catalytically active aspartate is pointing away from the active site. In a catalytically active state, the kinase is always in DFG-in conformation binding the magnesium ion that interacts directly with an oxygen atom of the β phosphate of ATP. In addition, the active state includes glutamate from the C-helix in a salt bridge with a lysine of the β3 strand. This salt bridge stabilizes the hydrogen bonds between lysine and oxygen atoms of the α and β phosphates of ATP [15].

When we look at the most recent molecular modeling studies where docking has been used for kinase inhibitor design, we only consider those studies where docking has been validated either by biological (in vitro) assays or/and X-ray crystallography. It is mandatory that if modeling data are published, and especially if there are predictions concerning a specific compound, these predictions must be supported by empirical data. In such a case where modeling is used to make and publish detailed activity predictions, it will create a situation where the given compounds, even the hypothetical ones, cannot be protected by patents.

Docking is basically just a method to create and score a protein-ligand binding pose. Indeed, the simplest way to use docking is to estimate a single compound binding mode like in the work of Lee et al. [16], which utilized docking together with

several other molecule methods. Although it is not mandatory, validation of the docking pose would increase the value of the study [17]. One simple approach was used by Ortuso et al. who combined several docking results (Glide XP) from X-ray structures of the Sgk1 kinase [18]. This approach yielded an average docking score which was used to identify a sub-micromolar Sgk1 inhibitor. Many research groups have used more complicated approaches and combined docking with binding free energy calculations and/or QSAR [19] or used a sequential approach with pharmacophore pre-screening before docking with different methods [20]. It is seldom that docking is used alone, and typically, docking is combined with one or several other modeling and screening methods. The reason for this complexity is quite simple: scoring functions are far from optimal, and typical docking results include a high number of false-positive and false-negative "hits" [12]. Due to this, kinase-specific scoring functions or rescoring have also been used resulting in the identification of a sub-micromolar FGFR1 inhibitor [21].

As mentioned above, the DFG domain conformation indicates if the kinase is in an active or inactive state. This DFG-domain description raises some questions that should be considered when carrying out virtual screening. The most important one is quite simple: Should we target DFG-in or DFG-out or some other conformations? Naturally, the simplest approach is to use whatever empirical structure is available. This is a valid option if one is ready to accept any type of inhibitor as a result. In many cases, researchers are more interested to find either type II or type 1½ inhibitor, especially, since it has been stated that better selectivity is reached if inactive kinase conformation is targeted [22]. As most of the empirical kinase structures are the DFG-in type [23], targeting inactive kinase conformation is not automatically an option. In theory, one can modify the kinase structure and use, for example, homology modeling or MD simulations to produce a DFG-out structure by using a catalytically active DFG-in conformation as a starting point. In practice, this approach is difficult to use and requires a substantial amount of pre-existing structural data [24]. Docking itself is a static approach, and structural errors outside of the protein binding site do not affect the outcome. Thus, one should be able to get viable results if the binding cavity itself has an appropriate conformation. This is probably also valid for induced-fit docking methods, if the used method is not based on MD simulations. However, one cannot use a classical MD-ensemble docking if the kinase structure has structural issues anywhere near the binding site, since those errors would easily be reflected to the binding site of the protein kinase.

One way to modify the kinase structure is to use an induced-fit protocol and modify the target kinase conformation so that structural features are as required. This approach was used to identify inhibitors against zeta-chain protein kinase 70 kDa (ZAP70) [25]. The gatekeeper residue methionine 414 was modified to resemble the structure of Janus kinase 2 (JAK2) by aligning ZAP70 to JAK2 binding sites. In addition, a potent JAK2 inhibitor was docked to the resulting structure, and the ZAP70/JAK2-inhibitor complex was relaxed by MD simulation procedure. The induced-fit ZAP70 structure was used for the docking campaign, and several low and sub-micromolar ZAP70 inhibitors were identified. This protocol proves that although the structure used for the docking campaign was not a classical homology

model structure but a modification of an X-ray, the docking protocol was still able to identify several validated hit compounds. It is a common situation with many high-affinity compounds that the protein-ligand complex is highly complementary. In such a case, docking is often unable to produce a proper binding pose for inhibitors which are structurally different from the inhibitor within the X-ray structure. As an example, one can look at the data from Pedreira et al., in which both normal docking and induced-fit docking approaches were unable to create proper binding modes for type 1.5 p38alfa MAP kinase inhibitors [26]. Only after manual modification of kinase conformation, a proper binding pose was constructed. This pose was validated by long MD simulations (3.6 μs) with three replicas for all studied systems. Unlike in many other modeling studies in which MD simulations are used to support the docking results, this extremely long MD simulation is truly validating the proposed docking poses. One cannot claim the same for those cases where MD simulation is either only single run and/or clearly shorter than 500–1,000 ns, as that timescale is just enough to cover protein side chain movements or so-called tier 1 movements [27, 28].

One additional point to discuss is related to the idea of targeting inactive kinase conformation. The question is if all DFG-in structures are also catalytically active kinase structures or if there are DFG-in-like structures which are catalytically inactive. It seems that this is the case, as a quite recent paper by Modi and Dunbrack [23] nicely demonstrates that only a small part of DFG-in conformations is catalytically active. To be catalytically active, the protein kinase should have all the structural features required for phosphorylation activity, including a proper setup to accommodate the ATP molecule and the magnesium ion. Indeed, there are several X-ray structures with DFG-in-like features but without proper conformation to accommodate the ATP and the metal ion. What is not known at the moment is whether these inactive DFG-in-like structures are thermodynamically distinct ones in vivo and thus biologically valid or whether the inactive DFG-in structures are just artifacts of the crystallization conditions. Current data indicate that the first option is valid, as combination of X-ray structure analysis and long-scale MD simulations with CDK2 was able to identify not only classical active and inactive kinase conformations but also several metastable states [29].

3   MD Simulations

As one can see, MD simulations are becoming an increasingly popular research tool to study both conformational aspects of protein kinases and for understanding drug-protein interactions. There are several factors which are making MD a true option, but the most important ones are the dramatically increased performance due to GPU implementation of software, better force fields, and especially the Markov State Modeling approach [30]. Around 10 years ago, most of the published MD simulations included at maximum 1 μs simulation time, but current studies can easily be based on data from an over 1 ms timeframe [31]. In our group, a routine simulation

speed in the case of protein kinase MD is around 500 ns/24 h/kinase, and several simulations (typically more than 10) are run simultaneously. This equals to around 5 microseconds produced MD data within 24 h. Together with ever-continuing work on new protein-specific and general force fields [32–36], this has allowed researchers to carry out long enough simulations with good accuracy for the kinase inhibitor complex. Naturally, currently available force fields are far from perfect, and there are several attempts to include polarizability and proton transfers within classical force fields [37, 38]. However, the current status of force field methods is good enough to allow high-quality simulations which are reproducing empirical data within a reasonable error margin.

Force field development is not the only reason why MD simulations are nowadays useful in drug design. Another breakthrough is a method called Markov State Models (MSMs). MSMs are kinetic models of the process under study, usually based on MD trajectory data. The aim of the MSM approach is to build a simplified model, easy to understand and simple enough that new insight can be gained. MSM is a coarse-grained representation of the more detailed molecular trajectories for quantitative comparisons [30]. The method builds a model with individual (metastable) states and detects how often conversion from one state to another is happening. MSMs often have thousands of states or even more. The critical factor is the transition from one state to another, and with faster transitions, shorter simulations are needed to construct an MSM. As Pande et al. explain in their excellent review, the specific challenges for building an MSM can be broken down into (1) how does one define states in a kinetically meaningful scheme and (2) how can one transition the matrix in an efficient manner. If done properly, the MSM will yield both a detailed enough model about the phenomenon under study and at the same time a confirmation that the given simulation time is long enough to construct such a model [30].

Kinase inhibitor design is a typical structure-based design process, utilizing structural biology and X-ray structures. However, several MD simulation studies have recently been able to reproduce most, if not all, relevant protein conformations within selected protein kinase families [31, 39, 40]. In addition, similar studies have detected previously unknown inactive kinase structures, which have either been later validated by structural biology approaches or confirmed by X-ray structure in related kinases. Sultan, Kiss, and Pande [41] used an accelerated molecular dynamics (AMD) to study seven Src kinase structures simultaneously. They also utilized an extension of the MSM method which allowed the authors to compare MD trajectories of seven Src kinases, namely, Fyn, Lyn, Lck, Hck, Fgr, Yes, and Bl kinase. The total length of AMD simulations exceeded several milliseconds. Results indicated that the kinase active state of the seven Src kinases is typically within 1–2 kcal/mol of the inactive conformation. In addition, kinase activation is slower than deactivation. The active-inactive transitions require several metastable intermediates, and potentially those conformations can be targeted by specific inhibitors.

Although docking is carried out in vacuum, water molecules can be considered during the docking procedure. Protein-ligand solvation and desolvation are the major sources of binding energy during protein-ligand binding [42]. Indirectly, water is

included in some of the scoring functions, like Glide XP [42, 43]. However, a more direct approach is to use information from the X-ray structures and especially MD simulations with explicit solvent molecules. One example of these approaches is the already mentioned work by Lee et al., which also considered water networks within the binding site [16]. Protein cavities are not always fully solvated, and dewetted regions can substantially affect the binding affinity of ligands and inhibitors. An example of this is demonstrated by Asquith et al. [7, 44] in two studies where water networks within GAK and EGFR kinases. In both of these studies, the WaterMap method was used to identify the effect of individual water molecules, and the pure docking score was not able to rationally explain the structure-activity relationships.

One must recognize that solvent effects are not independent of equally fundamental ligand ionization properties. An excellent example of how these are connected to each other is the study by Heider et al., which showed that pyridinylimidazole as GSK3β inhibitors were strongly affected by both preferred tautomer and solvent-related binding effects [45]. Naturally, one cannot reach these conclusions without a proper quantum mechanical evaluation of ligand behavior in solvent phase. This procedure, unfortunately, requires substantial computational resources and is thus not an option for a traditional virtual screening. It should therefore be limited to those cases where more traditional approaches are not satisfactory.

Protein ionization is typically kept fixed during all the modeling studies. This assumption has recently been challenged, as it is well known that protein side chain ionization does affect ligand binding, and protein dynamics and ionization are affected by the protein 3D environment, solvent, and ions nearby. In addition, since several side chains have their pK_a values near the physiological pH, the initially assigned protonation state might not be the one which is relevant for the phenomenon under examination. To solve this issue, Brooks et al. developed a method which combines classical MD simulation with explicit solvent for accurate molecular interactions, generalized Born implicit-solvent model for estimating the free energy of protein solvation, and a pH-based replica exchange scheme to significantly enhance both protonation and conformational state sampling [46, 47]. The method, named as hybrid-solvent continuous constant pH molecular dynamics with pH replica exchange (CpHMD), was used by Shen et al. to study, for example, how the c-Src kinase DFG domain flip (DFG-in vs. DFG-out) is affected by the protonation of Asp [48]. The authors showed that protonated DFG-aspartate is compatible only with DFG-out conformation, while unprotonated aspartate is possible with both DFG forms. This clearly underlines that ionization of all relevant residues must be properly assigned before MD simulation or any structure-based drug design method is used. They also used CpHMD to identify catalytically active but nucleophilic (neutral) lysine residues which can be targeted by covalent inhibitors (the interested reader should look at the very comprehensive Chap. 30 by Gehringer). In addition, within the same study, the authors were able to identify charged cysteine residues within kinases which existed even at a physiological pH [49].

Thanks to the recent development of force field parametrization, MD simulations are currently quite reliable. It is still quite common that results from the force field methods are not fully in line with empirical data, and this is especially true if we consider X-ray crystallography. Long timescale simulation studies with p38a MAP kinase inhibitors were recently used to explain the discrepancy between X-ray and NMR data [50]. According to the classical activation mechanism, supported by X-ray structures [51], p38a MAP kinase activation with a double-phosphorylated structure should include a large reorientation of the activation loop A. However, NMR studies indicated that double phosphorylation does not induce any major conformational rearrangements [52]. Simulations with CHARMM force field were conducted with ten replicas, and simulation time was varied between 500 ns and 1 μs, although in individual cases also longer simulation times were used. The results suggest that p38a predominantly samples conformations which are in contrast with the activation model obtained from X-ray crystallography. However, the authors analyzed crystal contacts and found several artifacts affecting the protein conformation, for example, an atypically long expression His-tag of a neighboring molecule bound to the hydrophobic docking groove of p38a MAP kinase. It is easy to agree with the statement of the Kuzmanic et al. [50] "These observations show how important it is to carefully analyze symmetry-related molecules and they call for caution in the interpretation of deposited X-ray structures, as they can be misleading."

Basically, the abovementioned conclusion can be drawn also from the studies dealing with Aurora kinase A (AurA) [53, 54]. By combining experimental data and MD simulations, it has been demonstrated that AurA activation by phosphorylation occurs without a population shift from the DFG-out to the DFG-in state and that the activation loop of the activated kinase remains highly dynamic. This is, once more, against the traditional view of the X-ray. Instead, molecular dynamics simulations and electron paramagnetic resonance experiments show that phosphorylation triggers a switch within the DFG-in subpopulation from an autoinhibited DFG-in substate to an active DFG-in substate, leading to catalytic activation.

4   Allosteric Control of Kinases

Most of the kinase inhibitors target the ATP binding site of the corresponding kinase protein. While the ATP binding site is highly conserved among the kinome, the so-called exosites are much more unique, although to some extent also conserved. The first successful kinase inhibitor targeting exosites was imatinib [55]. From the modeling point of view, this paradigm shift was quite big as it demonstrated that target protein conformation is not static and that kinase conformation can be modified by targeting exosites. While, at the moment, most of the new kinase inhibitors are targeting the ATP-binding cleft between the N- and C-lobes of the kinase, interest toward allosteric inhibitors is growing due to some very evident benefits. The biggest advantage is the fact that the allosteric binding site has no

high-affinity endogenous ligand. The second major benefit is that one can, at least theoretically, control the target kinase function in a more precise fashion. However, a lot of research is needed to understand how the allosteric control mechanism works. Currently the best molecular modeling method to tackle this question is naturally molecular dynamics, as all other methods, like docking, QSAR, and pharmacophore, only give a static image of the drug-receptor complex.

Allosteric effects have been explained by different theoretical frameworks, most of which are not explained here. One of the most recent theoretical approaches is the so-called "violin" model, specifically proposed for protein kinases by Kornev and Taylor [56–59]. This model is developed directly to explain the type III and type IV kinase inhibitors' mode of action. While more traditional theories of allosteric control rely on specific atomic interaction networks with a direct pathway from the allosteric site to the site of action, all of them have some caveats. The most notable is the high thermal motion of individual atoms within a protein. Unlike in the macroscopic world, thermal motion in the microcosmos is large enough to prevent simple one-pathway networks, and big parts of the information would be lost in the process. One can also easily understand the violin model based on the MD simulations. In the typical force field method, atoms and bonds are represented by ball and springs with corresponding spring constants and thus also with corresponding vibrations. These vibrations are, even at room temperature, strong enough to constantly break and re-make most of the interprotein interactions like H-bonds, ionic bonds, and hydrophobic (dispersion) interactions. As current force fields are accurate enough to reproduce a majority of the macroscopic parameters and spectra data, we can easily accept that these vibration and intramolecular motions are also represented accurately enough by modern all-atomic force fields.

Another important work dealing with allosterism, by McClendon et al. [58], is also based on MD simulations. The work includes microsecond scale MD simulations and the authors demonstrate that Protein Kinase A (PKA) has not just semi-rigid N- and C-lobes, but several semi-rigid communities interacting with each other and controlling in a rational way the function and activity of PKA. Correlated motions between these structurally contiguous communities are associated with a particular protein kinase function and/or a regulatory mechanism. A bit surprising is the finding that some well-known protein kinase motifs are split into different communities. The community maps are able to explain how different ligands induce long-distance allosteric coupling. These communities are also in agreement with the spine network [57].

Most of the kinase modeling studies are based on kinase domain structure alone, but there are also MD simulations which do include the regulatory units, like SH2 and SH3. A comprehensive study, combining MD simulations, free energy calculations, in vitro functional assays, and single point mutations, suggests that the SH2-kinase interactions are allosterically stabilizing the αC-helix of the c-Abl kinase domain [60]. One should recognize that while MD simulations were used with an unbiased classical all-atom AMBER-force field, the free-energy estimations were based on a hybrid coarse-grained model. A multidisciplinary approach combining simulations, functional assays, and mutagenesis has characterized the interdomain

coupling in the active SH3-SH2-Abl complex, suggesting that the SH2-KD interactions can allosterically stabilize the catalytically competent position of the αC-helix and thus exert control over the kinase activity [61]. The same system was also studied by using microsecond all-atom simulations and differential scanning calorimetry. The results from the dynamics of the SH3-SH2 tandem indicate a two-state switch, alternating between conformations observed in the autoinhibited and active complexes [62]. As a conclusion, computational studies of Abl and Src kinases regulation have indicated a complex interplay between the SH3 and SH2 domains, the SH2 linker, and the catalytic domain. These studies are in line with experimental results.

5 Discussion

Many important topics have not been discussed above, like phosphorylation effects and kinase dimerization. The main issue in modeling has hopefully been discussed sufficiently to draw some conclusions. The following conclusions are based on case studies in kinase modeling but, at the same time, are, after some modifications, generally applicable to all drug discovery type modeling efforts.

Successful modeling starts with the appropriate question or proper research hypothesis. This research question will lay the foundation for the selection of protein structures to be used. In the case of kinase modeling, one must know if potentially available protein structures (X-ray, Cryo-EM, or homology modeling) are in a biologically relevant state. The next critical point is the correct ligand/library preparation. Far too often, ligand tautomers/protomers are not based on detailed studies, but modelers trust too much in automatic procedures. The third critical point is too high confidence in docking and MD simulation results. Docking can, after all, create some kind of binding pose to almost all of the compounds in virtual screening libraries, although only a very small number of molecules are actually binding the target protein. The same is true for MD simulations. Many papers show single 100 ns simulations time stating that this is enough to identify binding/association/affinity. Since biological assays are usually done as triplicates, we should ask why this is not done with MD simulations [63, 64]. Instead of believing in one individual binding pose proposed by docking or short MD simulation, one should run several computational experiments with different setups, repeat MD simulations, and study how robust the proposed binding mode is to small changes in the system. At the same time, one should not think that empirical data are always superior over computational results. As discussed above, X-ray structures often do have issues affecting ligand structure, protein conformation, and structural determination, and sometimes the whole protein structure is wrong [65, 66]. This means that like in modeling and biological assays, one must look at all the structural biology data and combine information from different sources.

Most of us like good food and wine/beer/water. We also know that good food and drink cannot be created if we are using bad and rotten raw materials or dirty water.

The same is true for modeling and science in general. If the used method is not good, if protein structures are not adequate, or the ligands are not properly processed, one cannot obtain good data.

Compliance with Ethical Standards

Conflict of Interest: The author declares that he has no conflict of interest.

Funding: The author received no external funding during the preparation of this manuscript.

Ethical Approval: This article does not contain any studies with human participants or animals performed by the author.

Informed Consent: This article does not contain any studies which required DNA samples.

References

1. Fry DC, Kuby SA, Mildvan AS (1985) NMR studies of the MgATP binding site of adenylate kinase and of a 45-residue peptide fragment of the enzyme. Biochemistry 24:4680–4694. https://doi.org/10.1021/bi00338a030
2. Taylor SS, Kornev AP (2010) Protein kinases: evolution of dynamic regulatory proteins. Trends Biochem Sci 36:65–77. https://doi.org/10.1016/j.tibs.2010.09.006
3. Cramer RD, Patterson DE, Bunce JD (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967. https://doi.org/10.1021/ja00226a005
4. 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 464:427–430. https://doi.org/10.1038/nature08902
5. Durrant DE, Morrison DK (2017) Targeting the Raf kinases in human cancer: the Raf dimer dilemma. Br J Cancer 118:3–8. https://doi.org/10.1038/bjc.2017.399
6. Wood ER, Truesdale AT, McDonald OB, Yuan D, Hassell A, Dickerson SH, Ellis B, Pennisi C, Horne E, Lackey K, Alligood KJ, Rusnak DW, Gilmer TM, Shewchuk L (2004) A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib). Cancer Res 64:6652–6659. https://doi.org/10.1158/0008-5472.can-04-1168
7. Asquith CRM, Tizzard GJ, Bennett JM, Wells CI, Elkins JM, Willson TM, Poso A, Laitinen T (2020) Targeting the water network in cyclin G associated kinase (GAK) with 4-anilino-quin (az)oline inhibitors. ChemMedChem 15(13):1200–1215. https://doi.org/10.1002/cmdc.202000150
8. Hansch C, Fujita T (1964) p-σ-π analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626. https://doi.org/10.1021/ja01062a035
9. Meng EC, Shoichet BK, Kuntz ID (1992) Automated docking with grid-based energy evaluation. J Comput Chem 13:505–524. https://doi.org/10.1002/jcc.540130412
10. Roskoski R (2016) Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. Pharmacol Res 103:26–48. https://doi.org/10.1016/j.phrs.2015.10.021
11. Martz KE, Dorn A, Baur B, Schattel V, Goettert MI, Mayer-Wrangowski SC, Rauh D, Laufer SA (2012) Targeting the hinge glycine flip and the activation loop: novel approach to potent p38α inhibitors. J Med Chem 55:7852. https://doi.org/10.1021/jm300951u
12. Pantsar T, Poso A (2018) Binding affinity via docking: fact and fiction. Molecules 23:1899. https://doi.org/10.3390/molecules23081899

13. Forli S (2015) Charting a path to success in virtual screening. Molecules 20:18732–18758. https://doi.org/10.3390/molecules201018732

14. Huse M, Kuriyan J (2002) The conformational plasticity of protein kinases. Cell 109:275–282. https://doi.org/10.1016/s0092-8674(02)00741-9

15. Jacobs MD, Caron PR, Hare BJ (2007) Classifying protein kinase structures guides use of ligand-selectivity profiles to predict inactive conformations: structure of lck/imatinib complex. Proteins 70:1451–1460. https://doi.org/10.1002/prot.21633

16. Lee M, Balupuri A, Jung Y, Choi S, Lee A, Cho Y, Kang N (2018) Design of a novel and selective IRAK4 inhibitor using topological water network analysis and molecular modeling approaches. Molecules 23:3136. https://doi.org/10.3390/molecules23123136

17. Walter NM, Wentsch HK, Bührmann M, Bauer SM, Döring E, Mayer-Wrangowski S, Sievers-Engler A, Willemsen-Seegers N, Zaman G, Buijsman R, Lämmerhofer M, Rauh D, Laufer SA (2017) Design, synthesis, and biological evaluation of novel type I(1)/2 p38α MAP kinase inhibitors with excellent selectivity, high potency, and prolonged target residence time by interfering with the R-spine. J Med Chem 60:8027–8054. https://doi.org/10.1021/acs.jmedchem.7b00745

18. Ortuso F, Amato R, Artese A, D'antona L, Costa G, Talarico C, Gigliotti F, Bianco C, Trapasso F, Schenone S, Musumeci F, Botta L, Perrotti N, Alcaro S (2014) In silico identification and biological evaluation of novel selective serum/glucocorticoid-inducible kinase 1 inhibitors based on the pyrazolo-pyrimidine scaffold. J Chem Inf Model 54:1828–1832. https://doi.org/10.1021/ci500235f

19. Slynko I, Schmidtkunz K, Rumpf T, Klaeger S, Heinzlmeir S, Najar A, Metzger E, Kuster B, Schüle R, Jung M, Sippl W (2016) Identification of highly potent protein kinase C-related kinase 1 inhibitors by virtual screening, binding free energy rescoring, and in vitro testing. ChemMedChem 11:2084–2094. https://doi.org/10.1002/cmdc.201600284

20. Singh N, Tiwari S, Srivastava KK, Siddiqi MI (2015) Identification of novel inhibitors of Mycobacterium tuberculosis PknG using pharmacophore based virtual screening, docking, molecular dynamics simulation, and their biological evaluation. J Chem Inf Model 55:1120–1129. https://doi.org/10.1021/acs.jcim.5b00150

21. Wang Y, Dai Y, Wu X, Li F, Liu B, Li C, Liu Q, Zhou Y, Wang B, Zhu M, Cui R, Tan X, Xiong Z, Liu J, Tan M, Xu Y, Geng M, Jiang H, Liu H, Ai J, Zheng M (2019) Discovery and development of a series of Pyrazolo[3,4-d]pyridazinone compounds as the novel covalent fibroblast growth factor receptor inhibitors by the rational drug design. J Med Chem 62:7473–7488. https://doi.org/10.1021/acs.jmedchem.9b00510

22. Xu M, Yu L, Wan B, Yu L, Huang Q (2011) Predicting inactive conformations of protein kinases using active structures: conformational selection of type-II inhibitors. PLoS One 6: e22644. https://doi.org/10.1371/journal.pone.0022644

23. Modi V, Dunbrack RL (2019) Defining a new nomenclature for the structures of active and inactive kinases. Proc Natl Acad Sci U S A 116:6818–6827. https://doi.org/10.1073/pnas.1814279116

24. Bethke E, Pinchuk B, Renn C, Witt L, Schlosser J, Peifer C (2016) From type I to type II: design, synthesis, and characterization of potent pyrazin-2-ones as DFG-out inhibitors of PDGFRβ. ChemMedChem 11:2664–2674. https://doi.org/10.1002/cmdc.201600494

25. Zhao H, Caflisch A (2013) Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. Bioorg Med Chem Lett 23:5721–5726. https://doi.org/10.1016/j.bmcl.2013.08.009

26. Pedreira JGB, Nahidino P, Kudolo M, Pantsar T, Berger B-T, Forster M, Knapp S, Laufer S, Barreiro EJ (2020) Bioisosteric replacement of arylamide-linked spine residues with N-Acylhydrazones and selenophenes as a design strategy to novel dibenzosuberone derivatives as type I 1/2 p38α MAP kinase inhibitors. J Med Chem 63(13):7347–7354. https://doi.org/10.1021/acs.jmedchem.0c00508

27. Henzler-Wildman K, Kern D (2007) Dynamic personalities of proteins. Nature 450:964–972. https://doi.org/10.1038/nature06522

28. Frauenfelder H, Sligar S, Wolynes P (1991) The energy landscapes and motions of proteins. Science 254:1598–1603. https://doi.org/10.1126/science.1749933

29. Pisani P, Caporuscio F, Carlino L, Rastelli G (2016) Molecular dynamics simulations and classical multidimensional scaling unveil new metastable states in the conformational landscape of CDK2. PLoS One 11:e0154066. https://doi.org/10.1371/journal.pone.0154066

30. Pande VS, Beauchamp K, Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask. Methods 52:99–105. https://doi.org/10.1016/j.ymeth.2010.06.002

31. Meng Y, Gao C, Clawson DK, Atwell S, Russell M, Vieth M, Roux B (2018) Predicting the conformational variability of Abl tyrosine kinase using molecular dynamics simulations and Markov state models. J Chem Theory Comput 14:2721–2732. https://doi.org/10.1021/acs.jctc.7b01170

32. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11:3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

33. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, MacKerell AD (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. J Chem Theory Comput 8:3257–3273. https://doi.org/10.1021/ct300400x

34. Vanommeslaeghe K, MacKerell AD (2014) CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochim Biophys Acta 1850:861–871. https://doi.org/10.1016/j.bbagen.2014.08.004

35. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang J, Wang L, Lupyan D, Dahlgren MK, Knight JL, Kaus JW, Cerutti DS, Krilov G, Jorgensen WL, Abel R, Friesner RA (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. J Chem Theory Comput 12:281–296. https://doi.org/10.1021/acs.jctc.5b00864

36. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, Dahlgren MK, Mondal S, Chen W, Wang L, Abel R, Friesner RA, Harder ED (2019) OPLS3e: extending force field coverage for drug-like small molecules. J Chem Theory Comput 15(3):1863–1874. https://doi.org/10.1021/acs.jctc.8b01026

37. Asthana A, Wheeler DR (2013) A polarizable reactive force field for water to enable molecular dynamics simulations of proton transport. J Chem Phys 138:174502. https://doi.org/10.1063/1.4798457

38. Jing Z, Liu C, Cheng SY, Qi R, Walker BD, Piquemal J-P, Ren P (2019) Polarizable force fields for biomolecular simulations: recent advances and applications. Annu Rev Biophys 48:371–394. https://doi.org/10.1146/annurev-biophys-070317-033349

39. Paul F, Meng Y, Roux B (2020) Identification of druggable kinase target conformations using Markov model metastable states analysis of apo-Abl. J Chem Theory Comput 16:1896–1912. https://doi.org/10.1021/acs.jctc.9b01158

40. Sultan MM, Kiss G, Pande VS (2018) Towards simple kinetic models of functional dynamics for a kinase subfamily. Nat Chem 10:903–909. https://doi.org/10.1038/s41557-018-0077-9

41. Sultan MM, Denny RA, Unwalla R, Lovering F, Pande VS (2017) Millisecond dynamics of BTK reveal kinome-wide conformational plasticity within the apo kinase domain. Sci Rep 7:15604. https://doi.org/10.1038/s41598-017-10697-0

42. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196. https://doi.org/10.1021/jm051256o

43. 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 47:1739–1749. https://doi.org/10.1021/jm0306430

44. Asquith CRM, Maffuid KA, Laitinen T, Torrice CD, Tizzard GJ, Crona DJ, Zuercher WJ (2019) Targeting an EGFR water network with 4-anilinoquin(az)oline inhibitors for chordoma. ChemMedChem 14:1693–1700. https://doi.org/10.1002/cmdc.201900428

45. Heider F, Pantsar T, Kudolo M, Ansideri F, Simone AD, Pruccoli L, Schneider T, Goettert MI, Tarozzi A, Andrisano V, Laufer SA, Koch P (2019) Pyridinylimidazoles as GSK3β inhibitors: the impact of tautomerism on compound activity via water networks. ACS Med Chem Lett 10:1407–1414. https://doi.org/10.1021/acsmedchemlett.9b00177

46. Khandogin J, Brooks CL (2005) Constant pH molecular dynamics with proton tautomerism. Biophys J 89:141–157. https://doi.org/10.1529/biophysj.105.061341

47. Lee MS, Salsbury FR, Brooks CL (2004) Constant-pH molecular dynamics using continuous titration coordinates. Proteins 56:738–752. https://doi.org/10.1002/prot.20128

48. Tsai C-C, Yue Z, Shen J (2019) How electrostatic coupling enables conformational plasticity in a tyrosine kinase. J Am Chem Soc 141:15092–15101. https://doi.org/10.1021/jacs.9b06064

49. Liu R, Yue Z, Tsai C-C, Shen J (2019) Assessing lysine and cysteine reactivities for designing targeted covalent kinase inhibitors. J Am Chem Soc 141:6553–6560. https://doi.org/10.1021/jacs.8b13248

50. Kuzmanic A, Sutto L, Saladino G, Nebreda AR, Gervasio FL, Orozco M (2017) Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. Elife 6:e22175. https://doi.org/10.7554/elife.22175

51. Zhang Y-Y, Wu J-W, Wang Z-X (2011) Mitogen-activated protein kinase (MAPK) phosphatase 3-mediated cross-talk between MAPKs ERK2 and p38α. J Biol Chem 286:16150–16162. https://doi.org/10.1074/jbc.m110.203786

52. Tokunaga Y, Takeuchi K, Takahashi H, Shimada I (2014) Allosteric enhancement of MAP kinase p38α's activity and substrate selectivity by docking interactions. Nat Struct Mol Biol 21:704–711. https://doi.org/10.1038/nsmb.2861

53. 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 115: E11894–E11903. https://doi.org/10.1073/pnas.1811158115

54. 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:e32766. https://doi.org/10.7554/elife.32766

55. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science 289:1938–1942. https://doi.org/10.1126/science.289.5486.1938

56. Ahuja LG, Taylor SS, Kornev AP (2019) Tuning the "violin" of protein kinases: the role of dynamics-based allostery. IUBMB Life 71:685–696. https://doi.org/10.1002/iub.2057

57. Kornev AP, Taylor SS (2015) Dynamics-driven allostery in protein kinases. Trends Biochem Sci 40:628–647. https://doi.org/10.1016/j.tibs.2015.09.002

58. McClendon CL, Kornev AP, Gilson MK, Taylor SS (2014) Dynamic architecture of a protein kinase. Proc Natl Acad Sci U S A 111:E4623–E4631. https://doi.org/10.1073/pnas.1418402111

59. Kornev AP (2020) Allostery explained through synchronized oscillators and fractal networks. Biophys J 118:208a. https://doi.org/10.1016/j.bpj.2019.11.1248

60. Dölker N, Górna MW, Sutto L, Torralba AS, Superti-Furga G, Gervasio FL (2014) The SH2 domain regulates c-Abl kinase activation by a cyclin-like mechanism and remodulation of the hinge motion. PLoS Comput Biol 10:e1003863. https://doi.org/10.1371/journal.pcbi.1003863

61. Tse A, Verkhivker GM (2015) Molecular dynamics simulations and structural network analysis of c-Abl and c-Src kinase core proteins: capturing allosteric mechanisms and communication pathways from residue centrality. J Chem Inf Model 55:1645–1662. https://doi.org/10.1021/acs.jcim.5b00240

62. Fajer M, Meng Y, Roux B (2016) The activation of c-Src tyrosine kinase: conformational transition pathway and free energy landscape. J Phys Chem B 121:3352–3363. https://doi.org/10.1021/acs.jpcb.6b08409

63. Braun E, Gilmer J, Mayes H, Mobley D, Prasad S, Zuckerman D, Monroe J (2018) Best practices for foundations in molecular simulations [Article v1.0]. Living J Comput Mol Sci 1:5957. https://doi.org/10.33011/livecoms.1.1.5957

64. Grossfield A, Patrone P, Roe D, Schultz A, Siderius D, Zuckerman D (2018) Best practices for quantification of uncertainty and sampling quality in molecular simulations [Article v1.0]. Living J Comput Mol Sci 1:5067. https://doi.org/10.33011/livecoms.1.1.5067

65. Niedzialkowska E, Gasiorowska O, Handing KB, Majorek KA, Porebski PJ, Shabalin IG, Zasadzinska E, Cymborowski M, Minor W (2016) Protein purification and crystallization artifacts: the tale usually not told. Protein Sci 25:720–733. https://doi.org/10.1002/pro.2861

66. Cooper DR, Porebski PJ, Chruszcz M, Minor W (2011) X-ray crystallography: assessment and validation of protein–small molecule complexes for drug discovery. Expert Opin Drug Discov 6:771–782. https://doi.org/10.1517/17460441.2011.585154