RSC Medicinal Chemistry

ROYAL SOCIETY OF CHEMISTRY

REVIEW

Cite this: *RSC Med. Chem.*, 2021, **12**, 1491

New perspectives in cancer drug development: computational advances with an eye to design

Matteo Castelli,[a] Stefano A. Serapian,[a] Filippo Marchetti,[a] Alice Triveri,[a] Valentina Pirota,[a] Luca Torielli,[b] Simona Collina,[b] Filippo Doria,[a] Mauro Freccero[a] and Giorgio Colombo*[a]

Received 9th June 2021,
Accepted 6th July 2021

DOI: 10.1039/d1md00192b

rsc.li/medchem

Computational chemistry has come of age in drug discovery. Indeed, most pharmaceutical development programs rely on computer-based data and results at some point. Herein, we discuss recent applications of advanced simulation techniques to difficult challenges in drug discovery. These entail the characterization of allosteric mechanisms and the identification of allosteric sites or cryptic pockets determined by protein motions, which are not immediately evident in the experimental structure of the target; the study of ligand binding mechanisms and their kinetic profiles; and the evaluation of drug–target affinities. We analyze different approaches to tackle challenging and emerging biological targets. Finally, we discuss the possible perspectives of future application of computation in drug discovery.

Introduction

In 2020, the US Food and Drug Administration (FDA) approved 53 new products, the second-highest number in 20 years, just short of the record of 59 established in 2018. The interest in cancer is notable, with about 35% of approved products for this disease dominating the list.

One key factor that strongly contributed to this landmark achievement is the emergence of new knowledge and approaches for target identification, screening and drug design.

Computational studies are now recognized to play an increasingly crucial role in the evolution of drug discovery. The development of new hardware, software and their integration, coupled to the increasing amount of data on drug activities and the development of new algorithms to analyze them, can significantly contribute to reduce the costs and time required to advance a hit compound into a new administrable medicine.

The classic roles of computational chemistry and biology range from structure refinement to the generation of poses for ligands in targets, as well as from ligand or structure based virtual screening to similarity searches and *de novo* design. Most of these approaches are centered around the concept that drug candidates typically need to engage structurally well-defined active/binding sites on the target: specific side chains and cavities around such sites can be exploited to guide the modification of the initial lead *via* functional groups aimed to generate novel binding interactions.

Fundamental studies and technological advances have revealed that, in many cases, engaging an active site may not be sufficient to disarm the pathologic activity of a biological target. Changes in expression levels, mutations in sites distal from the active site, complex conformational changes, post-translational modifications, the emergence of drug resistance, have all been recognized to contribute to many pathological states.

Novel computational approaches, and biomolecular simulations in particular, are essential tools for understanding the molecular complexities and interconnections among all these mechanisms. On this basis they can provide new opportunities for cancer drug design and development.

As an example, recent work on epidermal growth factor receptor (EGFR), an important cancer oncogene, has shown how the combination of imaging technologies and simulations could reveal the determinants of assembly of ligand-free receptor polymer chains on the extracellular membrane, supporting the hypothesis that dysregulated species bear populations of symmetric and asymmetric kinase dimers that coexist in equilibrium. This structural characterization of the assembly bears clear relevance for future drug design efforts.[1]

Relevant applications entail revealing cryptic sites, that might not even be present in the crystallographic model of the target of interest, sampling complex free-energy landscapes, revealing the impact of mutations on the activation states of oncogenes, unveiling allosteric regulation

[a] *Department of Chemistry, University of Pavia, via Taramelli 12, 27100 Pavia, Italy. E-mail: g.colombo@unipv.it*
[b] *Department of Drug Sciences, Medicinal Chemistry and Pharmaceutical Technology Section, University of Pavia, via Taramelli 12, 27100 Pavia, Italy*

pathways, and estimating the kinetic parameters of drug binding and unbinding or drug residence times. Furthermore, it is worth noting that drug effects extend on multiple scales: the molecular interactions established by the drug with the target alter its function, impact its interaction network and eventually reverberate at larger multimolecular and cellular scales. Simulations that encompass and interconnect such different scales are becoming more and more important in drug design and will likely grow in their importance for our understanding of complex mechanisms and ultimately for our ability to predict pharmacological and biological effects.

In this review, we discuss how advanced computational approaches can help realize the promises of discovering, designing and optimizing novel cancer therapeutics. We will present our views on emerging methods and applications in this fascinating field at the crossroads between chemistry, biology, biophysics and computational sciences, outlining how they can be combined to extend the chemical space of new molecules for cancer treatment.

Is allostery so cryptic? Binding pocket detection and allosteric regulation

The investigation of allostery has become a hot field of investigation as it offers promising strategies for new inhibitor discovery. Understanding allosteric communication pathways has proven key in shedding light on functional regulation mechanisms, on the impact of mutations on the levels of enzymatic activation, and on the determinants of drug resistance. In this context, recent work by Galdadas and coworkers[2] combined equilibrium and nonequilibrium molecular dynamics simulations to characterize allosteric effects in two prototypical class A β-lactamases, TEM-1 and KPC-2. This work, while not strictly related to cancer, tackles the problem of drug resistance, which is one of the main challenges facing drug discovery. In particular, the authors look at how the dynamics and shape of the proteins respond to the presence or removal of an allosteric ligand. The picture provided by equilibrium and non-nonequilibrium simulations unveils communication pathways that connect very distant sites (more than 30 Å apart) in the proteins and, importantly, where more than 50% of clinically relevant amino acid substitutions concentrate on these identified signal transduction pathways.

A combination of classical equilibrium and enhanced sampling simulations[3] has been applied to the study of PI3Kα, a member of the phosphoinositide-3 kinase (PI3K) family (Fig. 1), which controls several cellular responses such as cell proliferation, survival, and motility by catalyzing the phosphorylation of the inositol lipid PIP2 at the 3′ position of the inositol ring, generating the signalling molecule PIP3. Mutations in PI3Kα cause enhanced signalling, exacerbating the above-mentioned cellular activities and contributing to oncogenesis.[4]

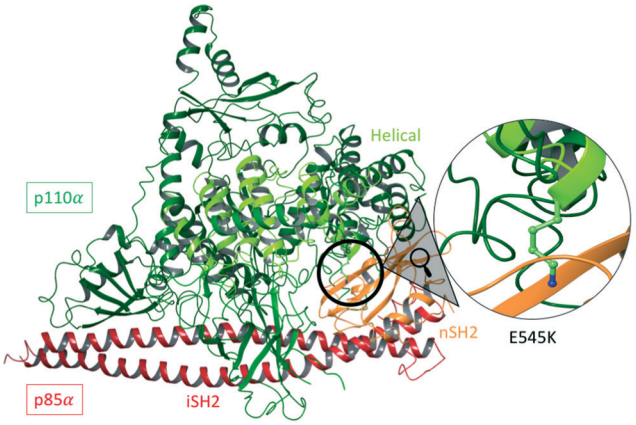

Fig. 1 PI3Kα subunit organization (PDB: 4OVU) with salient functional domains labelled. The mutation E545K lies at the interface between the nSH2 and helical domains of the p85α and p110α subunits, respectively. The inset zooms in on E545K rendered as ball-and-stick.

Here, the authors focused on one particularly important mutation for tumors, namely E545K. This mutation increases PI3Kα lipid kinase activity.[5]

Classical molecular dynamics (MD) and multiple walkers metadynamics simulations of the p110α (catalytic) and p85α (regulatory) subunits of PI3Ka indicate that the charge reversal determined by the mutation has complex consequences at the level of both protein structure and dynamics. E545K lies at the interface between the nSH2 (p85α) and helical (p110α) domains: the simulations indicate that two displacements originate from the mutation. One is the detachment of nSH2 domain from the helical domain, and the other is the sliding along the helical domain. The former motion leaves the activation-loop intact, as the loop's interactions with the iSH2 domain are maintained, which overall leaves the catalytic subunit unaffected. The latter motion, in contrast, directly influences the catalytic subunit by establishing new contacts between the positively charged Lys545 with residues of the linker that connects the nSH2 with the iSH2 domain. This in turn determines the onset of conformational changes in iSH2 that end up breaking up the regulatory contacts between this domain and the activation loop, thus increasing activity allosterically. Eventually, the end result of this perturbation is to unlock the activation-loop facilitating its exploration of active-like conformations, especially in the presence of PIP2.

These types of complex modes of action and regulation of activities, *via* the modulation of the cross-talk between ligands or mutation sites and the rest of the protein structure, have been observed in other multidomain proteins and appear to facilitate functionally oriented movements.[6,7]

Barros *et al.*[8] combined long-timescale simulations with Markov state models (MSM) to study how mutations in the tumor suppressor protein p53 determine which structural elements have a significant influence on the slowest motions of the protein (see Fig. 2). Because of its DNA-binding and regulation activities, mutations in p53 are found in the large

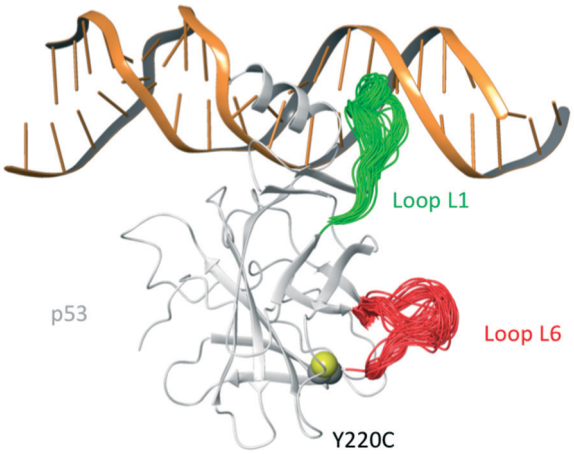

Fig. 2 p53 DNA-binding domain in complex with DNA (PDB: 1TSR), highlighting conformational variability in loops L1 (green) and L6 (red) (PDB: 2FEJ). The Y220C mutation is rendered as spheres.

majority of human cancers. Not surprisingly, this protein has become the focus of intense research and strategies to rescue its activity by pharmacological chaperones have emerged as promising options in cancer therapy.[9] Upon comparing the simulations and the energy landscapes of the wildtype p53 and of the Y220C cancer mutant, the study by Barros *et al.* uncovers loop L6 as a determinant of the slowest motions of the protein. Interestingly, here computational studies are validated by NMR relaxation studies. The Y220C mutation stabilizes alternative conformations to those present in all experimental models, whereby the loop is extended and locates further away from the DNA-interacting surface. This modification of the structural dynamics reverberates allosterically on the conformational landscape of the functionally-important loop L1, ultimately leading to inactivation of the mutant. Here, the use of extremely long MD simulations and MSM analysis are used in synergy to unveil conformational and kinetic differences between two forms of this important protein. Furthermore, the results highlight the presence of a novel cryptic pocket in Y220C that can be used to design compounds acting as chemical chaperones to restore the functional wildtype-like state of p53.

Enhanced sampling methods are being increasingly appreciated for their potential in drug discovery. Metadynamics, for instance, uses a (combination of) collective variable(s) (CVs) to guide complex transitions in the systems of interest. A notable example of the application of metadynamics to cancer drug discovery is the identification of the binding site and the characterization of the binding poses of the inhibitor SSR128129E (SSR) on fibroblast growth factor receptor 1 (FGFR1).[10] FGFR1, through interactions with its endogenous ligands FGFs is one of the key factors in tumor angiogenesis. SSR was identified *via* high-throughput screening and its binding site was hypothesized to be located in the D3 domain of FGFR1. Yet the unstable and highly dynamic nature of the domain prevented the definition of the precise binding spot. In an elegant application of metadynamics, the authors combined CVs related to ligand binding to CVs relevant for the folding of the domain. Using a range of CVs that describe not only the binding process but also the folding of the domain, the authors simulated the reversible binding of SSR to a hidden pocket formed by the extension of a helix in D3. The validity of the models was tested and corroborated by mutation of residues identified as important for the binding mechanism. Most importantly, the results of simulations were used to design more potent inhibitors.

These results show how the combination of different simulative approaches, followed by integration with experimental biochemical, mutational, and functional data, can offer potent new tools to gain high-resolution atomistic insight into the inner workings of allosteric communication in enzymes. It must be underlined here that the true potential of simulations in the quest for new active leads can be realized if new targetable binding sites are revealed and characterized. In this framework, targeting functionally relevant hinges or pockets (even if distal from the active site) can provide novel platforms to guide drug discovery. In many cases, potentially interesting sites are cryptic, *i.e.* they are not visible in proteins crystallized without a ligand and can be exposed in crystals only upon a binding event or a dynamic conformational rearrangement following a reaction. Work in the chaperone field is one example: the characterization of the mechanical coordination between binding and processing of the natural substrate at the nucleotide binding site and distal regions of Hsp90 or Hsp70 chaperones was used as the starting point for the identification of druggable pockets. The design of ligands with functional groups selected to target the complementary functionalities on such putative allosteric pockets generated a series of new modulators of the functions of the two chaperones. The novel leads showed interesting isoform selectivity properties and anticancer activities.[11–16]

In the cancer field, Spinello and coworkers[17] screened for allosteric binding sites on aromatase, an important enzyme in the development of breast cancer. Previous experimental findings in fact pointed to the possibility of allosterically inhibit the enzyme. The authors combined *in silico* screening, molecular dynamics and free energy simulations, supported by enzymatic and cell-based assays, to identify five leads that are demonstrated to inhibit the enzyme without directly competing with substrates directed at the active site (Fig. 3). In the context of breast cancer, the availability of alternative binding sites on one of the enzymes that are determinant for hormone synthesis may offer an important therapeutic alternative to estrogen deprivation therapy.

The same group employed long atomistic simulations NetWork Analysis (NWA) to the study of the Arp2/3 molecular machine, fundamental for cell motility and migration, whose aberrant functions favor cancer invasion and metastasis.[18] Here the authors identify the mechanistic elements that trigger the conformational transitions initiated at the ATP-allosteric binding site. The results show that while ATP-induced motions are ordered and synchronized, the binding

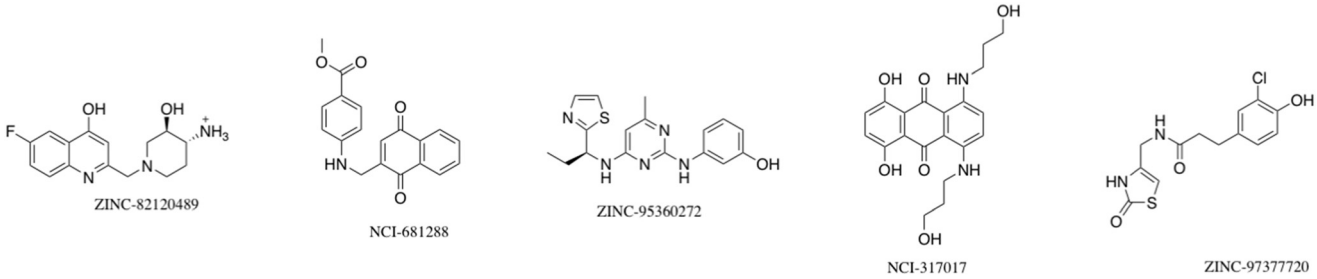

Fig. 3 Molecular structures of the five lead compounds capable of allosteric inhibition of the aromatase.

of allosteric inhibitors perturb functional motions by desynchronizing them. This ends up hindering the protein transition towards activation. Interestingly, the data establish a framework for the design of better ligands able to block infiltrative cancer migration.

Other elegant examples have appeared in the literature where cryptic pockets are identified and characterized based on the equilibrium fluctuations of apo proteins. An elegant review by the Bowman group reports on different computational and experimental tools and pipelines developed by the authors over the years to reveal targetable pockets from protein dynamics.[19] One key aspect the authors underline is the importance of correctly considering the relative weights of different structures sampled during simulations, which underlies the impact on function a certain molecule can have by modulating the balance of conformational populations.

In this context, the authors introduced a sampling algorithm called FAST (fluctuation amplification of specific traits) to enable extensive conformational sampling on commodity hardware.[20] This is achieved by balancing the trade-offs between focused searches around promising solutions (exploitation) and trying novel solutions (exploration). The algorithm is based on the observation that many physical properties have an overall gradient in conformational space, similar to the energetic gradients that are known to guide proteins to their folded states. The authors demonstrate this *via* a retrospective analysis of existing Markov state models (MSMs). FAST is shown to exploit such gradients through the recognition and amplification of structural fluctuations along gradients that optimize a selected physical property, *via* the overcoming of barriers that interrupt these overall gradients, and by rerouting to discover alternative paths when faced with insurmountable barriers. The authors prove that FAST efficiently identifies cryptic binding pockets, and unveils preferential paths between structures while providing the proper thermodynamics and kinetics of the events.

The same group further developed methods to identify potentially interesting binding pockets by segmenting protein structures into clusters of residues that undergo cooperative changes in their solvent exposure,[21] or to reveal allosteric relationships by identifying correlations among the rotameric and dynamic properties of proteins.[22]

These advanced approaches can also be combined to a number of experimental methods.[23]

The examples reported above show that the true potential of simulations in the quest for new active leads can even be enhanced if new targetable, cryptic binding sites are revealed and characterized. Besides facilitating the discovery of chemical tools to investigate complex mechanisms, these sites might provide unexplored opportunities to target proteins for which classic drug design approaches fail. These include difficult targets often deemed undruggable.

Simulative approaches can efficiently reveal connections between different sites and transient opening or closing of cryptic pockets. We should note here that this may not always be sufficient to support the design of ligands targeting that site. The first step in this direction is to explore the pocket to define which functional groups or substructures it might proficiently accommodate. MD-based mixed solvent simulations are the first step in this direction.[24] Inspired by X-ray and NMR experiments demonstrating that organic solvents can bind precisely at locations alternative to the active site, mixed solvent simulations can detect cryptic pockets and give access to the calculation of interaction energies between the target and ligand fragments. Seco *et al.*[25] showed that this principle could be successfully applied to different pharmacological targets.

The SILCS (Site Identification by Ligand Competitive Saturation) approach by the MacKerell group[26] introduced the use of a combination of small fragments in a mixture to map the preferential positioning of groups that are typically part of drug molecules, such as small aliphatic or aromatic groups, and hydrogen bond donors and acceptor. Since different ligands presenting the various functionalities are simultaneously present with the target in the simulation box, the method simulates the actual competition among the fragments and generates a 3D free energy map of fragment binding, highlighting the most favorable target–fragment interactions. When applied to the oncoprotein BCL-6, a driver of B-cell lymphoma, the SILCS maps recapitulate the crystallographic binding modes of peptides known to bind the protein. Building on these results, Cheng *et al.* developed drug-like inhibitors of BCL6 protein–protein interactions (PPIs), combining the SILCS approach with NMR-screening and medicinal chemistry evolution of initial leads.[27]

Bakan *et al.*[28] extended these concepts to quantitatively evaluate the druggability of a site. Specifically, these authors

combined the simulation of the binding dynamics of a set of diverse probe molecules representative of different approved drugs to the identification of potential binding sites and the evaluation of binding affinities as a function of the geometry and energetics of clusters of bound probes. The method was shown to correctly identify the binding site and affinity of known drugs for protein tyrosine phosphatase 1B (PTP1B), lymphocyte function-associated, antigen 1, and p38 mitogen-activated protein kinase. Interestingly, the method showed the ability to identify binding spots in very challenging cases.[28]

An efficient approach to identifying cryptic sites has been recently introduced by Gervasio and coworkers. The method, called SWISH (sample water interactions through scaled Hamiltonians), facilitates the exploration of hydrophobic and (partially) buried regions using a Hamiltonian replica exchange strategy that modifies the interaction of apolar carbons and sulfurs with water oxygen.[29,30] At higher replicas, the protein becomes less hydrophobic and this ultimately allows it exploration compared to plain MD. SWISH can be combined with probe molecules in the simulation box, whose work is to enter the otherwise transiently open cryptic pockets thus stabilizing them. This method is often coupled to the use of restraints between residues that are particularly important for the structural stability of the protein, since modifying residue–residue interactions may result in protein unfolding. Interestingly, SWISH can be combined to metadynamics both for examining the opening of the pockets and for modeling the diffusion of probes into the target.

The capacity of specific regions, distinct and/or distal from a classic active site, to host a small molecule requires the display of side chains able to determine favorable interactions with the ligand. Khazanov and Carlson[31] nicely pointed out that understanding the general composition of these sites is important to shed light on the druggability of the different elements of the proteome and understand its functional diversity. The authors analyzed 3295 non-redundant proteins with 9114 non-redundant binding sites to identify residues over-represented in binding regions *versus* the rest of the protein surface. To this end they took advantage of the Binding MOAD (Mother of All Databases) database,[32] one of the largest curated sets of protein–ligand complexes, developed in the Carlson group. On this basis they classified ligands as "valid", or biologically relevant, or "invalid", representing artifacts of molecules present *e.g.* in crystallization media and bound to the protein without any known function.

Contacts with the respective proteins are found to differ between the two sets of molecules. Interestingly, the identity of residues in biologically relevant binding sites differs from that of pockets that bind artefactual and non-functional small molecules. Furthermore, the composition of the "valid" binding is distinct from that of the rest of the protein surface. This type of knowledge and analysis can nicely complement methods for the discovery of novel binding sites.

Affinity for the target: evaluating proficient binding

Once functionally relevant novel binding sites have been discovered and leads identified, the key questions entail the evaluation of the affinity (binding free energy) and, possibly, the estimation of the residence time of the ligand in the binding site. Both aspects are strictly related to the key issue of target engagement.

The importance of methods for the estimation of target–ligand binding free energies has grown dramatically in recent years. In particular, the improvement in force fields, sampling methods, the speed up of MD simulations have made free energy estimation more and more accessible. Indeed, the use of methods to calculate absolute and relative free energies of binding is being increasingly incorporated in drug discovery protocols,[33–37] also at the industrial level. In most applications of interest for drug development, the focus is on the calculation of the relative free energy of binding of newly designed compounds with respect to an initial lead. In this context, most approaches are variations on the themes of free energy perturbation (FEP) and thermodynamic integration (TI), which require MD or Monte Carlo (MC) sampling to determine the free energy differences between related ligands.[36]

These methods are somewhat less efficient when the compounds to be analyzed are chemically very different, they belong to distinct chemical series, or large conformational changes are involved.[38,39] In such cases, collective-variable-based free energy calculation methods, such as metadynamics[40–42] or umbrella sampling,[43] are the preferred choices. These methods have been used to compute free energies of binding trajectories, even for allosteric systems, where they have been able to capture the details of the coupling between protein–ligand recognition and the onset of allosteric perturbation within the protein structure once the ligand bound.[13]

While generally valid, these methods are still computationally very intensive and require a suitable definition of the relevant collective variables. The latter point, in particular, requires an in-depth knowledge of the system under study and of the determinant degrees of freedom underlying binding mechanisms. These requirements still limit the routine application of enhanced sampling methods in drug discovery pipelines.

The recent resurgence of machine learning (ML) methods and the explosion of deep learning applications in the chemical sciences have inspired the combination of rigorous physical chemistry methods with data-driven methodologies (Fig. 4).

The accuracy of the predictions is one of the limiting factors in FEP simulations. In the best cases, such calculations can deliver predictions within 1–2 kcal mol^{-1} of the experimental value.[44] However, this limit may still prioritize compounds for synthesis that eventually turn out not to have desirable potency and selectivity profiles. Improving on this limitation could thus significantly reduce

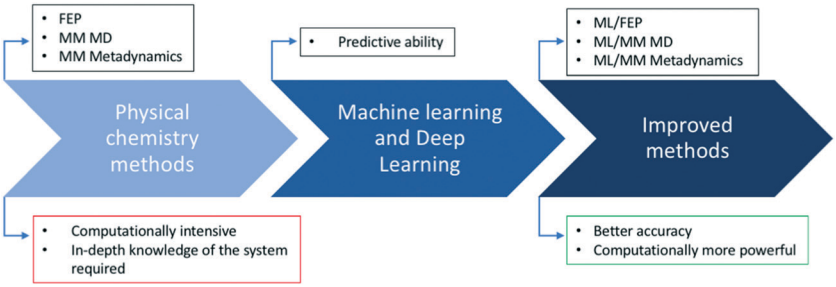

Fig. 4 A simplified scheme of how machine learning (ML) and deep learning approaches can improve standard physical chemistry methods.

the number and focus the chemotypes of proposed compounds in discovery campaigns, ensuring a more efficient and smooth evolution towards pharmacologically active molecules.

To improve the accuracy of FEP results, Scheen *et al.*[45] used machine learning regressions to evaluate empirical correction terms to be applied to FEP results. In this framework, the authors train a ML model to predict the error made in the evaluation of FEP (for the hydration process) using the set of experimental data present in the FreeSolv database. The calculated error is then used as an offset value to correct systematic errors made in normal FEP calculations. Naturally, this strategy assumes that given a sufficiently large training set, the model will be able to estimate the offset for a new set of predictions. Interestingly, when applied to the evaluation of hydration free energies of a set of compounds from the SAMPL4 competition, the mixed FEP/ML is shown to give better results than most free energy evaluation approaches, and to outperform pure ML methods.

A notable advancement is represented by the work of Rufa *et al.*[46] The problem they tackled is strictly linked to the limitations of existing classical-mechanics based force fields to capture the very complex physics of the interactions in biochemical systems. This problem is particularly aggravated in the estimation of torsional energetics: as new ligands are designed, a poor representation of their torsional profiles may have a large impact on binding free energy evaluations, especially if one considers that torsional motions can be coupled to other degrees of freedom. In the case of newly designed ligands, new torsional parameters often need to be calculated anew or refitted to high-level energy calculations. However, since significant conformational changes can take place upon transition from the unbound to the bound state, reparameterization of force field terms may not always be effective.[47] To overcome these problems, Smith *et al.* introduced the use of quantum machine learning potentials based on the use of neural networks.[48] In this framework, Rufa *et al.*[46] showed how hybrid machine learning/molecular mechanics (ML/MM) potentials can provide notable significant accuracy improvements in modeling protein–ligand binding affinities: they used a standard MM alchemical free energy calculation and then post-processed the results with a correction step to efficiently recover ML/MM free energies. They benchmarked their approach by studying kinase-inhibitor affinities showing a significant reduction in the errors in the estimation of free energies.

An additional source of errors in FEP calculations is the insufficient overlap between the various distributions used to take the system from the starting to the end state. Here, the use of ML approaches also appears to provide effective improvements over classical protocols. Targeted free energy perturbation (TFEP)[49] is one of the strategies to mitigate this problem, using a high-dimensional mapping in configuration space to increase the overlap of the distributions.

Wirnsberger *et al.*[50] reformulated TFEP in the framework of machine learning problem: here the mapping is modeled as a neural network that is optimized to increase overlap. The authors describe the possibility to apply this novel ML/TFEP strategy in normal periodic MD simulations, obtaining a reduction in the variance of the estimates of free energy values.

Machine learning approaches are being applied also to the calculation of absolute free energies. Evans and coworkers combined funnel metadynamics with an ML selected optimal pathlike variable to obtain accurate results for a set of 18 diverse ligands binding human epoxide hydrolase, a particularly complex target for drug discovery. Interestingly, the method demonstrated a good balance of computational cost and speed.[51]

This work highlights the importance of a proper selection of the collective variables required for enhanced sampling. There are indeed many difficult questions to face when setting out to pick CVs, especially considering that binding/molecular recognition problems are multi-dimensional: they range from the simple selection of initial configurations, or the selection of internal degrees of freedom in ligands and their potential couplings with receptor degrees of freedom, to the definition of the reaction coordinate for a protein–ligand binding reaction. Sultan and Pande[52] have recently proposed a data-driven approach inspired by the field of supervised machine learning to solve the problem of the selection of the "initial" CV(s). Using model peptides, they show how a different classifier can be used to reversibly sample slow structural transitions.

While based on the use of simple model systems, these combination strategies in which ML is coupled to physically rigorous treatments of the biological systems hold promise for future implementation in the evaluation of affinities in real drug design projects.

Finally, we envisage that ML and deep learning approaches may have significant impact on the prediction of the activities of allosteric ligands. Since they bind to pockets that are often far removed from the active site, direct correlations between the binding affinity values and effects on functions or cellular activity should not necessarily be expected. Indeed, the structure–activity relationships (SARs) of allosteric ligands are often complex. Indeed, some ligands may inhibit while others activate a protein activity: the question is how to prospectively discriminate such different effects. Using supervised ML approaches to post-process the results of long MD simulations of complexes between a ligand and an allosteric binder, Ferraro *et al.*[16] showed the possibility of discriminate between effective and non-effective allosteric inhibitors.

A further extension of this ML–MD approach, applied to a dataset of 133 ligands containing both inhibitors and activators of the biochemical activity of the protein, proved able to discriminate between the two groups with high accuracy.[53]

One of the major limitations in ML approaches is that no new knowledge is generated on the physical mechanisms underlying the modulation of protein function. The development of methods that couple the predictive activities of ML with the physical insights that physics-based methods can provide will most likely have a transformative impact on drug design.

Not only thermodynamics: residence times as compound/drug selection parameter

Recent advances in drug discovery have highlighted the kinetics of binding as a useful parameter for hit/lead selection. Indeed, residence times can be key determinants of efficacy.[54–56] Long residence time may extend the duration of the effect of a ligand and improve selectivity: if the drug engages one protein at a certain timepoint, it may be less available for establishing binding contacts with alternative targets, potentially minimizing side-effects. The role of residence times may be particularly relevant in allosteric drug discovery, where the ligand needs to be in contact with the receptor as long as necessary to trigger the dynamic signals that modulate functionally oriented motions. The consideration of the lifetime of a drug–target complex can thus improve the performances of drug discovery programs.

It comes as no surprise that in the last few years computational methodologies for calculating residence times have appeared and gained momentum in the drug discovery community.[57–59] Excellent in-depth reviews of the methods and general applications for these methods can be found in ref. 60 and 61.

One study involving the calculation of residence times for an important anticancer drug, dasatinib, binding to an important cancer target, Src kinase, is due to the Shaw group.[62] Furthermore, the authors studied another kinase inhibitor, namely PP1. Using unbiased, plain microsecond MD simulations, they could show the spontaneous binding of the two ligands to the active site, in poses very close to the ones observed in the respective crystal structures. The analysis of the trajectories allowed the authors to estimate the on-rates for the two compounds with results in close agreement with experiments.

De Cherchi *et al.*[63] also reported the use of plain MD to sample small molecule–protein recognition events. In their study of binding of the DADMe-immucillin-H to purine nucleoside phosphorylase, they ran multiple independent MD trajectories (14 runs amounting to a total of 13 microseconds of simulation time). The latter were then analyzed with an *ad hoc* machine learning algorithm to provide an estimate of the k_{on} rate close to experiment.

Enhanced sampling methods have been used to prioritize compounds based on their residence times: in this context, Callegari *et al.* ran comparative analyses among a series of compounds,[64] starting from the bound complex and extending until the ligand achieved the fully solvated state. The average simulation time taken by each species was then computed. This allowed a set of 10 arylpyrazole analogs targeting the Cdk8 protein to be correctly ranked into three classes according to short, medium, and long residence times as observed in the experiments (Fig. 5).

An alternative approach based on a different strategy to improve sampling is smoothed MD (sMD), which rescales the potential energy surface through the suitable choice of a smoothing factor. The end result is a potential energy surface with lower energy barriers among configurational basins, which makes it easier for a ligand to exit (or enter) a binding site. An advantage of sMD is that one does not have to set a reaction coordinate to guide unbinding. Indeed, the smoothing potential scaling is applied indiscriminately to the whole system. On the other hand, this requires restraints to be used on the protein to prevent artefactual unfolding events. The approach was used to study the unbinding of a series of ligands from HSP90, GRP78, A2A, and glucokinase, proving the possibility to correctly rank ligands according to their residence time.[65]

The problem of estimating binding kinetics has also been tackled with the application of Markov state models (MSM), whereby multiple trajectories are discretized through the definition of microstates according to select structural criteria or degrees of freedom. Subsequently, a transition probability matrix is constructed, which is eventually used to define the kinetic constants of the transitions among the microstates.[66,67] These approaches were then extended to learn how to adapt and increase sampling into the most interesting region of configurational space,[68] allowing to characterize complex drug–enzyme binding mechanisms.[69]

A nice example of combination of MSM analysis in drug discovery (though not strictly in the cancer area) is given by the paper by Hart and coworkers.[70] Here the authors use MSMs to identify hidden conformations and explore their role in determining TEM's specificity against

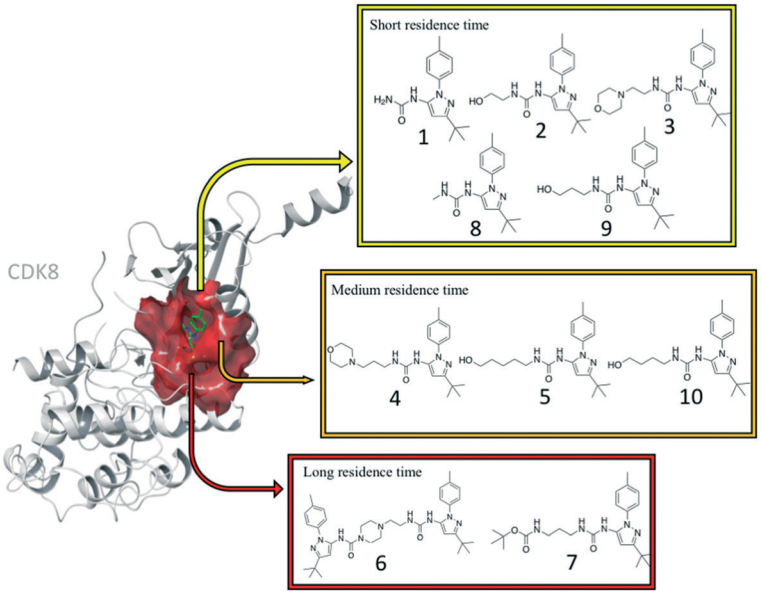

Fig. 5 Structure of Cdk8 in complex with compound **1** (PDB: 4F6S; sticks). The active site is represented in red. **1** and nine other arylpyrazole compounds reported by Callegari *et al.* are ranked according to their residence times.

ligands. These models are integrated with classical drug-design tools to generate "Boltzmann docking". In this framework, TEM specificity is correctly predicted by accounting for conformational heterogeneity. Interestingly, hidden states are identified whose populations correlate with activity against cefotaxime. The authors validate their model by mass spectroscopy and design novel variants to stabilize the hidden cefotaximase states showing that their populations predict activity against cefotaxime *in vitro* and *in vivo*.

The advent of multiscale methods is emerging as an important tool in the study of complex drug–receptor recognition problems. Jagger *et al.*[71] showed the possibility of ranking ligands according to their binding kinetics combining Brownian dynamics and classical MD with a milestoning model. Milestoning allows the calculation of the time evolution of complex processes, of which drug–receptor binding mechanisms are paradigmatic examples. In this framework, the method permits to study processes whose timescales largely exceed the ones accessible by plain MD. In milestoning, the system under exam is partitioned into cells by dividing hypersurfaces (milestones). The dynamics are reduced to transition events between the milestones that are computed *via* short MD simulations.[72]

The approach was successfully applied to the important enzyme superoxide dismutase in the study of the binding of superoxide anion (O_2^-), as well as to the N-terminal domain of troponin C with its natural calcium ion substrate (Ca^{2+}). The calculated k_{on} and k_{off} values appeared to be in good agreement with experimentally determined values.[73,74] One great advantage of milestoning is that it is highly parallelizable and computationally cheaper than comparable methods.

New challenges for new simulations: emerging nucleic acids targets

Recent advances in medicinal chemistry have brought nucleic acids (NA), especially RNA and specific tertiary-structure-forming NA motifs like G-quadruplexes, into the spotlight of computational drug discovery.[75] The importance of nucleic acid targets stems from the fact that they are involved in numerous diseases, ranging from cancer to neurodegeneration, from bacterial to viral infections. Targeting RNA in general, and specific tertiary motifs in particular, is exceptionally challenging due to their high conformational flexibility and the limited chemical diversity of the receptors against which to select binders. An additional challenge for computational approaches is represented by the necessity of optimizing force fields and simulation methodologies for these classes of molecules. To alleviate these problems, approaches ranging from reparameterization to the correction of force fields with experimentally derived data have been reported.[76–78] An excellent extensive review of how RNA has been studied and small molecules have been used to address its biology in disease was recently published by Falese *et al.*,[79] and we refer the interested reader to that paper for an in-depth analysis. Here, we will simply report a few examples of how simulations, ranging from docking to enhanced sampling, are supporting the improvement of our understanding related to the determinants of nucleic acid–ligand recognition.

In a notable example in which the target's conformational diversity was accounted for in the selection of viable ligands, Stelzer *et al.*[80] docked small molecules on an ensemble of RNA structures constructed combining NMR data and MD simulations. Specifically, the authors used experimental data from multiple sets of NMR residual dipolar coupling (RDC)

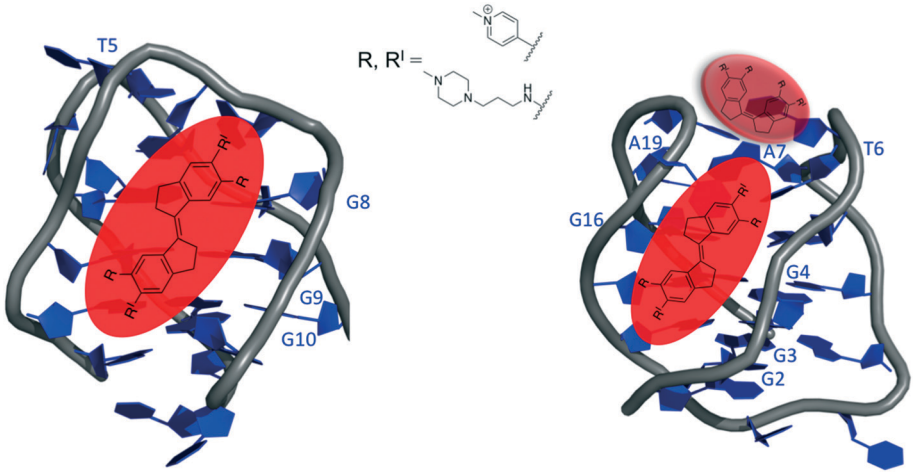

Fig. 6 Schematic representation of different binding modes of G4 ligands (PDB: 143D). Larger ovals represent binders with a *trans*-benzylpentane scaffold bound in groove regions; the smaller oval represents those with a *cis*-benzylpentane scaffold bound on the top of tetrads. Labelled bases are those deemed to be the most important for ligand binding.

data as filters to select conformers from a larger ensemble generated using unconstrained MD. The ensemble of docking targets is defined by the structures that satisfy all time averaged RDC data (Fig. 6). This data-driven selection is critical to focus the ensembles of RNA to a number of representative conformations amenable to subsequent docking studies. Indeed, the authors quantitatively predict binding energies for small molecules that bind different RNA conformations and discover new compounds that bind the target with interesting affinity and good activities.

The large ligand chemical space and target conformational space to explore are the two main challenges also in the case of ligand design for G-quadruplex motifs (G4s). G4s are structures formed in guanine-rich sequences, whose backbone is composed of stacked square-planar arrangements of four guanine bases, known as G-tetrads, stabilized by Hoogsteen hydrogen bonding.[81] G4s are found in prominent genomic regions involved in the regulation of biological processes, among which telomeres and oncogene promoters. Their location, biological role, and characteristic structure made them groundbreaking therapeutic targets for the development of anticancer drugs.[82–84]

Docking and molecular dynamics (MD) simulations have been applied to screen for ligands and investigate G4 affinities.[85–87]

Advanced methods based on enhanced sampling and absolute free energy calculations have recently been used to shed light on the mechanisms of ligand binding and associated G4 conformational responses. These phenomena typically occur on long-time scales, so other simulation methods can be used to overcome the limitations of classical MD, like metadynamics[88] or potential of mean force (PMF) calculations.[89]

In this framework, work by O'Hagan *et al.*[90] studied the G4 binding mechanisms of different rigid ligands. Using metadynamics simulations, they revealed a diversity of binding mechanisms a poses for various ligands. Interestingly, the small molecules are shown to have different effects on the conformational properties of the G4 receptor, unveiling a clear small-molecule-G4 cross-talk. Computational results were next corroborated by NMR and CD spectroscopy studies, which proved how predicted binding modes and affinities could be validated experimentally. Information on how a ligand can influence the structure of the G4 target is fundamental, both in the design of molecules with improved selectivity and activity profiles and in the interpretation of experimental data, improving our understanding of these fascinating NA motifs.

Nayis *et al.*[91] used computation to study the mechanism of G4-binding of Au–carbene compounds. These are known to stabilize G4s and interfere with telomere elongation, a phenomenon associated with uncontrolled cell replication and growth at the basis of tumor development. Here, the authors used MD simulations and absolute free energy calculation methods developed by Roux and coworkers,[89] to reveal different binding modes and mobilities of Au–carbene when targeting different G4 surfaces. Important for drug design and optimization is that the detailed analysis of energetic contributions revealed nonpolar and van der Waals interactions as the key factors driving binding. These results can clearly be useful to guide modifications to improve Au–carbene affinity and specificity for G-quadruplex binding.

Conclusion and perspectives

Molecular simulations have reached a high level of sophistication and accessibility. On the one hand, the development of more and more refined, accurate and reliable methods for sampling, free energy evaluation, and kinetic analysis of binding, makes computational chemistry a key part of the discovery of new candidate drugs. On the other hand, the impressive advance in technology, both from the

hardware and from the software points of view, has made computational experiments accessible on a routine basis for many applications. A striking example of this is represented by the recent simulation of the whole proteome of the SARS-CoV-2 virus[92] *via* distributed computing. Such simulations provide an unprecedentedly rich source of information for the design of new ligands revealing over 50 'cryptic' pockets, dramatically expanding the chemical space available for viral inhibition. Clearly, the same type or approach can be translated into the simulation of cancer related proteins. The consequence of these impressive advances is that a number of approaches which were substantially restricted to the academic realm have now become common practice even in the industrial setting.

Computational drug discovery is now advancing into the era of artificial intelligence and big data. Indeed, a continuously growing number of papers is now being published in which computational learning techniques are applied to disparate problems, from force-field corrections, to the evaluation of interaction energetics, to the prediction of the biological activities of drug candidates.[93]

It is clear that artificial intelligence will significantly impact drug development integrating the (massive) amounts of data available on the pharmacological properties of known drugs (ranging from activity to safety and metabolism) originating from disparate sources with the more classical physics-based modeling of structural, dynamic, and interaction properties. In this context, artificial intelligence approaches such as deep learning can help reveal non-obvious patterns correlating observed *in vivo* activities and the chemical determinants of binding or ligand-dependent regulation of functionally-oriented motions. This interconnection may also help anticipate the efficacy and safety of drug candidates. Finally, we envisage that the inclusion of data on genetic information, mutations, their familial frequency *etc.* may result in models that make it possible to adapt drug design to the demands of specific (patient dependent) protein variants, in a real personalized medicine perspective.

In summary, building on strong chemical and physical bases, new advances of artificial intelligence and big data can pave the way for better and more efficient drug development and optimization, with a strong impact on future drug discovery.

Conflicts of interest

There is no conflict of interest to declare.

Acknowledgements

Giorgio Colombo gratefully acknowledges funding from Fondazione AIRC under IG 20019 project.

References

1 L. C. Zanetti-Domingues, D. Korovesis, S. R. Needham, C. J. Tynan, S. Sagawa, S. K. Roberts, A. Kuzmanic, E. Ortiz-Zapater, P. Jain, R. C. Roovers, A. Lajevardipour, P. M. P. van Bergen en Henegouwen, G. Santis, A. H. A. Clayton, D. T. Clarke, F. L. Gervasio, Y. Shan, D. E. Shaw, D. J. Rolfe, P. J. Parker and M. L. Martin-Fernandez, *Nat. Commun.*, 2018, **9**, 4325.

2 I. Galdadas, S. Qu, A. S. F. Oliveira, E. Olehnovics, A. R. Mack, M. F. Mojica, P. K. Agarwal, C. L. Tooke, F. L. Gervasio, J. Spencer, R. A. Bonomo, A. J. Mulholland and S. Haider, *eLife*, 2021, **10**, e66567.

3 I. Galdadas, F. L. Gervasio and Z. Cournia, *Chem. Sci.*, 2020, **11**, 3511–3515.

4 D. A. Fruman, H. Chiu, B. D. Hopkins, S. Bagrodia, L. C. Cantley and R. T. Abraham, *Cell*, 2017, **170**, 605–635.

5 J. D. Carson, G. Van Aller, R. Lehr, R. H. Sinnamon, R. B. Kirkpatrick, K. R. Auger, D. Dhanak, R. A. Copeland, R. R. Gontarek, P. J. Tummino and L. Luo, *Biochem. J.*, 2007, **409**, 519–524.

6 E. Moroni, D. A. Agard and G. Colombo, *J. Chem. Theory Comput.*, 2018, **14**, 1033–1044.

7 S. A. Serapian, C. Sanchez-Martín, E. Moroni, A. Rasola and G. Colombo, *Trends Pharmacol. Sci.*, 2021, **42**(7), 566–576.

8 E. P. Barros, Ö. Demir, J. Soto, M. J. Cocco and R. E. Amaro, *Chem. Sci.*, 2021, **12**, 1891–1900.

9 A. C. Joerger and A. R. Fersht, *Annu. Rev. Biochem.*, 2016, **85**, 375–404.

10 C. Herbert, U. Schieborr, K. Saxena, J. Juraszek, F. De Smet, C. Alcouffe, M. Bianciotto, G. Saladino, D. Sibrac, D. Kudlinzki, S. Sreeramulu, A. Brown, P. Rigon, J.-P. Herault, G. Lassalle, T. L. Blundell, F. Rousseau, A. Gils, J. Schymkowitz, P. Tompa, J.-M. Herbert, P. Carmeliet, F. L. Gervasio, H. Schwalbe and F. Bono, *Cancer Cell*, 2013, **23**, 489–501.

11 I. D'Annessa, S. Sattin, J. H. Tao, M. Pennati, C. Sanchez-Martin, E. Moroni, A. Rasola, N. Zaffaroni, D. A. Agard, A. Bernardi and G. Colombo, *Chem. – Eur. J.*, 2017, **23**, 5188–5192.

12 S. Rinaldi, V. A. Assimon, Z. T. Young, G. Morra, H. Shao, I. R. Taylor, J. E. Gestwicki and G. Colombo, *ACS Chem. Biol.*, 2018, **13**, 3142–3152.

13 I. D'Annessa, S. Raniolo, V. Limongelli, D. Di Marino and G. Colombo, *J. Chem. Theory Comput.*, 2019, **15**, 6368–6381.

14 C. Sanchez-Martin, E. Moroni, M. Ferraro, C. Laquatra, G. Cannino, I. Masgras, A. Negro, P. Quadrelli, A. Rasola and G. Colombo, *Cell Rep.*, 2020, **31**, 107531.

15 I. R. Taylor, V. A. Assimon, S. Y. Kuo, S. Rinaldi, X. Li, Z. T. Young, G. Morra, K. Green, D. Nguyen, H. Shao, S. Garneau-Tsodikova, G. Colombo and J. E. Gestwicki, *Chem. Sci.*, 2020, **11**, 1892–1904.

16 M. Ferraro, E. Moroni, E. Ippoliti, S. Rinaldi, C. Sanchez-Martin, A. Rasola, L. F. Pavarino and G. Colombo, *J. Phys. Chem. B*, 2021, **125**, 101–114.

17 A. Spinello, S. Martini, F. Berti, M. Pennati, M. Pavlin, J. Sgrignani, G. Grazioso, G. Colombo, N. Zaffaroni and A. Magistrato, *Eur. J. Med. Chem.*, 2019, **168**, 253–262.

18 S. Laporte and A. Magistrato, *J. Phys. Chem. Lett.*, 2021, **12**(22), 5384–5389.

19 C. R. Knoverek, G. K. Amarasinghe and G. R. Bowman, *Trends Biochem. Sci.*, 2019, **44**, 351–364.
20 M. I. Zimmerman and G. R. Bowman, *J. Chem. Theory Comput.*, 2015, **11**, 5747–5757.
21 J. R. Porter, K. E. Moeder, C. A. Sibbald, M. I. Zimmerman, K. M. Hart, M. J. Greenberg and G. R. Bowman, *Biophys. J.*, 2019, **116**, 818–830.
22 S. Singh and G. R. Bowman, *J. Chem. Theory Comput.*, 2017, **13**, 1509–1517.
23 G. R. Bowman, E. R. Bolin, K. M. Hart, B. C. Maguire and S. Marqusee, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 2734–2739.
24 P. Ghanakota and H. A. Carlson, *J. Med. Chem.*, 2016, **59**, 10383–10399.
25 J. Seco, F. J. Luque and X. Barril, *J. Med. Chem.*, 2009, **52**, 2363–2371.
26 O. Guvench and A. D. MacKerell, Jr., *PLoS Comput. Biol.*, 2009, **5**, e1000435.
27 H. Cheng, B. M. Linhares, W. Yu, M. G. Cardenas, Y. Ai, W. Jiang, A. Winkler, S. Cohen, A. Melnick, A. MacKerell, T. Cierpicki and F. Xue, *J. Med. Chem.*, 2018, **61**, 7573–7588.
28 A. Bakan, N. Nevins, A. S. Lakdawala and I. Bahar, *J. Chem. Theory Comput.*, 2012, **8**, 2435–2447.
29 V. Oleinikovas, G. Saladino, B. P. Cossins and F. L. Gervasio, *J. Am. Chem. Soc.*, 2016, **138**, 14257–14263.
30 F. Comitani and F. L. Gervasio, *J. Chem. Theory Comput.*, 2018, **14**, 3321–3331.
31 N. A. Khazanov and H. A. Carlson, *PLoS Comput. Biol.*, 2013, **9**, e1003321.
32 R. D. Smith, J. J. Clark, A. Ahmed, Z. J. Orban, J. B. Dunbar and H. A. Carlson, *J. Mol. Biol.*, 2019, **431**, 2423–2433.
33 A. Pohorille, C. Jarzynski and C. Chipot, *J. Phys. Chem. B*, 2010, **114**, 10235–10253.
34 W. L. Jorgensen, *Science*, 2004, **303**, 1813.
35 W. L. Jorgensen, *Angew. Chem., Int. Ed.*, 2012, **51**, 11680–11684.
36 L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M. K. Dahlgren, J. Greenwood, D. L. Romero, C. Masse, J. L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D. L. Mobley, W. L. Jorgensen, B. J. Berne, R. A. Friesner and R. Abel, *J. Am. Chem. Soc.*, 2015, **137**, 2695–2703.
37 C. D. Christ, A. E. Mark and W. F. van Gunsteren, *J. Comput. Chem.*, 2010, **31**, 1569–1582.
38 M. Lapelosa, E. Gallicchio and R. M. Levy, *J. Chem. Theory Comput.*, 2012, **8**, 47–60.
39 J. Michel and J. W. Essex, *J. Comput.-Aided Mol. Des.*, 2010, **24**, 639–658.
40 F. L. Gervasio, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2005, **127**, 2600–2607.
41 V. Limongelli, M. Bonomi and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 6358–6363.
42 A. Cavalli, A. Spitaleri, G. Saladino and F. L. Gervasio, *Acc. Chem. Res.*, 2015, **48**, 277–285.
43 H.-J. Woo and B. Roux, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6825–6830.
44 R. Abel, L. Wang, D. L. Mobley and D. A. Friesner, *Curr. Top. Med. Chem.*, 2017, **17**, 2577–2585.
45 J. Scheen, W. Wu, A. S. J. S. Mey, P. Tosco, M. Mackey and J. Michel, *J. Chem. Inf. Model.*, 2020, **60**, 5331–5339.
46 D. A. Rufa, H. E. Bruce Macdonald, J. Fass, M. Wieder, P. B. Grinaway, A. E. Roitberg, O. Isayev and J. D. Chodera, bioRxiv, 2020, DOI: 10.1101/2020.07.29.227959.
47 E. Perola and P. S. Charifson, *J. Med. Chem.*, 2004, **47**, 2499–2510.
48 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.
49 C. Jarzynski, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **65**, 046122.
50 P. Wirnsberger, A. J. Ballard, G. Papamakarios, S. Abercrombie, S. Racanière, A. Pritzel, D. Jimenez Rezende and C. Blundell, *J. Chem. Phys.*, 2020, **153**, 144112.
51 R. Evans, L. Hovan, G. A. Tribello, B. P. Cossins, C. Estarellas and F. L. Gervasio, *J. Chem. Theory Comput.*, 2020, **16**, 4641–4654.
52 M. M. Sultan and V. S. Pande, *J. Chem. Phys.*, 2018, **149**, 094106.
53 F. Marchetti, E. Moroni, A. Pandini and G. Colombo, *J. Phys. Chem. Lett.*, 2021, **12**, 3724–3732.
54 R. A. Copeland, D. L. Pompliano and T. D. Meek, *Nat. Rev. Drug Discovery*, 2006, **5**, 730–739.
55 H. Lu and P. J. Tonge, *Curr. Opin. Chem. Biol.*, 2010, **14**, 467–474.
56 D. A. Schuetz, W. E. A. de Witte, Y. C. Wong, B. Knasmueller, L. Richter, D. B. Kokh, S. K. Sadiq, R. Bosma, I. Nederpelt, L. H. Heitman, E. Segala, M. Amaral, D. Guo, D. Andres, V. Georgi, L. A. Stoddart, S. Hill, R. M. Cooke, C. De Graaf, R. Leurs, M. Frech, R. C. Wade, E. C. M. de Lange, A. P. Ijzerman, A. Müller-Fahrnow and G. F. Ecker, *Drug Discovery Today*, 2017, **22**, 896–911.
57 D. A. Schuetz, M. Bernetti, M. Bertazzo, D. Musil, H.-M. Eggenweiler, M. Recanatini, M. Masetti, G. F. Ecker and A. Cavalli, *J. Chem. Inf. Model.*, 2019, **59**, 535–549.
58 D. B. Kokh, M. Amaral, J. Bomke, U. Grädler, D. Musil, H.-P. Buchstaller, M. K. Dreyer, M. Frech, M. Lowinski, F. Vallee, M. Bianciotto, A. Rak and R. C. Wade, *J. Chem. Theory Comput.*, 2018, **14**, 3859–3869.
59 P. Tiwary, V. Limongelli, M. Salvalaglio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E386.
60 S. Decherchi and A. Cavalli, *Chem. Rev.*, 2020, **120**, 12788–12833.
61 M. Bernetti, M. Masetti, W. Rocchia and A. Cavalli, *Annu. Rev. Phys. Chem.*, 2019, **70**, 143–171.
62 Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger and D. E. Shaw, *J. Am. Chem. Soc.*, 2011, **133**, 9181–9183.
63 S. De Cherchi, A. Berteotti, G. Bottegoni, W. Rocchia and A. Cavalli, *Nat. Commun.*, 2015, **6**, 6155.
64 D. Callegari, A. Lodola, D. Pala, S. Rivara, M. Mor, A. Rizzi and A. M. Capelli, *J. Chem. Inf. Model.*, 2017, **57**, 159–169.
65 L. Mollica, I. Theret, M. Antoine, F. Perron-Sierra, Y. Charton, J.-M. Fourquez, M. Wierzbicki, J. A. Boutin, G. Ferry, S. Decherchi, G. Bottegoni, P. Ducrot and A. Cavalli, *J. Med. Chem.*, 2016, **59**, 7167–7176.

66 N. Plattner and F. Noé, *Nat. Commun.*, 2015, **6**, 7653–7653.
67 I. Buch, T. Giorgino and G. De Fabritiis, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10184.
68 S. Doerr and G. De Fabritiis, *J. Chem. Theory Comput.*, 2014, **10**, 2064–2069.
69 S. D. Lotz and A. Dickson, *J. Am. Chem. Soc.*, 2018, **140**, 618–628.
70 K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross and G. R. Bowman, *Nat. Commun.*, 2016, **7**, 12965.
71 B. R. Jagger, C. T. Lee and R. E. Amaro, *J. Phys. Chem. Lett.*, 2018, **9**, 4941–4948.
72 P. Majek and R. Elber, *J. Chem. Theory Comput.*, 2010, **6**, 1805–1817.
73 L. W. Votapka and R. E. Amaro, *PLoS Comput. Biol.*, 2015, **11**, e1004381.
74 L. W. Votapka, B. R. Jagger, A. L. Heyneman and R. E. Amaro, *J. Phys. Chem. B*, 2017, **121**, 3597–3606.
75 I. Ortiz de Luzuriaga, X. Lopez and A. Gil, *Annu. Rev. Biophys.*, 2021, **50**, 209–243.
76 J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter and M. Otyepka, *Chem. Rev.*, 2018, **118**, 4177–4338.
77 S. Bottaro, G. Bussi, S. D. Kennedy, D. H. Turner and K. Lindorff-Larsen, *Sci. Adv.*, 2018, **4**, eaar8521.
78 J. Šponer, G. Bussi, P. Stadlbauer, P. Kührová, P. Banáš, B. Islam, S. Haider, S. Neidle and M. Otyepka, *Biochim. Biophys. Acta, Gen. Subj.*, 2017, **1861**, 1246–1263.
79 J. P. Falese, A. Donlic and A. E. Hargrove, *Chem. Soc. Rev.*, 2021, **50**, 2224–2243.
80 A. C. Stelzer, A. T. Frank, J. D. Kratz, M. D. Swanson, M. J. Gonzalez-Hernandez, J. Lee, I. Andricioaei, D. M. Markovitz and H. M. Al-Hashimi, *Nat. Chem. Biol.*, 2011, **7**, 553–559.
81 S. Neidle, *J. Med. Chem.*, 2016, **59**, 5987–6011.
82 S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discovery*, 2011, **10**, 261–275.
83 J. Spiegel, S. Adhikari and S. Balasubramanian, *Trends Chem.*, 2020, **2**, 123–136.
84 P. Kharel, G. Becker, V. Tsvetkov and P. Ivanov, *Nucleic Acids Res.*, 2020, **48**, 12534–12555.
85 D.-L. Ma, T.-S. Lai, F.-Y. Chan, W.-H. Chung, R. Abagyan, Y.-C. Leung and K.-Y. Wong, *ChemMedChem*, 2008, **3**, 881–884.
86 J. Luo, W. Wei, J. Waldispühl and N. Moitessier, *Eur. J. Med. Chem.*, 2019, **168**, 414–425.
87 V. Pirota, C. Platella, D. Musumeci, A. Benassi, J. Amato, B. Pagano, G. Colombo, M. Freccero, F. Doria and D. Montesarchio, *Int. J. Biol. Macromol.*, 2021, **166**, 1320–1334.
88 A. Laio and F. L. Gervasio, *Rep. Prog. Phys.*, 2008, **71**, 126601.
89 H.-J. Woo and B. Roux, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6825.
90 M. P. O'Hagan, S. Haldar, J. C. Morales, A. J. Mulholland and M. C. Galan, *Chem. Sci.*, 2021, **12**, 1415–1426.
91 A. Nayis, K. Liebl, C. V. Frost and M. Zacharias, *Biophys. J.*, 2021, **120**, 101–108.
92 M. I. Zimmerman, J. R. Porter, M. D. Ward, S. Singh, N. Vithani, A. Meller, U. L. Mallimadugula, C. E. Kuhn, J. H. Borowsky, R. P. Wiewiora, M. F. D. Hurley, A. M. Harbison, C. A. Fogarty, J. E. Coffland, E. Fadda, V. A. Voelz, J. D. Chodera and G. R. Bowman, *Nat. Chem.*, 2021, **13**, 651–659.
93 H. Zhu, *Annu. Rev. Pharmacol. Toxicol.*, 2020, **60**, 573–589.

