MedChemComm

REVIEW

View Article Online
View Journal | View Issue

Cite this: *Med. Chem. Commun.*, 2017, **8**, 534

Protein–ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling†

M. Bernetti,[ab] A. Cavalli[ab] and L. Mollica*[b]

Received 18th October 2016,
Accepted 25th January 2017

DOI: 10.1039/c6md00581k

rsc.li/medchemcomm

In the last three decades, protein and nucleic acid structure determination and comprehension of the mechanisms, leading to their physiological and pathological functions, have become a cornerstone of biomedical sciences. A deep understanding of the principles governing the fates of cells and tissue at the molecular level has been gained over the years, offering a solid basis for the rational design of drugs aimed at the pharmacological treatment of numerous diseases. Historically, affinity indicators (*i.e.* K_d and IC_{50}/EC_{50}) have been assumed to be valid indicators of the *in vivo* efficacy of a drug. However, recent studies pointed out that the kinetics of the drug–receptor binding process could be as important or even more important than affinity in determining the drug efficacy. This eventually led to a growing interest in the characterisation and prediction of the rate constants of protein–ligand association and dissociation. For instance, a drug with a longer residence time can kinetically select a given receptor over another, even if the affinity for both receptors is comparable, thus increasing its therapeutic index. Therefore, understanding the molecular features underlying binding and unbinding processes is of central interest towards the rational control of drug binding kinetics. In this review, we report the theoretical framework behind protein–ligand association and highlight the latest advances in the experimental and computational approaches exploited to investigate the binding kinetics.

Introduction

Among biological macromolecules, proteins represent a class of primary interest since they are involved in a plethora of cellular processes, ranging from structural and mechanical roles to signalling and regulation functions. To exert their biological functions, proteins typically need to directly interact with other molecules including proteins/peptides, nucleic acids, membranes, substrates, and small molecule ligands with high specificity and affinity.

A deeper understanding of protein functions at the atomic level stems from two prerequisites: the knowledge of their three-dimensional features, *i.e.* details of their tertiary and quaternary structures, and the comprehension of the mechanisms responsible for their interactions with ligands. In the last 20 years, protein structure determination has become a gold standard in the world of biophysical chemistry and a crucial step for the investigation of biochemical processes, assisted by an increasing availability of large scale facilities for protein production and data collection (*i.e.* X-ray diffraction and NMR spectroscopy). For these reasons, a large community has recently exploited the growing datasets of structural information to gain insight on the full description, characterisation, and quantification of the energetics that govern the formation of protein–protein and protein–ligand complexes.[1] In addition, the need for structural data and protein–ligand binding mechanisms for the rational design and development of new drugs demands a deep understanding of the nature of the molecular recognition.[2]

Since the primordial stages of pharmacology, it has become clear that a drug works only when bound to its target receptor.[3] Direct measurement of the extent to which a drug is bound to its receptor at equilibrium (*i.e.* the binding affinity) was not possible until long after the theory was first postulated, and identifiable and ultimately purifiable molecular receptors became routinely available, thus enabling direct measurements of binding affinity and guiding most early-stage discovery efforts. Historically, the dissociation constant K_d and its proxies IC_{50}/EC_{50} (*i.e.* the drug concentrations leading to the half-maximal inhibition of a biological activity) have been assumed for decades to be a valid surrogate for *in vivo* efficacy. However, recent studies have shown that the kinetics of drug–receptor binding could be just as important or even more important than affinity in determining drug

[a] *Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126 Bologna, Italy*
[b] *CompuNet, Istituto Italiano di Tecnologia, via Morego 30, 16163 Genova, Italy. E-mail: luca.mollica@iit.it*
† The authors declare no competing interests.

efficacy,[4–6] particularly when the duration of the pharmacological effect is a significant component of *in vivo* efficacy. The time a ligand spends at its receptor binding site, referred to as the drug–target residence time (t_r) and expressed as the inverse of the dissociation rate constant ($t_r = 1/k_{off}$), is often a better predictor of efficacy than is binding affinity.[4,6,7] A drug with a longer residence time on a given receptor can be kinetically selective for that receptor over another one, even if the affinity for both receptors is comparable. Conversely, the extended, non-physiological drug occupancy of the target receptor may cause toxicity;[8–10] thus, drugs with faster dissociation rates may be preferred in terms of drug safety, and may also lead to an increase in the therapeutic index (the ratio between a drug's toxic dose and its efficacious dose). From the association standpoint, a faster-binding drug, characterised by a high on-rate, may be desirable when targeting a short-lived receptor.[11]

While the concepts underlying the rational optimisation of binding affinity are relatively well understood, the same cannot be said for binding kinetics. This can be ascribed to a still poor comprehension of the molecular determinants of binding kinetics, compared to those of binding affinity, which are linked to the fundamental difficulty in characterising transient states. Binding affinity is calculated from the free energy difference between the bound and unbound states, which are both stable and typically easily observable. Association and dissociation rates rely on the height of the free energy barrier separating those states. However, the atomic arrangement of the drug and the receptor at this point of highest free energy (namely the transition state, TS) has only ephemeral existence and is thus difficult to identify. Understanding the molecular interactions taking place between drug and receptor at this difficult-to-observe transition state is thus of pivotal interest for moving towards a rational control of drug binding kinetics in drug discovery.[5]

The rational optimisation of protein–ligand interactions paves the way to the control of drug physiochemical behaviour with relevant therapeutic implications for drug efficacy and drug safety. In the present review, we discuss both the experimental and the theoretical techniques that are currently used for characterising protein–ligand and receptor–drug complexes (un)binding kinetics. In particular, we focus on the most recent studies that explicitly, functionally and/or structurally put a focus on the residence time of ligands in their protein binding sites as a crucial molecular determinant of their mode of action.

The theory behind protein–ligand association and dissociation processes

Thermodynamics and kinetics of protein–ligand complexes

For binding to occur, the initial contacts/collisions between a protein molecule and a ligand have to form an encounter complex, for which molecular diffusion plays a decisive role.[12] Molecular diffusion, which originates from molecular kinetic energy (or heat, thermal energy), is the entropy-driven process that guides the formation of this complex. In a protein–ligand–solvent system, the diffusion (or the random Brownian motions) of solute molecules has two origins: (i) the kinetic energy of the solute molecules themselves; and (ii) collisions of the large solute with the small water molecules, which move with different velocities in different random directions. At constant temperature and pressure, the motions of individual water molecules resulting from their atomic kinetic energy could lead to the maximisation of the solvent entropy. It seems likely that the heavy Brownian bombardment from a large amount of water molecules may play a role in facilitating the rotations, translations, and wanderings of the solute molecules and ultimately, the accidental collisions among them.[13] The long-range electrostatic attraction can promote association in the case of two solute molecules with opposite charges, thus allowing them to overcome the diffusion limit.[14]

The collision theory was the first attempt to provide an analytical description of the dependence of the rate constant of a reaction on the temperature and the activation energy. The relation is expressed by means of the Arrhenius equation:[15]

$$k = A\mathrm{e}^{-\frac{E_a}{RT}} \quad (1)$$

where E_a is the activation energy of the process and T is the temperature. The pre-exponential factor A, also called the frequency factor, is a constant that can be determined experimentally or numerically, and it describes the number of times two molecules collide. Notably, not every collision results in the expected product, since a proper orientation of the two colliding species is also required. Within the collision theory, the pre-exponential factor can be defined as follows:

$$A = {d_{AB}}^2 \sqrt{\frac{8k_B T}{\mu}} \quad (2)$$

where d_{AB} is the collision radius between two particles A and B, k_B the Boltzmann factor, T is the temperature and μ is the reduced mass of the system. The Arrhenius rate law has been widely used to determine the energies for the reaction barrier, ignoring any mechanistic considerations, such as whether one or more reactive intermediates are involved in the conversion of a reactant to a product. However, the collision theory deals with gases and does not account for structural complexities in atoms and molecules.

In order to resolve this discrepancy, the transition state theory was developed in the 30s by Henry Eyring, Meredith Gwynne Evans and Michael Polanyi. This theory moved from the foundations of thermodynamics to give a representation of the most accurate pre-exponential factor that yields the corresponding rate, assuming a special type of chemical equilibrium (quasi-equilibrium) between reactants and activated

transition state complexes.[16,17] There are three basic concepts behind the transition state theory: (i) rates of reactions can be studied by examining activated complexes (the so-called transition states) that lie near the saddle point of a potential energy surface; (ii) the activated complexes are in a special equilibrium (quasi-equilibrium) with the reactants; (iii) the activated complexes can convert into products, and kinetic theory can be used to calculate the rate of this conversion. Even within this theoretical framework, absolute reaction rate constants are extremely difficult to calculate. An accurate description of the potential energy profile of the system is required, and this is particularly challenging for the transition states, which are typically high-energy, ephemerally existing states. Nevertheless, transition state theory is extremely helpful in offering a qualitative description of how chemical reactions take place and it has been successfully applied to determine standard enthalpy, entropy and Gibbs energy of activation, once the rate constant has been experimentally determined (Fig. 1).[17] For the reaction between a ligand (L) binding to a protein (P) to form the bimolecular complex (PL), we can consider the simple equilibrium:

$$\mathrm{P} + \mathrm{L} \rightleftharpoons \mathrm{PL} \tag{3}$$

This two-state mechanism comprises a single elementary step—the binding (or unbinding) of the drug—without any intermediate states; it suffices to illustrate several key points, although drug binding may often involve, as already stressed, one or more intermediates.[7] The binding affinity is described by the dissociation constant K_d, which is the ratio of the product and reactants concentrations at equilibrium:

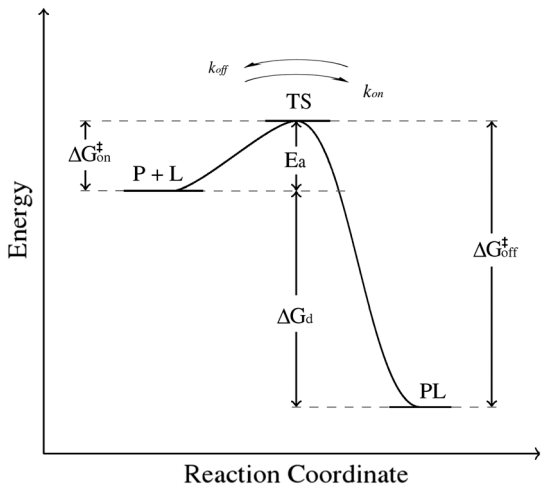

Fig. 1 A scheme of the energetic landscape of the complex (PL) formation between a protein (P) and a ligand (L). TS represents the transition state, E_a is the activation energy of the process, ΔG_d is the difference between the free energies of the reactants and of the product, $\Delta G^{\ddagger}_{on}$ the free energy difference between the reactants and TS, and $\Delta G^{\ddagger}_{off}$ the free energy difference between the product and the TS.

$$K_d = \frac{[\mathrm{P}][\mathrm{L}]}{[\mathrm{PL}]} \tag{4}$$

At equilibrium, K_d corresponds to the drug concentration at which half the receptor binding sites are occupied, and it is directly related to the free energy difference between the bound and unbound states, ΔG_d.

In contrast to K_d, which is determined solely by stable molecular interactions between the drug, receptor and solvent, the rate constants k_{on} and k_{off} depend on transient interactions along the binding pathway. Specifically, they are related to the highest free energy barrier—the transition state—that separates the bound and unbound states:

$$k_{on} \propto e^{-\frac{\Delta G^{\ddagger}_{on}}{RT}} \tag{5a}$$

$$k_{off} \propto e^{-\frac{\Delta G^{\ddagger}_{off}}{RT}} \tag{5b}$$

with $\Delta G^{\ddagger}_{on}$ being the free energy difference between reactants and the TS, and $\Delta G^{\ddagger}_{off}$ between the complex and the TS (Fig. 1). Thermodynamics and kinetics of binding are linked *via* the kinetics of half reactions (Fig. 1) as follows:

$$\mathrm{P} + \mathrm{L} \xrightarrow{k_{on}} \mathrm{PL} \tag{6a}$$

$$\mathrm{PL} \xrightarrow{k_{off}} \mathrm{P} + \mathrm{L} \tag{6b}$$

$$K_d = \frac{[\mathrm{P}][\mathrm{L}]}{[\mathrm{PL}]} = \frac{k_{off}}{k_{on}} \tag{6c}$$

This relation is particularly interesting as it highlights how, at least in principle, the free energies of the unbound and bound states and of the transition state can be varied independently of one another. Destabilising only the transition state decreases both rates without altering affinity; conversely, stabilising the transition state increases rates. Destabilising the bound state weakens affinity and increases the off-rate without altering the on-rate, whereas altering the energy of the unbound state affects the on-rate and the affinity only. In practice however, achieving any of these 'corner cases'—the 'pure' alteration of just two of the constants K_d, k_{on} and k_{off} without affecting the third—is difficult. Notably, changes in k_{off} can result in no measureable effect on K_d if there are compensatory changes in k_{on}. In such a case, ligand

optimisation based on affinity alone would hide the molecular determinants of binding kinetics behind an unchanging K_d. As such, QSAR models that explicitly resolve k_{on} and k_{off} can be more reliable than those based on binding affinity alone.[5,18,19] From this perspective, much attention has been recently focused on the impact of including kinetics information in the optimisation steps of a drug discovery campaign. The residence time, $t_r = 1/k_{off}$, and complex half-life, $t_{1/2} = \ln(2/k_{off})$, are useful common measures of binding kinetics. In a simple two-state mechanism, for instance, a typical nanomolar affinity drug that binds at the maximal, diffusion-limited on-rate ($k_{on} = 10^9$ M^{-1} s^{-1})[4] would have a residence time of ~1 s. It is worth noting that other than the on-rate, t_r and $t_{1/2}$ do not depend on drug concentration. Thus, even when the free concentration of a long-residence time drug falls below the K_d, a significant population of receptors will still be occupied by the drug and the pharmacological effect will be sustained.

We want to stress at this point that on- and off-rates represent elementary mechanistic steps only for a two-state process. In the more common, non-two-state mechanisms, the apparent rate constants consist of the multiple elementary rate constants describing the transitions between unbound, intermediate and bound states. Such a possibly composite nature—indeed, the mere existence of a non-two-state mechanism—must be kept in mind, especially when observing counterintuitive binding kinetics.[20]

Protein–ligand binding models

In the early stages of pharmaceutical research, the concept of shape complementarity in biomolecular recognition was first introduced by Fischer as early as 1894. The prerequisites of his "lock-and-key" (Fig. 2a) model are that both the protein and the ligand are rigid and that their binding interfaces perfectly match. As a result, only the correctly sized ligand (the key) can insert into the binding pocket (key hole) of the protein (the lock). Binding within the lock-and-key model is dominated by changes in entropy due to desolvation of the two solutes (most commonly a protein and a ligand in a drug discovery perspective) during the process of complex formation.[21,22] However, the experimental evidence that a protein is also able to bind to a ligand, even in the case where their initial shapes do not match well, cannot be explained. Increasing application and advances of experimental techniques, such as X-ray crystallography, NMR spectroscopy and fluorescence, to the study of biomolecular processes shed light on the extremely dynamic nature of macromolecules. In computational chemistry, mostly because of limited computational resources, early structure based drug discovery studies also relied on rigid conformations of the molecular targets of interest. However, over the years it has been clearly recognized that completely neglecting multiple conformations may lead to wrong evaluations. All these observations together contributed to raising awareness on protein conformational plasticity, as a matter of fact, and the concepts of "induced-fit" and "conformational selection" were established to include the more complex aspects involved in biomolecular recognition.[23]

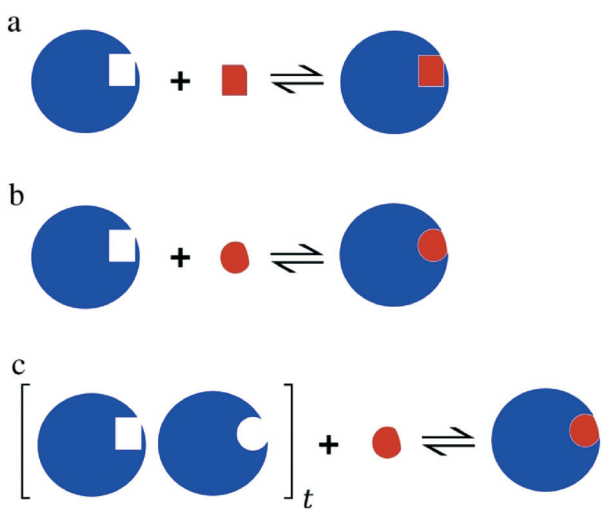

Fig. 2 The three different binding models of proteins and ligands, *i.e.* the lock and key (a), the induced fit (b), and the conformational selection (c). The protein is represented in blue and the ligand is represented in red.

In the induced fit model (Fig. 2b), the binding site in the protein is considered flexible and the interacting ligand induces a conformational change at the binding site. Because the induced fit mechanism takes into account only the conformational flexibility of the ligand-binding site, this model seems to be suitable for proteins showing merely minor conformational changes after the ligand binding. In addition, both the lock-and-key and the induced fit models treat the protein as a single, stable conformation under given experimental conditions. Typical induced fit binding has been demonstrated in the designed host–guest systems[24,25] through combinatorial chemistry, NMR spectroscopy and computational chemistry techniques. These studies, in conjunction with the kinetic model calculations, demonstrate that binding by induced fit requires a pre-existing complementarity between the interacting species.[26] The lack of perfect surface complementary between binding partners necessitates multiple tentative collisions: once initially favourable contacts (*i.e.* characterised by a negative enthalpy change) between the matched sites occur, they are supposed to be strong enough to provide the encounter complex with enough longevity so that induced fit takes place within a reasonable time.[26]

As most proteins are inherently dynamic, the conformational selection model (Fig. 2c) takes into account this inherent flexibility and relies on the free energy landscape theory of protein structure and dynamics.[27–30] It postulates that the native state of a protein does not exist as a single, rigid conformation but rather as a vast ensemble of conformational states/substates[31,32] that coexist at equilibrium with different populations, according to the distribution rules of statistical mechanics. Ligands bind the state of the protein that leads

to the most energetically favourable complex, causing a shift in the population distribution of these states.

We would like to emphasize that many efforts in the world of biophysical chemistry are still focused on the theory of biomolecular recognition; nonetheless, even more complicated mechanisms than those presented here may be plausible. A unified model in which induced-fit and conformational selection act in parallel could be also considered.[33]

Experimental techniques

The experimental determination of kinetic constants of chemical reactions has a longstanding tradition in physical chemistry, stemming from the pioneering work of Waage and Guldberg in 1864 and being an already established discipline in the early years of the 20th century with the work by van't Hoff. However, the relatively recent availability of molecular biology techniques to produce large amounts of purified proteins and to master several types of polypeptide structures and sequence labelling/editing allowed, only in the last two decades, the extensive and routine measurement of kinetic constants of protein–drug interactions at several levels of precision and accuracy.

Radioligands

One of the very first techniques used for the biochemical characterisation of the formation of specific drug–receptor complexes was the radioligand binding assay, especially for GPCRs.[34] For decades, this method has essentially been the only one allowing for the detection of the binding of high-affinity radioligands to their receptors. Sampling over time allows the determination of on- and off-rates of the radioligand binding. Although this type of radioligand binding assay mainly gives information about the labelled ligand itself, this method has been very valuable for the detection of different ligand binding steps and allosteric regulation.

As radioligand binding assays require separation of bound ligands from free ligands, the equilibrium shifts during this process and sometimes only a part of receptor–ligand complexes may remain bound to receptors. For the determination of the kinetics of nonradioactive competitive ligands, several assay formats have been proposed;[35] but as shown in these studies, the number of samples required increases dramatically and only indirect information regarding the ligands' residence time can be acquired. The scintillation proximity assay (SPA) allows for the study of radioligand binding in a homogeneous system, without the need to separate non-bound components.[36] This method was very useful for high throughput screening,[37] but it has not been employed further in radioligand binding kinetic studies. The use of dual-point IC_{50}/K_i value determination for characterisation of a competitor's residence time has been suggested[38] (*i.e.* an unlabelled competitor is co-incubated with a radioligand for either a short time or a long time, ranging from minutes to several hours, and their corresponding IC_{50} or K_i values are calculated). This approach can be used only for ligands with significantly long residence time. Another limitation of this method is the preparation of specific "SPA beads" for particular assays, which requires high-quality receptor preparations in large enough quantities for immobilisation.[37] As the technology for the preparation of assay plates and beads continues to develop, the SPA method may become a very interesting alternative to the available approaches.

A recent, notable example is given from the radioligand-based determination of unbinding kinetics that has been recently shown for the cholesterol-binding translocator protein (TSPO),[39] a protein that increases endogenous steroid levels, inducing beneficial effects in different pathological conditions. A relatively poor relationship between binding affinity and steroidogenic efficacy for compounds possessing phenylindolylglyoxylamide structure (PIGAs) prompted the authors to investigate the ligands' kinetics. In order to interpret the outcomes, they coupled a kinetic radioligand binding assay to a theoretical mathematical model for the description of equilibrium kinetics.[40] This allowed the observation of a positive correlation between drug–target t_r and compound ability to stimulate steroidogenesis and its anxiolytic activity.

Fluorescence

During the last decade, fluorescence methods have become serious alternatives to radioactive methods for detection, monitoring and visualization of processes connected with cells and their components. Fluorescent ligands were first used for obtaining information about receptor localization at the subcellular level about 40 years ago,[41] but their usage as an alternative to radioligands for the characterisation of ligand binding properties has not been very successful so far.[42] However, fluorescence has several features that can be successfully implemented to develop methods for the characterisation of ligand binding to receptors. Among these methods, assays using fluorescence polarization-based fluorescence anisotropy (FA), Förster/fluorescence resonance energy transfer (FRET), time-resolved-FRET (TR-FRET) and fluorescence correlation spectroscopy (FCS) have found wider recognition and applications in GPCR studies.[42,43] In particular, FCS measures fluctuations in the fluorescence intensity of fluorescently labelled particles diffusing through a small illuminated detection volume (the confocal volume), allowing free and receptor-bound ligands[44] to be distinguished based on their different diffusion rates. The advantage of FCS is that it can be used at a single cell level and actual quantities of ligands can be measured, as was recently demonstrated for different GPCRs.[45,46] Moreover, the localization of receptors, as well as the formation of complexes or aggregates of receptors, can be monitored by mean of FCS at the same time.[47]

The FA method is based on the phenomenon that the fluorescent probes, when excited by plane polarized light, emit light with a certain degree of polarization depending on the fluorophore's rotational freedom within the fluorescence

lifetime. Therefore, since the formation of a ligand–receptor complex increases the level of polarization, the ligand binding process can be measured and monitored in real-time by means of this technique. The FA method has been successfully used to characterise binding of ligands to peptide receptors such as chemokine CXCR4,[48] melanocortin MC4,[49] muscarinic[50] and serotoninergic[51] receptors and binding of nucleotides to G proteins.[51] FA assay is ratiometric, therefore its signal can only be detected if the ratio between the bound and free forms of the fluorescent species has significantly changed.[52] This happens if the concentrations of receptor and ligand are in a comparable range, and at the level of the dissociation constant of their interaction. High concentrations of receptors are usually difficult to achieve with natural tissues; therefore, first attempts to implement this method have been performed with reconstituted systems.[48]

Chromatography

Chromatography has traditionally been used as a separative technique. However, it may also be used to investigate the distribution properties *per se* of an analyte. An example of this application is the chromatographic determination of partition coefficients,[53] which provides a rapid alternative to traditional methods for high-throughput work. Giddings and Eyring published their statistical approach to describe molecular migration in chromatography in 1955, considering the separation process as a Poisson distribution and the chromatographic peak as the probability density function for the elution of a solute as a function of time (assuming a single mode of binding). The shape of a chromatographic peak may be completely characterised by its zero (the area) and first (center of gravity) moments.[54] Thus, by measuring the chromatographic profiles of a retained and an unretained compound, one can directly determine k_{on}, k_{off} and K_d (assuming the receptor loading of the stationary phase is known). To ensure the kinetics are measurable, it is necessary to employ flow rates (u) that are sufficiently fast that equilibrium on the column cannot be established, whereupon peak broadening through kinetic effects occurs and can be modulated by tailoring the temperature conditions and mobile-phase composition. Talbert and coworkers[55] illustrated the technique by determining the binding characteristic of L-tryptophan, an archetypal plasma binding probe, using dimethyl sulfoxide (DMSO) as the unretained reference time. A "critical ratio" (η) can be used that quantifies the kinetic peak broadening and allows the separation of compounds according to their kinetic properties:

$$\eta = \frac{\sigma_R / E(t_R)}{\sigma_0 / E(t_0)} \quad (7)$$

where $E(t_R)$ and $E(t_0)$ are the mean times of the chromatographic peak of a retained and an unretained compound, respectively, and σ_R and σ_0 are the variances of the corresponding peaks. For the kinetics to be measurable, this ratio must be greater than 1, thus allowing the estimate of a kinetic factor k that can be used for the prioritisation of the examined compounds:

$$k = u(\eta^2 - 1)\sigma_R/\eta^2 \quad (8)$$

Surface plasmon resonance spectroscopy

The major modern and routinely used technique for the determination of k_{on} and k_{off} of protein–ligand (un)binding process is surface plasmon resonance (SPR) spectroscopy. SPR is an optical-based method that measures the change in the refractive index near a sensor surface, and has the advantages of being label-free and capable of real-time quantification of protein–ligand binding kinetics and affinities.[56] The sensor surface is a thin film of gold on a glass support (with particular reference to Biacore instruments), which forms the floor of a flow cell through which an aqueous solution flows continuously. The protein receptor molecules are immobilized on the sensor surface, and the ligand (*i.e.* analyte) is injected into the aqueous solution to detect the binding reaction. As ligands bind to immobilized receptor molecules, an increase in the refractive index is observed. After a desired association time (*i.e.* when all binding sites are occupied), a solution containing no ligands is injected through the flow cell to dissociate the ligands from the proteins. As the ligands dissociate from the immobilized protein, a decrease in refractivity is observed. The time-dependent response unit (RU) curves can then be used to calculate the kinetic association rate constant k_{on} and the dissociation rate constant k_{off}.[22]

The capacity of SPR to measure the real-time binding data makes it well suited to binding kinetics analyses, although the mass transport limitation makes it difficult to accurately measure k_{on} values faster than $\sim 10^6$ M^{-1} s^{-1}, as stated by Van der Merwe *et al.*[57] Compared to isothermal titration calorimetry (ITC), SPR has the ability to measure higher binding affinities, typically in the ranges of 10^{-6}–10 μM. For SPR, the highly reproducible affinity measurements, in conjunction with precise temperature control, allow the estimation of binding enthalpy *via* van't Hoff analysis.[58]

Although this measurement is not as rigorous as in the case of ITC, the procedure has the practical advantage of requiring much smaller amounts of protein sample.[57] Moreover, although the traditional SPR technique is not well suited to high-throughput screening (HTS), recent developments in SPR instrumentation, sensor chip design and sample preparation strategies show that SPR has a high potential for HTS, as recently shown for membrane protein ligands.[56,59] However, protein immobilization may affect the conformational and translational/rotational entropies and it therefore may influence the association rate evaluation.[60]

Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is well suited for the identification and characterisation of

molecules that can be ideally placed between "target to tractable hit" and "tractable hit to candidate", when resonances of the ligand (*i.e.* chemical shifts) are observed and no labelling is required, minimal amounts of protein are needed and there is no protein size limitation.

Dynamic events can be observed by chemical exchange, which occurs when a nucleus exchanges between environments with structural characteristics that lead to a difference in the local magnetic environment. These may be intramolecular or inter-molecular processes, the former including motions of protein side chains, helix–coil transitions of nucleic acids, unfolding of proteins, conformational equilibria, tautomerization; the latter binding of ligands to macromolecules, protonation/deprotonation equilibria of ionizable groups.[61] The exchange can be slow, intermediate or fast according to the timescale (or frequencies) at which they occur relative to NMR observables.

A proof of concept for measuring the unbinding rates of ligands from proteins has been provided by Pons and coworkers measuring the calcium ion k_{off} when released from calmodulin.[62] The timescale of the process can be determined by exploring the static magnetic field dependence of the dispersion profile[63] obtained by means of the so-called spin echo experiments.[64] In this case, the exchange contribution to transverse relaxation (*i.e.* the magnetization decay measured perpendicular to the static magnetic field) for a two-state system (in the fast exchange limit) is given by the following:

$$R_{ex} = p_a p_b \Delta\omega^2 / k_{ex} \quad (9)$$

where p_a and p_b are the populations of the two states, $\Delta\omega$ is the frequency difference between the chemical shifts of the two sites and k_{ex} is the sum of the pseudo-first order rate constants for the direct and reverse processes. The off-rate for the highest-affinity calcium binding site of the synaptotagmin I C2A domain obtained from NMR data (2.0×10^3 s^{-1}) is similar to the one derived *via* laser photolysis on caged calcium.[65]

One pharmacologically relevant example of a protein–ligand kinetic study[66] has been performed on the interaction between the highly soluble chromogenic peptide substrate (substrate C) and the active human immunodeficiency virus-1 (HIV-1) protease (PR), a homodimer that constitutes a prime target of drugs directed against AIDS.[67] The system was observed to be in slow exchange; hence, the difference in chemical shift ($\Delta\omega$) between the free (P) and bound (PL) form peaks of the protease, satisfies the following condition:

$$2\pi\Delta\omega^2 \gg \left(k_{off} + k'_{on}\right)^2 \quad (10)$$

where $k'_{on} = k_{on}$ [P] is a pseudo-first-order rate constant. This allows the determination of the limiting values for k_{on} and k_{off}, $<2 \times 10^4$ M^{-1} s^{-1} and <10 s^{-1}, respectively. 3D NOESY[66] spectra of the complex allowed the determination of exchange rate of the substrate orientation relative to the two monomers of PR, leading to a protein-substrate k_{off} between 1 and 5 Hz, in agreement with the line broadening observed during the titration. This result implies that the substrate dissociates from PR during the domain reorientation, otherwise the k_{off} estimated from the NOESY experiment and line broadening would differ. In contrast, it was found that the inhibitor KNI529 (a small and more tightly bound ligand) reorients without dissociation from PR.[66]

The potential for utilising the ubiquitous fingerprinting ^1H–^{15}N HSQC NMR experiment[68,69] to determine thermodynamic quantities, such as dissociation constants and kinetics from a single set of spectra corresponding to a protein–protein or protein–ligand titration, is alluring.

A chemical shift titration method based on the line shape analysis has been recently developed, where co-variation of the total protein and ligand concentrations allows for the determination of precise K_d values for 1 : 1 protein–protein (*i.e.* ubiquitin-Mms2) interactions and of kinetics in the fast exchange regime.[70] Due to the general consensus that line shape analysis is mainly a qualitative, or semi-quantitative method,[71] the accuracy of line shape analysis in the present study was determined by conducting quantum mechanical NMR simulations of the chemical shift titration methods,[72] whereas theoretical and experimental precision for k_{off} has been determined using Monte Carlo statistics alongside classical line shape analysis. The lower k_{off} limit is determined by the amount of line broadening, and the resulting impact on signal to noise ratio. The upper k_{off} limit was chosen as it represents the approximate K_d value (1.8 mM), beyond which the biological relevance of an interaction becomes questionable. Despite the presented application being referred to as multidimensional NMR spectroscopy, its extension to monodimensional spectroscopy should be quite straightforward in the near future.

Interestingly, it has been recently demonstrated[73] that hydrogen bond scalar couplings (*i.e.* an observable that encodes information about the distance and connectivity of atoms) can yield information on the residence times of molecular complexes in solution. Hydrogen bond scalar couplings were discovered in the late 90s as direct evidence of the existence of hydrogen bonds, initially observed for the hydrogen bonds of nucleic-acid base pairs and protein secondary structures and successively for the molecular interfaces of nucleic acids,[74] protein–nucleic acid complexes,[75] and other protein–ligand complexes.[76] The molecular exchange of binding partners *via* the disruption and reformation of a complex alters the values of inter-molecular hydrogen bond scalar couplings. The residence time of a complex can be related to the apparent values of the inter-molecular hydrogen bond scalar coupling constants measured by quantitative J-modulation spin-echo difference experiments.[75,77] The observation of inter-molecular hydrogen bond scalar couplings is possible only if the residence time is sufficiently long ($\gtrsim$10–2 s).

X-ray scattering

Single crystal diffractive techniques have been used over many decades for the determination of the atomic level protein structure of biomolecules; therefore, X-ray scattering of molecular solutions could be used for the analysis of the binding properties of complexes. The scattering profile of such a mixture can be observed by means of small angle and wide angle X-ray scattering (SAXS an WAXS, respectively), and can be represented as a linear combination of individual scattering intensity contributions from the different molecular species.[78] In the absence of high resolution structural models, the one-dimensional SAXS data can be utilised to generate low-resolution (typically, 30–10 Å) three-dimensional models of macromolecules and their assemblies in solution. The minimum number of species can be obtained with singular value decomposition (SVD) or principal component analysis (PCA), although difficulties could arise from the presence of long-living transient protein–ligand structures.[79]

In perspective, the kinetics of protein–ligand interaction, as well as the dynamics, can be monitored *via* the time evolution of X-ray scattering of the molecules of interest in solution. Spatial resolution and unbiased retrieval of structural information are lost when the X-ray scattering pattern from an ensemble of randomly oriented and unequally spaced macromolecules is observed: nevertheless, changes in the overall dimension/mass can be measured while higher resolution information (relative position of different subunits, domains, or secondary structure elements) shows fingerprints in the WAXS region. In 2008 Cammarata and coworkers[80] demonstrated that in a time-resolved WAXS experiment (TR-WAXS), a laser pulse is used to trigger the protein structural change, and transient structures are then followed by delayed X-ray pulses. Structural changes occurring in the sample leave their 'fingerprints' measured in the differences between the signals before and after the laser initiates the reaction and can be monitored as a function of time. The authors explore the case of haemoglobin (Hb) interactions with carbon monoxide (CO), photolyzing the Hb–CO bonds by means of a laser pulse in the green light wavelength (527 nm), thus triggering a tertiary relaxation that is dominated by local structure changes. Strain at the Hb subunit interfaces drives a large-amplitude quaternary structural transition leading to T-state Hb. Finally, CO recombines bimolecularly with T-state Hb to regenerate the R-state. Data patterns representing tertiary relaxed proteins (200 ns) and the fully deoxygenated ones (100 μs) can be identified, thus demonstrating that the time resolution of TR-WAXS can reach the nanosecond timescale. Therefore, in the future, this technique can be successfully adapted, extended and applied to the case of ligands dissociating from proteins.

X-ray free-electron lasers (XFELs) represent an exciting new alternative if sub-picosecond time resolution is needed, despite the fact that they are still in an immature stage. XFELs are able to produce extremely short (~10 fs) X-ray pulses with a number of photons $\sim 10^2$–10^3 times higher than the ~100 ps pulses produced at synchrotrons.[81] XFELs have been used to perform the first steady-state serial femtosecond crystallography (SFX) experiments on protein crystals.[82,83] The signal evolution at timescales longer than a few tens of ps, can usually be attributed to exchanges in the population of a finite set of different molecular species, each with a well-defined time-independent structure. When a plausible kinetic model is available, it can be used to analyse the data in terms of a generalized approach (GA) that retrieves the fingerprint of a given physical species on the basis of its predicted time evolution.[84]

Computational studies

Binding and unbinding kinetics are emerging as key factors for predicting drug efficacy in living organisms. As mentioned in the previous section, there are several experimental techniques for studying (un)binding kinetics, but there are currently no efficient computational approaches for predicting absolute kinetic parameters. The few attempts reported in the literature have mainly been based on brute-force molecular dynamics (MD) simulations, which are highly demanding in terms of time and computational power. In MD, Newton's equation of motion is integrated for each atom at each step of the simulation to give a detailed, all atom description of the evolution of the system over time. However, MD has typically suffered from the so-called timescale problems, as accessible simulation times have traditionally been shorter than those required by slow biomolecular processes, such as binding/unbinding, to take place. Thus, several methods have been developed to more efficiently improve sampling and explore kinetics by means of computer simulations. Herein, we report on recent applications of computational methods for the investigation of binding kinetics.

Plain MD

Referring to standard simulations of ordinary length, Huang and Caflisch simulated the spontaneous dissociation of 6 small ligands from the FK506 binding protein (FKBP).[85] One hundred independent, relatively short runs, lasting up to 20 ns each, were carried out for each of the six ligands, whose affinities were in the mM range. A network analysis was built based on inter-molecular distances between the ligands' heavy atoms and eight residues in the active site. A cutoff distance of 15 Å was chosen between the ligand and the active site center of mass to designate the unbinding event. A single exponential fit was then applied to the cumulative distributions of the dissociation events recorded in each trajectory to calculate the unbinding time of each ligand. Low energy barriers between different binding poses, compared to the rate-limiting step of dissociation, led to the simple exponential time-dependence of dissociation kinetics.

One hardware-based solution that translates into a brute force approach to simulations of biomolecules has been applied using the massively parallel supercomputer ANTON, specifically devised by the D. E. Shaw Research group to

perform long MD simulations. By relying on such architecture, the first examples of unbiased MD for spontaneous protein–ligand binding simulations were reported, and kinetic information was extracted. Starting from a completely unbound state and without employing any driving forces, the binding processes to Src kinase by the anti-cancer drug Dasatinib and the inhibitor PP1 were investigated at atomic-level detail. Independent runs for a total of 150 μs simulation time were carried out, which successfully reproduced the correct, crystallographic binding mode for both ligands. Based on the single binding event recorded for Dasatinib and the three observed for PP1, an estimate for the on-rates was given, with a k_{on} of ~1.9 s^{-1} M^{-1} in fairly good agreement with the experimental value of ~5 s^{-1} μM^{-1} for the former ligand. Extensive MD simulations of the β2 and β1 adrenergic receptors, pharmaceutically relevant examples of G-protein coupled receptors, were also performed in order to elucidate the binding mechanism of known agonists and antagonists.[86] The direct calculation of the on-rate was complicated by varying ligand concentrations to mimic their partition into the lipid bilayer: for Alprenolol, k_{on} of 3.1×10^7 s^{-1} M^{-1} was calculated, which is in line with the experimental value of about 1.0×10^7 s^{-1} M^{-1}. Interestingly, the dominant pathway leading to the protein binding pocket was elucidated and explained as a two-step process. The desolvation prior to ligand entry in an extracellular vestibule is the first, rate-limiting step, which acts as a kinetic bottleneck for the subsequent ligands' access to the binding pocket. Importantly, both studies underline how water desolvation influences the kinetics of binding.

An elegant approach to the problem of kinetics in computational studies is the application of Markov State Models (MSMs), a powerful statistical tool that allows the extraction of both equilibrium thermodynamic and kinetic properties from long MD simulations.[87] A crucial step affecting the quality of an MSM is the identification of the so-called metastable states. To reach this end, the configurational space visited by the system must be grouped into its most relevant states. Classifying data items into groups is a general issue, which is typically addressed by means of the clustering technique. Different clustering algorithms exist and have been successfully applied to computer aided drug design over the years, ranging from docking pose classification[88,89] to analysis of MD simulations.[90–97] Construction of an MSM relies on clustering to give a coarse grained representation of the metastable states of a system. Geometrically similar configurations are typically assumed to possess similar kinetic properties. In this view, intra-state transitions are considered to be faster than inter-state transitions. The model is Markovian, meaning that a transition from state i to state j only depends on state i and not on previously visited states. Once the coarse grained representation has been identified, a transition matrix is built and the interconversion rate between the metastable states can be estimated. Although an adequate sampling is still needed as a starting point to build the model, a clear advantage is that the method can be applied to aggregates of long MD simulations carried out in parallel, instead of relying on a single, or a few, μs-long simulations.

A detailed reconstruction of the benzamidine binding process to trypsin was obtained from analysis of 495 independent MD runs, for a total simulation time of about 50 μs.[98] Construction of several MSMs allowed for the identification of metastable states, the rate-limiting step and the transition states involved. The on- and off-rates were estimated from the mean first passage time (MFPT) for the direct and inverse reactions, calculated as the average time taken to go from the unbound to bound and from the bound to unbound clusters, respectively. While evaluating k_{on}, the ligand concentration must also be taken into account. Both results deviate by at least one order of magnitude from the experimental values, with the typical larger uncertainty on the k_{off}, most likely due to poor sampling of dissociation events or to intrinsic force field limitations.

The same protocol was subsequently followed to study the carboxy-thiophene fragment association with AmpC β-lactamase.[99] Besides reproducing all of the known binding poses reported from crystal structures, a new one was identified that possessed peculiar kinetic properties. Specifically, the closure of a loop stabilises such a binding mode, providing an explanation for the highest residence time calculated for the cluster from the simulations.

Application of MSMs to MD simulations (65 and 200 ns long) of the lysine-, arginine- and ornithine-binding protein in the presence of their ligand L-arginine allowed the identification of the involvement of both conformational selection and induced fit models in the binding mechanism.[100] In a subsequent study on the choline-binding protein belonging to the same protein family, a flux analysis was combined with the previous framework. This allowed the quantitative characterisation of the contribution from the two models, resulting in the predominance of the conformational selection mechanism.[101]

Breaking down complex transitions into Markov model microstates representing apo and holo states of trypsin was the strategy recently adopted by Plattner and Noé to understand the binding mechanism of benzamidine.[102] The pathway leading to the association occurs on the same timescale of the interconversion of the apo states of the protein. This first step, dominated by conformational selection, is followed by a population shift towards a bound state, implying the induced-fit. On- and off-rates were evaluated from the MFPT, with the calculated k_{on} of $6.4 \pm 1.6 \times 10^7$ s^{-1} M^{-1} being in excellent agreement with the reported experimental value of 2.9×10^7 s^{-1} M^{-1}.

Another approach, based on classical MD simulations and machine learning, was exploited by Decherchi *et al.* to elucidate how a transition state analogue, DADMe-imucillin-H, associates with purine nucleoside phosphorylase.[103] Simulation data were gathered for a total of 13 μs and a kinetic model was derived, describing relevant intermediates and identifying three different routes leading to ligand association to the protein. The mean first time for binding was calculated from

the simulations, and the obtained value of 216 ± 101 ns compared well with 246 ns derived from experiments.

Recently, the adaptive multilevel splitting algorithm (AMS[104]), a rare event sampling technique, was applied for the first time to an MD simulation of protein–ligand unbinding.[105] The authors showed how the mean first passage time for the dissociative process can be estimated from the time spent by the system in nonreactive trajectory loops and in reactive paths. The method was applied to determine the unbinding time of trypsin from benzamidine. A total simulation time of about 2 μs was collected, leading to an estimate of the dissociation rate within the same order of magnitude as the experimental value.

Biased approaches

Although reaching higher timescales is becoming increasingly feasible, it is clear that the computational resources required are not easily affordable for most research groups. These methods are also unsuitable for industrial use, where dozens of compounds must be prioritised in the hit-to-lead and the lead optimisation phases. Importantly, since the residence time (t_r) of molecules can be in the order of seconds, minutes, or even hours, directly determining (un)binding rates *via* computational methods is extremely challenging. Thus, smarter algorithms and effective practical solutions are needed to tackle the problem of kinetic rates estimation.

In the effort to improve sampling and guide the exploration of the configurational space of a system towards its slow degrees of freedom, biased methods such as parallel tempering, umbrella sampling, steered molecular dynamics, temperature-accelerated molecular dynamics and metadynamics (typically referred to as enhanced sampling methods), random acceleration molecular dynamics and scaled molecular dynamics, among many others, have been specifically devised. Even if the application of such techniques is typically aimed at reconstructing free energy surfaces (FES), we report here on their employment in recent years for investigating protein–ligand binding.

In metadynamics (MetaD), exploration of the configurational space is guided by adding a bias potential along specific collective variables (CVs). Such variables must be chosen based on their ability to represent slow degrees of freedom and distinguish the different states of the system. The bias potential is added as a sum of Gaussian functions that are deposited on points of the CV space that are visited, thus discouraging the system from assuming configurations that have already been explored and favouring sampling in different regions of the CV space. In this way, the method allows large free energy barriers to be overcome and thus facilitates the efficient characterisation of complex FES. Different replicas of the system are simulated at the same time in the bias-exchange metadynamics (BE-META). In each replica, a different CV is biased and exchanges between different replicas are allowed.[106] A set of 7 CV (including hydrogen bond count and distances between ligand and target) was chosen for the simulation of the binding process of a peptide substrate to HIV-1 protease. Based on the BE-META simulations, thermodynamic and kinetic models were derived and association and dissociation rates were calculated, obtaining values consistent with kinetic experimental data available for similar ligands. The dissociation of a COX-2 selective inhibitor was simulated by Limongelli *et al.* by means of well-tempered MetaD, a formulation of MetaD that has been shown to offer faster convergence properties.[107] A path CV was used for biasing the gating through two helices, while the position and orientation of the ligand with respect to the protein were controlled by means of a distance and a dihedral angle CV. Besides the well-known binding mode for the ligand, a second basin was identified while simulating the ligand unbinding, which corresponds to a different pose inside the pocket with good thermodynamic stability. Thus, the longer residence time of the considered ligand was interpreted as related to the rearrangement between these two stable binding modes.

Tiwary and Parrinello recently formulated a methodology to determine kinetic properties from simulations biased *via* MetaD.[108–110] With a minimal additional computational cost, pathways, binding rates and rate-limiting steps can be derived. Thus, k_{on} and k_{off} were computed from MetaD simulations of benzamidine unbinding from trypsin. The results agreed well with the experimental values, and a detailed representation, including transition states and rate-limiting steps, was given. Application of this innovative methodology to different ligand–target complexes for calculating kinetic rates is extremely promising, but still rather time-consuming.

A system can be dragged from an initial to a final state acting along a reaction coordinate by means of steered molecular dynamics (SMD).[111] This computational technique involves the application of a potential – typically a constant velocity or force – to a subset of a system's particles along predefined degrees of freedom. Critical parameters, on the calibration of which particular care must be taken during setup, are the pulling velocity and force constant applied. The potential of mean force (PMF) can be reconstructed and the free energy difference along the chosen reaction coordinate calculated. In the context of protein–ligand binding, the method can be employed to pull a molecule out of its binding site and determine the corresponding PMF profile.

The unbinding of two drugs, Sunitinib and Sorafenib, from the human vascular endothelial growth factor receptor 2 (VEGFR2), was studied by Capelli *et al.* by means of SMD.[112] The distance between the ligand center of mass and the centroid of the Sorafenib binding site was used as a reaction coordinate to reconstruct the PMF profiles of the unbinding process for the two molecules. The analysis reveals that the PMF of Sorafenib (63.5 ± 7.2 kcal mol^{-1}) is higher than that of Sunitinib (29.6 ± 4 kcal mol^{-1}), qualitatively in agreement with the experimental k_{off} and residence time values (10^{-5} *vs.* 10^{-3} s^{-1}) reported for the two drugs.

Umbrella sampling (US)[113] was recently exploited to investigate the molecular features governing a mechanism of slow-onset inhibition.[114] In US, a harmonic restraining potential is

applied to a series of windows that are equally spaced along a reaction coordinate. The free energy profile in the considered space of the reaction coordinate can then be determined *via* a reweighting procedure, typically the weighted histogram analysis.[115] The partial nudged elastic band method[116] was applied to generate intermediate steps leading from the open- to the closed-conformations, both identified *via* X-ray crystallography, of *Mycobacterium tuberculosis* InhA's substrate-binding loop (SBL). The free energy profiles of different enzyme-inhibitor complexes were then determined *via* US. The closed state of SBL was shown to be uniquely stabilized by slow-onset inhibitors *via* an induced-fit mechanism, unveiling the molecular basis of the slow-onset kinetics of this class of compounds.

Site-directed mutagenesis experiments were combined with temperature-accelerated MD (TAMD)[117] by Guo *et al.* to identify relevant protein residues in the unbinding pathway of the antagonist ZM241385 from adenosine A2A receptor, a prototypical GPCR.[118] In TAMD, sampling is enhanced by harmonically tethering a CV to a fictitious particle that is subject to Brownian motion at a higher temperature. The unbinding events of the antagonist were simulated, applying

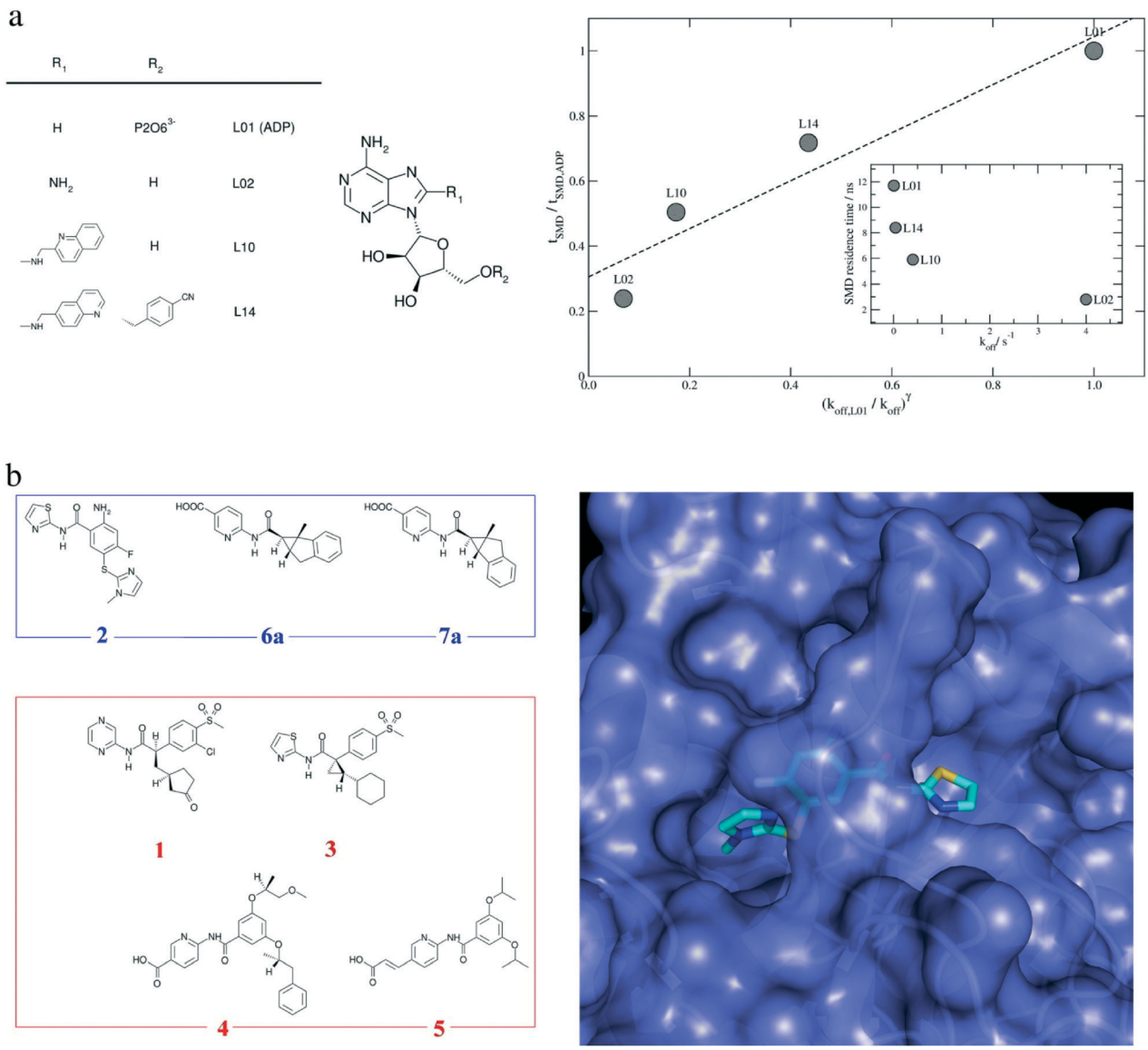

Fig. 3 (a) Grp78 ligand prioritisation *via* Scaled MD.[120] The ligand's scaffold is reported together with single ligand names and their substituents in positions R1 and R2. The scaled MD based *vs.* experimental normalized residence times are represented in the right part of the figure, as reported by Mollica *et al.* (the dashed line is the regression line of the four points represented in the graph). In the inset, the absolute values of residence time are reported. (b) GK1 ligands prioritised according to the scaled MD-based methodology described in the text are reported and grouped according to their molecular shape and computational residence times.[128] Ligands **2**, **6a** and **7a** (enclosed in the blue line) possess a linear shape and have a short computational residence time (respectively 29, 26 and 25 ns; the last two values are relative to a racemic mixture, whereas in the figure only the R entantiomer is represented for simplicity). Ligands **1**, **3**, **4** and **5** (enclosed in the red line) are T-shaped and have a longer computational residence time than the ones in the first group (105, 39, 93 and 99 ns, respectively).

TAMD, and contacts with protein residues along the dissociation pathway were registered. Subsequent mutagenesis of the residues identified allowed the determination of those more significantly affecting the dissociation characteristics of ZM241385. The authors noted how determinant the contribution from the MD simulations had been, as they recognized the importance of residues that would otherwise have gone unnoticed. This study undoubtedly highlights the potential of optimized kinetic properties in the rational design of new drugs.

Niu and co-workers[119] made use of random acceleration molecular dynamics (RAMD) to study protein–ligand unbinding. In RAMD, a force in a random direction is applied to a ligand and after a set of MD steps a new random force is applied if the molecule has not moved by a predefined distance, thus allowing a more extensive sampling of the conformational space. RAF kinase, a protein with a central role in the conserved RAS–RAFMEK–ERK signal transduction pathway for cell proliferation and survival, has been studied by the authors, with particular focus on the unbinding processes of the inhibitors PLX4720 and TAK-632 from B-RAF. The ligands could successfully be distinguished according to their binding energetics, in agreement with experimental kinetic data. Residue level information about the interactions that are able to retain the ligands for longer residence times was also provided by the study, suggesting new routes to the synthesis of novel, promising therapeutic compounds.

Exploiting kinetics in drug discovery: ligand prioritisation *via* scaled MD

The aforementioned methods are able to simulate dissociation events and can provide indications about the way dissociation occurs; however, rates seem to be elusive. Moreover, such advanced, system-dependent techniques are still unsuitable to be applied in a routine manner, as required in hit-to-lead or lead optimisation phases of a drug discovery campaign. This unavoidably calls for smarter algorithms and effective practical solutions to tackle the problem of exploiting kinetic information in the development of new drugs.

Recently, a novel computational method has been reported that addresses the challenge of predicting unbinding kinetics of ligands in a relative manner.[120] The requirement in terms of computational resources is lower compared to other approaches. Therefore, application in the hit-to-lead and lead optimisation phases of drug discovery is extremely promising. Rather than trying to predict the absolute off-rate value on individual complexes, the authors established a procedure to identify the correct k_{off}-based ordering relationship among congeneric compounds, which bind to a given target. The solution is rooted in the enhancement of the transition probability between different free energy minima during MD simulations by means of scaled potentials.[121–123] Thus, in the resulting scaled molecular dynamics (Scaled MD), the rupture of all the fundamental physical interactions that confer stability to a protein–ligand complex is facilitated, leading to unbinding in much shorter simulation timescales. Application of scaled potentials to the entire system is not advisable and may lead to loss of stability of the overall structure of the molecular target. A countermeasure to this has been the application of proper harmonic restraints that preserve the overall correct folding, while leaving the regions involved in the binding process unrestrained. For each protein–ligand system, multiple runs of scaled MD simulations are performed and stopped once the ligand is released from the binding site and fully solvated. Finally, a bootstrap analysis[124] on the simulated unbinding times per target is done in order to assess the statistical significance of the observations and to possibly decide whether to increase the number of runs per complex.

In the first application of this methodology, three different systems were analysed, namely the Heat Shock Protein 90, the 78 kDa Glucose-Regulated Protein and the Adenosine Receptor A2A. Four ligands could be ranked correctly for each system in agreement with experimental kinetic data obtained by means of SPR measurements.[125–127] Twenty unbinding simulations were carried out per ligand to determine the average unbinding times. This led to a linear distribution of rates for all the systems. Interestingly, the method was able to correctly reproduce the experimental ranking even when significant variations were observed both in terms of molecular volume and net charge of the ligands. The small and charged ATP could be distinguished well from other larger and globally neutral compounds unbinding from Grp78 (Fig. 3a).

In a subsequent publication,[128] the method was tested on a series of chemically unrelated ligands of glucokinase (GK, GK1, or hexokinase IV),[129] a system of primary pharmaceutical interest due to its involvement in Type 2 Diabetes Mellitus.[130] An interesting feature that emerged from this study is the influence of the ligand shape on the residence times. The compounds displaying a more pronounced linear shape (2, 6, and 7) exhibit a significantly shorter residence time than those with a T-shaped geometry (1, 3, 4, and 5), with an average t_r of ~27 ns for the former molecules and ~90 ns for the latter (Fig. 3b). This can be explained by inspection of the GK activators binding site – possessing a rather linear shape – which is easily accessible to ligands having a linear geometry, whereas molecules bearing other shapes than linear require an induced fit binding mechanism. Therefore, the present computational approach can help to discern among rather different scaffolds, and hence prioritise, for subsequent chemical synthesis, those most promising from residence time and lead efficacy standpoints. Furthermore, despite the loss of details associated with Scaled MD simulations, precious information can still be gained on the chemical features responsible for longer residence times, thus establishing structure–kinetics relationships (SKRs).

The applications discussed above highlight how the computational community is recently focusing much of the interest in studying the ligand binding process and determining the related rates. In principle, average binding times can be

directly calculated *via* plain MD if several binding events are observed.[103] However, this is currently only feasible for relatively fast binders, and even in this case, an enormous amount of computational resources is required; in view of this, calculating unbinding times is still unreachable. In the simplest picture, achieving the bound state first involves free diffusion of the two partners into the solvent to get in proximity, followed by overcoming the ΔG_{on} free energy barrier (Fig. 1). Dissociation relies instead on the disruption of all the interactions between the protein and the ligand in the bound state, a process that typically requires significantly higher energy barriers to be crossed (ΔG_{off}, Fig. 1). The transition states along binding pathways and the height of the corresponding free energy barriers can be determined *via* free energy based methods, such as MetaD and US. In the limit of a fully converged simulation, both binding and unbinding rates can be thus derived as indicated in eqn (5a) and (b). The error of the measured rates is then related to the uncertainty on the free energy barrier and on the accuracy of the pre-exponential factor. Scaled MD was used to accelerate unbinding events and successfully rank ligands consistently with their experimental k_{off}.[120] However, the present application does not allow the determination of absolute values of the off-rates. Unlike free energy based methods, in MSMs kinetic rates are directly calculated from a transition matrix. With an adequate underlying sampling, MFPTs for both unbound to bound and for bound to unbound transitions can be determined from these on- and off-rates, respectively. In general, while MSMs may gain in efficiency, more mechanistic details about the pathway may result from the application of free-energy based methods.

Mastering residence time as a novel tool in drug discovery

Maximising the thermodynamic affinity (*i.e.* K_d or IC_{50}) of a drug for its target has been the traditional strategy applied in the optimisation step of drug discovery programs. However, it is now well accepted that residence time, rather than affinity, often drives the efficacy of a drug *in vivo*.[7,35,131] Moreover, the vast majority of drug candidates fail in human clinical trials, owing to a lack of efficacy or an insufficient therapeutic index, often as a result of limited target engagement or concomitant drug binding to off-target receptors.[132,133] Currently, all these critical parameters are not routinely evaluated fully until late stages of drug discovery. This suggests that a conscious mastering of residence time could represent an essential step in the field. Improving the understanding of pharmacokinetics (PK) and pharmacodynamics (PD), while at the same time being aware of kinetic properties throughout the entire drug discovery process, could lead to an increase in the success rate of approved new drugs.[133,134] Prolonged occupancy of the designated target by a drug while minimising binding to off-target proteins (referred to as kinetic selectivity) is a promising strategy to improve a drug candidate's therapeutic index. Indeed, a significant fraction of the marketed drugs dissociate slowly from their targets, emphasizing the importance of the drug–target complex lifetime for *in vivo* drug activity. Although compelling arguments can be made for the tuning of drug–target kinetics in a drug discovery campaign, major barriers still exist for the implementation of this approach, including the lack of prospective tools that integrate drug–target residence time parameters with PK models to yield predictions of drug efficacy. Current PD models typically assume 'rapid equilibrium' between the target and the fraction of the drug in human plasma that is not protein-bound. Moreover, during drug discovery and optimisation, it is common to characterise compound activity with steady-state *in vitro* measurements, although recent advances in the field have been made to include kinetics in the PD models. This has been shown in a recently published study on the LpxC enzyme from *Pseudomonas aeruginosa*, in which a detailed analysis of dose response curve prediction for inhibitors of the enzyme was carried out in an animal model of infection.[135]

On the ligand side, reversible covalent binding to non-catalytic cysteines may represent a widely applicable method for obtaining prolonged residence times.[136–138] Drugs relying on intrinsically irreversible chemistry, such as acrylamides, are more likely to form permanent covalent adducts with off-target proteins.[139,140] These include both closely related targets (*e.g.* off-target kinases) and unrelated targets with hyper-reactive cysteines,[141] thus explaining why reversible covalent drugs are typically preferred. An approach to the discovery of such inhibitors was recently reported for a kinase-recognition scaffold.[142,143] A residence time of up to several hours was obtained by means of a covalent, but fully reversible, attachment of a cyanoacrylamide electrophile to cysteine in the ribosomal S6 *kinase* (RSK2).[143] The approach has been recently applied to research on B-cell chronic lymphocytic leukaemia and mantle cell lymphoma,[144,145] leading to the identification of inhibitors possessing remarkably slow off-rates, with residence times reaching one week.

Integrating both theory and experiment can be a successful strategy to gain a full description of the kinetic properties of a compound.[114,118] MD simulations were coupled by Seow and co-workers to experimental data in order to understand the peculiar activity of a ligand.[146] Three equipotent nanomolar antagonists (3D53, W54011, JJ47) of the inflammatory responses to C5a, a pro-inflammatory and chemotactic factor acting *via* the C5a receptor 1 (C5aR) on leukocytes, were compared. Surprisingly, the least drug-like antagonist 3D53 maintained potency in cells against higher C5a concentrations. Inhibition of macrophage responses was also much longer ($t_{1/2} \sim 20$ h), compared to W54011 or JJ47 ($t_{1/2} \sim 1$–3 h). In order to gain the atomic level details of the molecular features determining such unusually long residence times, 3D53 in complex with its receptor was simulated *via* MD. A unique ligand-induced conformational change was revealed, causing the antagonist to be trapped between specific transmembrane helices in the receptor. However, although 3D53 was orally more efficacious than W54011 or JJ47 in

preventing repeated agonist exposure inducing rat paw oedema over 24 hours, the oral bioavailability of the compound was negligible.

Conclusions and perspectives

In the present article, we provided a general overview on protein–ligand binding kinetics to introduce the reader to the underlying theory and the methods available to investigate this field. We highlighted how the advances in experimental and computational methods allowed a broader application of existing techniques and development of new approaches. Indeed, remarkable progress in technology has been made over the recent years. As a consequence, an increasing and heterogeneous volume of biological, pharmacological and structural data is becoming available. In this extremely dynamic context, big data analysis is emerging as a novel approach aimed at exploiting such precious information for drug discovery.[147]

We presented the most recent experimental and computational techniques employed in the characterisation of protein–ligand binding kinetics and we examined examples of their application. Recent works that put a strong focus on the usage of residence time as a therapeutic and functional index have been also highlighted. Indeed, the importance of including kinetics information in the drug discovery process is increasingly being recognized. In particular, since its first presentation,[148] the drug–target residence time model has been cited more than 600 times in the scientific literature. It has been applied broadly to drug discovery programmes, across multiple therapeutic areas, leading to the identification of numerous clinical-stage drugs.[149,150] Significant advances in the exploration of binding and unbinding mechanisms have been made and much effort has been focused on characterising the underlying kinetics. However, the current understanding of the factors influencing binding rates remains incomplete. Therefore, kinetic descriptors and predictors at the atomic level will likely be the next generation challenge in the field of drug discovery, as suggested by the studies reported in the present review. The future design of drugs possessing optimised binding kinetics will require a detailed characterisation of the entire drug–receptor binding pathway, the relevant metastable intermediate states involved and the binding mode reached inside the binding pocket. Computational methods based on MD, combined with the available experimental approaches, are emerging as extremely promising tools.[108,120,128] Indeed, identifying the molecular features and the relevant intermediate states involved in the binding process is becoming increasingly feasible. Such a complete understanding will provide the insights needed to guide a rational modulation of binding kinetics in the desired manner. For instance, functional groups can be inserted or modified in proper regions of the scaffold in order to facilitate interaction with specific residues of the protein along the binding pathway. Notably, such modifications on the active scaffold should be possibly conceived so that they do not affect binding affinity. Finally, in an almost orthogonal manner, solubilising groups can be properly added in order to also optimise pharmacokinetic properties.[151]

Altogether, the aspects discussed in the present review highlight the importance of considering kinetics, alongside thermodynamics, in the effort to both identify new drugs and improve efficacy. As outlined, gathering such information is becoming increasingly accessible to both the experimental and computational fields.

Acknowledgements

We gratefully acknowledge Dr. Matteo Masetti (University of Bologna, Italy), Dr. Grazisa Rossetti (INGM, Italy) and Doris Schuetz (University of Vienna, Austria) for having carefully read the manuscript and for critical suggestions.

References

1 R. Perozzo, G. Folkers and L. Scapozza, *J. Recept. Signal Transduction*, 2004, **24**, 1–52.
2 J. B. Chaires, *Annu. Rev. Biophys.*, 2008, **37**, 135–151.
3 J. N. Langley, *J. Physiol.*, 1905, **33**, 374–413.
4 R. A. Copeland, D. L. Pompliano and T. D. Meek, *Nat. Rev. Drug Discovery*, 2006, **5**, 730–739.
5 A. C. Pan, D. W. Borhani, R. O. Dror and D. E. Shaw, *Drug Discovery Today*, 2013, **18**, 667–673.
6 D. C. Swinney, *Curr. Opin. Drug Discovery Dev.*, 2009, **12**, 31–39.
7 H. Lu and P. J. Tonge, *Curr. Opin. Chem. Biol.*, 2010, **14**, 467–474.
8 S. A. Lipton, *Nature*, 2003, **428**, 473.
9 S. Ohlson, *Drug Discovery Today*, 2008, **13**, 433–439.
10 G. Vauquelin, S. Bostoen, P. Vanderheyden and P. Seeman, *Naunyn-Schmiedeberg's Arch. Pharmacol.*, 2012, **385**, 337–372.
11 W. Keighley, *Drug Discovery World Summer*, 2011, **12**, 39–45.
12 P. Bongrand, *Rep. Prog. Phys.*, 1999, **62**, 921.
13 Y.-H. Xie, Y. Tao and S.-Q. Liu, *J. Biomol. Struct. Dyn.*, 2013, **31**, 98–100.
14 M. Held, P. Metzner, J. H. Prinz and F. Noé, *Biophys. J.*, 2011, **100**, 701–710.
15 S. Arrhenius, *Z. Phys. Chem.*, 1889, **4**, 96–116.
16 D. G. Truhlar, B. C. Garrett and S. J. Klippenstein, *J. Phys. Chem.*, 1996, **100**, 12771–12800.
17 K. J. Laidler, *Theories of Chemical Reaction Rates*, McGraw-Hill, 1969.
18 C. F. Shuman, L. Vrang and U. H. Danielson, *J. Med. Chem.*, 2004, **47**, 5953–5961.
19 K. Andersson and M. D. Hämäläinen, *J. Chemom.*, 2006, **20**, 370–375.
20 R. A. Copeland, *Future Med. Chem.*, 2011, **3**, 1491–1501.
21 N. Smolin, A. Oleinikova, I. Brovchenko, A. Geiger and R. Winter, *J. Phys. Chem. B*, 2005, **109**, 10995–11005.
22 X. Du, Y. Li, Y.-L. Xia, S.-M. Ai, J. Liang, P. Sang, X.-L. Ji and S.-Q. Liu, *Int. J. Mol. Sci.*, 2016, **17**, 144.

23 R. Buonfiglio, M. Recanatini and M. Masetti, *ChemMedChem*, 2015, **10**, 1141–1148.
24 C.-E. Chang and M. K. Gilson, *J. Am. Chem. Soc.*, 2004, **126**, 13156–13164.
25 P. T. Corbett, L. H. Tong, J. K. M. Sanders and S. Otto, *J. Am. Chem. Soc.*, 2005, **127**, 8902–8903.
26 H. R. Bosshard, *News Physiol. Sci.*, 2001, **16**, 171–173.
27 H. Frauenfelder, S. G. Sligar and P. G. Wolynes, *Science*, 1991, **254**, 1598–1603.
28 D. W. Miller and K. A. Dill, *Protein Sci.*, 1997, **6**, 2166–2179.
29 J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins: Struct., Funct., Genet.*, 1995, **21**, 167–195.
30 K. Henzler-Wildman and D. Kern, *Nature*, 2007, **450**, 964–972.
31 D. D. Boehr, R. Nussinov and P. E. Wright, *Nat. Chem. Biol.*, 2009, **5**, 789–796.
32 R. Nussinov, B. Ma and C. J. Tsai, *Biophys. Chem.*, 2014, **186**, 22–30.
33 T. R. Weikl and C. Von Deuster, *Proteins: Struct., Funct., Bioinf.*, 2009, **75**, 104–110.
34 A. Rinken, S. Veiksina and S. Kopanchuk, *Pharmacol. Res.*, 2016, **113**(Pt B), 747–753.
35 D. Guo, J. M. Hillger, A. P. Ijzerman and L. H. Heitman, *Med. Res. Rev.*, 2014, **34**, 856–892.
36 S. Udenfriend, L. Gerber and N. Nelson, *Anal. Biochem.*, 1987, **161**, 494–500.
37 J. F. Glickman, A. Schmid and S. Ferrand, *Assay Drug Dev. Technol.*, 2008, **6**, 433–455.
38 C. E. Heise, S. K. Sullivan and P. D. Crowe, *J. Biomol. Screening*, 2007, **12**, 235–239.
39 B. Costa, E. Da Pozzo, C. Giacomelli, E. Barresi, S. Taliani, F. Da Settimo and C. Martini, *Sci. Rep.*, 2016, **6**, 18164.
40 H. J. Motulsky and L. C. Mahan, *Mol. Pharmacol.*, 1984, **25**, 1–9.
41 E. Melamed, M. Lahav and D. Atlas, *Nature*, 1976, **261**, 420–422.
42 R. Sridharan, J. Zuber, S. M. Connelly, E. Mathew and M. E. Dumont, *Biochim. Biophys. Acta, Biomembr.*, 2014, **1838**, 15–33.
43 L. Chen, L. Jin and N. Zhou, *Expert Opin. Drug Discovery*, 2012, **7**, 791–806.
44 S. J. Briddon and S. J. Hill, *Trends Pharmacol. Sci.*, 2007, **28**, 637–645.
45 R. Corriden, L. E. Kilpatrick, B. Kellam, S. J. Briddon and S. J. Hill, *FASEB J.*, 2014, **28**, 4211–4222.
46 R. H. Rose, S. J. Briddon and S. J. Hill, *Br. J. Pharmacol.*, 2012, **165**, 1789–1800.
47 K. Herrick-Davis, E. Grinde, A. Cowan and J. E. Mazurkiewicz, *Mol. Pharmacol.*, 2013, **84**, 630–642.
48 J. W. Jones, T. A. Greene, C. A. Grygon, B. J. Doranz and M. P. Brown, *J. Biomol. Screening*, 2008, **13**, 424–429.
49 S. Veiksina, S. Kopanchuk and A. Rinken, *Anal. Biochem.*, 2010, **402**, 32–39.
50 K. G. Huwiler, T. De Rosier, B. Hanson and K. W. Vogel, *Assay Drug Dev. Technol.*, 2010, **8**, 356–366.
51 L. Tõntson, S. Kopanchuk and A. Rinken, *Neurochem. Int.*, 2014, **67**, 32–38.
52 O. Nosjean, S. Souchaud, C. Deniau, O. Geneste, N. Cauquil and J. A. Boutin, *J. Biomol. Screening*, 2006, **11**, 949–958.
53 W. J. Lambert, *J. Chromatogr. A*, 1993, **656**, 469–484.
54 E. Grushka, *J. Phys. Chem.*, 1972, **76**, 2586–2693.
55 A. M. Talbert, G. E. Tranter, E. Holmes and P. L. Francis, *Anal. Chem.*, 2002, **74**, 446–452.
56 S. G. Patching, *Biochim. Biophys. Acta, Biomembr.*, 2014, **1838**, 43–55.
57 P. A. Van der Merwe, in *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*, ed. B. C. S. E. Harding, Oxford University Press, 2001, pp. 137–170.
58 B. E. Willcox, G. F. Gao, J. R. Wyer, J. E. Ladbury, J. I. Bell, B. K. Jakobsen and P. A. A. van der Merwe, *Immunity*, 1999, **10**, 357–365.
59 J. A. Maynard, N. C. Lindquist, J. N. Sutherland, A. Lesuffleur, A. E. Warrington, M. Rodriguez and S. H. Oh, *Biotechnol. J.*, 2009, **4**, 1542–1558.
60 P. L. Kastritis and A. M. J. J. Bonvin, *J. R. Soc., Interface*, 2013, **10**, 20120835.
61 J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance and N. J. Skelton, *Protein NMR Spectroscopy, Principles and Practice*, Elsevier Ltd, 2nd edn, 2007.
62 O. Millet, P. Bernadó, J. Garcia, J. Rizo and M. Pons, *FEBS Lett.*, 2002, **516**, 93–96.
63 O. Millet, J. P. Loria, C. D. Kroenke, M. Pons and A. G. Palmer, *J. Am. Chem. Soc.*, 2000, **122**, 2867–2877.
64 J. P. Loria, M. Rance and A. G. Palmer, *J. Am. Chem. Soc.*, 1999, **121**, 2331–2332.
65 J. H. Bollmann, B. Sakmann and J. G. Borst, *Science*, 2000, **289**, 953–957.
66 E. Katoh, J. M. Louis, T. Yamazaki, A. M. Gronenborn, D. A. Torchia and R. Ishima, *Protein Sci.*, 2003, 1376–1385.
67 J. Konvalinka, H. G. Kräusslich and B. Müller, *Virology*, 2015, **479–480**, 403–417.
68 G. Bodenhausen and D. J. Ruben, *Chem. Phys. Lett.*, 1980, **69**, 185–189.
69 G. A. Morris and R. Freeman, *J. Am. Chem. Soc.*, 1979, **233**, 760–762.
70 C. J. Markin and L. Spyracopoulos, *J. Biomol. NMR*, 2012, **54**, 355–376.
71 A. G. Palmer, C. D. Kroenke and J. P. Loria, *Methods Enzymol.*, 2001, **339**, 204–238.
72 S. A. Smith, T. O. Levante, B. H. Meier and R. R. Ernst, *Computer Simulations in Magnetic Resonance. An Object-Oriented Programming Approach*, 1994, vol. 106.
73 L. Zandarashvili, A. Esadze, C. A. Kemme, A. Chattopadhyay, D. Nguyen and J. Iwahara, *J. Phys. Chem. Lett.*, 2016, **7**, 820–824.
74 A. Dallmann and M. Sattler, *Curr. Protoc. Nucleic Acid Chem.*, 2014, **2014**, 7.22.1–7.22.19.
75 K. M. Anderson, A. Esadze, M. Manoharan, R. Brüschweiler, D. G. Gorenstein and J. Iwahara, *J. Am. Chem. Soc.*, 2013, **135**, 3613–3619.

76 F. Lohr, S. G. Mayhew and H. Ruterjans, *J. Am. Chem. Soc.*, 2000, **122**, 9289–9295.
77 L. Zandarashvili, D. Nguyen, K. M. Anderson, M. A. White, D. G. Gorenstein and J. Iwahara, *Biophys. J.*, 2015, **109**, 1026–1037.
78 A. T. Tuukkanen and D. I. Svergun, *FEBS J.*, 2014, **281**, 1974–1987.
79 J. Blobel, P. Bernad, D. I. Svergun, R. Tauler and M. Pons, *J. Am. Chem. Soc.*, 2009, **131**, 4378–4386.
80 M. Cammarata, M. Levantino, F. Schotte, P. A. Anfinrud, F. Ewald, J. Choi, A. Cupane, M. Wulff and H. Ihee, *Nat. Methods*, 2008, **5**, 881–886.
81 M. Levantino, B. A. Yorke, D. C. F. Monteiro, M. Cammarata and A. R. Pearson, *Curr. Opin. Struct. Biol.*, 2015, **35**, 41–48.
82 H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, R. B. Doak, F. R. N. C. Maia, A. V. Martin, I. Schlichting, L. Lomb, N. Coppola, R. L. Shoeman, S. W. Epp, R. Hartmann, D. Rolles, A. Rudenko, L. Foucar, N. Kimmel, G. Weidenspointner, P. Holl, M. Liang, M. Barthelmess, C. Caleman, S. Boutet, M. J. Bogan, J. Krzywinski, C. Bostedt, S. Bajt, L. Gumprecht, B. Rudek, B. Erk, C. Schmidt, A. Hömke, C. Reich, D. Pietschner, L. Strüder, G. Hauser, H. Gorke, J. Ullrich, S. Herrmann, G. Schaller, F. Schopper, H. Soltau, K.-U. Kühnel, M. Messerschmidt, J. D. Bozek, S. P. Hau-Riege, M. Frank, C. Y. Hampton, R. G. Sierra, D. Starodub, G. J. Williams, J. Hajdu, N. Timneanu, M. M. Seibert, J. Andreasson, A. Rocker, O. Jönsson, M. Svenda, S. Stern, K. Nass, R. Andritschke, C.-D. Schröter, F. Krasniqi, M. Bott, K. E. Schmidt, X. Wang, I. Grotjohann, J. M. Holton, T. R. M. Barends, R. Neutze, S. Marchesini, R. Fromme, S. Schorb, D. Rupp, M. Adolph, T. Gorkhover, I. Andersson, H. Hirsemann, G. Potdevin, H. Graafsma, B. Nilsson and J. C. H. Spence, *Nature*, 2011, **470**, 73–77.
83 I. Schlichting, *IUCrJ*, 2015, **2**, 246–255.
84 H. S. Cho, N. Dashdorj, F. Schotte, T. Graber, R. Henning and P. Anfinrud, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 7281–7286.
85 D. Huang and A. Caflisch, *PLoS Comput. Biol.*, 2011, **7**, e1002002.
86 R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis and Y. Shan, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13118–13123.
87 D. Shukla, C. X. Hernández, J. K. Weber and V. S. Pande, *Acc. Chem. Res.*, 2015, **48**, 414–422.
88 G. Bottegoni, W. Rocchia and A. Cavalli, *Methods Mol. Biol.*, 2012, **819**, 169–186.
89 G. Bottegoni, A. Cavalli and M. Recanatini, *J. Chem. Inf. Model.*, 2006, **46**, 852–862.
90 P. S. Shenkin and D. Q. Mcdonald, *J. Comput. Chem.*, 1994, **15**, 899–916.
91 Y. Li, *J. Chem. Inf. Model.*, 2006, **46**, 1742–1750.
92 B. Keller, X. Daura and W. F. Van Gunsteren, *J. Chem. Phys.*, 2010, **132**, 074110.
93 F. Noé, I. Horenko, C. Schütte and J. C. Smith, *J. Chem. Phys.*, 2007, **126**, 155102.
94 F. Haack, K. Fackeldey, S. Röblitz, O. Scharkoi, M. Weber and B. Schmidt, *J. Chem. Phys.*, 2013, **139**, 194110.
95 J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, *J. Chem. Theory Comput.*, 2007, **3**, 2312–2334.
96 R. T. McGibbon and V. S. Pande, *J. Chem. Theory Comput.*, 2013, **9**, 2900–2906.
97 R. D. Malmstrom, C. T. Lee, A. T. Van Wart and R. E. Amaro, *J. Chem. Theory Comput.*, 2014, **10**, 2648–2657.
98 I. Buch, T. Giorgino and G. De Fabritiis, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10184–10189.
99 P. Bisignano, S. Doerr, M. J. Harvey, A. D. Favia, A. Cavalli and G. De Fabritiis, *J. Chem. Inf. Model.*, 2014, **54**, 362–366.
100 D. Silva, G. R. Bowman, A. Sosa-peinado and X. Huang, *PLoS Comput. Biol.*, 2011, **7**, e1002054.
101 S. Gu, D. A. Silva, L. Meng, A. Yue and X. Huang, *PLoS Comput. Biol.*, 2014, **10**, e1003767.
102 N. Plattner and F. Noé, *Nat. Commun.*, 2015, **6**, 7653.
103 S. Decherchi, A. Berteotti, G. Bottegoni, W. Rocchia and A. Cavalli, *Nat. Commun.*, 2015, **2**, 1–10.
104 F. Cérou, A. Guyader, T. Lelièvre and D. Pommier, *J. Chem. Phys.*, 2011, **134**, 054108.
105 I. Teo, C. G. Mayne, K. Schulten and T. Lelièvre, *J. Chem. Theory Comput.*, 2016, **12**, 2983–2989.
106 F. Pietrucci and F. Marinelli, *J. Am. Chem. Soc.*, 2009, **131**, 11811–11818.
107 V. Limongelli, M. Bonomi, L. Marinelli, F. L. Gervasio, A. Cavalli, E. Novellino and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 5411–5416.
108 P. Tiwary, V. Limongelli, M. Salvalaglio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, E386–E391.
109 M. Salvalaglio, P. Tiwary and M. Parrinello, *J. Chem. Theory Comput.*, 2014, **10**, 1420–1425.
110 P. Tiwary and M. Parrinello, *Phys. Rev. Lett.*, 2013, **111**, 1–5.
111 S. Izrailev, S. Stepaniants, B. Isralewitz, D. Kosztin, H. Lu, F. Molnar, W. Wriggers and K. Schulten, *Comput. Mol. Dyn. Challenges, Methods, Ideas SE - 2*, 1999, vol. 4, pp. 39–65.
112 A. M. Capelli and G. Costantino, *J. Chem. Inf. Model.*, 2014, **54**, 3124–3136.
113 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
114 H. J. Li, C. T. Lai, P. Pan, W. Yu, N. Liu, G. R. Bommineni, M. Garcia-Diaz, C. Simmerling and P. J. Tonge, *ACS Chem. Biol.*, 2014, **9**, 986–993.
115 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.
116 C. Bergonzo, A. J. Campbell, R. C. Walker and C. Simmerling, *Int. J. Quantum Chem.*, 2009, **109**, 3781–3790.
117 L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.*, 2006, **426**, 168–175.
118 D. Guo, A. C. Pan, R. O. Dror, T. Mocking, R. Liu, L. Heitman, D. E. Shaw and A. P. IJzerman, *Mol. Pharmacol.*, 2016, **89**, 485–491.
119 Y. Niu, S. Li, D. Pan, H. Liu and X. Yao, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5622–5629.
120 L. Mollica, S. Decherchi, S. R. Zia, R. Gaspari, A. Cavalli and W. Rocchia, *Sci. Rep.*, 2015, **5**, 11539.

121 A. E. Mark, W. F. van Gunsteren and H. J. C. Berendsen, *J. Chem. Phys.*, 1991, **94**, 3808.
122 W. Sinko, Y. Miao, C. A. F. De Oliveira and J. A. McCammon, *J. Phys. Chem. B*, 2013, **117**, 12759–12768.
123 H. Tsujishitaj, I. Moriguchi and S. Hirono, *J. Phys. Chem.*, 1993, **97**, 4416–4420.
124 B. Efron, *Ann. Stat.*, 1979, **7**, 1–26.
125 A. T. MacIas, D. S. Williamson, N. Allen, J. Borgognoni, A. Clay, Z. Daniels, P. Dokurno, M. J. Drysdale, G. L. Francis, C. J. Graham, R. Howes, N. Matassova, J. B. Murray, R. Parsons, T. Shaw, A. E. Surgenor, L. Terry, Y. Wang, M. Wood and A. J. Massey, *J. Med. Chem.*, 2011, **54**, 4034–4041.
126 M. Congreve, S. P. Andrews, A. S. Doré, K. Hollenstein, E. Hurrell, C. J. Langmead, J. S. Mason, I. W. Ng, B. Tehan, A. Zhukov, M. Weir and F. H. Marshall, *J. Med. Chem.*, 2012, **55**, 1898–1903.
127 P. Schmidtke, F. Javier Luque, J. B. Murray and X. Barril, *J. Am. Chem. Soc.*, 2011, **133**, 18903–18910.
128 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.
129 K. Kamata, M. Mitsuya, T. Nishimura, J. I. Eiki and Y. Nagata, *Structure*, 2004, **12**, 429–438.
130 P. Zimmet, K. G. M. M. Alberti and J. Shaw, *Nature*, 2001, **414**, 782–787.
131 R. A. Copeland, *Expert Opin. Drug Discovery*, 2010, **5**, 305–310.
132 J. Arrowsmith, *Nat. Rev. Drug Discovery*, 2011, **10**, 328–329.
133 D. Cook, D. Brown, R. Alexander, R. March, P. Morgan, G. Satterthwaite and M. N. Pangalos, *Nat. Rev. Drug Discovery*, 2014, **13**, 419–431.
134 P. Morgan, P. H. Van Der Graaf, J. Arrowsmith, D. E. Feltner, K. S. Drummond, C. D. Wegner and S. D. A. Street, *Drug Discovery Today*, 2012, **17**, 419–424.
135 G. K. Walkup, Z. You, P. L. Ross, E. K. H. Allen, F. Daryaee, M. R. Hale, J. O. Donnell, D. E. Ehmann, V. J. A. Schuck, E. T. Buurman, A. L. Choy, L. Hajec, K. Murphy-benenato, V. Marone, S. A. Patey, L. A. Grosser, M. Johnstone, S. G. Walker, P. J. Tonge and S. L. Fisher, *Nat. Chem. Biol.*, 2015, **11**, 416–423.
136 E. Leproult, S. Barluenga, D. Moras, J. M. Wurtz and N. Winssinger, *J. Med. Chem.*, 2011, **54**, 1347–1355.
137 Q. Liu, Y. Sabnis, Z. Zhao, T. Zhang, S. J. Buhrlage, L. H. Jones and N. S. Gray, *Chem. Biol.*, 2013, **20**, 146–159.
138 J. Singh, R. C. Petter and A. F. Kluge, *Curr. Opin. Chem. Biol.*, 2010, **14**, 475–480.
139 T. Barf and A. Kaptein, *J. Med. Chem.*, 2012, **55**, 6243–6262.
140 A. S. Kalgutkar and D. K. Dalvie, *Expert Opin. Drug Discovery*, 2012, **7**, 561–581.
141 E. Weerapana, C. Wang, G. M. Simon, F. Richter, S. Khare, M. B. D. Dillon, D. A. Bachovchin, K. Mowen, D. Baker and B. F. Cravatt, *Nature*, 2010, **468**, 790–795.
142 R. M. Miller, V. O. Paavilainen, S. Krishnan, I. M. Serafimova and J. Taunton, *J. Am. Chem. Soc.*, 2013, **135**, 5298–5301.
143 I. M. Serafimova, M. A. Pufall, S. Krishnan, K. Duda, M. S. Cohen, R. L. Maglathlin, J. M. McFarland, R. M. Miller, M. Frödin and J. Taunton, *Nat. Chem. Biol.*, 2012, **8**, 471–476.
144 J. C. Byrd, R. R. Furman, S. E. Coutre, I. W. Flinn, J. A. Burger, K. A. Blum, B. Grant, J. P. Sharman, M. Coleman, W. G. Wierda, J. A. Jones, W. Zhao, N. A. Heerema, A. J. Johnson, J. Sukbuntherng, B. Y. Chang, F. Clow, E. Hedrick, J. J. Buggy, D. F. James and S. O'Brien, *N. Engl. J. Med.*, 2013, **369**, 32.
145 M. L. Wang, S. Rule, P. Martin, A. Goy, R. Auer, B. S. Kahl, W. Jurczak, R. H. Advani, J. E. Romaguera, M. E. Williams, J. C. Barrientos, E. Chmielowska, J. Radford, S. Stilgenbauer, M. Dreyling, W. W. Jedrzejczak, P. Johnson, S. E. Spurgeon, L. Li, L. Zhang, K. Newberry, Z. Ou, N. Cheng, B. Fang, J. McGreivy, F. Clow, J. J. Buggy, B. Y. Chang, D. M. Beaupre, L. A. Kunkel and K. A. Blum, *N. Engl. J. Med.*, 2013, **369**, 507–516.
146 V. Seow, J. Lim, A. J. Cotterell, M. Yau, W. Xu, R. Lohman, W. M. Kok, M. J. Stoermer, M. J. Sweet, R. C. Reid, J. Y. Suen and D. P. Fairlie, *Sci. Rep.*, 2016, **6**, 24575.
147 B. Chen and A. J. Butte, *Clin. Pharmacol. Ther.*, 2016, **99**, 285–297.
148 R. A. Copeland, *Nat. Rev. Drug Discovery*, 2015, **15**, 1–9.
149 K. P. Cusack, Y. Wang, M. Z. Hoemann, J. Marjanovic, R. G. Heym and A. Vasudevan, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 2019–2027.
150 G. S. Van Aller, M. B. Pappalardi, H. M. Ott, E. Diaz, M. Brandt, B. J. Schwartz, W. H. Miller, D. Dhanak, M. T. McCabe, S. K. Verma, C. L. Creasy, P. J. Tummino and R. G. Kruger, *ACS Chem. Biol.*, 2014, **9**, 622–629.
151 R. A. Copeland, *Nat. Chem. Biol.*, 2015, **11**, 451–452.

