JOURNAL
OF
THE ROYAL
SOCIETY

Interface

Flexibility and binding affinity in protein–ligand, protein–protein and multi-component protein interactions: limitations of current computational approaches

Pierre Tuffery and Philippe Derreumaux

References	**This article cites 189 articles, 29 of which can be accessed free** http://rsif.royalsocietypublishing.org/content/9/66/20.full.html#ref-list-1 **Article cited in:** http://rsif.royalsocietypublishing.org/content/9/66/20.full.html#related-urls

Subject collections	Articles on similar topics can be found in the following collections
	
	    bioinformatics (28 articles)
	    chemical physics (10 articles)
	    computational biology (178 articles)

Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

REVIEW

Flexibility and binding affinity in protein−ligand, protein−protein and multi-component protein interactions: limitations of current computational approaches

Pierre Tuffery[1] and Philippe Derreumaux[2,3,*]

[1]*INSERM UMR-S 973, Université Paris Diderot, 35 rue Hélène Brion, 75251 Paris cedex, France*
[2]*Laboratoire de Biochimie Théorique, CNRS, UPR 9080, Université Paris Diderot, Sorbonne Paris Cité, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France*
[3]*Institut Universitaire de France, 103, bd Saint-Michel, 75005 Paris, France*

The recognition process between a protein and a partner represents a significant theoretical challenge. *In silico* structure-based drug design carried out with nothing more than the three-dimensional structure of the protein has led to the introduction of many compounds into clinical trials and numerous drug approvals. Central to guiding the discovery process is to recognize active among non-active compounds. While large-scale computer simulations of compounds taken from a library (virtual screening) or designed de novo are highly desirable in the post-genomic area, many technical problems remain to be adequately addressed. This article presents an overview and discusses the limits of current computational methods for predicting the correct binding pose and accurate binding affinity. It also presents the performances of the most popular algorithms for exploring binary and multi-body protein interactions.

Keywords: flexibility; binding affinity; protein−ligand/protein; interactions; drug design; computational methods

1. INTRODUCTION

Top pharmaceutical companies use both biophysical and computational methods for small ligand screening and drug design. Biophysical methods including nuclear magnetic resonance (NMR), mass spectrometry and fluorescence-based techniques allow the qualitative detection of a small molecule binding to a target and the quantitative determination of physical parameters associated with binding [1]. Drug design methods include structure-based virtual screening, where the three-dimensional protein structure is known [2,3], and ligand/pharmacophore-based virtual screening in the absence of a known receptor structure in order to identify and exploit the spatial configuration of essential features that enable a ligand to interact with a specific receptor [4,5]. Recent years have also seen the emergence of chemogenomics with the aim of understanding the recognition between all possible ligands and the full space of proteins by using traditional ligand-based approaches and biological information on drug targets [6] or by requiring only protein sequence and chemical structure data [7].

In this review, we focus on the current approaches aimed at structure-based virtual screening that circumvent the time, labour and material costs associated with experimental binding essays and have led to success stories for specific targets [8–10]. In a typical virtual screening experiment, ligands can be generated de novo using combinatorial chemistry or taken from a library of chemical compounds. The resulting ensemble of thousands to millions of small molecules are then optimally docked to the target protein and subsequently ranked according to their calculated binding energies [11]. There are two main bottlenecks in this procedure. The first one is related to the propensity to generate native or native-like docking poses. In the first chapter of this review, we report our current understanding of biomolecular recognition and the methods

used for sampling the conformations of proteins and low molecular weight molecules separately. The second bottleneck in virtual screening is related to the correlation between the experimental and predicted binding affinities. Numerous studies predict false-positives or compounds with poor affinities. In the second chapter, particular emphasis is placed on advanced methods for predicting three-dimensional ligand-binding pockets, determining binding energies from physics-based interactions, and sampling the configurational space of protein–ligand complexes. Reviews on the effectiveness of empirical scores and knowledge-based functions to evaluate the interaction between ligands and rigid proteins with a continuum description of solvent can be found elsewhere [12–14].

Finally, the last chapter focuses on the field of binary and multi-component protein interactions. In many cases, the design of ligands must be envisioned in the context of a network of interacting molecules that have well-defined three-dimensional structures in isolation or become folded upon either binding to one partner or polymerization. Aggregation of several proteins constitutes a major societal challenge, as it is connected to human neurodegenerative diseases such as Alzheimer's disease (AD).

2. PROTEIN AND LIGAND PLASTICITY

It is well established that proteins, while adopting well-defined structures in aqueous solution, are in constant motion and display conformational heterogeneity [15,16]. Experimental and theoretical studies also show that proteins can fluctuate between open and closed forms in the absence of ligand [17] and protein domains are dynamic with movements including the hinge bending and the shear motions [18]. Looking at molecular recognition, involving the non-covalent association of ligands (either low molecular weight molecules or proteins) to large macromolecules with high affinity and specificity, two mechanisms have been considered for a long time. In the Fisher's 'lock-and-key' model, the protein displacements are limited to a few residues within the catalytic pocket and thus each partner essentially binds in its lowest free energy conformation [19]. In contrast, the Koshland's 'induced-fit' model posits that the bound protein conformation forms only after interaction with a binding partner [20].

Very recently, based on a large number of NMR observables and the energy landscape theory of protein structure and dynamics, a new molecular recognition paradigm has emerged. The so-called 'conformational selection' model postulates that many protein conformations including the bound state pre-exist, and the binding interaction leads to a Boltzmann population shift, redistributing the conformational states [21]. This concept is rather bad news for drug design and makes the docking exercise harder, because the available protein structures in the absence of small ligands or protein-binding partners are no longer the final targets. In addition, the size of the conformational ensemble for docking will depend on the shape of the free energy landscape and it is possible that very large

amplitude motions occur with virtually no expense in energy so that docking becomes very complicated. The 'conformational selection' paradigm is also challenged by intrinsically disordered proteins (IDPs), which are expected to represent 30 per cent of eukaryotic genome-encoded proteins with wholly or partially unstructured domains [22]. In many cases, IDPs undergo folding, in whole or in part, upon binding to their biological targets. This transition, referred to as 'coupled folding and binding' [23], may shift the minima in the conformational space by introducing new conformations.

There is a wide spectrum of theoretical approaches to tackle the functional motions of proteins ranging from normal mode analysis, which determines small vibration motion around a local minimum by means of the Wilson GF method [24] to molecular dynamics (MD)-based methods [25] that explore the configurational space by solving the Newton equations of motion. All these methods can use an all-atom or a coarser description of the target with explicit or implicit solvent representations. Coarse-grained (CG) models, which make use of beads to represent groups of atoms, reduce the number of degrees of freedom and extend the size of the systems to be studied. For example, one protein can be represented by interacting centres at the C_α atom positions (one-bead model), C_α atom and centroid positions of the side-chains (two-bead model), and three to six bead positions [26–31].

Normal mode analysis calculates with high accuracy the lowest vibrational frequency modes which very often resemble the conformational change between the unbound and bound protein forms [32]. By using a set of 20 proteins that undergo large conformational change upon association (>2 Å C_α RMSD), a single low-frequency normal mode was found that describes well the direction of the observed conformational change only in 35 per cent of the proteins studied starting from the unbound form [33]. Since these collective motions are related to the form and topology of the protein of interest [34], they are invariant to the details of the energy function [35] and can be obtained using elementary representations, opening the study of very large systems, such as the ribosome [36]. A more robust description of protein-collective motions can be obtained using 'consensus' normal modes from a set of related structures [37]. The drawback of normal mode analysis is that sampling is carried out in the vicinity of the starting structure and therefore it neither gives the real amplitude of the motion nor does it provide any information on the thermodynamics and kinetics of the transition.

It is possible to go beyond harmonic motions by various stochastic methods using constraint theory [38], activation–relaxation [39] or path-planning approaches [40,41]. The activation–relaxation technique in internal coordinate space, ARTIST [39] was found to be efficient for exploring conformational space in densely packed environments by successive identifications and crossings of well-defined saddle points connecting minima, i.e. energy minimized structures that are accepted/rejected using the Metropolis criterion. ARTIST is not sensitive to the heights of the barriers and can therefore move

through the configurational space. It lacks, however, a proper thermodynamical basis [39]. Path-planning is a classical problem in robotics. It consists of computing feasible motions for a mechanical system in a workspace cluttered by obstacles. Within this approach, molecules are modelled as articulated mechanisms. Groups of rigidly bonded atoms form the bodies of the mechanism and the articulations between bodies correspond to bond torsions. These torsions are the molecular degrees of freedom. The capabilities of path-planning for exploring protein motion were recently demonstrated for long flexible loops and domains [40,41].

At the other end of the spectrum, the most widely used approach for exploring the functional motions of proteins is all-atom MD simulation with explicit treatment of solvent and ions [25]. These simulations, which typically cover a timescale of 20–200 ns depending on the protein size and the available computer resources, cannot routinely explore large conformational changes occurring in the microsecond–millisecond timescale. Using Anton, a specially built supercomputer, atomic folding characterization of three proteins up to 56 amino acids was however recently reported for 100–1000 μs [42].

Other techniques offer a unique alternative to bridge detailed intermolecular interactions and motions occurring at larger spatial scales and longer timescales. The temperature replica exchange molecular dynamics (T-REMD) simulation consists of running N MD simulations in parallel (or replicas) at N increasing values of temperature [43]. At regular intervals, two MD, at adjacent temperatures, exchange their conformations according to the Metropolis criterion. The rationale underlying this method is that simulating at high temperatures allows the replicas to cross free energy barriers that are trapped at low temperatures. All of these sampled data can be used in the weighted histogram method (WHAM [44]) to obtain the full thermodynamics properties of the system, such as the heat capacity. Because the number of replicas scales with the number of degrees of freedom, all-atom T-REMD in explicit solvent is not routinely performed on large proteins consisting of more than 100 amino acids [45]. To accelerate sampling for large systems, we can resort to T-REMD with CG protein models [46] or use alternative all-atom approaches such as Hamiltonian replica exchange molecular dynamics (H-REMD) or temperature-accelerated molecular dynamics (TAMD). H-REMD uses several related Hamiltonians for different replicas, where only some of the terms of the potential energy function are modified across replicas through scaling parameters [47–49]. On the other hand, TAMD rapidly explores the important regions in the free energy landscape associated with a set of continuous collective variables (CVs). As CVs, we may select for instance hinge bending angles or low-frequency normal modes. By using CVs related to the Cartesian coordinates of the centres of contiguous domains, TAMD applied to the GroEL subunit, a 55-kDa, three-domain protein and the HIV-1 gp120 unit, has led to large-scale conformational change that may be useful in the development of inhibitors and immunogens [50]. It is also possible to perform all-atom explicit solvent metadynamics using a large number of CVs without any knowledge of the bound form, as described in §3, or run discrete (discontinuous) molecular dynamics (DMD) using all-atom [51] or CG [52–54] models. DMD does not require numerical integration of Newton's equations but rather computes and sorts collision times, resulting in an improved computational efficiency.

Finally, for large proteins, it would also be possible to follow a hierarchical procedure consisting of T-REMD simulations with a CG model and multiple short all-atom MD simulations in explicit solvent starting from the predicted lowest energy CG conformations [55].

In principle, the biologically active conformations of any low molecular weight molecule in isolation can be determined by MD-based or stochastic methods. While feasible for a small number of targets, they are too slow for high-throughout screening, and techniques using an ensemble of discrete states [56–58] or sampling dictionaries of rotatable bonds [59–61] are preferable. An example of such a fast technique for large-scale de novo peptide structure prediction is PEP-FOLD [62]. This WEB-server approach takes advantage of the concept of structural alphabet [63], in which protein backbones are described as a series of consecutive fragments of four residues. By predicting a limited series of local conformations along the sequence, the folding problem is turned into an assembly of rigid fragment conformations using a chain growth or greedy method [64,65]. Although the progressive assembly poses numerous questions, particularly for the analytical form of the Van der Waals interactions to be used since steric clashes occur much more frequently in a discrete space than in a continuous space, PEP-FOLD folded 25 peptides of 9–25 amino acids with α-helix, β-strand or random coil character accurately (RMSD of 2.3 Å on the peptide NMR rigid cores) and very rapidly (in a few minutes). This result opens new perspectives for the design of small peptides and even mini-proteins [66] and indicates that discretization is feasible for any ligand types and biomolecular systems, including polyRNA and oligosaccharides if it is accompanied by further consistent non-polarizable [67–74] and polarizable [75,76] force field improvements.

3. PROTEIN–LIGAND INTERACTION

An important and very active issue in protein–ligand recognition, where the ligand is a small molecule, is to predict three-dimensional ligand-binding pockets. Numerous structure-based methods, in particular for *in silico* screening of small compounds, have been developed to detect pockets, clefts or cavities in proteins [77–88]. Starting from the experimental structures, current approaches can be classified as geometric, energetic and with or without any consideration of evolutionary information.

Geometric approaches for locating cavities use either Voronoi diagrams such as CASTp [77] or fpocket [78] or three-dimensional grid-based approaches such as VICE [79], PocketPicker [81] and LigSite [82] that search for grid points that are not situated within the protein and satisfy some conditions. For instance, the scanning procedure in PocketPicker [81] comprises
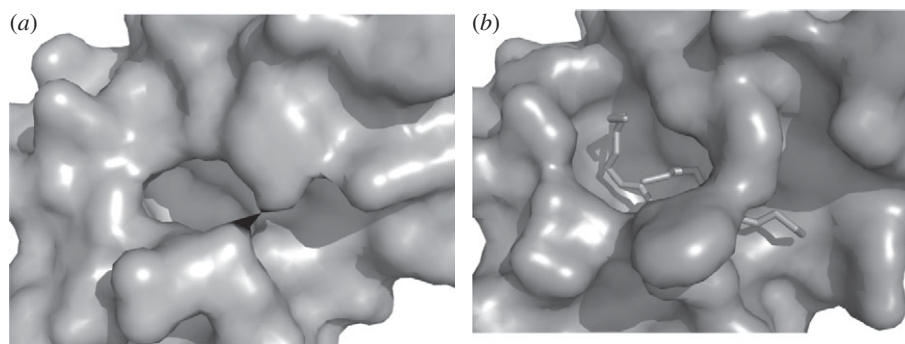
Figure 1. Protein adaptation to ligand binding. (*a*) Conformation of the phospholipase A2 without ligand (PDB 3p2p). (*b*) Conformation bound to an inhibitor (PDB: 5p2p). The protein orientations are strictly identical. Images generated using pymol.

the calculation of 'buriedness' of probe points installed in the grid to determine their atom environment. The accessibility of a grid probe is calculated by scanning the molecular surrounding along 30 search rays of length 10 Å and width 0.9 Å. As a result, the calculated buriedness indices range from 0 to 30 indicating a growing buriedness of the probe in a protein. The clustering of grid probes for pocket identification is restricted to those probes with buriedness indices ranging from 16 to 26. In LigSite [82], the protein is mapped onto a three-dimensional grid. A grid point is part of the protein if it is within 3 Å of an atom coordinate; otherwise it is solvent. Next, the x, y and z-axes plus the four cubic diagonals are scanned for pockets, which are characterized as a sequence of grid points, which start and end with the label protein and a period of solvent grid points in between. These sequences are called protein–solvent–protein (PSP) events. Pockets are then defined using as regions of grid points with a minimum number of PSP events. In practice, a threshold of two PSP events yields good results.

Energetic approaches are based on the calculation of interaction energies between the protein and one or a collection of probes or chemical groups. For instance, Laurie & Jackson [83] use a methyl probe at grid points, while An *et al.* [84] calculate a grid potential map using a carbon atom probe. Brenke *et al.* use an approach based on fast-Fourier techniques to map 16 small probes (ethanol, ispropanol, urea, etc.) to identify the hot spots of interaction, i.e. regions that are likely to bind small drug-like compounds with more affinity than the rest of a pocket [85]. Finally, combining evolutionary sequence conservations and three-dimensional structure-based methods has proven of interest for identifying surface cavities [80], and ConCavity [86] was found to outperform many approaches on a large set of single- and multi-chain protein structures.

The static view used by all previous pocket detection methods meets, however, three major limitations. Firstly, the methods will identify more than one candidate. Although the largest pocket frequently matches the experimental ligand-binding site (e.g. [87]), this rule cannot be generalized, and the question of ranking the pockets in terms of druggability remains to be determined. Secondly, algorithms will usually identify regions larger than the effective interacting area on the protein surface and a finer delimitation of pockets

is important in the context of *in silico* screening. Thirdly, pocket detection from ligand-free protein conformation can be misleading because the bound and unbound protein conformations can differ [88] as seen in figure 1. In addition, the traditional active site pocket concept meets its limits in the emerging field of protein–protein interaction inhibitor design where the notion of cavity becomes fuzzier and the observation of transient pockets at protein–protein interfaces for instance is particularly challenging [89,90]. The pocket-detection approach can be however tackled using simulation approaches. In particular, CG normal modes [91] and Brownian dynamics [92] have been shown to be useful for the prediction of functional sites. Recent approaches also start to track pockets in MD trajectories. The analysis of protein flexibility is expected to provide better understanding of pocket properties.

Along with pocket identification approaches, several docking methods are available to screen and evaluate a library of ligands [93–96]. These algorithms, which can allow full flexibility of the ligands, trade accuracy for CPU time and suffer therefore from three drawbacks.

Firstly, the limited sampling of the thermodynamically accessible protein conformations can be responsible for the failure to find the docked pose. One way of incorporating protein backbone flexibility is to consider a conformational ensemble using experimental data (for instance, different X-ray crystal or NMR-derived structures) and/or simulations (for instance, short MD simulations and normal mode analysis). This pre-generated protein ensemble is not sufficient when the target undergoes extensive rearrangement and only the crystal structure of its apo-structure is available [97]. However, inclusion of multiple protein and ligand conformations may also increase the ability of scoring functions to eliminate false-positives, as highlighted by the recent LigMatch approach [98].

Secondly, analysis of a large number of incorrect ligand–receptor docking poses indicates many other sources of error beyond receptor flexibility. They include improperly assigned histidine tautomers, charged states for aspartate, glutamate and histidine [96] and absence of water molecules in the binding region [99].

Thirdly, calculating binding free energy between a protein and a ligand in quantitative agreement with experiments, one of the most ambitious goals of structure-based drug design, is difficult to reach for high-throughput screening. In a recent study, Kim and Skolnick studied the quality of the BLEEP (knowledge-based), FlexX and X-Score (empirical) and AutoDock scores in various situations, including co-crystallized complex structures, cross-docking of ligands to their non-co-crystallized receptors, docking of thermally unfolded receptor decoys to their ligands and complex structures with 'randomized' ligand decoys. In all cases, the correlation of the raw docking score with the affinity is generally poor [100], confirming earlier studies [101] even in congeneric series [93]. Improving the correlation is a hard task and moves at a slow pace. In 2008, MedusaScore, based on models of physical interactions including van der Waals, solvation and hydrogen-bonding energies, was found to outperform the 11 scoring functions that are widely used in virtual screening for docking decoy recognition and binding affinity prediction [102]. However, the Pearson correlation coefficient between the score and the experimental dissociation constant pKd for the PDBBind database is only 0.63, matching the performance of the RosettaLigand procedure developed in 2006 [103]. Finally, by analysing the performance of six docking programs (DOCK, FlexX, GLIDE, ICM, PhDOCK and Surflex), Cross *et al.* [104] reported general trends in accuracy that are specific for particular protein families, suggesting that expert knowledge is critical for optimizing the accuracy of these methods.

All-atom MD simulations have long been discarded for virtual drug screening because they require a significant effort to calculate accurate residual charges and torsional energy barriers for all possible compounds, although progress has recently been made in this direction for small organic molecules [105]. More importantly, while the binding free energy can be deduced from the probability to find a ligand in interaction or not with its target during an MD simulation, exploring the relevant space of configurations is very time consuming because of the necessity to simulate rare events associated with crossing activation barriers. Yet, significant progresses have been achieved in making MD more tractable for drug design: increase of computer power and design of special machines [106–108] and development of two MD-based families to estimate in order to compute binding free energies: endpoint and pathway methods [109,110].

Special machines, such as Anton [106,107] or MD engines running on graphical processing units [108], have been able to sample the full binding pathways of how a drug finds its target binding site, for example the cancer drug dasatinib with the Src kinase protein within 15 μs [106], an irreversible agonist-β2 adrenoceptor complex within 30 μs [107], and the enzyme–inhibitor complex trypsin–benzamidine using a total time of 49.5 μs [108].

Endpoint free energy methods, such as the Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA) model, have received much attention because they benefit from computational efficiency as only the initial and final states of the system are evaluated. In the MM/PBSA approach, an explicit solvent simulation of the bound state is carried out. Then the simulation is post-processed to determine the enthalpic differences between the bound and unbound solute states. The solvation-free energy is the sum of a polar solvation term using the Poisson model and a non-polar term estimated by solvent-accessible surface area (SASA). The conformational entropy change is usually computed by normal mode analysis.

In a pioneering study, MM/PBSA was used to rank the binding affinities of 12 TIBO-like HIV-1 RT inhibitors. Good agreement between MM/PBSA results and experiments was obtained not only for the relative binding free energies, but also for the absolute ones, which have a root mean square deviation of 1.0 kcal mol^{-1} and a maximum error of 1.9 kcal mol^{-1} [111]. The quality of the agreement depends, however, on the target families studied. For instance, the free energy of binding between avidin and seven biotin analogues led to a mean absolute error of 2.3–4.5 kcal mol^{-1}, arising mainly from the entropy contribution [110]. Enhanced performance to recognize active among non-active compounds may be achieved by combining massive MD simulations of protein–ligand conformations obtained by molecular docking and MM/PBSA. This was confirmed by a recent study in which, trypsin, HIV-1 protease and acetylcholine esterase were each subjected to a 700 ps MD simulation using each of the top-ranking 1000 compounds obtained by docking. This pre-generated MD protein–ligand conformation ensemble was not, however, sufficient for cyclin-dependent kinase-2 [112].

There exist a few faster variants to the MM/PBSA method, such as MM/GBSA, in which the generalized Born (GB) solvent model is used to score the structures. A survey of recent literature reports various degrees of success. On the one hand, the docking poses a diverse set of pharmaceutically relevant targets, including CDK2, Factor Xa, thrombin and HIV-RT were re-evaluated using MM/GBSA and the correlation in all cases between the MM/GBSA results and the −log(IC50) experimental data, where IC50 represents the compound/substance concentration required for 50 per cent inhibition, was satisfactory [113]. Accurate prediction of the relative potencies of members of a series of kinase inhibitors was also reported using molecular docking and MM/GBSA scoring [114]. Overall, these methods are less expensive than those that use an explicit representation of the solvent. However, they can be fully unreliable when interfacial water molecules are present between the ligand and the cavity [99,115], and when the ligand-reorganization free energy is not taken into account. Using the X-linked inhibitor of apoptosis (XIAP) protein and 31 small ligands, major improvement was achieved when the free-energy change for ligands between their free- and bound-states, or ligand-reorganization free energy, was included in the MM/GBSA calculation, with the linear correlation coefficient value between the predicted and the experimentally determined affinities increasing from 0.36 to 0.66 [116].

Finally, it is of interest to compare the performances of the MM/PBSA and MM/GBSA methods. One recent study based on an ensemble of 59 ligands

interacting with six different proteins showed that MM/PBSA gave better correlations than MM/GBSA with experiment, and MM/PBSA performed better in calculating absolute, but not necessarily relative, binding free energies [117]. However, the performances of PBSA are very sensitive to the solute dielectric constant and the characteristics of the interface, while the performances of GBSA depend on the GB model used. Considering its ability to be used in MD simulations, GBSA could serve as a powerful tool in drug design.

The second MD-based family to calculate free energy includes the rigorous, but computationally expensive pathway MD methods [118]. The most widely known is the alchemical double decoupling method, where the interactions of the ligand with the protein and bulk solvent are progressively turned off. Alternatively, it is possible to use a potential of mean force (PMF) method, where the ligand moves along a reaction path from the binding site to the bulk solution [93,119]. In both of these computational approaches, various restraining potentials may be turned on and off to explore more efficiently the translational, rotational and conformational changes of the ligand and protein upon binding, and their effects are removed to yield an unbiased binding free energy. Alchemical decoupling approaches have been shown to provide reliable binding free energies within an accuracy of 1 kcal mol^{-1} if the protein conformational changes upon ligand binding are small, the ligand is uncharged and its solvation free energy is not very large. They are preferable if the ligand binds to buried sites or cavities, and a simple path for ligand association cannot be found. In contrast, PMF-based approaches are preferable if the solvation free energy of the ligand is very large [119]. Interestingly, in a recent blind test, the binding free energies of 50 neutral compounds to the JNK kinase were computed. The free energy computations correctly predicted two of the top five binders, as well as six of the 10 worst binders. The computed binding free energies range from −16 to −3 kcal mol^{-1}, while the experimental values range from −8.6 to −5.5 kcal mol^{-1}. The error in some cases amounts to 7 kcal mol^{-1}, emphasizing again the impact of large protein motion and/or wrong side-chain rotameric states on free energy calculations [119]. For large and flexible protein–ligand system, such as the protein plasmepsin of PM II and several *exo*-3-amino-7-azabicyclo[2,2,1] heptanes, the use of replica exchange-based free energy methods resulted in enhanced convergence, but still failed to reproduce experimental data [120].

Metadynamics or Hill's method is occupying a place of choice in pathway MD methods [121]. Metadynamics, which accelerates the sampling of rare events, maps out the free energy landscape as a function of a small number of CVs. This method is a modification of a standard MD simulation in which restraints are imposed on appropriately selected CVs of the system by a history-dependent potential. The history-dependent potential, by summing up Gaussian functions at regular time intervals along the trajectory, disfavours already visited configurations. Metadynamics was successfully applied to several ligand–protein interactions leading to a better understanding of specific interactions in

molecular recognition and binding affinities [122–125]. However, calculating the free energy projected on CVs is computationally intensive. Clearly, the accuracy of the free energy surface will depend on the force field used, but more importantly on the CVs used to describe the slow degrees of freedom associated with molecular recognition. In particular, it is important to explore the internal degrees of the protein (e.g. loop motions, large-scale motion and possible structured/disordered transitions in specific regions) and the degrees of freedom associated with the in and out motion of the ligand from the protein active site. The choice of these independent CVs is non-trivial and target–ligand-dependent.

Finally, by providing the energy and the position of transition states, the PMF method and metadynamics offer the perspective to extract not only thermodynamic, but also kinetic properties such as the association and dissociation constant rates that impact the efficacy and action time of drugs and have been mostly out of reach of simulations thus far [106–108,126].

4. PROTEIN–PROTEIN AND MULTI-COMPONENT PROTEIN INTERACTIONS

The field of protein–protein interactions has rapidly progressed in the past 10 years [127]. Proteins seldom act alone and most cellular functions are regulated through intricate protein–protein interaction networks. Large efforts have been made to unveil these interactions in a high-throughput manner, with the first interactome maps for several model organisms [128,129]. We can identify binary interactions—that involve only two proteins by multiple methods such as array technology, cross-linking study, cytoplasmic complementation assay, NMR, two hybrid or X-ray crystallography)—and multi-component interactions [130,131]. It is to be noted that if the experimental protein–protein structures are known, there are computer-based tools available to design proteins to bind faster and tighter to their protein-complex partner by electrostatic optimization between the two proteins [132,133].

Protein–protein docking aims to predict the three-dimensional structure of a complex from the knowledge of the structure of the individual proteins in aqueous solution. Many docking methods are now able to predict the three-dimensional structure of binary assemblies if the protein partners do not display important conformational changes between their bound and unbound forms [134–137]. However, large-scale motion upon binding is still a major and unresolved problem in the absence of experimental constraints. This motion generally involves the displacement or the internal rearrangement of loops or domains and can also be characterized by the simultaneous movement of several flexible parts (up to three in known cases). For instance, by using all-atom metadynamics, the transition between the bound and unbound structures of the cyclin-dependent kinase 5 (CDK5) reveals that the large-scale movement has a two-step mechanism: first, the αC-helix rotates by 45°, allowing the

interaction between Glu51 and Arg149; then the CDK5 activation loop refolds to assume the closed conformation [138]. As a result, the binding interface can be completely remodelled from the two unbound forms and it is therefore necessary to take flexibility into account at the beginning of the docking simulations. Collective movements of lower amplitudes have a less dramatic impact on the quality of the predictions but still can bias the results. Clearly, if the structure of the complex with homologous partners is available, this enhances the probability to have native-like conformations in the heap of states.

In the current docking programs, the first stage is aimed at a systematic exploration of the possible geometries of association and is performed either under rigid body approximation [139–142] using all-atom or CG representations or by using a limited sampling of protein backbone conformations. These latter methods involve: (i) the use of CG models [143] with either normal modes [144] or multi-copy approaches where discrete possible loop conformations are taken into account simultaneously [145]; (ii) multi-component docking of flexible domains that are kept rigid using geometric hashing [146]; (iii) Monte Carlo searches of backbone conformations and rigid-body degrees of freedom [147]; and (iv) pre-identification of binding interface and use of more time-consuming search approaches like MD, restricted in space [148].

The second refinement stage of most current docking programs introduces some flexibility by optimizing the side-chain interactions and the rigid body orientations and generates thousands of solution candidates that are ranked using a scoring function. Other programs or web servers introduce, however, backbone flexibility, and not only side-chains and rigid-body orientations. For example, in FiberDock, backbone mobility is modelled by an unlimited number of low and frequency normal modes [149], while in ATTRACT, it is modelled by the first few lowest frequency modes [143]. As reported in §3, precise evaluation of the binding free energy requires highly time-consuming exploration of all the details of the interaction at atomic precision and accurate information on binding affinities is therefore out of reach of all current docking methods.

The performances of state-of-the-art docking strategies producing rigid body solutions were recently evaluated by using a benchmark of 124 interacting pairs for which a high-resolution structure of the complex and the individual components exist [150]. Docking poses for all pairs were generated using FTDock [151] and ZDOCK3 [139], together with one of the most successful docking scoring schemes (pyDock [152]). While ZDOCK3 outperformed FTDOCK, ZDOCK3 obtained an acceptable solution among the top three for only 20 per cent of the cases. This low accuracy poses questions on the usefulness of these programs to study in high-throughput cross-docking manner protein–protein interactions without any knowledge of the interacting pairs [153]. The same conclusion was reached by a more recent study which concluded that there is a poor correlation between measured binding energies and nine commonly used scoring algorithms including pyDock for 81

complexes [154]. There is strong evidence that running docking experiments using homology models of the individual proteins is likely to decrease the success rate [153].

Coupling orientation procedures with flexibility of the protein backbone and the side-chains during the first stage of protein docking is a challenge. It is well known that side-chain rotamers are dependent on the main-chain conformation and side-chain rotamer transitions frequently occur at protein–protein interfaces [155]. Recently, a total of 64 groups and 12 web servers submitted docking predictions for 11 protein complexes and their performances were analysed in Critical Assessment of PRediction of Interactions (CAPRI) 2009 [156]. Overall, the evaluation reveals that eight groups produced high- and medium-accuracy models for six targets with medium conformational changes (1.5–2 Å; RMSD) upon interaction but exploring larger backbone and loop rearrangements, and improving the criteria for selecting promising solutions need to be addressed [137].

Significant progress was also reported by the Symmdock procedure for complexes with Cn symmetry (with a vertical n-fold axis of symmetry) by geometry-based docking [146] and the Rosetta 'fold-and-dock' procedure for symmetrical homodimeric complexes starting from fully extended monomer chains using symmetrical constraints [157]. A multi-scale approach based on Brownian dynamics simulations starting from X-ray crystal structures of the unbound proteins, incorporation of relevant biochemical data followed by all-atom MD simulations, in nine out of 10 cases yielded structures of protein–protein complexes close to those determined experimentally with the percentage of correct contacts > 30% and interface backbone RMSD < 4 Å; [158]. Good results could be obtained for large-scale simulations if multiple protein conformations were pre-stored in a library using various techniques. These include the all-atom activation–relaxation technique in internal coordinate space, ARTIST, with a continuum solvent model [159] and CG models coupled to Brownian dynamics [160] or MD simulations [161] if the protein fold is not too much maintained using the elastic network model.

Clearly, the field of protein interactions goes beyond docking and scoring two entities and three issues pose numerous challenges. Determining whether two given proteins interact is an extremely difficult *in silico* problem [162]. Current docking strategies often give high scores for proteins that do not interact experimentally. Little work has been reported in this cross-docking area, but an important result on protein surfaces is that the optimal orientations of the partners tend to lie within the largest cluster of docking results [163]. More recently, it was found that simply studying the interaction of all potential protein pairs within a dataset can provide significant insights into the identification of the correct interfaces [164]. The second issue lies in the development of algorithms able to dock proteins with well-defined three-dimensional structures beyond binary interactions. This field has attracted recent attention, leading to a combinatorial docking approach [165], the Rosetta 'fold-and-dock' procedure

[157]) and the HADDOCK multi-body docking program up to six molecules using experimental and/or bioinformatics to drive modelling [166].

Finally, and more complex and urgent is the case where random coil proteins aggregate first into soluble oligomers and then to insoluble amyloid fibrils, leading to misfolding diseases such as AD, the most common form of senile dementia. With increased life expectancy and an ageing population, the number of patients with AD is expected to reach 80 million in 2040. The hypothesis for AD causation for which the greatest clinical and experimental support exists is that oligomeric forms of the β-amyloid protein with 39–43 amino acids are the proximate neurotoxic agents [167]. Thus far, no disease-modifying treatments exist and current structural biology methods have failed to reveal three-dimensional target structures of the Aβ oligomers. This is an extremely difficult problem as the low molecular weight aggregates are transient and in dynamic equilibrium between many species ranging from dimers to dodecamers [168].

A number of theoretical studies have recently determined the transmembrane structures of Aβ oligomers [169], and the structures of Aβ oligomers in aqueous solution by using multiple simulation approaches described in §2 [170–176]. To increase our knowledge on the exact mechanism of action for known Aβ drugs, several modelling studies were conducted [177–182]. For instance, all-atom MD/REMD studies looked at the impact of organic molecules, such as naproxen, on Aβ dimers [178] and protofilaments [179] or the impact of N-methylated peptide-based inhibitors [180,181] and the Pittsburg compound on Aβ protofilaments [182].

Other modelling studies along with transmission electron microscopy and spectroscopic circular dichroism (CD) experiments, cell viability essays [183–185] and even *in vivo* experiments [186] helped in the rational design of beta-sheet ligands against Aβ42-induced toxicity. For instance, formation of Aβ42 oligomer in the presence of C-terminal Aβ inhibitors has been studied by the DMD approach starting from spatially separated monomeric mixtures of Aβ42 and inhibitors. It was found that Aβ31–42 and Aβ39–42 are leads for obtaining mechanism-based drugs for treatment of AD using a systematic structure–activity approach [184]. Additional DMD simulations suggest that region Asp1–Arg5, which is more exposed to the solvent in Aβ42 than in Aβ40 oligomers, is involved in mediating Aβ42 oligomer neurotoxicity [185].

Overall, there is increasing evidence that very long CG and all-atom simulations, if coupled to low-resolution experimental data, may soon provide the transient Aβ42/Aβ40 oligomeric structures at an atomic level of detail, a prerequisite for the discovery and optimization of more efficient drugs against AD.

5. CONCLUSIONS

We have reviewed current *in silico* methodologies for predicting the correct binding pose and affinity measures in protein–ligand and protein–protein complexes.

While an increasing number of studies have reported success stories in both areas [157,187,188], the size of the libraries requiring screening and the possibility of several interaction sites per protein still argue in favour of using mostly rigid body approximations and empirical scoring functions in order to select a small number of targets (<20) for experimental validation and further learning [14]. This procedure is not optimal for two reasons. Firstly, the methods cannot filter out all false-positives and expert knowledge is highly desirable to obtain compounds that bind selectively to their target receptors and do not cause side-effects by binding to other systems [104,189]. It is to be noted that false-positives are not available to the scientific community while they would help adjust the current methodologies. Secondly, optimization of the thermodynamic and kinetic properties is out of reach using standard computer resources if the experimental protein–ligand and protein–protein structures are not known.

The treatment of very large conformational changes in the receptor induced by ligand or protein-binding remains one of the biggest challenges in calculations of binding free energies. Finding the relevant rotational, translational and conformational degrees of freedom or CVs for a binary complex is far from being trivial, but this would be achievable for less than 20 candidates with the increase of computer power using a multi-scale approach that moves through different levels of complexity and precision. In the first step, approximate docking pathways could be sampled with rapid methods such as elastic network models, path-planning approaches and short replica exchange MD simulations based on CG representations of the systems. Next, these pathways could be refined and optimized with meta-dynamics or other rigorous techniques by retaining a full atomistic description of the system only in regions of interest while describing the rest of the system with elastic network models.

Much work remains to be done in the field of multiple-component docking of either well-defined structured proteins or proteins that fold, in whole or in part, upon binding or polymerization. All these stimulating challenges will undoubtedly involve extensive research.

REFERENCES

1 Renaud, J. P. & Delsuc, M. A. 2009 Biophysical techniques for ligand screening and drug design. *Curr. Opin. Pharmacol.* **9**, 622–628. (doi:10.1016/j.coph.2009.06.008)

2 Jorgensen, W. L. 2004 The many roles of computation in drug discovery. *Science* **303**, 1813–1818. (doi:10.1126/science.1096361)

3 Morra, G., Genoni, A., Neves, M. A., Merz Jr, K. M. & Colombo, G. 2010 Molecular recognition and drug-lead identification: what can molecular simulations tell us? *Curr. Med. Chem.* **17**, 25–41. (doi:10.2174/092986710789957797)

4 Clark, R. D. 2009 Prospective ligand- and target-based 3D QSAR: state of the art 2008. *Curr. Top. Med. Chem.* **9**, 791–810. (doi:10.2174/156802609789207118)

5 Dror, O., Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. 2009 Novel approach for efficient pharmacophore-based virtual screening: method and

applications. *J. Chem. Inf. Model* **10**, 2333–2343. (doi:10.1021/ci900263d)

6 Strombergsson, H. & Kleywegt, G. J. 2009 A chemogenomics view on protein–ligand spaces. *BMC Bioinform.* **10**(Suppl. 6), S13. (doi:10.1186/1471-2105-10-S6-S13)

7 Nagamine, N., Shirakawa, T., Minato, Y., Torii, K., Kobayashi, H., Imoto, M. & Sakakibara, Y. 2009 Integrating statistical predictions and experimental verifications for enhancing protein–chemical interaction predictions in virtual screening. *PLoS Comput. Biol.* **5**, e1000397. (doi:10.1371/journal.pcbi.1000397)

8 Clark, D. E. 2006 What has computer-aided molecular design ever done for drug discovery? *Exp. Opin. Drug Discov.* **1**, 103–110. (doi:10.1517/17460441.1.2.103)

9 Hardy, L. & Malikayil, A. 2003 The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.* **3**, 15–20.

10 Zoete, V., Grosdidier, A. & Michielin, O. 2009 Docking, virtual high throughput screening and *in silico* fragment-based drug design. *J. Cell. Mol. Med.* **13**, 238–248. (doi:10.1111/j.1582-4934.2008.00665.x)

11 Ghosh, S., Nie, A. & Huang, Z. 2006 Structure-based virtual screening of chemical libraries for drug discovery. *Curr. Opin. Chem. Biol.* **10**, 194–202. (doi:10.1016/j.cbpa.2006.04.002)

12 Guvench, O. & MacKerell Jr, A. D. 2009 Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* **19**, 56–61. (doi:10.1016/j.sbi.2008.11.009)

13 Rajamani, R. & Good, A. C. 2007 Ranking poses in structure-based lead discovery and optimization: current trends in scoring function development. *Curr. Opin. Drug Discov. Dev.* **10**, 308–315.

14 Kellenberger, E., Foata, N. & Rognan, D. 2008 Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J. Chem. Inf. Model* **48**, 1014–1025. (doi:10.1021/ci800023x)

15 Mittermaier, A. & Kay, L. E. 2006 New tools provide new insights in NMR studies of protein dynamics. *Science* **312**, 224–228. (doi:10.1126/science.1124964)

16 Lange, O. F. *et al.* 2008 Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475. (doi:10.1126/science.1157092)

17 Henzler-Wildman, K. A. *et al.* 2007 Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844. (doi:10.1038/nature06410)

18 Teague, S. J. 2003 Implications of protein flexibility for drug discovery. *Nat. Rev. Discov.* **2**, 527–541. (doi:10.1038/nrd1129)

19 Fisher, E. 1894 Einfluss der configuration auf die wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2984–2993.

20 Koshland, D. E. 1958 Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA* **44**, 98–104. (doi:10.1073/pnas.44.2.98)

21 Boehr, D. D., Nussinov, R. & Wright, P. E. 2009 The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796. (doi:10.1038/nchembio.232)

22 Oldfield, C. J., Cheng, Y., Cortese, M. S., Brown, C. J., Uversky, V. N. & Dunker, A. K. 2005 Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989–2000. (doi:10.1021/bi047993o)

23 Wright, P. E. & Dyson, H. J. 2009 Linking binding and folding. *Curr. Opin. Struct. Biol.* **19**, 31–38. (doi:10.1016/j.sbi.2008.12.003)

24 Derreumaux, P., Vergoten, G. & Lagant, P. 1990 A vibrational molecular force field of models compounds with biological interest. 1. Harmonic dynamics of crystalline urea at 123 K. *J. Comput. Chem.* **11**, 560–568. (doi:10.1002/jcc.540110504)

25 Karplus, M. & McCammon, J. A. 2002 Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652. (doi:10.1038/nsb0902-646)

26 Derreumaux, P. 1997 Folding a 20 amino acid αβ peptide with the diffusion process-controlled Monte Carlo method. *J. Chem. Phys.* **107**, 1941–1947. (doi:10.1063/1.474546)

27 Derreumaux, P. 1997 A diffusion process-controlled Monte Carlo method for finding the global energy minimum of a polypeptide chain.1. Formulation and test on a hexadecapeptide. *J. Chem. Phys.* **106**, 5260–5270. (doi:10.1063/1.473525)

28 Derreumaux, P. 2000 Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations. *Phys. Rev. Lett.* **85**, 206–209. (doi:10.1103/PhysRevLett.85.206)

29 Tozzini, V. 2005 Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **15**, 144–150. (doi:10.1016/j.sbi.2005.02.005)

30 Maupetit, J., Tuffery, P. & Derreumaux, P. 2007 A coarse-grained protein force field for folding and structure prediction. *Proteins* **69**, 394–408. (doi:10.1002/prot.21505)

31 Solernou, A. & Fernandez-Recio, J. 2011 pyDockCG: new coarse-grained potential for protein–protein docking. *J. Phys. Chem. B* **115**, 6032–6039. (doi:10.1021/jp112292b)

32 Tama, F. & Sanejouand, Y. H. 2001 Conformational change of proteins arising from normal modes calculations. *Protein Eng.* **14**, 1–6. (doi:10.1093/protein/14.1.1)

33 Dobbins, S. E., Lesk, V. I. & Sternberg, M. J. 2008 Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. *Proc. Natl Acad. Sci. USA* **105**, 10 390–10 395. (doi:10.1073/pnas.0802496105)

34 Nicolay, S. & Sanejouand, Y. H. 2006 Functional modes of proteins are among the most robust. *Phys. Rev. Lett.* **96**, 078104. (doi:10.1103/PhysRevLett.96.078104)

35 Tirion, M. M. 1996 Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* **77**, 1905–1908. (doi:10.1103/PhysRevLett.77.1905)

36 Tama, F., Valle, M., Frank, J. & Brooks III, C. L. 2003 Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. *Proc. Natl Acad. Sci. USA* **100**, 9319–9323. (doi:10.1073/pnas.1632476100)

37 Batista, P. R., Robert, C. H., Maréchal, J.-D., Hamida-Rebai, M. B., Pascutti, P. G., Bisch, P. M. & Perahia, D. 2010 Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis—application to the HIV-1 protease. *Phys. Chem. Chem. Phys.* **12**, 2850–2859. (doi:10.1039/b919148h)

38 Zavodszky, M. I., Lei, M., Thorpe, M. F., Day, A. R. & Kuhn, L. A. 2004 Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins* **57**, 243–261. (doi:10.1002/prot.20179)

39 Yun, M. R., Lavery, R., Mousseau, N., Zakrzewska, K. & Derreumaux, P. 2006 ARTIST: an activated method in internal coordinate space for sampling protein energy landscapes. *Proteins* **63**, 967–975. (doi:10.1002/prot.20938)

40 Cortés, J., Siméon, T., Remaud-Siméon, M. & Tran, V. 2004 Geometric algorithms for the conformational

analysis of long protein loops. *J. Comp. Chem.* **25**, 956–967. (doi:10.1002/jcc.20021)

41 Cortés, J., Siméon, T., Ruiz de Angulo, V., Guieysse, D., Remaud-Siméon, M. & Tran, V. 2005 A path-planning approach for computing large-amplitude motions of flexible molecules. *Bioinformatics* **21**(Suppl. 1), i116–i125. (doi:10.1093/bioinformatics/bti1017)

42 Shaw, D. E. *et al.* 2010 Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346. (doi:10.1126/science.1187409)

43 Sugita, Y. & Okamoto, Y. 1999 Replica exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **329**, 261–270. (doi:10.1016/S0009-2614(00)00999-4)

44 Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. 2004 The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021. (doi:10.1002/jcc.540130812)

45 De Simone, A., Zagari, A. & Derreumaux, P. 2007 Properties of the partially unfolded states of the prion protein. *Biophys. J.* **93**, 1284–1292. (doi:10.1529/biophysj.107.108613)

46 Chebaro, Y., Dong, X., Laghaei, R., Derreumaux, P. & Mousseau, N. 2009 Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent. *J. Phys. Chem. B* **113**, 267–274. (doi:10.1021/jp805309e)

47 Fukunishi, H., Watanabe, O. & Takada, S. 2002 On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: application to protein structure. *J. Chem. Phys.* **116**, 9058–9067. (doi:10.1063/1.1472510)

48 Mu, Y., Yang, Y. & Xu, W. 2007 Hybrid Hamiltonian replica exchange molecular dynamics simulation method employing the Poisson–Boltzmann model. *J. Chem. Phys.* **127**, 084119. (doi:10.1063/1.2772264)

49 Mu, Y. 2009 Dissociation aided and side chain sampling enhanced Hamiltonian replica exchange. *J. Chem. Phys.* **130**, 164107. (doi:10.1063/1.3120483)

50 Abrams, C. F. & Vanden-Eijnden, E. 2010 Large scale conformational sampling of proteins using temperature-accelerated molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **107**, 4961–4966. (doi:10.1073/pnas.0914540107)

51 Ding, F., Tsao, D., Nie, H. & Dokholyan, N. V. 2008 *Ab initio* folding of proteins with all-atom discrete molecular dynamics. *Structure* **16**, 1010–1018. (doi:10.1016/j.str.2008.03.013)

52 Urbanc, B., Borreguero, J. M., Cruz, L. & Stanley, H. E. 2006 *Ab initio* discrete molecular dynamics approach to protein folding and aggregation. *Meth. Enzymol.* **412**, 314–338. (doi:10.1016/S0076-6879(06)12019-4)

53 Sharma, S., Ding, F. & Dokhloyan, N. V. 2008 Probing protein aggregation using discrete molecular dynamics. *Front. Biosci.* **13**, 4795–4804. (doi:10.2741/3039)

54 Yun, S. & Guy, H. R. 2011 Stability tests on known and misfolded structures with discrete and all atom molecular dynamics simulations. *J. Mol. Graph. Model.* **29**, 663–675. (doi:10.1016/j.jmgm.2010.12.002)

55 Samiotakis, A., Homouz, D. & Cheung, M. S. 2010 Multiscale investigation of chemical interference in proteins. *J. Chem. Phys.* **132**, 175 101. (doi:10.1063/1.3404401)

56 Forcellino, F. & Derreumaux, P. 2001 Computer simulations aimed at structure prediction of supersecondary motifs in proteins. *Proteins* **45**, 159–166. (doi:10.1002/prot.1135)

57 Wei, C., Mousseau, N. & Derreumaux, P. 2006 The conformations of the amyloid-beta (21–30) fragment can be described by three families in solution. *J. Chem. Phys* **125**, 084911. (doi:10.1063/1.2337628)

58 Das, R. & Baker, D. 2008 Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382. (doi:10.1146/annurev.biochem.77.062906.171838)

59 Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. 2010 Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584. (doi:10.1021/ci100031x)

60 Bolton, E. E., Kim, S. & Bryant, S. H. 2011 PubChem3D: conformer generation. *J. Cheminform.* **3**, 4. (doi:10.1186/1758-2946-3-4)

61 Miteva, M. A., Guyon, F. & Tufféry, P. 2010 Frog2: efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **38**, W622–W627. (doi:10.1093/nar/gkq325)

62 Maupetit, J., Derreumaux, P. & Tuffery, P. 2009 PEP-FOLD: an online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **37**, W498–W503. (doi:10.1093/nar/gkp323)

63 Camproux, A. C., Gautier, R. & Tuffery, P. 2004 A hidden Markov model derived structural alphabet for proteins. *J. Mol. Biol.* **339**, 591–605. (doi:10.1016/j.jmb.2004.04.005)

64 Tuffery, P., Guyon, F. & Derreumaux, P. 2005 Improved greedy algorithm for protein structure reconstruction. *J. Comput. Chem.* **26**, 506–513. (doi:10.1002/jcc.20181)

65 Tuffery, P. & Derreumaux, P. 2005 Dependency between consecutive local conformations helps assemble protein structures from secondary structures using Go potential and greedy algorithm. *Proteins* **61**, 732–740. (doi:10.1002/prot.20698)

66 Maupetit, J., Derreumaux, P. & Tuffery, P. 2010 A fast method for large-scale de novo peptide and miniprotein structure prediction. *J. Comput. Chem.* **31**, 726–738.

67 Dauchez, M., Lagand, P., Derreumaux, P., Vergoten, G., Sekkal, M. & Sombret, M. 1994 Force field and vibrational spectra of oligosaccharides with different glycosidic linkages. *Spectrochim. Acta Part A—Mol. Biomol. Spectrosc.* **50**, 105–118. (doi:10.1016/0584-8539(94)80118-5)

68 Derreumaux, P., Wilson, K. J., Vergoten, G. & Peticolas, W. L. 1989 Conformational studies of neuroactive ligands. 1. Force field and vibrational spectra of crystalline acetylcholine. *J. Phys. Chem.* **93**, 1338–1350. (doi:10.1021/j100341a033)

69 Lagant, P., Derreumaux, P., Vergoten, G. & Peticolas, W. L. 1991 The use of ultraviolet resonance Raman intensities to test proposed molecular force fields for nucleic acid bases. *J. Comput. Chem.* **12**, 731–741. (doi:10.1002/jcc.540120610)

70 Jorgensen, W. L. & Tirado-Rives, J. 2005 Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl Acad. Sci. USA* **102**, 6665–6670. (doi:10.1073/pnas.0408037102)

71 Yildirim, I., Stern, H. A., Tubbs, J. D., Kennedy, S. D. & Turner, D. H. 2011 Benchmarking AMBER force fields for RNA: comparisons to NMR spectra for single-stranded r(GACC) are improved by revised χ torsions. *J. Phys. Chem. B* **115**, 9261–9270. (doi:10.1021/jp2016006)

72 Denning, E. J., Priyakumar, U. D., Nilsson, L. & Mackerell Jr, A. D. 2011 Impact of 2′-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* **32**, 1929–1943. (doi:10.1002/jcc.21777)

73 Fernandes, C. L., Sachett, L. G., Pol-Fachin, L. & Verli, H. 2010 GROMOS96 43a1 performance in predicting

oligosaccharide conformational ensembles within glyco-
proteins. *Carbohydr. Res.* **345**, 663–671. (doi:10.1016/j.
carres.2009.12.018)

74 Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P.,
Klepeis, J. L., Dror, R. O. & Shaw, D. E. 2010 New ref
for proteins: improved side-chain torsion potentials for
the Amber ff99SB protein force field. *Proteins* **78**,
1950–1958.

75 Ponder, J. W. *et al.* 2010 Current status of the AMOEBA
polarizable force field. *J. Phys. Chem. B* **114**, 2549–2564.
(doi:10.1021/jp910674d)

76 Baker, C. M., Anisimov, V. M. & MacKerell Jr, A. D.
2011 Development of CHARMM polarizable force field
for nucleic acid bases based on the classical Drude oscil-
lator model. *J. Phys. Chem. B* **115**, 580–596. (doi:10.
1021/jp1092338)

77 Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A.,
Turpaz, Y. & Liang, J. 2006 CASTp: computed atlas of
surface topography of proteins with structural and topo-
graphical mapping of functionally annotated resides.
Nucleic Acid Res. **34**, W116–W118. (doi:10.1093/nar/
gkl282)

78 Le Guilloux, V., Schmidtke, P. & Tuffery, P. 2009
Fpocket: an open source platform for ligand pocket
detection. *BMC Bioinform.* **10**, 168. (doi:10.1186/1471-
2105-10-168)

79 Tripathi, A. & Kellogg, G. E. 2010 A novel and efficient
tool for locating and characterizing protein cavities and
binding sites. *Proteins* **78**, 825–842. (doi:10.1002/prot.
22608)

80 Huang, B. & Schroeder, M. 2006 LIGSITEcsc: predicting
ligand binding sites using the Connolly surface and
degree of conservation. *BMC Struct. Biol.* **6**, 19.
(doi:10.1186/1472-6807-6-19)

81 Hendlich, M., Rippmann, F. & Barnickel, G. 1997 LIG-
SITE: automatic and efficient detection of potential
small-molecule binding sites in proteins. *J. Mol. Graph.
Model.* **15**, 359–363. (doi:10.1016/S1093-3263(98)00002-3)

82 Weisel, M., Proschak, E. & Schneider, G. 2007 Pocket-
Picker: analysis of ligand binding-sites with shape
descriptors. *Chem. Cent. J.* **1**, 1–17. (doi:10.1186/1752-
153X-1-7)

83 Laurie, A. & Jackson, R. 2005 Q-SiteFinder: an energy-
based method for the prediction of protein–ligand bind-
ing sites. *Bioinformatics* **21**, 1908–1916. (doi:10.1093/
bioinformatics/bti315)

84 An, J., Totrov, M. & Abagyan, R. 2005 Pocketome via
comprehensive identification and classification of ligand
binding envelopes. *Mol. Cell. Proteom.* **4**, 752–761.
(doi:10.1074/mcp.M400159-MCP200)

85 Brenke, R., Kozakov, D., Beglov, D., Hall, D., Landon,
M. R., Mattos, C. & Vajda, S. 2009 Fragment-based
identification of druggable 'hot spots' of proteins using
Fourier domain correlation techniques. *Bioinformatics*
25, 621–627. (doi:10.1093/bioinformatics/btp036)

86 Capra, J. A., Laskowski, R. A., Thornton, J. M., Singh,
M. & Funkhouser, T. A. 2009 Predicting protein ligand
binding sites by combining evolutionary sequence conser-
vation and 3D structure. *PLoS Comput. Biol.* **5**,
e1000585. (doi:10.1371/journal.pcbi.1000585)

87 Nayal, M. & Honig, B. 2006 On the nature of cavities
on protein surfaces: application to the identification of
drug-binding sites. *Proteins* **6**, 892–906. (doi:10.1002/
prot.20897)

88 Kahraman, A., Morris, R. J., Laskowski, R. A. &
Thornton, J. M. 2007 Shape variation in protein binding
pockets and their ligands. *J. Mol. Biol.* **368**, 283–301.
(doi:10.1016/j.jmb.2007.01.086)

89 Eyrisch, S. & Helms, V. 2007 Transient pockets on
protein surfaces involved in protein–protein interaction.
J. Med. Chem. **50**, 3457–3464. (doi:10.1021/jm070095g)

90 Eyrisch, S. & Helms, V. 2009 What induces pocket
openings on protein surface patches involved in
protein–protein interactions? *J. Comput. Aided Mol.
Des.* **23**, 73–86. (doi:10.1007/s10822-008-9239-y)

91 Ming, D., Cohn, J. D. & Wall, M. E. 2008 Fast dynamics
perturbation analysis for prediction of protein functional
sites. *BMC Struct. Biol.* **8**, 1–11. (doi:10.1186/1472-
6807-8-5)

92 Sacquin-Mora, S. & Lavery, R. 2006 Investigating the
local flexibility of functional residues in hemoproteins.
Biophys. J. **90**, 2706–2717. (doi:10.1529/biophysj.105.
074997)

93 Gilson, M. K. & Zhou, X. 2007 Calculation of protein–
ligand binding affinities. *Annu. Rev. Biophys. Biomol.
Struct.* **36**, 21–42. (doi:10.1146/annurev.biophys.36.
040306.132550)

94 Halperin, I., Ma, B., Wolfson, H. & Nussinov, R. 2002
Principles of docking: an overview of search algorithms
and a guide to scoring functions. *Proteins* **47**, 409–443.
(doi:10.1002/prot.10115)

95 Sousa, S. F. & Fernandez, P. A. 2006 Protein–ligand
docking: current status and future challenges. *Proteins*
65, 15–26. (doi:10.1002/prot.21082)

96 Totrov, M. & Abagyan, R. 2008 Flexible ligand docking
to multiple receptor conformations: a practical alterna-
tive. *Curr. Opin. Struct. Biol.* **18**, 178–184. (doi:10.
1016/j.sbi.2008.01.004)

97 McGovern, S. L. & Shoichet, B. K. 2003 Information
decay in molecular docking screens against holo, apo,
and modeled conformations of enzymes. *J. Med. Chem.*
46, 2895–2907. (doi:10.1021/jm0300330)

98 Kinnings, S. L. & Jackson, R. M. 2009 LigMatch: a mul-
tiple structure-based ligand matching method for 3D
virtual screening. *J. Chem. Inf. Model.* **49**, 2056–2066.
(doi:10.1021/ci900204y)

99 De Beer, S. B., Vermeulen, N. P. & Oostenbrick, C. 2010
The role of water molecules in computational drug
design. *Curr. Top. Med. Chem.* **10**, 55–66. (doi:10.
2174/156802610790232288)

100 Kim, R. & Skolnick, J. 2008 Assessment of programs for
ligand binding affinity prediction. *J. Comput. Chem.* **29**,
1316–1331. (doi:10.1002/jcc.20893)

101 Kroemer, R. T. 2007 Structure-based drug design: dock-
ing and scoring. *Curr. Prot. Pept. Sci.* **8**, 312–328.
(doi:10.2174/138920307781369382)

102 Yin, S., Biedermannova, L., Vondrasek, J. & Dokholyan,
N. D. 2008 MedusaScore: an accurate force field-based
scoring function for virtual drug screening. *J. Chem.
Inf. Model.* **48**, 1656–1662. (doi:10.1021/ci8001167)

103 Meiler, J. & Baker, D. 2006 ROSETTALIGAND:
protein-small molecule docking with full side-chain flexi-
bility. *Proteins* **65**, 538–548. (doi:10.1002/prot.21086)

104 Cross, J. B., Thompson, D. C., Rai, B. K., Baber, J. C.,
Fan, K. Y., Hu, Y. & Humblet, C. 2009 Comparison of
several molecular docking programs: pose prediction
and virtual screening accuracy. *J. Chem. Inf. Model.*
49, 1455–1474. (doi:10.1021/ci900056c)

105 Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O.
2011 SwissParam: a fast force field generation tool for small
organic molecules. *J. Comput. Chem.* **32**, 2359–2368.
(doi:10.1002/jcc.21816)

106 Shan, Y., Kim, E. T., Eastwood, M. P., Dror, R. O., See-
linger, M. A. & Shaw, D. E. 2011 How does a drug
molecule find its target binding site? *J. Am. Chem. Soc.*
133, 9181–9183. (doi:10.1021/ja202726y)

107 Rosenbaum, D. M. *et al.* 2011 Structure and function of an irreversible agonist-β(2) adrenoceptor complex. *Nature* **469**, 236–240. (doi:10.1038/nature09665)

108 Buch, I., Giorgino, T. & De Fabritiis, G. 2011 Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl Acad. Sci. USA* **108**, 10 184–10 189. (doi:10.1073/pnas.1103547108)

109 Mobley, D. L. & Dill, K. A. 2009 Binding of small-molecule ligands to proteins: what you see is not always what you get. *Structure* **17**, 489–498. (doi:10.1016/j.str.2009.02.010)

110 Weis, A., Katebzadeh, K., Söderhjelm, P., Nilsson, I. & Ryde, U. 2006 Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* **49**, 6596–6606. (doi:10.1021/jm0608210)

111 Wang, J., Morin, P., Wang, W. & Kollman, P. A. 2001 Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **123**, 5221–5230. (doi:10.1021/ja003834q)

112 Okimoto, N., Futatsugi, N., Fuji, H., Suenaga, A., Morimoto, G., Yanai, R., Ohno, Y., Narumi, T. & Taiji, M. 2009 High performance drug discovery: computational screening by combining docking and molecular dynamics simulations. *PLoS Comput. Biol.* **5**, e1000528. (doi:10.1371/journal.pcbi.1000528)

113 Guimaraes, C. R. W. & Cardozo, M. 2008 MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model.* **48**, 958–970. (doi:10.1021/ci800004w)

114 Lyne, P. D., Lamb, M. L. & Saeh, J. C. 2006 Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* **49**, 4805–4808. (doi:10.1021/jm060522a)

115 Moitessier, N., Englebienne, P., Lee, D., Lawandi, J. & Corbeil, C. R. 2008 Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.* **153**, S7–S26.

116 Yang, C. Y., Sun, H., Chen, J., Nikolovska-Coleska, Z. & Wang, S. 2009 Importance of ligand reorganization free energy in protein–ligand binding-affinity prediction. *J Am Chem Soc.* **131**, 13709–13721. (doi:10.1021/ja9039373)

117 Hou, T., Wang, J., Li, Y. & Wang, W. 2011 Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **51**, 69–82. (doi:10.1021/ci100275a)

118 Simonson, T., Archontis, G. & Karplus, M. 2002 Free energy simulations come of age: protein–ligand interaction. *Acc. Chem. Res.* **35**, 430–437. (doi:10.1021/ar010030m)

119 Deng, Y. & Roux, B.; 2009 Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B* **113**, 2234–2246. (doi:10.1021/jp807701h)

120 Steiner, D., Oostenbrink, C., Diederich, F., Zürcher, M. & van Gunsteren, W. F. 2011 Calculation of binding free energies of inhibitors to plasmepsin II. *J. Comput. Chem.* **32**, 1801–1812. (doi:10.1002/jcc.21761)

121 Laio, A. & Parrinello, M. 2002 Escaping free energy minima. *Proc. Natl Acad. Sci. USA* **99**, 12 562–12 566. (doi:10.1073/pnas.202427399)

122 Branduardi, D., Gervasio, F. L., Cavalli, A., Recanatini, M. & Parrinello, M. 2005 The role of the peripheral anionic site and cation–pi interactions in the ligand penetration of the human AChE gorge. *J. Am. Chem. Soc.* **127**, 9147–9155. (doi:10.1021/ja0512780)

123 Fiorin, G., Pastore, A., Carloni, P. & Parrinello, M. 2006 Using metadynamics to understand the mechanism of calmodulin/target recognition at atomic detail. *Biophys. J.* **91**, 2768–2777. (doi:10.1529/biophysj.106.086611)

124 Gervasio, F. L., Laio, A. & Parinello, M. 2005 Flexible docking in solution using metadynamics. *J. Am. Chem. Soc.* **127**, 2600–2607. (doi:10.1021/ja0445950)

125 Masetti, M., Cavalli, A., Recanatini, M. & Gervasio, F. L. 2009 Exploring complex protein–ligand recognition mechanisms with coarse metadynamics. *J. Phys. Chem. B* **113**, 4807–4816. (doi:10.1021/jp803936q)

126 Swinney, D. C. 2009 The role of binding kinetics in therapeutically useful drug action. *Curr. Opin. Drug Discov. Dev.* **12**, 31–39.

127 Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Röder, L., Euzenat, J., Rechenmann, F. & Jacq, B. 1999 Grasping at molecular interactions and genetic networks in *Drosophila melanogaster* using FlyNets, an Internet database. *Nucleic Acids Res.* **27**, 89–94. (doi:10.1093/nar/27.1.89)

128 Li, S. *et al.* 2004 A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540–543. (doi:10.1126/science.1091403)

129 Costanzo, M. *et al.* 2010 The genetic landscape of a cell. *Science* **327**, 425–431. (doi:10.1126/science.1180823)

130 Pache, R. A. & Aloy, P. 2008 Incorporating high-throughput proteomics experiments into structural biology pipelines: identification of the low-hanging fruits. *Proteomics* **8**, 1959–1964. (doi:10.1002/pmic.200700966)

131 Simonis, N. *et al.* 2009 Empirically controlled mapping of the *Caenorhabditis elegans* protein–protein interactome network. *Nat Methods* **6**, 47–54. (doi:10.1038/nmeth.1279)

132 Reynolds, K. A., Hanes, M. S., Thomson, J. M., Antczak, A. J., Berger, J. M., Bonomo, R. A., Kirsch, J. F. & Handel, T. M. 2008 Computational redesign of the SHV-1 β-lactamase/beta-lactamase inhibitor protein interface. *J. Mol. Biol.* **382**, 1265–1275. (doi:10.1016/j.jmb.2008.05.051)

133 Kiel, C., Filchtinski, D., Spoerner, M., Schreiber, G., Kalbitzer, H. R. & Herrmann, C. 2009 Improved binding of raf to Ras.GDP is correlated with biological activity. *J. Biol. Chem.* **284**, 31893–31902. (doi:10.1074/jbc.M109.031153)

134 Bonvin, A. 2006 Flexible protein–protein docking. *Curr. Opin. Struct. Biol.* **16**, 194–200. (doi:10.1016/j.sbi.2006.02.002)

135 Lensink, M. F., Méndez, R. & Wodak, S. J. 2007 Docking and scoring protein complexes: CAPRI 3rd edn. *Proteins* **69**, 704–718. (doi:10.1002/prot.21804)

136 Moreira, I. S., Fernandes, P. A. & Ramos, M. J. 2010 Protein–protein docking dealing with the unknown. *J. Comput. Chem.* **31**, 317–342.

137 Lensink, M. F. & Wodak, S. H. J. 2010 Docking and scoring protein interactions: CAPRI 2009. *Proteins* **78**, 3073–3084. (doi:10.1002/prot.22818)

138 Berteotti, A., Cavalli, A., Branduardi, D., Gervasio, F. L., Recanatini, M. & Parrinello, M. 2009 Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. *J. Am. Chem. Soc.* **131**, 244–250. (doi:10.1021/ja806846q)

139 Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R. & Weng, Z. 2007 Integrating statistical pair potentials into protein complex prediction. *Proteins* **69**, 511–520. (doi:10.1002/prot.21502)

140 Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. 2005 PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367. (doi:10.1093/nar/gki481)

141 Heifetz, A., Katchalski-Katzir, E. & Eisenstein, M. 2002 Electrostatics in protein–protein docking. *Protein Sci.* **11**, 571–587. (doi:10.1110/ps.26002)

142 Lyskov, S. & Gray, J. J. 2008 The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.* **36**, W233–W238. (doi:10.1093/nar/gkn216)

143 May, A. & Zacharias, M. 2008 Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking. *Proteins* **70**, 794–809. (doi:10.1002/prot.21579)

144 Zacharias, M. 2003 Protein–protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* **12**, 1271–1282. (doi:10.1110/ps.0239303)

145 Bastard, K., Prévost, C. & Zacharias, M. 2006 Accounting for loop flexibility during protein–protein docking. *Proteins* **62**, 956–969. (doi:10.1002/prot.20770)

146 Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. 2005 Geometry-based flexible and symmetric protein docking. *Proteins* **60**, 224–231. (doi:10.1002/prot.20562)

147 Wang, C., Bradley, P. & Baker, D. 2007 Protein–protein docking with backbone flexibility. *J. Mol. Biol.* **373**, 503–519. (doi:10.1016/j.jmb.2007.07.050)

148 Dominguez, C., Boelens, R. & Bonvin, A. M. J. J. 2003 HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737(38). (doi:10.1021/ja026939x)

149 Mashiach, E., Nussinov, R. & Wolfson, H. J. 2010 Fiber-Dock: a web server for flexible induced-fit backbone refinement in molecular docking. *Nucleic Acids Res.* **38**, W457–W461. (doi:10.1093/nar/gkq373)

150 Hwang, H., Pierce, B., Mintseris, J., Janin, J. & Weng, Z. 2008 Protein–protein docking benchmark version 3.0. *Proteins* **73**, 705–709. (doi:10.1002/prot.22106)

151 Gabb, H. A., Jackson, R. M. & Sternberg, M. J. 1997 Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **272**, 106–120. (doi:10.1006/jmbi.1997.1203)

152 Cheng, T. M., Blundell, T. L. & Fernandez-Recio, J. 2007 Electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins* **68**, 503–515. (doi:10.1002/prot.21419)

153 Mosca, R., Pons, C., Fernández-Recio, J. & Aloy, P. 2009 Pushing structural information into the yeast interactome by high-throughput protein docking experiments. *PLoS Comput. Biol.* **5**, e1000490. (doi:10.1371/journal.pcbi.1000490)

154 Kastritis, P. L. & Bonvin, A. M. J. J. 2010 Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.* **9**, 2216–2225. (doi:10.1021/pr9009854)

155 Guharoy, M., Janin, J. & Robert, C. H. 2010 Side-chain rotamer transitions at protein–protein interfaces. *Proteins* **78**, 3219–3225. (doi:10.1002/prot.22821)

156 Janin, J. 2010 Protein–protein docking tested in blind predictions: the CAPRI experiment. *Mol. Biosyst.* **6**, 2351–2362. (doi:10.1039/c005060c)

157 Das, R., André, I., Shen, Y., Wu, Y., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperski, T. & Baker, D. 2009 Simultaneous prediction of protein folding and docking at high resolution. *Proc. Natl Acad. Sci. USA* **106**, 18 978–18 983. (doi:10.1073/pnas.0904407106)

158 Motiejunas, D., Gabdoulline, R., Wang, T., Feldman-Salit, A., Johann, T., Winn, P. J. & Wade, R. C. 2008 Protein–protein docking by simulating the process of association subject to biochemical constraints. *Proteins* **71**, 1955–1969. (doi:10.1002/prot.21867)

159 Yun, M. R., Mousseau, N. & Derreumaux, P. 2007 Sampling small-scale and large-scale conformational changes in proteins and molecular complexes. *J. Chem. Phys.* **126**, 105101. (doi:10.1063/1.2710270)

160 Francia, F., Malferrari, M., Sacquin-Mora, S. & Venturoli, G. 2009 Charge recombination kinetics and protein dynamics in wild type and carotenoid-less bacterial reaction centers: studies in trehalose glasses. *J. Phys. Chem. B* **113**, 10389–10398. (doi:10.1021/jp902287y)

161 Hall, B. A. & Sansom, M. S. P. 2009 Coarse-grained MD simulations and protein–protein interactions: the Cohesin–Dockerin system. *J. Chem. Theory Comput.* **5**, 2465–2471. (doi:10.1021/ct900140w)

162 Deremble, C. & Lavery, R. 2005 Macromolecular recognition. *Curr. Opin. Struct. Biol.* **15**, 171–175. (doi:10.1016/j.sbi.2005.01.018)

163 Fernandez-Recio, J., Totrov, M. & Abagyan, R. 2004 Identification of protein–protein interaction sites from docking energy landscapes. *J. Mol. Biol.* **335**, 843–865. (doi:10.1016/j.jmb.2003.10.069)

164 Sacquin-Mora, S., Lavery, R. & Carbone, A. 2008 Identification of protein interaction partners and protein–protein interaction sites. *J. Mol. Biol.* **382**, 1276–1289. (doi:10.1016/j.jmb.2008.08.002)

165 Inbar, Y., Benyamini, H., Nussinov, R. & Wolfson, H. J. 2005 Combinatorial docking approach for structure prediction of large proteins and multi-molecular assemblies. *Phys. Biol.* **2**, S156–S165. (doi:10.1088/1478-3975/2/4/S10)

166 Karaka, E., Melquiond, A. S., de Vries, S. J., Kastritis, P. L. & Bonvin, A. M. 2010 Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi-body docking server. *Mol. Cell. Proteom.* **9**, 1784–1794. (doi:10.1074/mcp.M000051-MCP201)

167 Walsh, D. M. & Selkoe, D. J. 2007 Aβ oligomers—a decade of discovery. *J. Neurochem.* **101**, 1172–1184. (doi:10.1111/j.1471-4159.2006.04426.x)

168 Bernstein, S. L. *et al.* 2009 Amyloid-beta protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **1**, 326–331. (doi:10.1038/nchem.247)

169 Strodel, B., Lee, J. W., Whittleston, C. S. & Wales, D. J. 2010 Transmembrane structures for Alzheimer's Aβ(1–42) oligomers. *J. Am. Chem. Soc.* **132**, 13 300–13 312. (doi:10.1021/ja103725c)

170 Melquiond, A., Dong, X., Mousseau, N. & Derreumaux, P. 2008 Role of the region 23–28 in Aβ fibril formation: insights from simulations of the monomers and dimers of Alzheimer's peptides Aβ40 and Aβ42. *Curr. Alzheimer Res.* **5**, 244–250. (doi:10.2174/156720508784533330)

171 Wei, G., Mousseau, N. & Derreumaux, P. 2007 Computational simulations of the early steps of protein aggregation. *Prion* **1**, 3–8. (doi:10.4161/pri.1.1.3969)

172 Chebaro, Y., Mousseau, N. & Derreumaux, P. 2009 Structures and thermodynamics of Alzheimer's amyloid-β Aβ(16–35) monomer and dimer by replica exchange molecular dynamics simulations: implication for full-length Aβ fibrillation. *J. Phys. Chem. B* **113**, 7668–7675. (doi:10.1021/jp900425e)

173 Urbanc, B., Cruz, L., Teplow, D. & Stanley, H. E. 2006 Computer simulations of Alzheimer's amyloid-β protein

and assembly. *Curr. Alzheimer Res.* **3**, 493–505. (doi:10.2174/156720506779025170)

174 Urbanc, B., Betnel, M., Cruz, L., Bitan, G. & Teplow, D. B. 2010 Elucidation of amyloid beta-protein oligomerization mechanisms: discrete molecular dynamics study. *J. Am. Chem. Soc.* **132**, 4266–4280. (doi:10.1021/ja9096303)

175 Kim, S., Takeda, T. & Klimov, D. K. 2010 Mapping conformational ensembles of Aβ oligomers in molecular dynamics simulations. *Biophys. J.* **99**, 1949–1958. (doi:10.1016/j.bpj.2010.07.008)

176 Mitternacht, S., Staneva, I., Härd, T. & Irbäck, A. 2011 Monte carlo study of the formation and conformational properties of dimers of Aβ42 variants. *J. Mol. Biol.* **410**, 357–367. (doi:10.1016/j.jmb.2011.05.014)

177 Convertino, M., Pellarin, R., Catto, M., Carotti, A. & Caflisch, A. 2009 9,10-Anthraquinone hinders β-aggregation: how does a small molecule interfere with A β-peptide amyloid fibrillation? *Protein Sci.* **18**, 792–800.

178 Kim, S., Chang, W. E., Kumar, R. & Klimov, D. K. 2011 Naproxen interferes with the assembly of Aβ oligomers implicated in Alzheimer's disease. *Biophys. J.* **100**, 2024–2032. (doi:10.1016/j.bpj.2011.02.044)

179 Takeda, T., Kumar, R., Raman, E. P. & Klimov, D. K. 2010 Nonsteroidal anti-inflammatory drug naproxen destabilizes Aβ amyloid fibrils: a molecular dynamics investigation. *J. Phys. Chem. B* **114**, 15394–15402. (doi:10.1021/jp107955v)

180 Soto, P., Griffin, M. A. & Shea, J. E. 2007 New insights into the mechanism of Alzheimer amyloid-β fibrillogenesis inhibition by N-methylated peptides. *Biophys. J.* **9**, 3015–3025. (doi:10.1529/biophysj.107.112086)

181 Chebaro, Y. & Derreumaux, P. 2009 Targeting the early steps of Aβ 16–22 protofibril disassembly by N-methylated inhibitors: a numerical study. *Proteins* **75**, 442–452. (doi:10.1002/prot.22254)

182 Wu, C., Bowers, M. T. & Shea, J. E. 2011 On the origin of the stronger binding of PIB over thioflavin T to protofibrils of the Alzheimer amyloid-β peptide: a molecular dynamics study. *Biophys. J.* **100**, 1316–1324. (doi:10.1016/j.bpj.2011.01.058)

183 Hochdörffer, K. *et al.* 2011 Rational design of β-sheet ligands against Aβ 42–induced toxicity. *J. Am. Chem. Soc.* **133**, 4348–4358. (doi:10.1021/ja107675n)

184 Fradinger, E. A. *et al.* 2008 C-terminal peptides coassemble into Aβ42 oligomers and protect neurons against Aβ42-induced neurotoxicity. *Proc. Natl Acad. Sci. USA* **105**, 14175–14180. (doi:10.1073/pnas.0807163105)

185 Urbanc, B., Betnel, M., Cruz, L., Li, H., Fradinger, E. A., Monien, B. H. & Bitan, G. 2011 Structural basis for Aβ(1–42) toxicity inhibition by Aβ C-terminal fragments: discrete molecular dynamics study. *J. Mol. Biol.* **410**, 316–328. (doi:10.1016/j.jmb.2011.05.021)

186 Scherzer-Attali, R. *et al.* 2010 Complete phenotypic recovery of an Alzheimer's disease model by a quinone-tryptophan hybrid aggregation inhibitor. *PLoS ONE* **5**, e11101. (doi:10.1371/journal.pone.0011101)

187 Villoutreix, B. O., Eudes, R. & Miteva, M. A. 2009 Structure-based virtual ligand screening: recent success stories. *Comb. Chem. High Throughput Screen.* **12**, 1000–1016. (doi:10.2174/138620709789824682)

188 Jorgensen, W. L. 2009 Efficient lead discovery and optimization. *Acc. Chem. Res.* **42**, 724–733. (doi:10.1021/ar800236t)

189 Henrich, S., Feierberg, I., Wang, T., Blomberg, N. & Wade, R. C. 2010 Comparative binding energy analysis for binding affinity and target selectivity prediction. *Proteins* **78**, 135–153. (doi:10.1002/prot.22579)