CHAPTER 13

Computational Strategies and Challenges for Targeting Protein–Protein Interactions with Small Molecules

DANIELA GRIMME, DOMINGO GONZÁLEZ-RUIZ AND HOLGER GOHLKE*

Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Universitätsstr. 1, D-40225 Düsseldorf, Germany
*E-mail: gohlke@uni-duesseldorf.de

13.1 Introduction

In biological systems, function and malfunction ultimately originates from interacting rather than isolated molecules. As such, most cellular processes are carried out by multiprotein complexes,[1] and the organization of the ensemble of expressed proteins into functional units results in complex protein interaction networks.[2] Thus, protein–protein interactions (PPI) are a large and important class of potential therapeutic targets.[3,4]

Currently, three classes of protein–protein interaction modulators (PPIM) are known: (i) therapeutic antibodies, which are highly target-specific and stable in human serum.[5] As a downside, antibodies are not cell-permeable and show a lack of oral bioavailability; (ii) peptides derived from protein–protein interfaces (PPIF), which are applied as dimerization and interaction inhibitors.[6–11] However, poor metabolic stability and low bioavailability[9]

RSC Drug Discovery Series No. 23
Physico-Chemical and Computational Approaches to Drug Discovery
Edited by Javier Luque and Xavier Barril

Published by the Royal Society of Chemistry, www.rsc.org

limit their perspective for drug development; (iii) small-molecule PPIM,[12–15] which are less likely to suffer from the above limitations. Although considered almost impossible only a few years ago,[16] a number of recent studies have successfully demonstrated the application of (drug-like) PPI antagonists that bind *directly* to a PPIF.[3,12,17–19] In addition, inhibitors that influence PPI by binding to *allosteric* sites have emerged as promising alternatives.[20,21]

Approaches towards the identification of small-molecule PPIM can be classified into three general categories.[17] In principle, these approaches resemble those applied to find small-molecule inhibitors of "classical" (*e.g.*, enzyme) targets. First, interface-derived peptides may aid as lead structures in guiding the development of peptidomimetics.[8,9] Second, screening of (combinatorial) chemical libraries has proven successful in identifying small-molecule PPIM in a number of cases,[22–29] taking advantage of novel techniques such as cell-based translocation assays,[30] tethering,[31,32] or fragment-discovery approaches.[33–35] Notably, although structural information about the PPI to target was available in some of the cases, most of these screens were initially performed on non-targeted libraries.[17] Third, virtual screening of databases is a viable approach for finding small-molecule PPIM. However, this has been applied successfully only in a few cases to date (Table 13.1).

Recent advances in the understanding of the energetics and dynamics of PPI and methodological developments in the field of structure-based drug design methods may open new avenues to apply virtual screening and rational design approaches for finding small-molecule PPIM. These developments include (i) computational approaches to dissect binding interfaces in terms of energetic contributions of single residues (to identify "hot spot" residues), (ii) prediction of potential binding sites from unbound protein structures, (iii) recognition of allosteric binding sites as alternatives to directly targeting interfaces, (iv) docking approaches that consider protein flexibility and improved descriptions of the solvent influence on electrostatic interactions, and (v) data-driven docking approaches. In this updated and extended review,[36] we will describe and summarize these recent developments with a particular emphasis on their applicability to screen for or design small-molecule PPIM.

13.2 What Makes Protein–Protein Interfaces Difficult to Target?

The challenge of identifying small-molecule PPIM in general and through structure-based drug design methods owes much to the overall characteristics of PPIF. Compared to active sites of enzymes, PPIF are typically flat and devoid of deep binding sites for small molecules.[37,38] In addition, the majority of protein–protein complex interfaces are approximately 1600 ± 400 Å^2 in size,[39–42] with an average of 22 buried amino acid residues per binding partner.[39] This amount of buried surface area upon protein–protein complex formation greatly exceeds the potential binding area of small molecules.[17]

Table 13.1 Small-molecule PPIM identified in prospective studies by molecular modeling and virtual screening.

Biological system	*Methods/results*	*Ref.*
Bcl-x_L/2 and Bak-BH3 domains	1. Target modeled by homology; Virtual Screeing (VS) of MDL/ACD 3D database [193 833 compounds (cmpds.)]; DOCK 3.5; 1000 visually examined; 28 selected; found one with activity.	288
	2. Target modeled by homology; refinement by MD; VS of NCI 3D database (206 876 cmpds.); DOCK 4.01; 500 analyzed; 80 selected; 35 tested; seven inhibit the interaction.	289
	3. Ninety three compounds selected from commercial databases on the basis of their similarity to BH3Is (>80%) and solubility (log $P > 6.0$); ligand placed according to CSP; refinement with TreeDock.	290
	4. Innovative VS strategy with various filters; 45 compounds for biological evaluation; one compound showed inhibitory activity of IC_{50} 2.48 μM ($K_{i\ (calc)}$ 0.38 μM).	291
	5. *In silico* screening of compounds from the free database ZINC;[287] identified 17 sulfonamide derivatives as potential new inhibitors for biological evaluation.	292
CD4/MHC class II	DOCK 3.5; ACD screened (~150 000 cmpds.); 1000 best shape complementarity + 1000 best scored; three visual inspections; 41 compounds selected for cell-adhesion experiment; eight significantly inhibit activity.	293
T-cell response: myelin basic protein	1. NCI-3D DB screened (~150 000 cmpds.); DOCK 4.0.1; 1800 selected; screening of these with MCDOCK; 150 selected; 106 tested for toxicity; 39 screened in IL-2 functional secretion assay; seven "leads" identified.	294
XIAP: caspase-9info	Traditional Chinese medicinal herbs-3D DB (8221 cmpds.); DOCK; top 1000 re-scored with X-score (consensus scoring); 200 selected; 36 tested; five inhibit the interaction; Embelin is the most potent one.	295
Rac GTPase: guanine nucleotide exchange factors	UNITY search for compounds from NCI DB that fit into the binding site. FlexX to dock and rank selected compounds; 100 inspected; 15 further investigated; one inhibits the interaction.	296
PDZ protein interaction domain (PTEN and MAGI3)	From a known protein–peptide complex, scaffolds are derived and a (targeted) virtual library is generated; DOCK used to guide the synthesis.	297
PDZ domain of Dishevelled (Dvl)	Structure-based ligand screening (from the NCI) and NMR spectroscopy for final validation of potential inhibitors; found several inhibitors; the best has a binding affinity similar to the native binding partner, Fz; *in vivo* validation for blocking Wnt signaling.	298
Gp41 of HIV	ComGenex database (20 000 cmpds.); VS with DOCK 3.5; 200 cmpds. selected for visual inspection; 20 selected for biological assays.	299
MDM2-p53	Receptor-based pharmacophore model was developed based on the crystal structure and MD simulations; five novel small-molecule p53-MDM2 inhibitor scaffolds were identified; most potent inhibitor has K_i of 110 ± 30 nM.	300

Many PPIF also consist of non-contiguous regions in the primary protein sequence, which makes it difficult to rationally design binding site mimetics.[17] Furthermore, although smaller interfaces (<2000 Å^2) usually form a single epitope on the surface of each component protein, larger interfaces are generally composed of multiple epitopes.[43] In the latter case, many contacts distributed over a large surface may be required to yield a potent PPI inhibitor, as indicated by a synthetic mimic of a shallow, bowl-shaped protein surface designed by Hamilton and co-workers.[44]

With respect to the amino acid *composition* of PPIF of proteinase–inhibitor or antibody–antigen complexes, no difference was found in terms of polar, nonpolar, and charged proportions compared to the composition of solvent-accessible surface regions in general.[39,40,45] In the case of large interfaces (>5000 Å^2), hydrophobic residues were more abundant, while polar residues were more abundant in small interfaces (<1000 Å^2).[46] More specifically, aromatic amino acids (in particular, tyrosine) were found most frequently among the nonpolar residues,[47] whereas arginine was more abundant than aspartate, glutamate, and lysine in the case of charged amino acids.[39] Thus, given the general lack of differentiation in the amino acid *composition* of PPIF compared to other protein surface regions, finding small-molecule PPIM that preferentially bind to the former does not seem feasible at first glance.

The fact that PPIF were found to be mutually complementary with respect to their electrostatic potentials led to the proposal that long-range electrostatic interactions considerably influence binding to these epitopes.[48] This has been confirmed recently by identifying mutants of the β-lactamase inhibitor protein that lie >7 Å apart from the binding partner, yet show a contribution to binding by improving the overall electrostatic complementarity of the binding partners.[49,50] Likewise, libraries of closely related catalytic antibodies generated by phage display revealed that "second sphere" residues (*i.e.*, outside the active site) contributed favorably to both the binding of a transition state analog and catalytic efficiency.[51] Taking into account these long-range interactions may therefore provide a general means to aid in the design and tuning of binding interactions.

Several recent studies also point to the important role of water-mediated interactions in PPIF.[39,52,53] As such, it was shown that the close-packing found for atoms buried in PPIF can be extended to the majority of all interface atoms if solvent positions are taken into account.[39] Water molecules thus contribute to the close packing of atoms, which insures complementarity between the binding partners. Similarly, by comparison of knowledge-based direct and water-mediated contact potentials, partial solvation was found to be important in stabilizing charged groups in PPIF.[52] In fact, water-mediated polar interactions are as abundant at interfaces as are direct protein–protein hydrogen bonds, although the pattern of hydration varies between interfaces.[53] On the one hand, water molecules that form a ring around "dry" interfaces have been observed; on the other hand, "wet" interfaces may be permeated by water molecules. Overall, these findings indicate that proper accounting for

effects of (de-)solvation and specific water-mediated interactions is necessary for computational approaches towards the design and screening of small-molecule PPIM to be successful.

Another important contribution to the observed complementarity in PPIF arises from the inherent flexibility and plasticity of those regions.[54–58] Conformational variability thereby leads to an ensemble of substates around the native protein structure. The *distribution* of populations of these conformational substates depends on binding or other physical influences.[59,60] In view of small molecule-binding to PPIF, this leads to two implications. First, small-molecule PPIM can "induce" striking conformational changes, leading to grooves or pockets in PPIF that were not apparent in the unbound structure. Structural evidence for such a binding site plasticity has been found for Ro264550 binding to interleukin 2 (IL-2)[3] (Figure 13.1). Given that structural and thermodynamic studies further demonstrated that this portion of IL-2 is inherently flexible, one can expect that ligand binding *captured* a low-energy protein conformation instead of *inducing* a high-energy one.[14,61] In principle, it should be possible to detect these potential binding sites in the unbound state by computational means, which may give a hint as to the druggability of these interfaces. Above all, appropriately taking into account

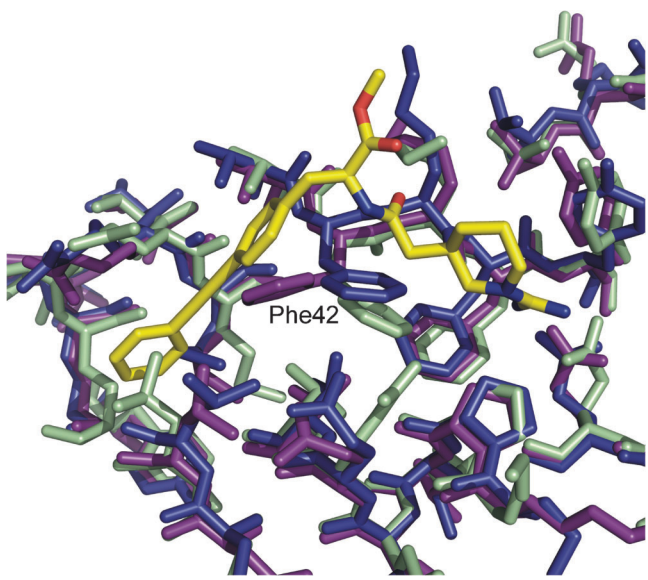

Figure 13.1 Pronounced plasticity observed in the PPI of interleukin 2 (IL2) upon binding of the small-molecule PPIM Ro264550 (*yellow*) (*magenta*: unbound IL2, PDB code 1m47; *gray*: bound IL2, PDB code 1m48). In addition, a conformation extracted from an ensemble of *unbound* IL2 structures is shown (*blue*) that most closely resembles the bound protein structure. The ensemble of unbound IL2 structures was generated by constrained geometric simulation using an enhanced version of FRODA[176] (C. Pfleger, H. Gohlke, unpublished results).

protein flexibility and plasticity in the process of small-molecule PPIM discovery is mandatory. Second, a prevailing distribution of different receptor conformations is also related to the functional adaptability of PPI partners. Studies have shown that residues participating in interactions with multiple binding partners often show a higher degree of flexibility and plasticity.[54,62,63] This implies that also different PPIM should be able to bind to such regions, leading to the notion that diverse molecule libraries need to be screened for potential binders.[64]

Although some of the challenges imposed by the structural and dynamical properties of PPIF also exist in the case of "classical" targets, the above described characteristics led to the view that the development of small molecules PPIM is difficult.[12] However, a more optimistic viewpoint is provided by energetic data on the stability of macromolecular complexes. Analyses of protein–protein complexes *versus* protein–peptide complexes show very similar thermodynamic behavior, despite considerable differences in the interface sizes.[65] Similarly, a "non-linear free energy" relationship with respect to atomic properties has been found recently in an investigation of the stability of macromolecular complexes, resulting in roughly constant binding free energies of the tightest complexes, independent of the interface size.[66] Both observations support an early view[67] that the actual interfaces that contribute to binding ("functional epitopes") have a similar size; these epitopes need to be distinguished from the "structural epitopes" that are given by the overall buried surface areas, which can vary considerably. Along these lines, extensive "alanine scanning" mutagenesis experiments in PPIF revealed that only a small subset of amino acids ("hot spots") within the overall interface contribute significantly to the binding affinity.[38,50,65,68–75] Recent structural analyses of protein–protein surfaces[76,77] further refined this picture such that protein–protein associations are now viewed as locally optimized, with clustered, networked, highly packed, and structurally conserved residues contributing dominantly and cooperatively[50,70,78,79] to the complex stability.[80]

These observations have far-reaching consequences for the discovery of small-molecule PPIM: if only a small number of amino acids within a PPIF provide the majority of the binding energy, it might not be necessary for small molecules to cover an entire PPIF. Instead, mimicking the smaller "functional epitope" should suffice for binding. This notion is not only supported by an increasing number of studies describing a successful development of small-molecule PPIM that bind in PPIF.[13,14,17,18,29,81] The striking observation that phage-display selections of small peptides[82–85] not targeted for protein–protein inhibition nevertheless bind at the protein hot spot indicates that these interface regions are particularly prone to binding, even for small molecules.

13.3 Computational Hot Spot Detection

Given that convergent binding apparently correlates with hot spots, these regions of PPIF should be considered primary targets for virtual screening or

target-oriented combinatorial chemistry.[86] Indeed, small-molecule mimics of hot spots have been found to inhibit PPI, albeit with generally weak affinities.[44,87–89] Although a plethora of experimental alanine scanning data exists for a wide range of PPI,[90,91] it is advisable not to overinterpret the data in terms of specific interactions between residues.[86] This is based on the fact that alanine mutations are perturbations to the free energy surfaces of the unbound state of the protein, the bound state, or both. If it can be assured that the mutation only influences the conformational ensemble of the complex (*i.e.*, it has no effect on the unbound state), measured binding free energy differences between mutant and wild-type protein can be related to specific contact differences. It is for this reason that computational approaches for detecting hot spots become increasingly important.[92] Furthermore, these approaches can be applied to predict hot spots also for (modeled or structurally known) protein–protein complexes for which no experimental mutagenesis data is available. A database, HotSprint, that collects computationally predicted hot spots has been introduced recently.[93]

As for techniques based on first principles, computational alanine scanning has been performed using the MM-PBSA approach[94] to estimate the individual contribution of each residue to the binding.[95] Here, explicit molecular mechanical energies are combined with continuum model-based solvation free energies (see below) and estimates of vibrational entropy changes to probe PPI. These terms are averaged over configurations of the molecular system obtained from high-quality molecular dynamics (MD) simulations. Applied to the "classical" example of experimental alanine scanning, the human growth hormone–receptor complex, the average unsigned error of calculated binding free energy differences obtained by this approach was ~ 1 kcal mol^{-1}.[96] If a proper thermodynamic cycle is employed, a full description of the structural and energetic consequences of a mutation upon the unbound and bound state is obtained.[96] However, this requires repeating the MD sampling of the unbound mutant protein and the mutated complex for each amino acid of interest, which is computationally expensive. Hence, approximations to generate the mutant ensembles have been introduced that only require the simulation of the wild-type proteins, which are then post-processed to introduce either mutations to alanine[95] or larger amino acids.[97] These methods provide a computationally inexpensive way of screening a large variety of possible modifications on either side of the interface. An approach along these lines is presented by Moreira *et al.*[98] To identify hotspot residues in PPIF, the molecular mechanics parmm94 force field and a continuum solvation approach with different internal dielectric constant values, depending on the type of amino acid, is used. After a MD simulation using a modified Generalized Born (GB) solvation model, the post-processing of the complexes, which follows a single-trajectory protocol, permits us to calculate effective energies of the complex and the interacting monomers. Overall, a success rate of 80% in predicting hot spots (binding free energy difference >4.0 kcal mol^{-1}), warm spots (binding free energy difference between 4.0 and 2.0 kcal mol^{-1}), and null spots (binding free energy difference < 2.0 kcal mol^{-1}) was reported for three

protein–protein complexes with 46 alanine mutations.[98,99] Recently, the CONCOORD/Poisson–Boltzmann surface area (CC/PBSA) approach[100] has been introduced, which mimics the MM-PBSA approach in general. However, in contrast to MM-PBSA, a conformational ensemble of protein–protein complex structures is generated using the CONCOORD program[101] rather than MD simulations. As a main advantage, protein flexibility is considered at a much lower computational cost. When tested on alanine mutants for the interface residues of the protein–protein complex Ras–RalGDS, however, a lower predictive power of CC/PBSA compared to MM/PBSA was found (see also below).[102]

As with experimental alanine scanning, the computational mimic inevitably leads to perturbations of the system under consideration.[86,103] In contrast, non-perturbing alternatives to determine the contribution of each residue to the binding free energy are provided by means of component analysis.[104–107] Here, contributions of molecular mechanical energies and solvation free energies are assigned to those atoms that participate in the respective interaction. Summing over atoms of a residue then yields the contribution to the binding free energy. Most importantly, these values are obtained without the need to make structural modifications in the binding partners. It is noted, however, that while the total binding free energy is a state function, free energy components, in general, are not, and are sensitive to the decomposition scheme chosen.[108–110] We pursued a free energy decomposition for the Ras–Raf and Ras–RalGDS protein–protein complexes recently.[104,111] For the first time, decomposition of the solvation free energy contribution was obtained by applying a GB model (see below). Compared to an analogous decomposition based on Poisson continuum electrostatics,[105,107] the GB approach allowed us to ''screen'' all residues of the binding partners at once, drastically lowering computational demand. Convincingly, squared correlation coefficients of 0.55 and 0.46 are found for both systems when comparing the calculated contributions to the binding free energy to experimentally determined binding free energy differences for alanine mutants in the PPIF. Thus, the applied decomposition scheme provides a means by which hot spots in PPIF can be determined rapidly and reliably. In addition, by extending the analysis to all residues of the binding partners, significant contributions to the binding free energy can be identified for single residues as far apart as 25 Å from the interface. This clearly indicates the presence of ''actions-at-a-distance'' in these systems.

Computationally cheaper alternatives to the first principle-based methods have been reported in terms of regression-based scoring functions that allow us to predict binding energy hot spots in PPIF. Here, contributions due to van der Waals and electrostatic energies, hydrogen bonds, water bridges, solvation free energies, and variations of the protein flexibility are combined linearly. The respective weighting factors of the energy terms are parameterized using a dataset of stability changes measured for single mutations in different proteins. In that respect, these functions resemble regression-based scoring functions for

protein–ligand interactions, first reported 17 years ago.[112] Widely used approaches in that respect are FoldX[113,114] and Robetta.[115,116] Encouragingly, although parameterized on alanine scanning data of monomeric proteins only, these functions also perform well if applied to predict alanine scanning results on protein interfaces, resulting in average unsigned errors of 0.9[113] and 1.1[116] kcal mol^{-1} between observed and calculated changes in binding energy. Hence, although not explicitly parameterized on protein interfaces, these functions seem to be general enough to also explain hot spot phenomena. These functions have been used to computationally redesign PPI specificity[117] and to validate homology modeled complexes of Ras and effector proteins.[118] The contributions of the single energy terms may be analyzed in more detail to better understand the thermodynamic characteristics of protein–protein recognition. Two results stand out. First, taking into account the fine details of the structure is crucial. In particular, explicitly modeling hydrogen bond strengths in an environment-dependent fashion considerably enhances the accuracy of hot spot predictions over the sole use of Coulomb electrostatics with a distance-dependent dielectric[116,119] (although it is noted that a more sophisticated treatment of electrostatics[120] may change this view). Second, accounting for (changes of) protein flexibility improved predictions for some complexes,[113,116] although the restricted accounts of flexibility clearly show limitations in more dramatic examples of interface plasticity such as the human growth hormone-receptor interface.[116]

As a knowledge-based approach, DrugScore^{PPI} is a fast and accurate method[102] for calculating relative binding free energies of Ala mutants in PPIF with respect to the wild-type complexes. For DrugScore^{PPI}, statistical pair-potentials have been derived from 851 complex structures and have been adapted against 309 experimental alanine scanning results. Available as a user-friendly webservice, DrugScore^{PPI} offers a fast and accurate prediction of hotspot residues in PPIF. When applied to an external test set of 22 alanine mutations in the interface of Ras–RalGDS, DrugScore^{PPI} significantly outperforms the CC/PBSA, FoldX, and Robetta methods with respect to predictive power and performs as good as the MM/GBSA method, which had been applied to a subset of 16 mutations.[102] Similarly, Tuncbag *et al.*[121] presented an empirical method that determines hot spot residues based on residue conservation, solvent accessibility, and statistical pairwise potentials for interface residues. Adjusted on 150 experimentally determined residues [58 (92) (non-)hot spot residues] and tested on an independent test set of 112 experimentally determined residues [54 (58) (non-)hot spot residues], they observed an accuracy of 70% to match with the experimental hot spot residues. The approach is available as a webserver, HotPoint.[122]

Finally, the machine-learning approach KFC for predicting hot spot residues in protein interfaces has been presented by Darnell *et al.*[123,124] It uses a combination of two physics-based and knowledge-based models characterizing shape specificity features (atomic density, residue size) and biochemical contacts (atomic contacts, hydrogen bonds, salt bridges). KFC

already shows a better predictive power than Robetta, as demonstrated for training and test datasets of 16 protein complexes. Still, the combination of KFC and Robetta's alanine scanning, termed KFCA, results in a statistically significant improvement in the accuracy of hot spot prediction with respect to KFC.

13.4 Predicting Potential Binding Sites in Protein–Protein Interfaces from Unbound Protein States

Although the discovery of interfacial hot spots led to the expectation that small-molecule complements of these regions could attain sufficient binding affinity, in many cases only micromolar inhibitors could be developed.[89] In turn, much more effective small-molecule PPIM have been found to bind to well-defined clefts or grooves in the interface.[3,38,42,89] Methods available for cleft detection are, among others, POCKET,[125] LIGSITE,[126] LIGSITE^{CSC},[127] SURFNET,[128] CAST,[129] PASS,[130] PocketPicker,[131] Fpocket,[132] and PocketAnalyzer (C. Pfleger, T. Jimenez Vaquero, H. Gohlke, unpublished results). Unfortunately, these clefts rarely occur in PPIF. However, in some cases, small molecules that bind to clefts not observed in the unbound protein could be identified experimentally.[3,61,133] This clearly demonstrates that inherent flexibility and plasticity is a hallmark of PPIF. Accordingly, detecting clefts in unbound protein interfaces by computational means will provide valuable starting points for the further rational design of small-molecule PPIM.[134]

As recent studies suggest, biomolecular recognition processes and flexibility (or changes of flexibility) of the binding partners are more fundamentally interrelated than acknowledged by the classical models[135] of "lock and key"[136] or "induced fit".[137] As such, the "conformational selection" model[60,138] proposes that proper conformations are "picked" by the binding from the ensembles of rapidly interconverting conformational species of the unbound molecules. This is supported by experimental evidence for the presence of conformational variability of binding partners prior to their association,[139] and yields an explanation as to why a single protein can bind multiple unrelated ligands at the same site.[64] This model also provides the foundation for computational investigations of conformational fluctuations of the unbound protein state, which may reveal conformational states adopted by the bound proteins (see below).

MD simulations offer the most direct computational approach to address the extent that conformational fluctuations of unbound proteins reflect the conformational changes upon association. While initially only rotamers of key side chains were investigated,[140] a recent analysis of 11 proteins by at least 4 ns long simulations has shown that a few key residues in protein interfaces frequently sample their bound state and may be critical in the early recognition of association.[141] In a more extensive study on 41 proteins for which the three-dimensional structures of bound and unbound state are known,[142] about half

of the short interface segments of unbound proteins were found to sample the bound state during a 5 ns simulation. These findings are striking because even in the absence of the binding partner, certain conformations of *substructural* parts resemble already known bound states. However, in no case in the latter study[142] do the proteins *as a whole* fluctuate closer to the bound state than the unbound state. This points to a limited sampling of adequate conformations due to insufficient simulation times as a primary reason, and possibly to inaccuracies of the underlying energetic descriptions of the systems. Encouragingly, however, for the "classical" target aldose reductase, complexed conformations of the protein could be identified from MD trajectories of the unbound protein state,[143] so predicting conformations of PPIF competent for binding of small-molecule PPIM might be equally feasible.

As a viable alternative to MD simulations, normal mode analysis (NMA)[144–146] and coarse-grained alternatives[147–149] have re-emerged as powerful methods for analyzing the dynamics of biomolecules from a structural perspective.[150] Here, an analytical solution to the equations of motions yields collective variables (normal modes) that describe the dynamics of the system. It is particularly interesting in view of predicting bound protein states from unbound ones in that usually a small subset of low-frequency normal modes (in many cases, even a single mode is sufficient) reliably describes the observed conformational changes.[151] One interpretation is that protein structures have evolved such that biologically relevant motions near the folded state predominantly occur along the directions of lowest-energy modes. Phrased differently, upon going from an unbound to a bound state, proteins most readily explore directions linked to a smooth energy ascent.[149] This provides an explanation as to why bound conformations may already exist in the ensemble of unbound proteins, as proposed by the "conformational selection" model. From a practical point of view, NMA is accordingly applied to identify potential conformational changes of proteins upon binding.[148,152,153] Macromolecular conformations generated through NMA may then be used in docking algorithms that account for protein flexibility.[154,155] We note, however, that due to the harmonic approximation inherent to NMA, transitions from one local minimum to another one are neglected by the method. This becomes particularly important for more localized motions that have been observed in the case of PPIF plasticity.[3,55,156,157] For this case, hybrid MD/NMA techniques[158] may be valuable, combining low-frequency modes from NMA, which describe the collective motions of the protein, with MD, which in turn accounts for more localized motions. That way, large-scale conformational changes of the system are amplified by the modes, while the MD contribution allows to escape the local minimum near the starting structure.

Constrained geometric simulation (CGS) is a computationally very efficient approach to model large-scale conformational changes in proteins. This approach considers all atoms of the macromolecule and is not restricted to exploring only the minimum near the starting structure.[159–161] CGS is based on flexibility concepts[162,163] that allow for the efficient and accurate location of

rigid and flexible regions within a macromolecule from a single, static structure. Here, 3D molecule-like bond networks (where bonds originate from covalent as well as non-covalent interactions in the protein) are analyzed with respect to the bond-rotational degrees of freedom.[164] This concept has been successfully applied to identifying collectively and independently moving regions in a series of proteins[164] or determining the change in protein flexibility upon protein–protein complex formation.[165] Coupled networks of covalent and non-covalent bonds within the protein are then used as input to two computational methods that explore the internal mobility of proteins. Thereby, the coupled bond network in the protein is preserved, and van der Waals overlaps are avoided. In the first method (ROCK),[159,160] correlated motions in flexible protein regions are explored by random-walk sampling of rotatable bonds, thereby leaving rigid regions undisturbed. In the second method (FRODA),[161] rigid protein regions are replaced by so-called ghost templates, which are then used to guide the movements of protein atoms. In both cases, generated protein conformations compare favorably with conformational ensembles determined by NMR.

As an application of CGS to identifying potential binding sites in PPIF, an enhanced version of FRODA was used to generate a conformational ensemble of the *unbound* state of IL-2 within a few hours on a single processor (C. Pfleger, H. Gohlke, unpublished results). IL-2 shows a pronounced interface plasticity upon binding of Ro264550.[3] As depicted in Figure 13.1, a conformation very similar to the one found in the *bound* IL-2 can be identified from these simulations. Unbound IL-2 is thus able to sample bound states even in the absence of the ligand. As this example indicates, these types of simulations may provide an efficient starting point for investigating the conformational variability of PPIF. For a successful prospective prediction of druggable clefts, the simulations need to be combined with screening for energetically accessible protein conformations and a geometrical detection of indentations in the interface region.

CONCOORD[101] is another method to predict protein flexibility based on geometrical considerations. Again, covalent and non-covalent interactions within the structure are translated into a set of geometrical constraints, which provides the starting point for the generation of an ensemble of new conformations. Using principal component analysis, the ''essential'' degrees of freedom of the structure can then be extracted from the ensemble. tCONCOORD[166] is a reimplementation of the CONCOORD method. Recently, Eyrisch *et al.* presented a protocol for identifying transient pockets in PPIF of BCL-X_L, IL-2, and MDM2 by applying the PASS algorithm[130] to MD snapshots.[133] In a second study, Eyrisch *et al.* showed that backbone movements *and* side-chains dynamics are important for transient pocket formation in protein interfaces.[167] When comparing transient pocket detection based on CONCOORD-, tCONCOORD-, and NMA-generated ensembles, only tCONCOORD was able to generate as many and as large pockets as obtained by MD simulations.[167]

13.5 Allosteric Binding Sites as Alternative Targets for Modulating Protein–Protein Interactions

Peptidic and synthetic mimics of protein surfaces and small-molecule PPIM targeting PPIF provide the most direct approach to disrupt PPI. In a more indirect way, allosteric mechanisms have been exploited as promising targets for modulating PPI, especially for cell-surface receptors.[20,38,168] Here, allosteric drugs modulate receptor activity through conformational changes in the receptor protein that are transmitted from the allosteric site to the effector coupling site. More specifically, allosteric modulators enrich certain subsets of conformations available to the protein in the global conformational ensemble[60,138] that differ in their biological binding/signaling properties.[20,169] This reinforces the role of protein dynamics as an entropic carrier of free energy of allostery.[139,170–172]

At least three advantages of allosteric PPIM over "direct" ones have been pointed out:[20] (i) the effect of allosteric modulators is saturable and less prone to overdosing; (ii) allosteric modulators can selectively tune responses only in tissues in which the endogenous agonist exerts its physiological effects; (iii) allosteric modulators have the potential for greater receptor subtype selectivity, based either on a mechanism related to the location of the allosteric sites or different degrees of cooperativity exerted by a modulator at each subtype. In the case of PPI, an additional advantage is that allosteric modulators need not bind at difficult-to-target PPIF regions but can address more pronounced binding sites either between protein subunits or in the interior of a protein.

The discovery of new allosteric sites is a challenging prerequisite to the rational development of allosteric PPIM. Several new allosteric sites have been identified by experimental means,[173] *e.g.* in glycogen phosphorylase,[174] protein tyrosine phosphatase 1B,[175] and HIV-1 reverse transcriptase.[176] The techniques applied include traditional high-throughput screening followed by X-ray crystallography, phage display with crystallography, and tethering.[177]

As an alternative to experiment-led discovery, several computational methods have demonstrated their capability to predict allosteric sites. In evolutionary trace analysis (ETA),[178,179] sequence and structural data are combined to infer the location of functional sites in proteins. Here, members of a protein family are first divided into functional classes based on their sequence identity tree. Then residues that are invariant within each class but vary among them are identified and mapped onto a representative structure. A cluster of these class-specific residues on the protein structure implies an evolutionarily privileged site that is responsible for the functional specificity of the individual family members. The method was applied to the family of regulator of G protein signaling (RGS) proteins,[180] which interact with G protein α subunit (G_α) proteins. A novel functional surface located next to but distinct from the interface between RGS and G_α was identified. Subsequent mutagenesis experiments[181] and crystal structure data[182] confirmed this surface to be the

interaction site for binding of the G protein effector subunit PDEγ. Interestingly, since some of the surface residues had profound effects on the regulation of G_α activity by PDEγ but did not directly interact with G_α, a form of allosteric communication among these residues was inferred.

The ETA exploits evolutionary information about *individual* residues to identify functional/allosteric protein sites. However, a hallmark of allosteric interaction is that it occurs between topographically distinct binding sites; this requires a reliable propagation of signals originating at the allosteric site to the functional one. Hence, if long-range "through-space" interactions can be neglected, the signal flow must proceed *via* a physically connected network of residues that link both sites. The statistical coupling analysis (SCA)[79] detects such coupling between two sites in a protein by analyzing the *co-evolution* of these positions in large and diverse multiple sequence alignments of a protein family. Applied to G-protein coupled receptors, the chymotrypsin class of serine proteases, hemoglobin, guanine-nucleotide-binding proteins, and RXR nuclear receptors, evolutionary conserved sparse networks of amino acid interactions could indeed be identified as structural motifs for allosteric communication in proteins retro- and prospectively.[183–185]

For a successful application of SCA, sub-alignments of members of a protein family of sufficient size and diversity are required so that coupling observed between sites reflects evolutionary constraints during evolution and not just historical relationships.[185] In those cases where the sequence information is insufficient for SCA but structural information about the protein is available, cooperative networks of residues within proteins can be predicted by the COREX approach.[186–188] Here, a large number of different conformational states of a protein are generated through the combinatorial unfolding of a set of predefined folding units. The correlation between the folding states of two residues then indicates the mutual susceptibility of each residue to perturbations at every other site, which in turn reveals the energetic coupling between those residues. Taken one step further, correlations between binding sites as a whole and the rest of the protein can be identified in addition to the pairwise residue correlations. Notably, the analysis of energetic couplings in dihydrofolate reductase revealed that perturbations at one site do not necessarily propagate through structure to the other site *via* a series of conformational distortions. Instead, perturbations exert an influence by affecting the distribution of folded and unfolded states in the ensemble, reinforcing the influence of dynamics on allosteric modulation.[189]

Although the described methods are exciting means to rationalize intramolecular communication, only a few allosteric sites predicted *de novo* have been reported so far.[173] Even if structural protein information is available, locating such sites is hampered by the fact that in many cases some degree of conformational adaptation is required for binding the allosteric modulator. Hence, potential allosteric binding sites may not be readily detectable in the unbound receptor from geometric considerations alone. However, additional structural features might provide a hint to the possibility

of opening up a binding site. This is demonstrated by the rather extreme example of inhibitor binding to the highly packed core region between helices 11 and 12 of β-lactamase (Figure 13.2).[61] Here, unfavorable ϕ/ψ angles of Leu220 located at the N-terminal end of helix 11 suggest conformational strain in the unbound structure, which may be relieved upon complex formation and then counterbalances the energetic cost of core disruption. In addition, a COREX analysis predicts the revealed binding site region to be relatively unstable.[56] Together, these observations suggest that helices 11 and 12, although well-packed, are more prone to an induced-fit adaptation than other core regions. Finally, the most compelling evidence for the existence of the cryptic site comes from the binding of a crystallization agent in the same site of a *homologous* β-lactamase structure.[190] Taken together, perhaps the most promising way to predict new allosteric sites is by, first, obtaining suggestions for potential sites from "crystallization artifacts" or already known ligands binding to related protein structures and, subsequently, confirming these potential allosteric sites computationally.

Finally, "interfacial inhibition" through uncompetitive inhibition has been elucidated recently[191–193] as another natural paradigm for interfering with

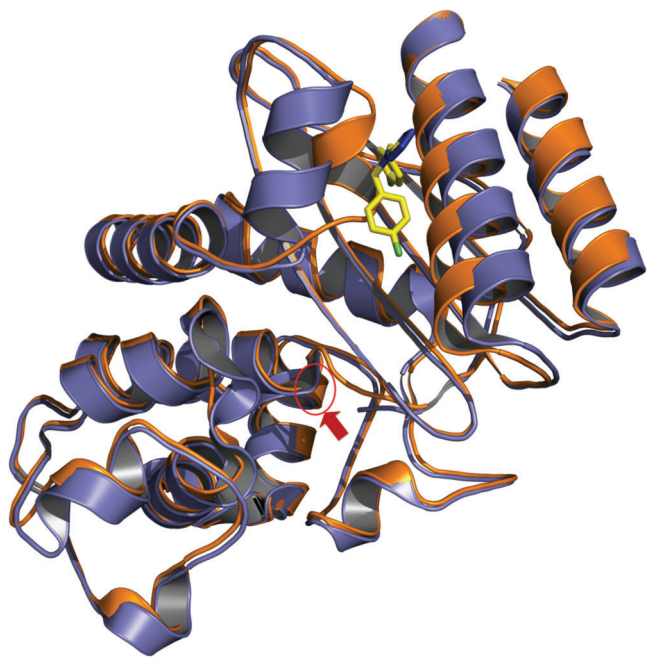

Figure 13.2 Conformational rearrangement of the backbone structure of β-lactamase (*orange*: unbound structure, PDB code 1zg4; *magenta*: bound structure, PDB code 1pzo) upon binding of an allosteric inhibitor (*yellow*). The location of the catalytically active serine is marked by a *red circle*.

macromolecular interactions. Here, targets are captured in dead-end complexes that are unable to complete their biological function. These intermediates display deeply curved surfaces with unbalanced energetic characteristics that are targeted by the inhibitor. However, such conditions are less likely to occur in the unbound proteins or completely bound complexes, and predictions of intermediate complex states as a prerequisite for identifying such sites are generally beyond current computational capabilities.

13.6 Docking for Targeting Protein–Protein Interfaces

Once potential binding sites have been identified, performing docking experiments or virtual screening (VS) is the next step in computer-aided drug development. In particular, this holds for PPI as a new class of targets because no large collections of known ligands may be available to successfully apply ligand-based approaches. Methodologically, current docking approaches face two main difficulties: dealing with solvent effects and protein flexibility. The impact of this situation on targeting PPI along with recent progress is discussed below.

13.6.1 Improved Descriptions of Solvent Effects

In the case of "classical" enzyme targets, a large number of successful applications of docking in VS has been reported.[194] By and large, these successes have been facilitated by steric constraints imposed by well-defined deep cavities that exist in these targets.[195] In such cases, the description of the complex energetic contributions to molecular recognition can be simplified. In fact, neglecting solvent effects on electrostatics did not have a significant effect on the success of some computer-aided drug-design programs.[194,196] However, a proper description of electrostatics is important in the case of PPIF, which are typically flat compared to enzyme targets. The effect of water on electrostatic interactions is twofold: it screens direct charge–charge interactions, and it solvates polar/charged groups.

Dealing thoroughly with solvent effects requires considerable computational resources. However, recent algorithmic developments together with increasing computational power now allow for a much more rigorous treatment of electrostatics. In particular, continuum models, which treat the solvent as being structureless, are now widely applied in computational biophysics.[197] In this approximation, Poisson's equation (PE) rigorously describes the electrostatics of a system consisting of a solute modeled as a distribution of charges in a low dielectric medium immersed in a high dielectric medium (typically water). In general, the PE can only be solved numerically for arbitrarily shaped molecules using, for example, finite difference techniques.[198] In order not to resolve the PE for every newly generated conformation in docking, Arora and Bashford[199] have introduced the Solvation Energy Density Occlusion (SEDO) approach. Here, the system is represented in terms of a solvation energy

density that is pre-computed for receptor and ligands prior to starting the docking simulation. Upon binding, the interacting region of both counterparts changes from a high dielectric medium to a low dielectric one. By neglecting a charge density rearrangement in the remaining high-dielectric region, one then only has to subtract the contribution arising from the newly occluded areas in the complex, which pays off a great gain in efficiency. The methodology was tested on two different data sets: a series of MHC class I protein–peptide complexes, and a congeneric series of HIV-1 protease–ligand complexes. The complexes with the small ligands of the HIV-1 protease yielded slightly better results than the peptides with the MHC class I protein, but all of them were in very good agreement with the results obtained when a non-modified PE approach was followed.

When a docking simulation runs, every new ligand conformation has to be evaluated. Practically, the scoring function complexity and implementation must be efficient. In this regard, the SEED approach by Majeux *et al.*[200] introduces an appealing treatment of the electrostatic contribution to the total binding free energy. Two important approximations are made. First, a simple distance-dependent dielectric model for the screened ligand–receptor interaction is used. Second, for both receptor and ligand, the main contribution to desolvation is considered to come from the removal of the first shell of water.[201] Totrov[202] has estimated that the first solvent layer contributes 66% to the total desolvation energy. Considering this, the receptor and ligand molecules are independently mapped onto a grid, and the corresponding desolvation is pre-computed at the centers of low dielectric probe-spheres rolled over the solvent accessible surface. This computation is done according to the Coulomb approximation of the electric displacement. During docking, only occluded areas in both counterparts have to be detected and summed to assess the total contribution. The approach was validated against solutions of the PE, showing very good correlations for every single contribution and the total electrostatic energy. The oncoprotein MDM2 was targeted to investigate the virtues of this approach. MDM2 binds to the p53 tumor suppressor, keeping it inactive.[203] 1,4-Benzodiazepines and dibenzocyclohexane were found computationally to yield the best binding energies. However, to our knowledge, these findings have not been validated experimentally.

The GB model is an approximation to the PE approach.[204] Here, space is also divided in regions of high (solvent) and low (solute) dielectric, but the reaction field energy is approximated by a pairwise sum over interacting charges. From the original Born theory, the electrostatic contribution to the free energy of solvation of a point charge q located in the center of a spherical cavity of radius R (Born radius) is given by eqn (13.1):

$$\Delta G_{\mathrm{pol}} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\frac{q^2}{R} \tag{13.1}$$

where ε is the dielectric permittivity of the medium.

The term ''generalized'' comes into play when considering more than two point charges and arbitrarily shaped cavities instead of spherical ones. Then, a smooth function that considers charge–charge interactions according to their location in the solute is required. To date, the formulation proposed by Still *et al.*[204] is the most widely used one to estimate solvation energies in this situation. An important parameter in the formulation is the effective Born radius, which measures the burial of an atom in the low-dielectric medium and depends on the atom's intrinsic Born radius and the arrangement of the rest of the atoms in the system. The way of estimating R has implications not only in terms of accuracy, but also in terms of efficiency. The different flavors of GB currently in use arise mainly from the way of defining and computing this parameter. The accuracy and efficiency of a variety of GB models for computing electrostatic solvation energies in comparison with the more rigorous PE model has been assessed by Feig *et al.*[205] The latest GB models yield results comparable to PE, although efficiency is still a concern for those models. A more in-depth review on GB models is available.[206]

Zou *et al.*[207] incorporated a GB model into DOCK.[208] In a further development of the ideas in this work, Liu *et al.*[209] have recently incorporated the more efficient, yet arguably less accurate, pairwise approximation proposed by Hawkins *et al.*[210] that takes into account atomic overlaps to compute Born radii. Despite the limited data set of only three systems used for the evaluation, acceptable results are obtained for the Born radii, considering the gain in efficiency ($\sim$8% of error for the receptor and 4% for the ligand). More interestingly, the GB enhanced scoring scheme performs better than the standard ''force field'' scoring function in DOCK in a VS experiment. This has, in fact, an important implication as these kinds of computational screenings are now commonplace in the early stages of current drug-design programs and suffer from high rates of false positives. Additionally, this re-implementation of the GB model into DOCK introduces a correction to properly treat possible void formations between the protein and ''misaligned'' ligands. This is one of the most problematic aspects of GB models: the boundary definition by means of spheres between solvent and solute may result in solvent-inaccessible, yet high-dielectric, voids in the interior of large biomolecules.[206] Interestingly, successful VS studies implementing these improvements in DOCK have been reported for glyceraldehyde-3-phosphate dehydrogenase,[211] lysosomal cysteine proteases,[212] and a PDZ protein interaction domain[213] as targets.

Aside from GB, other implicit solvation models have been adapted in some popular docking programs as a compromise between the required accuracy and the affordable computational effort.[214] In that respect, a particularly efficient implicit solvation model for computing the electrostatic part of the binding free energy in protein–ligand docking has been introduced recently.[215] With a similar performance in accuracy as GB, the mean pose calculation time by this model amounts to about 40 ms. On the other hand, when facing rather large VS experiments, it is now commonplace to follow a hierarchical

approach[216] where candidate compounds are pre-screened and selected with simpler scoring functions. This reduced database is subsequently re-ranked with more accurate approaches such as MM-PBSA.[217] Very encouragingly, by employing a modified protocol which includes using a GB model during the MD simulations, a very efficient PE solver, and a computational design based on a distributed-computing paradigm, a high-throughput variant of MM-PBSA has been introduced.[218] With this variant, more than 300 compounds were evaluated against three different targets overnight, using approximately 400 desktop computers.[219] As for the accuracy, statistically significant correlations to experimental data were obtained, with correlation coefficients >0.72.

13.6.2 Protein Flexibility in Protein–Ligand Docking

Proteins are inherently flexible, which provides the origin for their plasticity and enables them to conformationally adapt to a binding partner. However, current docking-based drug-design approaches generally treat the target protein as a rigid unit. Following this approximation, better results in VS experiments have been obtained when target structures extracted from complexes were used compared to "apo" structures.[220] On the other hand, "holo" structures appear to introduce a bias in the experiment which, in many cases, precludes finding chemically novel ligands.[221] If the adaptability of binding sites in enzymes is a real concern, taking into account protein flexibility appears even more critical in the case of PPIF, as they have been proven to be highly adaptive,[14] as discussed above.

From a practical point of view, challenges of incorporating protein flexibility into docking are twofold: first, one needs to detect what is flexible,[222] and second, this knowledge needs to be transformed into the docking algorithm. With respect to the former, options range from using experimental information to predicting the most relevant movements through computational methods such as MD, NMA, or a graph-theoretical approach.[164] The latter issue is in general far more open to creative approaches but must be guided by the unavoidable concern about efficiency.

13.6.2.1 Determining What is Moving and How

Protein flexibility comprises a range of possible movements, from single side-chains to drastic structural rearrangements as seen in calmodulin.[223] Depending on where the binding site is located, one or more of these movement types will be relevant for docking. Interestingly, a recent study by Zavodszky and Kuhn[224] assessed, for a large set of typical enzymatic targets, to what extent side-chain rotations of amino acids contribute to the flexibility within the binding site. Ligands from 63 different complexes comprising a total of 20 different enzymes were re-docked into the corresponding *apo* structure using their docking program SLIDE[225] that allows for protein side-chain

rotations. These side-chain rotations were proven to be necessary to correctly dock 54% of the ligands, but encouragingly, 95% of the rotations were smaller than 45°. The plausibility of every adaptation proposed by SLIDE was then evaluated by comparing the free with the resulting bound structures and by analyzing the geometry with PROCHECK.[226] Only 7% of the conformations were evaluated as unfavorable. Previously, Najmanovich *et al.*[227] had shown that there is no correlation between backbone and side-chain flexibility. They reported as well that rotations in side chains of up to three residues account for ~85% of all the cases where there is a conformational change upon ligand binding. For the case of PPI, it has been shown that the conformational change can be even more important for those residues involved in the interface.[228] Having the whole dynamic picture of a protein in hand would be the optimal situation to deal with the protein's flexibility. Although this is not the case for most of the interesting targets, there is hope that one can predict stable conformers even from the unbound structure due to the fact that the bound conformation is likely to be a pre-existing one in the free state[229] (see Section 13.4).

In summary, it is encouraging to see that with little effort much can be gained: many changes will be related only to side chains. Even more, our understanding of protein dynamics has enabled unbound structures as useful starting points for flexible docking. However, the challenge remains in considering every movement that occurs in the binding, independent of its range. In what follows, three approaches to predict flexible regions in proteins are described: automated conformation exploration restrained by experimental knowledge, derived from MD simulations, and through NMA.

Cavasotto and Abagyan[155] incorporated a receptor-ensemble docking approach in the frame of the IFREDA (ICM-flexible receptor docking algorithm) method for VS. The first step of this involves the *de novo* generation of alternative receptor conformations. Four protein kinases subfamilies were investigated. Forty random configurations of a ligand were generated in the binding site. The ligand, side chains in the binding site, and pre-selected loops known to undergo rearrangements are considered fully flexible. An *in vacuo* minimization is performed followed by a stochastic energy minimization and a final full minimization of the top ranked conformations. In this way, the area considered as flexible is enlarged, which yields plausible receptor conformations that are relevant for the binding site. Yet, the requirement for an exhaustive sampling is avoided. However, a drawback of the method is that flexible protein regions must be known in advance.

Zacharias[230] has proposed to use MD to study the movements of the protein before docking. From the MD one can calculate soft flexible degrees of freedom *via* principal component analysis; these soft modes can be afterwards incorporated into a flexible docking algorithm as additional variables to guide the movement of the protein. As only C_α atoms are considered, side-chain movements are not represented by the precalculated modes. However, the method has still proven useful in VS for pre-filtering databases because it is not

very computationally demanding. The author was also able to achieve successful docking results starting from an unbound protein structure where the ligand did not fit sterically. The method has the advantage that the explored protein conformations are fairly realistic, including solvent and counterion effects. In an attempt to also consider side-chain flexibility, Tatsumi combined MD with harmonic dynamics.[231] Collective movements are incorporated into the motion of C_α atoms by means of harmonic modes, whereas the motions of all other atoms are simulated by unbiased MD. This method is theoretically appealing, although inefficient, as one single docking takes 40 days of CPU time (using up to 16 CPUs in parallel).

NMA is a convenient and widespread method to study the dynamics of macromolecules in cases where only one structure is available (see above). Information about low-frequency normal modes are increasingly incorporated into docking procedures.[155,230] The aforementioned IFREDA, which needs information about side chains and backbone movements, was extended in that respect.[155] From NMA, ''relevant modes'' are selected, and protein structures are subsequently modified following them. The ''relevant modes'' are a means of focusing the general analysis to the binding site region, such that those modes that influence binding site atoms are selected. Although these are low-frequency modes, they do not correspond exactly with the lowest ones, so that intermediate-scale loop motions that might occur around the binding site are also captured. The methodology was validated by docking ligands into *apo* structures, where rigid-receptor docking had failed. Also, a small-scale VS showed larger enrichment factors than when performed with the rigid receptors.

Recently, we have developed a multi-scale modeling approach that combines concepts from rigidity theory, elastic network theory, and constrained geometric simulations. The approach is able to accurately predict protein conformational changes in three steps.[152,232] In the first step, the molecule is decomposed into rigid clusters using the FIRST approach.[164] Importantly, the composition of these clusters is not limited to residues adjacent in sequence or secondary structure elements. Instead, residues that are distant in primary sequence but close in the 3D structure may also be comprised in one cluster. In the second step, clusters are treated as rigid bodies, and the motions of these clusters are predicted by RCNMA (Rigid Cluster Normal Mode Analysis) using an elastic network representation of the coarse-grained protein. When applied to 10 proteins that show conformational changes upon complex formation, directions and magnitudes of the motions predicted by RCNMA agree well with experimentally determined ones, particularly if the movement is dominated by loop or fragment motions. In the last step, the NMSim module generates new conformers of the macromolecule using low-energy normal mode directions, predicted by RCNMA, and random direction components. The generated conformers are then iteratively corrected regarding steric clashes and constraint violations in order to generate stereochemically allowed conformations that lie preferentially in the subspace spanned by low-frequency

normal modes. The NMsim approach was validated on hen egg white lysozyme (HEWL). For this, experimentally determined structures and conformations from state-of-the-art MD simulations[233] were compared to conformations determined by FRODA,[161] CONCOORD,[234] and NMSim. Regarding residue fluctuations, NMSim results show a good agreement with those from MD simulations and experimental structures. With respect to the stereochemical quality, NMSim-generated structures have backbone torsion angle characteristics that are in remarkable agreement with the characteristics of 100 high-resolution experimental structures.

13.6.2.2 Incorporating Flexibility to Docking Algorithms

Once the dynamic properties of the protein are known, incorporating this information is the next challenge in algorithmic development. Comprehensive reviews on methods to deal with flexibility in docking have been published.[155,222,235,236] The main concern in this step is computational efficiency. If a number of conformations is known either from experiment or calculation, a trivial solution is to run a parallel docking against every structure.[237] The main potential drawback is that the dynamic picture might not be complete, and some relevant conformations might not be present.

Running parallel dockings with all available structures is generally affordable when the number of structures to consider is small. An interesting alternative to incorporate flexibility in a mean-field sense was accomplished within Autodock[238,239] by Österberg *et al.*[240] Using a grid-based approach for evaluating the interaction energy between ligand and receptor, they followed different strategies to combine representations of multiple target structures by merging their individual interaction energy grids. Taking either minimum values or potential averaged grid values did not perform well. In turn, weighting different grids according to a Boltzmann distribution assumption yielded the best results. The approach was evaluated on 21 complexes of peptidomimetic inhibitors with human immunodeficiency virus type 1 (HIV-1) protease, and they were able to correctly dock 20 of them using the weighted maps. As a drawback of the method, merging grid representations of different receptor conformations into one may lead to the situation where mutually exclusive combinations of receptor conformations are present. It can also be expected that the method reaches its limits in cases of larger protein mobility. The latest AutoDock version, AutoDock4,[241] also allows us to consider the flexibility of receptor side-chains explicitly. When tested in a redocking experiment with 188 diverse protein–ligand complexes, successful dockings were reported for complexes with 10 or less torsional degrees of freedom. In a cross-docking experiment with 87 HIV protease complexes, adding side-chain flexibility overall leads to more successful docking results but also raises problems like increased computational costs and an increased potential for false positives because of the larger search space. Another docking algorithm, FlipDock,[242] performs automated docking of flexible ligands into flexible

receptors, using the AutoDock force field. To represent ligand and receptor flexibility, a special data structure, the Flexibility Tree (FT),[243] is used. FlipDock was tested on 400 cross-dockings and showed a docking success rate of 93.5%, which compares favorably to the rate by AutoDock3.0 of 72%.

An alternative to the strategies of AutoDock to cope with protein mobility is to work explicitly with known (or predicted) conformations of the protein, but in a way that is more computationally efficient than trivially running several dockings in parallel. FlexE[244] incorporates a united protein description that handles similarities and differences in the ensemble of conformations. Structures are initially superimposed by backbone atoms. Then, united structures are created by combining the alternative side-chain conformations and backbone parts. A united structure is composed of instances, that is, conformationally different substructures. These structures are then used in an incremental docking approach. The ligand is placed fragment by fragment into the active site and possible interactions between the ligand and all instances are determined. In the final step, contributions from all instances are summed, whereby mutually incompatible instances are discarded in order to retain realistic protein structures. As there are several possible combinations of independent instances at each construction step, finding high-scoring independent sets of instances is the most time consuming step of the whole procedure. The authors report an improvement compared to running parallel dockings and merging the results (67% of the best-ranked solutions below 2.0 Å *vs*. 63%) with a considerable reduction of running time. Probably the most severe limitation of this approach is that it does not take into account changes in internal energy of the different protein conformations. This clearly favors open binding pockets that can accommodate large ligands, which form many favorable intermolecular interactions.

Along these lines, Wei *et al.* have pointed out that the largest impact on the improvement of their VS results was obtained by precisely including this contribution.[221] They have used an in-house modified version of DOCK[245] and an ensemble of experimental structures of the receptor as templates to represent conformational changes. The receptor is decomposed into rigid and flexible regions. For each of the flexible regions, there are several conformational possibilities according to the experimental data. In this sense, the receptor is multicomponent: a rigid component is combined with flexible ones. A depth-first search algorithm is used to scan through all possible conformations for the possibility to dock the ligand without steric clashes. If found, the score is computed by summing the contributions from every component. The best-fit conformation of each flexible receptor region is used to assemble the receptor conformation. They applied the method to identify known ligands of a hydrophobic cavity mutant of T4 lysozyme and the folate-binding pocket of thymidylate synthase from the Available Chemicals Directory. The inclusion of a weighted receptor conformational energy in the scoring function led to an improvement in the enrichments, particularly for lysozyme.

This example clearly shows that incorporating flexibility into a docking algorithm does not only involve sampling protein conformations, but also evaluating them[221,246] with respect to this energy. In this regard, it is worth emphasizing the importance of considering the free energy of the receptor's conformation and the stabilizing influence that a bound ligand can provide.[57]

13.6.3 Data-Driven Docking Approaches

Computational docking approaches fail due to two main reasons: insufficient sampling (which includes considering receptors as rigid) and simplifications in the scoring functions. Although rigorous descriptions of intermolecular interactions are available and could be readily incorporated, this comes together with an increase of computational demands. In fact, for most of the cases, the loss in efficiency does not pay off. To overcome this dilemma, another strategy is possible: supplement the scoring functions with experimental information. This is now a current trend in the field of protein–protein docking, as can be seen from the assessments of the last CAPRI round.[247] From there, a relevant conclusion is that "using prior knowledge of the protein regions that are likely to interact remains an important ingredient for achieving successful docking".

To supplement "pure" docking, there are two possibilities: first, to use the knowledge derived directly from the binding partners and, second, to incorporate experimental information about a particular complex. The first alternative, which has been reviewed by Fradera and Mestres,[248] includes the field of the so-called "tailored scoring functions".[249,250] There, the goal is to adapt general scoring functions to the particular target of interest by including information from known interactions of the protein with similar ligands. In general, one could say that the better the studied system is known, the better results can be expected, if the information is properly incorporated. Because PPI as targets do not enjoy such a wealth of information as is available for typical enzyme targets, the strategy cannot be fully exploited. Incorporating directly measured information from experiments is the alternative.

Methods that use experimental information to drive the docking of biomolecular complexes (typically protein–protein or nucleic acid–protein) have been reviewed by van Dijk *et al.*[251] The information that is currently being used in this field has two main sources: mutagenesis studies and different NMR properties (*i.e.*, chemical shifts perturbation, H/D exchange, residual dipolar couplings, diffusion anisotropy). Data derived from biochemical or biophysical experiments can aid docking on two different levels: first, by identifying binding sites and, second, by restricting the conformational search space, concentrating the sampling around native-like poses. The second objective is more challenging and requires data that incorporate very specific structural information describing the orientation of the interacting counterparts. The use of mutagenesis data, for example, is restricted to the first purpose, as the collected examples show.[251]

Since the very beginning, biomolecular NMR data have been connected to computational processing for structure elucidation. Structures are calculated by means of minimizing a hybrid energy function (eqn 13.2) that incorporates NMR measures (E_{data}, with a weighting factor w_{data}; *e.g.*, inter-proton distances derived from peak intensities arising from NOEs, torsion angles and hydrogen bond restraints from scalar couplings, bond orientations from residual dipolar couplings) and a force-field-based term (E_{ff}):[252]

$$E_{hybrid} = E_{ff} + w_{data}E_{data} \tag{13.2}$$

The accuracy of the obtained structures using this approach is highly dependent on the amount and the quality of data available, as has been acknowledged by Chen *et al.*[253] Since complete collection and assignments of structural restraints is a daunting task (and in some cases impossible), a look from the ''other side of the coin'' might prove useful: when experimental data are incomplete or inaccurate, directly deriving biomolecular (complex) structure from it may not be viable. The data may still serve, however, to guide a computational structure prediction tool or distinguish solutions generated by it.

With respect to the whole process, the experimental data can be employed before, at the same time, or after the computational sampling step. When used as a pre-filter, approximate poses are manually generated that are then computationally optimized.[254] However, if the data are incorporated at the search stage, on the one hand, the tedious and non-exhaustive manual generation can be avoided and, on the other hand, native-like configurations are likely to be searched. This last feature is also an advantage over methods that use the experimental data only for post-filtering of proposed solutions.[255]

NMR as a tool for investigating the conformation of bound ligands has been recently reviewed.[256] There, the quantitative analysis of transferred-NOEs and cross-correlated relaxation data are examined from the experimentalist's perspective. From the computational viewpoint, the far more interesting question is which easily obtainable NMR measurements can be incorporated to guide a docking algorithm. In this regard, chemical shift perturbations (CSP) and saturation transfer difference (STD) have already proven promising and will be discussed below. We will not consider here the case of complex structure elucidation using intermolecular NOEs that is applicable to tightly bound ligands, which is the standard protein NMR methodology.[257]

13.6.3.1 Guided Docking Using Chemical Shift Perturbations (CSP)

^{1}H–^{15}N heteronuclear single quantum correlation (HSQC) NMR is nowadays a well-established experiment. The most well-known application in the field of drug design is the so-called ''SAR (structure–activity relationship) by NMR''[33] approach. Upon ligand binding, the chemical shifts (CS) of the interacting

partners are affected due to a change in the environment. Provided that no large conformational changes occur during this process, the largest perturbations that can be observed, *e.g.* on the protein side, are due to the binding of the ligand. In the SAR by NMR approach, different low-affinity fragments are used to "explore" the binding site of a ^{15}N-labeled protein by monitoring CSP and mapping them onto the protein surface. Combining the information obtained from fragments that bind at different regions of the binding pocket, these fragments can be chemically linked to yield new molecules with increased binding affinity. ^{1}H–^{15}N HSQC spectra have also been used to determine that sulindac-derived inhibitors of the Ras–Raf interaction[258] bind directly to Ras. This is a relevant result not only with respect to the particular studied system, but it also shows the technique to be suitable for facilitating the task of finding binding sites at the target (see Section 13.4).

The HADDOCK approach[259,260] uses CSP information upon complexation for structure prediction of macromolecular complexes in a paradigmatic way. The underlying idea is that the size of the configurational/conformational search space can be significantly reduced once the residues involved in the intermolecular interactions are known. Information of this kind can be obtained from the analysis of CSPs. CSPs, however, do not reveal which residues interact with each other. At this point, docking comes into play. Here, the experimental information is introduced by means of ambiguous interaction restraints (AIRs), originally proposed by Nilges.[261] An AIR (eqn 13.3) is defined as an upper-bounded intermolecular distance that must be fulfilled upon complex formation. However, it does not require a particular residue pair to fulfill it, but a subset of pre-selected possible pairs.

$$d^{\mathrm{eff}}_{i(A),B} = \left(\sum_{m=1}^{N_{\mathrm{atoms},i(A)}} \sum_{k=1}^{N_{\mathrm{res},B}} \sum_{n=1}^{N_{\mathrm{atoms},k(B)}} \frac{1}{d^{6}_{m,n}} \right)^{-1/6} \tag{13.3}$$

Residues defined as "involved in the interaction" are taken as pairs (i, k), one from each counterpart A and B, respectively. The distance is computed for every atom m in residue i from the first protein to every atom n of residue k in the second protein. In this way, as soon as two atoms are in contact the restraint is satisfied. Schieborr *et al.*[262] have applied HADDOCK to the problem of protein–ligand docking. Owing to the large size difference between ligand and receptor, the authors modified the protocol, such that only strong effects of the ligand on the protein were considered. As a result, the impact of incorporating the AIR in this case is a restriction of the conformational search to the binding site area. However, no information about the mutual orientation of both binding partners is exploited. In fact, the authors acknowledge that the binding site of the protein could have been located also by mapping the strong CSP onto the protein structure. Thus, with this approach, the success in determining the native structure of the complex still depends on the force-field component of the scoring function.

It is stimulating that CSPs have been used successfully in structure refinement.[263–265] This opens a new perspective for the docking field. Refinement against CSP differs from structure calculations with distance restraints. In the case of CSP refinement, there are no pairwise distance restraints to fulfill, but a scoring function that minimizes differences between observed and calculated CS is required. This implies that an efficient method to compute theoretical CS is available.[266] Contributions to proton CS can be decomposed into local (diamagnetic and paramagnetic) and non-local contributions. In the latter case, effects from nearby aromatic rings (δ_{rc}) and other sources of magnetic anisotropy (δ_{mag}) as well as electrostatic and solvent effects (δ_{el}) are included (eqn 13.4). The local effects are approximated by the observed shifts in short peptides with corresponding secondary structure.

$$\delta_{total} = \delta_{local} + \delta_{rc} + \delta_{mag} + \delta_{el} \tag{13.4}$$

If there is no significant conformational change of the protein upon complexation, the observed CS differences between the free and complex spectra of the protein are due to the ligand. Compared to the SAR by NMR approach that exploits such information only qualitatively, one can now quantify these differences, such that one can deduce additional information about the orientation of different groups of the ligand. Aromatic rings, when present, constitute the main source of the contribution to the total CS difference. Accordingly, the orientation of the ring with respect to the protein can be determined because of the anisotropic nature of this effect. McCoy and Wyss[267] have explored the usefulness of these "ring current effects" within the frame of the program SHIFTS[268] as a post-filtering tool for elucidating the structure of protein–ligand complexes. Although native-like and close-to-native-like conformations were best ranked, the possibility of using such a method for guiding docking remains unclear. There are two concerns that would hamper the approach: first, in many cases errors in the prediction are in the same range as the observed changes upon complex formation and, second, CSP may also arise from conformational changes of the protein upon complex formation. González-Ruiz *et al.* recently presented a new approach that steers protein–ligand docking with quantitative NMR chemical shift perturbations.[269] This method is based on a hybrid scoring scheme that combines a weighted sum of DrugScore,[270] describing protein–ligand interactions, and Kendall's rank correlation coefficient,[271] which scores the agreement between experimentally determined and computed CSP for a given ligand pose. An efficient empirical model considering only ring-current effects is used for back-calculating CSP for a ligand pose. The hybrid scoring scheme was tested on 70 protein–ligand complexes with computed CSP reference data. Without CSP information, a docking success rate of 71% was achieved. This increased to 99% if CSP information was included. The approach should be helpful for protein–ligand complexes that are computationally difficult to predict, *e.g.* if a ligand binds to a flat binding site.

Instead of monitoring CSP from the protein side, Wang and Merz used a semiempirical method at the NMDO level developed to predict CS of ligand protons.[272] As an application they ranked manually generated poses of a FKBP12 inhibitor, finding a very good correlation between CSP and structural RMSDs. Despite the higher level of theory applied than in the SHIFTS approach, there are still concerns due to issues of protein flexibility. In addition, although sufficiently fast for a single calculation, a scoring function in a docking algorithm has to be evaluated millions of times, which makes this method unappealing for VS applications.

13.6.3.2 Impact of Data-driven Docking on Computational Targeting of PPI

Most of the current advances in designing small molecules to target PPI come, as expected, from experimental and not theoretical approaches.[273] However, as the docking field develops, more challenging examples can be targeted, which also includes PPI. Mixing experimental information concerning a particular system with current docking algorithms has proven not only feasible but also as the source for large improvements. This holds especially true in the particularly demanding field of protein–protein docking.

At the moment, quantitative information that can be derived from experiment (such as in the cases of CS and STD described) is used in combination with computational approaches to confirm and validate binding modes as a post-processing tool. Still, docking methods that rely on the comparison between computer-predicted and experimentally observed properties remain inefficient, which prevents them from being useful in VS experiments. Nevertheless, further developments of approaches that use experimental data at search time should be pursued. If used in a pre-processing step, using experimental information can help supplement computational approaches by restricting the search space to those regions that involve key interacting residues.

13.7 Summary and Outlook

Recent advances in computational approaches to detect PPI and identify small-molecule PPIM have been reviewed. The number of successful examples of computer-aided identification of agents that modulate PPI is still limited. However, there is significant progress in understanding and modeling molecular recognition properties of PPI, and we expect that in the next few years the influence of computational means on targeting PPI will increase considerably. This is reflected in the hit rates of around 15–20% obtained for four of the virtual screening experiments listed in Table 1,[289,293-295] which are much higher than if non-targeted libraries are screened.[17]

We have primarily focused on structure-based approaches that require knowledge of at least one of the interacting components of the protein–protein

complex. Although detailed structural information about the PPIF may not be available in all cases, and further challenges arise from the inherent plasticity of PPIF, this way seems to be more promising to pursue in our opinion than ligand-based approaches. On the one hand, for the latter methods to be successful, structure–activity relationships or pharmacophore models derived from lead structures obtained by experimental high-throughput screening need to be established. However, a sufficient amount of good-quality data that is usually required for training may not be available in many cases of novel protein–protein targets. On the other hand, structural knowledge and insights into the forces that act at a PPIF are not only critical to the prospective discovery or design of small-molecule PPIM. It will also allow us to understand the reasons for affinity and selectivity towards a given target retrospectively,[274] which will aid in guiding future efforts to improve both properties.

The approaches presented here form two key levels of what may become an integrated approach for finding small-molecule PPIM by computational means (Figure 13.3). Given the structure of the target, the first level comprises the prediction of potential binding sites, which includes hot spot analysis, cleft detection, and allosteric site detection. Subsequently, flexible, data-driven ligand docking with improved scoring will be applied in the frame of VS. Equally important, but not covered in this review, are aspects of target identification, target druggability, and ligand pre-selection. Even in the field of "classical" targets the first two topics have only been touched recently,[275,276] and computational approaches are emerging.[202,277–279] However, they may become even more important in the case of protein–protein targets due to the large variability of PPIF. As such, the interfaces of Bcl-2/Bak and p53/Mdm2 are highly hydrophobic whereas the Ras/Raf-1 kinase interface is largely polar.[274] Experiences gained from one PPIF may thus not be transferable to another case and identifying "dead-end targets" early will save time and expenses.

With respect to ligand pre-selection, computational approaches to filter for drug-like properties of small molecules are already widely used.[280,281] Here, the main focus is on the estimation of ADMET (absorption, distribution, metabolism, excretion, toxicity) properties. Considering, however, that the lack of well defined clefts or grooves in PPIF often results in PPIM with upper-limit potencies in the micromolar range,[89] a potential high degree of promiscuity of these PPIM on other targets is the consequence. Such non-specific, "promiscuous" compounds have been identified among screening hits,[282] lead compounds,[283] and known drugs.[284] A single, aggregation-based mechanism of action has been proposed to explain the observed effects.[285] Accordingly, the panel of approaches to estimate the pharmacological profile of a potential PPIM should be extended to methods[284,286] that allow us to identify and eliminate potentially promiscuous compounds at early stages. Finally, of heightened interest for the case of protein–protein targets would be schemes that pre-select compounds to be used in the virtual screening that bind to specific protein domains.[17] For this, however, a systematic identification of such entities by experiment is required first.

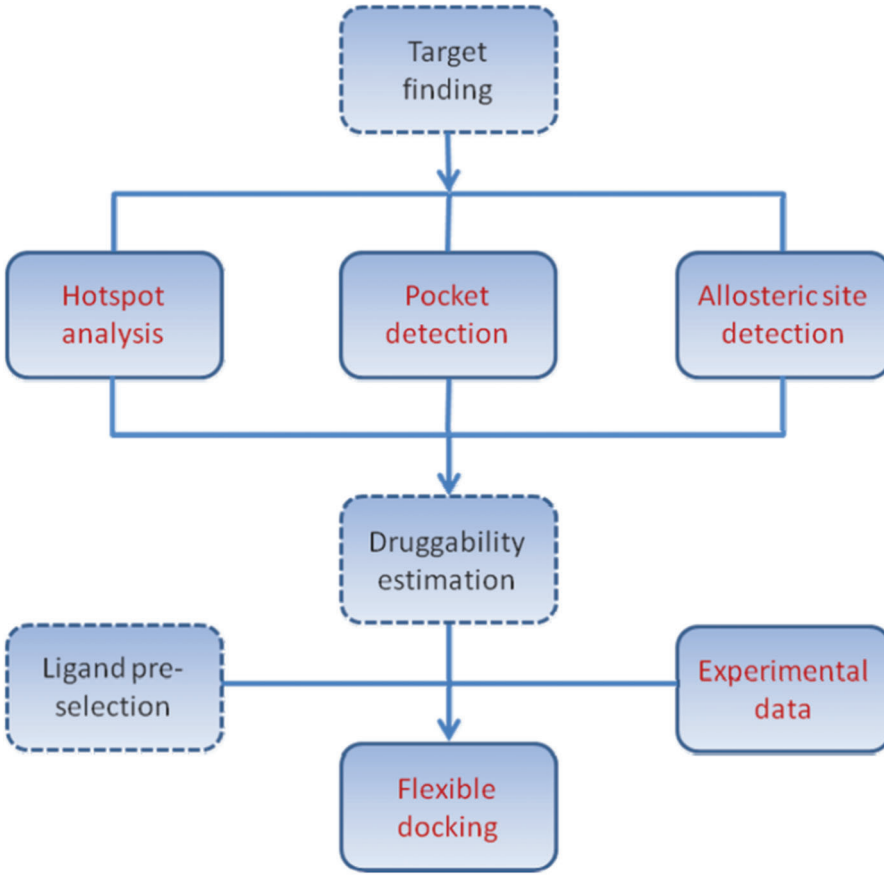

Figure 13.3 Flowchart describing an integrated approach to the computer-aided identification of small-molecule PPIM. Topics depicted in boxes with straight lines have been covered in this review.

Overall, being able to establish such a process will be a major breakthrough in the field, and in combination with progress in various experimental areas, we are awaiting exciting times for modulating PPI.

Acknowledgements

This study was supported by funds from Heinrich-Heine-University.

References

1. A. C. Gavin, M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak,

D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, *Nature*, 2002, **415**, 141.
2. P. Bork, L. J. Jensen, C. von Mering, A. K. Ramani, I. Lee and E. M. Marcotte, *Curr. Opin. Struct. Biol.*, 2004, **14**, 292.
3. M. R. Arkin, M. Randal, W. L. DeLano, J. Hyde, T. N. Luong, J. D. Oslob, D. R. Raphael, L. Taylor, J. Wang, R. S. McDowell, J. A. Wells and A. C. Braisted, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1603.
4. P. M. Fischer, *Drug Design Rev.*, 2005, 179.
5. L. Stockwin and S. Holmes, *Expert Opin. Biol. Ther.*, 2003, **3**, 1133.
6. R. Zutshi, M. Brickner and J. Chmielewski, *Curr. Opin. Chem. Biol.*, 1998, **2**, 62.
7. Z.-Y. Zhang, R. A. Poorman, L. L. Maggiora, R. L. Heinrikson and F. J. Kezdy, *J. Biol. Chem.*, 1991, **266**, 15591.
8. M. Liuzzi, R. Deziel, N. Moss, P. Beaulieu, A. M. Bonneau, C. Bousquet, J. G. Chafouleas, M. Garneau, J. Jaramillo and R. L. Krogsrud, *Nature*, 1994, **372**, 695.
9. V. Bottger, A. Bottger, S. F. Howard, S. M. Picksley, P. Chene, C. Garcia-Echeverria, H. K. Hochkeppel and D. P. Lane, *Oncogene*, 1996, **13**, 2141.
10. M. Rubinstein and M. Y. Niv, *Biopolymers*, 2009, **91**, 505.
11. J. Eichler, *Curr. Opin. Chem. Biol.*, 2008, **12**, 707.
12. L. Pagliaro, J. Felding, K. Audouze, S. J. Nielsen, R. B. Terry, C. Krog-Jensen and S. Butcher, *Curr. Opin. Struct. Biol.*, 2004, **8**, 442.
13. J. A. Wells and C. L. McClendon, *Nature*, 2007, **450**, 1001.
14. M. R. Arkin and J. A. Wells, *Nat. Rev. Drug Discovery*, 2004, **3**, 301.
15. T. Berg, *Curr. Opin. Drug Discovery Dev*., 2008, **11**, 666.
16. R. L. Juliano, A. Astriab-Fisher and D. Falke, *Mol. Interv.*, 2001, **1**, 40.
17. T. Berg, *Angew. Chem., Int. Ed.*, 2003, **42**, 2462.
18. A. G. Cochran, *Curr. Opin. Chem. Biol.*, 2001, **5**, 654.
19. H. Yin and A. D. Hamilton, *Angew. Chem., Int. Ed.*, 2005, **44**, 4130.
20. A. Christopoulos, *Nat. Rev. Drug Discovery*, 2002, **1**, 198.
21. G. Fuentes, J. Oyarzabal and A. M. Rojas, *Curr. Opin. Drug Discovery Dev.*, 2009, **12**, 358.
22. A. Degterev, A. Lugovskoy, M. Cardone, B. Mulley, G. Wagner, T. Mitchison and J. Yuan, *Nat. Cell Biol.*, 2001, **3**, 173.
23. D. L. Boger, J. K. Lee, J. Goldberg and Q. Jin, *J. Org. Chem.*, 2000, **65**, 1467.
24. S. Silletti, T. Kessler, J. Goldberg, D. L. Boger and D. A. Cheresh, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 119.
25. D. L. Boger, J. Goldberg, W. Jiang, W. Chai, P. Ducray, J. K. Lee, R. S. Ozer and C. M. Andersson, *Bioorg. Med. Chem.*, 1998, **6**, 1347.
26. J. Goldberg, Q. Jin, Y. Ambroise, S. Satoh, J. Desharnais, K. Capps and D. L. Boger, *J. Am. Chem. Soc.*, 2002, **124**, 544.

27. D. V. Erbe, S. Wang, Y. Xing and J. F. Tobin, *J. Biol. Chem.*, 2002, **277**, 7363.
28. W. Jahnke, A. Florsheimer, M. J. Blommers, C. G. Paris, J. Heim, C. M. Nalin and L. B. Perez, *Curr. Top. Med. Chem.*, 2003, **3**, 69.
29. S. Fletcher and A. D. Hamilton, *Curr. Top. Med. Chem.*, 2007, **7**, 922.
30. K. Almholdt, P. O. Arkhammar, O. Thastrup and S. Tullin, *Biochem. J.*, 1999, **137**, 211.
31. D. A. Erlanson, R. S. McDowell, M. M. He, M. Randal, R. L. Simmons, J. Kung, A. Waight and S. K. Hansen, *J. Am. Chem. Soc.*, 2003, **125**, 5602.
32. D. A. Erlanson, J. W. Lam, C. Wiesmann, T. N. Luong, R. L. Simmons, W. L. DeLano, I. C. Choong, M. T. Burdett, W. M. Flanagan, D. Lee, E. M. Gordon and T. O'Brien, *Nat. Biotechnol.*, 2003, **21**, 308.
33. S. B. Shuker, P. J. Hajduk, R. P. Meadows and S. W. Fesik, *Science*, 1996, **274**, 153.
34. A. C. Braisted, J. D. Oslob, W. L. DeLano, J. Hyde, R. S. McDowell, N. Waal, C. Yu, M. R. Arkin and B. C. Raimundo, *J. Am. Chem. Soc.*, 2003, **125**, 3714.
35. H. J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller and F. Mueller, *J. Med. Chem.*, 2000, **43**, 2664.
36. D. Gonzalez Ruiz and H. Gohlke, *Curr. Med. Chem.*, 2006, **13**, 2607.
37. F. K. Pettit and J. U. Bowie, *J. Mol. Biol.*, 1999, **285**, 1377.
38. O. Keskin, A. Gursoy, B. Ma and R. Nussinov, *Chem. Rev.*, 2008, **108**, 1225.
39. L. L. Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177.
40. J. Janin and C. Chothia, *J. Biol. Chem.*, 1990, **265**, 16027.
41. D. R. Davies, E. A. Padlan and S. Sheriff, *Annu. Rev. Biochem.*, 1990, **59**, 439.
42. P. Chene, *ChemMedChem*, 2006, **1**, 400.
43. P. Chakrabarti and J. Janin, *Proteins*, 2002, **47**, 334.
44. M. W. Peczuh and A. D. Hamilton, *Chem. Rev.*, 2000, **100**, 2479.
45. S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13.
46. F. Glaser, D. M. Steinberg, I. A. Vakser and N. Ben-Tal, *Proteins*, 2001, **43**, 89.
47. E. A. Padlan, *Proteins*, 1990, **7**, 112.
48. A. J. McCoy, V. Chandana Epa and P. M. Colman, *J. Mol. Biol.*, 1997, **268**, 570.
49. B. A. Joughin, D. F. Green and B. Tidor, *Protein Sci.*, 2005, **14**, 1363.
50. G. Schreiber and A. R. Fersht, *J. Mol. Biol.*, 1995, **248**, 478.
51. M. R. Arkin and J. A. Wells, *J. Mol. Biol.*, 1998, **284**, 1083.
52. G. A. Papoian, J. Ulander and P. G. Wolynes, *J. Am. Chem. Soc.*, 2003, **125**, 9170.
53. F. Rodier, R. P. Bahadur, P. Chakrabarti and J. Janin, *Proteins*, 2005, **60**, 36.

54. E. J. Sundberg and R. A. Mariuzza, *Structure*, 2000, **8**, R137.
55. W. L. DeLano, M. H. Ultsch, A. M. de Vos and J. A. Wells, *Science*, 2000, **287**, 1279.
56. I. Luque and E. Freire, *Proteins*, 2000, **4**, 63.
57. S. J. Teague, *Nat. Rev. Drug Discovery*, 2003, **2**, 527.
58. C.-S. Goh, D. Milburn and M. Gerstein, *Curr. Opin. Struct. Biol.*, 2004, **14**, 104.
59. B. Ma, H. J. Wolfson and R. Nussinov, *Curr. Opin. Struct. Biol.*, 2001, **11**, 364.
60. C.-J. Tsai, B. Ma and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9970.
61. J. R. Horn and B. K. Shoichet, *J. Mol. Biol.*, 2004, **336**, 1283.
62. V. A. Feher and J. Cavanagh, *Nature*, 1999, **400**, 289.
63. L. E. Kay, D. R. Muhandiram, N. A. Farrow, Y. Aubin and J. D. Forman-Kay, *Biochemistry*, 1996, **35**, 361.
64. B. Ma, M. Shatsky, H. J. Wolfson and R. Nussinov, *Protein Sci.*, 2002, **11**, 184.
65. W. E. Stites, *Chem. Rev.*, 1997, **97**, 1233.
66. N. Brooijmans, K. A. Sharp and I. D. Kuntz, *Proteins*, 2002, **48**, 645.
67. B. C. Cunningham and J. A. Wells, *J. Mol. Biol.*, 1993, **234**, 554.
68. L. Jin and J. A. Wells, *Protein Sci.*, 1994, **3**, 2351.
69. T. Clackson and J. A. Wells, *Science*, 1995, **267**, 383.
70. T. Clackson, M. H. Ultsch, J. A. Wells and A. M. de Vos, *J. Mol. Biol.*, 1998, **277**, 1111.
71. A. A. Bogan and K. S. Thorn, *J. Mol. Biol.*, 1998, **280**, 1.
72. Z. Hu, B. Ma, H. Wolfson and R. Nussinov, *Proteins*, 2000, **39**, 331.
73. J. Novotny, R. E. Bruccoleri and F. A. Saul, *Biochemistry*, 1989, **28**, 4735.
74. J. L. Kouadio, J. R. Horn, G. Pal and A. A. Kossiakoff, *J. Biol. Chem.*, 2005, **280**, 25524.
75. I. S. Moreira, P. A. Fernandes and M. J. Ramos, *Proteins*, 2007, **68**, 803.
76. B. Ma, T. Elkayam, H. Wolfson and R. Nussinov, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5772.
77. X. Li, O. Keskin, B. Ma, R. Nussinov and J. Liang, *J. Mol. Biol.*, 2004, **344**, 781.
78. D. E. Otzen and A. R. Fersht, *Protein Eng.*, 1999, **12**, 41.
79. S. W. Lockless and R. Ranganathan, *Science*, 1999, **286**, 295.
80. O. Keskin, B. Ma and R. Nussinov, *J. Mol. Biol.*, 2005, **345**, 1281.
81. A. W. White, A. D. Westwell and G. Brahemi, *Expert Rev. Mol. Med.*, 2008, **10**, e8.
82. S. S. Sidhu, W. J. Fairbrother and K. Deshayes, *ChemBioChem*, 2003, **4**, 14.
83. R. C. Pillutla, K. C. Hsiao, J. R. Beasley, J. Brandt, S. Ostergaard, P. H. Hansen, J. C. Spetzler, G. M. Danielsen, A. S. Andersen, R. E. Brissette, M. Lennick, P. W. Fletcher, A. J. Blume, L. Schaffer and N. I. Goldstein, *J. Biol. Chem.*, 2002, **277**, 22590.

84. N. C. Wrighton, F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe and W. J. Dower, *Science*, 1996, **273**, 458.
85. S. E. Cwirla, P. Balasubramanian, D. J. Duffin, C. R. Wagstrom, C. M. Gates, S. C. Singer, A. M. Davis, R. L. Tansik, L. C. Mattheakis, C. M. Boytos, P. J. Schatz, D. P. Baccanari, N. C. Wrighton, R. W. Barrett and W. J. Dower, *Science*, 1997, **276**, 1696.
86. W. L. DeLano, *Curr. Opin. Struct. Biol.*, 2002, **12**, 14.
87. P. L. Toogood, *J. Med. Chem.*, 2002, **45**, 1543.
88. D. A. Erlanson, R. S. McDowell and T. O'Brien, *J. Med. Chem.*, 2004, **47**, 3463.
89. Y. Li, Y. Huang, C. P. Swaminathan, S. J. Smith-Gill and R. A. Mariuzza, *Structure*, 2005, **13**, 297.
90. K. S. Thorn and A. A. Bogan, *Bioinformatics*, 2001, **17**, 284.
91. T. B. Fischer, K. V. Arunachalam, D. Bailey, V. Mangual, S. Bakhru, R. Russo, D. Huang, M. Paczkowski, V. Lalchandani, C. Ramachandra, B. Ellison, S. Galer, J. Shapley, E. Fuentes and J. Tsai, *Bioinformatics*, 2003, **19**, 1453.
92. N. Tuncbag, G. Kar, O. Keskin, A. Gursoy and R. Nussinov, *Brief Bioinform.*, 2009, **10**, 217.
93. E. Guney, N. Tuncbag, O. Keskin and A. Gursoy, *Nucleic Acids Res.*, 2008, **36**, D662.
94. J. Srinivasan, T. E. Cheatham, III, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401.
95. I. Massova and P. A. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133.
96. S. Huo, I. Massova and P. A. Kollman, *J. Comput. Chem.*, 2002, **23**, 15.
97. W. Wang and P. A. Kollman, *J. Mol. Biol.*, 2000, **303**, 567.
98. I. S. Moreira, P. A. Fernandes and M. J. Ramos, *J. Comput. Chem.*, 2007, **28**, 644.
99. I. S. Moreira, P. A. Fernandes and M. J. Ramos, *J. Phys. Chem. B*, 2006, **110**, 10962.
100. A. Benedix, C. M. Becker, B. L. de Groot, A. Caflisch and R. A. Bockmann, *Nat. Methods*, 2009, **6**, 3.
101. B. L. de Groot, D. M. van Aalten, R. M. Scheek, A. Amadei, G. Vriend and H. J. Berendsen, *Proteins*, 1997, **29**, 240.
102. D. M. Kruger and H. Gohlke, *Nucleic Acids Res.*, 2010, **38**, W480.
103. C. K. Vaughan, A. M. Buckle and A. R. Fersht, *J. Mol. Biol.*, 1999, **266**, 1487.
104. H. Gohlke, C. Kiel and D. A. Case, *J. Mol. Biol.*, 2003, **330**, 891.
105. G. Archontis, T. Simonson and M. Karplus, *J. Mol. Biol.*, 2001, **306**, 307.
106. J. Gao, K. Kuczera, B. Tidor and M. Karplus, *Science*, 1989, **244**, 1069.
107. Z. S. Hendsch and B. Tidor, *Protein Sci.*, 1999, **8**, 1381.
108. A. E. Mark and W. F. van Gunsteren, *J. Mol. Biol.*, 1994, **240**, 167.
109. G. P. Brady and K. A. Sharp, *J. Mol. Biol.*, 1995, **254**, 77.
110. S. Boresch and M. Karplus, *J. Mol. Biol.*, 1995, **254**, 801.

111. H. Gohlke and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 238.
112. H. J. Böhm, *J. Comput. Aided Mol. Des.*, 1994, **8**, 243.
113. R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369.
114. J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano, *Nucleic Acids Res.*, 2005, **33**, W382.
115. T. Kortemme, D. E. Kim and D. Baker, *Sci. STKE*, 2004, 2004, pl2.
116. T. Kortemme and D. Baker, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14116.
117. T. Kortemme, L. A. Joachimiak, A. N. Bullock, A. D. Schuler, B. L. Stoddard and D. Baker, *Nat. Struct. Biol.*, 2004, **11**, 371.
118. C. Kiel, S. Wohlgemuth, F. Rousseau, J. Schymkowitz, J. Ferkinghoff-Borg, F. Wittinghofer and L. Serrano, *J. Mol. Biol.*, 2005, **348**, 759.
119. T. Kortemme, A. V. Morozov and D. Baker, *J. Mol. Biol.*, 2003, **326**, 1239.
120. D. Bashford and D. A. Case, *Annu. Rev. Phys. Chem.*, 2000, **51**, 129.
121. N. Tuncbag, A. Gursoy and O. Keskin, *Bioinformatics*, 2009, **25**, 1513.
122. N. Tuncbag, O. Keskin and A. Gursoy, *Nucleic Acids Res.*, 2010, **38**, W402.
123. S. J. Darnell, L. LeGault and J. C. Mitchell, *Nucleic Acids Res.*, 2008, **36**, W265.
124. S. J. Darnell, D. Page and J. C. Mitchell, *Proteins*, 2007, **68**, 813.
125. D. G. Levitt and L. J. Banaszak, *J. Mol. Graphics*, 1992, **10**, 229.
126. M. Hendlich, F. Rippmann and G. Barnickel, *J. Mol. Graphics Model.*, 1997, **15**, 359.
127. B. Huang and M. Schroeder, *BMC Struct. Biol.*, 2006, **6**, 19.
128. R. A. Laskowski, *J. Mol. Graphics*, 1995, **13**, 323.
129. J. Liang, H. Edelsbrunner and C. Woodward, *Protein Sci.*, 1998, **7**, 1884.
130. G. P. Brady, Jr. and P. F. Stouten, *J. Comput. Aided Mol. Des.*, 2000, **14**, 383.
131. M. Weisel, E. Proschak and G. Schneider, *Chem. Cent. J.*, 2007, **1**, 7.
132. V. Le Guilloux, P. Schmidtke and P. Tuffery, *BMC Bioinf.*, 2009, **10**, 168.
133. S. Eyrisch and V. Helms, *J. Med. Chem.*, 2007, **50**, 3457.
134. S. Leis, S. Schneider and M. Zacharias, *Curr. Med. Chem.*, 2010, **17**, 1550.
135. Y. Levy, S. S. Cho, J. N. Onuchic and P. G. Wolynes, *J. Mol. Biol.*, 2005, **346**, 1121.
136. E. Fischer, *Ber. Dtsch. Chem. Ges.*, 1894, **27**, 2985.
137. D. E. Koshland, *Science*, 1967, **156**, 540.
138. C.-J. Tsai, S. Kumar, B. Ma and R. Nussinov, *Protein Sci.*, 1999, **8**, 1181.
139. B. F. Volkman, D. Lipson, D. E. Wemmer and D. Kern, *Science*, 2001, **291**, 2429.
140. S. R. Kimura, R. C. Brower, S. Vajda and C. J. Camacho, *Biophys. J.*, 2001, **80**, 635.
141. D. Rajamani, S. Thiel, S. Vajda and C. J. Camacho, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11287.

142. G. R. Smith, M. J. E. Sternberg and P. A. Bates, *J. Mol. Biol.*, 2005, **347**, 1077.
143. C. A. Sotriffer, O. Krämer and G. Klebe, *Proteins*, 2004, **56**, 52.
144. D. A. Case, *Curr. Opin. Struct. Biol.*, 1994, **4**, 285.
145. S. Hayward and B. L. de Groot, *Methods Mol. Biol.*, 2008, **443**, 89.
146. L. Skjaerven, S. M. Hollup and N. Reuter, *J. Mol. Struct.: THEOCHEM*, 2009, **898**, 42.
147. I. Bahar, A. R. Atilgan and B. Erman, *Fold. Des.*, 1997, **2**, 173.
148. F. Tama, F. X. Gadea, O. Marques and Y.-H. Sanejouand, *Proteins*, 2000, **41**, 1.
149. I. Bahar and A. J. Rader, *Curr. Opin. Struct. Biol.*, 2005, **15**, 586.
150. J. Ma, *Structure*, 2005, **13**, 373.
151. W. Zheng and S. Doniach, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13253.
152. A. Ahmed and H. Gohlke, *Proteins*, 2006, **63**, 1038.
153. F. Tama and Y. H. Sanejouand, *Protein Eng.*, 2001, **14**, 1.
154. M. Zacharias and H. Sklenar, *J. Comput. Chem.*, 1999, **20**, 287.
155. C. N. Cavasotto and R. A. Abagyan, *J. Mol. Biol.*, 2004, **337**, 209.
156. O. Livnah, E. A. Stura, S. A. Middleton, D. L. Johnson, L. K. Jolliffe and I. A. Wilson, *Science*, 1999, **283**, 987.
157. S. Atwell, M. Ultsch, A. M. De Vos and J. A. Wells, *Science*, 1997, **278**, 1125.
158. Z. Zhang, Y. Shi and H. Liu, *Biophys. J.*, 2003, **84**, 3583.
159. M. Lei, M. I. Zavodszky, L. A. Kuhn and M. F. Thorpe, *J. Comput. Chem.*, 2004, **25**, 1133.
160. M. I. Zavodszky, M. Lei, M. F. Thorpe, A. R. Day and L. A. Kuhn, *Proteins*, 2004, **57**, 243.
161. S. Wells, S. Menor, B. M. Hespenheide and M. F. Thorpe, *Phys. Biol.*, 2005, **2**, 1.
162. D. J. Jacobs and M. F. Thorpe, *Phys. Rev. Lett.*, 1995, **75**, 4051.
163. *Rigidity Theory and Applications*, ed. M. F. Thorpe and P. M. Duxbury, Kluwer/Plenum, New York, 1999.
164. D. J. Jacobs, A. J. Rader, L. A. Kuhn and M. F. Thorpe, *Proteins*, 2001, **44**, 150.
165. H. Gohlke, L. A. Kuhn and D. A. Case, *Proteins*, 2004, **56**, 322.
166. D. Seeliger, J. Haas and B. L. de Groot, *Structure*, 2007, **15**, 1482.
167. S. Eyrisch and V. Helms, *J. Comput. Aided Mol. Des.*, 2009, **23**, 73.
168. R. A. Laskowski, F. Gerick and J. M. Thornton, *FEBS Lett.*, 2009, **583**, 1692.
169. J.-P. Changeux and S. J. Edelstein, *Science*, 2005, **308**, 1424.
170. D. Kern and E. R. Zuiderweg, *Curr. Opin. Struct. Biol.*, 2003, **13**, 748.
171. F. Rousseau and J. Schymkowitz, *Curr. Opin. Struct. Biol.*, 2005, **15**, 23.
172. K. Gunasekaran, B. Ma and R. Nussinov, *Proteins*, 2004, **57**, 433.
173. J. A. Hardy and J. A. Wells, *Curr. Opin. Struct. Biol.*, 2004, **14**, 706.

174. N. G. Oikonomakos, V. T. Skamnaki, K. E. Tsitsanou, N. G. Gavalas and L. N. Johnson, *Structure*, 2000, **8**, 575.
175. C. Wiesmann, K. J. Barr, J. Kung, J. Zhu, D. A. Erlanson, W. Shen, B. J. Fahr, M. Zhong, L. Taylor, M. Randal, R. S. McDowell and S. K. Hansen, *Nat. Struct. Mol. Biol.*, 2004, **11**, 730.
176. L. A. Kohlstaedt, J. Wang, J. M. Friedman, P. A. Rice and T. A. Steitz, *Science*, 1992, **256**, 1783.
177. J. A. Hardy, J. Lam, J. T. Nguyen, T. O'Brien and J. A. Wells, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 12461.
178. O. Lichtarge, H. R. Bourne and F. E. Cohen, *J. Mol. Biol.*, 1996, **257**, 342.
179. O. Lichtarge and M. E. Sowa, *Curr. Opin. Struct. Biol.*, 2002, **12**, 21.
180. M. E. Sowa, W. He, T. G. Wensel and O. Lichtarge, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 1483.
181. M. E. Sowa, W. He, K. C. Slep, M. A. Kercher, O. Lichtarge and T. G. Wensel, *Nat. Struct. Biol.*, 2001, **8**, 234.
182. K. C. Slep, M. A. Kercher, W. He, C. W. Cowan, T. G. Wensel and P. B. Sigler, *Nature*, 2001, **409**, 1071.
183. G. M. Suel, S. W. Lockless, M. A. Wall and R. Ranganathan, *Nat. Struct. Biol.*, 2003, **10**, 59.
184. M. E. Hatley, S. W. Lockless, S. K. Gibson, A. G. Gilman and R. Ranganathan, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14445.
185. A. I. Shulman, C. Larson, D. J. Mangelsdorf and R. Ranganathan, *Cell*, 2004, **116**, 417.
186. V. J. Hilser and E. Freire, *J. Mol. Biol.*, 1996, **262**, 756.
187. E. Freire, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 10118.
188. E. Freire, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 11680.
189. H. Pan, J. C. Lee and V. J. Hilser, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 12020.
190. A. P. Kuzin, M. Nukaga, Y. Nukaga, A. M. Hujer, R. A. Bonomo and J. R. Knox, *Biochemistry*, 1999, **38**, 5720.
191. S. Robineau, M. Chabre and B. Antonny, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 9913.
192. L. Renault, B. Guibert and J. Cherfils, *Nature*, 2003, **426**, 525.
193. Y. Pommier and J. Cherfils, *Trends Pharmacol. Sci.*, 2005, **26**, 138.
194. B. K. Shoichet, *Nature*, 2004, **432**, 862.
195. P. Ferrara, H. Gohlke, D. J. Price, G. Klebe and C. L. Brooks, *J. Med. Chem.*, 2004, **47**, 3032.
196. W. L. Jorgensen, *Science*, 2004, **303**, 1813.
197. B. Honig and A. Nicholls, *Science*, 1995, **268**, 1144.
198. M. K. Gilson, K. Sharp, B. Honig, R. Fine and R. Hagstrom, *Biophys. J.*, 1987, **51**, A234.
199. N. Arora and D. Bashford, *Proteins*, 2001, **43**, 12.
200. N. Budin, N. Majeux and A. Caflish, *Biol. Chem.*, 2001, **382**, 1365.
201. M. Schaefer and M. Karplus, *J. Phys. Chem.*, 1996, **100**, 1578.
202. J. An, M. Totrov and R. Abagyan, *Genome Inform.*, 2004, **15**, 31.

203. J. Momand, G. P. Zambetti, D. C. Olson, D. George and A. J. Levine, *Cell*, 1992, **69**, 1237.
204. W. C. Still, A. Tempczyk, R. C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.*, 1990, **112**, 6127.
205. M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case and C. L. Brooks, III, *J. Comput. Chem.*, 2004, **25**, 265.
206. D. Bashford and D. A. Case, *Annu. Rev. Phys. Chem.*, 2000, **51**, 129.
207. X. Zou, Y. Sun and I. D. Kuntz, *J. Am. Chem. Soc.*, 1999, **121**, 8033.
208. J. Krumrine, F. Raubacher, N. Brooijmans and I. D. Kuntz, *Methods Biochem. Anal.*, 2003, **44**, 443.
209. H. Y. Liu, I. D. Kuntz and X. Q. Zou, *J. Phys. Chem. B*, 2004, **108**, 5453.
210. G. D. Hawkins, C. J. Cramer and D. G. Truhlar, *Chem. Phys. Lett.*, 1995, **246**, 122.
211. J. C. Bressi, C. L. Verlinde, A. M. Aronov, M. L. Shaw, S. S. Shin, L. N. Nguyen, S. Suresh, F. S. Buckner, W. C. Van Voorhis, I. D. Kuntz, W. G. Hol and M. H. Gelb, *J. Med. Chem.*, 2001, **44**, 2080.
212. M. Bogyo, S. Verhelst, V. Bellingard-Dubouchaud, S. Toba and D. Greenbaum, *Chem. Biol.*, 2000, **7**, 27.
213. N. Fujii, J. J. Haresco, K. A. P. Novak, D. Stokoe, I. D. Kuntz and R. K. Guy, *J. Am. Chem. Soc.*, 2003, **125**, 12074.
214. I. W. Davis and D. Baker, *J. Mol. Biol.*, 2009, **385**, 381.
215. A. Morreale, R. Gil-Redondo and A. R. Ortiz, *Proteins*, 2007, **67**, 606.
216. D. A. Case, T. E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668.
217. P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889.
218. S. P. Brown and S. W. Muchmore, *J. Chem. Inf. Model.*, 2007, **47**, 1493.
219. S. P. Brown and S. W. Muchmore, *J. Med. Chem.*, 2009, **52**, 3159.
220. S. McGovern and B. Shoichet, *J. Med. Chem.*, 2003, **46**, 2895.
221. B. Q. Wei, L. H. Weaver, A. M. Ferrari, B. W. Matthews and B. K. Shoichet, *J. Mol. Biol.*, 2004, **337**, 1161.
222. C. B. Rao, J. Subramanian and S. D. Sharma, *Drug Discovery Today*, 2009, **14**, 394.
223. D. Barnard, B. Diaz, L. Hettich, E. Chuang, X. F. Zhang, J. Avruch and M. Marshall, *Oncogene*, 1995, **10**, 1283.
224. M. I. Zavodszky and L. A. Kuhn, *Protein Sci.*, 2005, **14**, 1104.
225. V. Schnecke, C. Swanson, E. Getzoff, J. Tainer and L. A. Kuhn, *Proteins*, 1998, **33**, 74.
226. R. A. Laskowski, M. W. MacArthur, D. S. Moss and J. M. Thornton, *J. Appl. Crystallogr.*, 1993, **26**, 283.
227. R. Najmanovich, J. Kuttner, V. Sobolev and M. Edelman, *Proteins*, 2000, **39**, 261.
228. M. J. Betts and M. J. E. Sternberg, *Protein Eng.*, 1999, **12**, 271.

229. C. S. Goh, D. Milburn and M. Gerstein, *Curr. Opin. Struct. Biol.*, 2004, **14**, 104.
230. M. Zacharias, *Proteins*, 2004, **54**, 759.
231. R. Tatsumi, Y. Fukunishi and H. Nakamura, *J. Comput. Chem.*, 2004, **25**, 1995.
232. A. Ahmed and H. Gohlke, presented at the 1st International Conference on Computational & Mathematical Biomedical Engineering (CMBE09), Swansea, 2009.
233. A. N. Koller, H. Schwalbe and H. Gohlke, *Biophys. J.*, 2008, **95**, L04.
234. B. L. de Groot, G. Vriend and H. J. Berendsen, *J. Mol. Biol.*, 1999, **286**, 1241.
235. S. F. Sousa, P. A. Fernandes and M. J. Ramos, *Proteins*, 2006, **65**, 15.
236. M. Totrov and R. Abagyan, *Curr. Opin. Struct. Biol.*, 2008, **18**, 178.
237. H. A. Carlson, *Curr. Opin. Chem. Biol.*, 2002, **6**, 447.
238. T. Y. Lee, V. D. Le, D. Y. Lim, Y. C. Lin, G. M. Morris, A. L. Wong, A. J. Olson, J. H. Elder and C. H. Wong, *J. Am. Chem. Soc.*, 1999, **121**, 1145.
239. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785.
240. F. Österberg, G. M. Morris, M. F. Sanner, A. J. Olson and D. S. Goodsell, *Proteins*, 2002, **46**, 34.
241. G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell and A. J. Olson, *J. Comput. Chem.*, 2009, **30**, 2785.
242. Y. Zhao and M. F. Sanner, *Proteins*, 2007, **68**, 726.
243. Y. Zhao, D. Stoffler and M. Sanner, *Bioinformatics*, 2006, **22**, 2768.
244. H. Claussen, C. Buning, M. Rarey and T. Lengauer, *J. Mol. Biol.*, 2001, **308**, 377.
245. D. M. Lorber and B. K. Shoichet, *Protein Sci.*, 1998, **7**, 938.
246. X. Barril, C. Aleman, M. Orozco and F. J. Luque, *Proteins*, 1998, **32**, 67.
247. R. Mendez, R. Leplae, M. F. Lensink and S. J. Wodak, *Proteins*, 2005, **60**, 150.
248. X. Fradera and J. Mestres, *Curr. Top. Med. Chem.*, 2004, **4**, 687.
249. J. M. Jansen and E. J. Martin, *Curr. Opin. Chem. Biol.*, 2004, **8**, 359.
250. S. Radestock, M. Bohm and H. Gohlke, *J. Med. Chem.*, 2005, **48**, 5466.
251. K. Brejc, W. J. van Dijk, R. V. Klaasen, M. Schuurmans, J. van der Oost, A. B. Smit and T. K. Sixma, *Nature*, 2001, **411**, 269.
252. R. Abseher and M. Nilges, *J. Mol. Biol.*, 1998, **279**, 911.
253. J. H. Chen, W. Im and C. L. Brooks, *J. Am. Chem. Soc.*, 2004, **126**, 16038.
254. A. P. R. Zabell and C. B. Post, *Proteins*, 2002, **46**, 295.
255. A. Dobrodumov and A. M. Gronenborn, *Proteins*, 2003, **53**, 18.
256. T. Carlomagno, *Annu. Rev. Biophys. Biomol.*, 2005, **34**, 245.
257. G. Martorell, M. J. Gradwell, B. Birdsall, C. J. Bauer, T. A. Frenkiel, H. T. Cheung, V. I. Polshakov, L. Kuyper and J. Feeney, *Biochemistry*, 1994, **33**, 12416.

258. H. Waldmann, I. M. Karaguni, M. Carpintero, E. Gourzoulidou, C. Herrmann, C. Brockmann, H. Oschkinat and O. Muller, *Angew. Chem., Int. Ed.*, 2004, **43**, 454.
259. C. Dominguez, R. Boelens and A. M. J. J. Bonvin, *J. Am. Chem. Soc.*, 2003, **125**, 1731.
260. S. J. de Vries, A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar and A. M. Bonvin, *Proteins*, 2007, **69**, 726.
261. M. Nilges and S. I. O'Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1998, **32**, 107.
262. U. Schieborr, M. Vogtherr, B. Elshorst, M. Betz, S. Grimme, B. Pescatore, T. Langer, K. Saxena and H. Schwalbe, *ChemBioChem*, 2005, **6**, 1891.
263. G. M. Clore and A. M. Gronenborn, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5891.
264. J. Kuszewski, A. M. Gronenborn and G. M. Clore, *J. Magn. Reson., Ser. B*, 1995, **107**, 293.
265. K. Osapay, Y. Theriault, P. E. Wright and D. A. Case, *J. Mol. Biol.*, 1994, **244**, 183.
266. D. S. Wishart and D. A. Case, *Methods Enzymol.*, 2001, **338**, 3.
267. M. A. McCoy and D. F. Wyss, *J. Biomol. NMR*, 2000, **18**, 189.
268. K. Osapay and D. A. Case, *J. Am. Chem. Soc.*, 1991, **113**, 9436.
269. D. Gonzalez-Ruiz and H. Gohlke, *J. Chem. Inf. Model.*, 2009, **49**, 2260.
270. H. Gohlke, M. Hendlich and G. Klebe, *J. Mol. Biol.*, 2000, **295**, 337.
271. M. G. Kendall, *Biometrika*, 1938, **30**, 81.
272. B. Wang, K. Raha and K. M. Merz, Jr., *J. Am. Chem. Soc.*, 2004, **126**, 11430.
273. T. Berg, *Angew. Chem., Int. Ed.*, 2003, **42**, 2462.
274. S. K. Sharma, T. M. Ramsey and K. W. Bair, *Curr. Med. Chem.: Anti-Cancer Agents*, 2002, **2**, 311.
275. A. L. Hopkins and C. R. Groom, *Nat. Rev. Drug Discovery*, 2002, **1**, 727.
276. N. C. Meisner, M. Hintersteiner, V. Uhl, T. Weidemann, M. Schmied, H. Gstach and M. Auer, *Curr. Opin. Chem. Biol.*, 2004, **8**, 424.
277. P. J. Hajduk, J. R. Huth and S. W. Fesik, *J. Med. Chem.*, 2005, **48**, 2518.
278. J. Seco, F. J. Luque and X. Barril, *J. Med. Chem.*, 2009, **52**, 2363.
279. R. P. Sheridan, V. N. Maiorov, M. K. Holloway, W. D. Cornell and Y. D. Gao, *J. Chem. Inf. Model.*, 2010, **50**, 2029.
280. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3.
281. D. E. Clark and S. D. Pickett, *Drug Discovery Today*, 2000, **5**, 49.
282. S. L. McGovern, E. Caselli, N. Grigorieff and B. K. Shoichet, *J. Med. Chem.*, 2002, **45**, 1712.
283. S. L. McGovern and B. K. Shoichet, *J. Med. Chem.*, 2003, **46**, 1478.
284. J. Seidler, S. L. McGovern, T. N. Doman and B. K. Shoichet, *J. Med. Chem.*, 2003, **46**, 4477.

285. S. L. McGovern, B. T. Helfand, B. Feng and B. K. Shoichet, *J. Med. Chem.*, 2003, **46**, 4265.
286. B. Y. Feng, A. Shelat, T. N. Doman, R. K. Guy and B. K. Shoichet, *Nat. Chem. Biol.*, 2005, **1**, 146.
287. J. J. Irwin and B. K. Shoichet, *J. Chem. Inf. Model.*, 2005, **45**, 177.
288. J. L. Wang, D. Liu, Z. J. Zhang, S. Shan, X. Han, S. M. Srinivasula, C. M. Croce, E. S. Alnemri and Z. Huang, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7124.
289. I. J. Enyedy, Y. Ling, K. Nacro, Y. Tomita, X. Wu, Y. Cao, R. Guo, B. Li, X. Zhu, Y. Huang, Y. Q. Long, P. P. Roller, D. Yang and S. Wang, *J. Med. Chem.*, 2001, **44**, 4313.
290. A. A. Lugovskoy, A. I. Degterev, A. F. Fahmy, P. Zhou, J. D. Gross, J. Yuan and G. Wagner, *J. Am. Chem. Soc.*, 2002, **124**, 1234.
291. P. Mukherjee, P. Desai, Y. D. Zhou and M. Avery, *J. Chem. Inf. Model.*, 2010, **50**, 906.
292. A. M. Almerico, M. Tutone and A. Lauria, *J. Mol. Model.*, 2009, **15**, 349.
293. S. Li, J. Gao, T. Satoh, T. M. Friedman, A. E. Edling, U. Koch, S. Choksi, X. Han, R. Korngold and Z. Huang, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 73.
294. N. K. Koehler, C. Y. Yang, J. Varady, Y. Lu, X. W. Wu, M. Liu, D. Yin, M. Bartels, B. Y. Xu, P. P. Roller, Y. Q. Long, P. Li, M. Kattah, M. L. Cohn, K. Moran, E. Tilley, J. R. Richert and S. Wang, *J. Med. Chem.*, 2004, **47**, 4989.
295. Z. Nikolovska-Coleska, L. Xu, Z. Hu, Y. Tomita, P. Li, P. P. Roller, R. Wang, X. Fang, R. Guo, M. Zhang, M. E. Lippman, D. Yang and S. Wang, *J. Med. Chem.*, 2004, **47**, 2430.
296. Y. Gao, J. B. Dickerson, F. Guo, J. Zheng and Y. Zheng, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7618.
297. N. Fujii, J. J. Haresco, K. A. Novak, D. Stokoe, I. D. Kuntz and R. K. Guy, *J. Am. Chem. Soc.*, 2003, **125**, 12074.
298. D. Grandy, J. Shan, X. Zhang, S. Rao, S. Akunuru, H. Li, Y. Zhang, I. Alpatov, X. A. Zhang, R. A. Lang, D. L. Shi and J. J. Zheng, *J. Biol. Chem.*, 2009, **284**, 16256.
299. A. K. Debnath, L. Radigan and S. Jiang, *J. Med. Chem.*, 1999, **42**, 3203.
300. A. L. Bowman, Z. Nikolovska-Coleska, H. Zhong, S. Wang and H. A. Carlson, *J. Am. Chem. Soc.*, 2007, **129**, 12809.