

## On the binding affinity of macromolecular interactions: daring to ask why proteins interact

Panagiotis L. Kastritis and Alexandre M. J. J. Bonvin

*J. R. Soc. Interface* 2013 **10**, 20120835, published 12 December 2012

---

### References

**This article cites 340 articles, 72 of which can be accessed free**

<http://rsif.royalsocietypublishing.org/content/10/79/20120835.full.html#ref-list-1>

 **open access**

This article is free to access

### Subject collections

Articles on similar topics can be found in the following collections

[bioenergetics](#) (17 articles)

[biophysics](#) (287 articles)

[computational biology](#) (257 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)



**Cite this article:** Kastritis PL, Bonvin AMJJ. 2013 On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J R Soc Interface* 10: 20120835. <http://dx.doi.org/10.1098/rsif.2012.0835>

Received: 11 October 2012

Accepted: 21 November 2012

**Subject Areas:**

bioenergetics, biophysics, computational biology

**Keywords:**

dissociation constant, protein interaction models, protein complex modelling, protein–protein docking, scoring functions, structure–affinity relations

**Author for correspondence:**

Alexandre M. J. J. Bonvin

e-mail: [a.m.j.j.bonvin@uu.nl](mailto:a.m.j.j.bonvin@uu.nl)

# On the binding affinity of macromolecular interactions: daring to ask why proteins interact

Panagiotis L. Kastritis and Alexandre M. J. J. Bonvin

Bijvoet Center for Biomolecular Research, Faculty of Science, Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Interactions between proteins are orchestrated in a precise and time-dependent manner, underlying cellular function. The binding affinity, defined as the strength of these interactions, is translated into physico-chemical terms in the dissociation constant ( $K_d$ ), the latter being an experimental measure that determines whether an interaction will be formed in solution or not. Predicting binding affinity from structural models has been a matter of active research for more than 40 years because of its fundamental role in drug development. However, all available approaches are incapable of predicting the binding affinity of protein–protein complexes from coordinates alone. Here, we examine both theoretical and experimental limitations that complicate the derivation of structure–affinity relationships. Most work so far has concentrated on binary interactions. Systems of increased complexity are far from being understood. The main physico-chemical measure that relates to binding affinity is the buried surface area, but it does not hold for flexible complexes. For the latter, there must be a significant entropic contribution that will have to be approximated in the future. We foresee that any theoretical modelling of these interactions will have to follow an integrative approach considering the biology, chemistry and physics that underlie protein–protein recognition.

## 1. Historical perspective

In order to understand our current view of proteins and their interactions, one has to understand how previous knowledge about proteins was accumulated. The present work rests on the shoulders of our predecessors, who essentially determined the route of protein research in today's post-genomic era. It is truly amazing that we are able to routinely characterize and understand protein folding, dynamics and interactions to such an extent and at such detailed resolution. How did we end up with such a vast amount of data for protein molecules? Protein science is exactly 223 years old, which translates into 224 years of trying to understand the nature of protein molecules.

Antoine François, comte de Fourcroy (1755–1809), successfully distinguished several types of proteins back in 1789, including albumin, fibrin, gelatin and gluten. Some years later, Jöns Jacob Berzelius (1779–1848), in a letter to Gerardus Johannes Mulder (1802–1880) dated 10 July 1838, first suggested the term protein to describe a distinct class of biomolecules, stating:

The name protein that I propose for the organic oxide of fibrin and albumin, I wanted to derive from [the Greek word] *πρωτεϊος*, because it appears to be the primitive or principal substance of animal nutrition.

While at Utrecht University, The Netherlands, Mulder described the chemical composition of fibrin, egg albumin and serum albumin [1], which was pioneering work that led to the initial and critical observation that distinct proteins are composed of the same chemical elements: carbon, nitrogen, oxygen, hydrogen, phosphorus and sulphur. Additionally, Mulder successfully characterized

protein degradation products, such as leucine, determining an approximately correct molecular weight of the residue (131 Da) [2].

In 1902, Franz Hofmeister (1850–1922) and Emil Fischer (1852–1919), who spoke at a meeting in Karlsbad shortly after one another, independently announced that proteins are linear polymers consisting of amino acids linked by peptide bonds. The nature of the peptide bond in addition to the successful synthesis of the first optically active peptides by Otto Warburg in Fischer's laboratory were greatly influenced by the search for the 20 building blocks of proteins and prompted the investigation of the last few that were by that time still unknown: amino acid residues were recognized as protein constituents based on isolation from protein hydrolysates in a timeline of approximately 130 years [3] (leucine being the first, identified in 1819 [4], and threonine being the last, identified in 1936 [5]). The primary structure of the proteins was finally elucidated in 1949, when Fred Sanger sequenced bovine insulin [6].

In the late 1950s, John Kendrew determined the first crystal structure, that of sperm whale myoglobin [7], whereas Max Perutz determined the crystal structure of haemoglobin [8]. Both Kendrew and Perutz were protagonists in a blossoming era for X-ray crystallography, working closely together with William and Laurence Bragg, William Astbury and John Desmond Bernal. Interestingly, the crystal structure of haemoglobin is composed of four subunits, all non-covalently bound. Such a quarternary structure did not come as a surprise, since Theodor Svedberg had already determined the molecular weight of haemoglobin and, therefore, its subunit composition in the mid-1920s [9]. Therefore, one should not forget that the discovery of the quaternary structure (QS) preceded the discovery of the primary [6], secondary [10,11] and tertiary structures of proteins [7,8].

Whereas X-ray crystallography has proven to be the primary method for studying the atomic structure of biological macromolecules, nuclear magnetic resonance (NMR) spectroscopy allows both the three-dimensional structure and the dynamics of biomacromolecules to be probed. Kurt Wüthrich with his group outlined a framework for NMR structure determination of proteins in 1982 [12]. Two years later, the first de novo NMR structure of a protein in solution was determined—that of the bull seminal protease inhibitor [13], reported the same year as the Lac repressor headpiece [14]. In the following years, structures of a plethora of biomacromolecules have been determined by X-ray crystallography and NMR and, as of November 2012, approximately 87 000 structures have been deposited in the public repository of macromolecular structures, the Protein Data Bank (PDB database) [15,16].

Although the PDB already includes thousands of macromolecular complexes involved in protein–protein interactions, their importance in defining and orchestrating cellular processes was only recently appreciated [17,18]. A partial explanation could be that the Aristotelian concept of life that 'the whole is greater than the sum of its parts', erroneously considered as the central dogma of vitalism, seemingly contradicted the already established mechanistic view of molecular biology.

In the case of protein synthesis, it was known that macromolecular interactions must play a major role. Still, DNA replication, transcription and translation were unexplored areas in biology at that time and up to now have been

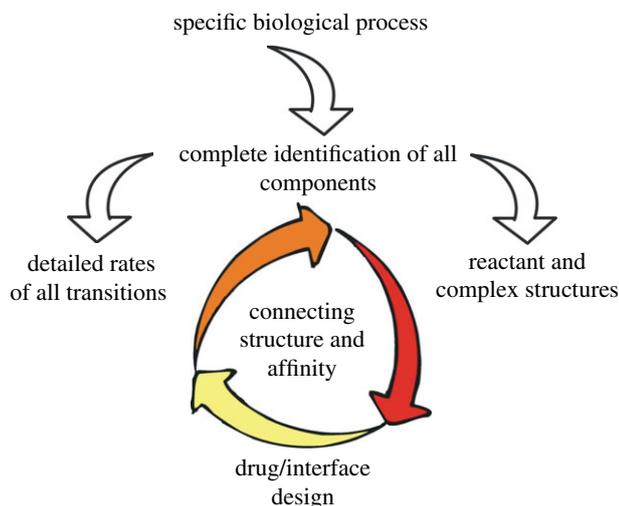
considered active areas of research. On the other hand, complete metabolic processes were characterized in detail, such as glycolysis [19], the Krebs cycle [20], cholesterol and fatty acid biosynthesis [21], which, again, erroneously led the community to believe that interactions were not essentially involved in the cellular metabolism. Subsequently, the dogma 'one gene/one enzyme/one function', framed by Beadle and Tatum [22], was being validated, stating that simple, linear connections are expected between the genotype and the phenotype of an organism. Therefore, up to the 1970s, macromolecular interactions were considered purification artefacts. For example, during the isolation and characterization of enzymes *in vitro*, several experimental difficulties arose as a result of protein–protein interactions, such as co-precipitation, which was believed to be contamination [23].

However, a unique observation back in 1958 by Frederic Richards gradually started to spark the interest in protein interaction phenomena [24]: Richards found that RNase A resulted in a cleaved product, RNase S, when a particular protease was used (subtilisin). RNase S is composed of two molecules, the S-peptide and the S-protein. When these are separated, no RNase activity is observed; however, when recombined in the test tube, the RNase activity is recovered [24]. Richards also foresaw the importance of the interactions of colicin molecules with their macromolecular substrates [25] and laid the foundations for the analysis of macromolecular interactions by implementing the well-known Lee & Richard's [26] algorithm for calculating accessible surface areas of biomolecules. In 1974, Robert Huber's group elucidated the crystal structure of the first protein–inhibitor complex [27]—that of bovine trypsin with its pancreatic trypsin inhibitor. Cyrus Chothia and Joël Janin [28] first characterized the structure and stability factors of the formed interface and concluded that the intrinsic interaction energy was simply proportional to the area of the interface, a first, rather coarse, but critical approximation to understand protein–protein binding. A few years later, in 1978, Shoshana Wodak and Joël Janin [29] implemented the first modelling algorithm for docking protein molecules.

In the following years, an increasing amount of data for protein–protein interactions was accumulated and dogmas about single protein function were being scrambled one by one: For DNA replication, which was thought to be catalysed by a single molecule in the 1960s [30], the involvement of other proteins (e.g. DNA helicase, DNA primase, single-strand binding proteins) was found to be essential for fulfilling this task apart from the polymerase [31]. For protein transport to the mitochondria, more than 20 proteins were identified as critical for this process [32]. In a meeting review published in *Cell* in 1992, Bruce Alberts & Miake-Lye stated that:

... cell biochemistry would appear to be largely run by a set of protein complexes, rather than proteins that act individually and exist in isolated species.

Consequently, to understand how the cell works, a holistic approach needs to be followed (shown in figure 1). Over the last 20 years, this approach has yielded on a daily basis fascinating results in both fundamental [33–39] and applied [40–43] research. The outcome is substantial not only for understanding life at the cellular level, but also for drug design: dissection of protein–protein interactions has opened routes to the production of therapeutics with novel



**Figure 1.** Methodology to follow in protein–protein interaction identification leading to drug/interface design.

functions aiming to cure, for example, amyloidosis-related diseases [44,45] and cancer [46,47].

## 2. Role of protein quaternary structure in a cell

The levels of protein structure were first portrayed by Linderstrøm-Lang & Schnellman [48], which defined QS as being the highest level of structural hierarchy described by the interactions of two or more non-covalently bound subunits that eventually form a functional molecule. QS was first used to designate obligate complexes, such as haemoglobin [8], and its main difference from non-obligate complexes lies in the nature of the interacting subunits: if the individual components of a complex can exist free in solution, then the complex is non-obligate; in contrast, if these subunits constitute an integral part of the structure and cannot be separated (or, if separated, the structure and function of the protein is irreversibly lost) then the complex is referred to as being obligate. Note that the definition of non-obligate and obligate interactions can also depend on the localization (for details, see §5.1).

Several sections in collective books [49–51], original publications [52] and critical reviews [53–58] have concentrated on describing the nature of both obligate and non-obligate interactions, whereas, more recently, reviews about the structure, function and modulation of non-obligate complexes have also appeared [59–62].

In this review, the focus will be on describing the structure and function of non-obligate protein–protein complexes in the context of recent findings, explaining the underlying theory of how and why proteins interact as well as the recently accumulated knowledge for their underlying affinity, describing the efforts to connect QS to binding affinity.

Along with the description of recent findings, fundamental past observations will be assessed and a critical view on modern models will be posed. The main motivation behind this is the central role that protein–protein interactions play in defining the fundamental functional and structural unit of all living matter, the cell. Since the biological function of a protein is defined by its interactions in the cell [63] and inappropriate interactions can lead to diseases such as amyloidoses [44,45] and cancer [46,47], development of

methods aiming to disrupt or modulate protein–protein interactions is critical [64]. Therefore, in order to successfully design drugs or interfaces with predefined properties, knowledge and understanding of binding affinity and its underlying contributing factors is deemed mandatory.

### 2.1. Determination of non-obligate quaternary structure at atomic resolution: how do proteins interact?

A plethora of non-obligate protein–protein complexes have been successfully determined using traditional techniques, such as X-ray crystallography and NMR spectroscopy. These techniques provide a detailed picture of how proteins interact at atomic resolution, meaning that their interfaces (defined as the regions involved in protein interactions) are well characterized and the contributing interactions documented. For example, water molecules important for the interaction can be described, as well as formed salt bridges, hydrogen bonds, degree of complementarity of the two partners directly linked with the strength of the van der Waals interactions, etc. Also, the shape of the interface can be examined and classified as being concave or convex, whereas the biochemical nature of the interface and the rim (the area in its close vicinity) is recognized by observing the contributing amino acid residues. Such analysis is trivial and very frequently used to compare properties of complexes of a different nature [65–69]. Despite that, it has been argued that the sizes of the datasets of derived protein–protein complexes have often been too small, which may lead to statistically unreliable conclusions [70]. Several tools of central importance are routinely used that are able to recognize structural parameters for protein–protein complexes [71], including NACCESS [26] for surface calculations and HBPLUS [72] for recognizing water molecules at the interface and the underlying contacts. Several webservers have also been designed to aid the annotation of macromolecular interfaces [73–77], such as PISA [74] ([http://www.ebi.ac.uk/msd-srv/prot\\_int/](http://www.ebi.ac.uk/msd-srv/prot_int/)), and comprehensive databases compiled, such as PICCOLO [77] (<http://www-cryst.bioc.cam.ac.uk/databases/piccolo>). Recognizing the interfacial region is of particular importance in protein–protein complexes since the biological function of the complex is in most cases directly related to the interactions made [78].

### 2.2. The concept of buried surface area and its inherent limitations

In protein–protein interactions, the buried surface area (BSA) is defined as the surface buried away from the solvent when two or more proteins or subunits associate to form a complex. The most widely used surface calculation method is the solvent-accessible surface introduced by Lee & Richards [26]. In this method, a probe sphere traces the solvent-accessible surface as it rolls over the protein. Protein atoms are assigned their corresponding van der Waals radii. The solvent-accessible surface area traced by the centre of the sphere can be considered as an expanded van der Waals surface of the molecule. In another method, if a water-sized probe sphere touches the protein surface, then this surface is defined as the contact surface (i.e. the contact point instead of the centre of the sphere is used to trace the surface). Since different methods have been developed to calculate and represent the protein surface to date [79–83], the area

calculated is clearly dependent on both the method used and the radii considered for the protein atoms and the probe sphere. For example, different van der Waals radii have been reported for atoms in biomacromolecules [84] and substantial differences in the algorithms used to calculate and represent molecular surfaces have been noted by Michael Connolly (<http://www.netsci.org/Science/Compchem/feature14.html>).

Besides that, another inherent limitation for the calculation of BSAs of protein–protein complexes lies in the fact that proteins do not associate as rigid entities, but may undergo small-to-large conformational changes upon binding. Therefore, in order to calculate BSA one has to know in detail the three-dimensional structures of the unbound states of the proteins that interact, and calculate the BSA according to

$$BSA = \sum_{n=1}^{N_{\text{comp}}} ASA_{\text{free}}^n - ASA_{\text{complex}}, \quad (2.1)$$

where  $ASA_{\text{free}}^n$  indicates the accessible surface area of the unbound molecules and  $ASA_{\text{complex}}$  that of the bound complex.

$$BSA = \sum_{n=1}^{N_{\text{comp}}^{\text{free}}} \left[ \frac{\sum_{a=1}^{N_{\text{conf}}^{\text{free}}(n)} ASA_{\text{free}}^{n,a}}{N_{\text{conf}}^{\text{free}}(n)} \right] - \frac{\sum_{b=1}^{N_{\text{conf}}^{\text{complex}}(n)} ASA_{\text{complex}}^b}{N_{\text{conf}}^{\text{complex}}(n)}. \quad (2.2)$$

However, since proteins undergo dynamic motions directly associated with their function [85] the surface area that is calculated using (2.1) represents an approximate value and not necessarily the expanded van der Waals surface that should be averaged over the surface formed by all conformations of the free reactants (assuming that conformations of the reactants are equally populated for simplification) and the bound structure: where  $N_{\text{comp}}$  indicates the total number of free components in the complex,  $a$  all possible representative conformations ( $1, 2, \dots, N_{\text{conf}}^{\text{free}}(n)$ ) of the free reactant  $n$ , and  $b$  all possible representative conformations ( $1, 2, \dots, N_{\text{conf}}^{\text{complex}}(n)$ ) of the complex.

However, although equation (2.2) is analytical, for simplification purposes, equation (2.1) is used. Hence, in BSA calculations, proteins are currently considered static and, when the unbound structures are not available, the accessible surface area is calculated from the separated components of the complex, therefore considering that proteins bind as rigid bodies. An interesting question about the definition of the functional surface of protein–protein interactions is whether functional solvent molecules or interacting ions and cofactors should be included in the calculations, since solvent has been proposed to functionally define the protein structure [86,87].

### 2.3. Non-covalent interactions formed in the interface and accepted approximations

During the study of the three-dimensional structure of a macromolecular complex in its bound conformation, molecular interactions present in the interface can be annotated. This annotation is an integral part of any structural analysis of a derived complex and has been recently critically reviewed [88]. One of the major inconsistencies found in the literature is the usage of different cut-offs for inter-residue interactions ranging from 5 to 14 Å [89–92]. Because of this, there is no consensus on the geometrical definition of non-covalent interactions [93–95]. Deviations in the cut-offs for specific interactions can also be found in the literature. Furthermore,

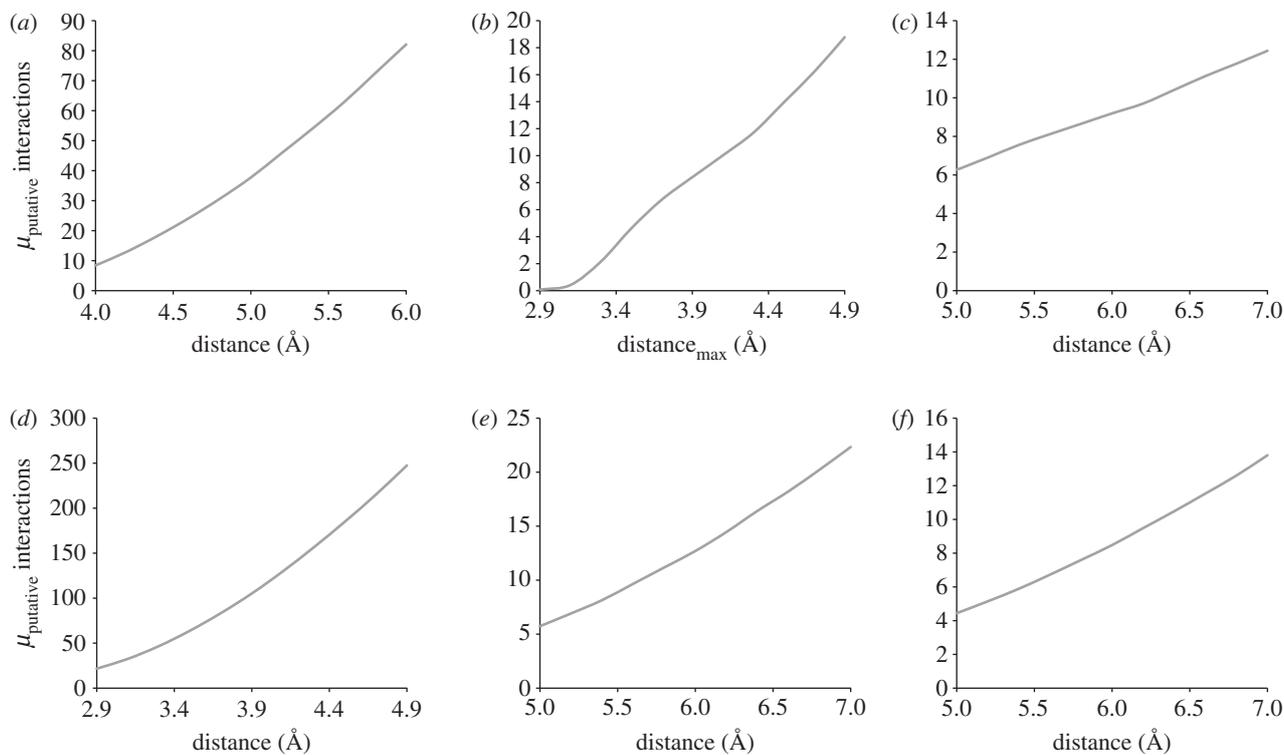
hydrophobic contacts can be analysed via a residue-based criterion (e.g. using the Kyte–Doolittle scale [96]) or an atom-based criterion, where hydrophobic contacts are defined between atoms within 5 Å from each other [77]. The distance between a donor and an acceptor atom to define a hydrogen bond also varies slightly between various web servers [74–77]. Other interactions, such as annotation of aromatic–sulphur or aromatic–aromatic interactions also follow different criteria [76,77] depending on the method used [97–101]. As a consequence, the different cut-offs used for analysing crystal structures hamper a direct comparison of annotated intermolecular interactions in the literature in a large-scale manner. Figure 2 illustrates how the number of interactions found for 195 protein–protein complexes [102,103] substantially changes by varying the cut-off by  $\pm 1$  Å [77]: their number changes as a function of distance in a, not entirely, linear manner. This also indicates that the number of interactions cannot simply be related to the binding strength and used to classify complexes as strong or weak binding, as also highlighted previously [102].

#### 2.3.1. Considerations for solvent effects

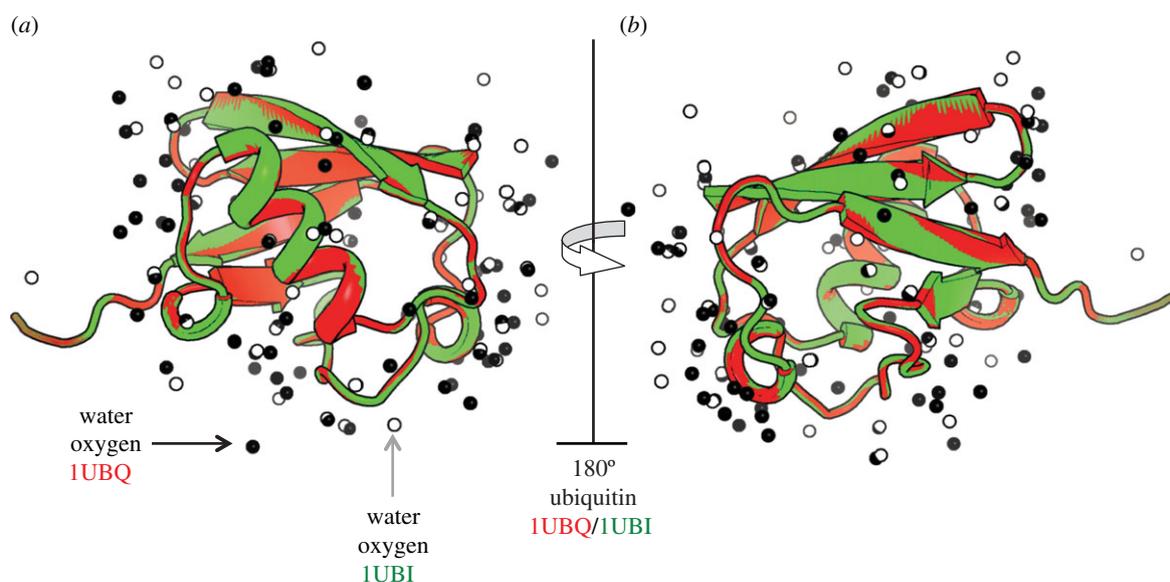
Since the release of the first crystal structure of a heteromeric complex [27]—that of trypsin with the pancreatic trypsin inhibitor (PTI)—the role of water has been clearly demonstrated: the side chain of Asp<sup>189</sup> of trypsin is in contact with the Lys<sup>15</sup> side chain of PTI via water-mediated hydrogen bonds. Its importance is also highlighted in the structure of trypsin in complex with the homologous inhibitor from soybean (STI), where the water molecule is absent, since the salt bridge is formed directly via the bulkier positively charged residue Arg of STI that substitutes Lys<sup>15</sup>. Apart from crystallography, various methods [104] can tackle not only the structure but also the dynamics of water molecules at protein surfaces and at interfaces of protein–protein complexes such as high-resolution neutron diffraction and multi-dimensional NMR. For example, buried water molecules for PTI observed in solution by NMR are in excellent agreement with crystallographic data [105].

Recently, several experimental [106] and theoretical [107] advances have provided deeper understanding in the structure of water around biomolecules. However, inconsistencies between the long-lived residence time of water molecules measured in solution and the NMR structures and positions of water molecules observed in protein crystals still exist [106]. Differences in water structure can even be seen between crystal structures of the same resolution (1.8 Å) and same space group (figure 3*a,b*). In a recent study, it was shown that the appearance of a catalytic water molecule in the electron density obtained by X-ray diffraction depends on whether the structure was determined under cryo- or ambient conditions [108].

Water molecules in the interface of protein–protein complexes may have structural and/or functional roles, depending on their interactions [86,109]. For example, water-mediated hydrogen bonds in an interface can contribute significantly to binding [110–112]. Water buried in the interface, filling interfacial ‘gaps’, has also been frequently reported [86], having an ambiguous role in modulating interfacial properties, since only a few H-bonds are formed and van der Waals interactions seem to dominate [112]. Interfacial water often participates in extensive water networks [113]; the latter have been observed in highly solvated interfaces, such as



**Figure 2.** Change in the number of intermolecular interactions for 195 protein–protein complexes using cut-offs  $\pm 1 \text{ \AA}$ .  $\mu$  corresponds to the average value calculated. (a) Hydrophobic contacts, (b) hydrogen bonds, (c) ionic, (d) van der Waals, (e) aromatic and (f)  $\pi$ -cation interactions.



**Figure 3.** (a,b) Crystallographically determined structures of ubiquitin (PDB entries 1UBQ and 1UBI), along with their corresponding crystallographic water molecules. Ubiquitin is shown in cartoon representation, whereas the oxygen atoms of water are shown as spheres.

in those of colicins in complex with their cognate or non-cognate immunity proteins [114] and in the barnase inhibitor barnase in complex with its cognate and non-cognate partner, barnase [115] and RNase S1 [116], respectively.

Water can also participate in allosteric phenomena [117]. Royer *et al.* [117] established that interfacial water of the dimeric haemoglobin from *Scapharca inaequivalvis* is modulating the molecule's allosteric cooperativity and contributes to fast communication between the subunits via vibrational energy transport that occurs on the 1–10 ps time scale [118]. Even in the self-assembly of amyloid fibrils, water is being considered as an active component in the process defining different interaction pathways [119]. One-dimensional

water wires at the interface of polar amyloidogenic proteins that are gradually expelled mediate the interaction of the forming fibrils [119], whereas, for hydrophobic peptides, the assembly of the two sheets and expulsion of water molecules occur nearly simultaneously [119]. Hydrophobic surfaces bind much faster (nearly 1000-fold) than hydrophilic ones, since trapped water creates a barrier to rapid assembly.

In order to obtain biophysical insights into the role of water in protein–protein interactions during the association process, most theoretical studies on protein folding and association deal mostly with hydrophobic interfaces [120,121], showing that hydrophobic dewetting is fundamental for the interaction. Yet, dewetting must occur rarely *in vitro* and *in vivo* since few

polar residues are enough to prevent the phenomenon [122]. On average, for protein–protein complexes approximately 70 per cent of the interfacial residues are hydrophilic.

The association mechanism of hydrophilic interfaces has only recently been investigated [113], showing that interfacial water may form an adhesive hydrogen-bond network between the interfaces at the encounter complex stage of association and consequently stabilize early intermediates before native contacts are formed. Note that this does not contradict Janin's observations for the percentage of hydration of protein–protein interfaces, which is around 25 per cent [66], since only a few residues will retain their water molecules in the product complex; the others will form hydrogen bonds and salt bridges with other polar residues and/or backbone atoms.

Overall, in years to come, the advent of both experimental and computational techniques to map the structure, position and dynamics of water molecules around proteins will allow the study of water–protein interactions in a more detailed manner, unveiling fundamental roles for water, currently either hypothesized or even unknown [86,109,122], and this in much more complicated environments, such as that of the cell itself [109,123].

### 3. Definition of binding affinity for macromolecular recognition

The binding of two proteins can be viewed as a reversible and rapid process in an equilibrium that is governed by the law of mass action. The binding affinity is the strength of the interaction between two (or more than two) molecules that bind reversibly (interact). It is translated into physico-chemical terms in the dissociation constant ( $K_d$ ), the latter being the concentration of the free protein that occupies half of the overall sites of the second protein at equilibrium.

By convention, the protein present in fixed and limited amounts will be termed the receptor protein (A), whereas the reaction component that is varied will be termed the ligand protein (B).

Certain assumptions inherent to any measurement of a protein–protein interaction should be considered:

- All interactions studied are assumed to be reversible and the association reaction is bimolecular; on the other hand, the dissociation reaction is unimolecular.
- The receptor protein must have a fixed concentration and, therefore, receptor molecules are equivalent and independent (do not interact).
- The interactions are measured at equilibrium.
- The two proteins that are measured in solution do not undergo any other chemical reactions and are assumed to exist only in their free or bound states.
- The measured affinity ( $K_d$ ) is proportional to the number of occupied receptor binding sites.

Therefore, for a simple reversible reaction between proteins A and B, one can write:



and, in more detail,



where [A] and [B] denote the concentrations of the free proteins (reactants), whereas [AB] denotes the concentration of their bound complex (product).  $k_{\text{on}}$  represents the association rate constant, measured in  $\text{M}^{-1}\text{s}^{-1}$ ;  $k_{\text{off}}$  represents the dissociation rate constant.

When the system is at equilibrium,  $K_d$  is defined as

$$K_d = \frac{[A][B]}{[AB]} = \frac{k_{\text{off}}}{k_{\text{on}}}. \quad (3.3)$$

One can re-write equation (3.2) in terms of total concentration of both proteins [A] and [B]. After applying the assumption for the conservation of mass, where

$$[A] = [A_t] - [AB] \quad (3.4)$$

and

$$[B] = [B_t] - [AB], \quad (3.5)$$

and introducing these in equation (3.3), one gets

$$K_d = \frac{([A_t] - [AB])([B_t] - [AB])}{[AB]}, \quad (3.6)$$

and, by re-arranging equation (3.6), this gives the fractional saturation (FS)

$$\frac{[AB]}{[A_t]} = \frac{[B_t] - [AB]}{K_d + [B_t] - [AB]} = \frac{[B]}{K_d + [B]}. \quad (3.7)$$

In other words, and according to equation (3.7), the FS corresponds to the fraction of the molecules of protein A that are saturated with the molecules of protein B.

By assuming that a single binding site is present, a rectangular hyperbola will be visible in a plot of FS  $[AB]/[A_t]$  versus [B]. Instead, one might highlight these binding events using a plot of FS  $[AB]/[A_t]$  versus  $\log[B]$ , or the well-known Scatchard plot, a plot of ligand bound/ligand free.

The Scatchard plot is the traditional method for analysing binding data where the concentration of the ligand [B] is measured. It is described by the following equation:

$$\frac{[B_b]}{[B]} = \frac{-[B_b]}{K_d} + \frac{n[A_t]}{K_d}, \quad (3.8)$$

where a straight line is derived for the simple model (one binding site is present) and  $n$  denotes the stoichiometry of the interaction (in the simple case,  $n = 1$ ) and  $[B_b]$  the concentration of the bound ligand. The straight line's characteristics are:  $x$ -intercept,  $n[A_t]$ ;  $y$ -intercept,  $n[A_t]/K_d$ ; slope,  $-1/K_d$ .

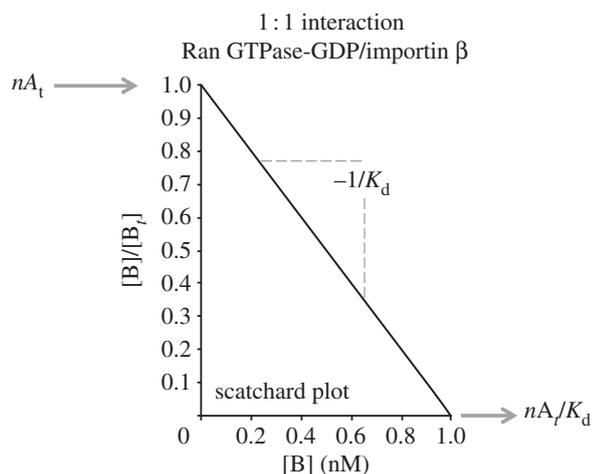
As an example, a simulated Scatchard plot for the 1 nM interaction between Ran GTPase–GDP and importin  $\beta$  is illustrated in figure 4, showing the abovementioned characteristics.

It is quite useful to assess the linearity of the Scatchard plot, since deviation from simple binding (and, therefore, distortion of the linearity of the plot) is expected to be the result of either multiple sites or non-specific binding, which may be difficult to distinguish in practice [124].

The binding affinity can also be translated in physical terms into the Gibbs free energy of dissociation ( $\Delta G_d$ ), which, for an interaction to occur, must be positive,

$$\Delta G_d = -RT \ln \frac{K_d}{c_0} = \Delta H_d - T\Delta S_d, \quad (3.9)$$

where  $c_0$  is the concentration that defines the standard state, being  $1 \text{ mol l}^{-1}$  by conventional criteria,  $R$  is the gas constant ( $8.3144 \text{ J K}^{-1} \text{ mol}^{-1}$  equal to  $1.9872 \text{ cal K}^{-1} \text{ mol}^{-1}$ ),  $T$  is the absolute temperature (kelvin), whereas  $\Delta H_d$ ,  $\Delta S_d$  and  $\Delta G_d$



**Figure 4.** Simulated scatchard plot for Ran GTPase-GDP and importin  $\beta$ . We assume a 1 : 1 interaction, having exactly 1 nM affinity (see text).

denote, respectively, the changes in enthalpy, entropy and binding free energy upon complex dissociation. The binding affinity is related to the Gibbs free energy of association ( $\Delta G_a$ ) as

$$\Delta G_a = -\Delta G_d. \quad (3.10)$$

Both free energies describe all the chemical and energetic factors involved in the dissociation and association reaction, respectively.

The free energy of binding,  $\Delta G_a$ , can be decomposed into two opposing general energies, one favouring the complexation of the unbound partners and one opposing it,

$$\Delta G_a = \Delta G_{\text{bond}} + \Delta G_{\text{entropy}}, \quad (3.11)$$

where  $\Delta G_{\text{bond}}$  and  $\Delta G_{\text{entropy}}$  denote the intrinsic 'non-bonded interaction energy' that includes all chemical forces acting on the interface of the complex and entropy, respectively, analogous to the physical enthalpy and entropy changes, respectively. Such simplification is useful for assessing the energy of macromolecular binding and has been rediscovered several times [28,125,126], from recognizing forces that participate in insulin dimerization [125] to analysis of cooperative effects of protein–protein interactions [127].

### 3.1. Experimental methods and associated errors

Understanding complex biochemical pathways requires quantitative *in vitro* analysis of protein–protein binding [128–130]. For the determination of the FS or binding parameters of a biological reaction between two proteins in such pathways, several methods have been developed [131,132], including NMR spectroscopy, equilibrium dialysis, dynamic light scattering, analytical ultracentrifugation, ultrafiltration, electrophoretic methods, differential scanning calorimetry, homogeneous time-resolved fluorescence, fluorescence correlation spectroscopy/fluorescence cross-correlation spectroscopy, spectroscopic assays, affinity capillary electrophoresis, bilayer interferometry, dual polarization interferometry, static light scattering and microscale thermophoresis. Overall, these methods can be classified in two general categories, namely direct (or separative) and indirect (non-separative) methods [133]. Direct methods measure the actual concentrations of the bound and free proteins, whereas indirect methods imply the concentrations from a signal that is being observed.

Gel filtration, ultracentrifugation, ultrafiltration or equilibrium dialysis are direct methods that can be used to measure binding of protein–protein interactions. Direct methods might be appropriate only for binding reactions exhibiting slow dissociation rates, since the process of separating the bound and free proteins must be faster than the rate of dissociation of the complex. If dissociation and separation of the reactants occur on similar time scales, these methods are inappropriate since the equilibrium will be disturbed by the separation of the reactants [133].

Optical methods, such as absorbance, resonance or fluorescence spectroscopy techniques, belong to the indirect methods, where the assumption is made that the measured signal is directly proportional to the concentration of the product, assuming that the proteins exist in only two states: the free and the bound populations, with each having its unique optical characteristic. Consequently, if  $O_B$  is the signal when protein B is present at a given concentration,  $O_0$  the signal in its absence, and  $O_{\text{sat}}$  the value at saturation of the reaction, one can measure the FS using

$$\text{FS} = \frac{O_B - O_0}{O_{\text{sat}} - O_0}. \quad (3.12)$$

Three of the most frequently used methods to measure the binding affinity of protein–protein interactions will be compared and discussed in more detail in the following, namely isothermal titration calorimetry (ITC) [134], surface plasmon resonance (SPR) [135] and fluorescence-based methods [136]. One should, however, bear in mind that more than 20 methods have been described in the literature for determining biomolecular binding kinetics [137]. The determination of the actual affinity clearly depends on the method used along with its inherent sensitivity and on the strength of the interactions that are being measured.

#### 3.1.1. Isothermal titration calorimetry

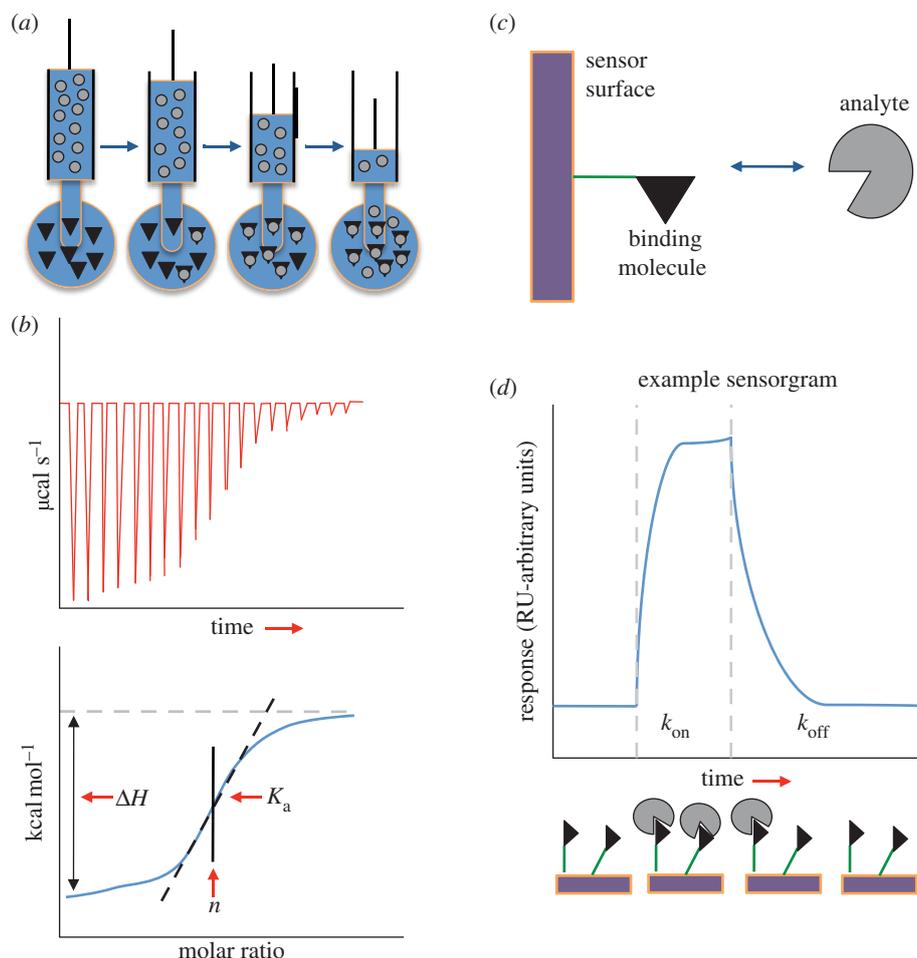
One of the most commonly used calorimetric approaches to study protein–protein interactions is ITC, which measures the heat uptake or release during a biomolecular interaction. An ITC experiment consists of successive additions of protein B to a solution of protein A, the latter contained in a reaction cell (figure 5a). Each addition leads to a specific amount of protein–protein complexes, as dictated by the binding affinity that can be observed by monitoring the heat release (or uptake; figure 5b).

Microcalorimetry reports on the enthalpy of association,  $\Delta H_a$ , that can be related directly to the dissociation enthalpy,  $\Delta H_d$ ; if the titration is performed at different temperatures, changes in heat capacity ( $\Delta C_p$ ) at constant pressure are also reported and are equal to

$$\Delta C_p = \frac{d(\Delta H_d)}{dT} = \frac{d(\Delta S_d)}{dT}, \quad (3.13)$$

where  $dT$  corresponds to the changes in the temperature.

What distinguishes ITC from the other techniques is that, besides measuring binding affinity, it also allows the enthalpy, entropy and change in heat capacity of the interaction ( $\Delta H_d$ ,  $\Delta S_d$  and  $\Delta C_p$ , respectively) to be determined. On the other hand, ITC cannot be used for very low- or very high-affinity protein–protein interactions since the change in heat capacity is not correctly captured by the method. However, some studies have reported affinity data obtained with ITC for very low-affinity complexes [102].



**Figure 5.** (a,b) Isothermal titration calorimetry and (c,d) surface plasmon resonance (SPR) techniques. (a) Titrations used to measure heat capacity changes and (b) calculation of  $K_a$ . (c) SPR method and (d) monitoring of the association/dissociation process of the mobile agent. See text for details.

### 3.1.2. Surface plasmon resonance

SPR is an optical method to measure the refractive index near a sensor surface. In Biacore, particularly, this surface forms the floor of a flow cell through which an aqueous solution can pass under continuous flow (figure 5c). In order to detect a binary interaction, one protein is immobilized onto the sensor surface. Its binding partner (the analyte) is injected into the aqueous solution through the flow cell. As the analyte binds to the immobilized partner, the accumulation of proteins on the surface results in an increase in the refractive index. Measurement of this change is performed and the result is plotted as response units (RUs) versus time (figure 5d). After a defined association time, a solution without the analyte is injected that dissociates the bound complex between the immobilized protein and the partner. During dissociation, a decrease in SPR signal (expressed in RUs) is observed. From these, kinetic constants can be retrieved; however, one should keep in mind that protein immobilization affects the conformational and rotational entropy, and, therefore, association rates. On the other hand, SPR has been shown to be the preferred method for characterizing the kinetics for protein–protein interactions, since most reported  $K_d$ s are determined by this method [136]. However, since diffusion is affected when using SPR, other methods should be used for  $k_{on}$  data collection [138].

### 3.1.3. Fluorescence-based methods

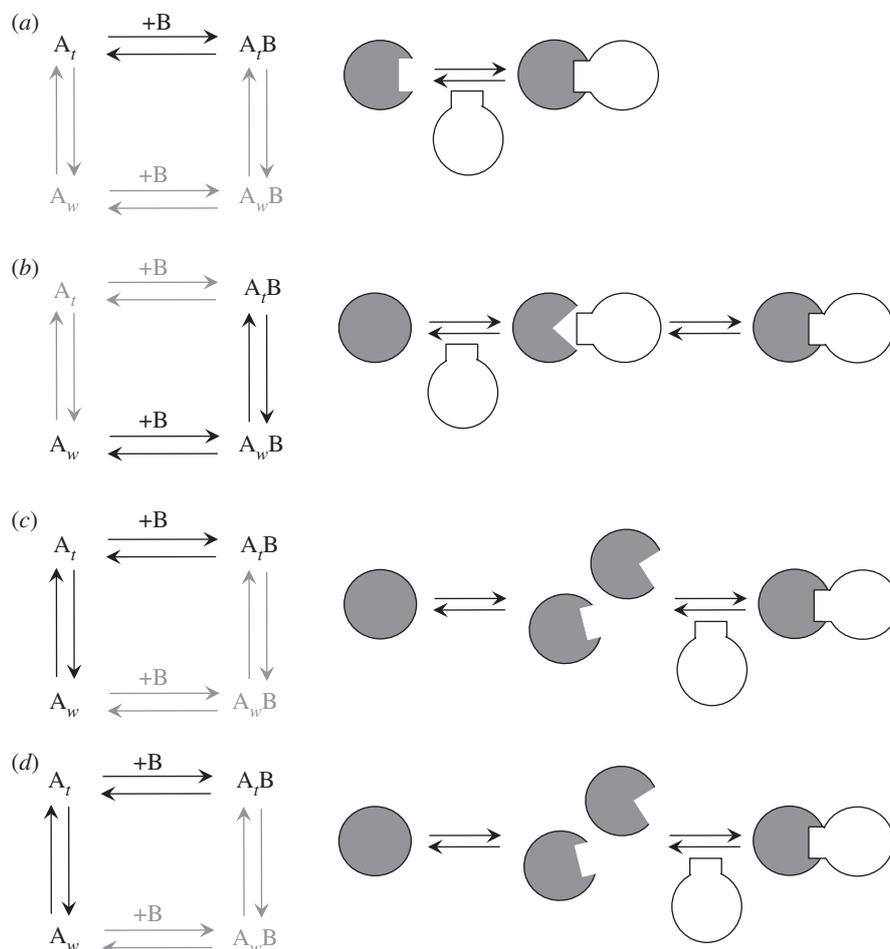
In most of these methods (e.g. fluorescence (de)polarization (FP) or Förster resonance energy transfer, competitive

binding assays are used in which a labelled ligand molecule is bound and subsequently displaced by any of a variety of competitive inhibitors [136]. A small amount of the labelled ligand is first bound to protein A and is subsequently displaced by titrating the unlabelled protein B. In that way, the inhibition constant  $K_i$  of the unlabelled ligand can be measured. Since the comparison is always of the  $K_i$  of the unlabelled inhibitor, the labelled one does not have to be physiological; therefore, any adverse effects that might appear in this system become unimportant. Since the  $IC_{50}$  is the concentration of inhibitor necessary to displace half the labelled ligand, if  $[A_i] \ll K_d$ ,  $IC_{50}$  is related to  $K_i$  by

$$K_i = \frac{IC_{50}}{1 + [L_l]/K_d}, \quad (3.14)$$

where  $[L_l]$  is the concentration of the labelled ligand and  $K_d$  is the equilibrium dissociation constant. For the determination of absolute affinities, measurement of the concentration of the labelled ligand is essential. Such methods, which fall into the category of spectroscopic methods, are very useful because additional information can be derived, such as structural data, binding distances between the fluorophore and the protein, etc. However, these are successful mostly for high-affinity interactions and are limited in studying more complicated equilibria.

This is mainly because the response is not a direct measure of binding, but rather proportional to it [133]. Overall, measurement of an affinity value for protein–protein complexes is always associated with the method used and



**Figure 6.** The three basic mechanisms proposed for molecular recognition: (a) lock and key, (b) induced fit, and (c) conformational selection (dynamic fit). On the left,  $A_t$  and  $A_w$  denote protein A in its tight (binding competent) and weak (binding incompetent) conformation. The chemical pathways that do not exist in each proposed model are indicated by light grey arrows and the way the binding occurs by black arrows. Note that protein B can also undergo conformational transitions; it is shown here rigid for simplicity.

the experimental conditions reported. For example, FP assays are homogeneous assays that give robust results if the size ratio between components of the complex is high [139]. For complexes of different natures, measurements are performed under different temperature, ionic strength and pH. These differences could lead to an observable variation over the reported data.  $K_d$  values are usually reported with standard errors of 20–50%, equivalent to 0.1–0.25 kcal mol<sup>-1</sup> for  $\Delta G_d$  [102]. Changes in temperature (18–35°C) or pH (5.5–8.5) can alter  $K_d$  by a factor of 2 or 10, respectively, corresponding to 0.3–1 in a logarithmic scale. In addition, the stoichiometry of the interaction ( $n$ ) can be determined with a precision of  $\pm 20\%$ , as reported by Wilkinson [133]. Moreover, incorrect corrections for non-specific binding, usage of a labelling method for proteins that may alter the binding behaviour of the complex, presence of non-binding contaminants or of contaminants that might enhance binding, etc. might also hamper the actual calculation of binding affinity. All these potential sources of errors must be treated carefully during measurement [140].

### 3.2. Conceptual models for biomolecular recognition

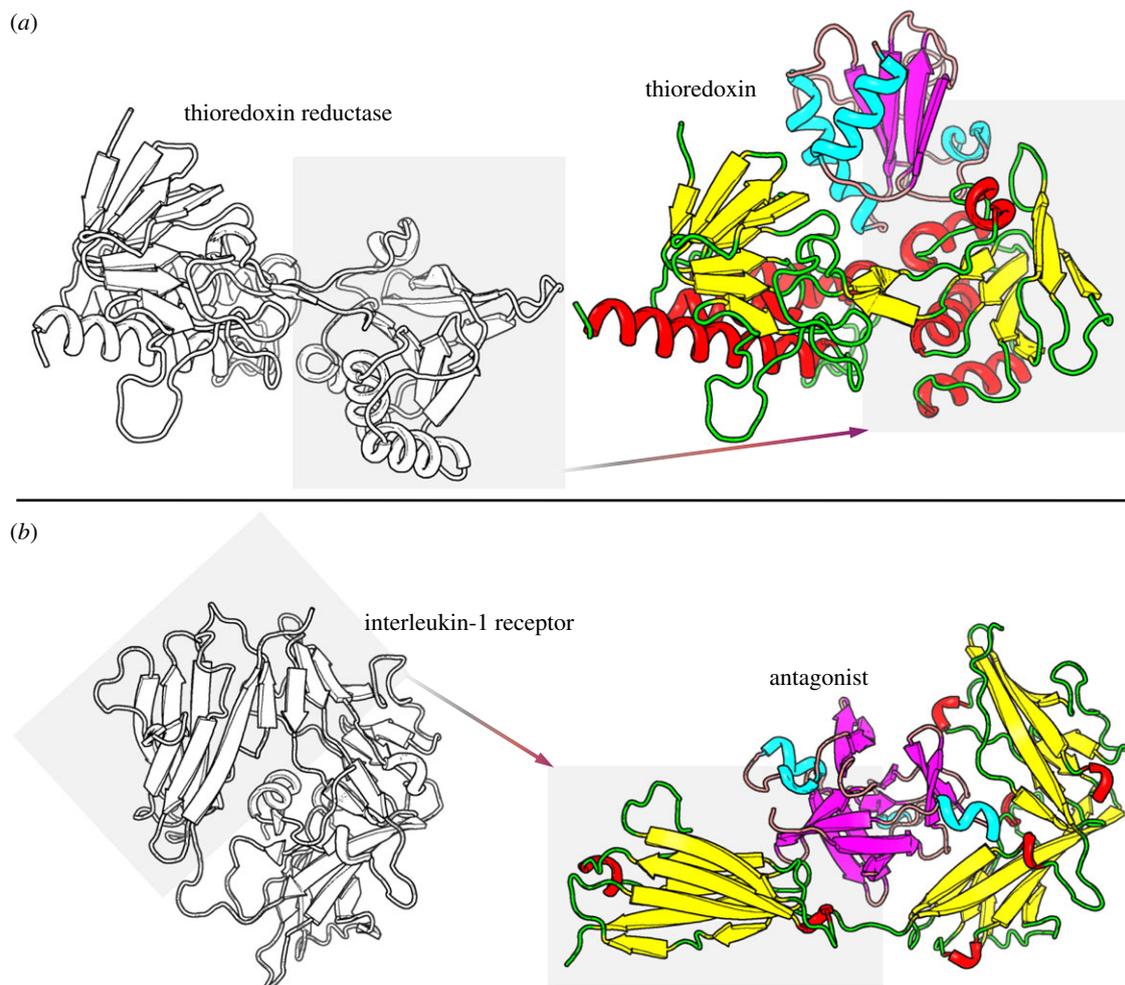
Since molecular recognition is a fundamental phenomenon governing all processes of life, different models that conceptually describe the process have been developed over the

last 130 years [141–146]. Three of the proposed mechanisms to describe binding are shown in figure 6a.

For proteins that interact in a rather rigid manner, a lock-and-key binding might occur [141], as hypothesized in 1894 by Emil Fischer [143]. The complex of trypsin with BPTI [27] is an example of such a lock-and-key mechanism: the interface of the unbound structures is nearly identical to that of their bound conformation (interface root mean square deviation (i-r.m.s.d.) is less than 0.3 Å). These interactions, along with other examples found in the literature [102], show that one plausible mechanism for protein binding is that one protein might be a (near) rigid complementary image of its partner protein.

A second mechanism describing molecular recognition is the induced-fit model, proposed by Koshland [142,147]. In induced fit, binding of one protein to the other induces specific conformational changes that result in the bound complex (figure 6b). The induced-fit model describes that:

- a precise orientation of catalytic groups is required for the reaction,
- proteins might cause an observable change in their binding interface, ranging from small side-chain or surface loop movements to large hinge movement of domains or even folding/unfolding events, and
- these changes will bring catalytic groups into the proper orientation.

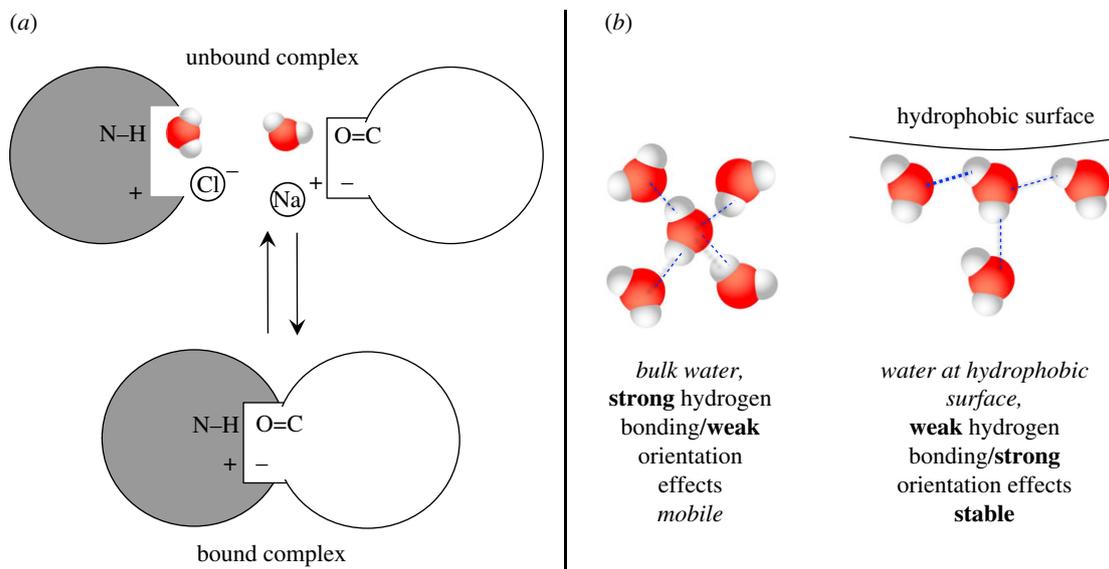


**Figure 7.** Conformational changes in protein–protein complexes; unbound conformations are shown in greyscale, whereas bound conformations are shown in colour code by assigning a secondary structure; (a) the complex between thioredoxin reductase and thioredoxin is illustrated in cartoon representation, and (b) the interleukin-1 receptor in complex with its antagonist; both complexes undergo extensive conformational changes upon ligand binding (see also text).

A third mechanism of molecular recognition is the fluctuation (dynamic) fit [143] (figure 6c), also recently rediscovered and termed (among others) conformational selection [148,149], conformational selectivity [149], population shift [150], selected fit [151] and pre-existing equilibrium [152]. For consistency with current literature, the conformational selection term will be used here. The conformational selection model hypothesizes that the reactants pre-exist in multiple conformations, the best fitting one of which will proceed to form the product complex. Conformational selection has been reviewed by both Koshland & Neet [153] and Citri [154] considering that it is either a useful addition to the induced-fit hypothesis or an alternative mechanism of macromolecular recognition: fluctuating protein molecules (the concept of protein motility) could provide a good basis for the conformational changes that occur during recognition, where one particular form that is able to bind the substrate will further proceed to react. Conformational selection has been observed in several macromolecular recognition events, even coupled with the induced-fit model [155–158], both in a simultaneous [157] and in a sequential manner [158]. Simultaneous occurrence of both mechanisms means that, depending on ligand concentrations, a shift in the recognition mechanism is observed. Hammes *et al.* [157] observed that at low ligand concentrations conformational selection dominates the binding process, whereas, by increasing the concentration, an induced-fit mechanism is observed.

Sequential occurrence of both processes simply implies that the conformation selected from the fluctuating biomolecules undergoes a subsequent structural rearrangement in the intermediate complex that then proceeds to the final bound form [158,159]. Although a clear distinction between induced-fit and conformational selection is hard to observe experimentally, both can be equally plausible for observed conformational changes. Conformational changes are illustrated here for the thioredoxin reductase–thioredoxin complex (PDB ID: 1F6M): thioredoxin undergoes a conformational change of 6 Å in backbone r.m.s.d., whereas the interface of the proteins differs by almost 5 Å, a result of a rotation of the nucleotide-binding domain by 67° (figure 7a). A more notable example is the complex formed between the antagonist of the interleukin-1 receptor and its receptor: when the receptor molecule is in its unbound conformation, its globular shape is maintained but the binding site is hindered by its C-terminal domain with which it strongly interacts. However, in the bound conformation of the complex, the C-terminal domain is displaced following a hinge motion, allowing the antagonist to bind in the active site. This motion results in an r.m.s.d. of the receptor molecule's backbone as large as 20 Å (figure 7b).

The concept of allostery, as originally proposed by Monod *et al.* [144], also falls into the conformational selection mechanism for molecular recognition. It states that proteins may exist in discrete interconvertible states independent of the ligand



**Figure 8.** Water in protein–protein interactions and the explanation of the Chothia–Janin theory for the affinity of protein–protein complexes; (a) intermolecular interactions are recovered in the bound conformation, being already present with the molecules of the solvent and its ions; (b) water at hydrophobic interfaces loses its entropy in comparison with bulk water, which is highly mobile.

structure and/or occupancy; the ratio of these different conformational states is determined by the thermal equilibrium. Presence of ligand merely shifts the equilibrium toward one state or another. This model quantified allosteric events and provided the thermodynamic basis for the dynamic-fit model, elaborated by Burgen [146] and others.

Clarification of which model prevails in macromolecular recognition has not yet been provided since all three distinct conceptual models have been observed experimentally. As a general scheme, one should bear in mind that all three mechanisms may exist both in a simultaneous or in a sequential manner, being recognition mechanisms that can cover a broad spectrum of binding events [157,158,160].

### 3.3. Overall determinants for binding affinity

Various structural determinants of the binding affinity of protein–protein complexes have been proposed throughout the years leading to the construction of different models [28,126,138,161–174], covering nearly all physico-chemical aspects of both the reactants and the product complex. All descriptors for binding affinity must meet four criteria in order to be related to binding affinity:

- They themselves, or their indirect/direct physical effects, must be generated in the complex structure and be absent or different in the unbound conformation of the reactant proteins. If this descriptor or its effect is always constant (its value does not change) between the free and bound forms of the proteins, it must not have any impact on binding affinity else the definition of binding affinity (see §3) will be violated.
- Descriptors that are related to the association of the complex are describing the  $k_{\text{on}}$  rate. Since the  $k_{\text{on}}$  rate is concentration dependent, at least one of the descriptors must also be concentration dependent.
- Descriptors related to the  $k_{\text{off}}$  and, therefore, the dissociation rate of the protein–protein complex must not be concentration dependent, since otherwise the definition of binding affinity would again be violated.

- Descriptors must be causal, since the observation of a correlation does not necessarily imply causality.

#### 3.3.1. Buried surface area

The BSA has been the primary descriptor to be related to binding affinity, and more specifically to the intrinsic bond (or interaction) energy,  $\Delta G_{\text{bond}}$ , according to the Chothia–Janin model [28]. Further justification has been provided by Miller *et al.* [175], who showed that BSA compensates for the area not buried intramolecularly within the potentially unstable subunits.

BSA is a macroscopic descriptor for the hydrophobic interactions of proteins and its magnitude has been estimated to be  $0.025 \text{ kcal mol}^{-1}$  per  $1 \text{ \AA}^2$  of hydrophobic surface removed from contact with water,

$$\Delta G_{\text{bond}} = 0.025 \cdot \text{BSA}. \quad (3.15)$$

This hydrophobic interaction is not only a favourable attraction of hydrophobic surfaces, but also expresses the gain in entropy of the water molecules released upon complexation (figure 8a). Since water molecules are less mobile near hydrophobic regions in the reactants, when the product complex is formed, water molecules will be released into the bulk solvent and gain mobility, and thus entropy (figure 8b). All other non-covalent interactions observed in the interface are theorized as negligible, since proteins are never in vacuum, but are highly solvated when unbound (figure 8a).

Therefore, all interactions of an interface are always satisfied, in both the unbound and bound conformations of the proteins, by contacting solvent molecules or protein residues, respectively. This model however neglects, for example, salt bridges or cation– $\pi$  interactions, because, even if counterions are present, the strength of the interaction might vary depending on the nature of the ion. Despite that, the Chothia–Janin model makes clear that the net contribution of non-covalent interactions, even if zero, must not be ignored because interactions determine the specificity of the complex. A highly specific interaction must reconcile with three criteria, all concerning interface complementarity:

- *Complementarity of ions.* If not all charged groups form salt bridges in the interface, the subunit association would require an ionic bond to the solvent (2–6 kcal mol<sup>-1</sup>) to be broken and, therefore, would highly destabilize the protein–protein complex.
- *Complementarity of hydrogen bonds.* A hydrogen bond that is not satisfied within the protein–protein interface would result in a large change in free energy (0.5–6 kcal mol<sup>-1</sup>) [176].
- *Steric complementarity.* Although van der Waals interactions are weak in nature, the number of atoms in the interface is large, and therefore they contribute to the specificity in a non-negligible manner.

The contribution of macroscopic descriptors of hydrophobic interaction (BSA, apolar BSA, polar BSA, number of atoms in the interface, etc.) to the binding affinity has been validated in a qualitative manner for a large number of complexes assembled [102,177]. For complexes that bind without obvious conformational change, these descriptors exhibit very significant relations to binding affinity, in an, almost, linear manner [102]. On the other hand, the affinities of complexes that undergo conformational changes are not in agreement with the Chothia–Janin theory [102]; therefore, hydrophobic interactions [28] must not be the only determinant for the intrinsic bond energy.

### 3.3.2. Hot spots and anchor residues

Warm- and hot-spot residues represent only a small fraction of interface, yet these residues contribute significantly to the binding free energy [161]. Warm and hot spots are defined as the residues whose mutation to alanine results in a destabilization of the bound state ensemble by 1–2 and 4 or more kcal mol<sup>-1</sup>, respectively. Null spots, in contrast, do not generate such a free energy difference. Experimentally, the contribution of a residue to the binding free energy can be assessed via alanine scanning mutagenesis, initially described by the Wells group [178,179]. A mutation to alanine essentially removes the side chain of the reference residue, leaving only the  $\beta$ -carbon. Subsequent kinetics analyses may provide clues regarding the role played by individual residues in protein binding. Note that a mutation to glycine might theoretically be a better solution because the whole side chain is removed. Nevertheless, mutations to glycine are not preferred as they might introduce local or global changes to the conformation (and dynamics) of the molecule.

Several algorithms have been developed [180–185] to identify hot-spot residues on protein–protein interfaces; these have been recently extensively reviewed [186–188]. Although they can be classified into two general classes (energy-based and feature-based methods), all are built on the following observations for the hot spots:

- They are most often found in central regions of the interface [161].
- Their amino acid composition differs from that of non-hot-spot residues [182].
- They are more conserved than non-hot spots [189].
- They are occluded from solvent [161,190].

Subsequently, the ‘water exclusion hypothesis’ (or O-ring theory [161]) has been proposed that may rationalize the role of the hot spots, whereas coupling of hot spots has also been

reported [191]. Briefly, hot spots that are buried in the interface are surrounded by polar regions of higher packing density. These regions occlude solvent and lower the local dielectric constant and consequently enhance the effect of dipole–dipole or ionic interactions in the formed complex [161,190]. Li & Liu [192] have also hypothesized a double water exclusion hypothesis, where hot spots are always water-free.

Hot-spot residues clearly demonstrate that hydrophobic interactions are not the absolute determinant for binding as described by Chothia and Janin. It is evident that the three complementarity principles mentioned above can be violated. Still, the hot-spot theory is qualitatively in line [190,193] with the Chothia–Janin theory [28] because bulkier residues tend to be found more frequently in hot spots, and these have the largest surface area [194].

Hot spots can affect either  $k_{\text{on}}$  or  $k_{\text{off}}$  (or both) [195], suggesting that the kinetic behaviour of the complex is affected in a different manner by specific hot spots. As an example, mutation of Arg<sup>17</sup> to Ala in the trypsin–PTI complex leads to a significant effect on both  $k_{\text{on}}$  and  $k_{\text{off}}$  rates, whereas Lys<sup>15</sup> to Ala has only a marginal effect on  $k_{\text{on}}$  but a similar destabilization effect on  $k_{\text{off}}$  to the Arg<sup>17</sup> to Ala mutation [196]. The Camacho group has proposed that amino acids that bury the largest solvent-accessible surface area after forming the complex have anchor side chains that are found in the free form in conformations similar to those observed in the bound complex [162]. Such anchors are proposed to reduce the number of possible binding pathways and therefore avoid structural rearrangements at the core of the binding interface. This would allow for a relatively smooth recognition process. Anchor residues must provide most of the specificity necessary for protein–protein recognition [196], whereas other important residues on the interface contribute to the stabilization (and, therefore, the off rate) of the formed complex [196]. Although the observed anchor residues can rationalize encounter complex selection, the transition from the recognition state to the final complex structure is difficult to determine computationally because of the increasing role of short-range interactions that may be harder to evaluate. In general, despite the fact that hot-spot residues are found in protein–protein interfaces, all evidence for their existence comes primarily from rigid and tight protein–protein interactions. This remains to be experimentally explored for transient complexes and complexes showing large conformational changes upon binding in particular [197].

### 3.3.3. Allosteric regulators and non-interface affinity modifiers

Although allostery has been defined initially as the regulation of a protein by a small molecule that differs from its substrate [144], the definition changed to account for regulation of a protein by a change in its tertiary structure/QS induced by a small molecule. In general, allosteric effects are now recognized as changes in the dynamics or structure of a protein by a modulator; the latter can be of any type, from a small molecule to another protein [198]. Such changes can shift the population of the inactive protein to its active form, thereby significantly altering its binding affinity, e.g. the binding of oxygen to haemoglobin. Examples of such ligands can be, besides oxygen, electron donor organic molecules (e.g. ATP), or post-translational modification events, such as phosphorylation, the latter

being the most common covalent protein modification to achieve allosteric control. Such modifications alter the binding affinity of the partners through changes in the dynamics and/or structure of the chains that interact. Therefore, not only interfacial or rim regions can affect the binding affinity of protein–protein interactions, but also modifications of sites remote from the interfacial region through any possible mechanism of allosteric regulation.

#### 4. Structure prediction of macromolecular complexes: is the docking problem still unsolved?

Although current structural biology tools have broadened our knowledge in single protein structure, function and dynamics, the situation differs substantially in the case of protein–protein complexes: owing to experimental limitations in probing protein–protein interactions [199] and solving the structure of biomolecular complexes [200] complementary computational approaches are often needed to assist experimentalists in investigating how two proteins of known structure interact and form a three-dimensional complex.

Protein–protein docking algorithms have been developed for this purpose. They use geometric, steric and energetic criteria to predict the atomic structure of a complex [64,201,202]. Every docking program incorporates two key elements:

- the search algorithm that samples configurational and conformational degrees of freedom and
- the scoring function that ranks the generated solutions.

Although predicting the structure of a complex by docking should be relatively simple for proteins that bind with near-rigid body manner and have highly complementary interface regions (such as trypsin–PTI [27] or barnase–barstar [115]), this is clearly not the case. Finding a correct solution for a biomolecular interaction at atomic resolution can be influenced by several factors inherent to any simulation of biomolecular recognition:

- Proteins are not static structures, as explained in §3.2. Their highly dynamic nature can cover the entire scale of conformational changes upon binding from small side-chain reorientations to unfolding/folding transitions. Next to that, different motions of the protein molecules can be exhibited in solution, such as hinge motions [203], secondary structure rearrangement [204], or even high plasticity of the interfacial region [205]. Although several methods can be used for predicting protein dynamics and/or conformational changes [206], none has been shown to perform reasonably well for proteins with different motions [207]. For example, protein motions can be experimentally monitored within a time scale of femtoseconds (e.g. with neutron scattering) to more than a second (e.g. with SAXS, SANS or H-D exchange), whereas molecular dynamics simulations can reach up to milliseconds (but not in a routine manner, being rather limited to nanoseconds–microseconds in most cases) for systems of small to medium size [208].
- The binding site is not always conserved or cannot always be identified [209–211]. Again, results show that most recent interface predictors can distinguish an interface

with fair accuracy [188]. However, for weak transient protein–protein complexes, interface prediction might fail [188,210,211].

- Current docking methods cannot distinguish whether two proteins will bind or not, i.e. predict the binding affinity. Docking programs will always yield some answer, independently of the affinity of the protein–protein interaction [212]. Recent studies have highlighted this fact, but, to date, no single docking program has been shown to be successful in identifying native complexes in cross-docking studies, except in the case of highly complementary interfaces [212–215]. Cross-docking is defined as the all-against-all binary docking procedure in which all combinations of proteins are docked to each other and the native complexes must be predicted.
- Scoring, defined as the selection of a preferred solution from the pool of generated conformers, has greatly improved during recent years [215], driven, among others, by blind prediction experiments such as CAPRI [216], the Critical Assessment of PRedicted Interactions (<http://www.ebi.ac.uk/msd-srv/capri/>). There are even strong critiques about scoring [177,217], even noting that it might be nearly random [218].

Most docking methods are successful for proteins that undergo minor-to-medium conformational rearrangements upon binding. For these systems, scoring functions can identify near-native models that can be subsequently refined [219,220]. Next to that, implementation of novel clustering algorithms [221–223] (clustering refers to the identification and classification of similar docking predictions into clusters) is allowing more efficient analysis of similar solutions, reducing both the computational time and the heterogeneity that could hinder identification of near-native poses.

Recently, there has been a trend in docking simulations to incorporate available experimental information into the docking and/or scoring process. This can dramatically reduce the conformational space to be sampled [202,224,225]. Such information can be used either *a priori* in docking, and therefore drive the docking procedure [226,227], as was originally done in HADDOCK [224], or *a posteriori*, meaning that generated solutions are filtered according to the experimentally observed attributes of the complex [225,228,229]. Recently, more groups are integrating experimental data coming from different sources and the idea of integrative docking [230,231], originally described in the initial HADDOCK publication [224], has become a matter of great importance in current molecular modelling research [232,233]. Integrative docking can be used either for modelling large macromolecular complexes [234], such as the nuclear pore [235] or other cellular machineries, using for example experimental data such as electron density maps [236], or for the detailed characterization of macromolecular assemblies of lower molecular weight using rather classical experimental information from NMR [224]. As an example, approximately 100 biomolecular structures of complexes determined using HADDOCK [236] in combination with various amounts of experimental data (mainly NMR) have been deposited into the PDB [15] as of November 2012.

Although docking is a powerful technique to predict the structure of a complex, based on its known constituents, prediction of the complex based on homology, the so-called template-based methods, is now rapidly increasing [237–239], as illustrated by novel theoretical applications

[211]. The Vakser group has recently claimed [211] that templates exist for nearly all complexes of structurally characterized proteins in the PDB, although the authors also report that such observations have not been validated for targets released during the CAPRI experiment. Also, Barry Honig's group has already shown that homologous interfaces can be identified for a vast number of protein–protein complexes and that the expected interface should, in principle, look similar to related ones that have been crystallographically determined [210]. This is, however, not always the case [240,241]. For example, the exact interaction geometry is less likely to be conserved as illustrated by the homologous complexes of the chemotaxis histidine kinase CheA with its phosphorylation target CheY for *Escherichia coli* and *Thermotoga maritima* [242]: in this system, a rotation of approximately 90° is observed between the formed interfaces [242]. In general, however, close homologues (30–40% or higher sequence identity) have been shown to interact in a rather similar manner [243,244].

#### 4.1. Is scoring in protein–protein docking related to binding affinity?

Several models have been developed to date for predicting the energetics of macromolecular complexes [28,126,138,161–173,245]. Although some have been very successful on small training sets [126,163], and even coupled to successful docking predictions [246,247], the published models did far less well on larger datasets [168,169,177] and their predictive value remains, in general, poor [177].

For algorithms developed for protein–protein docking coupled with binding affinity prediction, the classical model of Horton & Lewis [126], aimed at predicting binding affinity by decomposing the interface into its polar and apolar BSA, showed a very strong correlation with experimental measurements and crystal structures that were available at the time it was developed [126]. Nowadays, this model is clearly insufficient for binding affinity prediction, since the BSA is moderately correlated with the binding affinity, even for rigid binders ( $r = 0.54$  for 70 complexes) [102]. Another example is the algorithm based on the Freire equations [245] for describing binding free energy and modified for predicting binding affinity of a protein–peptide interaction by the Holmes group [247]. The algorithm did fairly well in predicting the actual energy of the reference structure even when coupled with docking; however, a lot of non-native poses generated had equivalent binding affinities, a common problem. The Holmes function assumes that the complex binds without any conformational change [245]. This contrasts with the current view of protein–peptide recognition indicating that, next to the multitude of conformations that a peptide can adopt in solution, folding events occasionally happen upon binding [61]. Another binding affinity predictor coupled with docking is the one developed by Ma and co-workers [165]. Their function ranked and scored the docking results for 10 protein complexes and, while it showed encouraging results, it did not succeed in ranking native solutions first [165]. As far as scoring functions in protein–ligand docking are concerned, these have been optimized mainly for drug design purposes. This means that an estimate of the binding affinity of the ligand can be obtained only in a qualitative and relative manner and for structurally similar ligands. In contrast, protein–protein docking scoring

functions have not been developed for predicting binding affinities [177], but rather for identifying the best solutions. Top-performing scoring functions in protein–protein docking [224,248–250] have proven to be reasonably reliable against blind cases in the CAPRI experiment [216,251], being able to identify models close to the experimentally determined ones. However, the same functions poorly predict experimentally measured binding affinities [177]. Next to that, scoring functions are not yet able to distinguish binders from non-binders, as shown by cross-docking simulations. A large-scale effort to predict designed interfaces that do actually bind was made by 28 different groups in a recent CAPRI experiment [168]. Results show that the algorithms can efficiently distinguish binders corresponding to experimentally determined structures from non-binders with designed interfaces. However, all scoring functions failed to predict the designed interface that actually binds from the remaining designs (86 in total) that do not.

#### 4.2. Structure–affinity models for protein–protein binding affinity prediction

Various sophisticated approaches for estimating the affinity of protein–protein interactions have been developed to date [252], some of which also include elaborate models that approximate the energetic contributions of the solvent [253]. However, in the context of macromolecular docking, where thousands of models may be generated, these methods are computationally prohibitive. Alternative, more approximate methods that mostly relate to changes in the solvent-accessible surface area upon binding have been proposed instead and these will be discussed in the following.

Since the initial model of Chothia & Janin [28] for predicting the interaction energy of protein–protein complexes, an extensive binding affinity benchmark has been assembled [102]. This dataset includes 144 protein–protein complexes of different affinities and amount of conformational changes to serve as a catalyst for coupling docking results to binding affinity prediction, or just for deriving new binding affinity predictors. Three original algorithms have been developed to date using this benchmark [169,170,172].

One has been developed using descriptors covering all possible combinations of residues in the interface for different binding conformations of the complexes (840 descriptors for 144 complexes in total) [172]. Using a genetic algorithm, these descriptors could be reduced to 378, most of which describe hydrophobic and steric interactions. This number is still much higher than the number of experimental data, indicating possible over-fitting.

Moal *et al.* [169] have designed a machine learning approach, combining four different machine learning methods. Although their results are fairly good for the training set, when the four methods were combined using a consensus approach, they yielded a correlation coefficient  $r$  with experimental measurements of 0.55, similar to the one that the simple BSA shares with the affinity of rigid complexes [102]. Another multiple regression model from the Weng group [170] exhibits a slightly higher correlation ( $r = 0.63$ ). However, the predictive power for affinities of antibody–antigen complexes is insignificant ( $r = 0.24$ ). Both methods cross-validated their algorithms using leave-one-out-cross-validation (LOOCV). The idea behind this cross-validation method is to predict the affinity of a single protein–protein complex from the dataset, based

on the optimized regression equation derived from all other complexes. There are some concerns about LOOCV:

- It tends to include unnecessary components in the model and has been shown [254] to be asymptotically incorrect.
- It does not work well for data with strong clusterization [255].
- It underestimates the true predictive error [256].

The Weng group used all the data for training [170]. No independent test set for validating the model was assembled. The model developed by Moal *et al.* [169], who did use an independent test set, did not hold any predictive capacity on this test set.

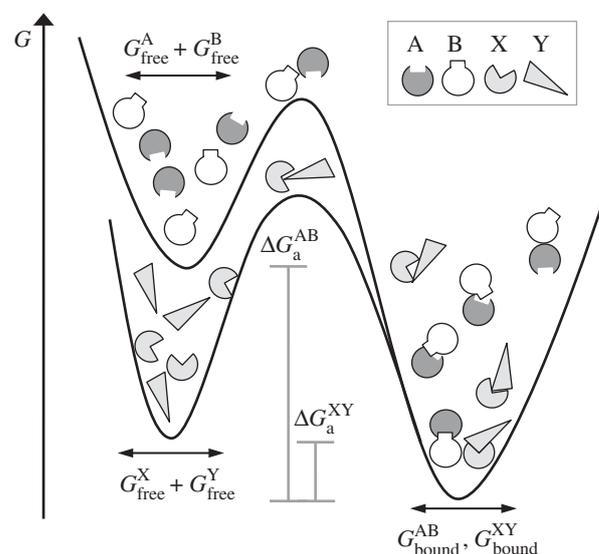
#### 4.2.1. Possible reasons for the limitations of current scoring and affinity prediction models: is there a theoretical prediction limit?

Different possibilities can account for the poor prediction of binding affinities using current biophysical models:

- The quality of the experimental data or the crystal coordinates might be ambiguous.
- Very few, if any, of the present models do account for conformational changes taking place upon binding or for the presence of cofactors that might be needed for binding.
- Allosteric regulation or more complicated kinetics of the complex (two-state kinetics, etc.) might hinder actual predictions. Current models only account for the simplest of the mechanisms—the lock-and-key model, as described in §3.
- Effects of pH, temperature, concentration and solvent are usually ignored.
- The performance (especially for affinity prediction models) depends on the quality and size of the set of experimental data used for testing, as well as on the diversity of the biological systems they represent.
- The current models only account for properties of the interface [102], or, rarely, from the rim region—the latter, if included, only for  $k_{\text{on}}$  prediction [138,257]. None account for contributions from the non-interface surface, which can play a significant role in modulating affinity (see §3.3).
- A final possibility is that linking a structure that has been determined in its crystalline state with the affinity that has been measured in solution state can introduce ambiguities in the derived results because of the different natures of the two states.

Overall, the ideal prediction limit that can be set for structure–affinity models (assuming that all modelling ambiguities are eliminated and results are only dependent on the measured data) must be within the experimental error, which, for a large dataset, can change  $K_d$  by a factor of 10–50, and  $\Delta G_d$  by 1.4–2.3 kcal mol<sup>-1</sup> [102,172].

Finally, one of the central reasons for current models' limitations could well be that the current scoring functions do not account for the underlying energetics of the free components. Figure 9 illustrates this point: assuming two different protein–protein complexes, A–B and X–Y with similar energies of their bound state but different energies of their free states, any model considering only the



**Figure 9.** Schematic of the energy landscape of two different protein–protein complexes.

bound state will predict similar binding affinities for those two complexes

$$\Delta G_a^{AB} = \Delta G_a^{XY}, \quad (4.1)$$

while, experimentally, they will have different affinities owing to the differences in their respective free states,

$$\Delta G_a^{AB,XY} = G_a^{AB,XY} - (G_{\text{free}}^{A,X} + G_{\text{free}}^{B,Y}). \quad (4.2)$$

The free state contribution is typically neglected in docking.

While docking scoring functions might not perform well in affinity prediction, this does not imply that they fail in scoring docking poses for which they have been developed. Indeed, most do show a strong performance in ranking and selecting high-quality models in the CAPRI competition [216,251].

An ideal scoring function that could also predict binding affinity should, in principle, be able to (indirectly or directly) account for the free energy of the unbound partners. Some binding affinity prediction algorithms can reasonably well describe the energy of a (near) rigid binding complex [258]. However, predicting the binding affinity of non-rigid binders will require a more detailed statistical–mechanical treatment in which the full ensemble of unbound structures for each partner, and their contribution to the free energy of the free state, should be considered. Such an approach should, in principle, be able to deal with more flexible molecules. A full description of the free energy conformational landscape of highly flexible or even (partially) unfolded molecules will remain out of our reach for the near future.

Overall, models developed to date describe the thermodynamics of an association reaction by its product only, ignoring reactants and possible accompanying structural changes. Novel functions will have to be developed that can predict the dissociation constant within the experimental error in order to have an actual use in modern drug discovery for protein–protein interactions. The availability of a protein–protein binding affinity benchmark [102] should foster the development and improvement of binding affinity prediction algorithms. Hopefully, in the not too distant future, binding affinity prediction and scoring will start to converge.

### 4.3. Prediction of kinetic rates

The association of protein–protein complexes is dictated by the rotational and translational diffusion of the partners, their surface properties, the electrostatic interactions that guide the interaction, as well as the solvent properties, which are, for example, at the origin of the hydrophobic effect. Several simple models have been constructed to predict  $k_{\text{on}}$ , mostly based on the Einstein–Stokes equation [259] and Poisson–Boltzmann calculations [260]. Although the limit of collision rate is approximately  $10^{10} \text{ M}^{-1} \text{ s}^{-1}$  (calculated by the Einstein–Stokes equation), no single protein–protein complex can achieve this without the aid of electrostatic steering [257]. This limit is three to six orders of magnitude above typically observed association rates, highlighting that most collisions do not lead to fruitful association. Much work has been done on the prediction and improvement of  $k_{\text{on}}$  rates especially for complexes whose association is assisted by charge interactions. Studies have revealed that enhancing electrostatic steering leads to a substantial increase in  $k_{\text{on}}$  [261–263], reaching the limits of the diffusion collision rate. One of the most recent models proposed for predicting association rates is TransComp [138]: it implements the transient-complex theory for predicting  $k_{\text{on}}$  and simulates the formation of a transient complex via diffusion where proteins have near-native separation and relative orientation but have not yet formed short-range interactions. The theory predicts that  $k_{\text{on}}$  is defined as

$$k_{\text{on}} = k_{\text{on}}^0 \exp\left(-\frac{\Delta G_{\text{el}}^*}{k_{\text{B}}T}\right), \quad (4.3)$$

where the basal rate constant for reaching the transient complex by random diffusion is included and the electrostatic interaction free energy of the transient complex. A moderation factor  $f$  is applied to  $\Delta G_{\text{el}}^*$ , when the latter is very negative, to correct for overestimation of  $k_{\text{on}}$

$$f = \left[1 + 10^{-4} \exp\left[-\frac{\Delta G_{\text{el}}^*}{k_{\text{B}}T}\right]\right]^{-1}. \quad (4.4)$$

The transition-state theory applied to protein–protein  $k_{\text{on}}$  rate prediction has been tested on 49 protein–protein complexes with known  $k_{\text{on}}$  rates ranging from  $2.1 \times 10^4$  to  $1.3 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$  [138]. The correlation between the predicted and experimental  $\log k_{\text{on}}$  has an  $r^2$  of 0.72, and the r.m.s.d. is 0.73, corresponding to a fivefold error in  $k_{\text{on}}$  prediction.

The method is valid so far for complexes for which the association rate is diffusion limited ( $k_{\text{on}} > \sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ) and the reactant proteins undergo negligible, if any, conformational rearrangements.

The Schreiber group developed a  $k_{\text{on}}$  prediction, the PARE function, 13 years ago [264], yielding comparable results to TransComp discussed above (G. Schreiber 2012, personal communication). Briefly, in PARE,  $k_{\text{on}}$  is determined using

$$\ln k_{\text{on}} = \ln k_{\text{on}}^0 - \frac{\Delta U}{RT} \left(\frac{1}{1 + \kappa a}\right), \quad (4.5)$$

where  $\ln k_{\text{on}}^0$  is the basal on rate of the interactions, and the electrostatic and salt influence is explicitly considered;  $\Delta U$  is the electrostatic energy of the interaction,  $R$  is the gas constant and  $T$  is the temperature.  $a$  is set to 6 Å and  $\kappa$  is the Debye–Hückel screening parameter relating to the ionic strength of the solution.

$U$  in equation (4.5) is calculated using

$$U = \frac{1}{2} \sum_{ij} \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \frac{e^{-k(r_{ij}-a)}}{1 + \kappa a}, \quad (4.6)$$

where  $i$  and  $j$  are atoms bearing charges and  $\epsilon$  is the dielectric constant of the medium.

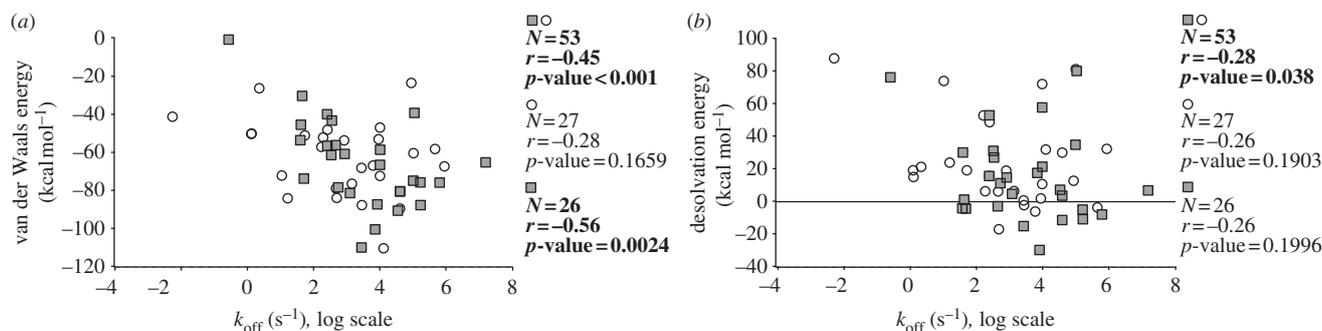
$\Delta U$  is therefore calculated by

$$\Delta U = U_{\text{complex}} - (U_{\text{protein1}} + U_{\text{protein2}}). \quad (4.7)$$

Note that, for proteins that bind with large conformational changes, disordered proteins being at the far end of the spectrum,  $k_{\text{on}}$  determines the binding affinity to a higher extent than  $k_{\text{off}}$  [265], whereas, for rigid complexes,  $k_{\text{off}}$  is the major determinant for binding affinity [265].

Engineering proteins to achieve desirable kinetic rates is non-trivial [257,263], even in non-crowding conditions [266]. For example, especially for protein–protein binding affinity engineering, charges have been shown to play multiple and complex roles in binding [267,268]. When present in remote areas from the interface, they could lead to the formation of non-specific complexes. For example, Tiemeyer *et al.* [269] showed that the surface charge distribution is very important for the orientation of proteins on lipid membranes. Significant effects of charges on  $k_{\text{on}}$  could be sometimes concomitant with effects on  $k_{\text{off}}$ , indicating that the association rate might be difficult to modulate in a significant and controlled manner independently of the dissociation rate [266]. This becomes even more challenging in *in vivo* conditions, where macromolecular crowding can also affect the association rates of protein–protein complexes [123,270–274]: increasing the rates by increasing the effective concentration, and decreasing the rates by decreasing the diffusion of the particles. In recent work, Ando & Skolnick [270] quantified the significant role of hydrodynamic forces in macromolecular motion and Elcock [271] highlighted their importance in protein–protein binding. Another mechanism that can affect  $k_{\text{on}}$  by altering the transition state is the introduction or deletion of steric clashes during association [275]. For example, the association rate constants of the IFN<sub>2</sub>–IFNAR2 complex changed when Ala<sup>19</sup> of IFN<sub>2</sub> (located at the interface) was replaced by a Trp [275]: this mutation introduced a repulsive interaction, resulting in a reduced  $k_{\text{on}}$ . However, in parallel, it also reduced  $k_{\text{off}}$  by the formation of a favourable interaction with Trp<sup>100</sup> on IFNAR2. While substantial progress has been made towards rationalizing the association effects [113,138,171,265,276], dissociation events are still not well understood. For  $k_{\text{off}}$ , breaking of short-range interactions between proteins and interfacial properties should be the rate-limiting step. The rate at which the two proteins diffuse away from each other (which will decrease with increased long-range electrostatic interactions) does not seem to affect  $k_{\text{off}}$  much [257]. However, van der Waals interactions only partially correlate with  $k_{\text{off}}$  rates when the dataset of Zhou and the affinity benchmark [98] are considered, and only for near-rigid binders (figure 10a).  $k_{\text{off}}$  should also depend on complex–solvent interactions, since in macromolecular crowding conditions hydrodynamic effects are dominant [270,277]. Indeed, a significant correlation is calculated between  $k_{\text{off}}$  and the desolvation energy, independently of the conformational change (figure 10b).

Currently, only one model has been proposed to predict the  $k_{\text{off}}$  of protein–protein interactions with reasonable



**Figure 10.** Correlations between some energetic components of the HADDOCK score [217] ((a) van der Waals interactions; (b) desolvation energy) and experimental  $k_{\text{off}}$  for 54 protein–protein complexes [98,133]. (Near) rigid binders are shown by grey squares, whereas flexible binders are shown by white circles.  $r$  denotes the correlation coefficient, whereas the  $p$ -value denotes the corresponding  $p$ -value ( $p$ -value < 0.05 is considered significant). Significant correlations are highlighted in bold.

accuracy [173]. The authors provided separate models for predicting  $k_{\text{on}}$ ,  $k_{\text{off}}$  and  $K_{\text{d}}$ , but the properties determining the  $K_{\text{d}}$  are different from those coming from the determinants after division of  $k_{\text{off}}$  by  $k_{\text{on}}$ . Also, for calculation of  $k_{\text{on}}$  rates the bound complex was used. This is counterintuitive since  $k_{\text{on}}$  describes the association of the unbound proteins.

## 5. Protein–protein interactions *in vivo*: the p53 example

Protein–protein complexes employ all kinds of attributes that large macromolecules may have in order to accomplish their functions within the cell. *Promiscuity*, *specificity*, *selectivity* and *binding affinity* are factors that can modulate protein–protein recognition and their combination is unique and case specific. These are defined as follows:

- *Specificity* is the ability of a protein to bind a single partner protein for performing a task.
- (*Binding*) *affinity* indicates the existence and strength of an interaction between proteins.
- *Promiscuity* (cross-reactivity/multi-specificity) denotes the ability of a single protein to perform multiple functions, thereby interacting with more than one partner in a specific manner.
- *Selectivity* defines a protein that is binding/using a range of other proteins, but some better than others.

Designing a protein–protein complex with a preferred attribute is very difficult since all these attributes are related to each other [64]. The tumour suppressor protein p53 is a great example for such a combination of properties [278]. p53 is an important hub in multiple signalling networks and is the protein most frequently involved in human cancer. It has been described as ‘the guardian angel of the cell’ [278,279]. p53 has a highly versatile structure, featuring every possible conformation, from ordered secondary structure elements and well-defined folds to completely disordered regions. Its core domain is always folded and binds to DNA and a few other proteins [280], whereas its two flanking regions are mostly in a disordered state undergoing disorder-to-order transitions [280–283]. These may bind hundreds of signalling proteins. A sequence segment within one of these regions exhibits chameleon features [282], meaning that it can adopt three different ordered conformations, excluding loop orientation ( $\alpha$ -helix,  $\beta$ -sheet with

flanking strands, beta-turn-like), depending on the partner with which it interacts. Therefore, in the cell, interactions of a specific protein binding site with many partners, such as the rigid core domain of p53, are likely to be mutually exclusive, resulting in competition for interactions among alternative partners [281]. Such competition must be a critical determinant for the specificity of the underlying interactions. The selectivity of such binding sites is determined by the relative binding affinities of alternative interaction partners and by the local concentrations of each protein. However, selectivity is extremely difficult to predict, either *in vivo* or *in vitro*, since a number of factors (such as post-translational modifications, subcellular localization and differences in subcellular distributions, interactions with additional proteins) may modify dramatically the interactions.

### 5.1. Cellular complexity, compartmentalization and crowding effects influence protein–protein interactions

As discussed previously, functional, structural and dynamic properties of the individual proteins influence binary interactions. In addition to protein variants coming from post-translational processes [284], alternative splicing [285] and other (e.g. genetic) factors that may influence gene expression [286], the structure of the interactome is one of the crucial factors underlying the complexity of life, from cells to complete organisms. It is highly dynamic with changes as a function of time, localization in the cell, as well as in response to environmental stimuli [287]. Even interactomes from cells derived from the same tissue, or synchronized cells, may substantially differ. As a consequence, a protein that can be found in different cellular compartments may exhibit different functions, different interactions, or discrete post-translational modifications. Therefore, not only the combination of promiscuity, specificity, selectivity and binding affinity for a specific protein–protein interaction defines the recognition but also all the endogenous and exogenous factors that influence the cell.

Protein interactions are governed by several forces, such as compartmentalization and electrostatic and hydrophobic effects. Co-localization, an endogenous property of the living cell, already increases the effective concentration of biomolecules, leading to non-specific protein interactions in their microcompartment. Co-localization, together with normal mechanisms of natural selection, can lead to the formation of

interacting domains, hetero- or homodimers [198]. For example, when proteins are not co-localized, a single mutation can lead to a change of a couple of kcal mol<sup>-1</sup>, but, since the concentration of proteins in the cells is usually in the micromolar to nanomolar range, binding is negligible. However, when proteins are co-localized (boosting greatly the effective concentration of proteins), a small effect on the dissociation constant can translate into substantial binding, since its value can be brought below the effective concentration of the co-localized partner (being approx. micromolar).

One of the most prominent endogenous properties of the cell is macromolecular crowding, a phenomenon that alters the properties of molecules in a solution when high concentrations of macromolecules such as proteins are present [274]. Macromolecular crowding enhances significantly interactions in a non-specific manner and is expected to affect both diffusion-limited and transition-state-limited association reactions, by decreasing or increasing their rates [123,272]. However, it is still unknown to what extent cellular heterogeneity and physiological properties of biological structures are affected, since no single experimental study has yet reported conclusive evidence on the role of macromolecular crowding. Models for macromolecular crowding should be developed in order to have a more realistic view of in-cell protein–protein interactions [271], given the available experimental data [272]. Towards this goal, novel in-cell NMR methodology [288,289] should contribute to our understanding of protein–protein interactions in different cellular environments and under different cellular conditions.

## 6. Design of protein–protein interfaces, modulators and inhibitors

Two general approaches deal with the modulation of protein–protein interactions, namely (i) redesign of the interface by genetic/protein engineering, aiming to alter properties of the protein–protein complex or even the specificity of an interaction, and (ii) inhibitor design, aiming to disrupt protein–protein interactions. A book on this topic has been published by Adler *et al.* [290].

### 6.1. Interface design of protein–protein complexes

Experimentally, redesign of natural protein–protein interfaces has been successfully applied in several cases [64,291], including various systems of structural [292] significance for interface design or major biological importance [293,294]. The current main goals of interface design are to increase the affinity and/or alter the specificity of an interaction [64,295]. Such studies include careful combination of experimental approaches (mutagenesis studies coupled with experimental measurements of affinity and determination of the structure of the complex of the derived variants) and theoretical methods (docking, interface/hot-spot prediction, free energy calculation, calculation of interfacial hydrophobicity, etc.) [292–294,296–300]. Most of the successful design methodologies include either promotion of dipole interactions between  $\alpha$ -helices [300–302] or binding of an  $\alpha$ -helix to the binding site of the target [293,296,303]. Several other methods have mapped known side-chain interactions from a crystal structure onto another protein that can then

be used as a scaffold [304,305]. Recently, homodimer designs composed of paired  $\beta$ -strands have been reported [292,297].

Several limitations of the designed binders have been reported, such as significant rotation of the binders in the crystal structure compared with the predicted orientation of the protein–protein complex, even by as much as 180° [294], or the existence of multiple low-energy binding conformations [296]. It has long been known that introducing changes to increase affinity of the partners might hamper the specificity of the interaction [306]. This may lead to the formation of interfaces with different properties from expected, despite successful engineering of the binders. Another issue regarding interface design is that new hydrogen-bond networks are daunting to design [307], whereas hydrophobic matching of the interface [292] can lead to aggregation. Finally, most of the designs reported to date have been aimed towards protein–protein complexes that have a high degree of surface complementarity. Engineering of more transient protein–protein interactions, with fewer concave interfaces, has not yet been reported. In general, by increasing the affinity of a given protein–protein complex, several other properties of the proteins might be influenced, from their individual stability or solubility to the complex's general properties, such as promiscuity and specificity. Therefore, design of protein–protein interfaces is a daunting task, since careful investigation of all altered properties of the reactants and the derived product should be reported in order to assess the modulation of the interfacial properties in detail.

A very notable example for the limitations of present scoring functions in interface design has been reported [168], where none of the current computational methods used to calculate energetic properties of protein–protein interfaces could discriminate designed complexes that were not able to bind from the one that actually binds. Therefore, although several components of scoring functions should be useful in discriminating designed interfaces from naturally occurring ones, such as the backbone conformational rigidity, electrostatic interactions or solvation energy [168], there is still a gap in our understanding of naturally occurring protein–protein interfaces compared with designed ones.

### 6.2. Small-molecule and peptide inhibitors of protein–protein complexes

The design of inhibitors of protein–protein interactions (protein–protein interaction inhibitors) is also being actively pursued [41,60,197, 308–313]. Some designed inhibitors, such as Navitoclax (ABT-263), an inhibitor of the Bcl-2 family of proteins, have even reached pre-clinical or clinical trials [314]. Most of the protein–protein interaction inhibitors target directly the interface of the complexes, the so-called interfacial inhibitors [315]. Note that this is not synonymous with orthosteric inhibitors, which in our understanding bind to the primary active site of an enzyme or ligand binding site of a receptor molecule [316]. Some inhibitors have been developed that bind at remote locations from the interface, preventing conformational changes required for the formation of the complex (allosteric or non-interfacial inhibitors) [315]. Since protein–protein interfaces are larger than classical enzymatic binding sites, inhibitors or modulators that have been designed for these are also larger in size [317]. Therefore, the traditional drug likeness rules set by Lipinski *et al.* [318] are not generally applicable for this class of inhibitors [319]. Properties of inhibitors of protein–protein interactions are still under investigation, although a

consensus seems to emerge [320–327]. Rationalization of the chemical space of protein–protein interaction inhibitors by using machine learning strategies or sets of molecular descriptors indicated that commercially available libraries are not sufficiently adequate for targeting protein–protein interactions [328], and their specificity, for example for p53-MDM2 inhibition [329], has not yet been fully elucidated. Current studies indicate that this class of inhibitors is generally lipophilic with a higher unsaturation index and ring complexity than common inhibitors [315]. However, whether lipophilicity is a consequence of the way these inhibitors are designed or of the nature of the interfaces that they target still remains to be explored.

Next to chemical substances aiming at disrupting protein–protein interactions, peptide inhibitors have also been reported [290,308,311,312,315,330]. It has become clear that protein–peptide interactions are of high abundance in the living cell, constituting 15–40% of all interactions [331]. Accordingly, discovery and development of protein–peptide inhibitors is of great interest. A few examples follow considering highly potent and selective cyclic [332,333] and other modified peptides [313,334,335]. Stapled peptides produced by connecting two structurally optimized amino acids have also been reported [336,337]. All have been reviewed recently [338–340]. Peptides that modulate protein–protein interactions may be rationally developed by mimicking one of the two partners involved in a protein–protein interaction [341], or directly derived from the screening of peptide sequences that do not originate from natural proteins [342]. Peptides may be antagonists of protein–protein interactions or inhibit the specified interaction. Examples are also available where the inhibitors shift the protein equilibrium, affecting oligomerization; such inhibitors are termed shiftides [343]. An example is the inhibition of HIV-1-IN by peptides derived from its cellular binding protein, LEDGF/p75 [343]: the derived peptides inhibit HIV-1-IN activity in a non-competitive manner, preventing DNA binding by shifting the HIV-1-IN oligomerization equilibrium towards its inactive tetrameric form rather than the active dimer. Notably, a lead chemical compound, now in clinical trials, ABT-737 [344], has been designed as a peptidomimetic, meaning that modulation of Bcl-XL by the BH3 peptide occurring in the cell was mimicked to derive an inhibitor with similar binding characteristics. However, peptide inhibitors have the disadvantage of being easily degraded and thus not orally available. Also, because of their nature, peptides can interact non-specifically with various targets when present in the cell [345]. Improvements in chemical peptide synthesis are required to allow easy chemical modifications

and use of non-natural amino acids [328,338,341], in order to possibly improve the stability and specificity of peptide inhibitors. Structurally improving peptides for binding specificity should also improve their advantage over small molecules as, in principle, owing to their nature, they should be easily accommodated in the interfaces of protein–protein interactions.

## 7. Conclusions

Despite past and current efforts in relating protein structure to binding affinity for protein–protein interactions, the underlying dissociation constants, measured *in vitro*, can still not be reproduced computationally within experimental error for a large dataset of protein–protein complexes. It has been evident that the main physico-chemical measure that relates to binding affinity for protein–protein interactions is the interface area. However, for protein–protein complexes that change significantly their conformation upon binding, even the interface area that is buried upon complexation is not related to binding affinity. Consequently, there must be a significant entropic contribution that will have to be approximated in the future by accounting for structural properties that may be connected to the complexation entropy.

Apart from the direct contributions from the interface that have already been modelled in a satisfactory manner (see §3.3 and 4), vibrational entropy, translational entropy, rotational entropy, conformational flexibility and solvent effects will also have to be accounted for—and, of course, also the effect of the crowded cellular environment. Spolar & Record [346] have attributed the large excess in entropy observed in flexible association to the conformational entropy after entropy decomposition into the abovementioned terms. The availability of an ever-increasing amount of structural and thermodynamic data for protein–protein complexes should stir developments in this research area and hopefully lead to a better understanding of the underlying relations.

Finally, most work so far has been concentrated on binary protein–protein interactions. Molecular associations including multi-component systems, allosteric interactions, multi-state kinetics, or even conformational transitions of membrane proteins are far from being sufficiently well understood to allow the derivation in a systematic manner of useful structure–affinity relations. We foresee that any theoretical modelling of these interactions in the future will have to follow an integrated approach considering the biology, chemistry and physics that underlie protein–protein recognition.

This work was supported by the Dutch Foundation for Scientific Research (NWO) through a VICI grant (no. 700.56.442) to A.M.J.J.B.

## References

- Mulder GJ. 1839 Ueber die Zusammensetzung einiger thierischen Substanzen. *J. Prakt. Chem.* **16**, 129–152. (doi:10.1002/prac.18390160137)
- Perrett D. 2007 From 'protein' to the beginnings of clinical proteomics. *Proteomics Clin. Appl.* **1**, 720–738. (doi:10.1002/prca.200700525)
- Tanford C, Reynolds J. 2004 *Nature's robots: a history of proteins*. New York, NY: Oxford University Press.
- Proust LJ. 1819 Recherches sur le principe qui assaisonne les fromages. *Ann. Chim. Phys.* **29**, 29–49.
- Meyer CE, Rose WC. 1936 The spatial configuration of  $\alpha$ -amino- $\beta$ -hydroxy- $n$ -butyric acid. *J. Biol. Chem.* **115**, 721–729.
- Sanger F. 1949 The terminal peptides of insulin. *Biochem. J.* **45**, 563–574.
- Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958 A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666. (doi:10.1038/181662a0)
- Perutz MF, Muirhead H, Cox JM, Goaman LC. 1968 Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: the atomic model. *Nature* **219**, 131–139. (doi:10.1038/219131a0)
- Svedberg T, Fahraeus R. 1926 A new method for the determination of the molecular weight of proteins.

- J. Am. Chem. Soc.* **48**, 430–438. (doi:10.1021/ja01413a019)
10. Pauling L, Corey RB. 1951 Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl Acad. Sci. USA* **37**, 235–240. (doi:10.1073/pnas.37.5.235)
  11. Pauling L, Corey RB. 1951 The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl Acad. Sci. USA* **37**, 251–256. (doi:10.1073/pnas.37.5.251)
  12. Wuthrich K, Wider G, Wagner G, Braun W. 1982 Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.* **155**, 311–319. (doi:10.1016/0022-2836(82)90007-9)
  13. Williamson MP, Havel TF, Wuthrich K. 1985 Solution conformation of proteinase inhibitor IIA from bull seminal plasma by <sup>1</sup>H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295–315. (doi:10.1016/0022-2836(85)90347-X)
  14. Kaptein R, Zuiderweg ER, Scheek RM, Boelens R, van Gunsteren WF. 1985 A protein structure from nuclear magnetic resonance data lac repressor headpiece. *J. Mol. Biol.* **182**, 179–182. (doi:10.1016/0022-2836(85)90036-1)
  15. Bernstein FC, Koetzle TF, Williams GJ, Meyer Jr EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977 The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542. (doi:10.1016/S0022-2836(77)80200-3)
  16. Berman HM. 2008 The Protein Data Bank: a historical perspective. *Acta Crystallogr. A* **64**, 88–95. (doi:10.1107/S0108767307035623)
  17. Alberts B, Miake-Lye R. 1992 Unscrambling the puzzle of biological machines: the importance of the details. *Cell* **68**, 415–420. (doi:10.1016/0092-8674(92)90179-G)
  18. Welch GR. 2012 'Fuzziness' in the cellular interactome: a historical perspective. *Adv. Exp. Med. Biol.* **725**, 184–190. (doi:10.1007/978-1-4614-0659-4\_11)
  19. Kresge N, Simoni RD, Hill RL. 2005 Otto Fritz Meyerhof and the elucidation of the glycolytic pathway. *J. Biol. Chem.* **280**, e3.
  20. Krebs HA. 1948 The tricarboxylic acid cycle. *Harvey Lect. Series* **44**, 165–199.
  21. Bloch K. 1963 The biological synthesis of unsaturated fatty acids. *Biochem. Soc. Symp.* **24**, 1–16.
  22. Beadle GW, Tatum EL. 1941 Genetic control of biochemical reactions in neurospora. *Proc. Natl Acad. Sci. USA* **27**, 499–506. (doi:10.1073/pnas.27.11.499)
  23. Srere PA. 2000 Macromolecular interactions: tracing the roots. *Trends Biochem. Sci.* **25**, 150–153. (doi:10.1016/S0968-0004(00)01550-4)
  24. Richards FM. 1958 On the enzymic activity of subtilisin-modified ribonuclease. *Proc. Natl Acad. Sci. USA* **44**, 162–166. (doi:10.1073/pnas.44.2.162)
  25. Konisky J. 1972 Characterization of colicin Ia and colicin Ib. Chemical studies of protein structure. *J. Biol. Chem.* **247**, 3750–3755.
  26. Lee B, Richards FM. 1971 The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400. (doi:10.1016/0022-2836(71)90324-X)
  27. Ruhlmann A, Kukla D, Schwager P, Bartels K, Huber R. 1973 Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. Crystal structure determination and stereochemistry of the contact region. *J. Mol. Biol.* **77**, 417–436. (doi:10.1016/0022-2836(73)90448-8)
  28. Chothia C, Janin J. 1975 Principles of protein–protein recognition. *Nature* **256**, 705–708. (doi:10.1038/256705a0)
  29. Wodak SJ, Janin J. 1978 Computer analysis of protein–protein interaction. *J. Mol. Biol.* **124**, 323–342. (doi:10.1016/0022-2836(78)90302-9)
  30. Bessman MJ, Lehman IR, Simms ES, Kornberg A. 1958 Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction. *J. Biol. Chem.* **233**, 171–177.
  31. Alberts BM. 1984 The DNA enzymology of protein machines. *Cold Spring Harb. Symp. Quant. Biol.* **49**, 1–12. (doi:10.1101/SQB.1984.049.01.003)
  32. Sollner T, Rassow J, Wiedmann M, Schlossmann J, Keil P, Neupert W, Pfanner N. 1992 Mapping of the protein import machinery in the mitochondrial outer membrane by crosslinking of translocation intermediates. *Nature* **355**, 84–87. (doi:10.1038/355084a0)
  33. Uetz P *et al.* 2000 A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627. (doi:10.1038/35001009)
  34. Barabasi AL, Oltvai ZN. 2004 Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113. (doi:10.1038/nrg1272)
  35. Sudhof TC. 1995 The synaptic vesicle cycle: a cascade of protein–protein interactions. *Nature* **375**, 645–653. (doi:10.1038/375645a0)
  36. Pawson T, Nash P. 2000 Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **14**, 1027–1047. (doi:10.1101/gad.14.9.1027)
  37. Meszaros B, Simon I, Dosztanyi Z. 2011 The expanding view of protein–protein interactions: complexes involving intrinsically disordered proteins. *Phys. Biol.* **8**, 035003. (doi:10.1088/1478-3975/8/3/035003)
  38. Sowa ME, Bennett EJ, Gygi SP, Harper JW. 2009 Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389–403. (doi:10.1016/j.cell.2009.04.042)
  39. Mizushima N, Komatsu M. 2011 Autophagy: renovation of cells and tissues. *Cell* **147**, 728–741. (doi:10.1016/j.cell.2011.10.026)
  40. Huang J, Schreiber SL. 1997 A yeast genetic system for selecting small molecule inhibitors of protein–protein interactions in nanodroplets. *Proc. Natl Acad. Sci. USA* **94**, 13 396–13 401. (doi:10.1073/pnas.94.25.13396)
  41. Arkin MR, Wells JA. 2004 Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug Discov.* **3**, 301–317. (doi:10.1038/nrd1343)
  42. Wells JA, McClendon CL. 2007 Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. *Nature* **450**, 1001–1009. (doi:10.1038/nature06526)
  43. Hanahan D, Weinberg RA. 2011 Hallmarks of cancer: the next generation. *Cell* **144**, 646–674. (doi:10.1016/j.cell.2011.02.013)
  44. Hard T, Lendel C. 2012 Inhibition of amyloid formation. *J. Mol. Biol.* **421**, 441. (doi:10.1016/j.jmb.2011.12.062)
  45. Eisenberg D, Jucker M. 2012 The amyloid state of proteins in human diseases. *Cell* **148**, 1188–1203. (doi:10.1016/j.cell.2012.02.022)
  46. Schreiber RD, Old LJ, Smyth MJ. 2011 Cancer immunoeediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570. (doi:10.1126/science.1203486)
  47. Lapenna S, Giordano A. 2009 Cell cycle kinases as therapeutic targets for cancer. *Nat. Rev. Drug Discov.* **8**, 547–566. (doi:10.1038/nrd2907)
  48. Linderstrom-Lang KU, Schnellman JA. 1959 Protein structure and enzyme activity. In *The enzymes* (ed. PD Boyer), pp. 443–510. New York, NY: Academic Press.
  49. Bahadur RP. 2010 A structural perspective on protein–protein interactions in macromolecular assemblies. In *Protein–protein complexes: analysis, modeling and drug design* (ed. M Zacharias), pp. 25–45. London, UK: Imperial College Press.
  50. Poupon A, Janin J. 2010 Analysis and prediction of protein quaternary structure. *Methods Mol. Biol.* **609**, 349–364. (doi:10.1007/978-1-60327-241-4\_20)
  51. Byrum S, Smart SK, Larson S, Tackett AJ. 2012 Analysis of stable and transient protein–protein interactions. *Methods Mol. Biol.* **833**, 143–152. (doi:10.1007/978-1-61779-477-3\_10)
  52. Mintseris J, Weng Z. 2005 Structure, function, and evolution of transient and obligate protein–protein interactions. *Proc. Natl Acad. Sci. USA* **102**, 10 930–10 935. (doi:10.1073/pnas.0502667102)
  53. Janin J, Bahadur RP, Chakrabarti P. 2008 Protein–protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180. (doi:10.1017/S0033583508004708)
  54. Janin J, Rodier F, Chakrabarti P, Bahadur RP. 2007 Macromolecular recognition in the Protein Data Bank. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 1–8. (doi:10.1107/S090744490603575X)
  55. Jones S, Thornton JM. 1996 Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20.
  56. Keskin O, Gursoy A, Ma B, Nussinov R. 2008 Principles of protein–protein interactions: what are the preferred ways for proteins to interact? *Chem. Rev.* **108**, 1225–1244. (doi:10.1021/cr040409x)
  57. Pereira-Leal JB, Levy ED, Teichmann SA. 2006 The origins and evolution of functional modules: lessons from protein complexes. *Phil. Trans. R. Soc. B* **361**, 507–517. (doi:10.1098/rstb.2005.1807)
  58. Teichmann SA. 2002 Principles of protein–protein interactions. *Bioinformatics* **18**(Suppl. 2), S249. (doi:10.1093/bioinformatics/18.suppl\_2.S249)

59. Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O. 2008 Architectures and functional coverage of protein–protein interfaces. *J. Mol. Biol.* **381**, 785–802. (doi:10.1016/j.jmb.2008.04.071)
60. Wilson AJ. 2009 Inhibition of protein–protein interactions using designed molecules. *Chem. Soc. Rev.* **38**, 3289–3300. (doi:10.1039/b807197g)
61. Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. 2010 Transient protein–protein interactions: structural, functional, and network properties. *Structure* **18**, 1233–1243. (doi:10.1016/j.str.2010.08.007)
62. Fleishman SJ, Baker D. 2012 Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149**, 262–273. (doi:10.1016/j.cell.2012.03.016)
63. Robinson CV, Sali A, Baumeister W. 2007 The molecular sociology of the cell. *Nature* **450**, 973–982. (doi:10.1038/nature06523)
64. Mandell DJ, Kortemme T. 2009 Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* **5**, 797–807. (doi:10.1038/nchembio.251)
65. Dey S, Pal A, Chakrabarti P, Janin J. 2010 The subunit interfaces of weakly associated homodimeric proteins. *J. Mol. Biol.* **398**, 146–160. (doi:10.1016/j.jmb.2010.02.020)
66. Rodier F, Bahadur RP, Chakrabarti P, Janin J. 2005 Hydration of protein–protein interfaces. *Proteins* **60**, 36–45. (doi:10.1002/prot.20478)
67. Bahadur RP, Chakrabarti P, Rodier F, Janin J. 2004 A dissection of specific and non-specific protein–protein interfaces. *J. Mol. Biol.* **336**, 943–955. (doi:10.1016/j.jmb.2003.12.073)
68. Bahadur RP, Chakrabarti P, Rodier F, Janin J. 2003 Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53**, 708–719. (doi:10.1002/prot.10461)
69. Lo Conte L, Chothia C, Janin J. 1999 The atomic structure of protein–protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198. (doi:10.1006/jmbi.1998.2439)
70. Ofran Y, Rost B. 2003 Analysing six types of protein–protein interfaces. *J. Mol. Biol.* **325**, 377–387. (doi:10.1016/S0022-2836(02)01223-8)
71. Ozbabacan SE, Engin HB, Gursoy A, Keskin O. 2011 Transient protein–protein interactions. *Protein Eng. Des. Sel.* **24**, 635–648. (doi:10.1093/protein/gzr025)
72. McDonald IK, Thornton JM. 1994 Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793. (doi:10.1006/jmbi.1994.1334)
73. Guharoy M, Pal A, Dasgupta M, Chakrabarti P. 2011 PRICE (PProtein Interface Conservation and Energetics): a server for the analysis of protein–protein interfaces. *J. Struct. Funct. Genomics* **12**, 33–41. (doi:10.1007/s10969-011-9108-0)
74. Krissinel E, Henrick K. 2007 Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797. (doi:10.1016/j.jmb.2007.05.022)
75. Reynolds C, Damerell D, Jones S. 2009 ProtorP: a protein–protein interaction analysis server. *Bioinformatics* **25**, 413–414. (doi:10.1093/bioinformatics/btn584)
76. Tina KG, Bhadra R, Srinivasan N. 2007 PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–476. (doi:10.1093/nar/gkm423)
77. Bickerton GR, Higuera AP, Blundell TL. 2011 Comprehensive, atomic-level characterization of structurally characterized protein–protein interactions: the PICCOLO database. *BMC Bioinf.* **12**, 313. (doi:10.1186/1471-2105-12-313)
78. Nooren IM, Thornton JM. 2003 Diversity of protein–protein interactions. *EMBO J* **22**, 3486–3492. (doi:10.1093/emboj/cdg359)
79. Shrake A, Rupley JA. 1973 Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–371. (doi:10.1016/0022-2836(73)90011-9)
80. Sanner MF, Olson AJ, Spehner JC. 1996 Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320.
81. Agishtein ME. 1992 Fuzzy molecular surfaces. *J. Biomol. Struct. Dyn.* **9**, 759–768. (doi:10.1080/07391102.1992.10507954)
82. Connolly ML. 1983 Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709–713. (doi:10.1126/science.6879170)
83. Yin S, Proctor EA, Lugovskoy AA, Dokholyan NV. 2009 Fast screening of protein surfaces using geometric invariant fingerprints. *Proc. Natl Acad. Sci. USA* **106**, 16 622–16 626. (doi:10.1073/pnas.0906146106)
84. Tsai J, Taylor R, Chothia C, Gerstein M. 1999 The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* **290**, 253–266. (doi:10.1006/jmbi.1999.2829)
85. Henzler-Wildman K, Kern D. 2007 Dynamic personalities of proteins. *Nature* **450**, 964–972. (doi:10.1038/nature06522)
86. Levy Y, Onuchic JN. 2006 Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 389–415. (doi:10.1146/annurev.biophys.35.040405.102134)
87. Zhang L, Yang Y, Kao YT, Wang L, Zhong D. 2009 Protein hydration dynamics and molecular mechanism of coupled water–protein fluctuations. *J. Am. Chem. Soc.* **131**, 10 677–10 691. (doi:10.1021/ja902918p)
88. Chakrabarti P, Bhattacharyya R. 2007 Geometry of nonbonded interactions involving planar groups in proteins. *Prog. Biophys. Mol. Biol.* **95**, 83–137. (doi:10.1016/j.pbiomolbio.2007.03.016)
89. Narayana SV, Argos P. 1984 Residue contacts in protein structures and implications for protein folding. *Int. J. Pept. Protein Res.* **24**, 25–39. (doi:10.1111/j.1399-3011.1984.tb00924.x)
90. Nishikawa K, Ooi T. 1986 Radial locations of amino acid residues in a globular protein: correlation with the sequence. *J. Biochem.* **100**, 1043–1047.
91. Bahar I, Jernigan RL. 1997 Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *J. Mol. Biol.* **266**, 195–214. (doi:10.1006/jmbi.1996.0758)
92. Zhang C, Kim SH. 2000 Environment-dependent residue contact energies for proteins. *Proc. Natl Acad. Sci. USA* **97**, 2550–2555. (doi:10.1073/pnas.040573597)
93. Barlow DJ, Thornton JM. 1983 Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885. (doi:10.1016/S0022-2836(83)80079-5)
94. Gallivan JP, Dougherty DA. 1999 Cation- $\pi$  interactions in structural biology. *Proc. Natl Acad. Sci. USA* **96**, 9459–9464. (doi:10.1073/pnas.96.17.9459)
95. Marcou G, Rognan D. 2007 Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **47**, 195–207. (doi:10.1021/ci600342e)
96. Kyte J, Doolittle RF. 1982 A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132. (doi:10.1016/0022-2836(82)90515-0)
97. Reid KSC, Lindley PF, Thornton JM. 1985 Sulphur-aromatic interactions in proteins. *FEBS Lett.* **190**, 209–213. (doi:10.1016/0014-5793(85)81285-0)
98. Bhattacharyya R, Pal D, Chakrabarti P. 2004 Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng. Des. Sel.* **17**, 795–808. (doi:10.1093/protein/gzh093)
99. McGaughey GB, Gagne M, Rappe AK. 1998  $\pi$ -Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **273**, 15 458–15 463. (doi:10.1074/jbc.273.25.15458)
100. Brandl M, Weiss MS, Jabs A, Suhnel J, Hilgenfeld R. 2001 C-H... $\pi$ -interactions in proteins. *J. Mol. Biol.* **307**, 357–377. (doi:10.1006/jmbi.2000.4473)
101. Zauhar RJ, Colbert CL, Morgan RS, Welsh WJ. 2000 Evidence for a strong sulfur-aromatic interaction derived from crystallographic data. *Biopolymers* **53**, 233–248.
102. Kastiris PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J. 2011 A structure-based benchmark for protein–protein binding affinity. *Protein Sci.* **20**, 482–491. (doi:10.1002/pro.580)
103. Hwang H, Vreven T, Janin J, Weng Z. 2010 Protein–protein docking benchmark, version 4.0. *Proteins* **78**, 3111–3114. (doi:10.1002/prot.22830)
104. Raschke TM. 2006 Water structure and interactions with protein surfaces. *Curr. Opin. Struct. Biol.* **16**, 152–159. (doi:10.1016/j.sbi.2006.03.002)
105. Billeter M. 1995 Hydration water molecules seen by NMR and by X-ray crystallography. *Progr. Mag. Res. Spectr.* **27**, 635–645. (doi:10.1016/0079-6565(95)01015-7)
106. Nucci NV, Pometun MS, Wand AJ. 2011 Site-resolved measurement of water–protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **18**, 245–249. (doi:10.1038/nsmb.1955)
107. Sterpone F, Stirmemann G, Laage D. 2012 Magnitude and molecular origin of water slowdown next to a protein. *J. Am. Chem. Soc.* **134**, 4116–4119. (doi:10.1021/ja3007897)
108. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T. 2011 Accessing protein conformational ensembles using room-temperature

- X-ray crystallography. *Proc. Natl Acad. Sci. USA* **108**, 16 247–16 252. (doi:10.1073/pnas.1111325108)
109. Ball P. 2008 Water as an active constituent in cell biology. *Chem. Rev.* **108**, 74–108. (doi:10.1021/cr068037a)
110. Janin J. 1999 Wet and dry interfaces: the role of solvent in protein–protein and protein–DNA recognition. *Structure* **7**, R277–279. (doi:10.1016/S0969-2126(00)88333-1)
111. Ahmed MH, Spyralis F, Cozzini P, Tripathi PK, Mozzarelli A, Scarsdale JN, Safo MA, Kellogg GE. 2011 Bound water at protein–protein interfaces: partners, roles and hydrophobic bubbles as a conserved motif. *PLoS ONE* **6**, e24712. (doi:10.1371/journal.pone.0024712)
112. van Dijk AD, Bonvin AM. 2006 Solvated docking: introducing water into the modelling of biomolecular complexes. *Bioinformatics* **22**, 2340–2347. (doi:10.1093/bioinformatics/btl395)
113. Ahmad M, Gu W, Geyer T, Helms V. 2011 Adhesive water networks facilitate binding of protein interfaces. *Nat. Commun.* **2**, 261. (doi:10.1038/ncomms1258)
114. Papadakos G, Wojdyła JA, Kleanthous C. 2012 Nuclease colicins and their immunity proteins. *Q. Rev. Biophys.* **45**, 57–103. (doi:10.1017/S0033583511000114)
115. Buckle AM, Schreiber G, Fersht AR. 1994 Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-Å resolution. *Biochemistry* **33**, 8878–8889. (doi:10.1021/bi00196a004)
116. Sevcik J, Urbanikova L, Dauter Z, Wilson KS. 1998 Recognition of RNase Sa by the inhibitor barstar: structure of the complex at 1.7 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 954–963. (doi:10.1107/S0907444998004429)
117. Royer Jr WE, Pardanani A, Gibson QH, Peterson ES, Friedman JM. 1996 Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin. *Proc. Natl Acad. Sci. USA* **93**, 14 526–14 531. (doi:10.1073/pnas.93.25.14526)
118. Gnanasekaran R, Xu Y, Leitner DM. 2010 Dynamics of water clusters confined in proteins: a molecular dynamics simulation study of interfacial waters in a dimeric hemoglobin. *J. Phys. Chem. B* **114**, 16 989–16 996. (doi:10.1021/jp109173t)
119. Reddy G, Straub JE, Thirumalai D. 2010 Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proc. Natl Acad. Sci. USA* **107**, 21 459–21 464. (doi:10.1073/pnas.1008616107)
120. Liu P, Huang X, Zhou R, Berne BJ. 2005 Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature* **437**, 159–162. (doi:10.1038/nature03926)
121. Chandler D. 2005 Interfaces and the driving force of hydrophobic assembly. *Nature* **437**, 640–647. (doi:10.1038/nature04162)
122. Ball P. 2008 Water as a biomolecule. *Chemphyschem* **9**, 2677–2685. (doi:10.1002/cphc.200800515)
123. Rosgen J. 2009 Molecular crowding and solvation: direct and indirect impact on protein reactions. *Methods Mol. Biol.* **490**, 195–225. (doi:10.1007/978-1-59745-367-7\_9)
124. Mendel CM, Mendel DB. 1985 'Non-specific' binding. The problem, and a solution. *Biochem. J.* **228**, 269–272.
125. Doty P, Myers GE. 1953 Low molecular weight proteins. Thermodynamics of the association of insulin molecules. *Discuss. Faraday Soc.* **13**, 51–58. (doi:10.1039/DF9531300051)
126. Horton N, Lewis M. 1992 Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**, 169–181. (doi:10.1002/pro.5560010117)
127. Erickson HP. 1989 Co-operativity in protein–protein association. The structure and stability of the actin filament. *J. Mol. Biol.* **206**, 465–474. (doi:10.1016/0022-2836(89)90494-4)
128. Phizicky EM, Fields S. 1995 Protein–protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**, 94–123.
129. Berggard T, Linse S, James P. 2007 Methods for the detection and analysis of protein–protein interactions. *Proteomics* **7**, 2833–2842. (doi:10.1002/pmic.200700131)
130. Dwane S, Kiely PA. 2011 Tools used to study how protein complexes are assembled in signaling cascades. *Bioeng. Bugs* **2**, 247–259. (doi:10.4161/bbug.2.5.17844)
131. Piehler J. 2005 New methodologies for measuring protein interactions *in vivo* and *in vitro*. *Curr. Opin. Struct. Biol.* **15**, 4–14. (doi:10.1016/j.sbi.2005.01.008)
132. Shoemaker BA, Panchenko AR. 2007 Deciphering protein–protein interactions. I. Experimental techniques and databases. *PLoS Comput. Biol.* **3**, e42. (doi:10.1371/journal.pcbi.0030042)
133. Wilkinson KD. 2004 Quantitative analysis of protein–protein interactions. *Methods Mol. Biol.* **261**, 15–32. (doi:10.1385/1-59259-762-9:015)
134. Ladbury JE, Chowdhry BZ. 1996 Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **3**, 791–801. (doi:10.1016/S1074-5521(96)90063-0)
135. Willander M, Al-Hilli S. 2009 Analysis of biomolecules using surface plasmons. *Methods Mol. Biol.* **544**, 201–229. (doi:10.1007/978-1-59745-483-4\_14)
136. Masi A, Cicchi R, Carloni A, Pavone FS, Arcangeli A. 2010 Optical methods in the study of protein–protein interactions. *Adv. Exp. Med. Biol.* **674**, 33–42. (doi:10.1007/978-1-4419-6066-5\_4)
137. Vuignier K, Schappler J, Veuthey JL, Carrupt PA, Martel S. 2010 Drug-protein binding: a critical review of analytical tools. *Anal. Bioanal. Chem.* **398**, 53–66. (doi:10.1007/s00216-010-3737-1)
138. Qin S, Pang X, Zhou HX. 2011 Automated prediction of protein association rate constants. *Structure* **19**, 1744–1751. (doi:10.1016/j.str.2011.10.015)
139. Berg T. 2003 Modulation of protein–protein interactions with small organic molecules. *Angew. Chem. Int. Ed. Engl.* **42**, 2462–2481. (doi:10.1002/anie.200200558)
140. Klotz IM. 1985 Ligand–receptor interactions: facts and fantasies. *Q. Rev. Biophys.* **18**, 227–259. (doi:10.1017/S0033583500000354)
141. Fischer E. 1894 Einfluss der configuration auf die Wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* **27**, 2985–2993. (doi:10.1002/cber.18940270364)
142. Koshland DE. 1958 Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl Acad. Sci. USA* **44**, 98–104. (doi:10.1073/pnas.44.2.98)
143. Straub FB, Szabolcsi G. 1964 [On the dynamic aspects of protein structure.] In *Molecular biology, problems and perspectives* (ed. AE Braunstein), pp 182–187. Moscow, Soviet Union: Izdat. Nauka. [In Russian.]
144. Monod J, Changeux JP, Jacob F. 1963 Allosteric proteins and cellular control systems. *J. Mol. Biol.* **6**, 306–329. (doi:10.1016/S0022-2836(63)80091-1)
145. Changeux JP, Thiery J, Tung Y, Kittel C. 1967 On the cooperativity of biological membranes. *Proc. Natl Acad. Sci. USA* **57**, 335–341. (doi:10.1073/pnas.57.2.335)
146. Burgen AS. 1981 Conformational changes and drug action. *Fed. Proc.* **40**, 2723–2728.
147. Koshland Jr DE. 1959 Enzyme flexibility and enzyme action. *J. Cell Comp. Physiol.* **54**, 245–258. (doi:10.1002/jcp.1030540420)
148. Ma B, Kumar S, Tsai CJ, Nussinov R. 1999 Folding funnels and binding mechanisms. *Protein Eng.* **12**, 713–720. (doi:10.1093/protein/12.9.713)
149. Brocklehurst K, Willenbrock SJ, Salih E. 1983 Effects of conformational selectivity and of overlapping kinetically influential ionizations on the characteristics of pH-dependent enzyme kinetics. Implications of free-enzyme pKa variability in reactions of papain for its catalytic mechanism. *Biochem. J.* **211**, 701–708.
150. Okazaki K, Takada S. 2008 Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proc. Natl Acad. Sci. USA* **105**, 11 182–11 187. (doi:10.1073/pnas.0802524105)
151. Weikl TR, von Deuster C. 2009 Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* **75**, 104–110. (doi:10.1002/prot.22223)
152. Bahar I, Lezon TR, Yang LW, Eyal E. 2010 Global dynamics of proteins: bridging between structure and function. *Annu. Rev. Biophys.* **39**, 23–42. (doi:10.1146/annurev.biophys.093008.131258)
153. Koshland Jr DE, Neet KE. 1968 The catalytic and regulatory properties of enzymes. *Annu. Rev. Biochem.* **37**, 359–410. (doi:10.1146/annurev.bi.37.070168.002043)
154. Citri N. 1973 Conformational adaptability in enzymes. *Adv. Enzymol. Relat. Areas Mol. Biol.* **37**, 397–648. (doi:10.1002/9780470122822.ch7)
155. Changeux JP, Edelstein S. 2011 Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* **3**, 19. (doi:10.3410/B3-19)
156. Csermely P, Palotai R, Nussinov R. 2010 Induced fit, conformational selection and independent dynamic

- segments: an extended view of binding events. *Trends Biochem. Sci.* **35**, 539–546. (doi:10.1016/j.tibs.2010.04.009)
157. Hammes GG, Chang YC, Oas TG. 2009 Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl Acad. Sci. USA* **106**, 13 737–13 741. (doi:10.1073/pnas.0907195106)
158. Wlodarski T, Zagrovic B. 2009 Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin. *Proc. Natl Acad. Sci. USA* **106**, 19 346–19 351. (doi:10.1073/pnas.0906966106)
159. Grunberg R, Leckner J, Nilges M. 2004 Complementarity of structure ensembles in protein–protein binding. *Structure* **12**, 2125–2136. (doi:10.1016/j.str.2004.09.014)
160. Zhou HX. 2010 From induced fit to conformational selection: a continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophys. J.* **98**, L15–L17. (doi:10.1016/j.bpj.2009.11.029)
161. Bogan AA, Thorn KS. 1998 Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9. (doi:10.1006/jmbi.1998.1843)
162. Rajamani D, Thiel S, Vajda S, Camacho CJ. 2004 Anchor residues in protein–protein interactions. *Proc. Natl Acad. Sci. USA* **101**, 11 287–11 292. (doi:10.1073/pnas.0401942101)
163. Audie J, Scarlata S. 2007 A novel empirical free energy function that explains and predicts protein–protein binding affinities. *Biophys. Chem.* **129**, 198–211. (doi:10.1016/j.bpc.2007.05.021)
164. Jiang L, Gao Y, Mao F, Liu Z, Lai L. 2002 Potential of mean force for protein–protein interaction studies. *Proteins* **46**, 190–196. (doi:10.1002/prot.10031)
165. Ma XH, Wang CX, Li CH, Chen WZ. 2002 A fast empirical approach to binding free energy calculations based on protein interface information. *Protein Eng.* **15**, 677–681. (doi:10.1093/protein/15.8.677)
166. Zhang C, Liu S, Zhu Q, Zhou Y. 2005 A knowledge-based energy function for protein–ligand, protein–protein, and protein–DNA complexes. *J. Med. Chem.* **48**, 2325–2335. (doi:10.1021/jm049314d)
167. Su Y, Zhou A, Xia X, Li W, Sun Z. 2009 Quantitative prediction of protein–protein binding affinity with a potential of mean force considering volume correction. *Protein Sci.* **18**, 2550–2558. (doi:10.1002/pro.257)
168. Fleishman SJ *et al.* 2011 Community-wide assessment of protein–interface modeling suggests improvements to design methodology. *J. Mol. Biol.* **414**, 289–302. (doi:10.1016/j.jmb.2011.09.031)
169. Moal IH, Agius R, Bates PA. 2011 Protein–protein binding affinity prediction on a diverse set of structures. *Bioinformatics* **27**, 3002–3009. (doi:10.1093/bioinformatics/btr513)
170. Vreven T, Hwang H, Pierce BG, Weng Z. 2012 Prediction of protein–protein binding free energies. *Protein Sci.* **21**, 396–404. (doi:10.1002/pro.2027)
171. Moal IH, Bates PA. 2012 Kinetic rate constant prediction supports the conformational selection mechanism of protein binding. *PLoS Comput. Biol.* **8**, e1002351. (doi:10.1371/journal.pcbi.1002351)
172. Tian F, Lv Y, Yang L. 2011 Structure-based prediction of protein–protein binding affinity with consideration of allosteric effect. *Amino Acids* **43**, 531–543. (doi:10.1007/s00726-011-1101-1)
173. Bai H, Yang K, Yu D, Zhang C, Chen F, Lai L. 2011 Predicting kinetic constants of protein–protein interactions based on structural properties. *Proteins* **79**, 720–734. (doi:10.1002/prot.22904)
174. Noskov SY, Lim C. 2001 Free energy decomposition of protein–protein interactions. *Biophys. J.* **81**, 737–750. (doi:10.1016/S0006-3495(01)75738-4)
175. Miller S, Lesk AM, Janin J, Chothia C. 1987 The accessible surface area and stability of oligomeric proteins. *Nature* **328**, 834–836. (doi:10.1038/328834a0)
176. Fersht AR *et al.* 1985 Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **314**, 235–238.
177. Kastrius PL, Bonvin AM. 2010 Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark. *J. Proteome Res.* **9**, 2216–2225. (doi:10.1021/pr9009854).
178. Cunningham BC, Wells JA. 1993 Comparison of a structural and a functional epitope. *J. Mol. Biol.* **234**, 554–563. (doi:10.1006/jmbi.1993.1611)
179. Clackson T, Wells JA. 1995 A hot spot of binding energy in a hormone–receptor interface. *Science* **267**, 383–386. (doi:10.1126/science.7529940)
180. Kortemme T, Baker D. 2002 A simple physical model for binding energy hot spots in protein–protein complexes. *Proc. Natl Acad. Sci. USA* **99**, 14 116–14 121. (doi:10.1073/pnas.202485799)
181. Guerois R, Nielsen JE, Serrano L. 2002 Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387. (doi:10.1016/S0022-2836(02)00442-4)
182. Ofra Y, Rost B. 2007 Protein–protein interaction hotspots carved into sequences. *PLoS Comput. Biol.* **3**, e119. (doi:10.1371/journal.pcbi.0030119)
183. Benedix A, Becker CM, de Groot BL, Caffisch A, Bockmann RA. 2009 Predicting free energy changes using structural ensembles. *Nat. Methods* **6**, 3–4. (doi:10.1038/nmeth0109-3)
184. Meireles LM, Domling AS, Camacho CJ. 2010 ANCHOR: a web server and database for analysis of protein–protein interaction binding pockets for drug discovery. *Nucleic Acids Res.* **38**, W407–W411. (doi:10.1093/nar/gkq502)
185. Grosdidier S, Fernandez-Recio J. 2008 Identification of hot-spot residues in protein–protein interactions by computational docking. *BMC Bioinf.* **9**, 447. (doi:10.1186/1471-2105-9-447)
186. Moreira IS, Fernandes PA, Ramos MJ. 2007 Hot spots: a review of the protein–protein interface determinant amino-acid residues. *Proteins* **68**, 803–812. (doi:10.1002/prot.21396)
187. Morrow JK, Zhang S. 2012 Computational prediction of protein hot spot residues. *Curr. Pharm. Des.* **18**, 1255–1265. (doi:10.2174/138161212799436412)
188. Fernandez-Recio J. 2011 Prediction of protein binding sites and hot spots. *WIREs: Comp. Mol. Sci.* **1**, 680–698. (doi:10.1002/wcms.45)
189. Hu Z, Ma B, Wolfson H, Nussinov R. 2000 Conservation of polar residues as hot spots at protein interfaces. *Proteins* **39**, 331–342.
190. DeLano WL. 2002 Unraveling hot spots in binding interfaces: progress and challenges. *Curr. Opin. Struct. Biol.* **12**, 14–20. (doi:10.1016/S0959-440X(02)00283-X)
191. Halperin I, Wolfson H, Nussinov R. 2004 Protein–protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure* **12**, 1027–1038. (doi:10.1016/j.str.2004.04.009)
192. Li J, Liu Q. 2009 ‘Double water exclusion’: a hypothesis refining the O-ring theory for the hot spots at protein interfaces. *Bioinformatics* **25**, 743–750. (doi:10.1093/bioinformatics/btp058)
193. Guharoy M, Chakrabarti P. 2009 Empirical estimation of the energetic contribution of individual interface residues in structures of protein–protein complexes. *J. Comput. Aided Mol. Des.* **23**, 645–654. (doi:10.1007/s10822-009-9282-3)
194. Janin J. 2009 Basic principles of protein–protein interaction. In *Computational protein–protein interactions* (eds R Nussinov, G Schreiber), pp. 1–20. New York, NY: Taylor & Francis Group.
195. Kimura SR, Brower RC, Vajda S, Camacho CJ. 2001 Dynamical view of the positions of key side chains in protein–protein recognition. *Biophys. J.* **80**, 635–642. (doi:10.1016/S0006-3495(01)76044-4)
196. Castro MJ, Anderson S. 1996 Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin. *Biochemistry* **35**, 11 435–11 446. (doi:10.1021/bi960515w)
197. Rudolph J. 2007 Inhibiting transient protein–protein interactions: lessons from the Cdc25 protein tyrosine phosphatases. *Nat. Rev. Cancer* **7**, 202–211. (doi:10.1038/nrc2087)
198. Kuriyan J, Eisenberg D. 2007 The origin of protein interactions and allostery in colocalization. *Nature* **450**, 983–990. (doi:10.1038/nature06524)
199. Sprinzak E, Sattath S, Margalit H. 2003 How reliable are experimental protein–protein interaction data? *J. Mol. Biol.* **327**, 919–923. (doi:10.1016/S0022-2836(03)00239-0)
200. Service R. 2005 Structural biology. Structural genomics, round 2. *Science* **307**, 1554–1558. (doi:10.1126/science.307.5715.1554)
201. Ritchie DW. 2008 Recent progress and future directions in protein–protein docking. *Curr. Protein Pept. Sci.* **9**, 1–15. (doi:10.2174/138920308783565741)
202. Vajda S, Kozakov D. 2009 Convergence and combination of methods in protein–protein docking. *Curr. Opin. Struct. Biol.* **19**, 164–170. (doi:10.1016/j.sbi.2009.02.008)
203. Hinsen K, Thomas A, Field MJ. 1999 Analysis of domain motions in large proteins. *Proteins* **34**, 369–382.

204. Tokuriki N, Tawfik DS. 2009 Protein dynamism and evolvability. *Science* **324**, 203–207. (doi:10.1126/science.1169375)
205. Sundberg EJ, Mariuzza RA. 2000 Luxury accommodations: the expanding role of structural plasticity in protein–protein interactions. *Structure* **8**, R137–R142. (doi:10.1016/S0969-2126(00)00167-2)
206. Sherwood P, Brooks BR, Sansom MS. 2008 Multiscale methods for macromolecular simulations. *Curr. Opin. Struct. Biol.* **18**, 630–640. (doi:10.1016/j.sbi.2008.07.003)
207. Weiss DR, Levitt M. 2009 Can morphing methods predict intermediate structures? *J. Mol. Biol.* **385**, 665–674. (doi:10.1016/j.jmb.2008.10.064)
208. Fenwick RB, Esteban-Martin S, Salvatella X. 2011 Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur. Biophys. J.* **40**, 1339–1355. (doi:10.1007/s00249-011-0754-8)
209. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES. 2004 Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci.* **13**, 190–202. (doi:10.1110/ps.03323604)
210. Zhang QC, Petrey D, Norel R, Honig BH. 2010 Protein interface conservation across structure space. *Proc. Natl Acad. Sci. USA* **107**, 10 896–10 901. (doi:10.1073/pnas.1005894107)
211. Kundrotas PJ, Zhu Z, Janin J, Vakser IA. 2012 Templates are available to model nearly all complexes of structurally characterized proteins. *Proc. Natl Acad. Sci. USA* **109**, 9438–9441. (doi:10.1073/pnas.1200678109)
212. Sacquin-Mora S, Carbone A, Lavery R. 2008 Identification of protein interaction partners and protein–protein interaction sites. *J. Mol. Biol.* **382**, 1276–1289. (doi:10.1016/j.jmb.2008.08.002)
213. Martin J, Lavery R. 2012 Arbitrary protein–protein docking targets biologically relevant interfaces. *BMC Biophys.* **5**, 7. (doi:10.1186/2046-1682-5-7)
214. Wass MN, Fuentes G, Pons C, Pazos F, Valencia A. 2011 Towards the prediction of protein interaction partners using physical docking. *Mol. Syst. Biol.* **7**, 469. (doi:10.1038/msb.2011.3)
215. Wass MN, David A, Sternberg MJ. 2011 Challenges for the prediction of macromolecular interactions. *Curr. Opin. Struct. Biol.* **21**, 382–390. (doi:10.1016/j.sbi.2011.03.013)
216. Lensink MF, Wodak SJ. 2010 Docking and scoring protein interactions: CAPRI 2009. *Proteins* **78**, 3073–3084. (doi:10.1002/prot.22818)
217. Kowalsman N, Eisenstein M. 2007 Inherent limitations in protein–protein docking procedures. *Bioinformatics* **23**, 421–426. (doi:10.1093/bioinformatics/btl524)
218. Felii E, Oliva B. 2010 How different from random are docking predictions when ranked by scoring functions? *Proteins* **78**, 3376–3385. (doi:10.1002/prot.22844)
219. Bonvin AM. 2006 Flexible protein–protein docking. *Curr. Opin. Struct. Biol.* **16**, 194–200. (doi:10.1016/j.sbi.2006.02.002)
220. Zacharias M. 2010 Accounting for conformational changes during protein–protein docking. *Curr. Opin. Struct. Biol.* **20**, 180–186. (doi:10.1016/j.sbi.2010.02.001)
221. Lorenzen S, Zhang Y. 2007 Identification of near-native structures by clustering protein docking conformations. *Proteins* **68**, 187–194. (doi:10.1002/prot.21442)
222. Rodrigues JP, Trellet M, Schmitz C, Kastiris P, Karaca E, Melquiond AS, Bonvin AM. 2012 Clustering biomolecular complexes by residue contacts similarity. *Proteins* **80**, 1810–1817. (doi:10.1002/prot.24078)
223. Kozakov D, Clodfelter KH, Vajda S, Camacho CJ. 2005 Optimal clustering for detecting near-native conformations in protein docking. *Biophys. J.* **89**, 867–875. (doi:10.1529/biophysj.104.058768)
224. Dominguez C, Boelens R, Bonvin AM. 2003 HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737. (doi:10.1021/ja026939x)
225. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernado P, Fernandez-Recio J. 2010 Structural characterization of protein–protein complexes by integrating computational docking with small-angle scattering data. *J. Mol. Biol.* **403**, 217–230. (doi:10.1016/j.jmb.2010.08.029)
226. Schmitz C, Vernon R, Otting G, Baker D, Huber T. 2011 Protein structure determination from pseudocontact shifts using ROSETTA. *J. Mol. Biol.* **416**, 668–677. (doi:10.1016/j.jmb.2011.12.056)
227. Schmitz C, Bonvin AM. 2012 Protein–protein HADDOCKing using exclusively pseudocontact shifts. *J. Biomol. NMR* **50**, 263–266. (doi:10.1007/s10858-011-9514-4)
228. Sgourakis NG, Lange OF, DiMaio F, Andre I, Fitzkee NC, Rossi P, Montelione GT, Bax A, Baker D. 2011 Determination of the structures of symmetric protein oligomers from NMR chemical shifts and residual dipolar couplings. *J. Am. Chem. Soc.* **133**, 6288–6298. (doi:10.1021/ja111318m)
229. Stratmann D, Boelens R, Bonvin AM. 2011 Quantitative use of chemical shifts for the modeling of protein complexes. *Proteins* **79**, 2662–2670. (doi:10.1002/prot.23090)
230. Fiset O, Lague P, Gagne S, Morin S. 2012 Synergistic applications of MD and NMR for the study of biological systems. *J. Biomed. Biotechnol.* **2012**, 254208. (doi:10.1155/2012/254208)
231. Cowieson NP, Kobe B, Martin JL. 2008 United we stand: combining structural methods. *Curr. Opin. Struct. Biol.* **18**, 617–622. (doi:10.1016/j.sbi.2008.07.004)
232. Habeck M, Nilges M. 2011 Combining computational modeling with sparse and low-resolution data. *J. Struct. Biol.* **173**, 419. (doi:10.1016/j.jsb.2011.01.002)
233. Alber F, Forster F, Korkin D, Topf M, Sali A. 2008 Integrating diverse data for structure determination of macromolecular assemblies. *Annu. Rev. Biochem.* **77**, 443–477. (doi:10.1146/annurev.biochem.77.060407.135530)
234. Karaca E, Melquiond AS, de Vries SJ, Kastiris PL, Bonvin AM. 2011 Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multibody docking server. *Mol. Cell Proteomics* **9**, 1784–1794. (doi:10.1074/mcp.M000051-MCP201)
235. Alber F *et al.* 2007 The molecular architecture of the nuclear pore complex. *Nature* **450**, 695–701. (doi:10.1038/nature06405)
236. Lasker K, Velazquez-Muriel JA, Webb BM, Yang Z, Ferrin TE, Sali A. 2011 Macromolecular assembly structures by comparative modeling and electron microscopy. *Methods Mol. Biol.* **857**, 331–350. (doi:10.1007/978-1-61779-588-6\_15)
237. Mosca R, Pons C, Fernandez-Recio J, Aloy P. 2009 Pushing structural information into the yeast interactome by high-throughput protein docking experiments. *PLoS Comput. Biol.* **5**, e1000490. (doi:10.1371/journal.pcbi.1000490)
238. Sinha R, Kundrotas PJ, Vakser IA. 2010 Docking by structural similarity at protein–protein interfaces. *Proteins* **78**, 3235–3241. (doi:10.1002/prot.22812)
239. Kundrotas PJ, Lensink MF, Alexov E. 2008 Homology-based modeling of 3D structures of protein–protein complexes using alignments of modified sequence profiles. *Int. J. Biol. Macromol.* **43**, 198–208. (doi:10.1016/j.ijbiomac.2008.05.004)
240. Elcock AH, McCammon JA. 2001 Identification of protein oligomerization states by analysis of interface conservation. *Proc. Natl Acad. Sci. USA* **98**, 2990–2994. (doi:10.1073/pnas.061411798)
241. Perica T, Chothia C, Teichmann SA. 2012 Evolution of oligomeric state through geometric coupling of protein interfaces. *Proc. Natl Acad. Sci. USA* **109**, 8127–8132. (doi:10.1073/pnas.1120028109)
242. Park SY, Beel BD, Simon MI, Bilwes AM, Crane BR. 2004 In different organisms, the mode of interaction between two signaling proteins is not necessarily conserved. *Proc. Natl Acad. Sci. USA* **101**, 11 646–11 651. (doi:10.1073/pnas.0401038101)
243. Aloy P, Ceulemans H, Stark A, Russell RB. 2003 The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.* **332**, 989–998. (doi:10.1016/j.jmb.2003.07.006)
244. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA. 2008 Assembly reflects evolution of protein complexes. *Nature* **453**, 1262–1265. (doi:10.1038/nature06942)
245. Murphy KP, Freire E. 1992 Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* **43**, 313–361. (doi:10.1016/S0065-3233(08)60556-2)
246. Audie J. 2009 Development and validation of an empirical free energy function for calculating protein–protein binding free energy surfaces. *Biophys. Chem.* **139**, 84–91. (doi:10.1016/j.bpc.2008.10.007)
247. Lavigne P, Bagu JR, Boyko R, Willard L, Holmes CF, Sykes BD. 2000 Structure-based thermodynamic analysis of the dissociation of protein phosphatase-1 catalytic subunit and microcystin-LR docked complexes. *Protein Sci.* **9**, 252–264. (doi:10.1110/ps.9.2.252)

248. Cheng TM, Blundell TL, Fernandez-Recio J. 2007 pyDock: electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins* **68**, 503–515. (doi:10.1002/prot.21419)
249. Camacho CJ, Zhang C. 2005 FastContact: rapid estimate of contact and binding free energies. *Bioinformatics* **21**, 2534–2536. (doi:10.1093/bioinformatics/bti322)
250. Pierce B, Weng Z. 2007 ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* **67**, 1078–1086. (doi:10.1002/prot.21373)
251. Lensink MF, Wodak SJ. 2010 Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins* **78**, 3085–3095. (doi:10.1002/prot.22850)
252. Wereszczynski J, McCammon JA. 2012 Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* **45**, 1–25. (doi:10.1017/S0033583511000096)
253. Li Z, Lazaridis T. 2012 Computing the thermodynamic contributions of interfacial water. *Methods Mol. Biol.* **819**, 393–404. (doi:10.1007/978-1-61779-465-0\_24)
254. Stone M. 1977 An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Stat. Soc. B* **38**, 44–47.
255. Eriksson L, Johansson E, Muller M, Wold S. 2000 On the selection of the training set in environmental QSAR analysis when compounds are clustered. *J. Chemometr.* **14**, 599–616.
256. Martens HA, Dardenne P. 1998 Validation and verification of regression in small data sets. *Chem. Intell. Lab. Syst.* **44**, 99–121. (doi:10.1016/S0169-7439(98)00167-1)
257. Schreiber G, Haran G, Zhou HX. 2009 Fundamental aspects of protein–protein association kinetics. *Chem. Rev.* **109**, 839–860. (doi:10.1021/cr800373w)
258. Tuffery P, Derreumaux P. 2012 Flexibility and binding affinity in protein–ligand, protein–protein and multi-component protein interactions: limitations of current computational approaches. *J. R. Soc. Interface* **9**, 20–33. (doi:10.1098/rsif.2011.0584)
259. Janin J. 1997 The kinetics of protein–protein recognition. *Proteins* **28**, 153–161.
260. Dong F, Zhou HX. 2006 Electrostatic contribution to the binding stability of protein–protein complexes. *Proteins* **65**, 87–102. (doi:10.1002/prot.21070)
261. Harel M, Spaar A, Schreiber G. 2009 Fruitful and futile encounters along the association reaction between proteins. *Biophys. J.* **96**, 4237–4248. (doi:10.1016/j.bpj.2009.02.054)
262. Phillip Y, Sherman E, Haran G, Schreiber G. 2009 Common crowding agents have only a small effect on protein–protein interactions. *Biophys. J.* **97**, 875–885. (doi:10.1016/j.bpj.2009.05.026)
263. Shaul Y, Schreiber G. 2005 Exploring the charge space of protein–protein association: a proteomic study. *Proteins* **60**, 341–352. (doi:10.1002/prot.20489)
264. Selzer T, Schreiber G. 1999 Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. *J. Mol. Biol.* **287**, 409–419. (doi:10.1006/jmbi.1999.2615)
265. Prakash MK. 2011 Insights on the role of (dis)order from protein–protein interaction linear free-energy relationships. *J. Am. Chem. Soc.* **133**, 9976–9979. (doi:10.1021/ja201500z)
266. Hugo N, Lafont V, Beukes M, Altschuh D. 2002 Functional aspects of co-variant surface charges in an antibody fragment. *Protein Sci.* **11**, 2697–2705. (doi:10.1110/ps.0209302)
267. Gibas CJ, Subramaniam S, McCammon JA, Braden BC, Poljak RJ. 1997 pH dependence of antibody/lysozyme complexation. *Biochemistry* **36**, 15 599–15 614. (doi:10.1021/bi9701989)
268. Sheinerman FB, Norel R, Honig B. 2000 Electrostatic aspects of protein–protein interactions. *Curr. Opin. Struct. Biol.* **10**, 153–159. (doi:10.1016/S0959-440X(00)00665-8)
269. Tiemeyer S, Paulus M, Tolan M. 2010 Effect of surface charge distribution on the adsorption orientation of proteins to lipid monolayers. *Langmuir* **26**, 14 064–14 067. (doi:10.1021/la102616h)
270. Ando T, Skolnick J. 2010 Crowding and hydrodynamic interactions likely dominate *in vivo* macromolecular motion. *Proc. Natl Acad. Sci. USA* **107**, 18 457–18 462. (doi:10.1073/pnas.1011354107)
271. Elcock AH. 2010 Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **20**, 196–206. (doi:10.1016/j.sbi.2010.01.008)
272. Zhou HX, Rivas G, Minton AP. 2008 Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37**, 375–397. (doi:10.1146/annurev.biophys.37.032807.125817)
273. Minton AP. 2001 The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.* **276**, 10 577–10 580. (doi:10.1074/jbc.R100005200)
274. Zimmerman SB, Trach SO. 1991 Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **222**, 599–620. (doi:10.1016/0022-2836(91)90499-V)
275. Harel M, Cohen M, Schreiber G. 2007 On the dynamic nature of the transition state for protein–protein association as determined by double-mutant cycle analysis and simulation. *J. Mol. Biol.* **371**, 180–196. (doi:10.1016/j.jmb.2007.05.032)
276. Alsallaq R, Zhou HX. 2008 Electrostatic rate enhancement and transient complex of protein–protein association. *Proteins* **71**, 320–335. (doi:10.1002/prot.21679)
277. Frembgen-Kesner T, Elcock AH. 2010 Absolute protein–protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: the role of intermolecular hydrodynamic interactions in barnase–barstar association. *Biophys. J.* **99**, L75–L77. (doi:10.1016/j.bpj.2010.09.006)
278. Brown CJ, Lain S, Verma CS, Fersht AR, Lane DP. 2009 Awakening guardian angels: drugging the p53 pathway. *Nat. Rev. Cancer* **9**, 862–873. (doi:10.1038/nrc2763)
279. Royds JA, Iacopetta B. 2006 p53 and disease: when the guardian angel fails. *Cell Death Differ.* **13**, 1017–1026. (doi:10.1038/sj.cdd.4401913)
280. Veprintsev DB, Freund SM, Andreeva A, Rutledge SE, Tidow H, Canadillas JM, Blair CM, Fersht AR. 2006 Core domain interactions in full-length p53 in solution. *Proc. Natl Acad. Sci. USA* **103**, 2115–2119. (doi:10.1073/pnas.0511130103)
281. Rajagopalan S, Andreeva A, Teufel DP, Freund SM, Fersht AR. 2009 Interaction between the transactivation domain of p53 and PC4 exemplifies acidic activation domains as single-stranded DNA mimics. *J. Biol. Chem.* **284**, 21 728–21 737. (doi:10.1074/jbc.M109.006429)
282. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. 2008 Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9**(Suppl. 1), S1. (doi:10.1186/1471-2164-9-S1-S1)
283. Lee H *et al.* 2000 Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *J. Biol. Chem.* **275**, 29 426–29 432. (doi:10.1074/jbc.M003107200)
284. Walsh CT, Garneau-Tsodikova S, Gatto Jr GJ. 2005 Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372. (doi:10.1002/anie.200501023)
285. Kim E, Goren A, Ast G. 2008 Alternative splicing: current perspectives. *Bioessays* **30**, 38–47. (doi:10.1002/bies.20692)
286. Wang WY, Barratt BJ, Clayton DG, Todd JA. 2005 Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.* **6**, 109–118. (doi:10.1038/nrg1522)
287. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000 Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257.
288. Sakakibara D *et al.* 2009 Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105. (doi:10.1038/nature07814)
289. Renault M, Tommassen-van Boxtel R, Bos MP, Post JA, Tommassen J, Baldus M. 2012 Cellular solid-state nuclear magnetic resonance spectroscopy. *Proc. Natl Acad. Sci. USA* **109**, 4863–4868. (doi:10.1073/pnas.1116478109)
290. Adler MJ *et al.* 2011 Small-molecule inhibitors of protein–protein interactions. Berlin, Germany: Springer.
291. van der Sloot AM, Kiel C, Serrano L, Stricher F. 2009 Protein design in biological networks: from manipulating the input to modifying the output. *Protein Eng. Des. Sel.* **22**, 537–542. (doi:10.1093/protein/gzp032)
292. Stranges PB, Machius M, Miley MJ, Tripathy A, Kuhlman B. 2011 Computational design of a

- symmetric homodimer using beta-strand assembly. *Proc. Natl Acad. Sci. USA* **108**, 20 562–20 567. (doi:10.1073/pnas.1115124108)
293. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, Baker D. 2011 Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332**, 816–821. (doi:10.1126/science.1202617)
294. Karanikolas J *et al.* 2011 A de novo protein binding pair by computational design and directed evolution. *Mol. Cell* **42**, 250–260. (doi:10.1016/j.molcel.2011.03.010)
295. Der BS, Kuhlman B. 2011 Biochemistry. From computational design to a protein that binds. *Science* **332**, 801–802. (doi:10.1126/science.1207082)
296. Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szymanski T, Dokholyan NV, Kuhlman B. 2010 Computational design of a PAK1 binding protein. *J. Mol. Biol.* **400**, 257–270. (doi:10.1016/j.jmb.2010.05.006)
297. Khakshoor O, Lin AJ, Korman TP, Sawaya MR, Tsai SC, Eisenberg D, Nowick JS. 2010 X-ray crystallographic structure of an artificial beta-sheet dimer. *J. Am. Chem. Soc.* **132**, 11 622–11 628. (doi:10.1021/ja103438w)
298. Potapov V, Reichmann D, Abramovich R, Filchinski D, Zohar N, Ben Halevy D, Edelman M, Sobolev V, Schreiber G. 2008 Computational redesign of a protein–protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J. Mol. Biol.* **384**, 109–119. (doi:10.1016/j.jmb.2008.08.078)
299. Kortemme T, Joachimiak LA, Bullock AN, Schuler AD, Stoddard BL, Baker D. 2004 Computational redesign of protein–protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379. (doi:10.1038/nsmb749)
300. Havranek JJ, Harbury PB. 2003 Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52. (doi:10.1038/nsb877)
301. Grigoryan G, Kim YH, Acharya R, Axelrod K, Jain RM, Willis L, Drndic M, Kikkawa JM, DeGrado WF. 2011 Computational design of virus-like protein assemblies on carbon nanotube surfaces. *Science* **332**, 1071–1076. (doi:10.1126/science.1198841)
302. Grigoryan G, Reinke AW, Keating AE. 2009 Design of protein–interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864. (doi:10.1038/nature07885)
303. Stewart ML, Fire E, Keating AE, Walensky LD. 2010 The MCL-1 BH3 helix is an exclusive MCL-1 inhibitor and apoptosis sensitizer. *Nat. Chem. Biol.* **6**, 595–601. (doi:10.1038/nchembio.391)
304. Liu S, Zhu X, Liang H, Cao A, Chang Z, Lai L. 2007 Nonnatural protein–protein interaction-pair design by key residues grafting. *Proc. Natl Acad. Sci. USA* **104**, 5330–5335. (doi:10.1073/pnas.0606198104)
305. Reynolds KA, Hanes MS, Thomson JM, Antczak AJ, Berger JM, Bonomo RA, Kirsch JF, Handel TM. 2008 Computational redesign of the SHV-1 beta-lactamase/beta-lactamase inhibitor protein interface. *J. Mol. Biol.* **382**, 1265–1275. (doi:10.1016/j.jmb.2008.05.051)
306. Demchenko AP. 2001 Recognition between flexible protein molecules: induced and assisted folding. *J. Mol. Recognit.* **14**, 42–61.
307. Joachimiak LA, Kortemme T, Stoddard BL, Baker D. 2006 Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface. *J. Mol. Biol.* **361**, 195–208. (doi:10.1016/j.jmb.2006.05.022)
308. Meireles LM, Mustata G. 2011 Discovery of modulators of protein–protein interactions: current approaches and limitations. *Curr. Top. Med. Chem.* **11**, 248–257.
309. Berg T. 2008 Small-molecule inhibitors of protein–protein interactions. *Curr. Opin. Drug Discov. Dev.* **11**, 666–674.
310. Fischer PM, Lane DP. 2004 Small-molecule inhibitors of the p53 suppressor HDM2: have protein–protein interactions come of age as drug targets? *Trends Pharmacol. Sci.* **25**, 343–346. (doi:10.1016/j.tips.2004.04.011)
311. Fry DC. 2012 Small-molecule inhibitors of protein–protein interactions: how to mimic a protein partner. *Curr. Pharm. Des.* **18**, 4679–4684.
312. D'Abamo CM, Archambault J. 2011 Small molecule inhibitors of human papillomavirus protein–protein interactions. *Open Virol. J.* **5**, 80–95. (doi:10.2174/1874357901105010080)
313. Cardinale D *et al.* 2011 Protein–protein interface-binding peptides inhibit the cancer therapy target human thymidylate synthase. *Proc. Natl Acad. Sci. USA* **108**, E542–E549. (doi:10.1073/pnas.1104829108)
314. Tse C *et al.* 2008 ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res.* **68**, 3421–3428. (doi:10.1158/0008-5472.CAN-07-5836)
315. Morelli X, Bourgeois R, Roche P. 2011 Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* **15**, 475–481. (doi:10.1016/j.cbpa.2011.05.024)
316. Pommier Y, Marchand C. 2011 Interfacial inhibitors: targeting macromolecular complexes. *Nat. Rev. Drug Discov.* **11**, 25–36. (doi:10.1038/nrd3404)
317. Bourgeois R, Basse MJ, Morelli X, Roche P. 2010 Atomic analysis of protein–protein interfaces with known inhibitors: the 2P2I database. *PLoS ONE* **5**, e9598. (doi:10.1371/journal.pone.0009598)
318. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 2001 Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **46**, 3–26.
319. Whitty A, Kumaravel G. 2006 Between a rock and a hard place? *Nat. Chem. Biol.* **2**, 112–118. (doi:10.1038/nchembio0306-112)
320. Domling A. 2008 Small molecular weight protein–protein interaction antagonists: an insurmountable challenge? *Curr. Opin. Chem. Biol.* **12**, 281–291. (doi:10.1016/j.cbpa.2008.04.603)
321. Arkin MR, Whitty A. 2009 The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Curr. Opin. Chem. Biol.* **13**, 284–290. (doi:10.1016/j.cbpa.2009.05.125)
322. Fischer PM. 2008 Computational chemistry approaches to drug discovery in signal transduction. *Biotechnol. J.* **3**, 452–470. (doi:10.1002/biot.200700259)
323. Fry DC. 2008 Drug-like inhibitors of protein–protein interactions: a structural examination of effective protein mimicry. *Curr. Protein Pept. Sci.* **9**, 240–247.
324. Fry DC, Vassilev LT. 2005 Targeting protein–protein interactions for cancer therapy. *J. Mol. Med. (Berl)* **83**, 955–963. (doi:10.1007/s00109-005-0705-x)
325. Neugebauer A, Hartmann RW, Klein CD. 2007 Prediction of protein–protein interaction inhibitors by chemoinformatics and machine learning methods. *J. Med. Chem.* **50**, 4665–4668. (doi:10.1021/jm070533j)
326. Pagliaro L, Felding J, Audouze K, Nielsen SJ, Terry RB, Krog-Jensen C, Butcher S. 2004 Emerging classes of protein–protein interaction inhibitors and new tools for their development. *Curr. Opin. Chem. Biol.* **8**, 442–449. (doi:10.1016/j.cbpa.2004.06.006)
327. Higuero AP, Schreyer A, Bickerton GR, Pitt WR, Groom CR, Blundell TL. 2009 Atomic interactions and profile of small molecules disrupting protein–protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.* **74**, 457–467. (doi:10.1111/j.1747-0285.2009.00889.x)
328. Chen TS, Keating AE. 2012 Designing specific protein–protein interactions using computation, experimental library screening, or integrated methods. *Protein Sci.* **21**, 949–963. (doi:10.1002/pro.2096)
329. Shangary S, Wang S. 2009 Small-molecule inhibitors of the MDM2-p53 protein–protein interaction to reactivate p53 function: a novel approach for cancer therapy. *Annu. Rev. Pharmacol. Toxicol.* **49**, 223–241. (doi:10.1146/annurev.pharmtox.48.113006.094723)
330. Koes DR, Camacho CJ. 2011 Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure. *Bioinformatics* **28**, 784–791. (doi:10.1093/bioinformatics/btr717)
331. Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB. 2005 Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* **3**, e405. (doi:10.1371/journal.pbio.0030405)
332. Leduc AM, Trent JO, Wittliff JL, Bramlett KS, Briggs SL, Chirgadze NY, Wang Y, Burris TP, Spatola AF. 2003 Helix-stabilized cyclic peptides as selective inhibitors of steroid receptor–coactivator interactions. *Proc. Natl Acad. Sci. USA* **100**, 11 273–11 278. (doi:10.1073/pnas.1934759100)
333. Li T, Saro D, Spaller MR. 2004 Thermodynamic profiling of conformationally constrained cyclic ligands for the PDZ domain. *Bioorg. Med. Chem. Lett.* **14**, 1385–1388. (doi:10.1016/j.bmd.2003.09.103)
334. Sievers SA *et al.* 2011 Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475**, 96–100. (doi:10.1038/nature10154)

335. Patra CR, Rupasinghe CN, Dutta SK, Bhattacharya S, Wang E, Spaller MR, Mukhopadhyay D. 2012 Chemically modified peptides targeting the PDZ domain of GIPC as a therapeutic approach for cancer. *ACS Chem. Biol.* **7**, 770–779. (doi:10.1021/cb200536r)
336. Walensky LD, Kung AL, Escher I, Malia TJ, Barbuto S, Wright RD, Wagner G, Verdine GL, Korsmeyer SJ. 2004 Activation of apoptosis *in vivo* by a hydrocarbon-stapled BH3 helix. *Science* **305**, 1466–1470. (doi:10.1126/science.1099191)
337. Gavathiotis E *et al.* 2008 BAX activation is initiated at a novel interaction site. *Nature* **455**, 1076–1081. (doi:10.1038/nature07396)
338. Benyamini H, Friedler A. 2010 Using peptides to study protein–protein interactions. *Future Med. Chem.* **2**, 989–1003.
339. Petsalaki E, Russell RB. 2008 Peptide-mediated interactions in biological systems: new discoveries and applications. *Curr. Opin. Biotechnol.* **19**, 344–350. (doi:10.1016/j.copbio.2008.06.004)
340. Jubb H, Higuero AP, Winter A, Blundell TL. 2012 Structural biology and drug discovery for protein–protein interactions. *Trends Pharmacol. Sci.* **33**, 241–248. (doi:10.1016/j.tips.2012.03.006)
341. Eichler J. 2008 Peptides as protein binding site mimetics. *Curr. Opin. Chem. Biol.* **12**, 707–713. (doi:10.1016/j.cbpa.2008.09.023)
342. Hecht I *et al.* 2009 A novel peptide agonist of formyl-peptide receptor-like 1 (ALX) displays anti-inflammatory and cardioprotective effects. *J. Pharmacol. Exp. Ther.* **328**, 426–434. (doi:10.1124/jpet.108.145821)
343. Hayouka Z *et al.* 2007 Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. *Proc. Natl Acad. Sci. USA* **104**, 8316–8321. (doi:10.1073/pnas.0700781104)
344. van Delft MF *et al.* 2006 The BH3 mimetic ABT-737 targets selective Bcl-2 proteins and efficiently induces apoptosis via Bak/Bax if Mcl-1 is neutralized. *Cancer Cell* **10**, 389–399. (doi:10.1016/j.ccr.2006.08.027)
345. Rajendran L, Knolker HJ, Simons K. 2010 Subcellular targeting strategies for drug design and delivery. *Nat. Rev. Drug Discov.* **9**, 29–42. (doi:10.1038/nrd2897)
346. Spolar RS, Record Jr MT. 1994 Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777–784. (doi:10.1126/science.8303294)