Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Structural Biology

Molecular origins of binding affinity: seeking the Archimedean point☆

Panagiotis L Kastritis and Alexandre MJJ Bonvin

Connecting three dimensional structure and affinity is analogous to seeking the 'Archimedean point', a vantage point from where any observer can quantitatively perceive the subject of inquiry. Here we review current knowledge and challenges that lie ahead of us in the quest for this Archimedean point. We argue that current models are limited in reproducing measured data because molecular description of binding affinity must expand beyond the interfacial contribution and also incorporate effects stemming from conformational changes/dynamics and long-range interactions. Fortunately, explicit modeling of various kinetic schemes underlying biomolecular recognition and confined systems that reflect *in vivo* interactions are coming within reach. This quest will hopefully lead to an accurate biophysical interpretation of binding affinity that would allow unprecedented understanding of the molecular basis of life through unraveling the *why's* of interaction networks.

Addresses
Bijvoet Center for Biomolecular Research, Science Faculty – Chemistry, Utrecht University, 3584CH Utrecht, The Netherlands

Corresponding author: Bonvin, Alexandre MJJ (a.m.j.j.bonvin@uu.nl)

Current Opinion in Structural Biology 2013, **23**:868–877

This review comes from a themed issue on **Protein-protein interactions**

Edited by **Joël Janin** and **Alexandre M J J Bonvin**

For a complete overview see the Issue and the Editorial

Available online 19th July 2013

0959-440X/$ – see front matter,

http://dx.doi.org/10.1016/j.sbi.2013.07.001

Introduction

Recognition processes between proteins involve functional interactions that underlie the cell's biology in a precise manner. Pathological conditions in cell physiology, leading, for example, to cancer or neurodegenerative diseases, always involve some degree of protein miscommunication. Despite current advances in the biophysical and biochemical methods used for the elucidation of the structure and kinetics of biomolecular interactions, the exact physicochemical basis of macromolecular recognition is still a matter of active discussion. For ease of reference, the relevant physicochemical quantities and constants are listed in Box 1. For a proper quantitative formulation of biomolecular recognition, availability of binding affinity data as well as atomic resolution structures of the protein–protein complexes and their free components is deemed crucial. In this review we ask the question: can we find the 'Archimedean point' in our odyssey for defining the binding affinity determinants of macromolecular recognition?

Archimedes (c. 287 BC–c. 212 BC), a famous Greek scientist and polymath, suggested during an argument that, given a sufficiently distant solid point away from the Earth (and a long enough lever), he could lift the whole earth: 'δῶς μοι πᾶ στῶ καὶ τὰν γᾶν κινάσω/give me somewhere to stand and I will move the earth'. The point where he would stand is called the 'Archimedean point', an eminent point from where any observer can quantitatively comprehend the subject of inquiry, which in our case, are structure–affinity relationships in protein–protein interactions.

The complexity of molecular recognition: the timescales of life

The extended range of dissociation (k_{off}, s^{-1}) and association (k_{on}, M^{-1} s^{-1}) rate constants (and their related equilibrium dissociation constant (K_d)) measured by *in vitro* assays directly reflects the various types of functional interactions in the cell. For example, protein-inhibitor complexes have a half-life ($1/k_{off}$) of days, even months — as measured, for example, by Vincent and Lazdunski [1] in the case of the interaction between trypsin and the pancreatic trypsin inhibitor, where the K_d is 60 fM at $T = 25°C$ and pH = 8. On the other side of the spectrum, electron transfer complexes that carry out redox reactions within a fraction of a second lead to transient interactions in the μM range. In the case of phosphorylation, or other post-translational modifications linked to metabolism regulation, the corresponding half-lives of the formed complexes diverge significantly, even in simple reactions (where one protein is the phosphodonor, usually a kinase, and the other the phosphoacceptor): for example, half-lives ranging from seconds for CheY and CheB [2] to several hours for OmpR and Spo0F [3] have been reported.

Binding affinity (expressed in physicochemical terms as the K_d) may span over 12 orders of magnitude, highlighting cellular function. For example, in the case

Box 1 Terminology

Partition function of a complex, Q	$Q = q_{int}\, q_{tr}\, q_{rot}\, q_{vib}\, q_{conf}\, q_{solv}$ [a]
Law of mass action [b]	$\alpha A + \beta B \rightarrow \gamma C$, $K_{eq} = [C]^{\gamma}/[A]^{\alpha}[B]^{\beta}$
Equilibrium dissociation constant	$K_d = c^0/K_{eq} = k_{off}/k_{on}$
Standard state (dissociation) free enthalpy	$\Delta G^{\circ}_{diss} = -RT \ln(K_d/c^0)$
(Dissociation) free energy, enthalpy, entropy [c]	$\Delta G_d = \Delta H_d - T\Delta S_d$
Entropy	$\Delta S_d = -d(\Delta G_d)/dT$
Heat capacity	$\Delta C_p = d(\Delta H_d)/dT$
Standard state	$p^0 = 1$ bar, $c^0 = 1$ mol L^{-1}
Gas constant	$R = 1.986$ cal mol^{-1} K^{-1}

[a] *Contributions*: q_{int}, interface; q_{tr}, translational; q_{rot}, rotational; q_{vib}, vibrational; q_{conf}, conformational; q_{solv}, solvent.
[b] Assuming $\alpha = \beta = \gamma = 1$, [X] and K_{eq} in M L^{-1} units.
[c] A positive ΔH_d favors association and a positive ΔS_d dissociation.

of reversible cell–cell adhesion processes, extremely low affinities are favored, in the mM [4] to μM range [5]. This is because recognition of cell surface molecules is multivalent and avidity-driven, and rapid focal adhesion turnover must mediate integrin signaling [4]. On the other side of the K_d spectrum, proteases, RNases and DNases, if not immediately neutralized and strictly regulated, will damage the cell irreversibly. That's why their inhibitors, for example, cognate inhibitors of Trypsin, Ribonuclease A, or Colicin E9, bind to their respective partners with K_d's lower than 6E−14 M.

An everlasting fondness: buried surface area and binding affinity

Modeling binding affinity is a complex problem, not only because of the timescales involved, but also in terms of understanding how the binding process occurs. Binding can take place via a simple lock-and-key (Fischer's) mechanism, without any obvious conformational change: for example, the binding of the bovine pancreatic trypsin inhibitor (BPTI) to trypsin with subpicomolar K_d follows a simple 1:1 monovalent and reversible two-state binding reaction. When comparing the crystal structures of the unbound conformers with that of the complex, hardly any changes in the conformation of their interface residues can be observed (root-mean-square-deviation (RMSD) < 0.3 Å). Many more complexes with known K_d's bind with only minor re-orientations of their side-chains, therefore, in a 'near-rigid' manner [6••]. Stein *et al.* [7] recently concluded, that Fischer's model holds when it comes to protein binding after studying >12 000 domain interactions. They also pointed out that, for flexible complexes, the bound state is often accessible via intrinsic motions of the free state, which would be consistent with a conformational selection mechanism. For this binding mechanism to occur, unbound conformations resembling the bound state must pre-exist.

For 'near-rigid' complexes, the Buried Surface Area (BSA) has been shown to relate to binding affinity with a Pearson's correlation coefficient $R = 0.54$ (P-value < 0.01) for 70 complexes with various functions [6••] (Figure 1a). This simple relation has a sound thermodynamic basis related to the hydrophobic effect for hydrocarbons [8,9]. Some assumptions are however needed to understand this contribution in protein–protein complexes (see below). In this model, the dissociation free energy ΔG_{diss} is approximated by

$$\Delta G_{diss} = -RT \ln\left(\frac{K_d}{c^0}\right) \sim \sum_i \alpha_i \mathrm{BSA}_i \qquad (1)$$

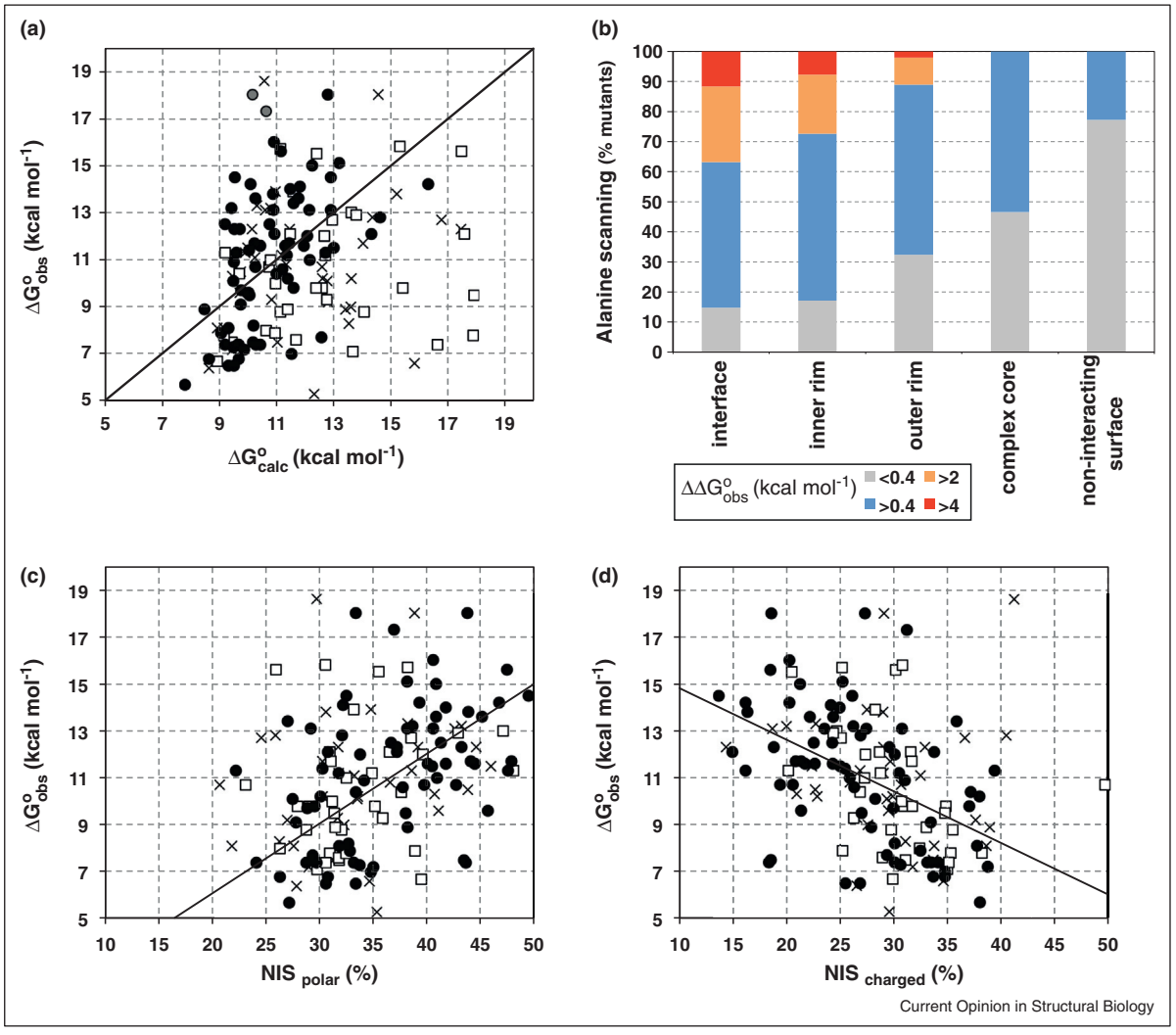
where $RT \sim 0.6$ kcal mol^{-1} at 298 K, c^0 is the concentration of the standard state (1 M by convention) and α_i is a hydration coefficient, which may be different for each atom type, and is expressed in kcal mol^{-1} Å^{-2}, similar to the surface tension. The BSA contains both hydrophilic ($\mathrm{BSA_{pol}}$) and hydrophobic surface fractions ($\mathrm{BSA_{apol}}$). The BSA-related part of Eqn 1 has also been split into polar and apolar terms, which yields improved correlations with ΔG_{diss} [10]. The exact values of the hydration coefficients have been a matter of debate even for simple systems [11].

A related concept in structure–affinity relationships is the binding efficiency, defined as the interaction energy per square ångström of BSA in the interface. The most efficient complexes (exhibiting high ΔG_{diss} and small BSA) generate up to 20 cal mol^{-1} Å^{-2} [12••], corresponding mostly to protein-inhibitor complexes, whereas the least efficient complexes can achieve efficiencies <25% of the maximal binding efficiency. Protein-inhibitor complexes often have a relatively small BSA (~1500 Å^2) and very low dissociation constants, whereas more 'flexible' complexes (flexible being used to denote complexes undergoing conformational changes upon binding), which bury larger surfaces, achieve smaller efficiencies. By considering a standard state $c^0 = 1$ M, a minimal contact area for a functional protein–protein interaction can be derived: Day *et al.* estimated it approximately 500 Å^2 [12••], reaching the same conclusion as a previous study by Janin who identified minimal functional interfaces of ~570 Å^2 from an analysis of crystal contact sizes [13].

Hot-spots in protein–protein interfaces: expanding the buried surface area model

Residues that, when substituted by alanine, have a major impact on the free energy of dissociation ΔG°_{diss} (>1.5 kcal mol^{-1}) are termed hot-spot residues (hot-spots). This was first reported by Clackson and Wells [14] who discovered that, in the human growth hormone-receptor interface, out of 26 mutations within the interface, six increased the K_d by a factor of 30, whereas the others did not have significant effects. Double-mutant cycle experiments have also shown that interface residues do display cooperativity [15]. The SKEMPI database [16] includes binding affinity data from over 700 alanine

Figure 1



Relationships between molecular properties and measured binding affinity. **(a)** Classical Buried Surface Area model ($\Delta G_{obs} = \alpha$BSA + β), relating the binding affinity to changes in the accessible surface area of the complexes; $R = 0.54$ (P-val $< 1E{-}4$, $N = 70$) for 'near-rigid' binders (iRMSD $\leq$ 1.0 Å, solid circles); $R = 0.16$ (P-val = 0.17, $N = 72$) for 'flexible' binders (iRMSD $>$ 1.0 Å, squares (1.0 Å $>$ iRMSD $\geq$ 1.5 Å) and crosses (iRMSD $>$ 1.5 Å) [6••]. **(b)** Alanine scanning mutagenesis data from SKEMPI [16] reveal that mutations on the entire surface, including the non-interacting surface, affect binding affinity; their impact decreases with increasing distance from the interface. **(c and d)** Percentages of polar (c) and charged (d) residues on the non-interacting surface show significant correlations with binding affinity for both rigid (polar: $R = 0.42$ (P-val $< 2E{-}4$, $N = 72$); charged: $R = -0.46$ (P-val $< 1E{-}4$, $N = 72$), respectively) and flexible complexes (polar: $R = 0.23$ (P-val $< 5E{-}2$, $N = 72$); charged: $R = -0.23$ (P-val $< 5E{-}2$, $N = 72$).

scanning experiments, next to other types of mutations, for 62 protein–protein complexes with known 3D structure. Looking at the location of these mutations using the Levy classification [17], hot-spots are always found in the interface and its direct periphery (Figure 1b). The Levy classification dissects the surface of protein–protein complexes according to changes in accessible surface area into: first, the interface region; second, the buried and third, exposed periphery; fourth, the complex interior and fifth, the complex exterior, beyond the exposed periphery. Hot-spots have been shown to typically bury more than 100 Å^2 of surface. They often correspond to residues with long side-chains such as Trp, Tyr or Arg. Considerable effects on binding affinity (2-fold changes, $\Delta\Delta G^{\circ}_{diss} > 0.4$ kcal mol^{-1}) have also been observed for sites distant from the interface, both in the protein core, but more interestingly, on the non-interacting surface (NIS) (Figure 1b). Recent studies by the Kalodimos group [18•,19] directly point to the role of conformational entropy in regulating binding affinity, providing a possible explanation for the effect of mutations remote from the interface: these remote mutations, by affecting the conformational entropy, are proposed to be possible on–off switches for the interaction. These studies have

provided the experimental proof for the existence of allosteric transitions, which do not involve conformational changes, as had been suggested in the 1980s [20].

Recently, we have discovered simple relationships between binding affinity and properties of the non-interacting surface (Figure 1c, d) for all complexes of the affinity benchmark [6••]: for instance, the percentages of polar and charged residues away from the interface show significant correlations with protein–protein binding affinity. Combining these with the simple BSA model for binding affinity (Figure 1a) leads to the formulation of a global surface model for binding affinity,[a] which accounts for the contribution of global physicochemical properties to the interaction strength, consistent with what alanine scanning data are reporting (Figure 1b).

The impact of conformational dynamics on binding affinity is substantial

A substantial fraction of protein–protein complexes undergo conformational rearrangements, from small local conformational changes to large disorder-to-order transitions. Some even display an increase in flexibility upon binding, as, for example, the core domain of p53 upon interaction with Hsp90 [21]. Conformational changes seem to be directly related to protein evolution: highly coevolving residues are frequently located in flexible regions [22]. The number of complexes exhibiting large conformational changes upon binding has been proposed by the Teichmann group to be much larger than currently estimated [23]. They associated relative protein accessibility to structural change: Using a simple power-law relationship with molecular weight, already discovered in the 1970s [24], they predicted the accessible surface area of well-folded proteins, proposing that observed deviations from it are linked to conformational changes. Another potential reason for the underestimation of conformational changes upon binding are the cryogenic temperatures regularly used in X-ray crystallography, which might introduce a bias toward reduced conformational motions. Room-temperature conditions should indeed be preferred, since they might reveal more functional motions [25]. Still, conformational changes between static structures at low temperatures can be observed and motions within one state could be different from conformational changes between two states. The reduced motions at cryogenic temperatures might mainly impact the proposed binding mechanisms (i.e. conformational selection versus induced fit), but not, *per se*, the conformational changes themselves.

For complexes that undergo 'large' conformational changes upon binding (iRMSD > 1.0 Å), the relation between BSA and binding affinity is masked by large changes in entropy that cannot be captured by the simple BSA model (Eqn 1, Figure 1a) [6••]. The latter nearly always overestimates the affinity, with some exceptions. This is in direct agreement with the pioneering report by Grünberg *et al.* [26] who showed that protein flexibility influences the thermodynamics of binding and may regulate protein–protein association.

Next to conformational changes within the interface, as measured by the interface RMSD (iRMSD) between free and bound states, disorder-to-order transitions are also often observed upon binding. In about one-third of the complexes of the binding affinity benchmark [6••], even for those cases classified as rigid, some residues in the interface are missing in either the free or the bound form, indicating differences in flexibility between those forms in the crystal state. The iRMSD measure alone does not capture those (since it can only be calculated on residues observed in both the free and bound states) and therefore appears to be a rather loose criterion for classifying flexibility in protein–protein interactions. The most flexible binders undergo extensive disorder-to-order transitions. For example, the complex between p38 MAPK and MAPK-activated protein kinase 2 [27] gains ~4000 Å^2 BSA from loops that are only observed in the bound conformation, while the iRMSD (measured on the segments common to the free and bound states) is only 1.9 Å.

Recently, various methods have been proposed to relate structural properties to conformational change [23,28]. One of their common features is that the free state's (intrinsic) flexibility directly relates to the observed conformational change. To date, no biophysical model has successfully modeled the contribution of conformational changes to binding affinity, except for the one proposed by Spolar and Record [29] who accounted for the entropy of folding per residue (~5.6 cal K^{-1}): They calculated the translational and rotational entropy along with a BSA-based hydrophobic contribution for several protein–protein complexes, and subtracted them from the measured ΔS° of association; if the result was near zero, the complexes were predicted as 'near-rigid' binders. The resulting excess entropy was attributed to conformational changes. This model compared well with estimates based on the entropy of folding per residue for those residues undergoing 'folding' upon binding. Considering the current extent of available affinity data [6••,30], a re-analysis to quantify and re-validate the extent of the contribution of conformational entropy to binding energy for various complexes should be possible. Explicit modeling of the energetics associated with conformational change has been estimated to decrease the error in binding affinity prediction [31]. The conformational entropy has been proposed to have a large contribution to the free energy of binding, reaching up to ~7 kcal mol^{-1} [32].

[a] Described in: Kastritis PL: Properties of the non-interacting surface modulate the binding affinity of protein–protein interactions. In *On the binding affinity of macromolecular complexes. Daring to ask why proteins interact*. PhD thesis, Utrecht University; 2012:161–208. *ISBN:* 978-9-03-935871-9.

Changes in conformational dynamics can be measured *in vitro* using NMR. Protein dynamics as measured by NMR T1, T2 and HET-NOE relaxation measurements can be interpreted in terms of conformational entropy [33–35]. Results on carbohydrate-protein interactions show that the conformational entropy contribution to ΔG_{diss} could be comparable in magnitude to that of the binding enthalpy [33]. Its quantification from NMR relaxation data is, however, not trivial and has been restricted so far to few biomolecular systems (e.g. [18•,36••,37]). To date, more than 30 structural ensembles of complexes determined by NMR, with known unbound structures and measured K_d's, are available in the Protein Data Bank (Table S1). Unfortunately, except for ubiquitin-related interaction data, none of these have been characterized in terms of conformational entropy albeit NMR relaxation data do exist for a few. An example of rigorous physicochemical modeling is available for the engineered complex between the Z(Taq) affibody and its binding partner (anti-Z(Taq) affibody) whose structure and that of its unbound constituents have both been solved by NMR [38]: the conclusion of this work is that the favorable hydrophobic surface desolvation upon complex formation is compensated by losses in translational and rotational entropy as well as unfavorable conformational entropy changes [38]. Such an analysis, however, requires data from Isothermal Titration Calorimetry (ITC) reporting on enthalpy, entropy, and heat capacity changes, as well as knowledge of the conformational changes and folding transitions from structural data.

Crowding effects on conformational dynamics and binding affinity remain elusive

Next to the classical conformational/flexibility changes and disorder-to-order transitions mentioned above, novel paradigms of complexes that form with *increased* or *altered* dynamics compared to their bound structures have been described (e.g. [21,37]) and extensively reviewed the last two years [39–41]. In addition, many interacting proteins are thought to function through molecular disorder (having no particular secondary structure), the so-called intrinsically disordered systems [42•,43].

The binding affinity of very flexible binders is related to k_{on} [44], in line with the proposed 'fly-casting' recognition mechanism for flexible complexes [45], where k_{on} is not diffusion-limited and displays an inverse dependence on solvent viscosity [46]. A controversy has been raised to whether disorder has an actual molecular explanation in the cellular context [42•,47•] and it is not just flexibility at its extreme, erroneously considered to be functional disorder [47•]. Accumulated data suggest that disorder could be exaggerated *in vitro* in the absence of crowding conditions (e.g. [48]). In contrast, another study showed that the accessible conformational space of the disordered polypeptide chain linked to functional recognition is mildly affected by crowding agents [49]. The impact of macromolecular crowding conditions on binding affinity is also controversial. Its effect might not be as severe as suggested from theory [50]. For example, the association rate of a β-lactamase to its protein inhibitor is ~25% lower *in vivo* than values reported from *in vitro* studies [51]. Highly electrostatic complexes are expected to show lower k_{on} in crowded conditions. To date, no conclusive evidence on the effect of macromolecular crowding on binding affinity (or conformational dynamics) is available since methods for *in vivo* binding affinity measurements are still in their infancy [52].

Current models for binding affinity prediction must include properties beyond the interface

Several models have been developed in the last years to model the binding affinity of macromolecular interactions (reviewed in [53•]). All consider only the bound conformation of the protein–protein complex, therefore ignoring the energetic contribution of the free state. These models only qualitatively correlate with binding affinity, at best [54•]. They are all relatively cheap in terms of computational costs compared to analytical calculations using molecular dynamics simulations in combination with PBSA calculations for instance [55]. MM-PBSA calculations have shown in a few specific cases to reasonably reproduce binding affinities, for example, in the case of the interaction between RAD51 and BRC peptides [56] and the H-Ras/C-Raf1 complex [57]. So far, none of the proposed models has achieved a better prediction than the simple model proposed by Horton and Lewis more than 20 years ago [10]. However, even this simple model often overestimates the binding affinity, in particular for flexible binders [6••]. The latter, typically in the mM–μM K_d range, often exhibit high BSA due to conformational changes. Simple BSA-based models considering only the bound form clearly overestimate the binding affinity by completely neglecting the loss of entropy for such systems. Current algorithms for affinity prediction rely heavily on interface interactions [53•,58]. They have been successful in specific cases [59], but, due to their complexity (e.g. large number of descriptors, support vector machine-based or neural network-based, *etc.*) (e.g. [60]), can often not fully explain the biophysical factors governing binding affinity. Besides that, they also unanimously neglect the role of the remainder of the surface, the non-interacting surface, whose global physicochemical properties are also related to binding affinity (Figure 1c, d) [6••]. The non-negligible role of the non-interacting surface is supported by experimental mutagenesis studies that have unanimously shown that both protein interior and the non-interacting surface do affect the free energy of binding (Figure 1b).

Other environment effects on protein–protein binding affinity that have not been included so far in any model are buffer conditions, Temperature, pH and salt concentration. Some models have included the dependence of

Figure 2

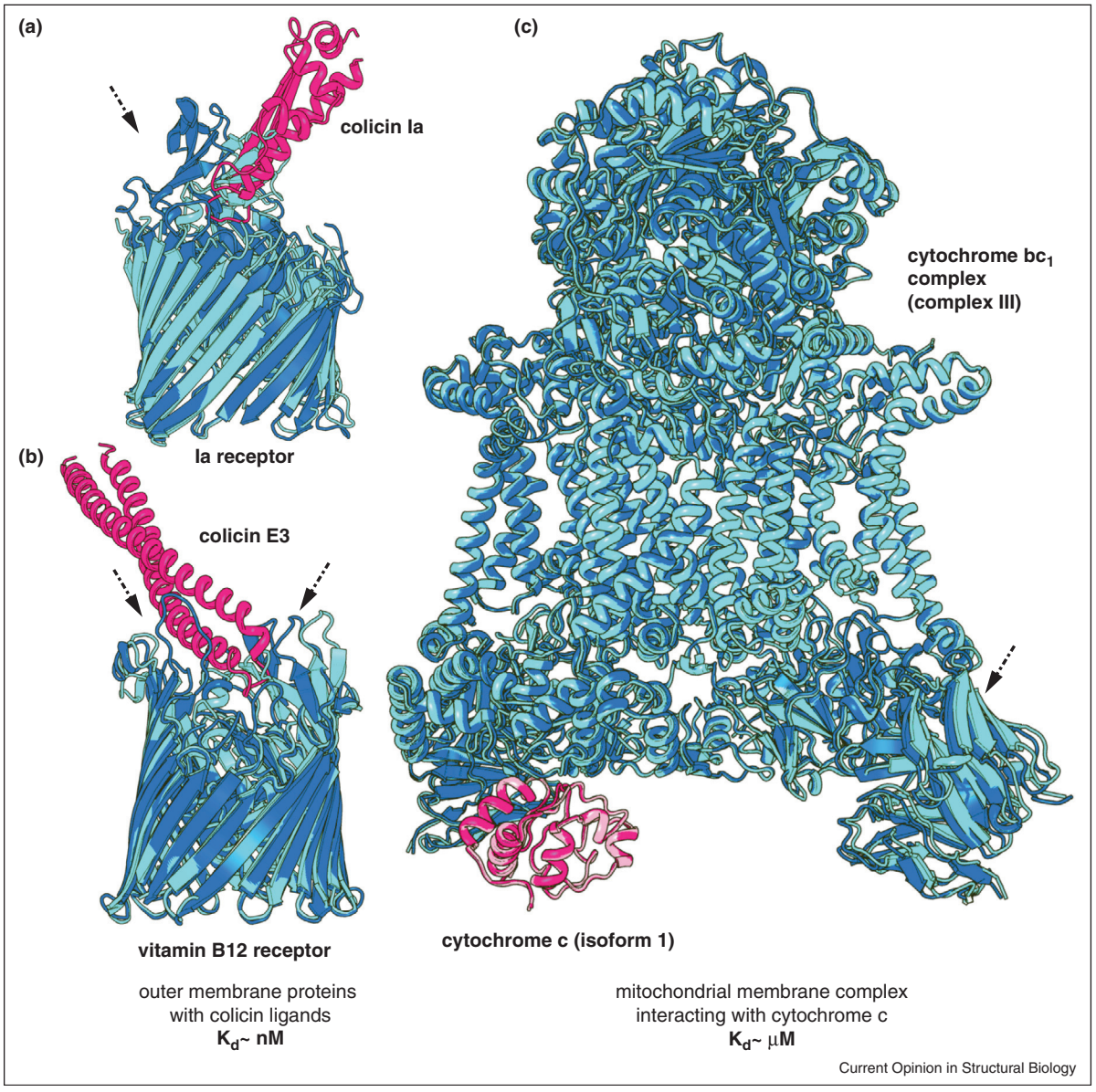


Illustration of three membrane complexes with bound and unbound structures and experimentally measured binding affinity present in the Protein Data Bank. Black arrows indicate the approximate location of conformational changes that occur upon binding. **(a)** Complex between colicin Ia (PDB ID: 2HDI, chain B) and its membrane receptor (PDB ID: 2HDF) ($K_d \sim$ 1E−10 M): domain rearrangements required for opening of the receptor are observed in the extracellular side, to allow the insertion of the colicin molecule. **(b)** Complex of colicin E3 (PDB ID: 1UJW, chain B) to the vitamin B12 receptor (PDB ID: 2GUF): The binding induces conformational changes that include loop reorientations as well as secondary structure rearrangements all over the interface in the extracellular side; the interaction is strong, in the nM range ($K_d \sim$ 9E−10 M). **(c)** bc_1 Complex (complex III) of the respiratory chain (PDB ID: 1KB9) in complex with cytochrome c oxidase (PDB ID: 1YCC): The interaction (PDB ID: 3CX5) may induce an allosteric change in the structure; the equilibrium dissociation constant is in the μM range.

k_{on} on the salt concentration using Debye–Hückel-like approximations, but not in the context of K_d modeling [61,62]. Especially, changes in pH could change the K_d by two orders of magnitude (~2.8 kcal mol^{-1}). Models to explicitly account for protonation/deprotonation events in protein–protein interactions are under active development [63,64], but have not yet been integrated in binding affinity models. Finally, as discussed above, crowding effects and molecular flexibility are not yet understood and both experimentalists and theoreticians will have to work together to understand their impact on binding affinity.

Conclusions and perspectives

Where the Archimedean point lies for understanding contributors to protein–protein binding affinity and if it actually exists

Figure 3

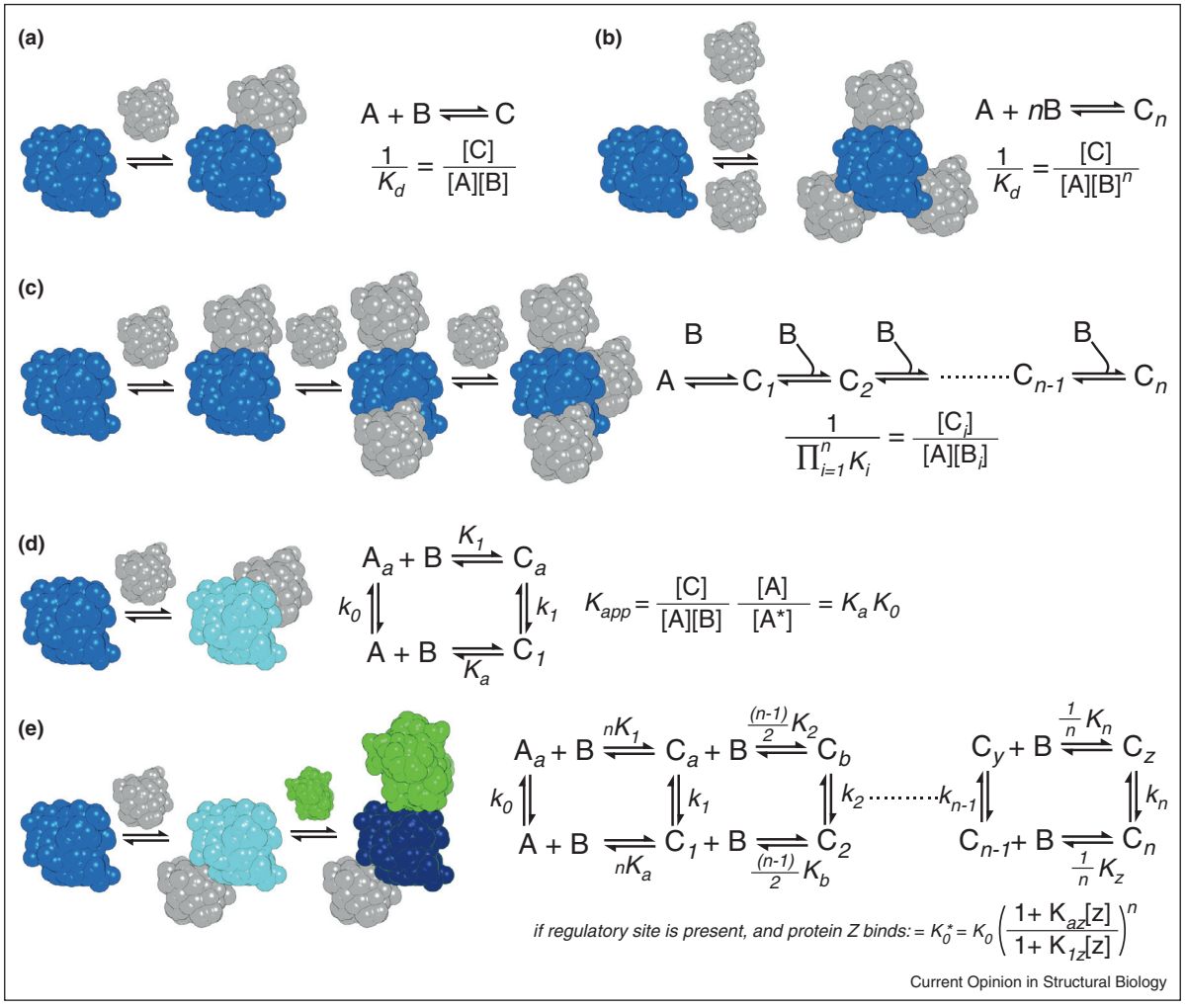


Illustration of kinetic models linked to various binding processes. Proteins A and B are represented in blue and gray spheres, respectively. Different (allosteric) states of protein A are indicated by various blue shades. A modulator protein Z is indicated in green. **(a)** 1:1 binding model. **(b)** Concerted binding. **(c)** Sequential binding. **(d and e)** Allosteric models with (d) single binding site and allosteric change and (e) multiple allosteric changes, including a modulator protein and sequential complexation. In the kinetic schemes, [A], [B], [C], [Z] denote the concentrations of protein A, protein B, their derived complex C and a modulating protein Z, respectively. The $K_{(d,i,a...z,1...n)}$'s denotes various equilibrium constants and $k_{(0-n)}$ the equilibrium constant for interconversion between states in the allosteric models d and e.

are still open questions. What we do know, however, is that, for 'near-rigid' complexes, changes in the accessible surface area are related to the binding affinity in a near quantitative manner. Further, for all types of complexes, that is, including flexible ones, global surface properties must play a role in binding affinity as well — but how is still an open question. Hot-spots are usually not observed outside the interface and its close periphery. Still, residues can contribute significantly to the binding free energy, even at distal sites on the non-interacting surface.

The impact of conformational changes or changes in molecular flexibility is now appreciated to be substantial but lacks a proper quantification. Pioneering structure–function studies of the complement system, an immune defense mechanism present in both vertebrates and invertebrates point to the fact that cellular signaling is based on a plethora of complex regulatory mechanisms for protein–protein interactions, mediated by domain rearrangements (up to ~100 Å), allosteric auto-activation controls, substrate-product binding and flexibility changes [65••]. Available and forthcoming crystallographic and NMR data, in combination with kinetic/thermodynamic measurements, should provide more precise information on the role of conformational changes and molecular flexibility on binding affinity.

There are further many open challenges to be addressed and opportunities to be taken to bring us closer to the Archimedean point. Modeling protein hydration, protonation and polarizability should lead to a more thorough understanding of interaction phenomena. Studying membrane systems with known bound and unbound structures (Figure 2) and modeling their affinities, for example, as the Honig group did for cadherin clustering [66••], should highlight the impact of the conformational confinement in the lipid environment. It is indeed expected that membrane anchoring is more frequent than anticipated in signaling cascades [67] and that the lipid environment has a role in regulating protein–lipid and protein–protein interactions [68••]. Furthermore, in order to tackle complex multi-component systems, we will have to reach beyond the 1:1 protein–protein interaction model for two-state kinetics for which most computational models have been developed so far, as more kinetic schemes currently proposed for macromolecular recognition await experimental data for modeling (Figure 3). Despite that, for multi-component assemblies, the thermodynamic data almost always concern binary subassemblies; obtaining reliable thermodynamic quantities for such systems is currently unrealistic, except in few specific cases may-be.

Finally, technological advances in binding affinity measurements are required to push the limits forward. We expect that methods to measure concentrations and rate constants *in vivo* will become more prevalent, even measuring in-cell pH changes in real time [69••], thereby hopefully improving both modeling and experimentation. For example, a wealth of data are available for actin polymerization *in vitro* [70•], but models developed for this exact phenomenon do not yet include effects of hundreds of protein modulators regulating the polymerization process. This complexity in protein–protein interactions will require modelers to think outside the box and unify available, accumulated and future knowledge from diverse research fields.

Acknowledgement

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.sbi.2013.07.001.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Vincent JP, Lazdunski M: **Trypsin-pancreatic trypsin inhibitor association. Dynamics of the interaction and role of disulfide bridges**. *Biochemistry* 1972, **11**:2967-2977.

2. Hess JF, Oosawa K, Kaplan N, Simon MI: **Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis**. *Cell* 1988, **53**:79-87.

3. Zapf J, Madhusudan M, Grimshaw CE, Hoch JA, Varughese KI, Whiteley JM: **A source of response regulator autophosphatase activity: the critical role of a residue adjacent to the Spo0F autophosphorylation active site**. *Biochemistry* 1998, **37**:7725-7732.

4. Vaynberg J, Fukuda T, Chen K, Vinogradova O, Velyvis A, Tu Y, Ng L, Wu C, Qin J: **Structure of an ultraweak protein–protein complex and its crucial role in regulation of cell morphology and motility**. *Mol Cell* 2005, **17**:513-523.

5. van der Merwe PA, Davis SJ: **Molecular interactions mediating T cell antigen recognition**. *Annu Rev Immunol* 2003, **21**:659-684.

6. •• Kastritis PL, Moal IH, Hwang H, Weng Z, Bates PA, Bonvin AM, Janin J: **A structure-based benchmark for protein–protein binding affinity**. *Protein Sci* 2011, **20**:482-491.
Using 144 protein–protein complexes with known bound and unbound structures along with the measured binding affinity, the authors rationalize the effect of conformational change on binding affinity and demonstrate the quantitative relation of buried surface area and free energy change.

7. Stein A, Rueda M, Panjkovich A, Orozco M, Aloy P: **A systematic study of the energetics involved in structural changes upon association and connectivity in protein interaction networks**. *Structure* 2011, **19**:881-889.

8. Chothia C: **Hydrophobic bonding and accessible surface area in proteins**. *Nature* 1974, **248**:338-339.

9. Hermann RB: **Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area**. *J Phys Chem* 1972, **76**:2754-2759.

10. Horton N, Lewis M: **Calculation of the free energy of association for protein complexes**. *Protein Sci* 1992, **1**:169-181.

11. Janin J: **Ångströms and calories**. *Structure* 1997, **5**:473-479.

12. •• Day ES, Cote SM, Whitty A: **Binding efficiency of protein–protein complexes**. *Biochemistry* 2012, **51**:9124-9136.
The authors examine both statistically and experimentally the relationship between binding affinity and interface size and assess the efficiency of protein–protein binding. They conclude, among others, that a minimal contact area of ~500 Å^2 is required for a stable complex.

13. Janin J: **Specific versus non-specific contacts in protein crystals**. *Nat Struct Biol* 1997, **4**:973-974.

14. Clackson T, Wells JA: **A hot spot of binding energy in a hormone–receptor interface**. *Science* 1995, **267**:383-386.

15. Reichmann D, Rahat O, Cohen M, Neuvirth H, Schreiber G: **The molecular architecture of protein–protein binding sites**. *Curr Opin Struct Biol* 2007, **17**:67-76.

16. Moal IH, Fernandez-Recio J: **SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models**. *Bioinformatics* 2012, **28**:2600-2607.

17. Levy ED: **A simple definition of structural regions in proteins and its use in analyzing interface evolution**. *J Mol Biol* 2010, **403**:660-670.

18. • Tzeng SR, Kalodimos CG: **Protein activity regulation by conformational entropy**. *Nature* 2012, **488**:236-240.
The first experimental evidence for allosteric transitions in protein–DNA recognition involving dynamics rather than conformational changes, derived using a combination of biochemical, NMR and ITC data.

19. Tzeng SR, Kalodimos CG: **Allosteric inhibition through suppression of transient conformational states**. *Nat Chem Biol* 2013, **9**:462-465.

20. Cooper A, Dryden DT: **Allostery without conformational change. A plausible model**. *Eur Biophys J* 1984, **11**:103-109.

21. Park SJ, Borin BN, Martinez-Yamout MA, Dyson HJ: **The client protein p53 adopts a molten globule-like state in the presence of Hsp90**. *Nat Struct Mol Biol* 2011, **18**:537-541.

22. Jeon J, Nam HJ, Choi YS, Yang JS, Hwang J, Kim S: **Molecular evolution of protein conformational changes revealed by a**

network of evolutionarily coupled residues. *Mol Biol Evol* 2011, **28**:2675-2685.

23. Marsh JA, Teichmann SA: **Relative solvent accessible surface area predicts protein conformational changes upon binding**. *Structure* 2011, **19**:859-867.

24. Janin J: **Surface area of globular proteins**. *J Mol Biol* 1976, **105**:13-14.

25. Fraser JS, van den Bedem H, Samelson AJ, Lang PT, Holton JM, Echols N, Alber T: **Accessing protein conformational ensembles using room-temperature X-ray crystallography**. *Proc Natl Acad Sci USA* 2011, **108**:16247-16252.

26. Grünberg R, Nilges M, Leckner J: **Flexibility and conformational entropy in protein–protein binding**. *Structure* 2006, **14**:683-693.

27. White A, Pargellis CA, Studts JM, Werneburg BG, Farmer BT 2nd: **Molecular basis of MAPK-activated protein kinase 2:p38 assembly**. *Proc Natl Acad Sci USA* 2007, **104**:6353-6358.

28. Karaca E, Bonvin AM: **A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes**. *Structure* 2011, **19**:555-565.

29. Spolar RS, Record MT Jr: **Coupling of local folding to site-specific binding of proteins to DNA**. *Science* 1994, **263**:777-784.

30. Dey S, Pal A, Chakrabarti P, Janin J: **The subunit interfaces of weakly associated homodimeric proteins**. *J Mol Biol* 2010, **398**:146-160.

31. Kamisetty H, Ramanathan A, Bailey-Kellogg C, Langmead CJ: **Accounting for conformational entropy in predicting binding free energies of protein–protein interactions**. *Proteins* 2011, **79**:444-462.

32. Wand AJ, Moorman VR, Harpole KW: **A surprising role for conformational entropy in protein function**. *Top Curr Chem* 2013, http://dx.doi.org/10.1007/128_2012_418. (in press).

33. Akke M, Brueschweiler R, Palmer AG III: **NMR order parameters and free energy: an analytical approach and its application to cooperative calcium(2+) binding by calbindin D9k**. *J Am Chem Soc* 1993, **115**:9832-9833.

34. Li Z, Raychaudhuri S, Wand AJ: **Insights into the local residual entropy of proteins provided by NMR relaxation**. *Protein Sci* 1996, **5**:2647-2650.

35. Yang D, Kay LE: **Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding**. *J Mol Biol* 1996, **263**:369-382.

36. •• Diehl C, Engstrom O, Delaine T, Hakansson M, Genheden S, Modig K, Leffler H, Ryde U, Nilsson UJ, Akke M: **Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3**. *J Am Chem Soc* 2010, **132**:14577-14589.
The authors use a variety of structural and biochemical methods to measure and compare (conformational) entropy and enthalpy of binding and conclude that changes in conformational entropy in protein–carbohydrate interactions are comparable in magnitude to the binding enthalpy.

37. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ: **The role of conformational entropy in molecular recognition by calmodulin**. *Nat Chem Biol* 2010, **6**:352-358.

38. Dogan J, Lendel C, Hard T: **Thermodynamics of folding and binding in an affibody:affibody complex**. *J Mol Biol* 2006, **359**:1305-1315.

39. Baldwin RL, Rose GD: **Molten globules, entropy-driven conformational change and protein folding**. *Curr Opin Struct Biol* 2013, **23**:4-10.

40. Marsh JA, Teichmann SA, Forman-Kay JD: **Probing the diverse landscape of protein flexibility and binding**. *Curr Opin Struct Biol* 2012, **22**:643-650.

41. Tzeng SR, Kalodimos CG: **Protein dynamics and allostery: an NMR view**. *Curr Opin Struct Biol* 2011, **21**:62-67.

42. • Uversky VN, Dunker AK: **The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure**. *F1000 Biol Rep* 2013, **5**:1.
An opinion article about the function of intrinsically disordered proteins — the authors claim that function is inherent in disordered proteins and provide various examples to support their claims.

43. Uversky VN: **Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes**. *Chem Soc Rev* 2011, **40**:1623-1634.

44. Prakash MK: **Insights on the role of (dis)order from protein–protein interaction linear free-energy relationships**. *J Am Chem Soc* 2011, **133**:9976-9979.

45. Sugase K, Dyson HJ, Wright PE: **Mechanism of coupled folding and binding of an intrinsically disordered protein**. *Nature* 2007, **447**:1021-1025.

46. Rogers JM, Steward A, Clarke J: **Folding and binding of an intrinsically disordered protein: fast, but not 'diffusion-limited'**. *J Am Chem Soc* 2013, **135**:1415-1422.

47. • Janin J, Sternberg MJ: **Protein flexibility, not disorder, is intrinsic to molecular recognition**. *F1000 Biol Rep* 2013, **5**:2.
An opinion article, published together with Ref. [42], contradicting the idea of functional disorder. It provides various examples from the literature that highlight that flexibility and not disorder is required for function, especially *in vivo*.

48. Dedmon MM, Patel CN, Young GB, Pielak GJ: **FlgM gains structure in living cells**. *Proc Natl Acad Sci USA* 2002, **99**:12681-12684.

49. Szasz CS, Alexa A, Toth K, Rakacs M, Langowski J, Tompa P: **Protein disorder prevails under crowded conditions**. *Biochemistry* 2011, **50**:5834-5844.

50. Zhou HX: **Influence of crowded cellular environments on protein folding, binding, and oligomerization: biological consequences and potentials of atomistic modeling**. *FEBS Lett* 2013, **587**:1053-1061.

51. Phillip Y, Kiss V, Schreiber G: **Protein-binding dynamics imaged in a living cell**. *Proc Natl Acad Sci USA* 2012, **109**:1461-1466.

52. Phillip Y, Schreiber G: **Formation of protein complexes in crowded environments — from in vitro to in vivo**. *FEBS Lett* 2013, **587**:1046-1052.

53. • Kastritis PL, Bonvin AM: **On the binding affinity of macromolecular interactions: daring to ask why proteins interact**. *J R Soc Interface* 2013, **10**:20120835.
An extensive review covering all aspects of protein–protein recognition serving as a primer for the next generation of modelers of affinity based on protein structure.

54. • Kastritis PL, Bonvin AM: **Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark**. *J Proteome Res* 2010, **9**:2216-2225 Corrigendum in: J Proteome Res 2011, **10**:921–922.
Using a dataset of known protein–protein complexes and related measured binding data, this article proves for the first time that current biophysical models are not able to relate their calculated energies to measured K_d's.

55. Wereszczynski J, McCammon JA: **Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition**. *Q Rev Biophys* 2012, **45**:1-25.

56. Cole DJ, Rajendra E, Roberts-Thomson M, Hardwick B, McKenzie GJ, Payne MC, Venkitaraman AR, Skylaris CK: **Interrogation of the protein–protein interactions between human BRCA2 BRC repeats and RAD51 reveals atomistic determinants of affinity**. *PLoS Comput Biol* 2011, **7**:e1002096.

57. Gohlke H, Case DA: **Converging free energy estimates: MM-PB(GB)SA studies on the protein–protein complex Ras-Raf**. *J Comput Chem* 2004, **25**:238-250.

58. Fleishman SJ, Whitehead TA, Strauch EM, Corn JE, Qin S, Zhou HX, Mitchell JC, Demerdash ON, Takeda-Shitaka M, Terashi G *et al.*: **Community-wide assessment of protein-interface modeling suggests improvements to design methodology**. *J Mol Biol* 2011, **414**:289-302.

59. Pierce BG, Weng Z: **A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes**. *Protein Sci* 2013, **22**:35-46.

60. Moal IH, Agius R, Bates PA: **Protein–protein binding affinity prediction on a diverse set of structures**. *Bioinformatics* 2011, **27**:3002-3009.

61. Qin S, Pang X, Zhou HX: **Automated prediction of protein association rate constants**. *Structure* 2011, **19**:1744-1751.

62. Schreiber G, Shaul Y, Gottschalk KE: **Electrostatic design of protein–protein association rates**. *Methods Mol Biol* 2006, **340**:235-249.

63. Jensen JH: **Calculating pH and salt dependence of protein–protein binding**. *Curr Pharm Biotechnol* 2008, **9**:96-102.

64. Mitra RC, Zhang Z, Alexov E: ***In silico* modeling of pH-optimum of protein–protein binding**. *Proteins* 2011, **79**:925-936.

65. •• Forneris F, Wu J, Gros P: **The modular serine proteases of the complement cascade**. *Curr Opin Struct Biol* 2012, **22**:333-341.
Authors combine recent structure–function knowledge of the complement system in the light of affinity data in a clear and comprehensible manner, concluding that conformational plasticity of interacting partners is essential to drive biological activity of the complement system (that is, to identify and eliminate pathogens or cellular debris).

66. •• Wu Y, Vendome J, Shapiro L, Ben-Shaul A, Honig B: **Transforming binding affinities from three dimensions to two with application to cadherin clustering**. *Nature* 2011, **475**:510-513.
Combining multiscale molecular dynamics simulations and theory of protein–protein binding, the authors convert K_d measured in 3D (Liters, L) to K_d measured in 2D (m^2), explicitly considering structure and dynamics of membrane-bound molecules.

67. Groves JT, Kuriyan J: **Molecular mechanisms in signal transduction at the membrane**. *Nat Struct Mol Biol* 2010, **17**:659-665.

68. •• Weingarth M, Prokofyev A, van der Cruijsen EA, Nand D, Bonvin AM, Pongs O, Baldus M: **Structural determinants of specific lipid binding to potassium channels**. *J Am Chem Soc* 2013, **135**:3983-3988.
The authors combine molecular dynamics simulations, solid-state NMR and functional assays to show that the KcsA membrane channel's specificity is regulated by its anionic lipid environment.

69. •• Modi S, Nizak C, Surana S, Halder S, Krishnan Y: **Two DNA nanomachines map pH changes along intersecting endocytic pathways inside the same cell**. *Nat Nanotechnol* 2013, **8**:459-467.
This article reports mapping of pH changes within cellular compartments in real-time by using simultaneously different DNA nanomachines as molecular sensors that have been programmed to enter living cells via different pathways.

70. • Ditlev JA, Mayer BJ, Loew LM: **There is more than one way to model an elephant. Experiment-driven modeling of the actin cytoskeleton**. *Biophys J* 2013, **104**:520-532.
An opinion article in the same line as this one, describing efforts of modeling the actin cytoskeleton and showing that experimentalists and modelers must work together to optimize both experimentation and generation of testable hypotheses, respectively.