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 phospho-donor, usually a kinase, and the other the phospho-acceptor): 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

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

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 (BSA_{pol}) and hydrophobic surface fractions (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