Biochemistry 412

Protein-Protein Interactions II

March 28, 2008
Protein interfaces are diverse in size, shape, composition and solvent content. In this figure, interfacial sidechain atoms are color coded as hydrophobic (orange), polar (green) or charged (red). Mainchain atoms are in blue and atoms outside the interface are gray. Water, when present in the crystal structure, is included as blue spheres. Both sides of each interface are shown. (a) The small and very hydrophobic interface formed between the erythropoietin receptor and an engineered mimic of erythropoietin (PDB code 1e0b). (b) The highly charged and well-solvated interface between interleukin-4 and its ρ receptor (PDB code 1iar). (c) The very large binding interface formed between tissue factor and factor Vlla (PDB code 1dan). All of the color figures in this article were created using the free molecular graphics program PyMOL [54].

Some ways that mutations can destabilize protein-protein interactions

Destabilizing perturbations in protein–protein interactions can arise from numerous sources as a consequence of mutation. Some mechanisms involve static interactions at the site of substitution: (a) loss of optimal van der Waals contacts; (b) loss of electrostatic pairings, which provide substantial binding energy compared to solvent; (c) loss of essential electrostatic pairings, which are not net-stabilising relative to solvent; and (d) loss of buried nonpolar surface area. However, others involve ensemble phenomena, which are not easily recognized or interpreted: (e) a discrete local conformational change; (f) local unfolding at the interface; (g) increased entropy of unbound states; (h) aggregation; and (i) global unfolding. In particular, changes in interface dynamics (e–g) are very hard to demonstrate or rule out experimentally because of the limitations of available technology. These effects, which relate to the plasticity of protein interfaces, account for much of the current uncertainty in the interpretation of alanine scanning data.

Four state model for the formation of protein complexes


Free-energy profile describing the pathway for the formation of a protein–protein complex (AB) from the free proteins A and B via the encounter complex AB*, the transition state AB† and the intermediate AB** (see Scheme 1). Lines 1 and 2 describe the interaction between a pair of proteins in the absence and in the presence of favourable electrostatic forces, respectively.
Today's topics:

• isothermal titration calorimetry

• phage display

• discovery of small molecule antagonists of protein-protein interactions
If you want to be “hard core” and really understand protein-protein interactions, you need to know more than just the free energies of association. You (ultimately) will need to know something about enthalpies, entropies, and heat capacities, too.
FIGURE 1. Three-dimensional density. Three-dimensional solvent number density distribution around myoglobin is shown as a slice computed from a molecular dynamics trajectory. The solvent density is overlaid with an average structure of myoglobin and contoured at 0.005 Å⁻³ (blue), 0.01 Å⁻³ (green), 0.02 Å⁻³ (yellow), and 0.035 Å⁻³ (red). The bulk solvent density is 0.033 Å⁻³. (a) Density from simulation, (b) density from prediction. Reproduced with permission of the authors from ref 24.

FIGURE 3. Comparison of simulation and experiment. A view of the hydration sites contained in the 39 independent Protein Data Bank entries for sperm whale myoglobin (blue dots) compared to the hydration number density maxima generated in the molecular dynamics simulation (yellow circles). Note the generally poor agreement between many discrete representations of hydration structure around the protein. Reproduced with permission of the authors from ref 22.

FIGURE 4. Global picture of protein hydration. This figure shows a one-dimensional slice through the protein-water interface. Markings “in” and “out” correspond to the interior and exterior of the protein. There are four distinct regions in the interface: (A) solvent-free region; (B) region of interpenetration of protein and solvent; (C) local maxima in the solvent density, “hydration sites”; and (D) remote region, where no distinct density maxima are found, but the diffusion rate is perturbed.

When two proteins form a complex, solvent must be displaced from the interfacial regions and the conformational freedom (configurational entropy) of the main chain and side chain atoms will change also.
The Gibbs free energy change, $\Delta G$, of an association reaction is temperature dependent and is described by

$$
\Delta G(T) = \Delta H(T_R) + \int_{T_R}^{T} \Delta C_p dT - T \Delta S(T_R)
$$

$$
- T \int_{T_R}^{T} \Delta C_p d \ln T
$$

(1)

$\Delta H$ and $\Delta S$ are the change in enthalpy and entropy, respectively, $\Delta C_p$ is the heat capacity change, and $T_R$ is an appropriate reference temperature. If $\Delta C_p$ is temperature-independent in the temperature interval of interest, Eq. (1) simplifies to

$$
\Delta G(T) = \Delta H(T_R) - T \Delta S(T_R)
$$

$$
+ \Delta C_p [T - T_R - T \ln(T/T_R)]
$$

(2)

Equations (1) and (2) show that the free energy of association has an enthalpy and an entropy component. The changes of enthalpy and entropy depend on temperature according to

\[ \Delta C_p = \frac{d(\Delta H)}{dT} = T \frac{d(\Delta S)}{dT} \]  

(3)

To characterize the thermodynamics of a binding reaction means to determine \( \Delta G, \Delta H \) and \( \Delta S \) at a given reference temperature and to obtain \( \Delta C_p \) to predict the change of the above three parameters with temperature.

Formalisms for Characterizing Binding Affinities

For a protein (P), ligand (A), and complex (P • A)

\[ P + A \xrightleftharpoons[k_d]{k_a} P \cdot A \]

where \([P]_{\text{total}} = [P] + [P \cdot A]\]

The association constant: \[ K_a = [P \cdot A]/[P][A] = k_a/k_d \]

The dissociation constant: \[ K_d = 1/K_a = [P][A]/[P \cdot A] \]

…please note that \(K_d\) has units of concentration, and so when \(K_d = [A]\) then \([P] = [P \cdot A]\), and thus \(K_d\) is equal to the concentration of the ligand A at the point of half-maximal binding.
Isothermal Titration Calorimetry Yields $\Delta H$ of Binding

Concentration independence of the interaction between Ras and the Ras-binding domain (RBD) of Ral guanine nucleotide dissociation stimulator (RalGDS). (a) Calorimetric titration of Ras into 6 µM (upper curve) and 50 µM (lower curve) RalGDS. The baseline was subtracted from the raw data and an off-set of 0.5 µcal/s was added to the upper curve. (b) The theoretical curves fitted to the integrated data yield $K_d = 1.2$ µM and $\Delta H^\circ = -15.0$ kcal/mol for RalGDS at 6 µM (open circles), and $K_d = 1.1$ µM and $\Delta H^\circ = -14.8$ kcal/mol for RalGDS at 50 µM (closed squares). Titrations conducted at protein concentrations varying by nearly 90% exhibit similar heats of reaction and dissociation constants. This indicates that, at both concentrations, the proteins possess the same oligomeric form and, in this system, the proteins are monomeric. Reproduced with permission from [15].
...and when you have $\Delta H$ and $\Delta G (= -RT\ln K_a)$, knowing the temperature you can calculate $\Delta S$. 
Some examples of experimentally-measured thermodynamic quantities for interacting proteins, measured using isothermal titration calorimetry:

<table>
<thead>
<tr>
<th>Thermodynamic binding parameters.</th>
<th>ΔH (kcal/M)</th>
<th>TΔS (kcal/M)</th>
<th>ΔTm (°C)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein–protein interactions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fungal xylanase + xylanase-inhibiting protein</td>
<td>−11.6</td>
<td>−1.5</td>
<td></td>
<td>[9]</td>
</tr>
<tr>
<td>Ubiquitin-fused complement-type repeat domains</td>
<td>−11.5</td>
<td>−3.9</td>
<td></td>
<td>[10]</td>
</tr>
<tr>
<td>5 and 6 + receptor-associated protein domain 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ubiquitin-fused complement-type repeat domains</td>
<td>−4.9</td>
<td>4.1</td>
<td></td>
<td>[10]</td>
</tr>
<tr>
<td>5 and 6 + receptor-associated protein domain 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferri-cytochrome c + ferri-cytochrome c peroxidase</td>
<td>−2.6</td>
<td>5.5</td>
<td></td>
<td>[11*]</td>
</tr>
<tr>
<td>(Ferri-cytochrome c + ferri-cytochrome c peroxidase) in trisaccharides (site 1)</td>
<td>−2.8</td>
<td>8.5</td>
<td></td>
<td>[12*]</td>
</tr>
<tr>
<td>(Ferri-cytochrome c + ferri-cytochrome c peroxidase) in trisaccharides (site 2)</td>
<td>−0.3</td>
<td>8.4</td>
<td></td>
<td>[12*]</td>
</tr>
<tr>
<td>Peripheral subunit-binding domain of dihydrolipoyl acetyltransferase + pyruvate decarboxylase</td>
<td>−8.4</td>
<td>4.5</td>
<td>4.8</td>
<td>[13]</td>
</tr>
<tr>
<td>Peripheral subunit-binding domain of dihydrolipoyl acetyltransferase + dihydrolipoyl dehydrogenase</td>
<td>2.2</td>
<td>14.8</td>
<td>6</td>
<td>[13,14]</td>
</tr>
<tr>
<td>Ras-binding domain of Raf + Ras</td>
<td>−5.2</td>
<td>4.5</td>
<td></td>
<td>[15]</td>
</tr>
<tr>
<td>Ras-binding domain of RalGDS + Ras</td>
<td>−14.5</td>
<td>−6.2</td>
<td></td>
<td>[15]</td>
</tr>
<tr>
<td>Ras-binding domain of Raf + Rap</td>
<td>−3.7</td>
<td>4.8</td>
<td></td>
<td>[15]</td>
</tr>
<tr>
<td>Ras-binding domain of RalGDS + Rap</td>
<td>−14.5</td>
<td>−4.7</td>
<td></td>
<td>[15]</td>
</tr>
</tbody>
</table>

Note: isothermal titration calorimetry also directly yields $n$, the stoichiometry of binding.

The maximal affinity of ligands and “ligand efficiency”

The maximal free energy contribution per non-hydrogen atom $\approx -1.5$ kcal/mol. The maximal total binding energy is 15-20 kcal/mol.

One more thing re ligands:

Fig. 1. Free energy of binding (in kcal/mol) for ligands and enzyme inhibitors plotted as a function of the number of nonhydrogen atoms in the ligand. See Table 1. A line with slope of 1.5 kcal/mol and an intercept of 0 is included as a visual aid to analysis. $\triangle$, Metal ions or metalloenzymes; $\blacktriangle$, small anions; $\circ$, natural ligands; $\bullet$, enzyme inhibitors.

Phage Display

An M13 phagemid vector designed for phage display. A phagemid vector contains origins of single-stranded (f1 ori) and double-stranded (322 ori) DNA replication and a selective marker, such as the β-lactamase gene (Amp') which confers resistance to ampicillin. For phage display, the phagemid also contains a cassette consisting of a promoter that drives transcription of an open reading frame encoding a secretion signal and the displayed protein fused to an M13 coat protein. The vector replicates in E. coli as a double-stranded plasmid, but coinfection with a helper phage results in the production of single-stranded DNA that is packaged into phage particles. The phage coat contains five different proteins, and polypeptides can be displayed as either amino-terminal fusions (with P3, P8, P7, or P9) or carboxy-terminal fusions (with P6, P8, or P3). The phage particles can be used in binding selections, and binding clones can be amplified by recycling through an E. coli host.

[as referenced in Sidhu (2000)]
*Curr. Opin. Biotechnol.* 11, 610]
Fig. 2. A simplified view of M13 bacteriophage assembly. Newly-synthesized coat proteins (white cylinders) are imbedded in the inner membrane (IM) with their N termini in the periplasm and their C termini in the cytoplasm. Single-stranded viral DNA is extruded through a pore complex (yellow cylinders) that spans the inner membrane and the outer membrane (OM). Coat proteins also interact with the pore complex, where they surround the DNA and are thus transferred from the bacterial membrane into the assembling phage coat. In this way, newly-assembled phage particles are extruded from the cytoplasm to the extracellular environment without lysis of the E. coli host. A heterologous protein (red circle) will be displayed on the phage surface if it is fused to a coat protein that can successfully incorporate into the phage coat.

Figure 1. In vitro selection with phage display. Polypeptides are displayed on phage particles that also contain the encoding DNA. Display is usually achieved by fusion to either the gene-3 minor coat protein which is located at one end of the particle (shown) or the gene-8 major coat protein which covers the length of the particle. The library is incubated with an immobilized target to select for binders, and nonbinding phage are removed by washing. Bound phage are then eluted and amplified in E. coli. Amplified phage pools can be subjected to additional rounds of selection, or alternatively, the sequences of individual binding polypeptides can be determined by sequencing the encapsulated DNA.

Fig. 1. Structure of a filamentous bacteriophage. (A) A diagram of the bacteriophage particle shows the single-stranded DNA core surrounded by a proteinaceous coat. The coat is composed mainly of P8 molecules that cover the length of the particle. The ends of the particle are capped by minor coat proteins: P3 and P6 at one end, and P7 and P9 at the other end. (B) Fiber X-ray studies reveal that the P8 molecules are arranged in a repeating array (PDB entry 1IFJ). A small section of the array is shown, with the P8 molecules rendered as ribbons. The N termini are exposed on the particle surface while the C termini are buried at the core. (C) The phage coat contains several thousand copies of P8, but only approximately five copies each of the four minor coat proteins. Each of the five different coat proteins has been successfully used as a platform for the functional display of heterologous polypeptides as either N- or C-terminal fusions. References are provided in the main text. (B), Figs. 4 and 5 were generated using the computer program INSIGHTII (Molecular Simulations, Inc., San Diego).

“Shotgun” alanine scanning with statistical free energies from wt/Ala phage ratios!

Figure 2. Shotgun alanine scan of hGH. A) Following either a “structural” or “functional” selection (white or black bars, respectively), the wt/Ala ratios at 19 mutated positions were determined by sequencing several hundred binding clones. Ratios greater than or less than one indicate mutations that are deleterious or beneficial to the selected trait, respectively. The wt/Ala ratios can be used to calculate the change in the free energy of binding for each Ala mutant relative to wt (ΔΔG_{Ala-wt}). 8) ΔΔG_{Ala-wt} values determined by shotgun scanning (y axis) are in good agreement with those determined by conventional alanine scanning (x axis) that involved the analysis of purified point-mutated proteins with surface plasmon resonance spectroscopy. The least squares linear fit of the data is shown (y = -0.01 + 1.0x; R = 0.88). C) When mapped onto the structure of hGH (PDB accession code 3HHR), the shotgun-scan results reveal a compact functional epitope composed of a small subset of contact side chains that contribute most of the binding energy to the interaction with hGHbp. Residues are colored according to their statistical ΔΔG_{Ala-wt} values obtained from the shotgun alanine scan: red, ΔΔG_{Ala-wt} > 2.0 kcal mol⁻¹; orange, 1.3 < ΔΔG_{Ala-wt} < 2.0 kcal mol⁻¹; yellow, 0.6 < ΔΔG_{Ala-wt} < 1.3 kcal mol⁻¹; cyan, ΔΔG_{Ala-wt} < 0.6 kcal mol⁻¹. The shotgun-scanning data were taken from ref. [11] by Weiss et al., while the conventional alanine-scan data were from ref. [10] by Cunningham and Wells. See the main text for further details.

Other interesting (and related) topics:

- yeast “two-hybrid” technology
- antibody engineering
- small molecule antagonists
The **yeast two-hybrid system** was originally developed by S. Fields & O.-K. Song [Nature 340, 245 (1989)]. It is an alternative means of linking a physical association based on protein-protein recognition with the underlying sequences encoding the protein partners. As with phage display, DNA sequencing of the genes for proteins emerging from the selection process allows the protein sequences to be identified.

**Bottom line:** phage display is usually used to characterize libraries containing mutants of a single protein with respect to each other, and yeast two-hybrid is usually used to characterize libraries containing genes for different wild type proteins. Both methods are normally concerned with stratifying the variants with respect to their affinities towards a given test protein.
Antibodies can almost always be used to provide an antagonist to a protein-protein interaction. Hence, they are popular as drugs for diseases where antagonizing such an interaction in the body* would be beneficial.

[*Note: obviously, only extracellular interactions are normally accessible to Abs.]
Figure 3 | Representative antibody formats. The modular domain architecture of immunoglobulins has been exploited to create a growing range of alternative antibody formats that spans a molecular-weight range of at least 12–150 kDa and a valency (n) range from monomeric (n = 1), dimeric (n = 2) and trimeric (n = 3) to tetrameric (n = 4) and possibly higher. For simplicity, all antibody formats are shown as being monospecific; that is, having one or more copies of identical antigen-binding sites. However, formats with a valency of two or more have also been used to create antibodies that have two or more (up to the valency of the format) distinct antigen-binding sites, which bind different antigens or different epitopes on the same antigen. The building block that is most frequently used to create novel antibody formats is the single-chain variable (V) domain antibody fragment (scFv), which comprises V domains from the heavy and light chain (V_H and V_L domain) joined by a peptide linker of up to ~15 amino-acid residues. Both the V_H-domain–peptide-linker–V_L-domain topology and the V_H-domain–peptide-linker–V_L-domain topology have been widely used. Constant (C) and V domains, inter-chain disulphide bonds (green bars) and glycosylation are depicted as in FIG. 2, and peptide and chemical linkers are shown as orange and red lines, respectively. C_H, C domain of immunoglobulin heavy chain; C_L, C domain of immunoglobulin light chain; dsFv, disulphide-stabilized scFv.
Antibodies are great, but what about small molecule antagonists of protein-protein interactions?

>>> Some people would rather take their medicine by popping a pill than by being jabbed with a needle!

Big historical problem: pharmaceutical companies were historically prejudiced against finding such drugs because of the belief that the protein-protein interface was too large and too “flat” to allow a small molecule to compete with the natural protein ligand. But….. Remember *ligand efficiencies*?? Also, protein interface surfaces were discovered to be unusually adaptive.
**Figure 1** | Examples of protein–protein interface hotspots. Alanine-scanning mutational analysis (replacing each amino acid, in turn, with alanine) was carried out on the contact surfaces of four pairs of interacting proteins. The resultant change in the binding free energy compared with interactions involving the wild-type protein (ΔΔG) is shown by colour coding amino-acid residues, from red (the most-disruptive changes) to green (those having little or no effect). Therefore, in each case, most of the free energy is contributed by a small number of residues (red), and this region is known as the hotspot	extsuperscript{10}. VEGF, vascular endothelial growth factor; Z domain, a derivative of a domain from *Staphylococcus aureus* protein A. (Image courtesy of W. DeLano, DeLano Scientific, Palo Alto, California.)

How proteins interact with natural protein partners vs. a small molecule antagonist

**IL-2:**

**Bcl-X<sub>L</sub>:**

**HDM2:**

**HPV-18/11 E2:**

[Diagram showing interactions between natural ligands and small molecule ligands for IL-2, Bcl-X<sub>L</sub>, HDM2, and HPV-18/11 E2.]

Green = contact surface for natural ligand

Orange = contact surface for small molecule

Interaction surfaces can be very adaptive --

Side chain (and some main chain) movements can create binding sites for small molecules that are not apparent in the 3D structures of complexes with the natural ligand.

(see, for example, http://www.nature.com/nature/journal/v450/n7172/extref/nature06526-s1.mov)
Ligand efficiency for small molecule antagonists of protein-protein interactions is surprisingly uniform and well below that maximum predicted efficiency of -1.5 kcal/mol per non-H atom [see earlier slide by Kuntz et al (1999)], but still good enough to allow drugs to be developed.

**Figure 5 | Relationship between compound potency and size for small molecules that inhibit protein–protein interactions.** For the highest-affinity fragments and small molecules that target protein–protein interfaces, the binding free energy (−ΔG) is plotted against the number of non-hydrogen atoms. Kₐ values were converted to free energy (kcal per mol) using standard-state conditions of 1 M concentration at a temperature of 300 K. Where a direct binding dissociation constant was not available or was not the lowest measured, the Kᵢ or IC₅₀ was used instead. The slope can be described by \( y = 0.24x \), and the correlation coefficient is 0.77. The linear relationship implies that there is a uniform ligand efficiency for these targets. Note that the first occurrence of Compound 1 (black circle) denotes the compound discussed in this article and depicted in Fig. 2, and the second occurrence (orange circle) denotes a molecule with a different chemical structure.