Lecture #8
The Cell as a Machine

High Specificity and Reversibility

In considering the problem of transcription factor binding in the nucleus and the great specificity that is called for to transcribe the right and not the wrong gene, there is a need to understand the important elements of protein-protein bonding. The treatment of the problem is slightly different than what we have discussed. As we will show today, the lifetime of large complexes is too long for an active cell to function. The basic problem is the dissociation of a complex once formed.

Signal Transduction From the Periphery to the Nucleus

There are a variety of characterized pathways that signal the differentiation of a cell from state A to state B. One of the simplest changes in cell state is the transition from G0 to G1 with the addition of serum hormones to cells grown without serum. The hormone receptors at the plasma membrane cause changes in the nucleus. These are typically designated as signaling pathways and they often involve the successive activation of a series of linked kinases that ultimately activate the transcription of specific genes in the nucleus.

Consideration of Solution Binding Equilibria (adapted from notes from Harold Erickson, Duke)
Consideration of the Entropy term and the implication for multiple interactions.

1. $RT \ln K_A = - \Delta G_A$ (Note that favorable association requires a negative free energy of association.)

This free energy describes all of the chemical and energetic factors involved in the association reaction. It is extremely useful to break this term down into two opposing energies. The importance of explicitly separating these two terms has been recognized since the work of Doty and Myers on insulin dimerization in 1953; it has been rediscovered about every ten years since; it was a crucial point in the analysis of Chothia and Janin; and it forms the basis for the analysis of cooperativity (Erickson, 1989). The two terms are the intrinsic bond energy, which includes all the chemical forces acting across the subunit interface, and the intrinsic subunit entropy (expressed as a free energy).

2. $\Delta G_A = \Delta G_{bond} + \Delta G_s$

The term ΔG_{bond} is the intrinsic bond energy. All the chemical forces intrinsic to the protein-protein interface are included in the term ΔG_{bond} ; it is a negative number for all favorable associations; the stronger the bond, the larger its absolute value. The C & J analysis is directed at estimating this bond energy and stating all of the chemical forces that contribute to it. The term ΔG_s is the intrinsic subunit entropy, expressed in units of free energy. It may be considered the free energy required to immobilize a subunit in a dimer or polymer, independent of the type or number of bonds formed. Free energy is required

because entropy (translational and rotational entropy of the protein subunit) is lost when the subunit is immobilized in the polymer. The magnitude of this term is still under discussion. Chothia and Janin, and most previous authors, estimated it to be 103-145 kJ/mol. However, these calculations did not include any compensation for motion or vibration of subunits in the bonded state. In Appendix I of Erickson, 1989, a much lower value is calculated, 29-45 kJ/mol. Surprisingly, the intrinsic entropy depends very little on the size and shape of the protein subunit (these dimensions enter the calculation as a logarithm), so a single value can be used for all protein association reactions. In the present analysis we will use the value DG s = +29 kJ/mol. Let's put this in perspective and illustrate the calculations with some numbers. A typical modest protein-protein association will have a $K_D = 10^{-6}$ M (so $K_A = 10^6\ M^{-1}$). In eq. 1, R has the value 8.31 J/deg.mol, and T is the absolute temperature = 300 degrees. 1a. RT ln K A = –DG A = 2.5 kJ/mol.X ln (10^6) = 34.5 kJ/mol

2a. $\Delta G_A = \Delta G_{bond} + \Delta G_s = -34.5$ kJ/mol; $\Delta G_s = +29$ kJ/mol; thus $\Delta G_A = -34.5 = \Delta G_{bond} + 29$, which gives $\Delta G_{bond} = -63.5$ kJ/mol

Thus the net free energy of association, –34.5 kJ/mol, is very similar in magnitude to the intrinsic entropy term, 29 kJ/mol; and the intrinsic bond energy is the sum of the two, –63.5 kJ/mol. The intrinsic bond energy must be sufficient to compensate for the entropic energy loss, and to produce the favorable association constant. Note that DG bond is a negative free energy, indicating that it favors association. DG s is a positive number, opposing association. The nature of the protein-protein bond, à la Chothia and Janin. The crucial observation of C & J (actually this is based on work of F. Richards several years earlier) was that the interface between subunits looked exactly like the interior of a protein molecule - in particular it consisted of closely packed atoms with no spaces. This means that the two surfaces must fit together very snugly. Water was completely excluded, and there were van der Waals contacts between the surface atoms across the interface. It was also clear that hydrogen bonds and ionic bonds (salt bridges) were made across the interface. C & J then pondered the question of how each of the many interactions would contribute to the free energy of the protein-protein bond. The key to this analysis was to realize that when the protein-protein bond is broken the interface surfaces are not placed in a vacuum, but are exposed to the solvent. Many of the bonds that are formed across the interface when the subunits are together are made to solvent molecules when the subunits are apart.

The conclusion from this diagram, incredible as it may sound at first, is that (a) ionic bonds, (b) hydrogen bonds, and (c) van der Waals forces make little contribution to the net free energy of the protein-protein bond. The reason is, these bonds are not lost when the subunits are separated, but are just replaced with equivalent bonds to the solvent. C & J argue that an ionic bond (or H- or vdW bond) to the solvent is similar in free energy to the bonds across the interface, so there is little change in free energy when one bond is replaced by the other. The nature of the protein-protein bond is enormously simplified if we simply set all these contributions to zero. Saying that the net contribution of ionic bonds is zero does not mean that they are ignored. These bonds play a crucial role in protein association, namely in determining the specificity of the interaction. Protein association is not promiscuous: e.g., pancreatic trypsin inhibitor associates very strongly with trypsin, but

not at all with most proteins. The basis for the high specificity is that the interfaces must be precisely complimentary. This complimentarity comprises three features. (The classic C&J theory is given first; italics introduce corrections from more recent analysis, especially Wells' - see reading list.) a) Ionic complimentarity. All possible salt bridges must be made. If there is a minus charge group on one subunit, there must be a plus charge opposite on the other side. Otherwise the subunit association would require breaking an ionic bond to solvent (8.3-25 kJ/mol or 2-6 kcal/mol) and replacing it with nothing. Burying a charge might be very expensive in free energy and could strongly destabilize the protein-protein interaction. b) Hydrogen bond complimentarity. The argument for H bonds is exactly analogous to that for ionic bonds. Any hydrogen bond donor on one subunit must find a hydrogen bond acceptor on the opposite. Losing a hydrogen bond would cost 4 - 25 kJ/mol, depending on the chemical nature and charge of the buried group (see Fersht et al., 1985 for this important experimental analysis, and also Wells for exceptions where H-bonds can be eliminated without much cost). c) Steric complimentarity. When separated, the subunit interfaces are covered with H_2O, the water molecules making van der Waals contacts with all exposed atoms on the surface. The van der Waals bonds are less than 4 kJ per atom, but there art lots of atoms. In order not to lose the energy of the van der Waals interactions, contacts with solvent must be replaced by contacts with the other subunit when the bond is made. This means that the subunits have to fit together very snugly. Separating a pair of atoms even 1 Å will eliminate the van der Waals energy.

It costs energy to create a hole in the interface (nature abhors a vacuum). Steric complimentarity is perhaps the most important contribution to specificity. Loss of a large, hydrophobic aa leaving a hole in the center of the interface can greatly weaken binding; Mutation of a small aa to a bulky one, especially if it is in a pocket needed for a large aa from the other interface, can completely eliminate binding (Wells, see below).

However, there is one important consistency between the C&J picture and the analysis of Wells' group – the important aa's are all concentrated in the hot spot, and the non-contributing ones are distributed around the edges. The aa's in the hot spot are apparently all contributing to the binding energy with approximately the full 0.025 kcal/mol-$Å^2$. The idea of complimentarity certainly needs to be modified, since it is now obvious that some (but not all) partners for H-bonds and ionic bonds can be eliminated from one side of the interface without seriously disrupting the bond. Most likely the remaining partner undergoes an induced fit to make the bond to solvent or to another aa in the interface (see Altwell...Wells, 1997 for demonstration of this kind of plasticity). Some holes in the interface can probably be tolerated. However, the most obvious candidate for the most important aa was W104, which is right in the center of the interface and buries the most surface area. This indeed turns out to be the most important aa in the binding. Presumably when it is deleted the hole it leaves cannot be filled by induced fit movements, and the missing van der Waals interactions destabilizes the bond by >4.5 kcal/mol.

An important problem that can be addressed with the knowledge of the protein-protein interface. A particularly valuable probe for microtubule studies is tubulin labeled with biotin, a small molecule of ~200 daltons that (a) can be attached covalently to reactive lysines or other groups on the surface of the protein, and (b) can be subsequently labeled with antibodies for localization. What would this biotin molecule do to tubulin assembly if it were attached to an amino acid (a lys) in the middle of the tubulin-tubulin interface.

Tubulin has a MW of 50,000. How could tubulin be labeled to minimize reaction at the interface?

C. Kinetics

Let's approach kinetics from the point of a receptor binding its ligand (growth factor, GF). We assume that the receptors on the cell surface are present in very low concentration relative to the GF. There are two questions.

1. Lifetime of the empty receptor. If the receptor is empty, how long will it take to be occupied? This depends on the concentration of GF, but in many cases is limited simply by diffusion of GF, and does not depend on the bond energy.

2. Lifetime of the complex. Once a receptor has bound a GF, how long will the complex last before the GF dissociates? This is determined largely by the bond energy.

k2 M - 1 s - 1

k s - 1

-1

The lifetime of the empty receptor is the reciprocal of the association rate: $t_E = 1/k_2[GF]$. This is the average time it will take before an empty receptor is occupied (actually it is the time for 1/e of the receptors to be occupied, we will ignore this fine point). The lifetime of the empty receptor depends only on the concentration of GF and k_2.

$t_E = 1/k_2[GF] = 1/(2 \times 10^6)[GF]$

The second order association rate, k_2, is a key parameter. In principal k_2 could vary considerably depending on the protein pairs, and would require experimental measurement for each case. In fact it turns out that a large number of protein-protein associations occur at the same rate, $k_2 = 2 \times 10^6\ M^{-1} s^{-1}$. In the next section we will explain that this is the generic, diffusion limited rate constant for protein-protein interaction. There are some proteins that associate faster and some slower, so experimental measurements are important when they are available. But in the absence of experimental data, this is a very good guess. Assume: $k_2 = 2 \times 10^6\ M^{-1} s^{-1}$ - for diffusion-limited protein-protein association If the growth factor binding to its receptor is diffusion limited, here are some lifetimes to occupy a receptor for difference concentrations of GF. [GF] t_E GrFac 10^{-9} M = 500 s; 10^{-8} M = 50 s; (Actin) 10^{-6} M = 0.5 s

2. The lifetime of the complex is $t_C = 1/k_{-1}$, the reciprocal of the rate of dissociation of GF from the receptor complex. t_C is completely independent of the concentration of free GF but it depends very much on the bond energy (for associations with diffusion limited k_2, K_D is determined completely by k_{-1}). It frequently happens that one does not know the kinetic constant, k_{-1}, but the equilibrium binding constant, K_D, is known. Now we can make a good guess, using the principle above that the association rate is diffusion limited, and equal to $k_2 = 2 \times 10^6\ M^{-1} s^{-1}$, for many protein-protein associations. For these associations, the dissociation rate is determined directly by K_D, and we can estimate k_{-1} directly from K_D. $t_C = 1/k_{-1} = 1/K_D k_2 = 1/K_D (2 \times 10^6)$. Again some examples, starting with the very strong complex of trypsin-trypsin inhibitor. K_D t_C Trp-TrIn 10^{-13} 5×10^6 s (58 days) GF 10^{-9} 500 s (8 min) actin 10^{-6} 0.5 s It is interesting now to consider the association-dissociation as a cyclic event. Consider a GF binding to its receptor. If the GF concentration is 10^{-9}, equal to K_D, the lifetime of the empty receptor is

500 s, and once a complex is formed its lifetime is also 500 s. This is another way of saying that when the concentration of P_1 is equal to K_D, the receptor if 50% occupied. If the concentration of GF is now increased ten-fold to 10^{-8} M, the lifetime of the complex remains unchanged at 500 s, but the empty receptor now binds ligand in only 50 s. The receptor is ~90% occupied.

3. The diffusion-limited rate constant for protein-protein association. Koren and Hammes (1976) surveyed a number of protein associations, and found that a large number of them had $k_2 = 0.5 - 5 \times 10^6 M^{-1} s^{-1}$. To some in the field this rate seemed incredibly fast. Northrup and Erickson (1992) resolved this question by using Brownian dynamics, a computer simulation treating the protein subunits as Brownian particles. It turns out the very slow rate, $5 \times 10^3 M^{-1} s^{-1}$, would be appropriate if proteins behaved as spheres in a vacuum - where they would bounce apart and separate upon every unproductive encounter. In contrast, the protein molecules are in a diffusive environement of water. The Brownian dynamics showed that in water each encounter resulted in multiple collisions. If there is an unproductive encounter the particles didn't bounce apart, they simply diffused a short distance. Because of this "diffusive entrapment" the proteins have a high probability of rotating to a new position and bumping into each other again and again. The conclusion of the analysis was that $k_2 = 2 \times 10^6 M^{-1} s^{-1}$ is the generic, diffusion limited rate constant for protein-protein interaction.

It is important to note that there are some protein associations that occur 10-100 times slower - these have an additional energy barrier to complex formation. Some associations are also 10-100 times faster. At least some of these very fast reactions have been characterized as due to electrostatic steering. Charge groups on the surface of the subunits steer them into correct alignment as they are approaching each other.

Cooperative association

What happens to K_D if you double the intrinsic bond energy? A hypotherical example would be to compare binding a bivalent Ab to binding of the monovalent Fab. Specifically, let's assume we have a HEL dimer, with the two HEL subunits held together so they can each bind the Fab of the IgG. HEL Fab Assume that the monovalent Fab binds with $K_{A\ Fab} = 4.5 \times 10^{-7} M^{-1}$ HEL Fab What is $K_{A\ IgG}$? To make the calculation easy we will introduce two assumptions, neither of which are really valid.

1. Assume that the Ab is rigid. (actually the Fab fragments on an IgG have considerable rotational flexibility).

2. Assume that the crosslinked HEL dimer is also rigid, and that the crosslinking presents the HEL epitopes so that the rigid IgG can bind both of them without strain or distortion.

Essentially the problem asks what happens to K_A if we double the intrinsic bond energy. Is K_A also doubled? Do you square it? Both of these quick guesses have been proposed for certain cases, but they don't work. We need to consider the intrinsic bond energy to do the calculation. Most important, we have to pay close attention to the intrinsic subunit entropy.

1. Calculate $DG_{bond\ Fab} = -RTlnK_A$ –29 kJ/mol = –40 – 29 = –69 kJ/mol.

2. Calculate $DG_{bond\ IgG}$. This is simply 2 x $DG_{bond\ Fab}$ = – 33.4 kJ/mol.

3. Calculate $K_{A\ IgG}$: – RTln $K_{A\ IgG} = DG_{bond\ IgG}$ – 29 = – 33.4 + 29 = – 26.4 kcal/mol

4. $K_{A\ IgG} = \ln(-26{,}400) / RT = 1.3 \times 10^{19}$ M.

This is an enormous enhancement in binding affinity , from 4.5×10^{7} to 1.3×10^{19}, from doubling the intrinsic bond energy. The key to this effect, and to this calculation, is to realize that the entropy tax is paid only once, while the intrinsic bond energy is fully counted twice . Since the IgG is assumed to be rigid, it is fully immobilized when the first Fab binds. The second Fab then binds "for free" and all of its energy goes to increasing K. Of course this is an oversimplification, since it depends critically on the assumptions of (a) rigidity of the Ab, and (b) the perfect fit to the HEL dimer. In a real Ab the second Fab will still have substantial rotational entropy after the first one has bound. This will have to be compensated when the second head binds (but it should be much smaller than 29 kJ/mol). Also an IgG binding to a virus will probably have to strain a bit to make the bivalent attachment. Nevertheless, experimental measurements of Abs binding to viruses show 1,000 X greater affinity for IgG compared to Fab.

PROBLEMS:

1. If we have a transcription factor that is already bound to DNA near a transcription initiation site, then an enhancer can bind to it and activate transcription. From in vitro measurements of binding, the enhancer binds to the transcription factor with a binding constant of 10^{6} M^{-1} in the presence of an enhancer DNA piece only 20 base pairs long. The enhancer binds to the appropriate DNA sequence with an affinity of 10^{4} M^{-1}. Because binding to the DNA is a prerequisite for the enhancer binding to the transcription factor and chromosomal DNA has mobility, the entropy term for the enhancer-transcription factor binding is not the full 29 kJ/mole (assume it is 12 kJ/mole). Estimate the binding constant of chromosomal DNA-bound enhancer for the transcription factor.
2. What is the off-rate of the enhancer-transcription factor binding (assume that the on rate is decreased by 10 fold from a diffusion controlled on rate, because enhancer binds in complex with DNA and DNA diffuses quite slowly)? When the enhancer-transcription factor bond is broken, what is the rate of the dissociation of the enhancer from the DNA (assume here that the on rate is diffusion controlled)?