Principles of protein-protein interactions

Biophysical Chemistry 1, Fall 2010

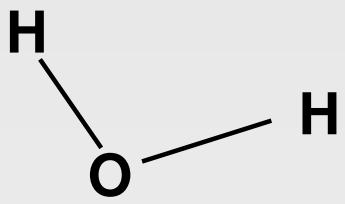
Fundamentals of biological thermodynamics

Basics of protein-protein recognition and interactions

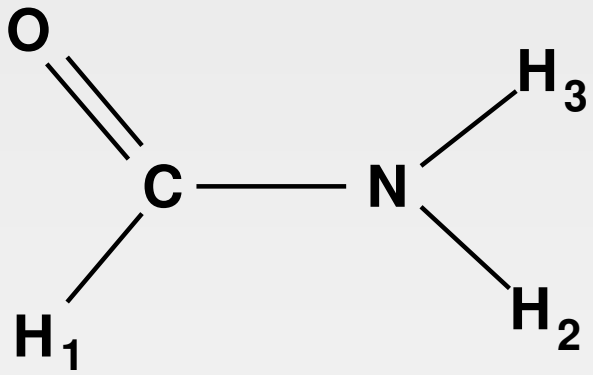
Reading assignment: Slater, Chap. 3 (handout)

What does a force field look like?

$$U = \sum_{\text{bonds}} K_b(b - b_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{impropers} K_w w^2$$
$$+ \sum_{torsions} K_\phi \cos(n\phi) + \sum_{nonbonded\ pairs} \left\{ 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \frac{q_i q_j}{r} \right\}$$



water

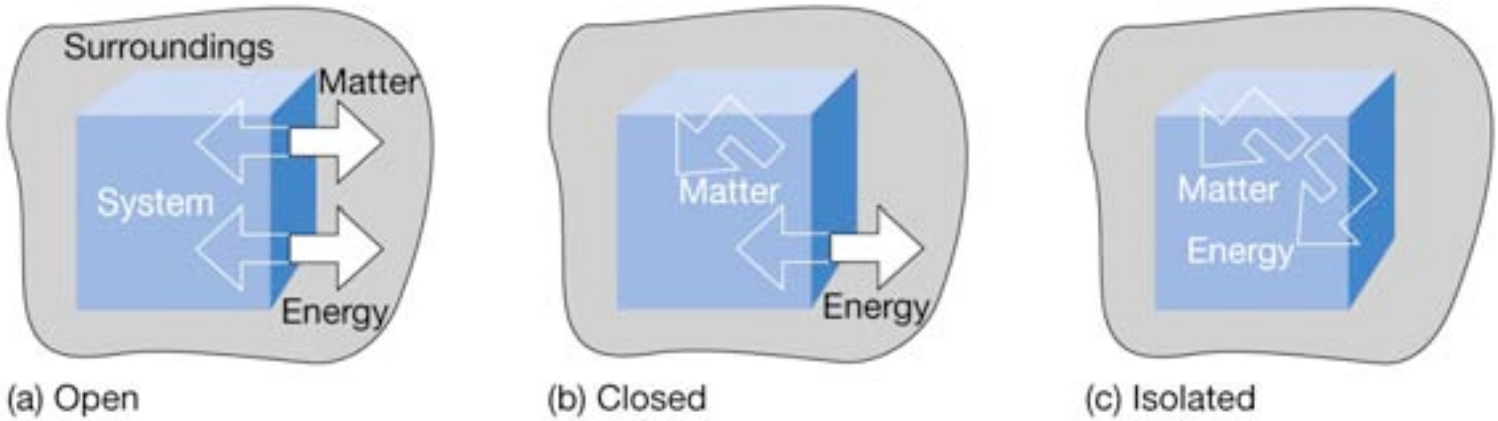


formamide

Lightening intro to biological thermdynamics

A **system** is defined as the matter within a defined region of space (i.e., reactants, products, solvent)

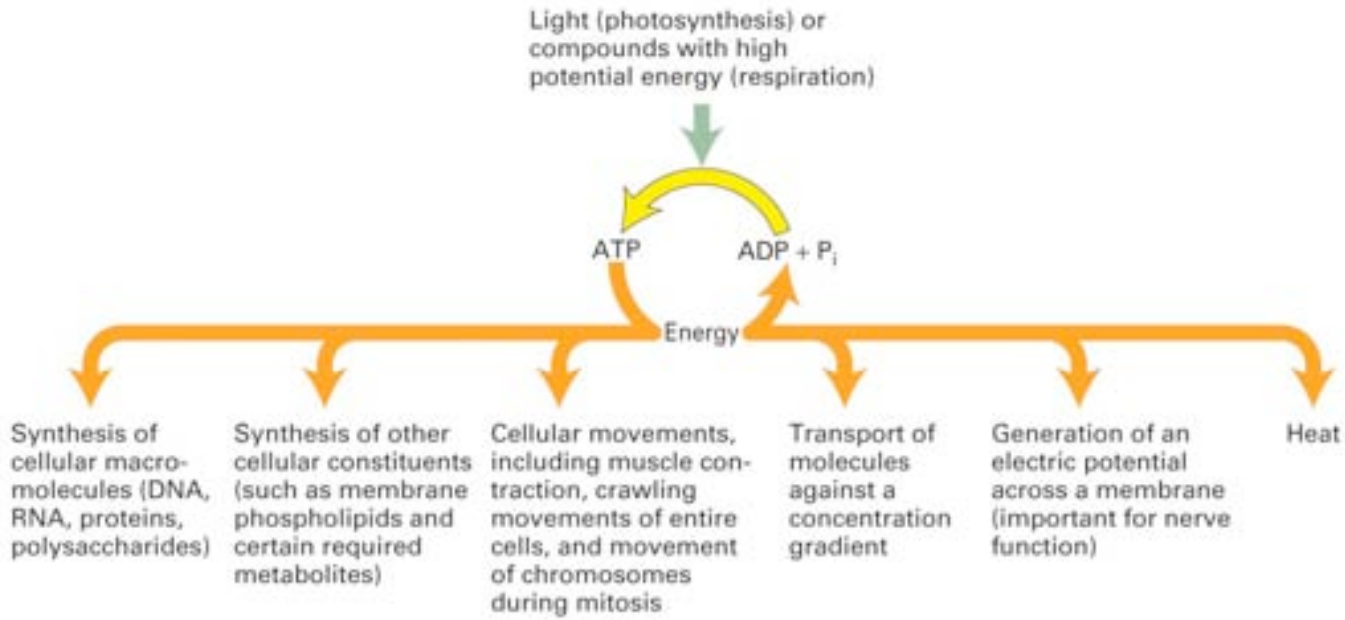
The matter in the rest of the universe is called the **surroundings**



The first law of thermodynamics

The total energy of a system and its surroundings is constant

In any physical or chemical change, the total amount of energy in the universe remains constant, although the form of the energy may change.



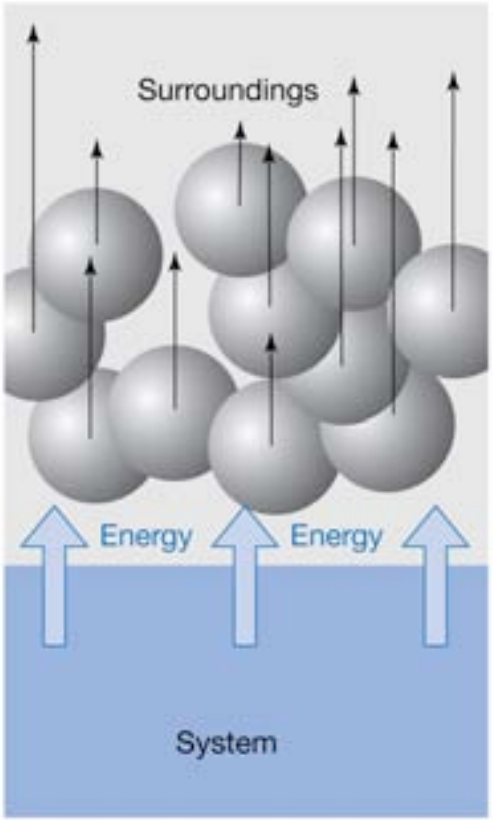
What is “U” (internal energy)?

The **internal energy** of a system is the total kinetic energy due to the motion of molecules (translational, rotational, vibrational) and the total potential energy associated with the vibrational and electric energy of atoms within molecules or crystals.

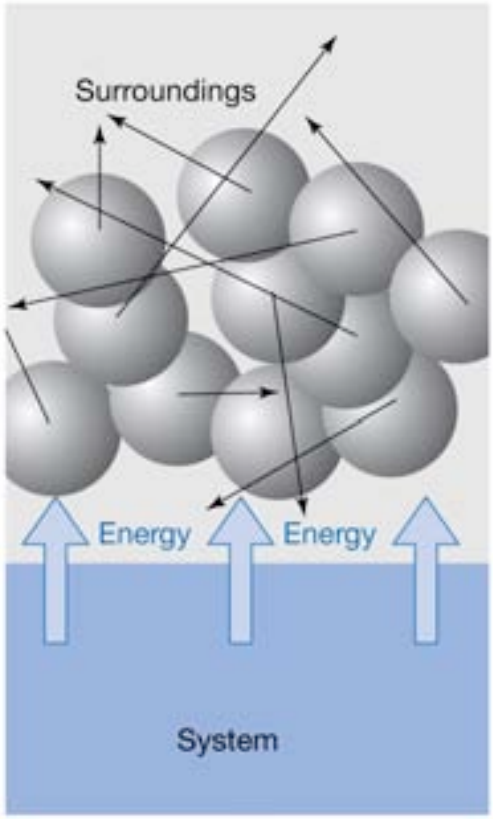
U is a state function, that is, its value depends only on the current **state of the system**

Energy is heat + work

$$\Delta U = W + Q$$



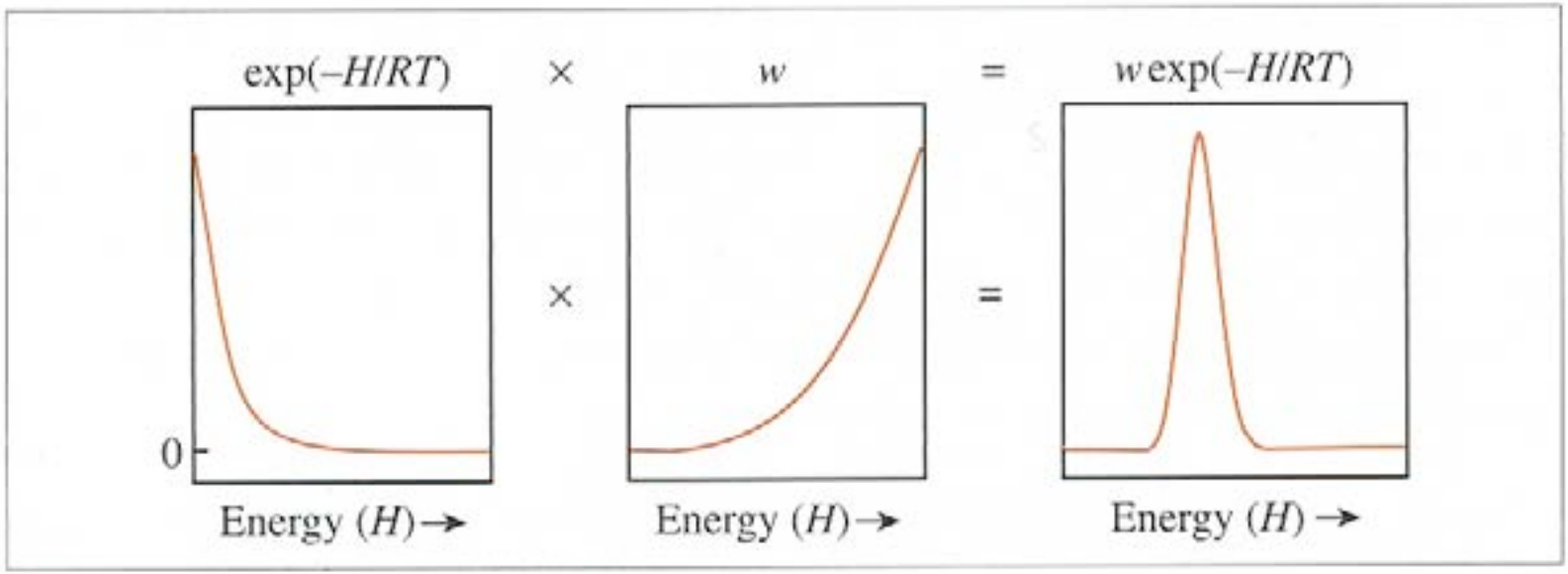
Work involves the non-random movement of particles



Heat involves the random movement of particles

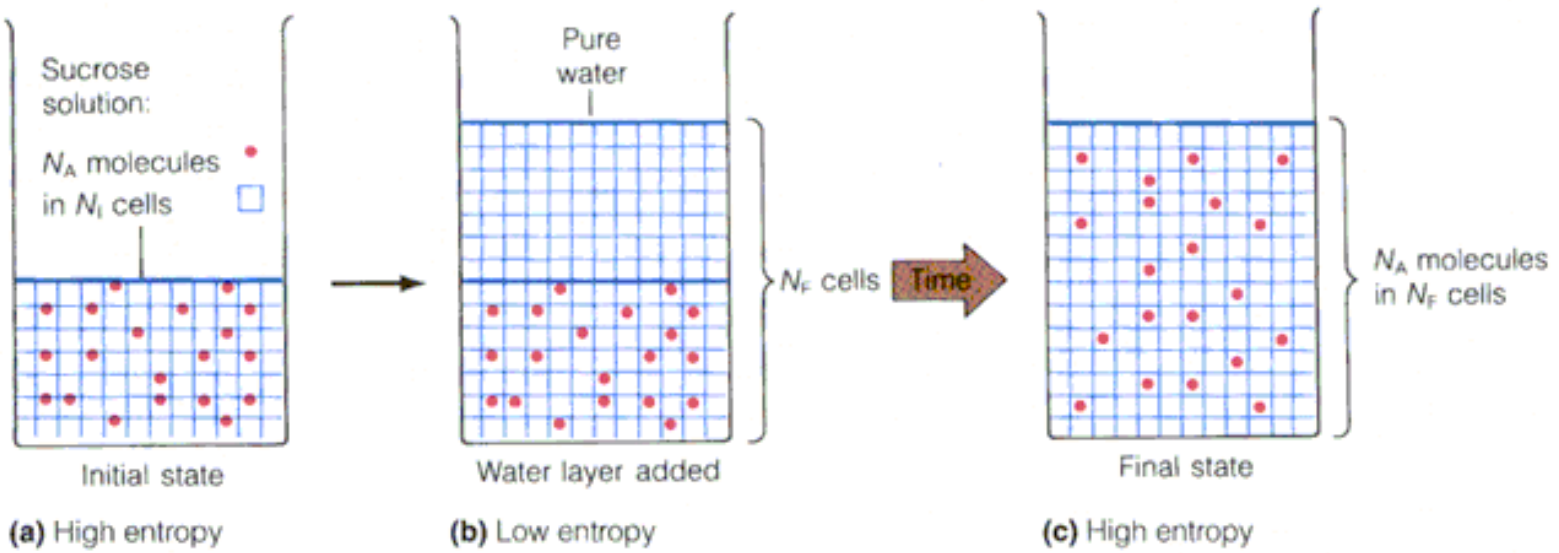
Entropy and probability (qualitative!)

$p(H) = w \exp(-H/RT)$



The Boltzmann entropy

$$S = k \ln N$$



Energy & entropy: the math

(good reading: J.C. Slater, “Introduction to Chemical Physics”; Dover, Chapter III)

- First law of thermodynamics:

$$dU = dQ - dW \text{ or } \Delta U = \int dU = \int dQ - \int dW \tag{1}$$

- Second law of thermodynamics:

$$dS \geq dQ/T \text{ or } TdS \geq dU + dW \tag{2}$$

Connections to microscopic properties

Let p_i be the probability (fraction) of micro-state i. Then we can postulate a connection to the entropy:

$$\boxed{S = -k \sum_i p_i \ln p_i} \tag{3}$$

This is large when the system is "random". For example, if $p_i = 1/W$ (same for all i), then $S = k \ln W$. This entropy is also additive (or "extensive"). Consider two uncorrelated systems that have a total number of states W_1 and W_2. The total number of possibilities for the combined system is $W_1 W_2$. Then:

$$S = k \ln(W_1 W_2) = k \ln W_1 + k \ln W_2 = S_1 + S_2 \tag{4}$$

The canonical ensemble: temperature

Now consider dividing an isolated system (whose total energy U is therefore fixed) into a number of subsystems, each of which could have its own internal energy E_i, but where there is thermal contact between the subsystems, so that energy can be transferred among them. The fixed total energy is

$$U = \sum_i E_i p_i$$

where p_i is the probability that subsystem i will have energy E_i. Let us find the most probable configuration by maximizing the entropy, subject to the constraint of constant total energy and that $\sum p_i = 1$:

$$dS = 0 = -k \sum dp_i (\ln p_i) + k\beta \sum E_i dp_i - ka \sum dp_i \tag{5}$$

Here a and β are undetermined multipliers. The only general solution is when the coefficients of the dp_i terms add to zero:

$$\ln p_i = a - \beta E_i$$

$$\boxed{p_i = \frac{\exp(-\beta E_i)}{\sum \exp(-\beta E_i)}} \tag{6}$$

Connections to clasical thermodynamics

The Lagrange multiplier a is just the denominator of Eq. 6. To figure out what β is, we connect this back to thermodynamics:

$$dS = k\beta \sum_i dp_i E_i = k\beta\, dQ \;\Rightarrow \beta = 1/kT$$

The denominator of Eq. 6 is called the partition function, and all thermodynamic quantities can be determined from it and its derivatives:

$$\boxed{Z \equiv \sum \exp(-\beta E_i)}$$

$$\begin{aligned} A &= U - TS = -kT \ln Z \\ S &= -(\partial A/\partial T)_V = k \ln Z + kT(\partial \ln Z/\partial T)_V \\ U &= -(\partial \ln Z/\partial \beta);\; C_V = T\left(\frac{\partial^2 (kT \ln Z)}{\partial T^2}\right) \end{aligned}$$

Connections to classical mechanics

We have implicitly been considered a discrete set of (quantum) states, E_i, and the dimensionless partition function that sums over all states:

$$Z_Q = \sum_i e^{-\beta E_i}$$

How does this relate to what must be the classical quantity, integrating over all phase space:

$$Z_C = \int e^{-\beta H(p,q)} dpdq$$

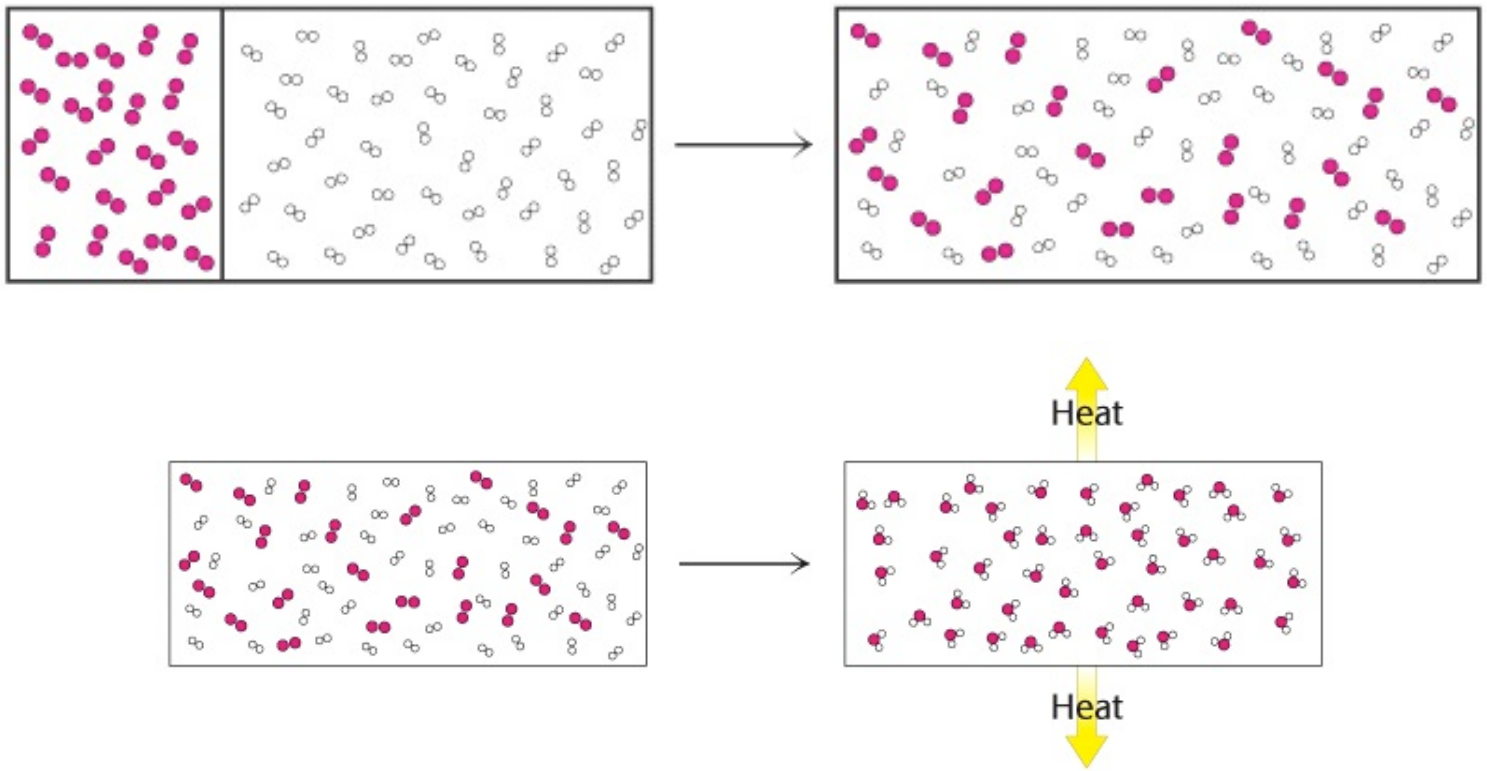
Z_c has units of $(energy \cdot time)^{3N}$ for N atoms. The Heisenberg principle states (roughly): $\Delta p \Delta q \simeq h$, and it turns out that we should "count" classical phase space in units of h:

$$Z_Q \simeq Z_c / h^{3N}$$

For M indistinguishable particles, we also need to divide by $M!$. This leads to a discussion of *Fermi, Bose* and *Boltzmann* statistics....

The second law of thermodynamics

The total entropy of a system and its surroundings always increases for a spontaneous process



The concept of *free energy*

The Gibbs free energy (ΔG)

$$\Delta S_{total} = \Delta S_{system} + \Delta S_{surroundings}$$

$$\Delta S_{surroundings} = -\Delta H_{system}/T$$

$$\Delta S_{total} = \Delta S_{system} - \Delta H_{system}/T$$

$$-T\Delta S_{total} = \Delta H_{system} - T\Delta S_{system}$$

$$\boxed{\Delta G = \Delta H_{system} - T\Delta S_{system}}$$

For a reaction to be spontaneous, the entropy of the universe, ΔS_{total}, must increase

$$\Delta S_{system} > \Delta H_{system}/T \quad \text{or} \quad \Delta G = \Delta H_{system} - T\Delta S_{system} < 0$$

The **free energy** must be **negative** for a reaction to be **spontaneous!**

Free energy is enthalpy minus entropy

$$\Delta G = \Delta H - T\Delta S$$

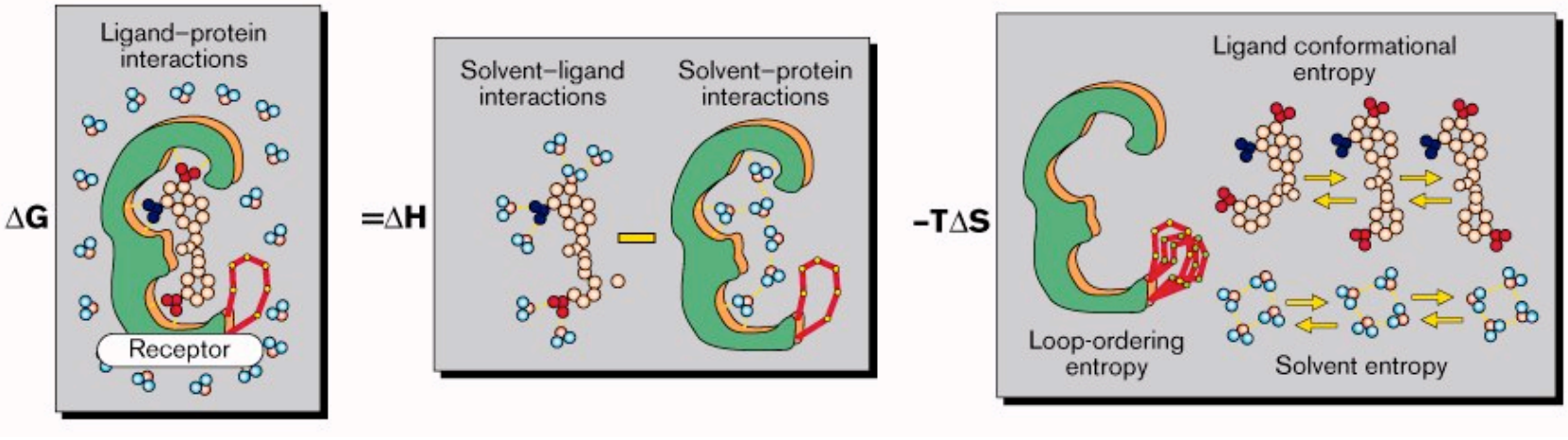
The Enthalpic term

- Changes in bonding
  - van der Waals
  - Hydrogen bonding
  - Charge interactions

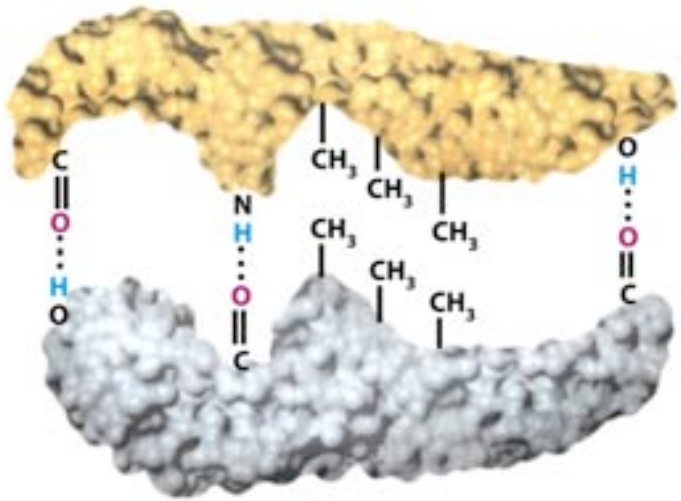
The Entropic term

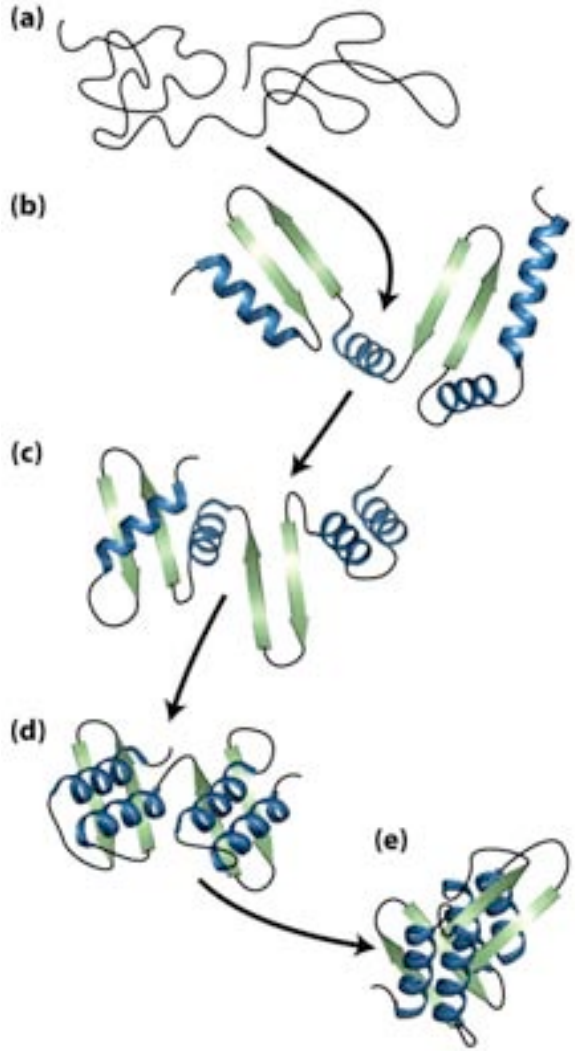
- Changes the arrangement of the solvent or counterions
- Reflects the degrees of freedom
- Rotational & Translational changes

How to think about protein/ligand interactions



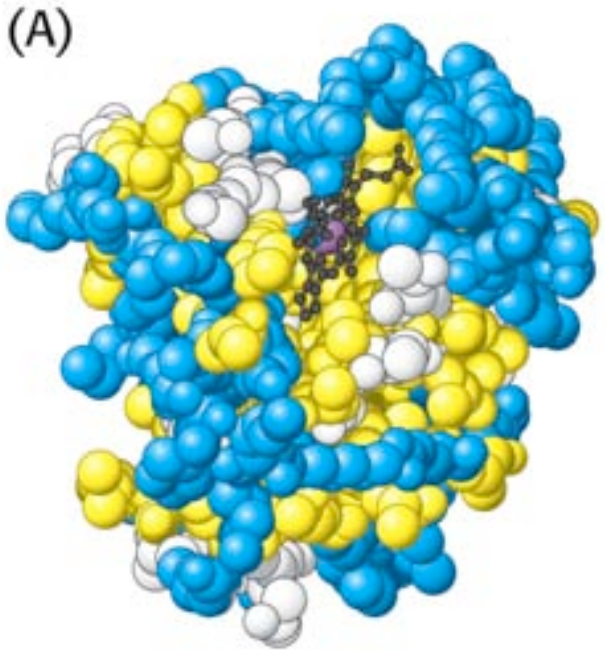
Protein folding vs. protein-protein interactions



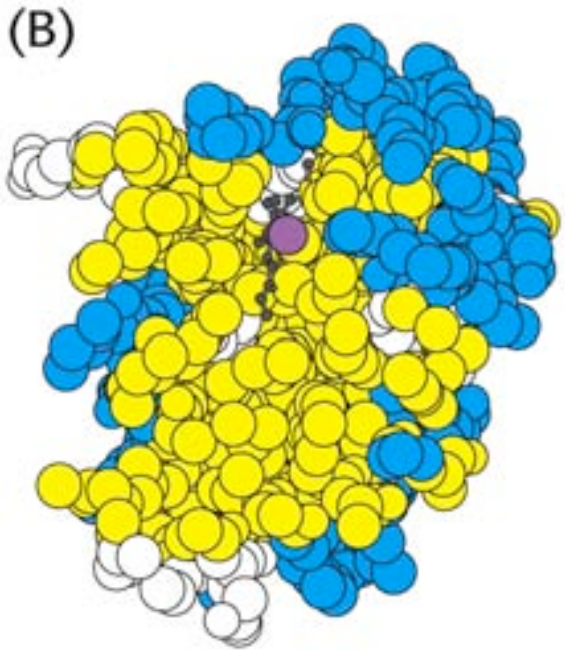


Hydrophilic and hydrophobic residues

Amino acid distribution



space-filling

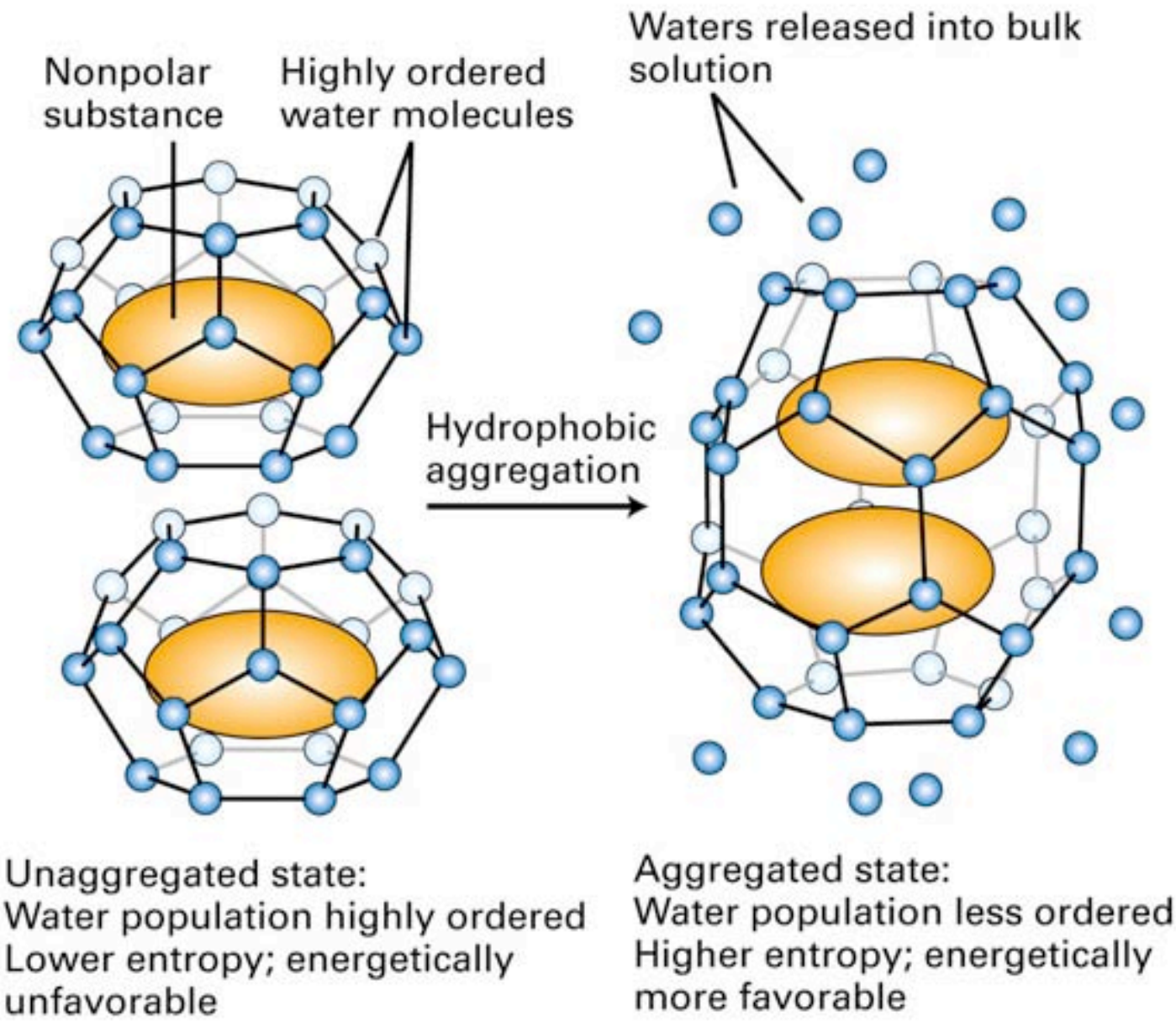


cross-section

nonpolar

polar

The hydrophobic effect



Back to structural biology: protein-protein interactions

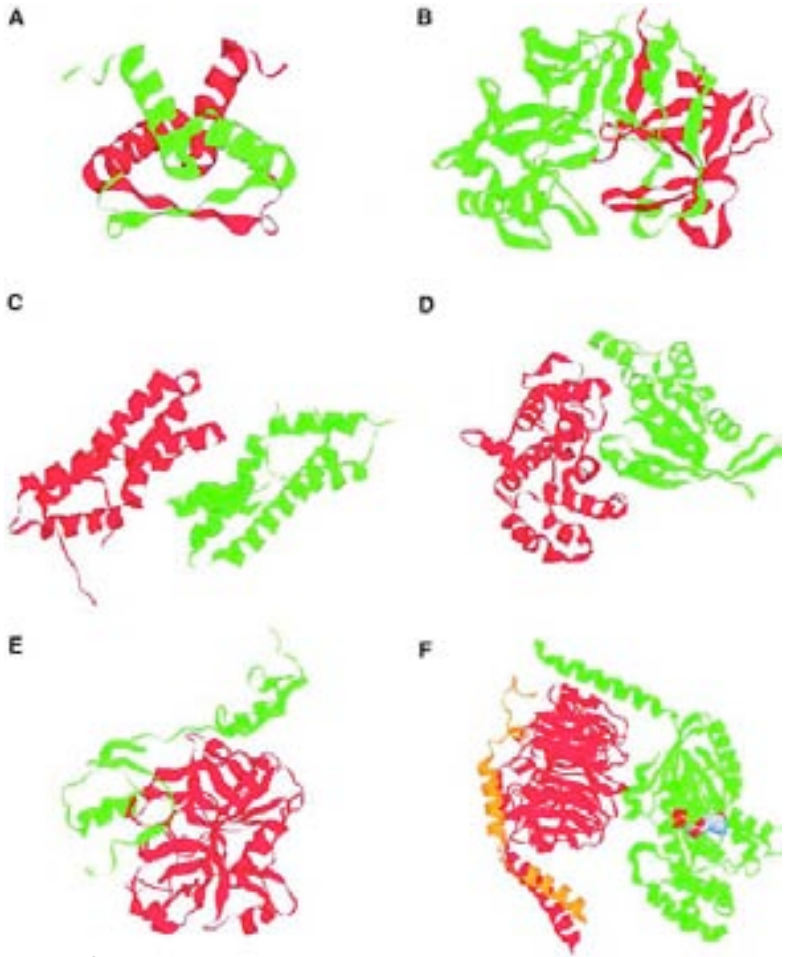


Figure 1. Examples of different types of protein–protein interactions : (A) obligate homodimer, P22 Arc repressor; (B) obligate heterodimer, human cathepsin D that consists of a non-homologous light (red) and heavy (green) chain; (C) non-obligate homodimer, sperm lysin; (D) non-obligate heterodimer, RhoA (green) and RhoGAP (red) signalling complex; (E) non-obligate permanent heterodimer, thrombin (red) and rodniin inhibitor (green); (F) non-obligate transient heterotrimer, bovine G protein, i.e., the interaction between Ga (green) and Gb (red, orange) is transient. The proteins in an obligate interaction are not found as stable structures on their own *in vivo*.

.

Published in: Irene M.A. Nooren; Janet M. Thornton. *EMBO J.* (2003) **22**, 3486-3492.

Multi-protein complexes are very common

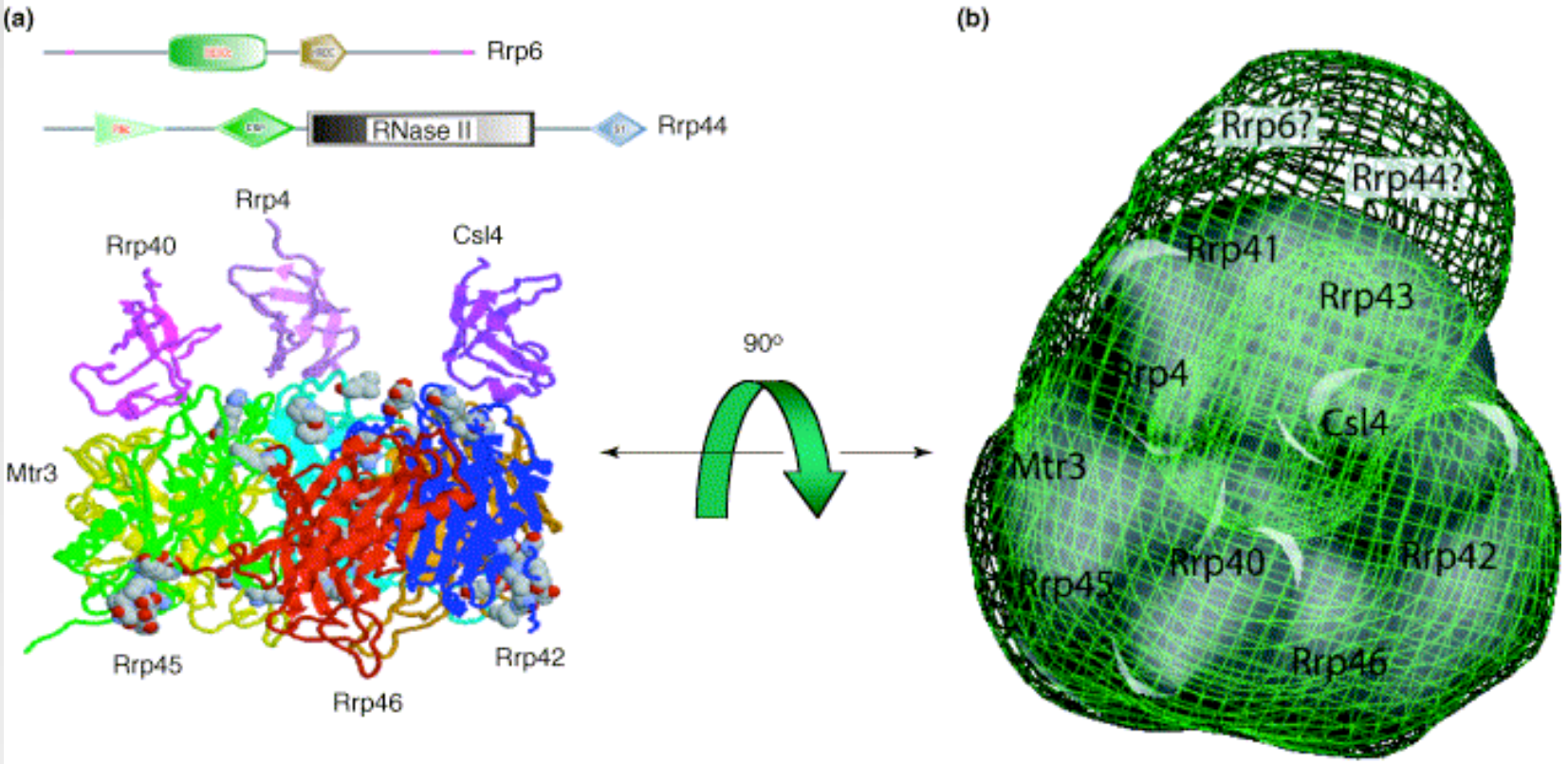


Figure 3. Putative structure through modeling and low-resolution EM. **(a)** Exosome subunits. The top of the panel shows the domain organization of two subunits present in the complex, but lacking any detectable similarity to known three-dimensional structures. The model for the nine other subunits (bottom) was constructed by predicting binary interactions using InterPReTS and building models based on a homologous complex structure using comparative modeling. **(b)** EM density map (green mesh) with the best fit of the model shown as a gray surface and the predicted locations of the subunits labeled. The question marks indicate those subunits for which no structures could be modeled

.

Published in: Robert B. Russell *et al. Curr Opin Struct Biol.* **2004**, 14, 313-324..

Instrisincally disordered proteins

- A large fraction of cellular proteins are thought to be natively disordered, or unstable in solution.
- The structures of disordered proteins are not necessarily random. Rather, the disordered state has a significant residual structure. In the disordered state, a protein exists in an ensemble of conformers.
- In many cases, these regions constitute only certain parts or domains of the whole protein.
- While disordered on their own, their native conformation is stabilized upon binding.
- The global fold of disordered proteins does not necessarily change upon binding to different partners; however, local conformational variability can be observed, thereby complicating predictions of protein interactions.

Proteins are flexible

Proteins are flexible.

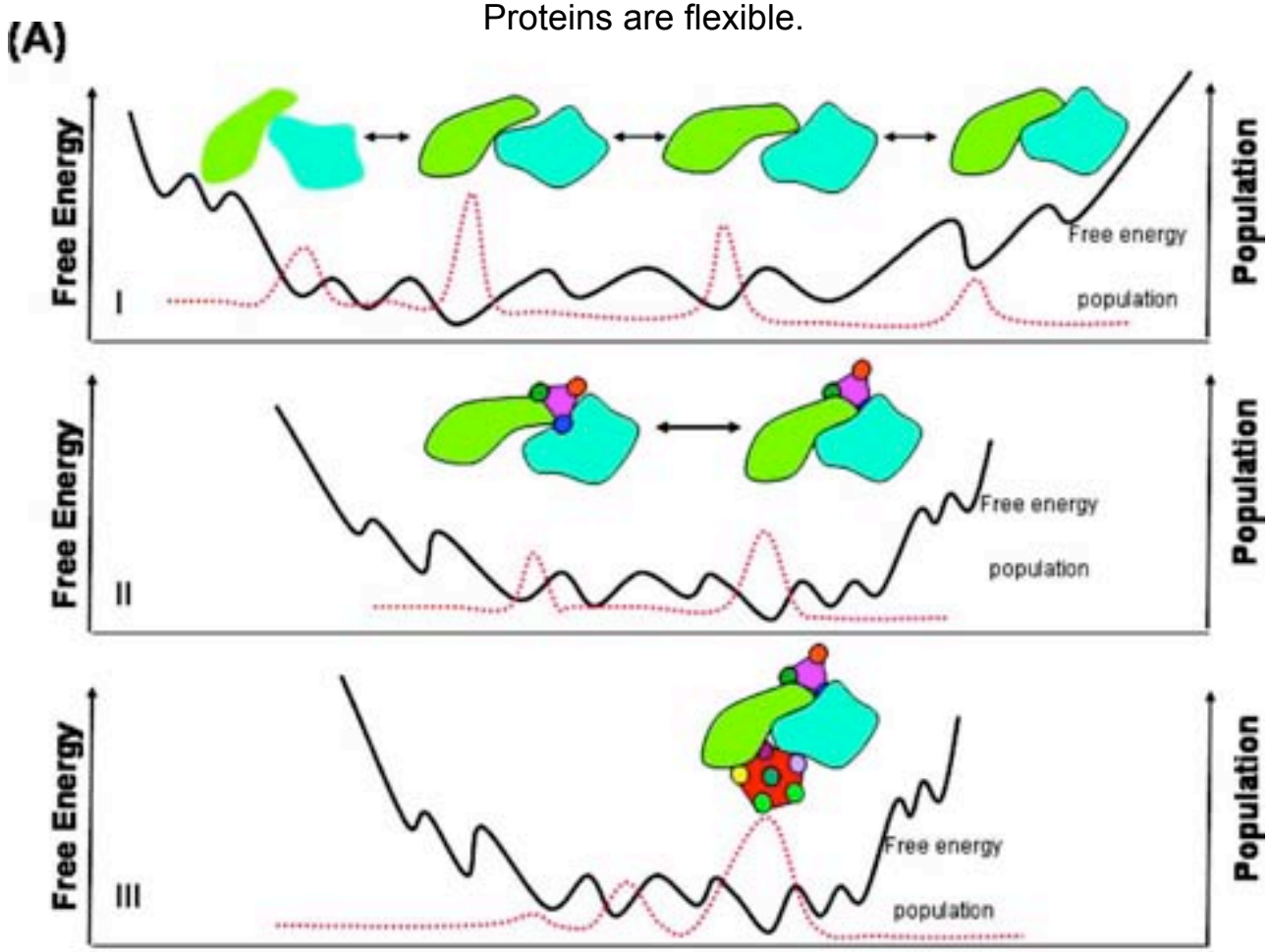


Figure 2 (A) The free energy landscape of a protein may change upon binding to another protein. Binding may induce a shift in the distribution of the populations of the conformational states of the protein; consequently, the relative population of the conformer with an altered binding site shape at another location on the protein surface may increase. The solid black line refers to the free energy landscape, and the dashed red line refers to the relative populations. (I) Distribution of the substates of the protein conformations, presenting several binding possibilities. (II) When a ligand binds at the first binding site, it shifts the conformational energy landscape and the distribution of the populations to favor selective binding at a second, allosteric site. (III) The final dominant conformer recognizes both ligands.

Allosteric structural changes

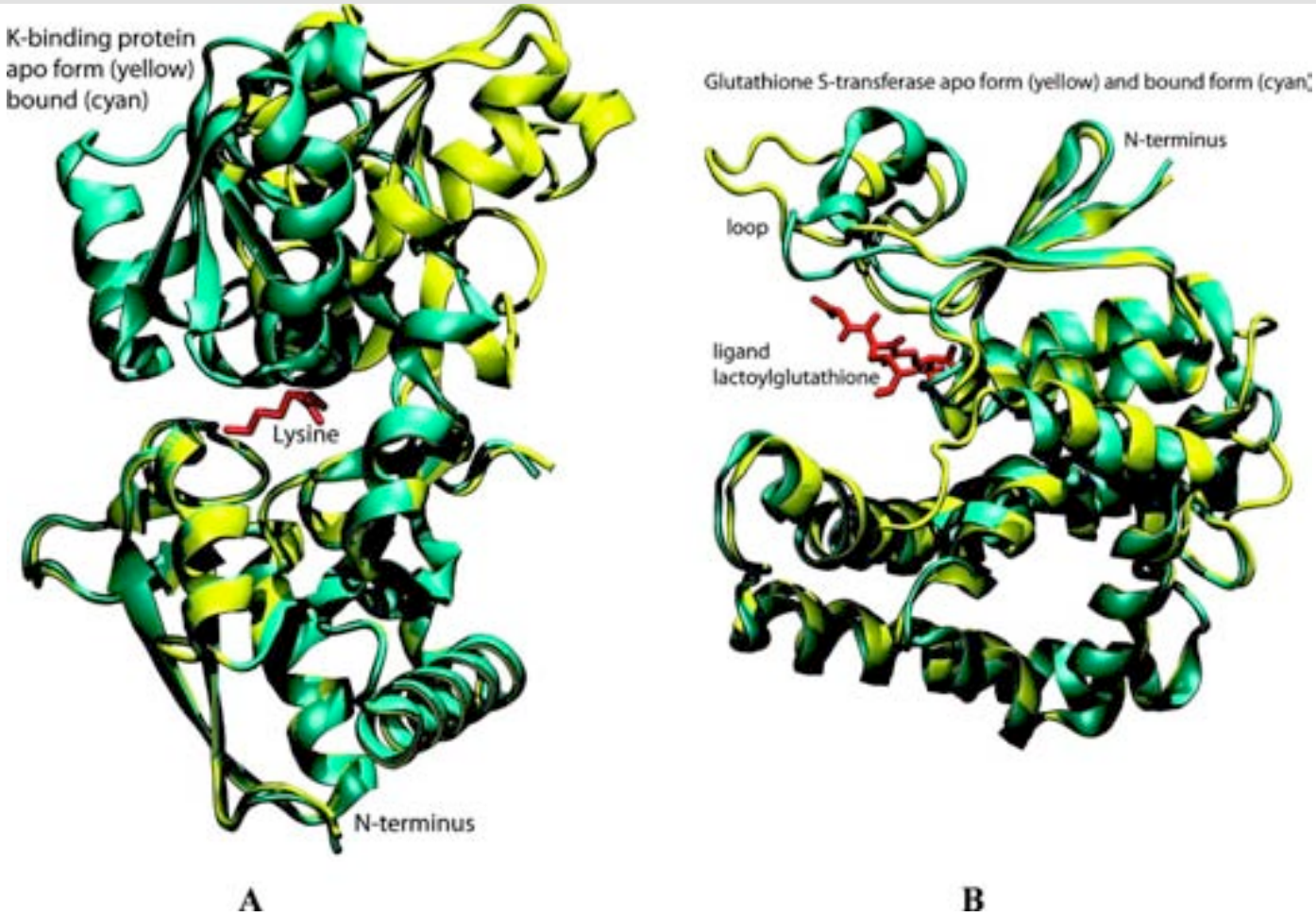


Figure 3 Comparisons of proteins in bound, complexed states versus in the free (apo) states. (A) The conformational changes undertaken by the K-binding protein (PDB IDs: 2lao (yellow) and 1lst (cyan)). The free structure (yellow) closes up and becomes stabilized when it is bound (cyan structure) to its ligand. The ligand, shown in red, belongs to the cyan structure. This is a domain motion example. (B) Glutathione S-transferase-I in free and bound forms (PDB IDs: 1aw9 (shown in cyan) and 1axd (yellow), respectively). The ligand introduces a conformational change in the loop.

The idea of “hot spots”

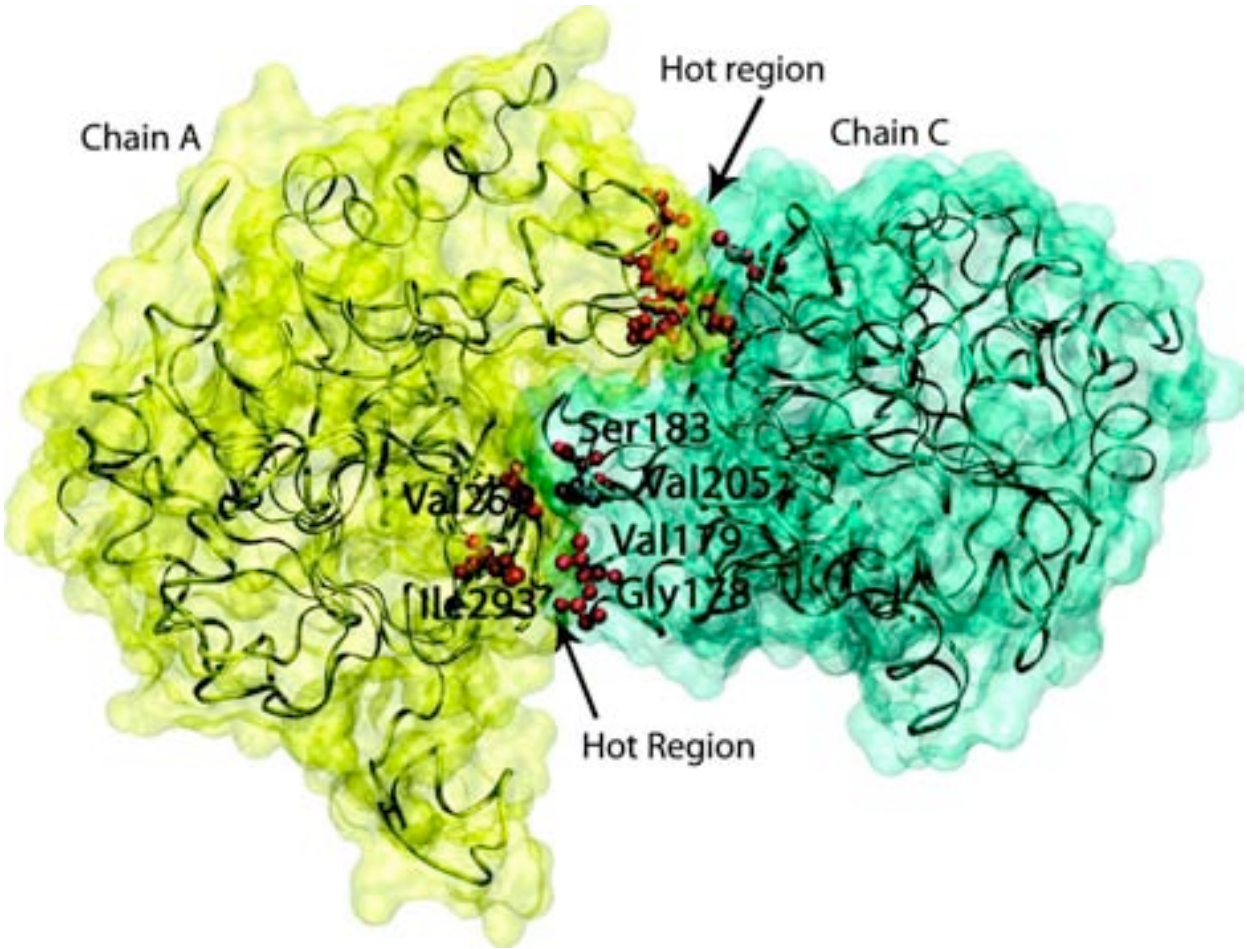


Figure 5 Crystal structure of a complex displaying the hot regions between two M chains of the human muscle L-lactate dehydrogenase (PDB ID: 1i10). Two interacting chains are shown in yellow and cyan. The hot spots (red), shown in ball and stick representation, are residues whose substitution by Ala leads to a significant ($\Delta\Delta G \geq 2$ kcal/mol) drop in the binding free energy (Clackson & Wells; *Science* **1995** 267, 383). There are two hot regions in this interface of the homodimer. The figure illustrates that hot spots are in contact with each other and form a network of interactions forming *hot regions*.

Sequences differ, structures are the same

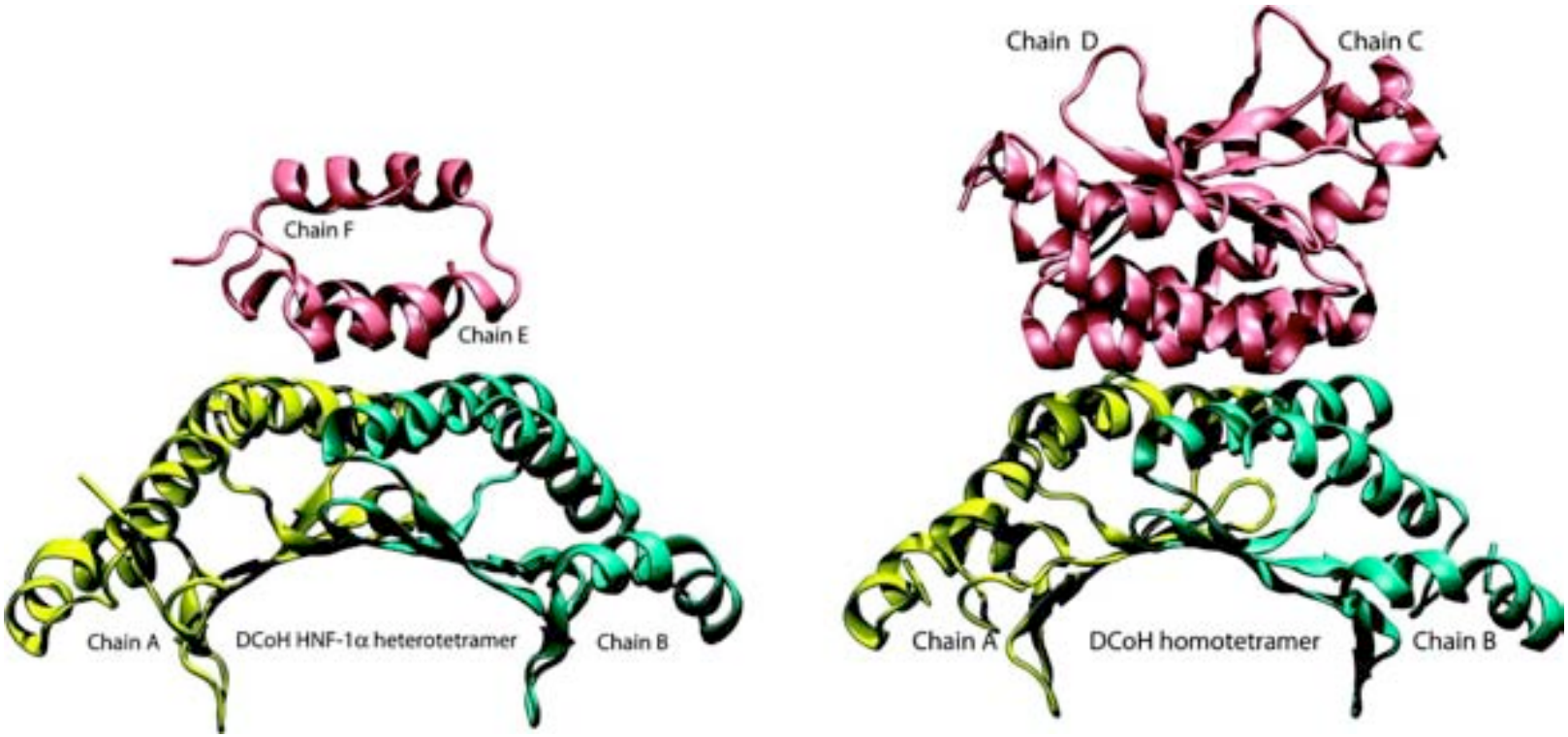


Figure 9 Example of multiple proteins binding at the same site on the protein surface, dimerization cofactor of hepatocyte nuclear factor (DCoH). DCoH serves as an enzyme and a transcription coactivator. The left figure is the crystal structure of hepatocyte nuclear factor dimerization domain, HNF-1, bound to a DCoH dimer (PDB ID: 1F93, Chains A, B of DCoH, and Chains E, F of HNF-1). In order to act as a coactivator, DCoH binds to HNF 1. The figure on the right displays the enzymatic form of the protein DCoH forming dimers of dimers (shown Chains A, B, C, and D, PDB ID: 1DCH).