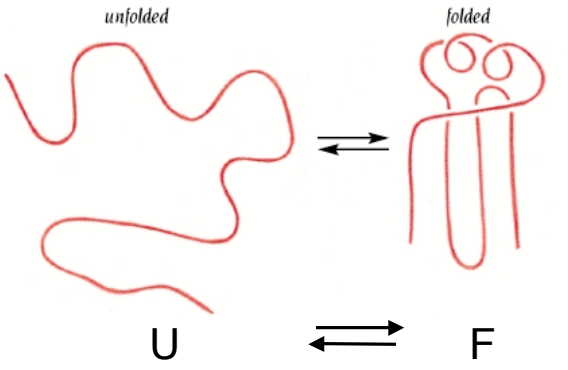
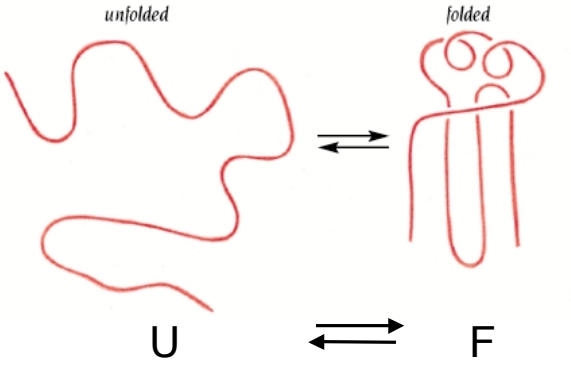
Protein Folding



What stabilizes a given fold?

Okay, so the information needed to determine the correctly folded protein structure is contained in the primary sequence -- but what stabilizes a folded protein?

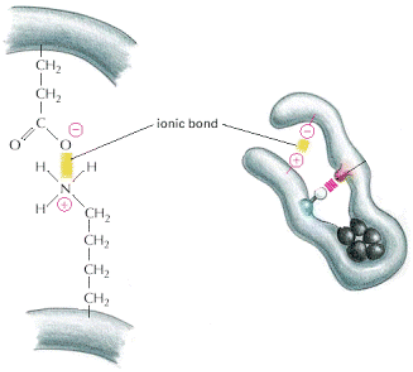
Thermodynamic Forces that Stabilize the Folded State



Ionic, H-Bonds, Van der Waal's Forces, Hydrophobic Effect

Many of you may have learned that there are four fundamental forces in Nature: strong and weak forces, as well as gravitational and electromagnetic forces. The first two are forces that hold nuclei together and they are very very strong but operate only at very short distances (femtometers). Gravitational forces on the other hand, are very very weak forces but they act at very long distances and are only significant for really big things like planets (well actually gravity is also important to medium sized things like baseballs, just not to really really little things like molecules) . The types of non-covalent interactions that help stabilize the folded state of a protein are all manifestations of the same fundamental force -- electromagnetic forces. We're going to review three types of non-covalent interactions we previously talked about, focusing on how they relate to proteins, and then later talk about the hydrophobic effect. The hydrophobic effect isn't a force per se, but a net force.

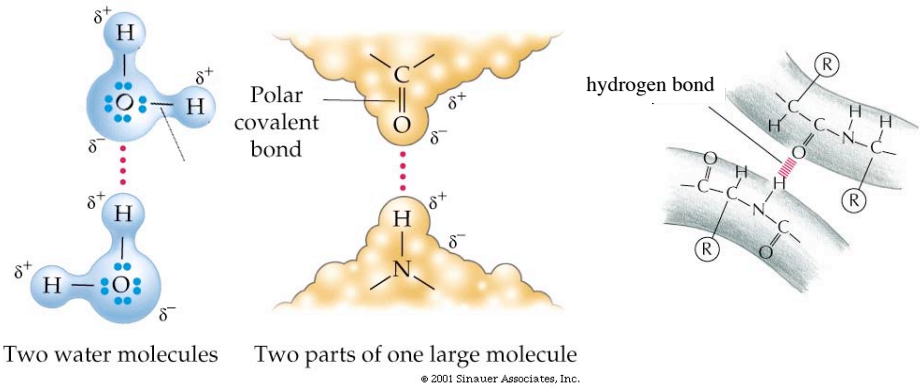
Thermodynamic interactions that stabilize the folded state of a protein



Ionic bonding (salt bridges)

We know that opposite charges attract and like charges repel. We also said that some side chains of amino acids are either fully protonated or fully deprotonated at pH 7 and therefore act as monopoles (full positive or negative charges) in a protein. The attractive force between side chains with opposite charges is inversely dependent on the polarity of the environment around the two charges (For those of you who will take more chemistry, you will learn much more about this force -- described by Coulomb's law) . For example, the attractive interaction between two charges buried in the nonpolar interior of a protein is approximately twenty times stronger than the attractive interaction between two charges in water. The dipoles of the water molecules surrounding each charge align in such a way that the charges are partially attenuated, decreasing the attraction between them. Another way of looking at it: the charges interact so favorably with water that they don't have much left for each other. Because there is no water on the interior of a protein, and because most of the buried side chains are nonpolar (for reasons you will learn), the interactions between side chains of opposite charge are highly favorable. We call ionic bonds in proteins salt bridges.

Thermodynamic interactions that stabilize the folded state of a protein

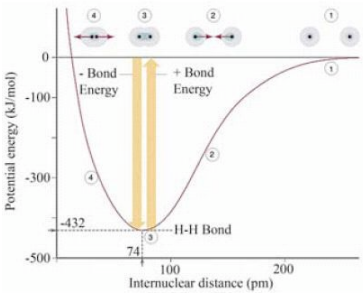


Hydrogen bonds in a protein molecule

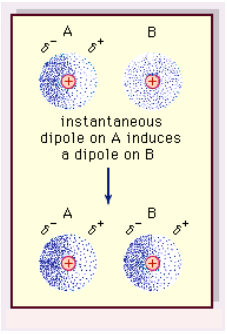
Hydrogen bonds are a particularly strong form of dipole-dipole interaction. Because atoms of different elements differ in their tendencies to hold onto electrons -- that is, because they have different electronegativities -- all bonds between unlike atoms are polarized, with more electron density residing on the more electronegative atom of the bonded pair. Separation of partial charges creates a dipole, which you can think of as a mini-magnet with a positive and a negative end. In any system, dipoles will tend to align so that the positive end of one dipole and the negative end of another dipole are in close proximity. This alignment is favorable. Hydrogen bonds are dipole-dipole interactions that form between heteroatoms in which one heteroatom (e.g. nitrogen) contains a bond to hydrogen and the other(e.g. oxygen) contains an available lone pair of electrons. You can think of the hydrogen in a hydrogen bond as being shared between the two heteroatoms, which is highly favorable. Hydrogen bonds have an ideal X-H-X angle of 180°, and the shorter they are, the stronger they are.

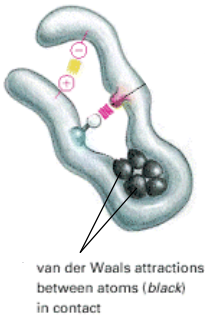
Hydrogen bonds play an important role in the formation of secondary structure. Alpha helices are hydrogen bonded internally along the backbone whereas beta strands are hydrogen bonded to other beta strands. Side chains can also participate in hydrogen bonding interactions. You should be able to list the side chains that can participate in hydrogen bonds now that you know the structures of the side chains. Because hydrogen bonds are directional, meaning the participating dipoles must be aligned properly for a hydrogen bond to form (another way of saying it is that the hydrogen bonding angle must be larger than about 135°, with an optimum of 180°), and because unfavorable alignment of participating dipoles is repulsive, hydrogen bonds between side chains play key roles in determining the unique structures that different proteins form.

Thermodynamic interactions that stabilize the folded state of a protein



Leonard-Jones Potential



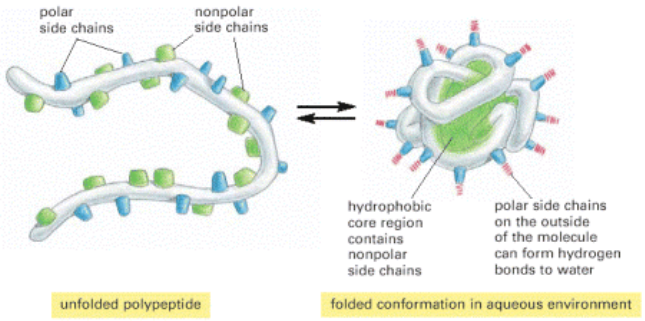


van der Waal's forces

Everyone is familiar with the concept of opposite charges attracting, and it is easy to see how ionic bonds or dipole-dipole interactions can be attractive. What may be less intuitive is that even non-polar molecules form favorable electrostatic interactions with one another. Everyone learns that methane boils at a lower temperature than butane (butane is the liquid that is put in lighters), which boils at a lower temperature than octane. We could rationalize these differences by pointing out that as you increase the mass of something it is generally harder to vaporize (higher boiling). However, this explanation doesn't allow us to rationalize the difference in boiling point between n-pentane and neo pentane, which have exactly the same mass. N-pentane is an extended molecule and has a boiling point of 36 °C whereas neo pentane, which is compact and has much less surface area, boils at 10 °C. N-pentane molecules are capable of packing together in such a manner that each methylene group on one molecule is close to a methylene group on another molecule. As the groups approach one another, the electron clouds become distorted so that there are transient, complementary dipoles induced in opposing surfaces. The favorable electrostatic interactions between these flickering dipoles are weak and act only at very short distances, but they add up so that molecules with lots of surface area (and no steric barriers to bringing surfaces very close) interact strongly. That is why n-pentane is so much harder to vaporize than neopentane.

The curve in the picture above (Leonard-Jones Potential) shows how the potential energy between two atoms varies as a function of the distance between the atoms. The potential energy between two nonpolar groups in a molecule varies in the same way. When the nonpolar groups are at a distance, they don't interact (because they need to be close to distort one another's electron clouds, producing those favorable flickering dipoles). As they approach more closely, the potential energy decreases (which is favorable) until a minimum is reached. Forcing the groups any closer causes a steep rise in energy (because the electron clouds start to overlap, which is unfavorable). We call the induced dipole-induced dipole interactions between nonpolar surfaces London dispersion forces, or van der Waals' interactions.

Thermodynamic interactions that stabilize the folded state of a protein

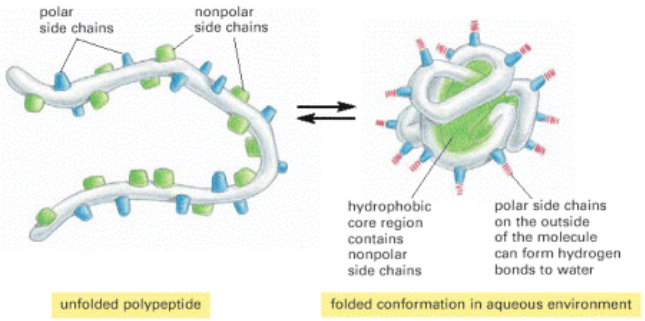


The hydrophobic effect

We have talked about three major types of attractive interactions that stabilize protein structures. There is another very important driving force for protein folding, however. This driving force is called the "hydrophobic effect", which can be described as the tendency for nonpolar surfaces to interact with each other rather than with water -- they segregate. This leads to the burial of nonpolar side chains in the interior of proteins, which in turn leads to a "collapse" of the protein from an extended coil to a more compact, globular structure. This hydrophobic collapse happens *not* because van der Waals' interactions are favorable (they are, but so are dipole-induced dipole interactions between nonpolar surfaces and water), but because it is unfavorable for water molecules to organize around a nonpolar surface. From the standpoint of the nonpolar molecule, it is at least as favorable to interact with water as it is to interact with another molecule like itself. From the standpoint of water, however, the same cannot be said. Dipole-dipole interactions (particularly of the hydrogen-bonding kind) are stronger than dipole-induced dipole interactions. Thus, it is unfavorable from a hydrogen bonding standpoint for the hydrogen bonds between water molecules to be disrupted because a nonpolar molecule is in the way. In order to compensate for this unfavorable situation, the water molecules become highly ordered around the nonpolar surface, forming the best possible hydrogen bonds (short, linear, etc.). The water molecules around a nonpolar surface hydrogen bond like the water molecules in ice. Now the attractive interactions between the water molecules are maximized, but the degree of order of the system has increased dramatically in order for this to occur. As you will learn, disorder is more favorable than order when it comes to energetics.

Continued on next page . . .

Thermodynamic interactions that stabilize the folded state of a protein

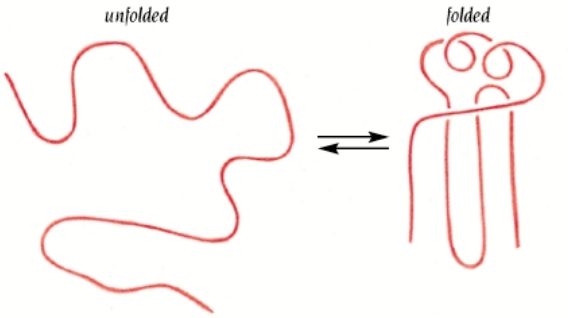


The hydrophobic effect

Burying the nonpolar surfaces in the interior of a protein creates a situation where the water molecules can hydrogen bond with each other without becoming excessively ordered. Thus, the energy of the system goes down.

Therefore, an important factor governing the folding of any protein is the distribution of its polar and nonpolar amino acids. The nonpolar (hydrophobic) side chains in a protein such as those belonging to phenylalanine, leucine, isoleucine, valine, methionine and tryptophan tend to cluster in the interior of the molecule (just as hydrophobic oil droplets coalesce in water to form one large droplet). In contrast, polar side chains such as those belonging to arginine, glutamine, glutamate, lysine, etc. tend to arrange themselves near the outside of the molecule, where they can form hydrogen bonds with water and with other polar molecules. There are some polar amino acids in protein interiors, however, and these are very important in defining the precise shape adopted by the protein because the pairing of opposite poles is even more significant than it is in water.

How much more stable is the folded state than the unfolded state?



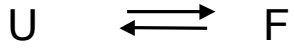
U ⇌ F

NOT MUCH!

You can think of protein folding as a chemical reaction where the equilibrium is defined by the concentrations of the product (which is the folded state) and the reactant (the unfolded state). You might wonder how much product there is compared to reactant. That is, how far to the right does the equilibrium lie, and what does that mean in energetic terms?

You might surmise from Anfinsen's experiment that the folded state of a protein is more stable than the unfolded state (or it would not occur spontaneously), but you might also suspect that it doesn't take much energy to unfold a protein. You don't have to heat a protein very high to denature it. It all depends on the protein, but some proteins denature a few degrees above body temperature (In fact, when you get a fever some of your proteins can actually denature or misfold, and your body makes special stress response proteins called "heat shock" proteins that help these proteins fold correctly so that the fever doesn't do even more damage.)

How do we understand this equilibrium?



$K_{eq} = \frac{[F]}{[U]}$

$\Delta G° = -RT\ln K_{eq}$

G = Gibbs Free Energy
Δ = Change

Recall 3 Cases:

$K_{eq} > 1$ rxn favors folded state

$K_{eq} < 1$ rxn favors unfolded state

$K_{eq} = 1$ equal mixture of U and F

3 Cases:

$\Delta G° < 0$ rxn favors folded state

$\Delta G° > 0$ rxn favors unfolded state

$\Delta G° = 0$ equal mixture of U and F

This slide is to remind you that all reversible chemical processes can be described by an equilibrium constant that is simply the ratio of the products over the reactants. For a protein folding equilibrium, a ratio greater than one favors the folded state, a ratio less than one favors the unfolded state, and a ratio of one means that you have a fifty:fifty mixture of folded and unfolded states.

AT EQUILIBRIUM (and at standard state), there is a relationship between the chemical equilibrium and the change in free energy (Gibbs energy) that occurs as a result of the chemical reaction. That is, the change in free energy that occurs is related to the natural log of the equilibrium constant by a constant, R, and temperature (in Kelvin). Basically, all we are doing by representing equilibrium in terms of delta G $^{\circ}$ is to express the equilibrium constant in units of energy. If the equilibrium constant is 1, the change in free energy is zero. If the equilibrium constant is greater than 1, the change in free energy is less than zero. A change in Gibbs free energy that is less than zero means that a reaction is favorable. That is what you need to remember at this point but in a few lectures David will tell you more about Gibbs freeenergy when you're not at equilibrium.

Components of ΔG

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

H Enthalpy: a measure of heat

$\Delta H < 0$ gives off heat; exothermic

$\Delta H > 0$ absorbs heat; endothermic

$\Delta H = \sum BE_{broken} - \sum BE_{formed}$

S Entropy is a measure of disorder

$\Delta S > 0$ system is more disordered

$\Delta S < 0$ system is more ordered

$\Delta S = S_{final\ state} - S_{initial\ state}$

We have just learned that the change in Gibbs free energy is related to the equilibrium constant. The change in Gibbs energy can also be broken down into two different components. One component is called delta H, the change in enthalpy, which is a measure of the heat given off (or taken up) in a chemical reaction. One way of calculating delta H is by taking the difference between the sum of bond energies of those bonds that are broken ($\sum BE_{broken}$)and the sum of bond energies of those bonds that are formed ($\sum BE_{formed}$) during a reaction. When you make a covalent bond, heat is given off because the potential energy of the system has decreased. Likewise, you need to put in energy (heat) in order to break a covalent bond. When $\sum BE_{broken}$ is smaller than $\sum Be_{formed}$, heat is given off, the overall change in enthalpy is negative, making the reaction exothermic.

The other component of the change in Gibbs energy is called delta S, the change in entropy, which is a measure of disorder. Entropy relates to the number of ways energy can be distributed in a system. The more ways you can distribute energy, the higher the entropy. If you consider an unfolded protein, there are millions of ways that protein can be unfolded and so the entropy of the protein is high (we're not considering the effect of water here). However, there is only one way that the protein can be folded properly. Therefore, when a protein folds, the entropy of that protein, considered apart from the water, goes way down. Because entropy is a state function (only depends on initial and final states), the change in entropy (delta S) of a given reaction would be $S_{final\ state} - S_{inital\ state}$. Later when you take more advanced physical chemistry, you will learn about how S values are calculated using statistical thermodynamics.

(Notes for this slide continues on the next page.)

Components of ΔG

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

H Enthalpy: a measure of heat

$\Delta H < 0$ gives off heat; exothermic

$\Delta H > 0$ absorbs heat; endothermic

$\Delta H = \sum BE_{broken} - \sum BE_{formed}$

S Entropy is a measure of disorder

$\Delta S > 0$ system is more disordered

$\Delta S < 0$ system is more ordered

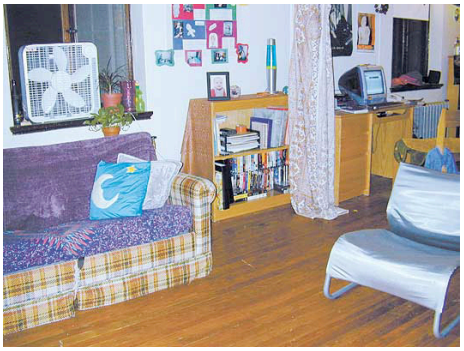
$\Delta S = S_{final\ state} - S_{initial\ state}$

Recall that for a reaction to proceed (that is, to favor the products), the change in Gibbs energy must be negative. Therefore, either the change in enthalpy must be negative (and dominant over an unfavorable entropy term) or the change in entropy must be positive (and dominant over an unfavorable enthalpy term), or the change in enthalpy must be negative and the change in entropy must be positive.

2nd Law of Thermodynamics

$$\Delta S_{Total} > 0$$

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

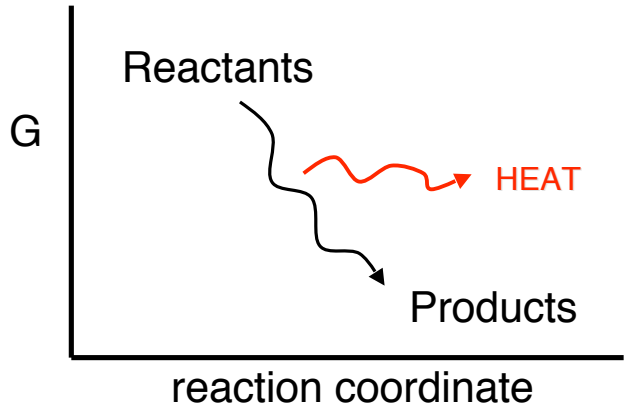


The total entropy of the universe is always increasing.

The first law of thermodynamics, the law of conservation of energy, says that energy can't be created or destroyed, just interconverted between forms. The second law of thermodynamics says that the entropy of the universe is always increasing. (The laws of thermodynamics are so-called because they have never been falsified, and we believe them to be fundamental to our understanding of the universe.) Entropy is a measure of disorder and the universe tends towards disorder because disorder is more probable.

It is possible to observe chemical reactions that are associated with a decrease in entropy. If that occurs, however, there must be an increase in entropy elsewhere in the universe. In the case of protein folding as you will see, the system refers to the protein itself, and the surroundings refers to the water around it. We are going to talk more about enthalpy and entropy on the next slides.

Some chemical reactions "go" because ΔH is negative (*exothermic*)



But . . . not all reactions that go spontaneously are exothermic *(give off heat)*.

Some chemical reactions proceed to products because the change in enthalpy is negative, meaning they give off heat. These reactions are called exothermic reactions. Not all reactions that proceed spontaneously are exothermic, however. Remember that reactions "go" spontaneously if delta G is negative and delta H (enthalpy) is only one of the components that can influence whether delta G is negative. Some chemical reactions occur spontaneously, but instead of giving off heat, these reactions take up heat (delta H is positive). Reactions that proceed spontaneously but require heat, which they absorb from the surroundings, are called endothermic reactions.

Ice melting is an *endothermic* process

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

$\Delta H > 0$ $\Delta S > 0$

H_2O (solid) $\xrightarrow[\Delta G < 0]{\text{Above } 0^\circ \text{ C}}$ H_2O (liquid)

Ice melting is a spontaneous process at temperatures above freezing. That means that the Gibbs free energy change for ice melting must be negative above zero degrees. We know that ice melting is an endothermic process, which involves an uptake of heat. How do we know this? If you hold an ice cube in your hand, your hand becomes cold and the ice starts to melt. Heat is being withdrawn from your hand to break the hydrogen bonds that hold ice together and so your hand becomes cold. There are hydrogen bonds in liquid water also, but they are not as organized or as strong as the hydrogen bonds in ice and so the enthalpy change is positive.

If ice melting is endothermic and the change in enthalpy is positive, then the entropy change must be positive as well (because the Gibbs energy change is negative above zero degrees). Ice is a highly ordered structure -- a crystal -- in which all the hydrogen bonds are satisfied simultaneously and each water molecule is located at a defined position within the crystal lattice. In liquid water, the water molecules are moving around, and you can't infer the relative position of one molecule if you know the position of another one, as you can with ice. The molecules in water have many more ways of existing than the water molecules in ice, and so the entropy of water is much higher. Therefore, **$\Delta S > 0$ when ice melts into water.**

Heat is given off when bonds are formed
(*i.e. water freezing is exothermic*)

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

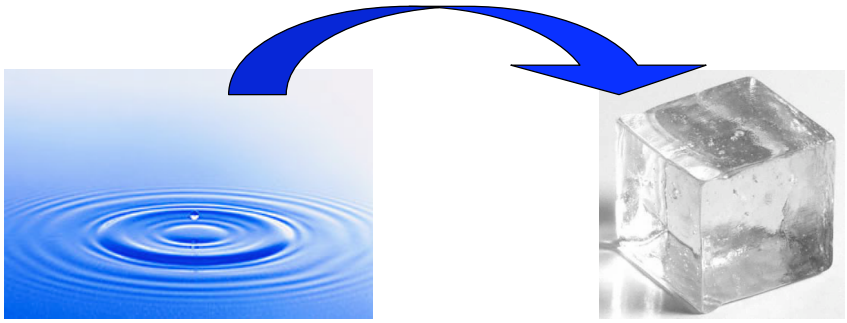
$$\Delta H < 0 \quad \Delta S < 0$$

H_2O (liquid) $\xrightarrow[\Delta G < 0]{\text{Below } 0^\circ \text{ C}}$ H_2O (solid)

We just learned that ice melting above zero degrees is a spontaneous process -- i.e., the change in Gibbs energy is negative. Well, water freezing below zero degrees is also a spontaneous process -- meaning the change in Gibbs energy for freezing is also negative. The components of the change in Gibbs free energy are different, however. As water freezes, the water molecules become more ordered and so the entropy change is negative. The enthalpy change is also negative because short, strong hydrogen bonds form. Thus, heat is given off when water freezes. Thus, both ice melting and water freezing are spontaneous processes at the appropriate temperatures. In one case, the process is driven by entropy and in the other case it is driven by enthalpy.

Interactions between biological molecules that occur spontaneously can also be driven either by entropy or enthalpy.

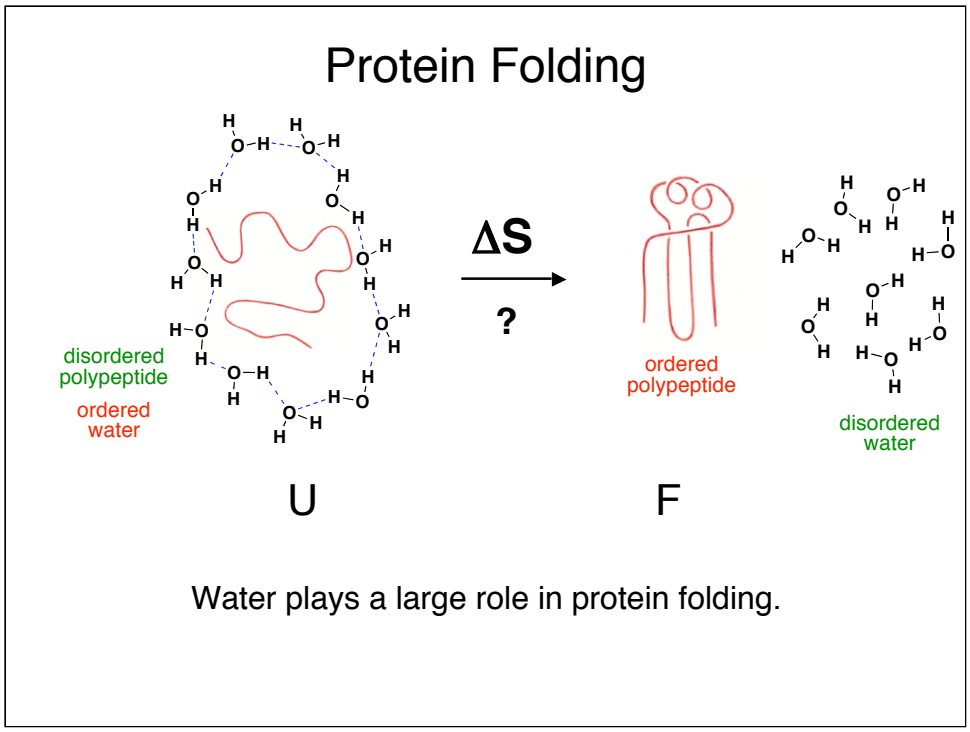
Does this violate the 2nd law of thermodynamics?



No: the heat released upon freezing creates more disorder in the surroundings

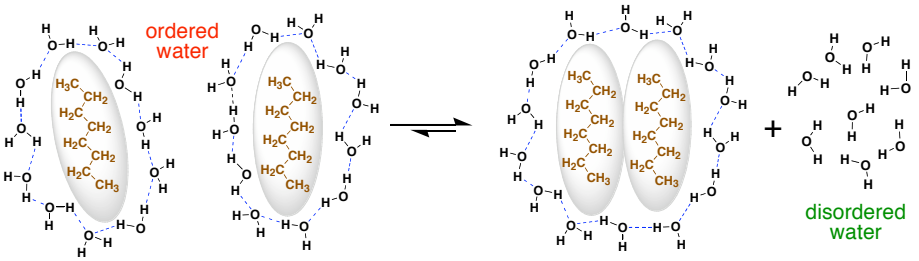
$\Delta H < 0$ $\Delta S_{system} < 0$, but $\Delta S_{total} > 0$

You might wonder whether water freezing violates the second law of thermodynamics, which states that entropy is always increasing. The answer is that it does not. While the water molecules in the ice (system) are becoming more ordered, the molecules in the surroundings (for example, air) are becoming less ordered. The heat given off in the exothermic process of freezing causes the molecules in the surroundings to dance around more.



Let's go back to protein folding now that we know about Gibbs free energy, enthalpy and entropy. In order for folding to occur spontaneously, the change in free energy of the process must be negative. If you consider the protein in isolation, you might imagine that protein folding is an exothermic process in which folding is driven by a favorable, negative change in enthalpy. You might imagine this because you know that the entropy of the polypeptide chain decreases dramatically when you go from the millions of shapes that the unfolded protein can adopt to the one properly folded shape. Therefore, the entropy change from the standpoint of the protein alone is negative ($\Delta S_{system} < 0$). You also know that the protein can form enthalpically favorable hydrogen bonds, electrostatic interactions, and van der Waals interactions, and so it might not seem unreasonable to conclude that protein folding is driven by enthalpy. However, as we have mentioned, the hydrophobic effect plays a major role in protein folding in water. We are now in a position to understand the hydrophobic effect a little bit better. When you put a nonpolar molecule in water, the water molecules around that nonpolar surface become much more ordered (organized) than in the bulk liquid. These water molecules, therefore, become "ice-like" -- highly organized -- in an effort to maximize their hydrogen bonds (and minimize the unfavorable effects of disrupting the hydrogen bonds to the water molecules in order to interact with the nonpolar surface). The "ice-like" structure of the water molecules is not simply one layer thick because the order in the first layer is propagated to surrounding layers. That means that everywhere there is a nonpolar side chain, there is a mini iceberg surrounding it. We have already learned that the entropy of ice is lower than the entropy of water. The entropy of water surrounding a nonpolar surface is lower than the entropy of "bulk water". If the protein can bring the hydrophobic side chains together, then the water molecules surrounding the nonpolar side chains are released into bulk water, which, like ice melting, results in increased disorder ($\Delta S_{surroundings} >> 0$). The change in Gibbs energy reflects a balance between competing effects, with the increased entropy of water due to the "hydrophobic effect" playing a very significant role.

The Hydrophobic Effect



The release of disordered water upon aggregation causes an increase in the entropy of the universe ($\Delta S_T > 0$).

We can redefine the hydrophobic effect as the tendency of nonpolar side chains to become buried because that leads to increased entropy of water. This effect is a major driving force for protein folding. It results, as we mentioned earlier, in what we call the "hydrophobic collapse" of a polypeptide chain. Directional interactions such as hydrogen bonds and the desire to avoid unfavorable steric or electrostatic interactions then leads to a precise shape being adopted. In the process, almost all water molecules are excluded from the interior of a protein.

Kinetics - what path does the unfolded chain take to get to the folded state?

Levinthal's Paradox:

- each amino acid has 9 conformations (Ramachandran plot)

- Conversion between conformations: 10^{-12}sec

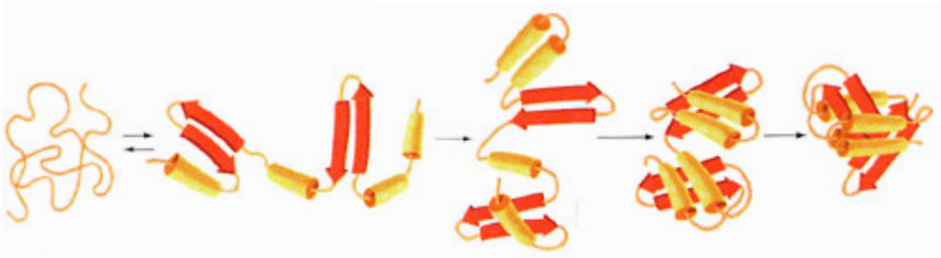
- 150 aa chain

~ 10^{150} conformations
To search all 10^{150} conformations requires 10^{138} sec ~ 10^{130} years!!

But - protein folds in ~0.1-~1000 sec. HOW?

Cyrus Levinthal was a physicist at MIT who became interested in protein folding (and became a biologist). He pointed out that proteins cannot fold by sampling all possible conformations that can be adopted by the peptide chain along their way to finding the lowest energy conformation. In fact, they can't sample even a fraction of those conformations and fold in a reasonable amount of time. He noted that each amino acid can be assumed to have nine backbone conformations (three conformations each around the phi and psi angles and one around the amide). These conformations interconvert on a picosecond timescale, which is pretty fast. If you assume that there are 150 amino acids in a given protein, you can calculate that it would take more years for a single protein to fold than the universe is old (and the universe is estimated to be about 15,000,000 years old, which is pretty old). Proteins fold in milliseconds to minutes, however. That means, as Levinthal pointed out, that there must be protein folding "pathways."

Speculation: Folding process is directed to avoid irrelevant/incorrect conformations

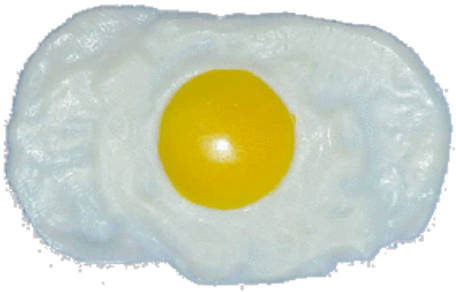


One idea : Domains guide folding

Many people now study protein folding pathways. One reason to study how proteins fold is that we would like to be able to predict protein structure from primary sequence. If we could do that, we could predict the structures of all the gene sequences in various genome databases. We don't know either the structures or functions of the proteins that many of these genes encode. Since structure often provides clues to function, if we could predict the structures, we might have some idea what the proteins do. We would at least be able to make educated guesses that could be tested.

We already mentioned the concept of "hydrophobic collapse", that polypeptide chains undergo a rapid collapse in water in an effort to bury hydrophobic side chains. It is believed that this process is not completely chaotic, but that there are small units of secondary structure that form rapidly and then collapse. These elements of secondary structure then reorganize to yield a properly folded protein. It is possible to predict for many stretches of amino acids whether they are likely to form helices or sheets based on the distribution, or periodicity, of polar and nonpolar residues in a particular stretch and also based on which amino acids are present. Some amino acids are known to be found more often in helical regions and some are found more often in sheets.

Humpty Dumpty's Fate:



Denaturation is often irreversible.

We learned that Anfinsen was able to refold urea-denatured ribonuclease A by removing the urea. When you fry an egg, you denature the proteins in the egg. If you put the fried egg in water, it doesn't go back to the way it was before. Does that make Anfinsen wrong? Why can't we refold Humpty Dumpty if Anfinsen is right?

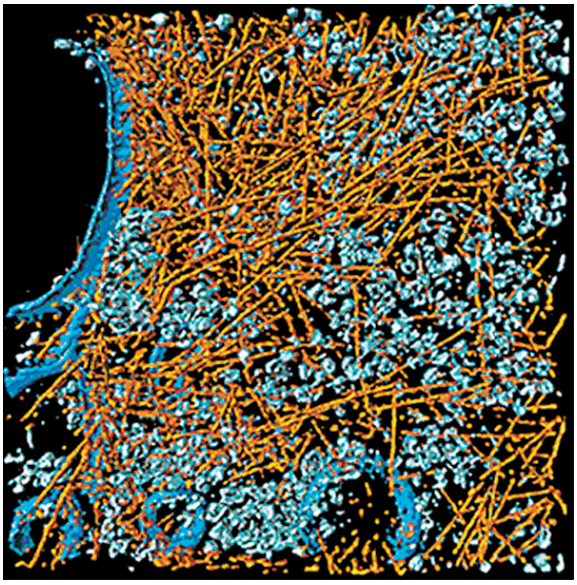
Well, it is worth pointing out that Anfinsen worked with ribonuclease under conditions of high dilution. There wasn't much protein there (but as we pointed out, that didn't matter because he used enzymatic activity, which is extremely sensitive, as a measure of folding).

Proteins in cells are densely packed.

Concentration of total protein inside cells = 200-300g/L

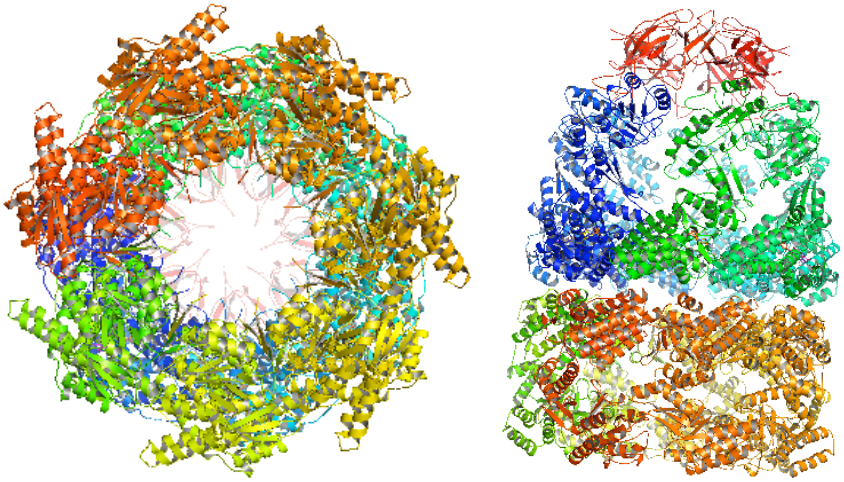
Cellular interiors are 20-30% volume occupied by macromolecules

(Red blood cells contain 350g/L of hemoglobin alone)



The conditions inside a cell are not at all like the conditions in Anfinsen's test tube. Inside a cell the total concentration of protein is about 200-300 g/L. That's like the equivalent of half a pound of salt dissolved in your bottle of seltzer water, except that salt is extremely water soluble and proteins aren't. They can stick to themselves and to each other and to other biomolecules. One third of the volume inside a cell is filled with macromolecules. You might imagine that it would be hard for a protein to fold correctly in such an environment, with so little room and all those other things to stick to. And so it is. Bacterial cells have certain proteins which actually help other proteins fold correctly.

Anfinsen's Cage



Chaperones provide a secluded environment for proteins to fold in.

Proteins that help other proteins fold are called chaperones. The structure of one of them, GroE, is shown on this slide. What this chaperone does is to provide a cavity into which an unfolded protein goes. The cavity is just big enough for one protein, so the protein can fold under conditions of high dilution. Sometimes, this kind of chaperone is referred to as an Anfinsen cage.

Take Home Messages

- The side chains of the amino acids determine the shape and properties -- and thus the functions of proteins.
- Folded proteins are marginally stable. The hydrophobic effect plays a huge role in the stability of the folded protein.
- YOU MUST UNDERSTAND how to apply the equation: $\Delta G = \Delta H - T\Delta S$
- Proteins do not sample every possible conformation in the time it takes them to fold, and sometimes folding must be catalyzed.