060123 Quiz 2 Morphology of Complex Materials

1) For a protein of N = 100 amino acid residues with j ~ 8 conformations (rotational isomeric states) per residue the number of possible conformations for the protein is j^N or about 10^{89} [Pande papers on the web page]. All but one of these states corresponds to the non-native or unfolded state. It would require 10^{66} years to explore all of these conformations [R. H. Pain, *Mechanisms of Protein Folding*, Oxford Press 1994].
   a) Explain how the concept of an energy landscape can be used to resolve this problem with protein folding.
   b) We saw several simulation results that show proteins "exploring" the energy landscape. Describe what "exploring" means in this context.
   c) What distinguishes the "native state" from the "unfolded state"?
   d) Explain how Anfinsen used β-mercaptoethanol, urea and RNAase to demonstrate an intermediate state in the folding of RNAase to the native state. (Show the structure of β-mercaptoethanol and urea and explain how β-mercaptoethanol could interact with RNAase to control folding.)
   e) What are stopped-flow and continuous flow kinetics? Explain the problems associated with these two experiments.

2) On Friday we considered circular dichroism as a method to monitor protein folding.
   a) What is the difference between linearly-polarized and unpolarized light?
   b) Explain how plane-polarized light can be produced from two circularly polarized beams.
   c) For linearly polarized light explain the difference between absorption and refraction. (Absorption coefficient and the index of refraction).
   d) For plane polarized light that is passed through a protein solution explain the state of polarization of the exiting light and explain why the exiting light has this state of polarization.
   e) List 2 other methods that can be used to monitor protein folding.

3) Proteins in cells (in vivo) probably fold by significantly different pathways compared to in vitro proteins. For example the concentration in vitro is generally on the order of 1 mg/ml while the total protein concentration in cells is on the order of 350 mg/ml.
   a) A protein in its native state is 5 to 10 nm in diameter and is extremely dense, while in the unfolded state may be 50 to 100 nm with an extremely loose structure. Such a protein might weigh 5×10^{-19} mg. At what concentration will unfolded and folded proteins have strong interactions between different chains (assume spheres and calculate the overlap concentration)?
   b) Comment on the problems that might be encountered due to concentration of proteins in the cell.
   c) Explain the following statement: "Eukaryotic genes (taken from higher cells which contain nuclei and internal organelles), when transferred into prokaryotes (bacteria, like *E. Coli*), can be expressed to form protein, but they often misfold and aggregate in the bacterial cells and form structures called **inclusion bodies**." (from St. Johns Web page)
   d) Explain what a "chaperone" is and how it might assist in protein folding.
   e) What type of protein interaction do chaperones generally enable during folding?

ANSWERS: 060123 Quiz 2 Morphology of Complex Materials

1) a) The energy landscape, as shown by Pande and Rokhsar [*Proc. Nat. Acad. Sci*. **95** 1490 (1998)], indicates that regions of conformation exist that are highly preferred energetically and that the protein molecule will gravitate towards these regions as it explores the energy landscape. It has been found from computer simulations that the energy landscape forms a kind of funnel where the protein can reach the lowest energy, native state by a number of different paths along this funnel. Local energy minima along this pathway might include various intermediate states and particularly a disorganized dense state referred to as the Molten Globule which has been referred to by Pande as a "liquid" state compared to the "crystalline" native state. In this line of thought the unfolded state is a "gas". Pande shows the energy landscape as a plot of number of "native" contacts" versus total number of contacts.

You should notice several inconsistencies in Pande's proposal that the native state/molten golubule/and unfolded state correspond to states in normal atomic or molecular systems. First, normal systems display entropy associated with free motion of subunits, that is, a gas molecule has translational entropy that is significantly larger than that associated with a mer unit in a polymer. Secondly, Pande's approach, although giving lip service to Polymer Science, largely ignores a large body of literature, a field in fact, that has extensively dealt with the physical chemistry of phase transitions in chain molecules including at least 4 noble prizes. This would seem to be a major oversight even for a biologist. The unfolded state is a liquid state and not a gas state. The molten globule and native states involve only one molecule so these are not phase transitions but isomeric states of a molecule. There are many tracks to attack what has been published by Pande and you should mention something of this from your previous training in Materials Science. The analogy with nucleation of phases in the *Current Opinion in Structural Biology* **8** 88 (1998) article is particularly troubling if you think about the details of this comparison and if you have any knowledge of homogeneous nucleation theory. There is a reason (you should recognize quickly) that Pande goes no further than to suggest this idea. (This does not mean that you do not appreciate Pande's contribution.)

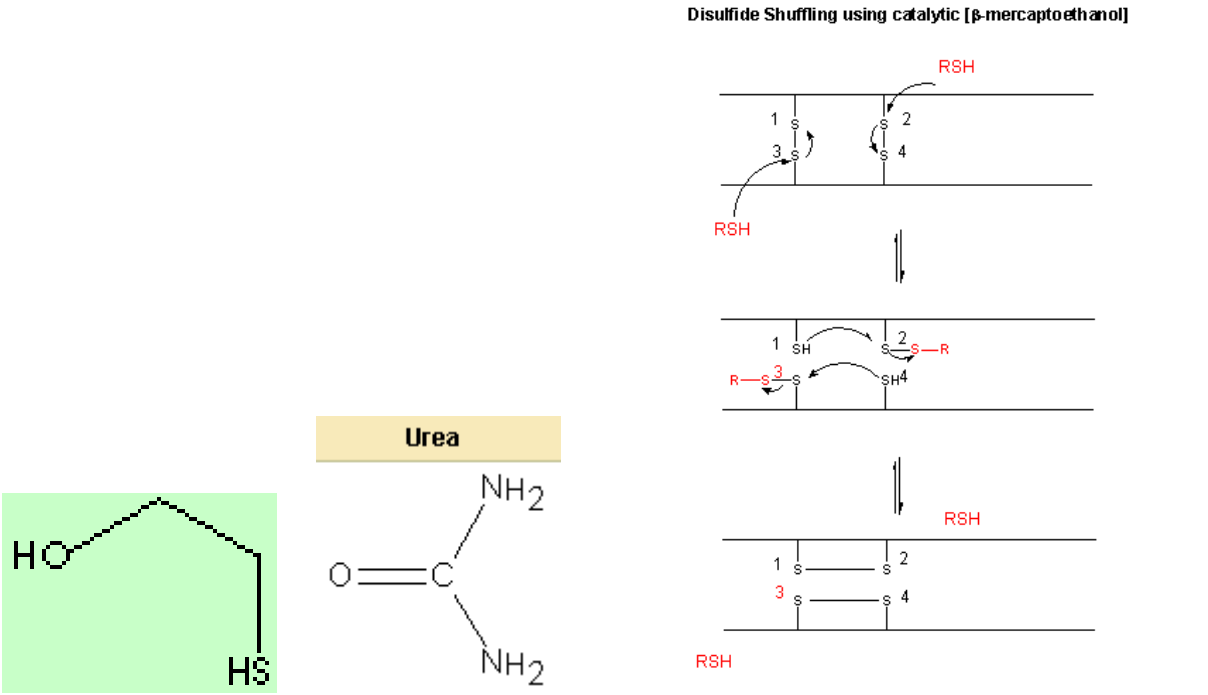
b) Exploring the energy landscape, in the context of Pande's plots of fraction of native contacts versus total number of contacts would mean swapping of partners between binary interactions of residues in the protein in a more or less random way except that the energy of the resulting state would be used to give the probability of that swap happening. This approach is common in Monte Carlo conformational simulations of synthetic polymers for instance and has been commonly used for more than 35 years. In terms of 2d graphics this probing of the energy landscape becomes a molecular dance. the polypeptide chain shakes and translates until it finds native conformations or low energy conformations that lock in parts of the chain. Finally, we usually see one part of the molecule that swings wildly, even 180 degree swings of certain residues to give the final native state. This was the antennae in the simulation we saw in class. The final native state is locked-in due to the low free energy though some slight structural vibrations can be seen.

c) The native state is defined by biological functionality. If the protein is active it is in its native state. It is also implied that the native state is the lowest energy state, generally

the smallest and densest packed state. The unfolded state is similar to a synthetic polymer in solution except that certain secondary structures exist in the unfolded state making the structure simpler and easier to recognize than a true random state such as would be seen in a synthetic polymer. Put differently, the unfolded state displays large and more irregular chain persistence compared to a normal polymer.

d) Anfinsen chose a simple protein that had 4 disulfide bonds that held the tertiary structure together in the native state, RNAase. He understood that he could denature this protein by breaking these disulfide bonds using urea. Mercaptoethanol is capable of bonding to these cystine groups to lock-out disulfide bonds. The mercaptoethanol can also destabilize a bonded intermediate state if a small amount of mercaptoethanol is added. Using this tool Afinsen was the first to reversibly denature and refold a protein to a native state structure making the mystery of tertiary structure controllable.

Figure: CATALYTIC SHUFFLING OF DISULFIDES WITH BETA-MERCAPTOETHANOL



e) Stopped-flow experiments are intended to study protein folding generally using spectroscopic techniques, circular dichroism or flourescense. In stopped flow the unfolded protein (perhaps too acidic to fold) is mixed rapidly (1 ms) with a diluent or other solution that encourages folding, the mixed solution is then flowed into an observation chamber. This apparatus is used due to the high cost of pure protein. It is the most expedient method with a small amount of protein. A continuous flow experiment allows for a more rapid observation of the folding but at the cost of using much more protein. There is a problem with using smaller tubes in a continuous flow experiment in that mixing is not effective in tiny tubes and it is difficult to have high linear velocities in small tubes. For the continuous-flow experiment the observation time is only limited by the velocity so folding on micro-second time scales are apparently possible.

2) a) The electric field vector for linearly polarized light is normal to the direction of propagation and restricted to a plane that includes the direction of propagation. For unpolarized light the electric field vector is normal to the direction of propagation but is not restricted otherwise.

b) If two circularly polarized light contain one component (y for instance) that is out of phase by +90° while the other circularly polarized light contains the same (y) component that is out of phase by -90° then summing the two circularly polarized beams will lead to a plane polarized beam in the x-direction.

c) Absorption leads to a reduction in the amplitude (and intensity) of the light. Refraction changes the speed and leads to a phase shift when compared with a non-refracted beam.

d) The protein solution displays circular dichroism so plane polarized light becomes elliptically polarized after passing though the solution. This is understood by considering that the plane polarized light can be decomposed into two circularly polarized beams of opposite phase (180 degrees for the y component for instance). The circularly dichroic solution preferentially absorbs one of the two circular polarized components leading to a phase shifted elliptically polarized beam. The major axis of polarization is offset from that of the plane polarized incident beam due to circular birefringence.

e) UV absorption and fluorescence are also used.

3) a) The overlap concentration is the concentration within a protein molecule. To calculate ϕ^* we need the mass divided by the occupied volume of the protein. Here we have been given the mass and size. Assuming that the size is the diameter the 5 nm native state has an overlap concentration of 7.6 mg/ml while the 100 nm unfolded state 0.001 mg/ml. Above the overlap concentration significant overlap between different molecules are expected (aggregation is likely).

b) The hydrophobic segments of proteins will aggregate or agglomerate non-reversibly. Inclusion bodies may form that can not be broken apart.

c) While the prokaryotic cells can apparently produce the peptide sequence, the polypeptides can not fold properly in the absence of other features, perhaps including molecular chaperones that provide an *aggregation-free-zone* where proteins can fold into their native states.

d) There are two main types of molecular chaperones. Heat shock proteins (70 kg/mole) Hsp-70, and chaperonins. Heat shock proteins act as the polypeptide chain emerges from the ribosome site binding to hydrophobic sections of the polypeptide. Heat shock proteins act as single protein chains not as quartenary structures. They bind with peptides when ATP is also bound and release proteins after cleaving ATP, using energy. Chaperonins are smaller Chaperonin-60 and chaperonin-10 for instance, but act in large quartenary structures, for instance GroEL and GroES discussed in class. Chaperonins act away from the ribosome, for example in the mitochondria or other cellular structures. Chaperonins also bind ATP and cleave ATP on releasing the native state proteins.

e) Chaperonins enable hydrophobic interactions that couldn't occur in the cytoplasm since folding that leads to hydrophobic interactions often involves revealing hydrophobic segments of the polypeptide to the cellular environment where these sites would likely bond with other hydrophobic groups on other proteins and lead to aggregation.