Effect of denaturant and protein concentrations upon protein refolding and aggregation: A simple lattice model

PRASHANT GUPTA, CAROL K. HALL, AND ANNE C. VOEGLER

Department of Chemical Engineering, Box 7905, North Carolina State University, Raleigh, North Carolina 27695

(RECEIVED January 7, 1997; ACCEPTED September 5, 1998)

Abstract

We present a study of the competition between protein refolding and aggregation for simple lattice model proteins. The effect of solvent conditions (i.e., the denaturant concentration and the protein concentration) on the folding and aggregation behavior of a system of simple, two-dimensional lattice protein molecules has been investigated via dynamic Monte Carlo simulations. The population profiles and aggregation propensities of the nine most populated intermediate configurations exhibit a complex dependence on the solution conditions that can be understood by considering the competition between intra- and interchain interactions. Some of these configurations are not even seen in isolated chain simulations; they are observed to be highly aggregation prone and are stabilized primarily by the aggregation reaction in multiple-chain systems. Aggregation arises from the association of partially folded intermediates rather than from the association of denatured random-coil states. The aggregation reaction dominates over the folding reaction at high protein concentration and low denaturant concentration, resulting in low refolding yields at those conditions. However, optimum folding conditions exist at which the refolding yield is a maximum, in agreement with some experimental observations.

Keywords: computer simulation; folding intermediates; folding pathways; protein aggregation; protein folding

Protein aggregation is a serious problem in the biotechnology, biomedical, pharmaceutical, and food industries. In biotechnology, protein aggregation can limit the recovery of recombinant protein from inclusion bodies, making it a sizable obstacle to successful commercialization of recombinant technologies (Baldwin, 1975; Kim & Baldwin, 1982; Marston, 1986; Kane & Hartley, 1988; Mitraki & King, 1989; Gierasch & King, 1990; Georgiou & Bowden, 1990). In biomedical research, diseases such as Alzheimer's disease (Simmons et al., 1994; Selkoe, 1995), Down's syndrome (Massry & Glasscock, 1983), cataracts (Clark & Steele, 1992), prion disease (Moore & Melton, 1997), sickle cell disease (Eaton & Hofrichter, 1990), and light-chain disease (Gallo et al., 1996) are thought to be caused in part by abnormal protein aggregation. In the pharmaceutical industry, protein aggregation can interfere with the maintenance of protein stability during the packaging, shipping, storage, and administration of protein-based drugs (Ahern & Manning, 1992; Costantino et al., 1995). In the food industry, protein aggregation can lead to the production of off-flavors and unwanted color changes (Persson & Gekas, 1994). Finally, in biochemistry research, protein aggregation is a fact of life, and often

times a nuisance. According to King (1989), "Hundreds of papers in the biochemistry literature describe experiments on protein denaturation or folding where, at some point, the protein 'aggregated' or 'precipitated,' ending the experiment."

Despite the obvious practical importance of protein aggregation (Sadana, 1995), our understanding of its physical basis is far from complete. The general picture that has emerged thus far is that intramolecular hydrophobic residue–residue interactions, which favor the folded state, compete with inter-molecular hydrophobic residue–residue interactions, which favor the aggregated state. The winner of this competition depends on the protein concentration (increasing protein concentration increases the likelihood of inter-residue contacts and hence favors aggregation) and on the denaturant concentration (decreasing denaturant concentration increases the strength of both inter- and intra-residue attractions). The research community is divided over the identity of the aggregating species. Some groups take the view that the aggregating species is the denatured state; this then allows them to treat proteins as model heteropolymers and to construct phase diagrams in the protein concentration/denaturant concentration plane (Fields et al., 1992; Dill & Stigter, 1993, 1995; De Young et al., 1993a, 1993b). Other groups take the view that the aggregating species is a partially refolded intermediate (London et al., 1974; Zettlmeissl et al., 1979; Goldberg & Zetina, 1980; Mitraki & King, 1989). Their idea is

Reprint requests to: Carol K. Hall, Department of Chemical Engineering, Box 7905, North Carolina State University, Raleigh, North Carolina 27695; e-mail: hall@turbo.che.ncsu.edu.

that, as the protein folds along the well-defined pathway towards the native state, certain partially folded intermediates become highly populated, depending upon the denaturant concentration. Because the intermediates have exposed hydrophobic surfaces, they can associate in an off-pathway reaction leading to the aggregated state. The latter view of aggregation has also been adopted to a certain extent by research workers (Kelly, 1996; Wetzel, 1996) aiming to understand the deposition of normally soluble proteins as insoluble amyloid fibrils in the so-called protein deposition diseases, or amyloidoses (Sipe, 1992), one of which is Alzheimer's disease. The nature of folding pathways and intermediates is thought to depend strongly on denaturant concentration, with different intermediates being populated at different denaturant concentrations (Schmid, 1992). This suggests that it might be possible to prevent aggregation by altering denaturant concentration so as to avoid the formation of aggregation-prone intermediates. An understanding of the effect of solvent conditions upon folding intermediates and aggregation could, therefore, play a crucial role in formulating a strategy aimed at improving recombinant protein yield.

In this paper, we describe a study aimed at determining how the competition between protein refolding and aggregation is influenced by both denaturant concentration and protein concentration. Because the slow nature of aggregation processes precludes the use of highly realistic models of protein molecules, we have adopted a simple lattice approach, following the work of Lau and Dill (1989). In this approach, a protein chain is modeled as a unique sequence of hydrophobic (H) and polar (P) beads arranged on a two-dimensional square lattice. This model captures what is arguably the most essential feature of the interactions that drive protein folding and aggregation, namely the hydrophobic effect, yet is simple enough to allow the simulation of a many-chain system. Despite the obvious rudimentary nature of such models, they have proven to be quite useful because they allow investigators to extract the general theoretical principles that underlie the connection between a protein's sequence and its native structure, folding pathways, kinetics, etc. (Dill et al., 1995). In considering the results of such studies, it is helpful to adopt a figurative, as opposed to a literal, interpretation of the model's features. For example, although the beads on a lattice protein are often referred to as individual residues, it would be more appropriate to think of them as groups or clusters of either hydrophobic or polar residues along the protein chain.

In our previous papers (Gupta & Hall, 1995, 1997), we studied the folding-unfolding behavior of an isolated model protein chain with a specific sequence of H and P beads. Exhaustive conformational searches were carried out in order to quantify the folding-unfolding transition. Dynamic Monte Carlo simulations were performed at various denaturant levels to elucidate the folding pathways. In this paper, we extend our dynamic Monte Carlo simulations to multiple-chain systems comprised of model protein chains with the same sequence in order to observe aggregate formation at various protein and denaturant concentrations.

Highlights of the results reported in this paper are the following. Aggregation is seen to arise from the association of partially folded intermediates (chains with >60% native structure) rather than from the association of denatured random-coil states. The size of the aggregates increases as the protein concentration increases. The multichain environment stabilizes some partially folded intermediates that are not even present on the isolated chain's folding pathway. We consider this to be a significant finding. It suggests that

protein folding pathways are strongly perturbed by the presence of other proteins. It also suggests that aggregation occurs at high protein concentration, not solely because protein-protein contacts are more likely, but also because the crowded environment makes it more likely that an aggregation-prone intermediate will appear on the folding pathway and be stabilized. Analysis of the variation in the population of the nine most-populated folding intermediates with protein and denaturant concentrations provides us with a reasonably clear picture of how and why changes in protein and denaturant concentrations affect the populations of partially-folded open (i.e., aggregation-prone) intermediates in our model. We find that changing solution conditions has a twofold effect upon the population of any given intermediate in that it can tend to favor the forward reaction (intermediate formation) because of enhanced intermolecular interactions, while simultaneously tending to favor the reverse reaction (intermediate disappearance) because of enhanced intramolecular interactions, or vice versa. We find that there are some general rules that govern how changes in protein and denaturant concentrations affect the aggregation and folding processes; these rules are outlined in the Discussion. The observed refolding yield is low at high protein concentrations where the aggregation reaction begins to dominate, and at very low denaturant concentrations where the interchain HH attraction becomes very strong, trapping partially folded protein chains in aggregates. There exist optimum levels of denaturant and protein concentrations where the refolding yield is highest.

We begin in the Model and sequence section by describing the lattice model that we have employed to study folding and aggregation. In the following section we review those key results from our earlier papers (Gupta & Hall, 1995, 1997) that are crucial to understanding what we present here. In the section following that, we present the details of our multichain simulations. In Results, we discuss the influence of denaturant and protein concentration upon folding intermediates and aggregation in the first subsection, and describe aggregate formation and refolding yields in the next subsection. Conclusions are presented in the Discussion.

Model and sequence

Because the potential energy surface for a protein molecule is extremely complex (Dill, 1990) and because most proteins fold on a time scale of milliseconds to minutes, the computational resources currently available preclude highly realistic modeling of folding phenomena. Realistic modeling of aggregation phenomenon is even more difficult, because it involves multiple chains and even longer time scales. To circumvent these problems we have employed a simple model of a protein molecule that captures the essence of intra- and interchain protein interactions, but at the same time allows for simulation of multiple-chain systems.

A two-dimensional square lattice model proposed by Lau and Dill (1989) has been employed. The protein molecule is represented as a sequence of hydrophobic (H) and polar (P) beads arranged in a self-avoiding random walk on a square lattice. A bead can be thought of as representing a group of amino acid residues. To incorporate the hydrophobic effect, two non-bonded H beads that are situated next to each other in a given chain conformation are said to attract each other with a potential of mean force of strength ϵ . We define the dimensionless parameter ϵ^* as $\epsilon^* = \epsilon/kT$, where T is the absolute temperature and k is the Boltzmann constant. The sign of ϵ^* is always negative to indicate that the HH interaction is attractive in nature. The magnitude of ϵ^* , $|\epsilon^*|$, is a

Fig. 1. The native state of the 20-bead protein. The H beads form a compact hydrophobic core that is surrounded by the P beads. There are eight HH contacts in the native conformation.

measure of the denaturant concentration in the system. Low values of $|\epsilon^*|$ imply that the HH interactions are weak, and hence, the denaturant concentration is high; whereas high values of $|\epsilon^*|$ imply that the HH interactions are strong, and hence, the denaturant concentration is low. The conformation with the maximum number of HH contacts is the native state. This simple model captures what is arguably the most important feature of the intraprotein interaction, namely the hydrophobic effect. The protein concentration in the solution is characterized by the packing fraction inside the simulation box, ρ , defined to be the number of occupied lattice sites in the simulation box divided by the total number of lattice sites.

Because protein refolding and aggregation depend upon the folding pathways and intermediates, we devised a model protein sequence capable of forming a unique native state and well-defined partially folded intermediates that are aggregation prone. Our 20-bead hypothetical protein sequence contains 12 P beads and 8 H beads. The native state for this sequence, shown in Figure 1, contains 8 HH contacts. In the native state, all of the hydrophobic beads are tightly packed inside a hydrophobic core that is surrounded by a shell of polar beads.

Review of previous results on isolated protein model

In our previous papers (Gupta & Hall, 1995, 1997), we carried out exhaustive conformational searches on the hypothetical sequence to characterize the folding–unfolding transition. In an exhaustive conformational search, every possible conformation of the protein molecule is enumerated on the computer. We also carried out dynamic Monte Carlo simulations on this sequence to elucidate the folding pathmaps and intermediates at various levels of denaturant concentration. Here we briefly review those key findings that are relevant to multiple-chain simulations.

Exhaustive searches of the conformational space of the isolated 20-bead protein sequence allowed us to exactly determine the native state as well as the canonical partition function, Z (Gupta & Hall, 1995). The canonical partition function, in turn, is used to

calculate ensemble-averaged equilibrium properties such as radius of gyration, $\langle R_g^2 \rangle^{1/2}$, average energy, $\langle E \rangle$, specific heat, $\langle C \rangle$, etc. The exhaustive conformational searches revealed that, at low denaturant concentration, the model chain is in an unfolded random-coil state characterized by a large radius of gyration and a small number of HH contacts. As the denaturant concentration is decreased, the chain folds into the compact native state with 8 HH contacts. Figure 2 shows the location of the folding–unfolding transition by displaying the average number of HH contacts as a function of $|\epsilon^*|$.

Dynamic Monte Carlo simulations were performed on the isolated 20-bead protein sequence to elucidate the folding pathmaps and intermediates at various levels of denaturant concentration (Gupta & Hall, 1997). The folding process for the hypothetical protein was seen to be cooperative in nature in that the protein chain does not start folding until the middle section of the chain has folded correctly. Once the middle section is formed, the rest of the chain zips up into the native state relatively quickly. Also, the model protein exhibits well-defined folding pathways and intermediates which change with the denaturant concentration.

The folding pathway for a protein is defined as the ordered sequence of conformations that the protein molecule assumes as it progresses towards the native state from a random-coil conformation. We define a conformation to be completely characterized by all the non-bonded HH contacts present in the conformation, i.e., two conformations having identical HH contacts are deemed identical even if their HP and PP contacts are different. The 20-bead hypothetical chain is capable of forming 14 different HH contacts. In any given conformation, each of these contacts can either be “on” or “off.” Thus, the model protein can assume at the most 2^{14} distinct conformations. Due to excluded volume and connectivity constraints, however, the actual total number of possible conformations is less than 2^{14} .

To determine the folding pathway, the ordered sequence of conformational states obtained from the simulation is examined for the

Fig. 2. Folding transition for the 20-bead model chain. This graph shows the average number of HH contacts vs. the dimensionless HH contact energy, $|\epsilon^*|$. At low values of $|\epsilon^*|$, the chain is in a random coil state characterized by a small number of HH contacts; as $|\epsilon^*|$ is increased, the chain folds into the native state with eight HH contacts.

presence of "loops" and all the loops are removed. The conformation sequence is said to contain a "loop" when a conformation present at time step t is also found at a later time step $t + t'$. The resulting loop of length t' is then removed from the sequence. After removing all the loops, we are left with the path taken by the protein molecule to go from a random coil state to the native state.

The populations of the folding intermediates are observed to be highly dependent upon the denaturant concentration ($|\epsilon^*|$). During the course of the folding process, the protein chain sometimes gets trapped in non-native conformations that require breaking an HH contact to fold further. At high denaturant concentration (low $|\epsilon^*|$), the probability of breaking an HH contact is fairly high and, therefore, the protein chain can break out of a trap fairly easily. However, at low denaturant concentrations (high $|\epsilon^*|$), the probability of breaking an HH contact is extremely small and, thus, the protein chain is forced to stay in such trap conformations for longer periods of time. Therefore, the population in misfolded trap configurations is observed to be higher at low denaturant concentrations. For a detailed discussion of folding intermediates and kinetic traps, we refer the reader to our previous paper (Gupta & Hall, 1997). Finally, the rate of folding was found to decrease as $|\epsilon^*|$ increases, so that the protein took longer to fold at low values of denaturant concentration.

Dynamic Monte Carlo simulations of multichain systems

To study protein aggregation, dynamic Monte Carlo simulations were carried out on multiple-chain systems at various denaturant concentrations and protein concentrations. The simulation system contained 25 to 40 protein chains (500 to 800 beads). The move set employed for the simulations reported here is almost identical to the one that we used for our isolated chain simulations (Gupta & Hall, 1995) and included three-bond moves, four-bond moves, and chain-end rearrangements. A three-bond move displaces two beads in the middle of the chain, while a four-bond move displaces three beads and creates a random local configuration in the middle of the chain. The only change that we made to our previous move set is that we added a translation move in which a randomly chosen chain is translated rigidly by a unit distance in a randomly chosen direction. The simulation time was measured in terms of "Monte Carlo time steps." A Monte Carlo time step is defined as N consecutive attempted moves, where N is the number of beads in the simulation system. Therefore, for a simulation system consisting of 40 chains with 20 beads each, one Monte Carlo step is equivalent to 800 consecutive attempted moves.

Simulations were carried out at various refolding solution "conditions." The condition of a refolding solution is characterized by two parameters, the HH attraction strength ($|\epsilon^*|$) and the packing fraction (ρ). The dimensionless parameter, $|\epsilon^*|$, characterizes both interchain and intrachain HH attraction. The simulations were conducted at 22 different sets of conditions. We chose three values of $|\epsilon^*|$ (4, 6, and 8) and seven values of packing fraction (0.05, 0.1, 0.25, 0.4, 0.52, 0.68, and 0.8), giving rise to 21 combinations. In addition, we carried out simulations at a packing fraction of 0.59 and an $|\epsilon^*|$ value of 4. Each simulation was started from a random initial configuration that was generated by first "growing" the required number of chains in our simulation box and then relaxing the system for a quarter of a million time steps at an $|\epsilon^*|$ value of zero (i.e., no HH attraction). This ensured that the initial configuration was sufficiently random and completely denatured. At this point the HH attraction was "turned on" and the simulation was

started. To specify the stopping criterion for the simulation, we monitored running averages of the number of folded chains, the number of aggregated chains, and the total number of HH contacts in the system. We divided the time sequence obtained from the simulation into three equal parts and stopped the simulation only if the ensemble averages for the last third of the simulation were within 2.5% of the ensemble averages for the middle third of the simulation. Also, the ensemble averages were calculated only after the simulation had completed two million time steps. Therefore, every simulation had to run for at least for 2×10^6 time steps (about 10^9 attempted moves). It turned out that the shortest simulation ran for about 5 million time steps.

Our simulations indicate that aggregation takes place on time scales that are two to four orders of magnitude longer than refolding time scales, depending upon the denaturant concentration and protein concentration. For example, the average folding time for the isolated chain was around 3×10^4 time steps at $|\epsilon^*| = 4$ and about 2×10^5 at $|\epsilon^*| = 8$. The corresponding aggregation simulations had to be run for a few million time steps (several billion attempted moves) at $|\epsilon^*| = 4$, and a few billion time steps (several trillion (10^{12}) attempted moves) at $|\epsilon^*| = 8$. The longest simulations required over a month of CPU time for each run on a DEC-Alpha 3000 workstation. Naturally, during such long simulations, some of the chains underwent the folding-unfolding transition thousands of times.

Because some of the simulations had to be run for extremely long periods of time, it was not possible to conduct multiple simulation runs for each one of the 22 solution conditions that we studied. For solution conditions that required relatively short simulations, we conducted several simulation runs for each condition. We conducted fewer runs for conditions that required longer simulation times. We conducted only one run at solution conditions that required extremely long simulations. As a result, we were only able to compute error bars by calculating the standard deviation of observed values for those conditions where we conducted multiple runs.

Results

In this section we present the results obtained from the dynamic Monte Carlo simulations of multichain systems. The next subsection describes the influence of protein and denaturant concentrations upon the populations and aggregation propensities of the nine most-populated configurations. In the second subsection, we discuss the refolding yield and aggregation as a function of denaturant and protein concentration.

Folding intermediates and aggregation

We begin this subsection with a general description of the factors that determine the populations of any intermediate and the effects that variations in protein and denaturant concentrations have on these factors.

The population of any configuration over the course of a simulation is determined by both the rate of formation of the configuration and the average residence time of the configuration once it has been formed. These two factors are each influenced by a set of complex molecular mechanisms. The rate of formation depends upon how "well-traveled" the pathway branches are that give rise to the configuration in question. The probability of a pathway branch being traveled itself depends upon the relative populations

of all the “precursor” configurations. The average residence time of a configuration depends upon how aggregation prone that configuration is and whether or not the configuration is a “kinetic trap,” defined to be a steep local minimum on the potential energy surface. Once a protein chain gets caught in a kinetic trap configuration, it finds it very hard to escape because the only way to do so is to break HH contacts. Breaking HH contacts becomes progressively more difficult as denaturant concentration is lowered; hence, the kinetic traps become “deeper” at low denaturant concentration. It is quite conceivable that a configuration may have a low rate of formation but may be a deep kinetic trap and have a long residence time, thus exhibiting a high population over the course of simulation. It is also conceivable that a configuration may have a high rate of formation but may have a very short residence time, thus exhibiting a low overall population over the course of the simulation. These phenomena are further complicated by the fact that both the denaturant concentration and the protein concentration have a profound effect upon the folding pathways as well as on the stability of individual configurations.

To understand how changes in the denaturant and/or protein concentration affect the interplay between aggregation and refolding, we investigated the stability of open configurations. Open configurations are the ones with relatively large radii of gyration and, as a consequence, contain several exposed H beads that makes them aggregation prone. Compact configurations, on the other hand, are tightly folded upon themselves; they have relatively small radii of gyration and, hence, contain few exposed H beads, which implies that they are not very aggregation prone. Solution conditions that encourage either the formation of, or preservation of, open configurations tend to favor aggregation.

As shown in the plots and discussion that follow, we find that the stabilities of both open configurations and compact configurations are profoundly influenced by solution conditions. Changing solution conditions often has a twofold effect upon the stability of a given configuration in that it can simultaneously cause the configuration’s rate of formation to increase because of enhanced intramolecular attractions and its residence time to decrease because of enhanced intermolecular attractions, or vice versa. For example, consider the effect that variation in the denaturant concentration has on the stability of open configurations. Low denaturant concentration (high $|\epsilon^*|$) discourages the formation of open configurations because the strong intrachain HH attraction energy causes the chains to collapse upon themselves very quickly. At the same time, low denaturant concentration preserves open configurations because the strong interchain HH attraction energy causes the open configurations to participate in very stable aggregates. High denaturant concentration (low $|\epsilon^*|$), on the other hand, favors formation of open configurations because the weak intrachain HH attraction energy associated with forming more compact configurations is not sufficient to overcome the configurational entropy loss. At the same time, high denaturant concentration discourages the preservation of open configurations because the low interchain HH attraction energy is insufficient to maintain aggregation.

Changing the protein concentration also has a twofold effect upon the stability of open configurations. High protein concentration discourages the formation of open configurations because the excluded volume constraints force the chain to pack more tightly. At the same time, high protein concentration encourages the preservation of open configurations because the higher probability of forming interchain HH contacts leads to the formation of very stable aggregates.

Configurations that are also kinetic traps have an additional set of responses to changes in denaturant or protein concentration. As mentioned earlier, kinetic traps become deeper at low denaturant concentration (high $|\epsilon^*|$); this means that low denaturant concentration encourages the preservation of kinetic trap configurations. High protein concentration also tends to preserve kinetic traps due to increased aggregation and excluded volume constraints.

To study how the population profiles of the configurations in our multichain simulations vary with denaturant and protein concentration, we monitored the configuration of each chain at every time step and the identity of chains that participate in aggregates. (A chain is defined to be participating in an aggregate if it forms at least one interchain HH contact.) From this data, we calculated the occurrence probabilities for all configurations that were encountered, defined to be the probability of finding a randomly chosen chain at a randomly chosen time step in that particular configuration. If there are n_t chains in a given conformation at time step t , the occurrence probability for that configuration can be calculated as

$$P_{\text{occurrence}} = \frac{1}{N\tau} \sum_{t=1}^{\tau} n_t,$$

where N is the total number of chains and τ is the total number of time steps in the simulation.

The dependence of the occurrence probabilities on denaturant and protein concentration for the nine most-populated non-native configurations (labeled A through I) is illustrated in Figures 3–5. Configurations A, B, G, H, and I are found on the folding pathmap for an isolated chain (Gupta & Hall, 1997). Configurations C and F also have significant residence times in isolated chain folding simulations; but, because they occur mostly in “loops,” they do not

Fig. 3. Probabilities of finding the model chain in a given configuration at $|\epsilon^*| = 4$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

Fig. 4. Probabilities of finding the model chain in a given configuration at $|\epsilon^*| = 6$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

appear on the path maps. Configurations *D* and *E* do not have significant residence times during isolated chain folding simulations, yet they emerge as two of the nine most-populated species in a multichain environment, as seen in Figures 3–5. Figures 3–5 depict the occurrence probability of the nine configurations at the packing fractions enumerated in the Dynamic Monte Carlo simulations section at $|\epsilon^*| = 4$ (Fig. 3), 6 (Fig. 4), and 8 (Fig. 5). The

Fig. 5. Probabilities of finding the model chain in a given configuration at $|\epsilon^*| = 8$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

configurations are shown underneath the x axis, the packing fraction is displayed along the y axis, and the z axis represents the probabilities. As mentioned in the Dynamic Monte Carlo simulation section, we calculated error bars for those conditions where we conducted multiple simulation runs. We find that at $|\epsilon^*| = 4$, the results from different runs at any given value of packing fraction are within $\pm 5.62\%$ of the average value. At $|\epsilon^*| = 6$, the results are within $\pm 6.56\%$ of the average. For the sake of clarity, we have omitted these error bars from the charts and have only plotted the average values.

We begin our description of Figures 3, 4, and 5 by focusing on how the populations of the three relatively compact configurations, labeled *A*, *B*, and *C*, change with denaturant concentration and with protein concentration. We can see that the population of configuration *A* decreases sharply as the denaturant concentration is reduced from $|\epsilon^*| = 4$ to $|\epsilon^*| = 6$. The change from $|\epsilon^*| = 6$ to $|\epsilon^*| = 8$, however, has very little effect on the population. Configuration *A* is a near-native state, with one of the chain ends in a dangling random configuration; six of the eight native HH contacts are made and there are no incorrect HH contacts that need to be broken. To pin down the chain end and reach the native state causes a decrease in configurational entropy but results in the formation of two new HH contacts. At high denaturant concentration (low $|\epsilon^*|$), the energy decrease associated with formation of an HH contact is not beneficial enough to compensate for the accompanying entropy loss; therefore, the probability of finding a chain in configuration *A* remains relatively high. As the denaturant concentration is reduced ($|\epsilon^*|$ is increased), the balance between the energy decrease and entropy loss shifts; hence, configuration *A*'s probability of occurring decreases. As the protein concentration is increased, the unfolded chain end in configuration *A* becomes more and more likely to get entangled with other chains, thus resulting in increased residence times for this intermediate.

Configuration *B* in Figures 3, 4, and 5 is an example of a kinetic trap. This configuration has seven HH contacts, six of which are native. The non-native HH contact must be broken before the protein can fold into its native state. The population of configuration *B* decreases with increasing denaturant concentration (decreasing $|\epsilon^*|$) as well as with increasing protein concentration, a trend that is the opposite of the trend exhibited by configuration *A*. The higher occurrence probability at low denaturant concentration is probably due to the fact that configuration *B* is a kinetic trap. Because breaking an HH contact is more difficult at low denaturant concentration, the residence time of kinetic trap configurations is high at these conditions, resulting in relatively high occurrence probabilities. The decrease in occurrence probability with increasing protein concentration is contrary to our previous observation that high protein concentration tends to preserve kinetic traps. This contradiction can perhaps be explained by our observation that configuration *A* is a precursor to the formation of configuration *B* in the folding pathway of single-protein systems (Gupta & Hall, 1997). At low values of protein concentration, the unfolded tail in configuration *A* has lower excluded volume restrictions and more freedom to “flop around,” and hence, is more likely to misfold into configuration *B*, thus giving rise to relatively high populations of configuration *B*. At high values of protein concentration, the unfolded tail of configuration *A* is relatively frozen, and hence, is not as likely to fold into configuration *B*.

Configuration *C* in Figures 3, 4, and 5 presents another example of a kinetic trap. The lower half of the chain is in a native-like conformation and contains four native HH contacts, while the

upper half is completely misfolded and contains three nonnative HH contacts. To get into the native state, the chain needs to break the three nonnative HH contacts, making this configuration a kinetic trap. The population profile for this configuration indicates that its probability of occurrence remains relatively constant across protein concentrations. Configuration *C* has only one exposed H bead, and is not very aggregation prone; it is largely stabilized by intrachain associations. Because interchain associations do not play a significant role in the stability of this configuration, the concentration of protein chains in the surrounding solution has little impact on its occurrence probability. Configuration *C*'s probability of occurrence does not show a significant dependence upon the denaturant concentration either. We suspect that this behavior is the result of complex interplay among underlying factors that affect the stability of protein configurations and cannot point to any one factor that could explain this behavior.

Configurations *D* and *E* are partially folded intermediates that are not found on the folding pathways of the isolated model protein chain. Configuration *D* exhibits a pocket containing four exposed H beads, and configuration *E* has two exposed H beads, rendering both of these configurations aggregation prone. These configurations present an interesting case in that they emerge as two of the nine most populated configurations in a multichain environment, evidently by the virtue of their ability to form aggregates. They are stabilized primarily by interchain interactions in the multichain environment. Although the exact dependence of their population profiles on the solution conditions appears quite complex, their probabilities of occurrence do seem to be lower at $|\epsilon^*| = 8$ than at the other two $|\epsilon^*|$ values (except at very high protein concentrations). The occurrence probabilities generally increase with increasing protein concentrations because aggregation is more likely at higher protein concentrations. Once again, we believe that the complex dependence upon solution conditions is a reflection of the complexity of underlying factors affecting the stability of partially folded species.

Configurations *F* and *G* are somewhat open and very aggregation prone. Configuration *F* contains five exposed H beads, while configuration *G* contains four exposed H beads. Because these configurations contain only native-like HH contacts, they are not kinetic traps. We can see from Figures 3–5 that their probabilities of occurrence increase monotonically as the protein concentration is increased, which can be attributed to the fact that open configurations are more likely to aggregate at higher protein concentrations. If we ignore the populations at the highest value of packing fraction (0.8), we also see that their probabilities of occurrence are higher at high denaturant concentration (low $|\epsilon^*|$) and decrease slightly as the denaturant concentration is decreased. This is because, as mentioned earlier, the probability of finding an open configuration is higher at high denaturant concentration (low $|\epsilon^*|$).

Configurations *H* and *I* represent an extreme case. Both of these configurations are "open" in that they both have relatively large radii of gyration and contain significant stretches of exposed H beads. Both of these chains contain six native HH contacts and four exposed H beads. Even though excluded volume constraints generally discourage formation of such open configurations, the residence time for these two configurations is fairly high because they are extremely aggregation prone. In addition, both of these configurations are kinetic traps because, given our set of Monte Carlo moves, it is not possible for a protein chain in either of these two configurations to fold into the native state without breaking any existing HH contacts. At low $|\epsilon^*|$ the population profiles ex-

hibit maxima at moderate protein concentrations. At higher values of $|\epsilon^*|$, however, the behavior becomes more complex. The probabilities of occurrence seem to decrease with decreasing denaturant concentration (increasing $|\epsilon^*|$), which is similar to the behavior exhibited by configurations *F* and *G*. This behavior can again be attributed to the lower probability of finding an open configuration at low denaturant concentration.

Most of the chains that participate in aggregation contain a substantial amount of native-like structure, implying that aggregation results from association between partially folded intermediates rather than between denatured (random coil) proteins. Figure 6 shows the average number of HH contacts per chain for proteins that have aggregated vs. packing fraction. It can be seen that the aggregated chains exhibit a substantial number of native HH contacts and relatively few nonnative HH contacts. In general, the number of native contacts decreases with increasing packing fraction. An example of an aggregate is shown in Figure 7. This aggregate contains 10 protein chains and was observed at $|\epsilon^*| = 8.0$ at a packing fraction of 0.68. It is evident from Figure 7 that most of the participating chains contain stretches of native-like secondary structure.

To characterize the aggregation behavior of the nine most-populated configurations, we define a parameter termed the "aggregation propensity" as the conditional probability that a particular configuration will participate in an aggregate given that the configuration has already been formed. The aggregation propensity is a measure of how aggregation prone a configuration is. If there are n_t chains in a given configuration at a time step t , and a_t chains out of those are participating in aggregates (which means that they exhibit at least one interchain HH contact), then we can define the aggregation propensity, P_{agg} as

$$P_{agg} = \frac{1}{\tau} \sum_{t=1}^{\tau} \frac{a_t}{n_t}.$$

The aggregation propensities of the nine configurations discussed previously are shown in Figures 8–10. These three-

Fig. 6. Average number of HH contacts in aggregated chains at $|\epsilon^*| = 4.0, 6.0$, and 8.0 . The aggregated chains exhibit a substantial number of native-like HH contacts.

Fig. 7. Example of a large aggregate involving 10 protein molecules. Some of the participating chains are in recognizable partially folded intermediate states, while the others are simply trapped in some random configurations due to excluded volume interactions from neighbors.

dimensional bar charts depict the aggregation propensities at the packing fractions enumerated previously for $|\epsilon^*| = 4$ (Fig. 8), 6 (Fig. 9), and 8 (Fig. 10).

Although the aggregation propensity profiles in Figures 8, 9, and 10 appear quite chaotic and cannot be explained by pointing to any single factor, some general patterns can be observed. The aggregation propensities for configurations A, B, and C (which are compact, nonaggregation prone) seem generally to increase with increasing protein concentration at $|\epsilon^*| = 4$ (Fig. 8). This is because a protein chain is more likely to come in contact with another chain at high protein concentrations. As $|\epsilon^*|$ is increased, the nearly monotonic pattern observed at $|\epsilon^*| = 4$ begins to fluctuate. Configurations D and E are among the more aggregation-prone species on Figures 8–10; this is to be expected, because they are observed only in a multichain environment and are primarily stabilized by interchain associations. Configurations F through I have high propensities for aggregation because they are very open configurations and have at least four exposed H beads each.

Examining the difference between the aggregation propensities of compact and open configurations reveals an interesting pattern. At $|\epsilon^*| = 4$ (Fig. 8), there is no significant difference between the aggregation propensities of compact configurations (A, B, and C) and open, aggregation-prone configurations (D through I). This difference becomes considerable, however, at $|\epsilon^*| = 6$ (Fig. 9), and then becomes less significant again at $|\epsilon^*| = 8$ (Fig. 10). We believe that this variation arises from a fundamental difference

Fig. 8. Aggregation propensities of various configurations at $|\epsilon^*| = 4$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

between the forces that drive aggregation for compact and for aggregation-prone configurations. Aggregation-prone configurations welcome aggregation because the interchain HH contacts that are made contribute substantially to the configuration's stability. Compact configurations (e.g., configurations A, B, and C) neither welcome nor discourage aggregate formation because they are structurally so close to the global free energy minimum (the native state) that aggregate formation does not offer them much extra stability. This means that compact configurations tend to disengage

Fig. 9. Aggregation propensities of various configurations at $|\epsilon^*| = 6$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

Fig. 10. Aggregation propensities of various configurations at $|\epsilon^*| = 8$. The y axis represents the packing fraction, ρ , whereas the z axis represents probability.

easily from an aggregate unless they are trapped by excluded volume restrictions. At $|\epsilon^*| = 4$, the aggregation propensity of compact configurations is high because the aggregation tendency associated with the increased availability of aggregation partners outweighs the disengagement tendency associated with weaker HH binding. At $|\epsilon^*| = 8$, the aggregation propensity of compact configurations is high because the difficulty of disengaging from an aggregate outweighs the decreased likelihood of finding aggregation partners with exposed H beads. At $|\epsilon^*| = 6$, all of these effects tend to cancel each other, yielding relatively low aggregation propensities of compact configurations.

From Figures 8, 9, and 10, we can see that almost all of the aggregation propensity profiles display a peak at a packing fraction of 0.4 (fourth bar from the front for every configuration on the bar charts). This is true for every configuration at every value of denaturant concentration shown in the three figures. Because the peak is observed at a packing fraction of 0.4, regardless of the value of $|\epsilon^*|$, it suggests that this behavior arises out of the packing effect and is not related to the hydrophobic interaction.

A packing fraction of 0.4 falls in the threshold region where the model protein solution crosses over from the dilute to the semidilute regime (de Gennes, 1979). Any polymer solution is said to be in the dilute regime when the polymer chains are separate. As the polymer concentration in the solution is increased, the chains start to interpenetrate and the polymer solution is said to be in the semi-dilute regime. Clearly, this is not a sharp threshold, but rather a region where this crossover happens. To calculate the packing fraction at which the model protein solution crosses over from the dilute to the semidilute regime, we use the fact that the average end-to-end distance for an isolated 20-bead chain, D_e , is 7.7 (Gupta & Hall, 1995). This implies that the "volume" occupied by the isolated chain in two dimensions is roughly equal to $(\pi D_e^2)/(4) = 46.566$. For the 20-bead chain on a lattice, the highest packing fraction that would allow a "volume" of 46.566 per chain is $20/46.566 = 0.429$. Therefore, at around a packing fraction of 0.429, the protein chains begin to interpenetrate, which gives rise to a

sharp increase in aggregation propensity, regardless of the ϵ value. This does not, however, explain why the aggregation propensity drops as the packing fraction is increased further.

Refolding yield

In this section we describe how the refolding yield and fractional aggregation are affected by the protein and denaturant concentrations. The refolding yield is defined to be the ratio of the number of correctly folded chains to the total number of chains in the simulation box, and the fractional aggregation is defined to be the ratio of the number of chains participating in aggregates to the total number of chains in the simulation box. Figure 11 presents the refolding yield and fractional aggregation as a function of the protein and denaturant concentrations. (Note that these two values do not sum to one at any condition because unfolded, nonaggregated chains may also be present in the system.) The refolding yield is high at low values of packing fraction (protein concentration) and goes down sharply as the packing fraction is increased; the fractional aggregation is low at low protein concentration, but increases rapidly as the protein concentration is increased. At higher protein concentrations, the chains are in close physical proximity to each other and it is easier for a partially folded intermediate to form HH contacts with other partially folded intermediates, thus aggregating. This type of behavior is in general agreement with experimental observations (Zettlmeissl et al., 1979) that the protein refolding yield is higher in dilute solutions than in concentrated solutions. For example, in their studies of the aggregation versus refolding of porcine muscle lactate dehydrogenase, Zettlmeissl et al. found that the refolding yield drops sharply and the aggregate formation increases concomitantly as the initial protein concentration is increased.

Of the three denaturant concentrations shown in earlier results, the highest refolding yield and, hence, the lowest aggregation in Figure 11 is observed at the intermediate value of denaturant concentration, $|\epsilon^*| = 6.0$, followed by 8.0 and 4.0, respectively. We believe that this effect can be explained by appealing to the argu-

Fig. 11. Refolding yield and fractional aggregation vs. protein concentration for the 20-bead chain. At low protein concentration, the refolding yield is high but drops sharply as the protein concentration is increased.

ments pertaining to the competition between intra- and intermolecular attractions presented in the Folding intermediates and aggregation section. At low values of $|\epsilon^*|$, the chains are more likely to form open configurations (due to weak intramolecular attractions) and, hence, are more likely to form aggregates. At high values of $|\epsilon^*|$, the chains are less likely to form open configurations (due to strong intramolecular attractions), but those that do form are very stable due to the strong intermolecular attractions. It turns out that an intermediate value of $|\epsilon^*|$ provides the optimum condition where the refolding yield is maximized.

"Critical" solution conditions at which the refolding yield is a maximum have been observed experimentally (Jaenicke, 1974, 1987; London et al., 1974; Jaenicke & Rudolph, 1980; Goldberg, 1982). Jaenicke and coworkers studied the refolding of lactate dehydrogenase (Jaenicke, 1974, 1987; Jaenicke & Rudolph, 1980) and found that the refolding yield was sensitive to the solution pH and went through a maximum at a pH value of 8. Because the solution pH affects hydrophobic interaction, a variation in pH is equivalent to a variation in $|\epsilon^*|$ in our simple model. Therefore, the observed maximum in refolding yield is qualitatively similar to our results. One reason why more experiments do not show a maximum in their refolding yield as a function of denaturant concentration could be that the very high denaturant concentration regime is often not explored experimentally.

Critical solution conditions at which the refolding yield is a minimum have also been observed. Goldberg (1982) found that exposing a refolding β -galactosidase solution to 3 M urea causes a very high degree of aggregation, leading to a minimum in refolding yield. London et al. (1974) also report a minimum in the refolding curve at a critical concentration of 3 M urea for tryptophanase. We believe that the experimentally observed minima in the refolding yield for the above mentioned proteins may be caused by the excessive population of some very aggregation-prone intermediates at critical denaturant concentrations. More recently, De Young et al. (1993b) have observed that apomyoglobin exhibits a minimum in solubility at 2.4 M urea. Our simulations do not indicate any solution conditions where the refolding yield is minimized; this may be a limitation of the simple model that we have employed. Interestingly enough, a mean-field model of Stigter and Dill (1993) in which aggregation is assumed to arise from the association of random coil states, rather than from the folding intermediates observed here, is able to account for this phenomenon.

Another interesting feature evident from Figure 11 is that the refolding yield (which generally decreases as the protein concentration increases) seems to increase slightly at low to moderate protein concentration before it goes down sharply. Thus, there appears to be an optimum protein concentration that maximizes the refolding yield. This peak in the refolding yield is observed at $|\epsilon^*| = 6$ and 8, but not at $|\epsilon^*| = 4$. The sharp drop in the refolding yield with increasing protein concentration can be explained by the fact that when the protein chains are packed tightly together, it is easier for them to form interchain HH contacts and aggregate. The slight increase in the refolding yield with protein concentration that precedes this can be explained by the fact that the simulation is started with a perfectly denaturing solvent ($\epsilon = 0.0$). This means that, at higher protein concentrations, the chains are already in somewhat more compact states than they would have been in had they been isolated. When the HH interaction is "turned on," it is easier for the chains to form intrachain HH contacts. This facilitates formation of compact configurations that serves to help the chain along its folding pathway, thus affording less chance for the

aggregation reaction to occur. At $|\epsilon^*| = 4.0$, however, the protein chains undergo folding-unfolding transitions continuously, and by unfolding repeatedly the chains lose the advantage of having started from a compact state. Thus, the refolding curve for $|\epsilon^*| = 4$ shows a monotonic drop in refolding yield with increasing protein concentration.

Discussion

We have employed a simple lattice model to study the competition between refolding and aggregation in multiprotein systems. The protein model is admittedly limited in that we only look at two-dimensional chains on a square lattice, and that solvent molecules are not explicitly included. However, this model captures what is arguably the most important feature of folding phenomena, namely the hydrophobic interaction.

Our model multichain simulations exhibit the competition between refolding and aggregation that one would expect for real proteins. The most-populated partially folded configurations range from compact (nonaggregation-prone) intermediates to open (highly aggregation-prone) intermediates. The aggregation-prone intermediates are stabilized by interchain interactions that hamper the progress toward the native state.

Aggregation seems to arise from association among partially folded intermediates, not random coil states. The aggregating species are found to contain a substantial number of native HH contacts. In addition, the nine most populated species exhibit substantial amounts of native-like structure.

The following general rules describe how changes in denaturant and protein concentration affect the refolding and aggregation process: (1) low denaturant concentration (high $|\epsilon^*|$) discourages the formation of open (noncompact) configurations because the strong intrachain HH interaction energy causes the chains to collapse upon themselves very quickly and, at the same time, preserves open configurations because the strong interchain HH interaction energy causes the open configurations to participate in very stable aggregates. (2) High denaturant concentration encourages aggregation because the weak intrachain HH attractions allow the protein chains to unfold out of the native state, thereby increasing the availability of partially folded intermediates that can serve as aggregation partners and, at the same time, reduces the stability of aggregates, again because of the weak interchain HH attraction. (3) High protein concentration discourages the formation of open configurations because the excluded volume constraints force the chains to pack more tightly and, at the same time, preserves open configurations because the greater opportunity for interchain HH contacts results in the formation of very stable aggregates. (d) Kinetic traps become progressively "deeper" as denaturant concentration is lowered, resulting in higher populations of kinetic trap configurations. (4) Kinetic trap configurations that also have exposed H beads are observed to be highly aggregation prone.

Multiple-chain systems introduce intermediates that are not observed along the folding pathmaps of isolated protein chains. These intermediates are stabilized mostly by interchain interactions, and hence, exhibit high aggregation propensities.

The aggregation reaction is generally seen to be more dominant at high protein concentration and low denaturant concentration. However, the simulations indicate that there are some optimum conditions at intermediate denaturant and protein concentrations that result in the highest refolding yield (lowest fractional aggregation) in agreement with some experimental observations. Al-

though the exact parameters required to obtain the optimum refolding yield are expected to be strongly dependent upon the protein sequence, we believe that the very existence of such optimal refolding conditions is a characteristic of the refolding and aggregation phenomena itself.

Acknowledgments

The authors are grateful to Professors Ken Dill and Jonathan King, and to Dr. Jeffrey Cleland for the guidance that they have provided. This work was sponsored by the National Science Foundation (Grant No. CTS 9208590 and Grant No. CTS 9704044), and by the North Carolina Biotechnology Center (Grant No. 9113-ARG-0205).

References

- Ahern TJ, Manning MC. 1992. *Stability of protein pharmaceuticals*. New York: Plenum Press.
- Baldwin RL. 1975. Intermediates in protein folding reactions and the mechanism of protein folding. *Annu Rev Biochem* 44:453–475.
- Clark J, Steele J. 1992. Phase separation inhibitors and prevention of selenite cataract. *Proc Natl Acad Sci USA* 89:1720–1724.
- Costantino HR, Langer R, Klibanov AM. 1995. Aggregation of lyophilized pharmaceutical protein, recombinant human albumin: Effect of moisture and stabilization by experiments. *Biotechnology* 13:493–496.
- de Gennes P-G. 1979. *Scaling concepts in polymer physics*. Ithaca, New York: Cornell University Press.
- De Young LR, Dill KA, Fink AL. 1993b. Aggregation and denaturation of apomyoglobin in aqueous urea solutions. *Biochemistry* 32:3877–3886.
- De Young LR, Fink AL, Dill KA. 1993a. Aggregation of globular proteins. *Acc Chem Res* 26:614–620.
- Dill KA. 1990. Dominant forces in protein folding. *Biochemistry* 29:7133–7155.
- Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding—A perspective from simple exact models. *Protein Sci* 4:561–602.
- Dill KA, Stigter D. 1993. Theory for protein solubilities. *Fluid Phase Equilib* 82:237–249.
- Dill KA, Stigter D. 1995. Modeling protein stability as heteropolymer collapse. *Adv Protein Chem* 46:59–104.
- Eaton WE, Hofrichter J. 1990. Sick cell hemoglobin polymerization. *Adv Protein Chem* 40:63–279.
- Fields GB, Alonso DOV, Stigter D, Dill KA. 1992. Theory of the aggregation of proteins and copolymers. *J Phys Chem* 96:3974–3981.
- Gallo G, Goni F, Boctor F, Vidal R, Kumar A, Stevens FJ, Frangione B, Ghiso J. 1996. Light chain cardiomyopathy. Structural analysis of the light chain tissue deposits. *Am J Pathol* 148:1397–1406.
- Georgiou G, Bowden GA. 1990. Inclusion body formation and the recovery of aggregated recombinant protein. In: Prokop A, Bajpas RK, Ho C, eds. *Recombinant DNA technology and applications*. New York: McGraw Hill. pp 333–356.
- Gierasch LM, King J, eds. 1990. *Protein folding: Deciphering the second half of the genetic code*. Washington, DC: American Association for the Advancement of Science.
- Goldberg ME, Zetina CR. 1980. Importance of inter-domain interactions in the structure, function and stability of the F_1 and F_2 domains isolated from the β_2 subunit of *E. coli* tryptophan synthase. In: Jaenicke R, ed. *Protein folding*. Amsterdam: Elsevier North Holland Biomedical Press. pp 469–484.
- Goldberg ME. 1982. *Dynamic aspects of conformation changes in biological macromolecules*. Dordrecht, The Netherlands: Reidl Publications. pp 57–65.
- Gupta P, Hall CK. 1995. Computer simulation of protein refolding pathways and intermediates. *AIChE J* 41:985–990.
- Gupta P, Hall CK. 1997. Effect of solvent conditions upon refolding pathways and intermediates for a simple lattice protein. *Biopolymers* 42:399–409.
- Jaenicke R. 1974. Reassociation and reactivation of lactic dehydrogenase from the unfolded subunits. *Eur J Biochem* 46:149–155.
- Jaenicke R, Rudolph R. 1980. Folding and association of oligomeric enzymes. In: Jaenicke R, ed. *Protein folding*. Amsterdam: Elsevier North Holland Biomedical Press. pp 525–546.
- Jaenicke R. 1987. Folding and association of proteins. *Prog Biophys Mol Biol* 49:117–237.
- Kane JF, Hartley DL. 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends Biotechnol* 6:95–101.
- Kelly JW. 1996. Alternative conformations of amyloidogenic proteins govern their behavior. *Curr Opin Struct Biol* 6:11–17.
- Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem* 51:459–489.
- King J. 1989. Deciphering the rules of protein folding. *C & EN April* 10:32–54.
- Lau KF, Dill KA. 1989. A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 22:3986–3997.
- London J, Skrzynia C, Goldberg M. 1974. Renaturation of *E. coli* tryptophanase after exposure to 8 M urea. Evidence for the existence of nucleation centers. *Eur J Biochem* 47:409–415.
- Marston FAO. 1986. The purification of eukaryotic polypeptides synthesized in *E. coli*. *Biochem J* 240:1–12.
- Massry S, Glasscock R. 1983. *Textbook of nephrology*. Baltimore, Maryland: Williams & Wilkins.
- Mitraki A, King J. 1989. Protein folding intermediates and inclusion body formation. *Biotechnology* 7:690–697.
- Moore RC, Melton DW. 1997. Transgenic analysis of prion disease. *Mol Hum Reprod* 3:529–544.
- Persson KM, Gekas V. 1994. Factors influencing aggregation of macromolecules in solution. *Process Biochem* 29:89–98.
- Sadana A. 1995. Protein refolding and inactivation during bioseparation: Bioprocessing implications. *Biotechnol Bioeng* 48:481–489.
- Schmid FX. 1992. Kinetics of unfolding and refolding of single-domain proteins. In: Creighton TE, ed. *Protein folding*. New York: W.H. Freeman & Co. pp 197–241.
- Selkoe DJ. 1995. Deciphering Alzheimer's disease: Molecular genetics and cell biology yield major clues. *J NIH Res* 7:57–64.
- Simmons LK, May PC, Tomoselli KJ, Rydel RE, Fuson KS, Brigham BF, Wright S, Lieberburg I, Becker GW, Brems DN, Li WY. 1994. Secondary structure of amyloid β peptide correlates with neurotoxicity *in vitro*. *Mol Pharmacol* 45:373–379.
- Sipe JC. 1992. Amyloidosis. *Annu Rev Biochem* 61:947–975.
- Stigter D, Dill KA. 1993. Theory for protein solubilities. *Fluid Phase Equilib* 82:237–249.
- Wetzel R. 1996. For protein misassembly, it's the "I" decade. *Cell* 86:699–702.
- Zettlmeissl G, Rudolph R, Jaenicke R. 1979. Reconstitution of lactic dehydrogenase. Non-covalent aggregation vs reactivation. Physical properties and kinetics of aggregation. *Biochemistry* 18:5567–5571.