Simulations of Reversible Protein Aggregate and Crystal Structure

Sugunakar Y. Patro and Todd M. Przybycien
Applied Protein Biophysics Laboratory, Howard P. Isermann Department of Chemical Engineering, Rensselaer Polytechnic Institute, Troy, New York 12180-3590 USA

ABSTRACT We simulated the structure of reversible protein aggregates as a function of protein surface characteristics, protein-protein interaction energies, and the entropic penalty accompanying the immobilization of protein in a solid phase. These simulations represent an extension of our previous work on kinetically irreversible protein aggregate structure and are based on an explicit accounting of the specific protein-protein interactions that occur within reversible aggregates and crystals. We considered protein monomers with a mixture of hydrophobic and hydrophilic surface regions suspended in a polar solvent; the energetic driving force for aggregation is provided by the burial of solvent-exposed hydrophobic surface area. We analyzed the physical properties of the generated aggregates, including density, protein-protein contact distributions, solvent accessible surface area, porosity, and order, and compared our results with the protein crystallization literature as well as with the kinetically irreversible case. The physical properties of reversible aggregates were consonant with those observed for the irreversible aggregates, although in general, reversible aggregates were more stable energetically and were more crystal-like in their order content than their irreversible counterparts. The reversible aggregates were less dense than the irreversible aggregates, indicating that the increased energetic stability is derived primarily from the optimality rather than the density of the packing in the solid phase. The extent of hydrophobic protein-protein contacts and solvent-exposed surface area within the aggregate phase depended on the aggregation pathway: reversible aggregates tended to have a greater proportion of hydrophobic-hydrophobic contacts and a smaller fraction of hydrophobic solvent-exposed surface area. Furthermore, the arrangement of hydrophobic patches on the protein suface played a major role in the distribution of protein contacts and solvent content. This was readily reflected in the order of the aggregates: the greater the contiguity of the hydrophobic patches on the monomer surface, the less ordered the aggregates became, despite the opportunities for rearrangement offered by a reversible pathway. These simulations have enhanced our understanding of the impact of protein structural motifs on aggregate properties and on the demarcation between aggregation and crystallization.

INTRODUCTION

Protein aggregates are ubiquitous in a multitude of biological, analytical, and processing operations. Although aggregate formation may be desirable in situations such as inclusion body formation because of the ease of recovery (Kane and Hartley, 1988) and protection against proteolysis (Hellebust et al., 1989), it is strongly undesirable in processes such as protein crystallization, in vitro folding, and formulation and delivery operations, because of the bioactivity losses (Becker et al., 1987), immunogenic implications (Lewis et al., 1969a,b; Moore and Leppert, 1980), altered half-lives, and solubility and diffusional limitations (Cleland et al., 1993) associated with aggregates. The success of industrial protein-processing operations and analytical characterization efforts depends on the ability to influence aggregate formation and redissolution behavior which, in turn, depends on an understanding of aggregate structure. Because specific sites on protein surfaces are observed to mediate self-association (Brems, 1988; Brems et al., 1992; Casal et al., 1988; Goldenberg et al., 1983; Mitraki et al., 1991; Przybycien and Bailey, 1989; Rinas et al., 1992; Wetzel et al., 1991), studies aimed at understanding protein aggregates should account for the physicochemical nature of the protein surface and the role of specific protein surface patches and solvent in the formation of intermolecular interactions.

Although protein structure simulations provide a powerful means of relating monomer surface characteristics to aggregate properties, few earlier simulation studies have considered the impact of specific interaction sites. We have used a lattice-based Monte Carlo algorithm to simulate the ultimate structure of protein aggregates, taking into account specific hydrophobic and hydrophilic sites on the model protein surface. We assessed the following aggregate properties: density, porosity, distribution of intermonomer contacts, nature and extent of surface area exposed to solvent, and order.

In our previous work (Patro and Przybycien, 1994), we examined the properties of aggregates generated under kinetically irreversible conditions, mimicking protein precipitation and inclusion body formation; a summary of the salient results follows. The distribution of hydrophobic and hydrophilic sites and their orientation on the monomer surface had a significant impact on the final structure of simulated aggregates. An increase in the extent of monomer hydrophobic surface area resulted in aggregates with higher densities and lower system free energies. In addition, these effects were accompanied by increases in the number of hydrophobic-hydrophobic contacts per monomer aggre-

Received for publication 3 July 1995 and in final form 29 February 1996.
Address reprint requests to Dr. Todd M. Przybycien, Bioseparations Research Center, Howard P. Isermann Department of Chemical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180-3590. Tel.: 518-276-2793; Fax: 518-276-4030; E-mail: przybt@rpi.edu.

0006-3495/96/06/2888/15 $2.00

gated, and decreases in the solvent-exposed hydrophobic surface area of the aggregates. For a given extent of monomer hydrophobic surface area, grouping monomer hydrophobic surfaces in a single contiguous stretch resulted in lower aggregate densities and lower short-range order in aggregates. Although more favorable hydrophobic-hydrophobic contact energies produced aggregate structures with higher densities, the number of unfavorable protein-protein contacts increased as well; increasing the magnitude of the entropic penalty accompanying the immobilization of monomers in the aggregate phase produced the opposite effect.

Our present work represents an extension of this model to aggregates formed by reversible pathways; if the displacement from equilibrium is small, protein aggregation may proceed under thermodynamically reversible conditions (De Young et al., 1993a). Because a reversible aggregation process allows a multitude of aggregate structure states to be sampled, we expected aggregates generated under these conditions to exhibit greater order than those generated under irreversible conditions. In our assessment of the physical properties of the reversible aggregates generated, we made comparisons to both the irreversible case and to the protein crystallization literature. Although the model itself is not representative of protein crystallization, comparison of the simulation results to the protein crystallization literature enabled us to relate our results to physically observed phenomena. In addition, such a comparison has provided a unique perspective of the influence of protein molecular architecture on the demarcation between aggregates and crystals.

MODEL DESCRIPTION AND EXECUTION

A complete description of the lattice model and simulation mechanics was presented in our earlier work (Patro and Przybycien, 1994); the salient features of the model and modifications for the reversible case are summarized in this section. We abstracted protein molecules as hexagons, with different sides representing distinguishable hydrophobic and hydrophilic patches. Fig. 1 shows our conceptualization of an aggregating protein system after perturbation by an aggregation-inducing agent. The resulting closed, constant-volume system can be subdivided into two regions: the solution region, with negligible protein-protein (p,p) contacts, and the solvated aggregate region, shown as a lattice in the figure. The objective of the simulation is to generate this lattice-based aggregate region, which evolves as the simulation proceeds; the structural configuration of the lattice at the end of the simulation is akin to a two-dimensional cross section of a protein aggregate.

The simulation algorithm, in brief, is as follows. At the beginning of the simulation, the aggregate region is completely filled with the solvent. A lattice site is chosen at random and an attempt is made to transfer a protein monomer with a randomly selected rotational orientation from the solution phase to the aggregate phase. This transfer is accompanied by the simultaneous back-transfer of an equivalent volume of solvent from the aggregate phase to the solution phase. The success of such a transfer attempt is governed by the associated system free energy change, $\Delta G_{sys(fwd)}$, for the overall transfer process, which is given as the sum of 1) the energetic contribution resulting from the formation and breakup of protein-protein (p,p), protein-solvent (p,s), and solvent-solvent (s,s) contacts in the system and 2) the configurational free energy penalty, ΔG_{config}, that accounts for the reduction in the entropy of the monomer upon confinement to the aggregate phase. $\Delta G_{sys(fwd)}$ for a single protein monomer transfer from the solution phase to any solvent-occupied lattice position surrounded by n previously transferred protein monomers is given by (Patro and Przybycien, 1994)

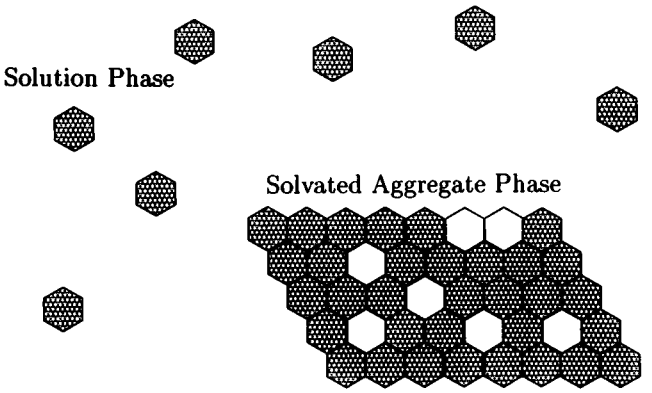

FIGURE 1 Solution and aggregate phases of the system. Dispersed protein monomers are shown to indicate the solution phase, and the solvated aggregate phase is shown in the form of a lattice. The filled hexagons in the lattice represent protein monomers within the aggregate, and the unfilled hexagons represent the entrapped solvent.

$$\Delta G_{sys(fwd)} = \sum_{i=1}^{n} \Delta\Delta G(p,p)_i + \Delta G_{config} = \Delta G_{ppt} + \Delta G_{sol} + \Delta G_{config} = -\Delta G_{sys(rev)}, \quad (1)$$

where $\Delta\Delta G(p,p)_i$ is the free energy change for the formation of the ith (p,p) interaction in solution. Note that the $\Sigma_{i=1}\Delta\Delta G(p,p)_i$ term is context dependent and can be further parsed into ΔG_{ppt} and ΔG_{sol}, the changes in the free energies of the precipitate and solution phases, respectively. Expressions for ΔG_{ppt} and ΔG_{sol} that explicitly account for the formation and disruption of the various intermolecular interactions during the transfer process were given previously (Patro and Przybycien, 1994).

To simulate protein aggregates formed under thermodynamically reversible conditions, we modified our original model to allow for microscopic reversibility. In the event a protein-occupied lattice site is chosen during the selection process, a reverse transfer is attempted. The free energy

change for the reverse transfer, $\Delta G_{sys(rev)}$, is just the negative of that for the forward direction as indicated in Eq. 1.

For every forward or reverse transfer attempt, ΔG_{sys} is computed. If $\Delta G_{sys} \leq 0$, the transfer attempt is always successful. However, if $\Delta G_{sys} > 0$, we use Metropolis sampling to determine whether the transfer or reverse transfer is allowed (Patro and Przybycien, 1994; Pattou et al., 1991). ΔG_{sys} is a function of the extent of aggregation, $\xi(\tau)$, which in turn is a function of the number of transfer attempts or the Monte-Carlo time, τ. The total system free energy change, ΔG^t_{sys}, is given by the deterministic sum

$$\Delta G^t_{sys}(\xi) = \sum_{\tau=0}^{\tau} \Delta G_{sys}(\xi(\tau)). \quad (2)$$

Note that the extent of aggregation is not vectorial: microscopic reversibility implies that $\xi(\tau)$ is not monotonic in τ. This allows for an extensive sampling of aggregate configuration space.

Monomer addition or removal occurs without regard to any presumptions about whether the target lattice site is internal or external to the aggregate. Such an unrestricted monomer addition process is a statistical device we use to construct the final aggregate structure; this is similar to the statistical devices used to explore vapor-liquid equilibria (Panagiotopoulos, 1987). Monomer addition and removal in the "interior" of the lattice may be thought of as a functional equivalent of internal rearrangement and aging.

The simulation is terminated when the total system free energy is minimized. This corresponds to the equilibrium condition and is given by

$$\left.\frac{dG^t_{sys}}{d\xi}\right|_{eqm} = \left.\frac{d\Delta G^t_{sys}(\xi)}{d\xi}\right|_{eqm} = 0. \quad (3)$$

At equilibrium, the aggregate structure is complete and the fraction of the lattice filled at this point in the simulation is termed the "equilibrium density." The reversible simulation process results in a final state that is very close to the global free energy minimum, whereas a kinetically irreversible simulation run results in a final state that is the minimum along a given irreversible pathway. The fraction of the lattice filled at the end of an irreversible simulation run is termed the "jamming limit." Recognition of the minimum free energy state is achieved in hindsight. We run the simulations until a total of 300 million transfer attempts are made. Within this period, the system was empirically verified to have achieved the lowest free energy state accessible in finite time. At this point, we retrace our steps in Monte Carlo time along the aggregation pathway to regenerate the aggregate lattice state corresponding to the minimum ΔG^t_{sys}. As a further verification of the attainment of equilibrium (Durbin and Feher, 1991), we compute the forward and backward rates of transfer. The equilibrium densities had standard deviations ranging between 0.01% and 0.12% of the mean for a window of 90,000 time points about the equilibrium time, indicating that the transfer rates were essentially equivalent. For this calculation, a nine-point quadratic smoothing of equilibrium densities was done as described by Savitzky and Golay (1967; Steinier et al., 1972).

We chose seven different model monomer types, as shown in Fig. 2, to study the influence of both the extent and distribution of hydrophobic (HΦ) and hydrophilic (HF) sites on the protein surface. Only HΦ and HF sites were considered in this simplified model; no electrostatic interactions were considered, although these interactions may also play an important role in self-association (Cudney et al., 1994; Takahashi et al., 1993). This is similar in spirit to the HP chain models used to capture the essence of intramolecular self-interaction effects on protein folding (Shortle et al., 1992; Yue and Dill, 1992). The free energy changes associated with the formation of (p,p) interactions in solution are computed from the corresponding interaction energies in vacuo; these interaction energies are listed in Table 1. We estimated ΔG_{config} to be about $+1\ RT$ kcal/mol and assumed it to be invariant for each monomer transfer. A comprehensive discussion of these energy estimates appeared in our previous work (Patro and Przybycien, 1994).

Although the dimensionality of the lattice may impact phase transition behavior, a two-dimensional model suffices to predict qualitative trends. The size of the lattice restricts the size of the largest possible unique aggregate structure; therefore, a 64×64 lattice was chosen to minimize size effects while maintaining reasonable computing times. Periodic boundary conditions were used to minimize edge effects. The extensive sampling of aggregate structure states during the course of a reversible simulation requires about two orders of magnitude more CPU time than the corresponding kinetically irreversible simulation. The reversible simulations were therefore executed in triplicate rather than in groups of 500, as done for the irreversible case. Run-to-run variations were found to be negligible. The code for the simulations was written in FORTRAN, and simulations were run on a UNIX-based IBM RISC System/6000. A supplementary random number generation routine was written in C and interfaced with the rest of the code.

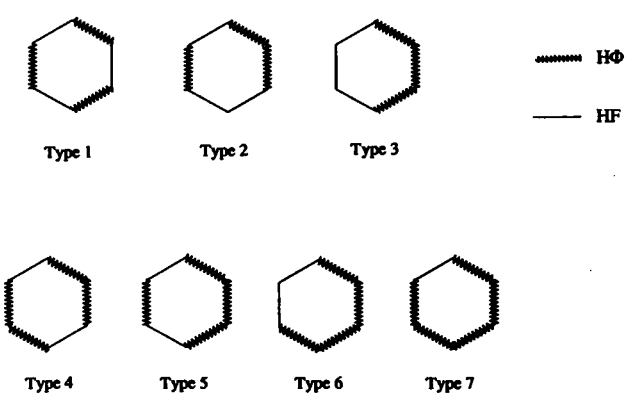

FIGURE 2 Monomer types used in simulations.

TABLE 1 Interaction free energies in vacuo and in solution

	$\Delta G(x, y)$* (kcal/mol)			$\Delta\Delta G(p, p)^{\#}$ (kcal/mol)	
	HΦ	HF	S	HΦ	HF
HΦ	+15.5			−4.5	
HF	+13.5	−0.6		+7.0	+6.4
S	+7.0	−6.5	−6.0	—	—

*$\Delta G(x,y)$ corresponds to (x,y) contact formation in vacuo.
#$\Delta\Delta G$(p,p) corresponds to (p,p) contact formation in solution; $\Delta\Delta G(p,p') = \Delta G(p,p') + \Delta G(s,s) - (\Delta G(p,s) + \Delta G(p',s))$.

RESULTS AND DISCUSSION

The aggregate properties we assessed included density, porosity, protein-protein contacts within the aggregate phase, nature, and extent of solvent-exposed surface area of the agggregates and long- and short-range order of the aggregate phase. Because the reversible self-association process adopted for the simulations is more akin to a crystallization process, in that the system is allowed to sample various possible structure states along the aggregation pathway in search of the minimum free energy state, we anticipated that reversible aggregates would exhibit greater order than irreversible aggregates. Concurring with our predictions, the structures generated reversibly attained much lower system free energies and were more crystal-like compared to those generated irreversibly for all the model protein monomer types studied. Therefore, in comparing the reversible and irreversible cases, we interpreted aggregate properties in terms of crystals and precipitates, respectively. That crystallization and precipitation are related is given by the experimental finding that the formation of protein crystals is sometimes preceded by the formation of precipitates, which slowly dissolve in favor of crystals (McPherson, 1990).

In the midst of our comparisons, we noted a subtle error in our previous simulations for the irreversible case. Although this error does not affect any of the conclusions drawn in the earlier work, it does have an impact on the magnitude of several of the numerical results. All comparisons made in the present work are to corrected results. A more complete description of the nature of the error and a summary of corrections to results not explicitly included in comparisons here are given in the Appendix.

System energetics

Fig. 3 shows in schematic form the typical trajectory for ΔG^t_{sys} as a function of τ for the reversible simulation pathway. Initially, upon perturbation by a solubility-reducing agent, the system is in a nonequilibrium state. We have identified four distinct phases during the simulation approach to a new equilibrium state: 1) nucleation, 2) reversible growth, 3) relaxation, and 4) fluctuation. Fig. 4 illustrates the state of the aggregate lattice at an instant in Monte Carlo time during each of these phases for the aggregation of type 1 monomers.

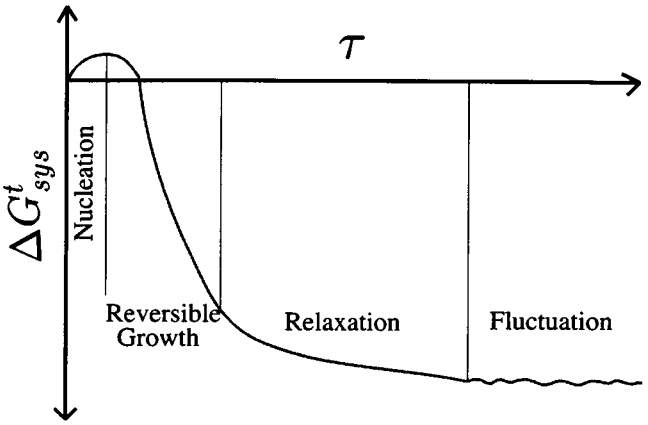

FIGURE 3 The system free energy profile during the four phases of the simulations.

During the nucleation phase, the aggregate region consists primarily of isolated protein monomers, and ΔG^t_{sys} is positive because of the loss in configurational entropy. Aggregation is driven by the Metropolis sampling process in this phase of the simulation, as each isolated transfer is energetically unfavorable. The positive nature of ΔG^t_{sys} in the first phase is observed in all nucleation-controlled processes, such as protein crystallization (De Young et al., 1993b; Kam et al., 1978; Mikol et al., 1990). McPherson (1990) explained this initial increase in system free energy as being essential to the creation of a solid phase with stable nuclei and compared this energy barrier with the activation energy of a chemical reaction. During this phase of the simulation, each protein monomer can potentially serve as a nucleation site for further aggregation; however, because of the incorporation of microscopic reversibility, some of these monomers redissolve into the solution phase and only the remaining monomers can act as stable nucleii to facilitate further monomer addition. When the aggregating phase contains a sufficient number of these stable nuclei, the system reaches a critical state beyond which monomer addition is favorable, leading to a decrease in the system free energy. The attainment of this critical state, denoted by a maximum in ΔG^t_{sys}, constitutes the end of the nucleation phase.

The phase immediately after the nucleation phase is the reversible growth phase. Monomer addition is almost always energetically favorable during this phase. In this phase, the aggregate phase experiences rapid addition of monomers with simultaneous rearrangement. This rearrangement, however, is overwhelmed by monomer addition, leading to a steep increase in the equilibrium density and a steep decrease in ΔG^t_{sys} as more and more favorable (p,p) contacts are formed.

As reversible growth continues, the aggregate phase reaches a state where the rate of monomer addition and rearrangement are comparable; the aggregate must undergo significant rearrangement to allow any further net addition of monomers. Such a change in the system behavior indi-

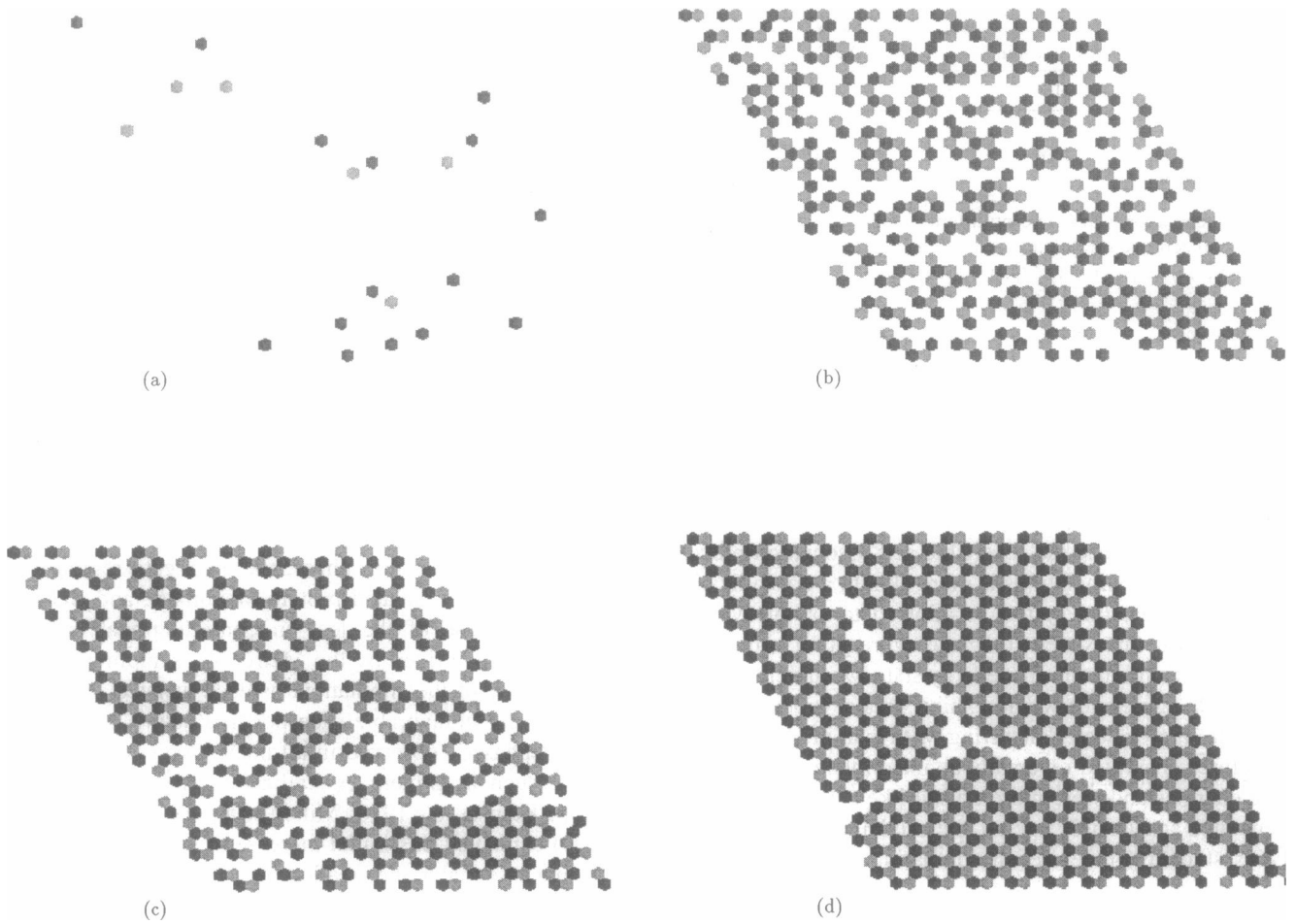

FIGURE 4 Sample aggregate structures in the (*a*) nucleation, (*b*) reversible growth, (*c*) relaxation, and (*d*) fluctuation phases for type 1 monomers. The two shades represent the two unique rotational orientations of monomer type 1.

cates the onset of the relaxation phase and can be seen as a change in the slope of the free energy profile in the schematic. During this phase, the organized and unorganized regions of aggregate structure that are out of register with their neighbors are rearranged to facilitate the further addition of monomers in a more ordered fashion. This part of the growth is analogous to the aging or ripening of aggregates in that less viable structures are replaced by more favorable structures (Kabalnov and Shchukin, 1992). It should be noted that the reversible growth and relaxation phases together constitute the generic growth process, and our distinction is based merely on the rate of decrease in ΔG^{t}_{sys} during the simulation.

At the end of the relaxation phase, the system is close to its minimum energy state. Any further decrease in ΔG^{t}_{sys} can be achieved only at the expense of significant rearrangement. The aggregate phase primarily undergoes rearrangement with negligible net addition of monomers. Both ΔG^{t}_{sys} and the equilibrium density fluctuate within a narrow range because of the Metropolis sampling algorithm. In principle, this phase can extend indefinitely. However, we chose the final aggregate configuration as that which corresponds to the minimum ΔG^{t}_{sys} between the onset of this phase and 300 million time steps. The equilibrium density was determined by computing the fraction of the lattice that is filled when the system reaches its minimum free energy state. The fluctuation phase is the longest phase in our simulation, and the extent of system rearrangement during this phase is primarily dependent on the monomer type.

A comparison of our reversible and irreversible simulation results indicates that equilibrium aggregates at the end of the relaxation phase are more ordered or more crystallike than their irreversible counterparts in all respects. Reversible aggregation pathways are inherently more conducive to structures with greater order.

Phase times and their monomer dependence

Fig. 5 shows the duration of each phase in terms of the number of transfer attempts for each of the monomer types studied. Note the scaling of each of the phase times and that nucleation time $\ll$ reversible growth time $<$ relaxation time.

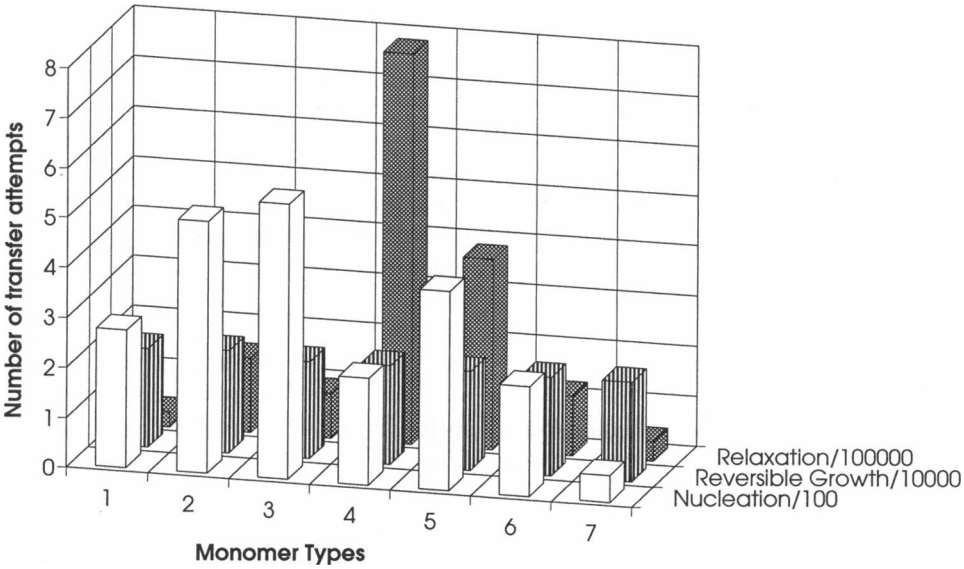

FIGURE 5 Monte Carlo time spent by aggregates of different types in the nucleation, reversible growth, and relaxation phases. The time is given in terms of number of transfer attempts.

The nucleation time is a function of the number of favorable contacting surfaces and their arrangement on the surface. In our simulations, a greater number of favorable contacting surfaces, analogous in a sense to a greater concentration of monomers, reduces the duration of the nucleation phase, the height of the free energy barrier for nucleation, and number of stable nuclei required for growth to proceed. The number of isolated monomers at the end of the nucleation phase were 95, 172, 182, 83, 141, 80, and 21 for types 1–7, respectively, and the corresponding heights of the critical energy barrier for nucleation, normalized by RT, were 87.4, 73.2, 83.2, 45.0, 49.8, 42.0, and 13.4, respectively. In their modeling and experimental studies of lysozyme crystallization, Kam et al. (1978) observed that the height of the free energy barrier to nucleation and the size of the critical nuclei, again analogous to the number of stable nuclei, decrease with increasing protein concentrations. Although this result may seem intuitive because a greater number of favorable contacting surfaces facilitates the formation of (p,p) contacts to overcome the barrier, it has more interesting implications for protein crystallization. A large number of favorable contacting surfaces per monomer is analogous to a highly supersaturated solution; in this situation, the energetic driving force for aggregation is large, leading to a brief nucleation phase but inhibiting the later development of order (McPherson, 1990).

In addition, we observed a surface distribution effect on the nucleation times. Monomer types 1 and 4 have the highest symmetries and consequently have small nucleation times, as there are many possible favorable protein-protein contact combinations compared to other monomer types of equivalent hydrophobicities. Type 3 has all of its hydrophobic patches concentrated in a contiguous stretch such that one successful HΦ-HΦ contact limits options for subsequent HΦ-HΦ contacts. This resulted in a longer nucleation time for monomer type 3 than for monomer types 1 or 2. Monomer type 5 has the longest nucleation time of the type 4, 5, and 6 group, as it too has three HΦ patches in one stretch; however, the additional opposing single patch provides more flexibility than type 3, resulting in a shorter nucleation time for type 5 than for type 3. Monomer types 6 and 7 have short nucleation times despite the contiguity of their hydrophobic surfaces. Four or five contiguous HΦ patches are less restrictive than three, and in these cases, the effect of the extent of hydrophobic surface area outweighs that of the monomer configuration.

The duration of the reversible growth phase is monomer-independent, because this phase is characterized by rapid monomer addition with little back-transfer. However, it should be noted that there is some rearrangement as well, because of the microscopic reversibility, which helps build structures with greater order compared to those generated under kinetically irreversible conditions.

Because the relaxation phase mainly involves rearrangement, the duration of this phase and the onset of the fluctuation phase depend strongly on the monomer configuration and the ease with which the most-ordered aggregate configuration can be attained. Monomer types 1 and 7 have short relaxation times because, although the maximum possible order for these aggregates is high, the repeat unit configurations are compatible with the lattice. On the other hand, the long relaxation phase for monomer type 4 may be attributed to conflicts between the large number of different repeat units possible. Monomer types 3 and 6 have short relaxation times because the inherent order in the repeat units is low, speeding the onset of the fluctuation phase. The magnitude of the fluctuations in ΔG^{t}_{sys} for low-order monomer types is necessarily larger than those for other monomer types of equivalent hydrophobicities. During this extended fluctuation phase, the aggregate structure may undergo significant rearrangements because of conflicts between several favorable arrangements. This is analogous to the behavior of some membrane proteins, the crystal space groups of which evolve during crystallization (Brzozowski and

Tolley, 1994). However, the surface characteristics of monomer types 3 and 6 are not very conducive to generating structures with high order, and they may have several shallow energetic minima close to the global minimum, which may result in prolonged structural perturbations without significant order development. This concurs with the experimental observation that crystallization times for proteins may be as long as several months (McPherson, 1990), although the delay in nucleation may also contribute to lengthy crystallization times.

Equilibrium density

Fig. 6 shows the variation of equilibrium density with the final ΔG^t_{sys} for each monomer type. Jamming limits for each monomer type for kinetically irreversible simulations are shown for comparison as collections of 500 data points. For the kinetically irreversible case, the simulation pathway is deterministic, and as many as 500 runs were required to characterize the statistical distribution of the data.

It should be noted that the extent and distribution of hydrophilic and hydrophobic sites on the protein surface affect aggregate properties under both kinetically controlled and thermodynamically controlled scenarios. Both equilibrium density and jamming limit increase with an increase in the monomer hydrophobic surface area, as predicted in our earlier work (Patro and Przybycien, 1994) and in the aggregation simulations of Fields et al. (1992). In addition, the equilibrium densities for different monomer types are clearly distinguishable in Fig. 6, indicating that the distribution of sites had an impact on equilibrium aggregates to an even greater extent than for irreversible aggregates. For example, monomer types 1, 2, and 3 are equally hydrophobic, yet they have very different equilibrium densities and ultimate ΔG^t_{sys} values. This may have interesting implications for the cocrystallization of proteins. The lens crystallins serve as an example of such a system. The eye lens is composed of α-, β-, and γ-crystallins, which are believed to be evolutionally related, and they must be maintained in a crystalline configuration for normal function; the structural integrity of the lens is jeopardized in the event these proteins aggregate. The usual degradation of the lens with age is believed to be due to the self-aggregation and intercrystallin aggregation of α-crystallins (Liang and Li, 1991). Under such circumstances, a knowledge of how the protein surface characteristics affect the ease of crystallization may be helpful in controlling the cocrystallization process and preventing aggregation-induced disorders.

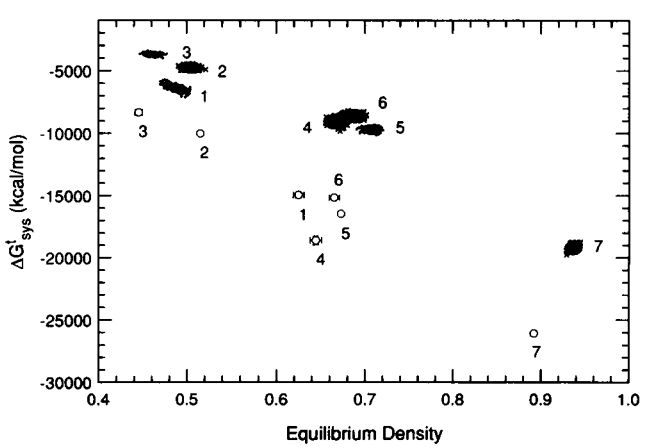

FIGURE 6 Variation of system free energy with equilibrium density or jamming limit as functions of monomer type. The labels correspond to the monomer types given in Fig. 2. The circles represent average values for the reversible simulations with error bars, where visible, representing ± SD of the mean. The patches represent collections of 500 data points for the irreversible simulations.

Another surface distribution effect is manifested by monomer types 3 and 6. These monomer types give aggregates that are energetically less stable than those from other monomer types of equivalent hydrophobicities, for both kinetically and thermodynamically controlled cases. Because monomer types 3 and 6 have contiguous HΦ patches, it can be inferred that such a structural motif is not as conducive to generating an energetically stable and highly ordered structure as the more distributed configurations. Although HΦ-HΦ interactions are treated as favorable interactions in our simulations, a contiguous stretch of hydrophobic patches prevents the effective usage of all hydrophobic surfaces for contacts; once one HΦ-HΦ contact is made, options for subsequent favorable interactions are more limited, leading to less ordered structures. The opportunity for extensive rearrangement afforded by the reversible pathway cannot overcome the steric constraints imposed by the monomer configuration.

The patch distribution effect may also explain the difficulty encountered during the crystallization of membrane proteins that are characterized by extensive hydrophobic exteriors. Large membrane proteins such as transport enzyme complexes are usually easier to crystallize than their smaller homologs, indicating that the difficulty in crystallization is not due to the size of the protein but rather to the more hydrophobic nature of the smaller proteins; the smaller proteins, unlike the larger ones, do not have protruding hydrophilic surfaces (Yoshikawa et al., 1992). Similarly, it is easier to crystallize a membrane protein complex with interspersed hydrophilic sites than a hydrophobic subunit of the complex (Yu et al., 1972). Nonionic surfactants, detergents, or their derivatives are usually employed to overcome the difficulty encountered in crystallizing membrane proteins (Brzozowski and Tolley, 1994; Cudney et al., 1994; Garavito et al., 1986; Garavito and Picot, 1991; Yoshikawa et al., 1992) and in achieving two-dimensional crystals of membrane proteins (Vonck and van Bruggen, 1990; Yeager, 1994). These surfactants are believed to act by binding to the hydrophobic patches on the protein surface and screening them from each other (Bam et al., 1995; Cleland et al., 1993; Sluzky et al., 1992; Tandom and Horowitz, 1989).

Fig. 7 illustrates this patch distribution effect with the simulated equilibrium structures for monomer types 1 and

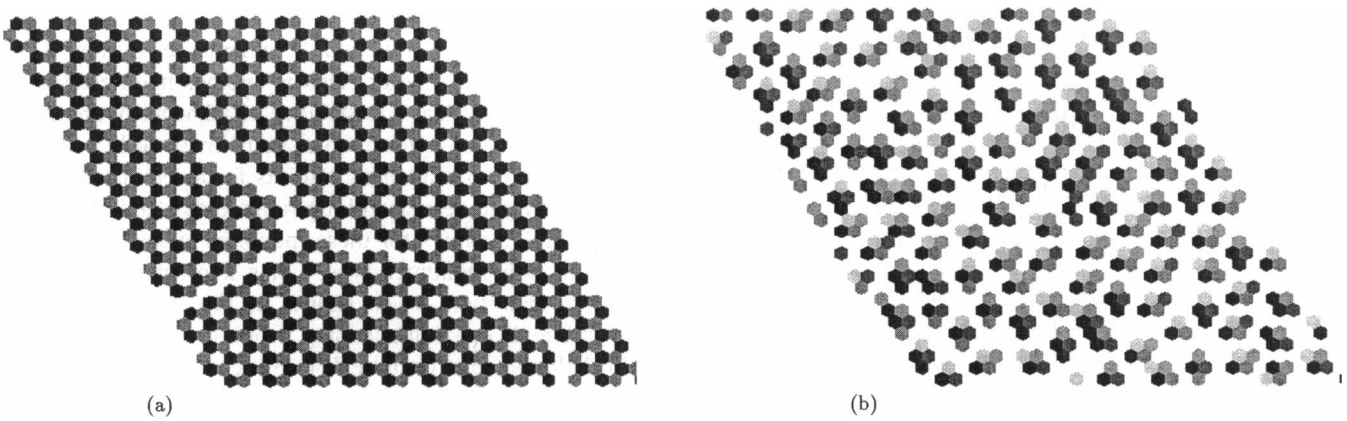

FIGURE 7 Sample equilibrium aggregate structures for (*a*) type 1 and (*b*) type 3 monomers. Different shades represent different rotational orientations of the two monomer types.

3. It is clearly seen from the figure that a surface configuration with HF surfaces interspersed between HΦ surfaces, such as monomer type 1, is more conducive to highly ordered structures than one with contiguous HΦ surface, as in monomer type 3. The predominant repeating structural unit for monomer type 3 appears to be a tetramer. Larger structures for this monomer type resemble, in an abstract sense, the linear aggregates that are believed to hinder crystallization (Carter et al., 1988; Kadima et al., 1990; Kam et al., 1978). It is important to note that the minimum system free energy state for monomer type 3 is not a regular structure that can extend infinitely in space. This observation suggests that certain proteins may not crystallize simply because their surface configurations are not conducive to a regular structural packing in space and may explain why efforts to crystallize some proteins have not been fruitful to date. Myelin basic protein, a highly hydrophobic protein with a molecular mass of only 18.5 kDa, serves as a good example. Failure to crystallize this protein has been attributed to its predominantly random coil structure and the coexistence of a population of structurally heterogeneous species in crystallization media (Sedzik, 1994; Sedzik and Kirschner, 1992), indicating that both patch distribution effects as well as conformational flexibility effects may oppose crystallization.

In addition to the surface distribution effects, we observed that the equilibrium aggregates were energetically more stable than kinetically irreversible aggregates for all the monomer types studied, even though their equilibrium densities were smaller than their jamming limits. An exception to this trend is seen for the type 1 monomer, the symmetry of which matches that of the lattice. Together, these findings clearly indicate that the lower system free energy for equilibrium aggregates is primarily due to the formation of a regular, repeating structure that maximizes the number of favorable contacts per monomer.

Protein-protein contacts within aggregates

Table 2 shows the mean (p,p) contact-type distribution for equilibrium aggregates. The data exhibit all of the general trends observed in our earlier simulations of kinetically irreversible aggregates. As monomer hydrophobicity increases, the fraction of HΦ-HΦ contacts actually decreases, reflecting an increased energetic tolerance for mismatches. And, for a given level of hydrophobicity, the more uniform the distribution of HΦ and HF patches, the greater the percentage of HΦ-HΦ contacts, again reflecting the packing difficulties encountered in accommodating contiguous patches.

One notable difference between these results and those of the kinetically controlled case is that the fraction of HF-HF contacts exceeds the fraction of HF-HΦ contacts in all cases. This is not surprising, as the reversible pathway is able to distinguish between the slight energetic advantage of HF-HF contacts relative to HF-HΦ contacts that we have built into our simulations, whereas the irreversible pathway can lock in unfavorable interactions. A reversible aggregation pathway leads to more optimal packing, as expected.

In addition, equilibrium aggregates had a greater proportion of HΦ-HΦ contacts than irreversible aggregates for all monomer types studied. This is a consequence of the fact that the HΦ-HΦ interaction is the only favorable (p,p) interaction in our simulations and is not meant to imply that crystal interactions are exclusively hydrophobic. In fact, crystal interactions are usually weak, as very strong interactions can compensate for unfavorable mismatched contacts (Crosio et al., 1992; Durbin and Feher, 1991; Kam et al., 1978), and, in some sense, make the process irreversible. This observation is in accord with recent studies of protein crystal contacts wherein it was reported that nonpolar interactions occur to a lesser extent in crystal contacts than is expected by mass action (Dasgupta, 1994).

We explicitly investigated the impact of (p,p) interaction energies on the order in irreversible aggregates. Our results indicated that small interaction energies favor more ordered

TABLE 2 Distribution of (p, p) contacts as a function of monomer type for equilibrium simulations*

Monomer type	Contact type (HF, HF) (%)	(HF, HΦ) (%)	(HΦ, HΦ) (%)
1	0.0 ± 0.0	0.0 ± 0.0	100.0 ± 0.0
2	3.3 ± 0.5	0.2 ± 0.2	96.5 ± 0.6
3	5.4 ± 0.2	0.0 ± 0.0	94.6 ± 0.2
4	0.2 ± 0.1	0.1 ± 0.0	99.7 ± 0.1
5	6.2 ± 0.2	0.6 ± 0.1	93.1 ± 0.1
6	12.1 ± 0.3	1.0 ± 0.1	86.9 ± 0.4
7	8.7 ± 0.2	5.7 ± 0.1	85.6 ± 0.1

*Results given as mean ± SD of triplicate runs.

structures. A recent database of (p,p) contacts within crystals and oligomers for a set of 225 proteins also indicates that interactions in crystals are weak and mainly polar in nature (Dasgupta, 1994).

Fig. 8 shows the number of HΦ-HΦ contacts normalized by the number of protein molecules aggregated, (HΦ-HΦ)/$\langle j \rangle$, for different monomer types. The organization of the figure is similar to that of Fig. 6. The (HΦ-HΦ)/$\langle j \rangle$ values for the equilibrium aggregates for all monomer types studied were higher than those for the kinetically irreversible aggregates and are very close to the theoretical maximum of 1.5 for types 1 through 3, 2.0 for types 4 through 6, and 2.5 for type 7, indicating that favorable contacting surfaces are used more effectively in equilibrium aggregate structures. As noted in our earlier simulations, increasing the monomer hydrophobic fraction results in a greater absolute number of (HΦ-HΦ) contacts and lower system free energies.

Because the HΦ-HΦ contact is the only favorable contact in our model, structures with less favorable ΔG^t_{sys} values should be characterized by fewer HΦ-HΦ contacts per monomer aggregated. However, as shown in Fig. 8, the (HΦ-HΦ)/$\langle j \rangle$ values for monomer types 3 and 6 are not

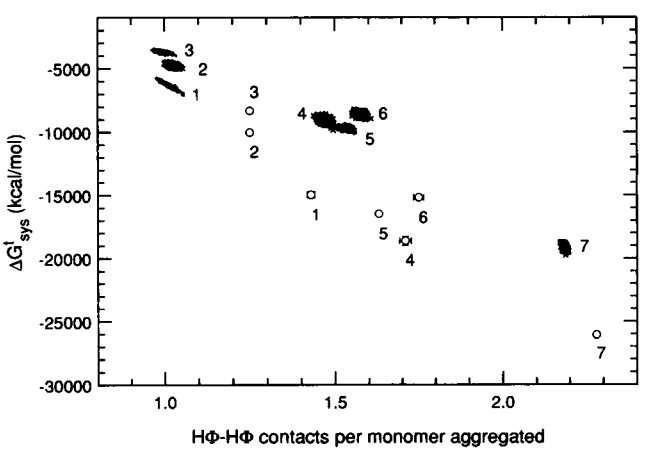

FIGURE 8 Relationship between system free energy change and number of $H\Phi - H\Phi$ contacts per monomer aggregated as a function of monomer type. The presentation of data for both the reversible and irreversible simulations is similar to Fig. 6.

smaller than other types with equivalent monomer hydrophobic contents, even though the magnitudes of their ΔG^t_{sys} values are much smaller. This is another manifestation of the reduced physical accessibility of adjacent hydrophobic sites once a given hydrophobic site is involved in a contact. A contiguous arrangement restricts further monomer addition, leading to decreased long-range order, and ultimately, higher ΔG^t_{sys} values for these monomer types. In contrast, even though monomer type 5 has a greater equilibrium density than type 4, as shown in Fig. 6, it has a less favorable ΔG^t_{sys} value arising from a lower fraction of HΦ-HΦ contacts; because HΦ-HΦ contacts are indicative of short-range order in our simulations, the higher ΔG^t_{sys} value observed for type 5 arises because of the lack of short-range order. These findings indicate that both long-range and short-range order have to be maximized to drive the system toward its global minimum.

Solvent-accessible surface area

We computed the losses in the total solvent-accessible surface area (SAS) and in the hydrophobic solvent-accessible surface area (HΦSAS) for each aggregate structure we generated. In addition to providing insights into the driving force for aggregation, a comparison of buried HΦSAS for protein aggregates and crystals may shed some light on the demarcation between aggregates and crystals. Furthermore, information about solvent accessibility can provide insights into atomic mobilities in solid-phase proteins (Harvey et al., 1984; Higuchi et al., 1984; Sheriff et al., 1985).

The contribution of hydrophobic interactions to protein behavior has received much attention since Kauzmann's landmark paper (1959). The reluctance of nonpolar groups to be exposed to water is believed to be the major driving force for protein folding (Chothia, 1976, 1985; Dill, 1985; Dill et al., 1989; Nozaki and Tanford, 1971; Tanford, 1978; Wolfenden, 1983; Wolfenden et al., 1981) and for protein aggregation (Fields et al., 1992). Our simulations anticipated this in that the HΦ-HΦ interaction was assigned as the only favorable protein-protein interaction and that hydrophobic surface burial was the only means available to the system for free energy minimization. Accordingly, we observed that equilibrium aggregates have limited solvent-exposed HΦ surface areas (data not shown). However, the extents of exposure were different for reversible and irreversible aggregates, and increased with the HΦ content of the monomer; this indicates that protein aggregation, like folding, represents a balance of exposed and buried hydrophobic surface areas. In addition, our results show that the optimum extent of HΦ burial is affected by the aggregation pathway, the monomer surface configuration, and the protein-protein interaction energies (Patro and Przybycien, 1994).

Table 3 lists the losses in HΦSAS and SAS upon aggregation for all monomer types studied. Despite lower densities, equilibrium aggregates had greater losses in HΦSAS, suggesting greater order. Islam and Weaver (1990) reported that the

overall losses in SAS upon crystallization varied between 14% and 52% for 39 monomeric proteins and between 8% and 58% for 19 dimeric proteins. The overall SAS losses observed in our simulations were biased to a much higher range, because the exposed hydrophobic fractions of our model monomers were higher than those of globular proteins (Miller et al., 1987). The wide type-to-type variation in our results confirms that protein surface characteristics significantly influence the extent of surface burial.

Porosity

Fig. 9 shows the equilibrium pore-size distributions for all of the monomer types studied. All of the simulated equilibrium aggregate structures, except those for monomer types 1 and 3, were characterized by a continuous distribution of pore sizes up to the largest pore observed for that particular monomer type. Overall porosity, or the solvent content of the aggregate structures, is inversely related to the equilibrium density and decreases with increasing monomer hydrophobicity.

Aggregates for monomer type 1 had a repetitive structural motif containing a pore of size 1 as well as a single large pore varying in size from 400 to 500, which mainly resulted from the mismatch between the size of the repeating unit and the lattice dimensions; it is not possible to build a perfectly regular structure for monomer type 1 with all favorable contacts on a lattice of dimensions 64 × 64. The lattice dimensions impose an inherent defect on aggregates for type 1 monomers.

Equilibrium aggregate structures for type 3 monomers have essentially a single large pore, as shown in Fig. 9 *a*. The presence of all of the monomer hydrophobic patches in a contiguous stretch on the type 3 monomer surface limited the long-range order. This inability to establish a continuous aggregate phase resulted in a single pore that percolates (Feder, 1980). A similar lack of long-range order was also observed for type 6 structures. This was reflected in the finding that the largest pore sizes observed for structures of monomer types 3 and 6 were larger than those observed for structures of monomers with equivalent hydrophobicity. The increased hydrophobicity of monomer type 6, however, decreased the overall porosity, leading to a nonpercolating large pore.

The mean solvent content for the equilibrium aggregates for monomer types 1, 2, and 3 are 37, 48, and 55 vol%, respectively; these monomer types have hydrophobic fractions that are similar to the average value for globular proteins (Miller et al., 1987). Matthews (1968) computed the volume percentage solvent in 116 different crystalline globular proteins and found values ranging between 27 and 65 vol%, with an extreme observed value of 95 vol% for tropomyosin. He and Carter (1992) reported 78 vol% solvent content for hSA crystals. Furthermore, Crosio et al. (1992) observed six different crystal forms of pancreatic ribonuclease to result from different precipitating conditions, with the crystal solvent content varying from 43 to 59 vol%.

Despite greater overall porosities, reversible aggregates, in general, had smaller maximum pore sizes than their irreversible counterparts, suggesting that the repeating unit of a highly ordered structure will likely have an optimum and well-defined solvent content. Although a high solvent content may adversely affect crystal resolution (Schick and Jurnak, 1994), an optimum solvent content may play a vital role in maintaining the native structure of the protein or enzyme. These solvent channels may facilitate diffusion of ligands, substrates, ions, coenzymes, and substrates into and out of the crystal and allow the use of crystalline enzymes as biocatalysts (St. Clair and Navia, 1992). Furthermore, solvent content and solid phase stability are related: a critical solvent content was also shown to stabilize solid-state proteins against thermal denaturation at 100°C (Franks et al., 1988), and it plays an important role in developing lyophilized protein formulations (Chen, 1992).

Order in protein crystals and aggregates

We have made qualitative comments about the order in protein aggregates at several points during the discussion of our simulation results. In this section, we provide a more quantitative view of the order in aggregate structures for selected monomer types.

We computed quantitative estimates of both long-range and short-range order for equilibrium aggregates of monomer types 1, 4, and 7. The equilibrium density was used as a measure of long-range order, and the (p,p) contact distribution, especially the extent of HΦ-HΦ contact formation,

TABLE 3 Percentage losses in HΦSAS and overall SAS upon aggregation*

	Equilibrium simulations		Irreversible simulations	
Monomer type	% loss in HΦSAS	% loss in SAS	% loss in HΦSAS	% loss in SAS
1	95.4 ± 0.7	47.7 ± 0.3	67.7	37.6
2	83.2 ± 0.2	43.1 ± 0.3	68.5	42.9
3	83.1 ± 0.5	43.9 ± 0.3	66.4	43.0
4	85.3 ± 0.8	57.1 ± 0.6	73.5	62.3
5	81.7 ± 0.1	58.5 ± 0.1	76.8	65.9
6	87.3 ± 0.6	67.0 ± 0.8	78.9	69.7
7	91.3 ± 0.2	88.9 ± 0.2	87.4	93.5

*Results given as mean ± SD of triplicate runs for equilibrium simulations.

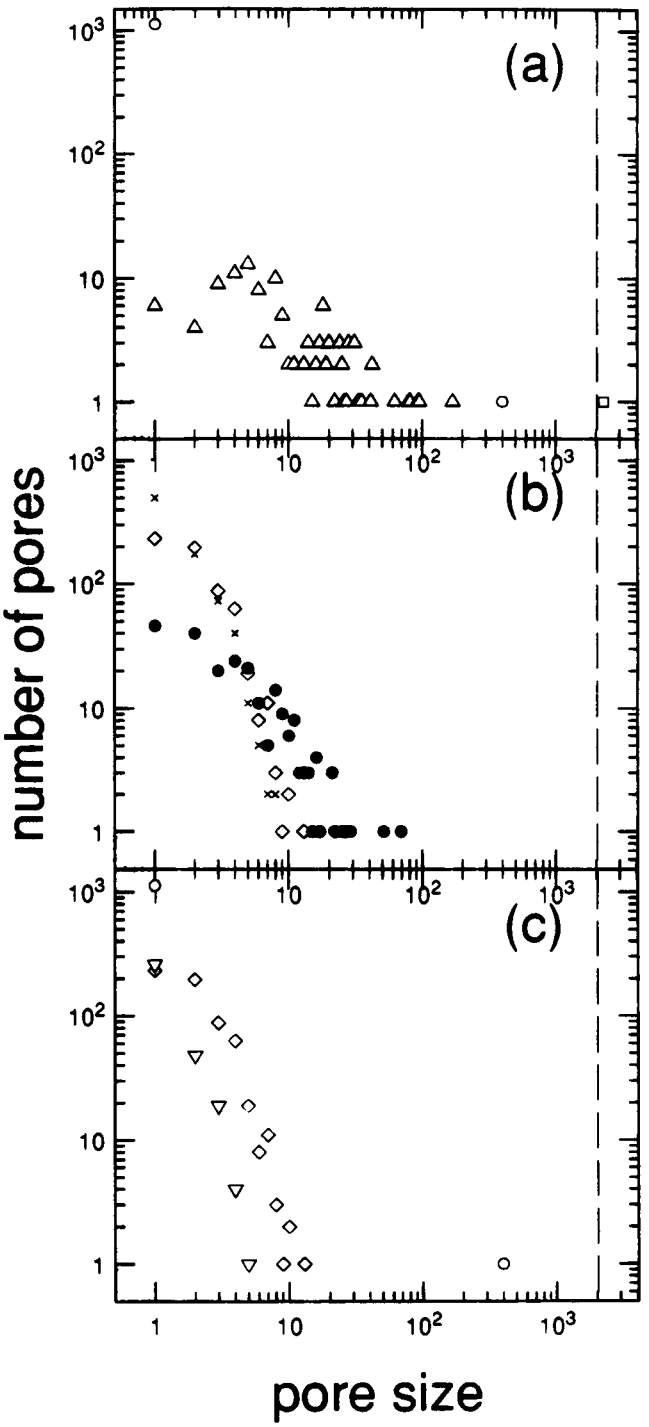

FIGURE 9 Equilibrium pore size distributions for different monomer types. (*a*) Monomers with three hydrophobic patches, types 1–3; (*b*) monomers with four hydrophobic patches, types 4–6; (*c*) a monomer with three hydrophobic patches (type 1), four hydrophobic patches (type 4), and five hydrophobic patches (type 7). *a* and *b* determine the effect of monomer surface architecture, and *c* demonstrates the impact of extent of monomer hydrophobic surface area. Monomer types 1 through 7 are represented by the symbols ○, △, □, ◇, ×, ●, and ▽, respectively. Pore sizes are measured by number of lattice points included within a single continuous pore.

was used as a measure of short-range order. Monomer type 7 is an exception; the optimum structure consists of only HΦ-HΦ contacts with a repeating unit in which six protein molecules surround a solvent molecule, leading to the formation of pores of size 1. Therefore, the short-range order for type 7 structures, at the resolution of the repeating unit, was computed from the number of pores of size 1. For monomer types 1, 4, and 7, we computed the ratios of simulated to perfect structure equilibrium density, number of HΦ-HΦ contacts, and the overall ΔG^t_{sys}, and expressed them as the percentage short-range, long-range, and overall order, respectively. Table 4 shows these data. In general, the ratio $\Delta G^t_{sys}/\Delta G^t_{sys(crystal)}$ is a good estimate of the overall order of aggregates, and this should roughly equal the product of the fractional short-range order and the fractional long-range order. This relationship, however, is less accurate for irreversible aggregates and highly hydrophobic monomers such as type 7 monomers, because of the presence of many unfavorable (p,p) contacts in addition to HΦ-HΦ contacts.

In all of the cases studied, equilibrium aggregate structures had greater order than irreversibly formed aggregate structures. In general, increasing the number of hydrophobic patches on the monomer surface resulted in decreased overall order. Aggregates of monomer type 7 have the least order, because the presence of a large number of hydrophobic surfaces per monomer provides a barrier to successful reorganization, even for the equilibrium aggregates. Once again, this finding explains why highly hydrophobic proteins such as membrane proteins are so resistant to crystallization. Also listed in Table 4 are the order estimates for structures at the end of the relaxation phase for the equilibrium simulations. Aggregate structures at this point in the simulation are analogous to reversible aggregates that precede the formation of crystals. As expected, the order of these structures is intermediate between that for equilibrium aggregates and irreversibly formed aggregates, indicating that reversible conditions are conducive to more ordered structures and that significant reordering occurs in the growth phase as well.

CONCLUSIONS

Through our simulations, we investigated the general influence of protein surface characteristics on resulting aggregate properties under thermodynamically reversible and kinetically irreversible conditions. Our results indicate that protein monomer surface characteristics profoundly affect the structure and morphology of protein aggregates. Aggregates formed under equilibrium conditions were more stable and crystal-like than their irreversible counterparts, facilitating comparisons to protein crystals and to physically observed phenomena occurring during protein aggregation and crystallization.

Although the system free energy minimization was expected to occur primarily through hydrophobic surface burial, interestingly, there was an optimum extent for this

TABLE 4 Order computation for aggregate structures

Monomer type		% order*		
		Final structure (equilibrium)	End of relaxation (equilibrium)	Final structure (irreversible)
1	Long-range	93.7	72.9	73.0
	Short-range	95.7	75.5	67.7
	Overall	89.0	53.0	37.9
4	Long-range	85.9	84.9	89.2
	Short-range	85.3	83.5	73.5
	Overall	72.1	68.0	35.2
7	Long-range	95.7	99.9	90.7
	Short-range	44.4	37.3	37.1
	Overall	69.8	59.0	51.2

*Long-range order = $(1 - |\Delta\rho|/\rho_{crystal})$, where $\rho_{crystal}$ is the equilibrium density of the optimum structure and $\Delta\rho = |\rho - \rho_{crystal}|$. Short-range order is computed based on HΦ − HΦ contacts as explained in the text, for types 1 and 4; for type 7, it was computed based on the pore size distribution. Overall order = $\Delta G^t_{sys}/\Delta G^t_{sys(crystal)}$. For optimum structures of types 1, 4, and 7: equilibrium densities are 0.67, 0.75, and 0.857; and $\Delta G^t_{sys(crystal)}$ are −16815.8, −25829.4, and −37422.5 kcal/mol, respectively.

burial, which is determined by the monomer surface configuration and the nature of the aggregation pathway. This optimum hydrophobic burial demarcates the structure of protein crystals from that of aggregates. The random, deterministic nature of a kinetically driven aggregation process may lead to entrapment in a local free energy minimum, whereas a process occurring reversibly leads to the global minimum with a corresponding highly organized structure. This explains why protein crystallization is most successfully carried out at low supersaturations; high supersaturation conditions provide a kinetic driving force with an inherent irreversibility.

We noted that either a concentration of hydrophobic surfaces in a contiguous stretch or a large extent of hydrophobicity for monomers restricts both long- and short-range order, and this finding correlates well with the difficulties encountered in attempting to crystallize highly hydrophobic membrane proteins. The majority of the (p,p) contacts that make up equilibrium aggregates and irreversibly formed aggregates were HΦ-HΦ in our simulations.

That the organization of protein surfaces controls the organization of aggregates and has been exploited by nature is evident from a variety of examples. The eye lens is made up of a mixture of crystallins, and by choosing proteins with slightly varying sequences that favor different packing densities, nature is able to fine-tune the extent of association, or the packing density, which, in turn, regulates the refractive index gradient requisite for normal function of the eye (Fernald and Wright, 1983; Sergeev et al., 1988; White et al., 1989); Sergeev et al. (1988) have attributed the difference in the packing densities of two highly homologous calf γ crystallins to the variation in their sequence at positions 103 and 155. Hence, it is likely that specific interaction sites are involved in the evolutional optimization of crystallins. The transparency of the eye lens is also the result of the long-range self-association of certain γ-crystallins. White et al. (1989) have reported that specific protein interactions may also be responsible for limiting these long-range interactions, which, in turn, would lead to inhomogeneous packing on the scale of light wavelength, causing cataract formation. An example of nature's selection of specific interaction sites to prevent in vivo aggregation can be seen in the binding of chaperonins to nascent polypeptide chains during protein folding. Fenton et al. (1994) reported that chaperonin GroEL functions through a specific binding site and identified a putative polypeptide-binding site and a highly conserved residue, Asp87, which is essential for ATP hydrolysis and subsequent polypeptide release. Specific protein interactions are responsible for molecular diseases as well. Sickle cell anemia, a molecular disease first described by Pauling et al. (1949), is a consequence of aggregation of deoxygenated sickle cell hemoglobin into fibers. The specific replacement of glutamic acid by valine in the β6 position is believed to cause this aggregation (Ingram, 1959); a double mutation, β6 Glu → Val and β73 Asp → Asn, leads to a milder version of the disease, because the latter modification, β73 Asp → Asn, supposedly reduces the ease of aggregation (Bookchin et al., 1970). Perturbations that have an impact on the driving force for aggregation or affect the organization of interaction sites, or both, can dramatically alter aggregation behavior; in terms of our simple model, these perturbations translate into changes in either the site-site interaction energies or the monomer type, or both.

In addition to these examples, there is experimental evidence that subtle sequence modifications lead to successful crystallization when the efforts to crystallize the wild-type molecule have failed (D'Arcy, 1994; McElroy et al., 1992), suggesting the critical role played by monomer surface architecture in controlling protein self-association. Specific sites may also be exploited for crystallization via the matrix coprecipitation-cocrystallization technique, in which specially designed polycyclic aromatic charged ligands bind to specific sites on the protein surface, leading to compaction of the protein into crystallizable configurations, as well as providing support for an interconnected network (Conroy and Lovrien, 1992). Because the natural selection of proteins is based on function and not on the ease with which they can be crystallized, subtle alterations of protein surface characteristics may be effected via protein engineering to aid in protein crystalli-

TABLE 5 Distribution of (p,p) contacts as a function of monomer type and $\Delta\Delta G(H\Phi,H\Phi)$

	Monomer type	$\Delta\Delta G(H\Phi,H\Phi)$ (kcal/mol)	(*HF*,*HF*) (%)	(*HF*,*HΦ*) (%)	(*HΦ*,*HΦ*) (%)
Simulated results	1	−4.5	0.69	9.31	90.00
	2	−4.5	12.78	7.38	79.84
	3	−4.5	15.63	7.10	77.27
	4	−4.5	8.33	13.07	78.60
	5	−4.5	10.60	11.73	77.67
	6	−4.5	12.33	12.19	75.48
	7	−4.5	5.84	16.27	77.89
	1	−1.0	0.002	0.008	99.99
	1	−2.0	0.009	0.170	99.82
	1	−3.0	0.05	3.03	96.92
	1	−4.0	0.32	8.76	90.92
	1	−5.0	1.39	10.42	88.19
Calculated results*	1, 2, 3		2.023×10^{-3}	7.344×10^{-4}	99.997
	4, 5, 6		5.056×10^{-4}	3.672×10^{-4}	99.9991
	7		8.091×10^{-5}	1.469×10^{-4}	99.9998

*Probability of pair-wise contacts made in isolation.

zation. Whereas analyses pertaining specifically to crystal contacts (Dasgupta, 1994) may allow a fine-tuning of mutations to promote crystallization, our simulation results may provide some very basic guidance as to what surface architectures might be most productively exploited.

APPENDIX

During the preparation of this manuscript as a follow-up to "Simulations of Kinetically Irreversible Protein Aggregate Structure," which appeared in the May 1994 issue of the *Biophysical Journal* (Patro and Przybycien, 1994), we noticed a subtle typographical error in our simulation code. This error caused the solvent-solvent interaction energy in the free energy calculation at one specific site of a transferred monomer to be set to zero. The net result of this error was to impose a small amount of artificial order on the simulations. This error occurred as we shifted the program to a new platform on campus; we discovered the error when modifying the code to consider the reversible aggregate case.

We have re-executed all of the simulations from the original paper. Although the magnitudes of some of the properties originally reported have changed on correction of the error, the trends noted in the first paper remain essentially identical. All comparisons with the kinetically irreversible case reported in this paper are made to the corrected simulations. We continue here with a complete summary of the corrected numerical results for the kinetically irreversible case; in reference to figures and tables in the following discussion, (I) denotes the irreversible simulation paper and (II) the present paper.

Jamming limit: In Figs. 7 (I) and 8 (I), only the numerical values have changed. Corrected data from Fig. 7 (I) are shown in Fig. 6 (II).

(p,p) contacts in aggregates: All of the trends observed in Figs. 9 (I) and 10 (I) remain intact, with changes only in the numerical values. Corrected data from Fig. 9 (I) are shown in Fig. 8 (II). In Table 2 (I), except for type 1 monomer, 77–80% of the contacts are HΦ-HΦ. The corrected version of this table is given at the end of this Appendix (Table 5).

Solvent-accessible surface area: For monomer types 1 through 7, 67–87% of the monomer hydrophobic surface was buried upon aggregation. Overall losses in SAS for types 1 through 6 were 38, 43, 43, 62, 66, and 70%, respectively. For type 1 monomers, overall losses in SAS varied from 30 to 39% as $\Delta\Delta G(H\Phi,H\Phi)$ was varied from −5 to −1 kcal/mol.

Pore size distribution: In contrast to Fig. 11 (I), the corrected pore size distributions for monomer types 1 and 2 were essentially continuous. The following numerical values for the largest pore-size changed upon correction: type 3 structures had a percolating pore of size ranging from 2114 to 2256, types 4 and 5 had no pores of size greater than 30, the largest pore for type 6 aggregates was 139, and the largest pore for type 7 aggregates was of size 6 and was observed only seven times in 500 runs. In contrast to Fig. 12 (I), the corrected pore size distribution for $\Delta\Delta G(H\Phi,H\Phi) = -4$ kcal/mol was also continuous.

This work was supported in part by National Science Foundation grant CTS-9211666.

REFERENCES

Bam, N. B., T. W. Randolph, and J. L. Cleland. 1995. Stability of protein formulations: investigation of surfactant effects by a novel EPR spectroscopic technique. *Pharm. Res.* 12:2–11.

Becker, G. W., R. R. Bowsher, W. C. Mackellar, M. L. Poor, P. M. Tackitt, and R. M. Riggin. 1987. Chemical, physical and biological characterization of biosynthetic human growth hormone. *Biotechnol. Appl. Biochem.* 9:478–487.

Bookchin, R. M., R. L. Nagel, and H. M. Ranney. 1970. The effect of β^{73} Asn on the interactions of sickling hemoglobins. *Biochim. Biophys. Acta.* 221:373–375.

Brems, D. N. 1988. Solubility of different folding conformers of bovine growth hormone. *Biochemistry.* 27:4541–4546.

Brems, D. N., L. A. Alter, M. J. Beckage, R. E. Chance, R. D. DiMarchi, L. K. Green, H. B. Long, A. H. Pekar, J. E. Shields, and B. H. Frank. 1992. Altering the association properties of insulin by amino acid replacement. *Protein Eng.* 5:527–533.

Brzozowski, A. M., and S. P. Tolley. 1994. Poly(ethylene)glycol monomethyl ethers—an alternative to poly(ethylene) glycols in protein crystallization. *Acta. Crystallogr.* D50:466–468.

Carter, C. W., E. T. Baldwin, and L. Frick. 1988. Statistical design of experiments for protein crystal growth and the use of a precrystallization assay. *J. Cryst. Growth.* 90:60–73.

Casal, H. L., U. Kohler, and H. H. Mantsch. 1988. Structural and conformational changes of β-lactoglobulin B: an infrared spectroscopic study of the effect of pH and temperature. *Biochim. Biophys. Acta.* 957:11–20.

Chen, T. 1992. Formulation concerns of protein drugs. *Drug Dev. Ind. Pharm.* 18:1311–1354.

Chothia, C. 1976. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* 105:1–14.

Chothia, C. 1985. Principles that determine the structure of proteins. *Annu. Rev. Biochem.* 53:537–572.

Cleland, J. L., M. F. Powell, and S. J. Shire. 1993. The development of stable protein formulations: a close look at protein aggregation, deamidation, and oxidation. *Crit. Rev. Ther. Drug Carrier Syst.* 10:307–377.

Cleland, J. L., and D. I. C. Wang. 1993. Cosolvent effects on refolding and aggregation. *In* Biocatalysis Design for Stability and Specificity, ACS Symposium Series, Vol. 516. M. Himmel and G. Georgiou, editors. American Chemical Society, Washington, DC.

Conroy, M. J., and R. E. Lovrien. 1992. Matrix coprecipiting and cocrystallizing ligands (MCC ligands) for bioseparations. *J. Cryst. Growth.* 122:213–222.

Crosio, M., J. Janin, and M. Jullien. 1992. Crystal packing in six crystal forms of pancreatic ribonuclease. *J. Mol. Biol.* 228:243–251.

Cudney, B., S. Patel, K. Weisgraber, Y. Newhouse, and A. McPherson. 1994. Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr.* D50:414–423.

D'Arcy, A. 1994. Crystallizing proteins—a rational approach? *Acta Crystallogr.* D50:469–471.

Dasgupta, S. 1994. Statistical analyses of nonbonded interactions at protein crystal and oligomer interfaces. Ph.D. thesis. Rensselaer Polytechnic Institute, Troy, NY.

De Young, L. R., A. L. Fink, and K. A. Dill. 1993a. Aggregation and denaturation of apomyoglobin in aqueous urea solutions. *Biochemistry.* 32:3877–3886.

De Young, L. R., A. L. Fink, and K. A. Dill. 1993b. Aggregation of globular proteins. *Acc. Chem. Res.* 26:614–620.

Dill, K. A. 1985. Theory for the folding and stability of globular proteins. *Biochemistry.* 24:1501–1509.

Dill, K. A., D. O. V. Alonso, and K. Hutchinson. 1989. Thermal stabilities of globular proteins. *Biochemistry.* 28:5439–5449.

Durbin, S. D., and G. Feher. 1991. Simulation of lysozyme crystal growth by the Monte Carlo method. *J. Cryst. Growth.* 110:41–51.

Feder, J. 1980. Random sequential adsorption. *J. Theor. Biol.* 87:237–254.

Fenton, W. A., Y. Kashi, K. Furtak, and A. L. Horwich. 1994. Residues in chaperonin groEL required for polypeptide binding and release. *Nature.* 371:614–619.

Fernald, R. D., and S. E. Wright. 1983. Maintenance of optical quality during crystalline lens growth. *Nature.* 301:618–620.

Fields, G. B., D. O. V. Alonso, D. Stigter, and K. A. Dill. 1992. Theory for the aggregation of proteins and copolymers. *J. Phys. Chem.* 96: 3974–3981.

Franks, F., R. H. M. Hatley, and H. L. Friedman. 1988. The thermodynamics of protein stability: cold destabilization as a general phenomenon. *Biophys. Chem.* 31:307–315.

Garavito, R. M., Z. Markovic-Housley, and J. A. Jenkins. 1986. The growth and characterization of membrane protein crystals. *J. Cryst. Growth.* 76:701–709.

Garavito, R. M., and D. Picot. 1991. Crystallization of membrane proteins: a minireview. *J. Cryst. Growth.* 110:89–95.

Goldenberg, D. P., D. H. Smith, and J. King. 1983. Genetic analysis of the folding pathway for the tailspike protein of phage P22. *Proc. Natl. Acad. Sci. USA.* 80:7060–7064.

Harvey, S. C., M. Prabhakaran, B. Mao, and J. A. McCammon. 1984. Phenylalanine transfer RNA: molecular dynamics simulation. *Science.* 223:1189–1191.

He, X. M., and D. C. Carter. 1992. Atomic structure and chemistry of human serum albumin. *Nature.* 358:209–215.

Hellebust, H., M. Murby, L. Abrahmsen, M. Uhlen, and S. O. Enfors. 1989. Different approaches to stabilize a recombinant fusion protein. *Biotechnology.* 7:165–168.

Higuchi, Y., M. Kusunoki, Y. Matsuura, N. Yasuoka, and M. Kakudo. 1984. Refined structure of cytochrome c_3 at 1.8 Å resolution. *J. Mol. Biol.* 172:109–139.

Ingram, V. M. 1959. Abnormal human haemoglobins. III. The chemical difference between normal and sickle cell haemoglobins. *Biochim. Biophys. Acta.* 36:402–411.

Islam, S. A., and D. L. Weaver. 1990. Molecular interactions in protein crystals: solvent accessible surface and stability. *Protein Struct. Funct. Genet.* 8:1–5.

Kabalnov, A. S., and E. D. Shchukin. 1992. Ostwald ripening theory: applications to fluorocarbon emulsion stability. *Adv. Colloid. Interface Sci.* 38:69–97.

Kadima, W., A. McPherson, M. F. Dunn, and F. A. Jurnak. 1990. Characterization of precrystallization aggregation of canavalin by dynamic light scattering. *Biophys. J.* 57:125–132.

Kam, Z., H. B. Shore, and G. Feher. 1978. On the crystallization of proteins. *J. Mol. Biol.* 123:539.

Kane, J. F., and D. L. Hartley. 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli. Trends Biotechnol.* 6:95–101.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* 14:1–63.

Lewis, U. J., E. V. Cheever, and B. K. Seavey. 1969a. Aggregate-free human growth hormone. I. Isolation by ultrafiltration. *Endocrinology.* 84:325–331.

Lewis, U. J., D. C. Parker, M. D. Okerlund, R. M. Boyar, M. Litteria, and W. P. Vanderlaan. 1969b. Aggregate-free human growth hormone. II. Physico-chemical and biological properties. *Endocrinology.* 84: 332–339.

Liang, J. N., and X. Li. 1991. Interaction and aggregation of lens crystallins. *Exp. Eye Res.* 53:61–66.

Matthews, B. W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* 33:491–497.

McElroy, H. E., G. W. Sisson, W. E. Schoetlin, R. M. Aust, and J. E. Villafranca. 1992. Studies on engineering crystallizability by mutation of surface residues of human thymidylate synthase. *J. Cryst. Growth.* 122:265–272.

McPherson, A. 1990. Current approaches to macromolecular crystallization. *Eur. J. Biochem.* 189:1–23.

Mikol, V., E. Hirsch, and R. Giege. 1990. Diagnostic of precipitant for biomacromolecule crystallization by quasi-elastic light-scattering. *J. Mol. Biol.* 213:187–195.

Miller, S., J. Janin, A. M. Lesk, and C. Chothia. 1987. Interior and surface of monomeric proteins. *J. Mol. Biol.* 196:641–656.

Mitraki, A., B. Fane, C. Haase-Pettingell, J. Sturtevant, and J. King. 1991. Global suppression of protein folding defects and inclusion body formation. *Science.* 253:54–58.

Moore, W. V., and P. Leppert. 1980. Role of aggregated human growth hormone (hGH) in development of antibodies to hGH. *J. Clin. Endocrinol.* 51:691–697.

Nozaki, Y., and C. Tanford. 1971. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. *J. Biol. Chem.* 246:2211–2217.

Pangiotopoulos, A. Z. 1987. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol. Phys.* 61:813–826.

Patro, S. Y., and T. M. Przybycien. 1994. Simulations of kinetically irreversible protein aggregate structure. *Biophys. J.* 66:1274–1289.

Pattou, D., B. Maigret, M. C. Fourniezaluski, and B. P. Roques. 1991. Computational analysis of conformational behavior of cholecystokinin fragments I—Cck4, Cck5, Cck6 and Cck7 molecules. *Int. J. Pept. Protein Res.* 37:440–450.

Pauling, L., H. A. Itano, S. J. Singer, and I. C. Wells. 1949. Sickle cell anemia, a molecular disease. *Science.* 110:543–548.

Przybycien, T. M., and J. E. Bailey. 1989. Aggregation kinetics in salt-induced protein precipitation. *AIChE J.* 35:1779–1790.

Rinas, U., L. B. Tsai, D. Lyons, G. M. Fox, G. Stearns, J. Fieschko, D. Fenton, and J. E. Bailey. 1992. Cysteine to serine substitutions in basic fibroblast growth factor: effect on inclusion body formation and proteolytic susceptibility during in vitro folding. *Biotechnology.* 10:435–440.

Savitzky, A., and M. Golay. 1967. Smoothing and differentiation of data by simplified least square procedures. *Anal. Chem.* 36:1627–1639.

Schick, B., and F. Jurnak. 1994. Crystal growth and crystal improvement strategies: extension of the diffraction resolution of crystals. *Acta Crystallogr.* D50:563–568.

Sedzik, J. 1994. DESIGN: a guide to protein crystallization experiments. *Arch. Biochem. Biophys.* 308:342–348.

Sedzik, J., and D. K. Kirschner. 1992. Is myelin basic protein crystallizable? *Neurochem. Res.* 17:157–166.

Sergeev, Y. V., Y. N. Chirgadze, S. E. Mylvaganam, H. Driessen, C. Slingsby, and T. L. Blundell. 1988. Surface interactions of γ-crystallins in the crystal medium in relation to their association in the eye lens. *Proteins.* 4:137–147.

Sheriff, S., W. A. Hendrickson, R. E. Stenkamp, L. C. Sieker, and L. H. Jensen. 1985. Influence of solvent accessibility and intermolecular contacts on atomic mobilities in hemerythrins. *Proc. Natl. Acad. Sci. USA.* 82:1104–1107.

Shortle, D., H. S. Chan, and K. A. Dill. 1992. Modeling the effects of mutations on the denatured states of proteins. *Protein Sci.* 1:201–215.

Sluzky, V., A. M. Klibanov, and R. Langer. 1992. Mechanism of insulin aggregation and stabilization in agitated aqueous solutions. *Biotech. Bioeng.* 40:895–903.

St. Clair, N. L., and M. A. Navia. 1992. Cross-linked enzyme crystals as robust biocatalysts. *J. Am. Chem. Soc.* 114:7314.

Steinier, J., Y. Termonia, and J. Deltour. 1972. Comments on smoothing and differentiation of data by simplified least square procedure. *Anal. Chem.* 44:1906–1909.

Takahashi, T., S. Endo, and K. Nagayama. 1993. Stabilization of protein crystals by electrostatic interactions as revealed by a numerical approach. *J. Mol. Biol.* 234:421–432.

Tandom, S., and P. M. Horowitz. 1989. Reversible folding of rhodanese. *J. Biol. Chem.* 264:9859.

Tanford, C. 1978. The hydrophobic effect and the organization of living matter. *Science.* 200:1012–1018.

Vonck, J., and E. F. J. van Bruggen. 1990. Electron microscopy and image analysis of two-dimensional crystals and single molecules of alcohol oxidase from *Hansenula polymorpha. Biochim. Biophys. Acta.* 1038: 74–79.

Wetzel, R., L. J. Perry, and C. Veilleux. 1991. Mutations in human interferon γ affecting inclusion body formation identified by a general immunochemical screen. *Biotechnology.* 9:731–737.

White, H. E., H. P. C. Driessen, C. Slingsby, D. S. Moss, and P. F. Lindley. 1989. Packing interactions in the eye lens. *J. Mol. Biol.* 207:217–235.

Wolfenden, R. 1983. Waterlogged molecules. *Science.* 222:1087.

Wolfenden, R., L. Andersson, P. M. Cullis, and C. C. B. Southgate. 1981. Affinities of amino acid side chains for solvent water. *Biochemistry.* 20:849–855.

Yeager, M. 1994. Crystallization of membrane proteins: in situ two-dimensional crystallization of a polytopic membrane protein: the cardiac gap junction channel. *Acta. Crystallogr.* D50:632–638.

Yoshikawa, S., K. Shinzawa-Itoh, H. Ueda, T. Tsukihara, Y. Fukumoto, T. Kubota, M. Kawamoto, K. Jukuyama, and H. Matsubara. 1992. Strategies for crystallization of large membrane protein complexes. *J. Cryst. Growth.* 122:298–302.

Yu, C. A., L. Yu, and T. E. King. 1972. Preparation and properties of cardiac cytochrome c^1*. *J. Biol. Chem.* 247:1012–1019.

Yue, K., and K. A. Dill. 1992. Inverse protein folding problem: designing polymer sequences. *Proc. Natl. Acad. Sci. USA.* 89:4163–4167.