Simulations of Nucleation and Early Growth Stages of Protein Crystals

Andrzej M. Kierzek,* Wojciech M. Wolf,# and Piotr Zielenkiewicz*
*Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02–106 Warsaw, and #Institute of General and Ecological Chemistry Technical University of Łódź, 90-924 Łódź, Poland

ABSTRACT Analysis of known protein crystal structures reveals that interaction energies between monomer pairs alone are not sufficient to overcome entropy loss related to fixing monomers in the crystal lattice. Interactions with several neighbors in the crystal are required for stabilization of monomers in the lattice. A microscopic model of nucleation and early growth stages of protein crystals, based on the above observations, is presented. Anisotropy of protein molecules is taken into account by assigning free energies of association (proportional to the buried surface area) to individual monomer-monomer contacts in the lattice. Lattice simulations of the tetragonal lysozyme crystal based on the model correctly reproduce structural features of the movement of dislocation on the (110) crystal face. The dislocation shifts with the speed equal to the one determined experimentally if the geometric probabilty of correct orientation is set to 10^{-5}, in agreement with previously published estimates. At this value of orientational probability, the first nuclei, the critical size of which for lysozyme is four monomers, appear in 1 ml of supersaturated solution on a time scale of microseconds. Formation of the ordered phase proceeds through the growth of nuclei (rather then their association) and requires nucleations on the surface at certain stages.

INTRODUCTION

A lack of methods that are able to predict protein crystallization conditions or at least help in the automated search for them is the bottleneck in protein structure determination. Knowledge about the mechanism of crystallization of already crystallized molecules would not only help to design crystallization conditions for others, but also strongly contribute to our understanding of protein aggregation, which is ubiquitous in the cell and biotechnological applications. For these reasons, several sophisticated experimental techniques have been applied to investigate protein crystallization. Electron microscopy (Durbin and Feher, 1990) and atomic force microscopy (Konnert et al., 1994; Malkin et al., 1995) have been used to study the growth of crystal faces in several proteins. Light scattering and neutron scattering experiments have been performed to study aggregation phenomena in protein solutions under crystallization conditions (Eberstein et al., 1994; Georgalis et al., 1995; Muschol and Rosenberger, 1995; Niimura et al., 1995).

To date the experimental techniques are unable to provide insights into the very early stages of the crystallization process. Thus for studying nucleation and early stages of growth, theoretical approaches and computer simulations must be applied. The huge body of analytical approaches resulting from the classical nucleation theory describe the crystallization process on the mesoscopic/macroscopic scale. The vast majority of computer simulations were performed on ionic or small molecular systems. Representative examples are simulations performed by Meakin (1988). He used diffusion-limited cluster aggregation and reaction-limited cluster aggregation for colloidal gold particles. The main disadvantage of those methods, including the classical nucleation theory, is that they assume isotropicity and a spherical shape of the modeled species. Contrary to that assumption, protein molecules are highly anisotropic. Therefore, one of the most important properties to be considered during the simulation of protein aggregation is the mutual orientation of individual protein molecules. The numerical parameter that accounts for this feature is called *orientational probability*. It simply describes the probability that two molecules of a certain shape are aligned in proper orientation to form a defined interface when they meet in space. For large molecular systems like proteins, the values of orientational probabilities are low and strongly affect aggregation kinetics.

To the best of our knowledge, there are only two reports in the literature on theoretical models that have been applied to the study of protein crystallization and in which the shape of protein molecules has been taken into consideration. Tissen et al. (1994) developed a model to study protein crystallization on the basis of the 3D structure of a particular protein. They applied Stokesian dynamics and continuum hydrodynamics to protein molecules, which were represented as rigid bodies. The model correctly predicted diffusion coefficients of various proteins. Possible applications to the study of nucleation and crystal growth have been suggested. Recently, Patro and Przybycien (1996) simulated the structure of reversible protein aggregates as a function of protein surface characteristics. The protein-protein interaction energies and the entropic penalty were responsible for immobilization of the protein molecule in the solid phase. To perform Monte Carlo simulations on a 2D lattice, the real molecules were idealized by hexagons. Very general aspects of protein crystallization and aggregation were discussed.

Received for publication 10 February 1997 and in final form 12 May 1997.
Address reprint requests to Dr. Piotr Zielenkiewicz, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5a, 02-106 Warsaw, Poland. Tel.: 48-22-658-47-03; Fax: 48-39-12-16-23; E-mail: piotr@ibbrain.ibb.waw.pl.

0006-3495/97/08/571/10 $2.00

The model presented in this paper attempts to describe nucleation and the early stages of growth of a model protein of known crystal structure. It is assumed that the protein-protein interactions leading to crystal formation can be described within the same formalism as the one developed for the protein-protein recognition process. Chothia and Janin (1975) have shown that the major negative term in the free energy of association comes from the (de)solvation contribution, which can be estimated as proportional to the surface area buried upon complex formation. The empirical dependence of the free energies of association on the surface areas buried when expressed with the use of atomic solvation parameters (Eisenberg and McLachlan, 1986) gives very good agreement with experimentally measured association constants (Horton and Lewis, 1992). Therefore we assume that the free energy of association can be evaluated as a sum of two terms of opposite signs: one (favoring association) proportional to the surface area buried on the interface, and the constant term related to the rotational and translational entropy loss of subunits upon their aggregation. We observe one important difference between protein crystal growth and protein-protein interaction leading to the formation of functional complexes: the surface areas buried on the interface between two monomers in the crystal are small (Janin and Rodier, 1995) and the calculated interaction energy is smaller than the value of the entropic term. However, in the crystal environment, a single molecule is surrounded by several symmetry-related molecules. Interactions with many neighbors allows the molecule to overcome the entropy loss and fixes the molecule in the crystal lattice. Adding molecules to neighboring positions in the lattice increases the surface area of this molecule buried in interfaces with surrounding molecules and leads to its stabilization in the lattice.

In this paper we present lattice simulations based on the above mechanism, which reproduce the experimentally observed behavior of the dislocation on the (110) face of tetragonal lysozyme crystal. The method is then used to study nucleation and the early stages of growth of this crystal.

FORMULATION OF THE MODEL

In the first stage of our simulation, we generate the complex of protein molecules that reproduces the full set of the intermolecular interactions in the crystal form under investigation. The complex contains a single protein monomer surrounded by the monomers in the crystal lattice, which form with it at least one intermolecular distance shorter than 4.5 Å. The starting monomer in the generation of that complex will subsequently be referred to as the central molecule. The energy of interaction of each of surrounding monomers with the central one is estimated as proportional to the surface area buried on the interface upon complex formation.

In the second stage of simulation, the three-dimensional lattice is built. The protein molecules are represented as points occupying nodes of the lattice. Each node, together with its neighbors, represents a cluster of molecules built around the central molecule, as calculated at the beginning of the simulation. Edges connecting nodes filled in with molecules describe intermolecular interfaces. An integer variable representing discrete orientational states is assigned to each molecule. Two molecules are considered to be interacting if they occupy neighboring nodes and have the same orientational state. The edge between neighboring nodes represents one of the interfaces in the crystal. The interaction energy of that pair of molecules is assigned half of the value calculated for the corresponding pair of molecules in the crystal environment. The interaction energy is then subtracted from the entropic penalty and assigned to each of the interacting monomers as its free energy of association. A schematic diagram of the model is shown in Fig. 1.

By using this representation of the system, several features of protein crystallization can be studied. Sizes and shapes of the first stable nuclei can be studied by random generation of aggregates in the center of the lattice. Behavior of the dislocation on the crystal face and kinetics of nucleation can be analyzed by simulating diffusion as the random walk of points in the lattice. The following sections will describe this approach in detail.

Generation of the crystal environment

The Protein Data Bank (Bernstein et al., 1977) entry file was carefully examined and edited to exclude all atoms that do not belong directly to the protein, i.e., all water, substrate, and cofactor molecules as well as ions were not taken into consideration. Then the "pure" protein molecule was transformed into the crystallographic coordinate framework. All interatomic distances between this molecule and molecules surrounding it in the crystal were generated and examined. For that purpose, 26 crystallographic unit cells having a face or edge in common with the original unit cell as well as the crystallographic and noncrystallographic (if applicable) symmetry operators have been generated and applied. If any of the intermolecular distances between the central and surrounding molecules were shorter than 4.5 Å, the latter was regarded as significantly contributing to the association energy and included in the initial complex. Calculations were performed using the in-house-developed program MICELL (written by W. M. Wolf with the use of libraries from the CCP4 suite; Collaborative Computational Project, 1994).

Energy calculations

For each surrounding molecule, its interaction energy with the central one was assumed to be proportional to the change of accessible surface area upon formation of the interface between the molecules. Accessible surface areas for monomers and each of the dimers were calculated using

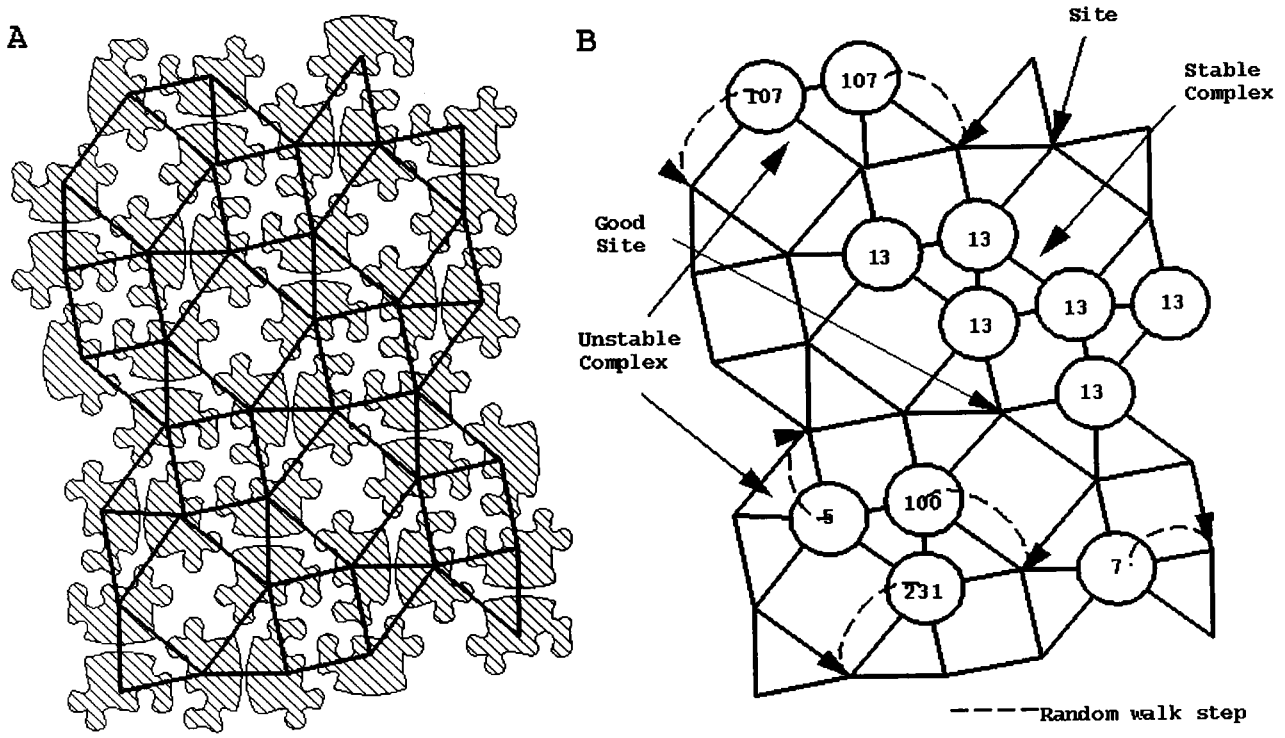

FIGURE 1 (*A*) Two-dimensional schematic diagram of the lattice set-up. Graphics primitives portray protein molecules in the crystal. Each node of the lattice represents a position of the molecule. Edges represent "protein bonds," i.e., interfaces between molecules that have at least one intermolecular distance shorter than 4.5 Å. (*B*) During simulations, protein molecules are simplified to points with randomly assigned orientational states (*circles with numbers*). For the hexamer labeled as "stable complex," 1) all six molecules have the same orientational state, which means that they are properly oriented to form interactions; and 2) each monomer forms many interactions, which lets it overcome the entropic cost of aggregation. The node of the lattice that is the neighbor of at least one molecule belonging to the hexamer is considered a site on the surface of that complex. One of the sites has the property that when a molecule arrives at that site, with proper orientational state, it forms enough interactions to overcome the entropic cost of aggregation. This site is labeled a "good site." There are two "unstable complexes" shown on the picture. In the trimer, all molecules have different orientational states, so they are considered to be improperly oriented to form any interactions. The dimer is considered unstable because single interaction is not sufficient to overcome the entropic penalty of aggregation. All monomers that do not belong to stable complexes are moved randomly to unoccupied neighboring nodes. During the moves, orientational states are randomly changed.

the analytical approach of Richmond (1984), with a probe radius of 1.4 Å. Energy was calculated according to the equation

$$\Delta G^{\circ}_{\text{interaction}} = \sum_{\text{atoms i}} \sigma_i (A_{\text{id}} - A_{\text{im}}) \qquad (1)$$

where A_{id} and A_{im} are the accessible surface areas of the ith atom in the dimer and monomer, respectively, and σ_i is the atomic solvation parameter specified for the type of ith atom. Values of atomic solvation parameters fitted to reproduce free energies of transfer of amino acid side-chain analogs from octanol to water were taken from the work of Eisenberg and McLachlan (1986).

Lattice coordinate system

The principle of the system that was used to define positions of the nodes in the lattice is based on the very well known feature that the crystals are built from a huge number of repetitive unit cells. Within the unit cell, molecules are related to each other by symmetry operators. The minimum set of symmetry operators required to build contents of the unit cell starting from an asymmetrical unit is called a *crystallographic space group.* Therefore, four integer-type coordinates are sufficient to unequivocally describe the position of a particular molecule in the crystal. The first three numbers define the origin of the unit cell in which the molecule is located, and the fourth number represents the symmetry operator. Thus the symbol (n_a, n_b, n_c, n_z) means that the particular molecule is related to the starting one by the symmetry element represented by the symmetry operator n_z and is located in the unit cell, the origin of which is defined by vector $\mathbf{r} = n_a\mathbf{a} + n_b\mathbf{b} + n_c\mathbf{c}$. Starting from the initial orthonormal Protein Data Bank-type atomic parameters and knowing the coordinates of the Nth molecule defined by (n_a, n_b, n_c, n_z), it is possible to generate Cartesian (x, y, z) coordinates of the atoms of that molecule. Furthermore, it is possible to derive four-dimensional vectors defining, for any molecule in the lattice, coordinates of all its neighbors with at least one intermolecular distance shorter than 4.5 Å. These vectors can be found by analysis of complexes built, as described in the previous section,

around each molecule in the unit cell. Table 1 shows the vectors derived for the tetragonal form of lysozyme crystal. To obtain coordinates of the ith neighboring node of the lattice node with coordinates (n_a, n_b, n_c, n_z), one should find in the table the ith vector for the molecule defined by coordinates (0, 0, 0, n_z). Values of the first three coordinates of the vector should be than added to n_a, n_b, n_c, respectively. The operator number of the node being calculated is simply taken from the vector in the table. Using this algorithm, it is possible to find lattice coordinates of all neighboring nodes of any node in the lattice.

Lattice simulations

For any configuration of monomers in the lattice, free energy of association for any monomer can be calculated according to the following equation:

$$\Delta G^{\circ}_{assoc}(k) = \Delta G^{\circ}_{rot,trans} - \sum_{j} \delta(k, j)\Delta G^{\circ}_{inter}(j) \qquad (2)$$

where $\Delta G^{\circ}_{assoc}(k)$ is the free energy of association of the kth monomer; $\Delta G^{\circ}_{rot,trans}$ is the entropic penalty related to the loss of rotational and translational degrees of freedom; $\delta(k, j)$ is equal to 1 if the jth neighboring node of molecule k is occupied by the monomer with the same orientational state as k, and to 0 otherwise; $\Delta G^{\circ}_{inter}(j)$ is half of the value of the interaction energy between the central molecule and its jth neighbor in the crystal environment.

For a complex of monomers, a node that is a neighbor of at least one node occupied by the monomer of that complex will be referred to as a site on the surface of this complex. It is convenient to introduce the term "fixing potential" of the site. For a given site of the particular complex, its fixing potential is equal to the value of the free energy that the molecule added at this site with the same orientational state as monomers in the complex would have. The sites that have negative values of fixing potential will be referred to as good sites.

Diffusion of the protein molecules in the solution was simulated by the random walk of points in the lattice. At the beginning of simulation, the defined number of monomers were assigned to random nodes in the lattice. The ratio of the number of monomers and the number of nodes in the lattice was equal to the volume fraction of the protein studied under crystallization conditions. In the given simulation step, the free energy of all monomers was evaluated. Then each monomer with a positive value of free energy was moved to a random neighboring node of the lattice (if that node was not occupied), and its orientational state was randomly changed. Monomers with negative energies were considered immobilized in the solid phase and were not moved. As the distances between the central molecule and surrounding ones are approximately equal to the diameter of the molecule, the time step of the simulation was calculated as the time required by the molecule to move a distance equal to its diameter, using the equation for mean squared

TABLE 1 Four-dimensional vectors defining positions and orientations of the neighboring molecules in the lattice

Coordinates of the starting molecule*	Coordination of the neighboring molecules which, with the respective central molecule, form distances shorter than 4.5 Å
(0, 0, 0, 1) [X, Y, Z]	a. (0, 0, −1, 1)
	b. (0, 0, 1, 1)
	c. (0, 0, −1, 3)
	d. (0, 0, 0, 3)
	e. (−1, 0, −1, 4)
	f. (−1, 0, 0, 4)
	g. (0, 0, 1, 7)
	h. (0, 0, 0, 8)
(0, 0, 0, 2) [−X, −Y, Z + 0.5]	a. (0, 0, −1, 2)
	b. (0, 0, 1, 2)
	c. (−1, −1, 0, 4)
	d. (−1, −1, 1, 4)
	e. (0, −1, −1, 3)
	f. (0, −1, 0, 3)
	g. (0, 0, 1, 8)
	h. (0, 0, 1, 7)
(0, 0, 0, 3) [−Y + 0.5, X + 0.5, Z + 0.75]	a. (0, 0, −1, 3)
	b. (0, 0, 1, 3)
	c. (0, 1, 0, 2)
	d. (0, 1, 1, 2)
	e. (0, 0, 0, 1)
	f. (0, 0, 1, 1)
	g. (0, 0, 1, 5)
	h. (0, 0, 1, 6)
(0, 0, 0, 4) [Y + 0.5, −X + 0.5, Z + 0.25]	a. (0, 0, −1, 4)
	b. (0, 0, 1, 4)
	c. (1, 0, 0, 1)
	d. (1, 0, 1, 1)
	e. (1, 1, −1, 2)
	f. (1, 1, 0, 2)
	g. (0, 0, 1, 6)
	h. (0, 0, 0, 5)
(0, 0, 0, 5) [−X + 0.5, Y + 0.5, −Z + 0.75]	a. (0, 0, 1, 5)
	b. (0, 0, −1, 5)
	c. (0, 1, 1, 7)
	d. (0, 1, 0, 7)
	e. (1, 1, 1, 8)
	f. (1, 1, 0, 8)
	g. (0, 0, −1, 3)
	h. (0, 0, 0, 4)
(0, 0, 0, 6) [X + 0.5, −Y + 0.5, −Z + 0.25]	a. (0, 0, 1, 6)
	b. (0, 0, −1, 6)
	c. (1, 0, 0, 8)
	d. (1, 0, −1, 8)
	e. (0, 0, 1, 7)
	f. (0, 0, 0, 7)
	g. (0, 0, −1, 4)
	h. (0, 0, −1, 3)
(0, 0, 0, 7) [Y, X, −Z]	a. (0, 0, 1, 7)
	b. (0, 0, −1, 7)
	c. (0, 0, 0, 6)
	d. (0, 0, −1, 6)
	e. (0, −1, 0, 5)
	f. (0, −1, −1, 5)
	g. (0, 0, −1, 1)
	h. (0, 0, −1, 2)
(0, 0, 0, 8) [−Y, −X, −Z + 0.5]	a. (0, 0, 1, 8)
	b. (0, 0, −1, 8)
	c. (−1, −1, 0, 5)
	d. (−1, −1, −1, 5)
	e. (−1, 0, 1, 6)
	f. (−1, 0, 0, 6)
	g. (0, 0, −1, 2)
	h. (0, 0, 0, 1)

* Crystallographic symmetry operators like in space group $P4_32_12$ are given in square brackets.

displacement of a Brownian particle:

$$\Delta t = \langle (x)^2 \rangle / 6D \quad (3)$$

The values of the diffusion constant D (102 $\mu m^2\ s^{-1}$) and the corresponding hydrodynamic radius a (2.09 nm) were taken from Eberstein et al. (1994). The mean square displacement $\langle (x)^2 \rangle$ was assumed to be equal to $4a^2$. The hydrodynamic radius was used instead of the average radius of a molecule, as determined according to the crystal structure. The rationale is that the actual Brownian particle that moves in the solvent is a protein molecule surrounded by the hydratation shell. The calculated time step of our simulation is 28 ns.

Periodic boundary conditions were set by simply "wrapping the lattice around." This means that nodes occupying one side of the lattice were connected to the nodes occupying the opposite side.

APPLICATION OF THE MODEL AND RESULTS

The tetragonal crystal form of hen egg white lysozyme has been chosen as a model system. The crystal environment was generated using coordinates stored in the 5LYZ entry (Diamond et al., 1974) of the Protein Data Bank. In the complex obtained, the central molecule was surrounded by eight neighbors. Table 2 lists values of accessible surface areas buried in the interfaces and corresponding interaction energies. In all simulations, the value of the entropic penalty was arbitrarily set to 12 kcal/mol. The value of entropy loss upon complex formation is a subject of debate among different authors. Experimental studies suggest values varying between 7 and 15 kcal/mol at room temperature (Erickson and Pantaloni, 1981; Erickson, 1989). The calculations using methods of statistical mechanics estimate the entropy loss at ~15 kcal/mol (Janin, 1995). In general, values ranging from 0 up to 30 kcal/mol are used in the literature. The choice of 12 kcal/mol will also be discussed later.

Behavior of the dislocation on the crystal face

To test the model, we attempted to reproduce the movement of the dislocation on the (110) face of the crystal, which was observed in detail by the atomic force microscopy method of Konnert et al. (1994). The authors reported that the face grows by the movement of the dislocation, which is two molecules high. The lower layer of monomers was observed to be more extended than the upper one. We have measured the positions of the step edge on the pictures made by Konnert et al. at 1-min intervals. The constant speed estimated according to these measurements is equal to 0.3 μm/min.

TABLE 2 Surface areas buried in the interfaces and calculated association free energies

Letter code of the interface*	Surface area buried in the interface (Å^2)	Calculated free energy of association (kcal/mol)
a	37.5	−0.6
b	37.5	−0.6
c	548.5	−12.0
d	341.3	−7.6
e	341.3	−7.6
f	548.5	−12.0
g	1104.4	−21.9
h	657.0	−16.5

* The letter codes of interfaces correspond to letter codes of neighboring nodes in Table 1.

In the lattice with the size 15 × 15 × 15 unit cells, the layer of unit cells parallel to the (110) face was filled with monomers. Then an additional layer of unit cells with dimensions 2 × 15 was added to the surface of the previous one. As can be seen in Fig. 2, the height of a step built in such a way is two molecules.

At the start of the simulation, 216 monomers were added at random positions of the lattice, reproducing the protein concentration of 5 mg/ml that Konnert et al. used in their experiments. In the course of the simulation, whenever the monomer was immobilized in the solid phase, a new one was added randomly to the lattice to keep the concentration constant.

As the simulations proceeded, points representing monomers aggregated at the edge of the step but not on the flat surface of the crystal. Fig. 2 shows the configuration of the monomers in the lattice after simulation. As can be seen, the lower layer is more extended than the upper one. Thus we conclude that our model qualitatively reproduces the behavior of the (110) face of a tetragonal lysozyme crystal.

Several simulations were performed with different numbers of orientational states. We have found that in our simulations, the step moves with reasonable speed if the number of orientational states is set to 10^5. After 9.7×10^7 iterations of this simulation, 160 monomers were fixed at the edge of the step. The maximum range of the step is three unit cells. The distance between corresponding atoms in the unit cells in the direction of the step movement is 112 Å. Thus the dislocation moved ~336 Å during the simulation. As the step moved with the constant speed, we simply divided this distance by the number of program iterations multiplied by the time step of the simulation. We obtained the value 0.7 μm/min.

We believe that agreement within the order of magnitude with the experimentally determined speed of the step (0.3 μm/min) is sufficient to conclude that our model reproduces the kinetics of movement of the dislocation on the (110) face of tetragonal lysozyme crystals.

Nucleation

It was not possible to obtain in a reasonable time any stable nuclei in the lattice, by using the random-walk approach with the number of orientational states equal to 10^5. To find the size of the smallest stable complex and to estimate the time of its appearance, the following calculations were executed.

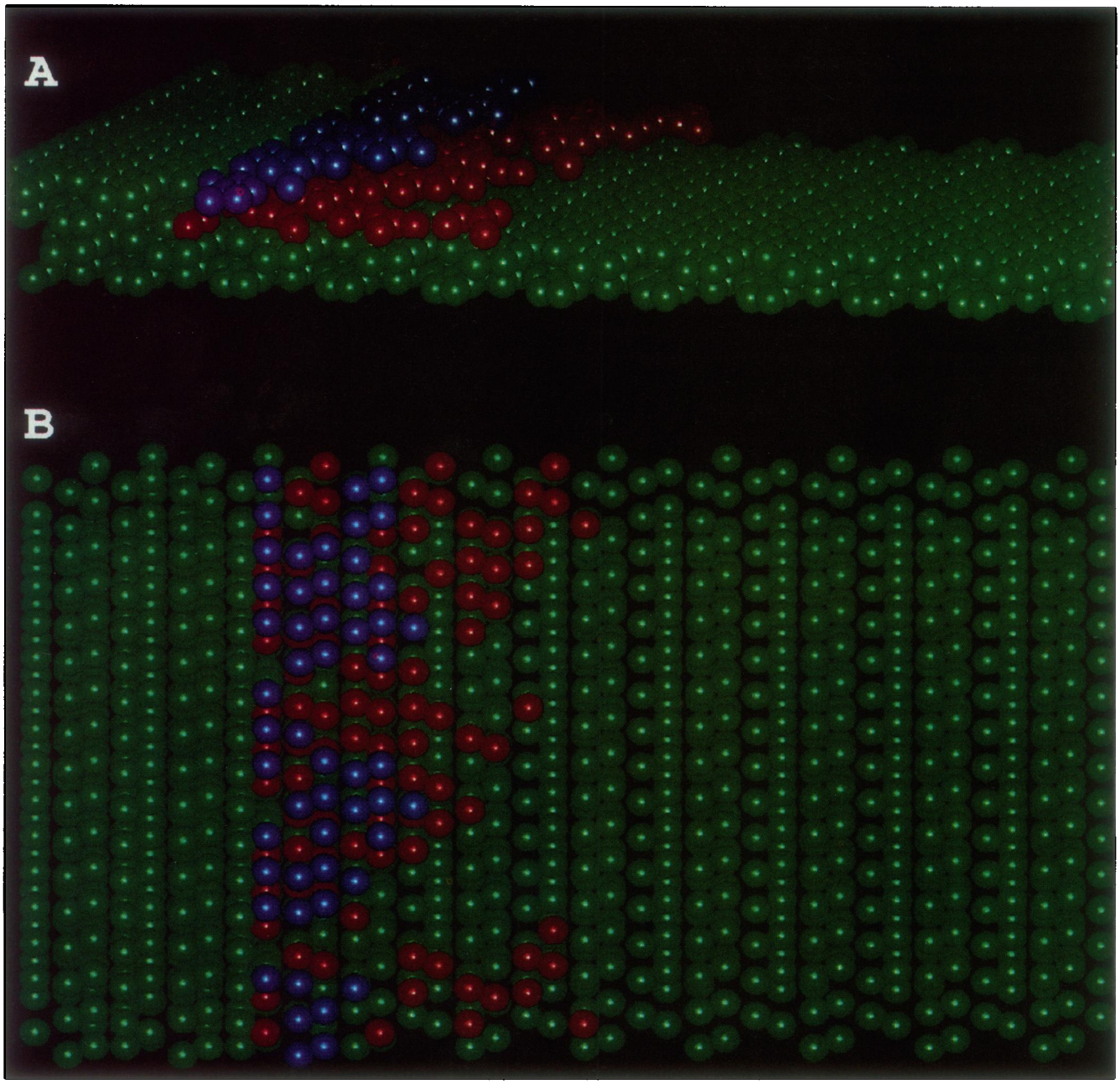

FIGURE 2 Configuration of the points in the lattice before and after simulation of the behavior of the dislocation on the (110) face of the crystal. (*B*) Top view of the configuration shown in *A*. The positions of the protein molecules in the lattice are represented by spheres. Green spheres represent the (110) face of the crystal with a two-molecule-high dislocation, which was built into the lattice before simulation. Positions of all molecules that had negative energies after simulation are represented by blue and red spheres. Red was used for molecules that aggregated in the lower layer; blue was used for molecules that aggregated in the upper layer. Molecules that were not immobilized in the solid face are not shown.

First, random 3-mers, 4-mers, . . . , 10-mers were generated in the following way. The first monomer ($k = 1$) was added at the center of the lattice. Then a complex of the size $k + 1$ was created by adding the monomer at a random site of the complex of the size k. A total of 10^7 3-mers, 4-mers, . . . , 10-mers were built. For each complex generated, the free energies were evaluated. If all molecules had negative energies, the aggregate was considered stable and its number of good sites was evaluated. The fraction of complexes with good sites was then calculated as the ratio of the number of complexes with at least one good site on its surface to the number of stable complexes.

We have not found any complex with a size smaller than four molecules that met the condition that all monomers have negative energy. To further support this result, all possible trimers were generated and examined. None of them were stable. None of the stable tetramers had any good sites on its surface. There were no stable pentamers. The

good sites could be found on the surfaces of the stable hexamers and the greater complexes. For 6-mers, 7-mers, 8-mers, 9-mers, and 10-mers, the fractions of complexes with good sites were 0.15, 0.35, 0.51, 0.82, and 0.74, respectively. Thus the smallest stable complex is the tetramer, but it cannot grow further without the additional nucleation of two monomers on its surface.

To estimate the time scale of appearance of the first stable complex, we ran random-walk simulations, with the number of orientational states set to 1. The number of molecules put into the lattice of the size $15 \times 15 \times 15$ was set to 1000 to reproduce the volume fraction 0.037, which has been found by Georgalis et al. (1995) to be optimal for lysozyme crystallization in 0.54 M NaCl and 0.1 M Na-acetate buffer (pH 4.2). Then random-walk steps were performed until the stable tetramer was obtained. This experiment was repeated 1000 times. The average number of iterations required to obtain a stable tetramer was 26. The time scale of nucleation was then estimated in the following way. If the number of orientational states is set to N, any k-mer formed in the lattice is stable if all of its monomers have chosen the same of all N orientational states. Furthermore, if the volume of the system is increased n times, the expected time of occurrence of the stable k-mer becomes n times shorter. Thus we estimate the expected time of occurrence of the stable k-mer in the lattice with the V nodes and number of orientational states set to N, denoted as $T(k, V, N)$, using

$$T(k, V, N) = jN^{k-1} tV_0/V \quad (4)$$

where j is the average number of program iterations required to obtain a stable k-mer in the lattice with V_0 nodes, the number of orientational states is set to 1, and t is the time step of the simulation.

As we assume that one node of our lattice corresponds to the volume of a sphere with a radius equal to the hydrodynamic radius of a lysozyme molecule (9.1293 nm^3), 1 ml of solution is represented by a lattice having 2.62×10^{19} nodes. Using the results of the simulations described above and setting the number of orientational states to 10^5, we estimate the time required for the stable tetramer to appear in a 1-ml drop of solution as 0.76 μs.

Early stages of growth

Simulations of the early stages of growth started from the stable tetramer placed in the center of the lattice and were carried out using the random-walk scheme. The size of the lattice was set to $15 \times 15 \times 15$ unit cells, and the number of molecules was set to 1000, which corresponds to volume fraction 0.037. Because of the fact that nucleations on the surfaces of small complexes are required for growth, it was not possible to perform simulations on a reasonable time scale when the number of orientational states was set to 10^5. For that reason, simulations were executed for the number of orientational states set to 100. This value was high enough to ensure that new tetramers are not formed in the lattice and that nucleation of the two monomers occurs only on the surfaces of complexes without good sites.

During calculations, the size of the growing complex and the number of simulation steps were recorded whenever a new molecule(s) was immobilized on the surface of the growing nuclei. Then the number of simulation steps required to obtain a complex of the given size, if the orientational probability is 10^5, was calculated using the following expression:

$$t(N) = \sum \Delta t_{sim} (O/O_{sim})^{\Delta S} \quad (5)$$

The summation runs over all first N observations of the size and time collected during simulation, Δt_{sim} and ΔS are the time and size differences between subsequent observations, and O/O_{sim} is the ratio of the true number of orientational states (10^5) and the value applied in the simulation. The expression simply says that as the true number of orientational states is O/O_{sim} times higher than the value applied in the calculations, the time required for the attachment of ΔS monomers will be $(O/O_{sim})^{\Delta S}$ times longer that the time observed during simulation. As we considered the growth of single nuclei, the volume of the crystallization batch was not included in the expression.

According to our results, the time required for the tetramer to grow to the complex of 50 molecules was 8×10^{12} simulation steps, which corresponds to 228×10^3 s. The time scale was determined by the requirement for nucleations on the surfaces of complexes that lack good sites. On the "growth trajectory" leading from the tetramer to the complex of size 50, there were, on the average, about eight complexes lacking good sites.

DISCUSSION

Parameters of the model

Except for the experimentally known diffusion constant and protein radius, the parameters of our model are interaction energies, entropic penalty, and the number of orientational states.

As written above, interaction energies were calculated assuming that the free energy of association of protein monomers is proportional to the area buried in the interface between monomers. The validity of this approach applied to calculations of the free energy of specific protein complexes was discussed in detail by other authors (Horton and Lewis, 1992; Juffer et al., 1995). The question that should be asked here is whether the approach is still valid when we consider the solution under crystallization conditions. To answer this question, we must consider the dependence of the model on its parameters.

It is clear that parameters that determine the size of the smallest stable complex and the number of good sites on the surfaces of complexes are not values of interaction energies alone, but the differences between interaction energies and the entropic penalty. For the values of these differences that correspond to entropic penalties lower than 11 kcal/mol, the

dimers become stable. There is evidence which suggests that the dimer intermediate is not formed during the nucleation and growth of a tetragonal lysozyme crystal. First, in the light-scattering experiments (Georgalis et al., 1995), the most populated fraction observed in the solution is monomers. This simply shows that lysozyme is not a dimer under crystallization conditions.

The other evidence is provided by the observation of growth of the (110) face. If all kinds of dimers were intermediates on the lysozyme crystallization path, the single molecules could form favorable interactions on the flat surfaces of the (110) crystal face. If that is true, the atomic force microscopy observations would show that the face is rough as it incorporates single molecules and the growth of the step would not be observed. It is also possible that only the most stable dimer is the intermediate on the crystallization pathway. Stability of this dimer does not imply incorporation of single monomers on the (110) face, as the patches of lysozyme surface that are required to form the interface of this dimer (interface g in Table 2) are not exposed on the surface of the face. In this case the step would grow by the incorporation of such dimers. Fig. 3 shows the dimer on the edge of the step. One molecule lies in the lower layer, and a second in the upper one. Thus, in the case of growth by incorporation of this kind of dimers, the two layers would have equal lengths, which is not the case, according to atomic force microscopy observations.

There are also other observations that strongly suggest that dimers are not formed during lysozyme crystallization. Janin and Rodier (1995) analyzed surfaces of protein-protein interfaces found in 152 crystals and compared them with the surfaces of interfaces in the randomly generated complexes. They have shown that in many crystals interface areas were in the same range as in the random complexes (i.e., below 1200 Å^2). As can be seen in Table 2, the tetragonal lysozyme crystal falls into this category. It is very unlikely that the dimers with interface surface areas comparable to the random ones are stable in solution. This is further justified by the random energy model concept, which was recently applied by Janin (1996) to an analysis of protein complexes.

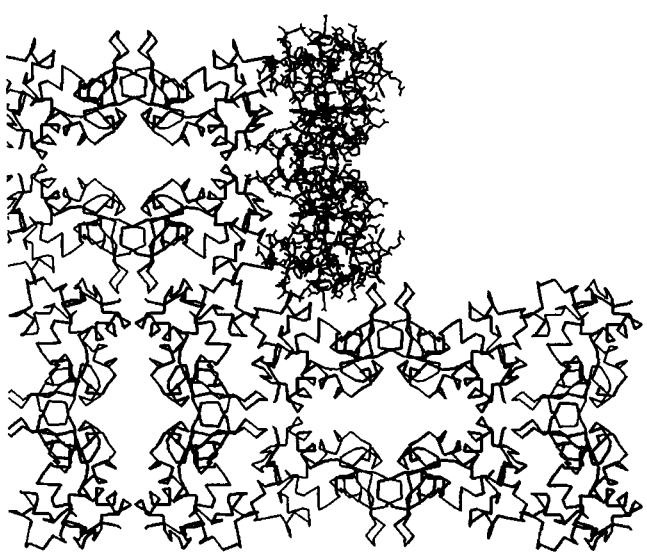

FIGURE 3 Dimer with the largest monomer-monomer interface area placed on the edge of the dislocation. Molecules of the crystal face are represented by the traces of the main chains. Molecules of the dimer are shown in an "all-atom" representation.

The largest differences between interaction energies and the entropic penalty at which qualitative behavior of the step is in agreement with experimental data corresponds to the entropic penalty value of 19 kcal/mol. Above this value, the step requires nucleation of two molecules to grow, as there are no good sites on its edge. To check how this alters the behavior of the step, additional simulations of its movement were performed. Entropic penalties were set at 20 kcal/mol in both cases, and numbers of orientational states were set to 100 and 1000. In the simulation with the number of orientational states set to 100, the face grew because of nucleations on the flat surfaces. The second simulation was stopped after no stable nuclei were observed on the flat surface of the face or on the edge of the step after 5.6×10^7 iterations of the program. As one can see in this case, the time scale of the process is inconsistent with experimental data. If the number of iterations reaches 10^7, at least one layer of monomers is expected to appear on the edge of the step to reproduce the experimentally determined speed of the step movement.

As shown above, our model reproduces qualitative behavior of the step for the values of differences between interactions and entropic penalty corresponding to entropic penalties in the range of 11–19 kcal/mol. The speed of the movement of the step for this parameter is governed by the number of orientational states. The value of this parameter for which the model reproduces experimental speed of dislocation, within the order of magnitude, is 10^5. The value of 10^5 lies within the range of orientational probabilities calculated from the Brownian dynamics simulations by Northrup and Erickson, 1992.

An additional constraint on the parameters is provided by the analysis of the time scales of nucleation. For the entropic penalty value of 13 kcal/mol, the smallest stable complex is the hexamer. If the number of orientational states is set to 1, the average number of program iterations required to nucleate a hexamer is 2476. For the number of orientational states set to 10^5, Eq. 4 yields a time of nucleation equal to 7.2×10^9 s, which is longer than the time of lysozyme crystal growth. Thus for the interaction energies listed in Table 2, the entropic penalty cannot exceed the value of 12 kcal/mol.

As follows from the above discussion that the applied, arbitrary values of energetic parameters are within the range adequately describing the experimentally known facts. Moreover, application of different parameters from this range does not significantly affect conclusions about the early events in crystallizing protein solutions.

Mechanisms of lysozyme crystallization

The size of the first complex that has the crystal order and in which interaction energies overcome entropic penalties for immobilization of molecules is four monomers. Appearance of this complex requires the encounter of four properly oriented molecules in space. As the number of distinct mutual orientations is large (10^5), the probability of tetramer formation is very small. Thus the rate-limiting factor during nucleation is the anisotropy of protein molecules. Despite the very low probability of their formation, the first stable nuclei appear on a time scale of 10^{-7} s. This time scale is dependent on the volume of the crystallization batch. In greater volumes, the number of "attempts" to form the first stable complex is higher. The time scale of 10^{-7} s is the result for a volume of 1 ml.

The tetramer has no sites at its surface at which a single monomer could form interactions that allow it to overcome the entropic penalty of aggregation. Thus the tetramer can grow only by additional nucleation of the two monomers on its surface.

Several simulations of growth performed by us show that the time scales of formation of larger complexes are not critically dependent on which particular aggregates in the growth pathway require nucleation on the surface. It should also be noted that larger complexes, on the average, have a higher probability of possessing good sites on their surface. At present we cannot exclude the possibility of existence of pathways that are much faster than those obtained in our simulations, i.e., pathways in which, for geometrical reasons, each complex has a good site. Examining this possibility will be the subject of further study in our laboratory. Let us note, however, that the picture obtained in our simulations is consistent with the interpretation of Bessho et al. (1994), based on fitting equations from the polynuclear growth theory to experimental data on lysozyme crystallization. The predicted sizes of the smallest stable nuclei were three or four molecules, and the value of the parameter describing the order of growth kinetics suggested surface nucleations or dimer adhesions. The sizes of compact nuclei present in crystallizing lysozyme solutions on the minute time scale determined by Georgalis et al. (1995) are also in at least qualitative agreement with the model presented in this paper.

In the very early stages of growth described here, the fraction of higher order aggregates in solution is so low that the possibility of their direct interaction can be excluded. The mechanism of intermediate stages of crystal growth is probably more complex and will be the subject of further development of the model.

CONCLUSIONS

The model presented in this paper concentrates on specific protein-protein interactions leading to the formation of structured nuclei in crystallizing protein solutions. We have intentionally excluded the nonspecific aggregation phenomena leading to, e.g., fractal formation in solution (Georgalis et al., 1995), as the possibility of such structures rearranging to highly ordered crystals seems unlikely. It has been shown that accounting for specific interactions alone can describe both the experimentally known behavior of the face growth as well as predict the size of the critical nucleus. Moreover, the very early stages of growth of the nuclei can be qualitatively described. It is suggested that the early stages of growth of the ordered phase are strongly determined by the geometrical constraints. If so, one can expect the differently ordered crystals of various protein molecules to have different scenarios of growth. We believe that analyzing this diversity of mechanisms for known crystal molecules and relating these data to the solution content affecting monomer-monomer interactions will provide further insights into the mechanisms of protein crystallization.

We are grateful to Drs. Y. Georgalis and D. Płochocka for critical comments on the manuscript.

REFERENCES

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., and M. D. Brice. 1977. The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bessho, Y., M. Ataka, M. Asai, and T. Katsura. 1994. Analysis of the crystallization kinetics of lysozyme using a model with polynuclear growth mechanism. *Biophys. J.* 66:310–313.

Chothia, C., and J. Janin. 1975. Principles of protein-protein recognition. *Nature.* 256:705–708.

Collaborative Computational Project, Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* D50:760–763.

Diamond, R., D. C. Phillips, C. C. F. Blake, and A. C. T. North. 1974. Real-space refinement of the structure of hen egg-white lysozyme. *J. Mol. Biol.* 82:371–382.

Durbin, S. D., and G. Feher. 1990. Studies of crystal growth mechanism of proteins by electron microscopy. *J. Mol. Biol.* 212:763–774.

Eberstein, W., Y. Georgalis, and W. Saenger. 1994. Molecular interactions in crystallizing lysozyme solutions studied by photon correlation spectroscopy. *J. Cryst. Growth.* 143:71–78.

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.

Erickson, H. P. 1989. Co-operativity in protein protein association. The structure and stability of the actin filament. *J. Mol. Biol.* 206:465–474.

Erickson, H. P., and D. Pantaloni. 1981. The role of subunit entropy in cooperative assembly-nucleation of microtubules and other two dimensional polymers. *Biophys. J.* 34:293–309.

Georgalis, Y., P. Umbach, J. Frank, D. M. Soumpasis, and W. Saenger. 1995. Protein crystallization screening through scattering techniques. *Adv. Colloid Interface Sci.* 58:57–75.

Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein. Sci.* 1:169–181.

Janin, J. 1995. Elusive affinities. *Proteins.* 21:30–39.

Janin, J. 1996. Quantifying biological specificity: the statistical mechanics of molecular recognition. *Proteins.* 25:438–445.

Janin, J., and F. Rodier. 1995. Protein-protein interaction at crystal contacts. *Proteins.* 23:580–587.

Juffer, A. H., F. Eisenhaber, S. J. Hubbard, D. Walther, and P. Argos. 1995. Comparison of atomic solvation parametric sets: applicability and limitations in protein folding and binding. *Protein Sci.* 4:2499–2509.

Konnert, J. H., P. D. Antonio, and K. B. Ward. 1994. Observation of growth steps, spiral dislocations and molecular packing on the surface of lysozyme crystals with the atomic force microscope. *Acta Crystallogr.* D50:603–613.

Malkin, A. J., Yu. G. Kuznetsov, T. A. Land, J. J. DeYoreo, and A. McPherson. Mechanisms of growth for protein and virus crystals. 1995. *Nature Struct. Biol.* 2:956–959.

Meakin, P. The growth of fractal aggeragates and their fractal measures. 1988. *Phase Transitions.* 12:336–480.

Muschol, M., and F. Rosenberger. 1995. Intercactions in undersaturated and supersaturated lysozyme solutions: static and dynamic light scattering results. *J. Chem. Phys.* 103:1004–1012.

Niimura, N., Y. Minezaki, M. Ataka, and T. Katsura. 1995. Aggregation in supersaturated lysozyme solution studied by time-resolved small-angle neutron scattering. *J. Cryst. Growth.* 154:136–142.

Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulations. *Proc. Natl. Acad. Sci. USA.* 89:3338–3342.

Patro, S. Y., and T. M. Przybycien. 1996. Simulations of reversible protein aggregate and crystal structure. *Biophys. J.* 70:2888–2902.

Richmond, T. J. 1984. Solvent accessible surface area and excluded volume in proteins. *J. Mol. Biol.* 178:63–89.

Tissen, J. T. W. M., J. Drenth, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1994. Microhydrodynamics simulation of protein crystallization. I. Static calculation. *Biophys. J.* 67:1801–1805.