ELSEVIER

Biophysical Chemistry 77 (1999) 123–137

Biophysical Chemistry

Lattice simulations of protein crystal formation

Andrzej M. Kierzek[a], Piotr Pokarowski[b], Piotr Zielenkiewicz[a,*]

[a] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland
[b] Institute of Mathematics, Polish Academy of Sciences, Sniadeckich 8, 00-950 Warsaw, Poland

Received 25 August 1998; received in revised form 8 January 1999; accepted 2 February 1999

Abstract

A new algorithm is presented for the lattice simulation of protein crystal growth. The algorithm allows the calculation of the size distribution of microcrystals in the volume and timescale of experiments and within the framework of the previously-published microscopic model [A.M. Kierzek, W.M. Wolf, P. Zielenkiewicz, Biophys. J. 73 (1997) 571–580]. Simulations for the tetragonal lysozyme crystal show that there are two critical sizes in the development of ordered phase. The first one corresponds to the size of the smallest stable complex which, in the case of the tetragonal lysozyme crystal, is the particular tetramer. In a volume of 5 μl the tetramer appears in the millisecond timescale. The second critical radius of approximately 100 monomers is only reached by a few of all the smallest stable complexes formed in the solution. The model predicts that out of 10^7 tetramers which appear in solution, only eight reach the size of 100 monomers within 8 h. After exceeding the second critical radius the microcrystals grow to the size of 10^4 monomers in the minute timescale and are thus assumed to quickly lead to macroscopic crystals. The predicted number of crystals formed during 8 h of nucleation is in qualitative agreement with arrested nucleation experiments. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Protein crystallisation; Lattice simulations; Lysozyme

1. Introduction

In spite of the recent advances in protein crystallisation, such as surface response methods for optimising crystallisation conditions [1] or the modification of charged residues [2], crystal growth remains the slowest step in protein structure determination by means of X-ray crystallography. A better understanding of the mechanisms of protein crystal growth would lead to the optimisation of the currently used crystallisation methods. For example, the knowledge about the size distribution of aggregates during the crystallisation of some model proteins would help to improve protocols for monitoring the process by means of scattering methods. This, in turn, would shorten the necessary time to decide if the protein would crystallise under the given conditions.

---

* Corresponding author. Tel.: +48-22-6584-703; fax: +48-22-6584-636; e-mail: piotr@ibbrain.ibb.waw.pl

0301-4622/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII: S0301-4622(99)00019-8

Many experimental approaches, such as electron microscopy [3], atomic force microscopy [4], light and neutron scattering [5–8] and calorimetry [9], have been used to study crystallisation of a few model proteins. There has also been a significant effort on the theoretical side. To understand the effective interactions controlling charged protein aggregation, Soumpasis and Georgalis [10] computed the potential of mean force for the lysozyme–water–NaCl system within a hypernetted chain approximation. Many authors attempted to interpret the data on protein crystallisation using physicochemical concepts developed for the description of the solutions of small molecules, for example, applications of classical nucleation theory [11,12] and the recent work of Wolde and Frankel [13] where protein crystal nucleation was modelled as the process of critical density fluctuation.

The theories borrowed from small molecule crystal growth and colloidal physics do not take into account the dependence of the free energy of protein–protein association on the complex shapes of the protein molecules. If the shapes of the protein monomers are not explicitly included in the model then the well known fact that associating proteins tend to form interfaces with the largest surface area buried [14] is neglected. Also, the orientational effect, i.e. the probability that protein monomers which are close together in space are properly oriented to form a defined interface is omitted. As the reported values of orientational probabilities vary in the range 10^{-5}–10^{-7} [15–18], they influence the kinetic results by several orders of magnitude.

The above simplifications can be avoided if detailed knowledge about the structure of the crystal is used to provide insights into the mechanism of its growth. One way in which this is attempted is the statistical analysis of the properties of protein–protein interfaces [19–21]. Another is to build the model of crystal growth in such a way that the anisotropy of protein molecules (generally or the detailed shape of the model protein) is considered. In 1996, Patro and Przybycien [22] performed 2D Monte Carlo simulations in which protein monomers were represented by hexagons, the edges of which modelled surface patches with different properties. Very general aspects of protein aggregation were studied by this approach. Recently, Pellegrini et al. [23] were able to explain the non-uniform distributions of space groups found in the Brookhaven Protein Data Bank [24] by Monte Carlo simulations in which nuclei were assembled from three dimensional rigid bodies with different surface patches. In 1991, Durbin and Feher [16] reproduced the behaviour of the faces of a tetragonal lysozyme crystal by the Monte Carlo simulation in which the interactions on three main types of protein–protein contacts in the crystal were used as the free parameters of the model. The most detailed representation of the protein structure was used by Tissen at al. [25]. They used triangulated surfaces of the model proteins and performed Stokesian dynamics and continuum hydrodynamics calculations. Due to the computational complexity of the model, only results concerning the diffusion coefficients of the proteins under investigation were reported so far.

In our previous paper we used lattice simulations to study nucleation and early growth stages of a tetragonal lysozyme crystal [18]. The protein molecules were modelled as points occupying the nodes of a 3D lattice. The edges of the lattice represented contacts between protein molecules in the crystal. Contact energies were calculated assuming that they are proportional to the change of accessible surface area during the formation of the interface. A discrete orientational state was assigned to each of the monomers in the lattice and interactions were considered only between the molecules which were properly oriented. Such a representation of the system allowed us to use interaction energies calculated according to the analysis of the crystal structure on the atomic level, to account for orientational effects and, at the same time, to avoid the computational complexity of explicit protein structure treatment during random walk simulations.

The model was used to reproduce the behaviour of the dislocation on the (110) face of a tetragonal lysozyme crystal which was observed in detail by atomic force microscopy [4]. In our previous paper it was shown that the behaviour of the dislocation can only be reproduced for the very narrow range of the difference between the entropic penalty and interaction energies and for the orientational probability (reciprocal of the number of discrete orientational states) 10^{-5}. When applied to nucleation, the model showed that the smallest stable complex with the crystal order is the particular tetramer (see Fig. 1) and that it appears in the solution in the microsecond time scale. Even for the representation of the system used in our work it was not possible to generate ‘growth trajectories’ for a large number of tetramers in the time scale of the experiments.

This work presents a new algorithm which allows the application of the above model to the calculation of the size distribution of microcrystals in the time scale of hours. Its application to a tetragonal lysozyme crystal shows that only a small fraction of independently growing first stable complexes (tetramers) reaches the critical size of approximately 100 monomers and then become macroscopic crystals. The predicted number of macroscopic crystals obtained after 8 h of nucleation is in reasonable agreement with the results of arrested nucleation experiments [26,27].

2. Methods

Although detailed formulation of our model is given in the previous paper [18], it is necessary to restate its basic assumptions in order to describe the new algorithm.

2.1. The model

The simulation begins with an atomic level analysis of the environment of the protein monomer in the crystal under investigation. The environment is defined as the complex containing the protein monomer, which will be referred to as the central molecule, and all molecules in the crystal which form with it at least one intermolecular contact shorter than 4.5 Å. The interaction free energy is calculated for each of the interfaces formed by the central molecule. It is assumed that the interaction free energy is proportional to the change of accessible surface area during the formation of the interface. Details concerning the generation of the crystal environment and energy calculations are given in the previous paper; Table 1 lists the values obtained for tetragonal lysozyme crystal.

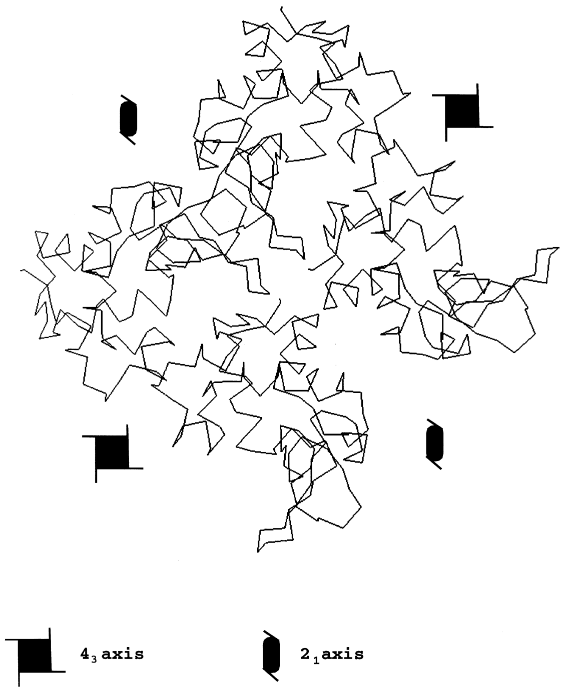

Fig. 1. The tetramer which is the smallest stable complex of the tetragonal lysozyme crystal. The complex is shown along *c* crystal direction. Positions of crystalographic axes are shown on the picture.

Table 1
Surface areas buried in the interfaces and calculated association free energies for the tetragonal lysozyme crystal

Letter code of the interface[a]	Surface area buried at the interface (Å^2)	Calculated association free energy (kcal mol^{-1})
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

[a] There are eight interfaces in the crystal environment of the molecule in the tetragonal lysozyme crystal. For details see [18].

In the second stage of the simulation, protein monomers are represented as points occupying nodes of a three dimensional lattice. The lattice is built in such a way that each edge corresponds to one of the interfaces found in the crystal environment. To account for the orientational effects an integer variable representing the discrete orientational state is assigned. Only monomers which occupy neighbouring nodes and have the same orientational state are counted as interacting. Thus, association free energy can be calculated for any monomer in the lattice using the following equation:

$$\Delta G^0_{assoc}(k) = \Delta G^0_{rot,trans} - \sum_j \delta(k,j)\Delta G^0_{inter}(j) \tag{1}$$

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

Let us introduce two definitions which will be useful in further discussions: (i) a *site* on the surface of a complex is any empty node adjacent to the complex; and (ii) a *good site* is any site which meets the condition that a molecule arriving at this site with the same orientational state as the molecules belonging to the complex would have a negative free energy of association.

We have studied several features of the crystallisation process using the above representation of the system. The *smallest stable complex* with crystal order was found as the smallest (with the minimum number of monomers) cluster of monomers in the lattice in which all the monomers had negative association free energies. To simulate crystal growth, the diffusion of protein molecules was modelled as a random walk of points in the lattice. In each step of the simulation, monomers with positive energies were moved to random neighbouring nodes and their orientational states were randomly changed. Monomers with negative energies were considered as immobilised in the solid phase and were not moved. The time step of the simulation was calculated to be 28 ns using the equation:

$$\Delta t = \langle (x)^2 \rangle / 6D \tag{2}$$

The values of the diffusion constant D (102 μm s^{-1}) and the corresponding hydrodynamic radius a (2.09 nm), were taken from Eberstein et al. [5]. The mean squared displacement $\langle (x)^2 \rangle$ was assumed to be equal to $4a^2$. Thus, the time step of our simulation is the average time in which the protein, together with the solvation shell, moves the distance equal to its diameter.

The method is described in detail in our previous paper; Fig. 2 shows the schema of the simulation.

2.2. Limitations of the random walk algorithm

Due to the very low orientational probability, it was not computationally possible to simulate any events requiring the interaction of more than one molecule at the given time step.

In the simulations of the tetragonal lysozyme crystal, a single molecule arriving at a good site of the complex is properly oriented with the probability of 10^{-5}. If the complex lacks good sites, its growth requires the meeting of two properly oriented molecules at neighbouring sites in the same time step. The probability that two molecules will be properly oriented is 10^{-10}, five orders of magnitude lower than the previous one. Thus, the expected number of random walk steps required for achieving only the proper orientation of two monomers in the same time step is of the order of 10^{10}. It is not possible to compute that number of random walk steps in our system (the number which can be computed in a reasonable time is of the order of 10^7).

For the above reasons, a random walk algorithm could only be directly applied to the simulation of the behaviour of dislocation on the crystal face. In this case, the edge of the step always contained good sites. It was not possible to simulate the process of aggregation of four monomers during the formation of the smallest stable complex as this process requires three molecules to have the same orientation (orientational state of the fourth molecule) at the same time step. Additionally, it was also impossible to directly simulate the further growth of the tetramer due to the fact that it lacks good sites and thus the surface nucleation of two molecules is required for its growth. Larger complexes without good sites were also found.

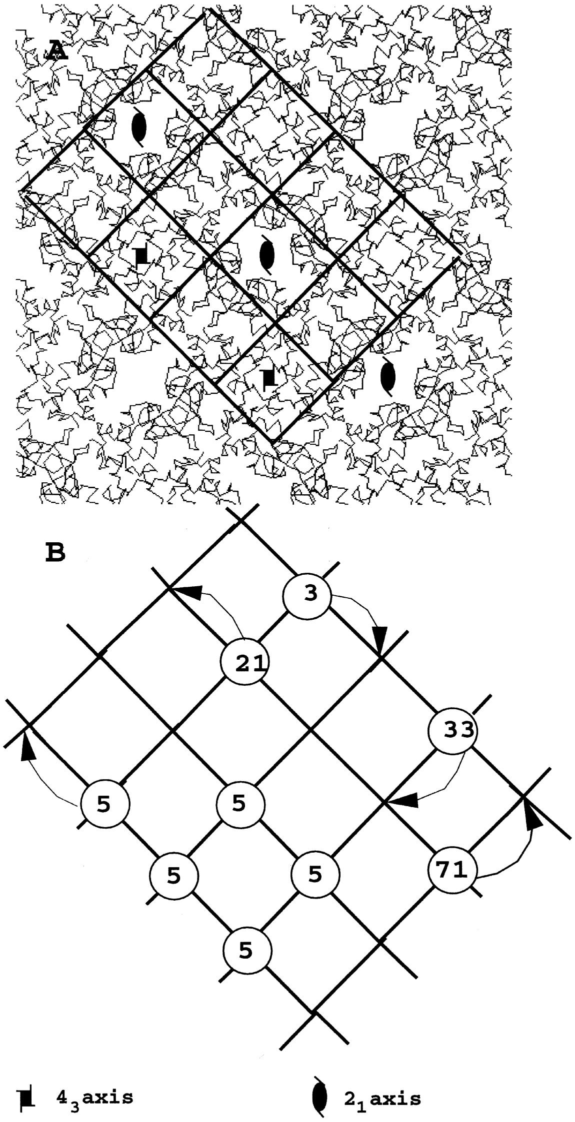

Fig. 2

Schema of the random walk simulation. (a) Two-dimensional slice of the tetragonal lysozyme crystal perpendicular to the c direction. Part of the lattice built according to the analysis of the crystal structure is shown. As one can see the node of the lattice represents the position of the monomer; edges represent intermolecular interfaces. (b) During simulation, monomers are represented as points occupying nodes of the lattice. A discrete orientational state is assigned to each monomer. If the monomers occupy neighbouring nodes and have the same orientational states, the interaction free energy corresponding to the particular interface (lattice edge) is subtracted from the entropic penalty for aggregation. The monomers with positive energies are moved to random neighbouring nodes. The picture shows two monomers which occupy neighbouring nodes, but different orientational states, and are treated as non-interacting. In the cluster of five monomers, four do interact with each other and form a stable complex. The fifth molecule forms only single interaction with others, which is too small to overcome entropic cost. This molecule will be moved in the random walk timestep.

In our previous paper, we showed how the time of the occurrence of the tetramer and the growth curve from the tetramer can be approximated using the results of random walk simulations with higher orientational probabilities. We were able to show only a single ‘growth trajectory’, even using these approximations. In this paper we present a new algorithm to extrapolate results of random walk simulations to longer times, larger systems and smaller orientational probabilities. This algorithm is fast enough to calculate approximately 10^7 growth trajectories and thus it can be applied to the calculation of the size distribution of microcrystals in crystallising lysozyme solutions.

2.3. Time of appearance of the first stable complex

Let us consider a lattice having V_0 nodes and occupied by N_0 monomers with orientational probabilities O_0. Before the smallest stable complex appears, all the configurations of the monomers in which all monomers occupy different nodes of the lattice have equal probability of occurrence in the step of a random walk simulation. Therefore, the probability that the configuration of monomers in the given random walk step contains at least one smallest stable complex is defined as follows:

$$P(O_0,V_0) = k/K \tag{3}$$

where k is the number of configurations which was found to contain at least one smallest stable complex out of K randomly generated ones. Using the value $P(O_0,V_0)$ one can calculate the probability of the occurrence of the configuration containing the smallest stable complex in the same lattice with the same number of monomers if the orientational probability is decreased to O. This is given by the equation:

$$P(O,V_0) = (O/O_0)^{S-1} P(O_0,V_0) \tag{4}$$

where S is the size of the smallest stable complex. The equation states that $P(O_0,V_0)$ is decreased $(O/O_0)^{S-1}$ times as $S-1$ monomers must, during the same time step, align to the remaining one with orientational probability (O/O_0) times smaller than O_0.

The lattice with a large volume having V nodes can be treated as V/V_0 independent lattices containing V_0 nodes. Thus, the probability of finding a configuration containing at least one smallest stable complex in the lattice with V nodes and N monomers with orientational probability O is given by the equation:

$$P(O,V) = 1 - [1 - P(O,V_0)]^{V/V_0} \tag{5}$$

if the volume fraction in the large lattice is equal to the volume fraction in the small lattice $(N/V = N_0/V_0)$.

The appearance of the configuration containing the smallest stable complex, in the single step of a random walk simulation, can be treated as a success of Bernoulli trial with the probability $P(O,V)$. In the sequence of Bernoulli trials, the waiting time for the first success has the geometric distribution [28]:

$$P(t=T) = p(1-p)^T, \quad t = 0,1,2,3,... \tag{6}$$

where t is the waiting time expressed in the number of trials, T is the particular value of this time, and p is the probability of success. The mean of this distribution equals $(1-p)/p$. Therefore, the mean waiting time for the appearance of the smallest stable complex is given by the equation:

$$T(O,V) = [1 - P(O,V)]/P(O,V) \tag{7}$$

The accuracy of the above equations was tested by comparison with the results of the random walk simulations in the following way. First, 10^6 random configurations of 777 monomers were generated in the tetragonal lysozyme lattice of the size $15 \times 15 \times 15$ unit cells (volume fraction 0.029). The orientational probability was set to 1. Among these, 2157 configurations were found to contain at least one smallest stable complex. Using Eqs. (3)–(5) and (7) we calculated the mean waiting time for the appearance of the tetramer in the lattice of the size $30 \times 30 \times 30$ unit cells filled with 6221 monomers (volume fraction 0.029) and having the orientational probability 0.1. The result was 57 950 time steps. Then we performed 321 random walk simulations in the same lattice ($30 \times 30 \times 30$ unit cells, 6221 monomers, orientational probability 0.1). Each simulation was stopped when the first configuration containing at least one smallest stable complex appeared and the number of random walk steps executed was recorded. The mean number of random walk steps which had to be performed in order to obtain the smallest stable complex was 32 390 with a standard deviation of 26 495. The result obtained by application of Eqs. (3)–(5) and (7) lies within the standard deviation of the mean results of random walk simulations.

As one can see, the application of Eqs. (3)–(5) and (7) allows the calculation of the probability of the tetramer occurrence in the system with an eightfold greater volume and orientational probability 10 times smaller than in the system for which random configurations were generated and examined. Therefore, the results of the examination of 10^6 random configurations followed by the application of the above equations can be used to estimate the probability of the appearance of the smallest stable complex in the lattice representing a macroscopic volume and for very low orientational probability.

2.4. Fast algorithm for crystal growth simulation

Our new algorithm simulates growth from the single smallest stable complex. The result of this simulation will be referred to as the *growth trajectory*.

At the beginning of the simulation the smallest stable complex is placed in the lattice and the *time of the simulation* is set to 0. Then, all the sites on the surface of the complex are found. For each site the energy of a properly oriented molecule arriving at the site is calculated in order to determine if it is a good site. Then, each of the good sites are processed in the following way. The appearance of a properly oriented monomer in the good site in the given random walk step is considered as the success of the Bernoulli trial with probability p. Therefore, the number of random walk steps one should wait for the attachment of the monomer to the good site has a geometric distribution with parameter p: $P(t = T) = p(1-p)^T$. For each of the good sites the time of attachment of the monomer at this site is generated as a random number from the above distribution. The value of p was calculated as:

$$p = ON/V \tag{10}$$

which is the product of the orientational probability and the volume fraction of the system. By setting the probability p equal for all good sites, we neglect the fact that the sites with different numbers of unoccupied neighbouring nodes have a different diffusion accessibility in the random walk simulation. This simplification does not significantly influence the results in the case of low orientational probabilities.

Finally, the monomer is added to the good site of the smallest stable complex for which the time of appearance of the monomer was the shortest. This time is added to the time of the simulation. In the next iteration of the algorithm, the list of good sites is updated by checking if any neighbouring node of the last added monomer becomes a good site. The complex grows by reiterating the above steps.

If, in the course of the simulation, a complex without good sites is found the program looks for *surface nucleation sites*. The surface nucleation site is any pair of neighbouring sites for which two molecules, added to these sites with the proper orientation, would form interactions exceeding the entropic penalty for aggregation. The probability of the appearance of two properly oriented molecules at the surface nucleation site is calculated as:

$$p = (O \times N/V)^2 \tag{11}$$

The mean waiting time for the appearance of two molecules in the surface nucleation site under consideration is generated as the random number from the geometric distribution with the parameter p. Then, two monomers are added to the surface nucleation site for which the shortest time was obtained and the time of the simulation is increased by its value.

As a complex having good sites can also grow by the surface nucleation of two monomers, considering the surface nucleation sites only for complexes without good sites is a simplification done for the sake of computational efficiency. For low orientational probabilities, the probability of the attachment of two molecules is several orders of magnitude lower than the probability of the attachment of single one and thus the simplification does not influence results.

For the crystal under investigation, there is no stable complex which does not contain both good sites and surface nucleation sites. Thus, it is not necessary to consider the extremely unlikely event of the surface nucleation of three molecules.

In all simulations, random numbers U from the uniform distribution were generated using the random number generator of Marsaglia et al. [29] with the cycle 2^{144}. The random numbers Z from the geometric distribution were generated using the following formula $Z = \lfloor \beta \ln U \rfloor$ where $\beta = 1/\ln(1-p)$ and p is the parameter of geometric distribution [30].

The summary of the new algorithm is given in Fig. 3.

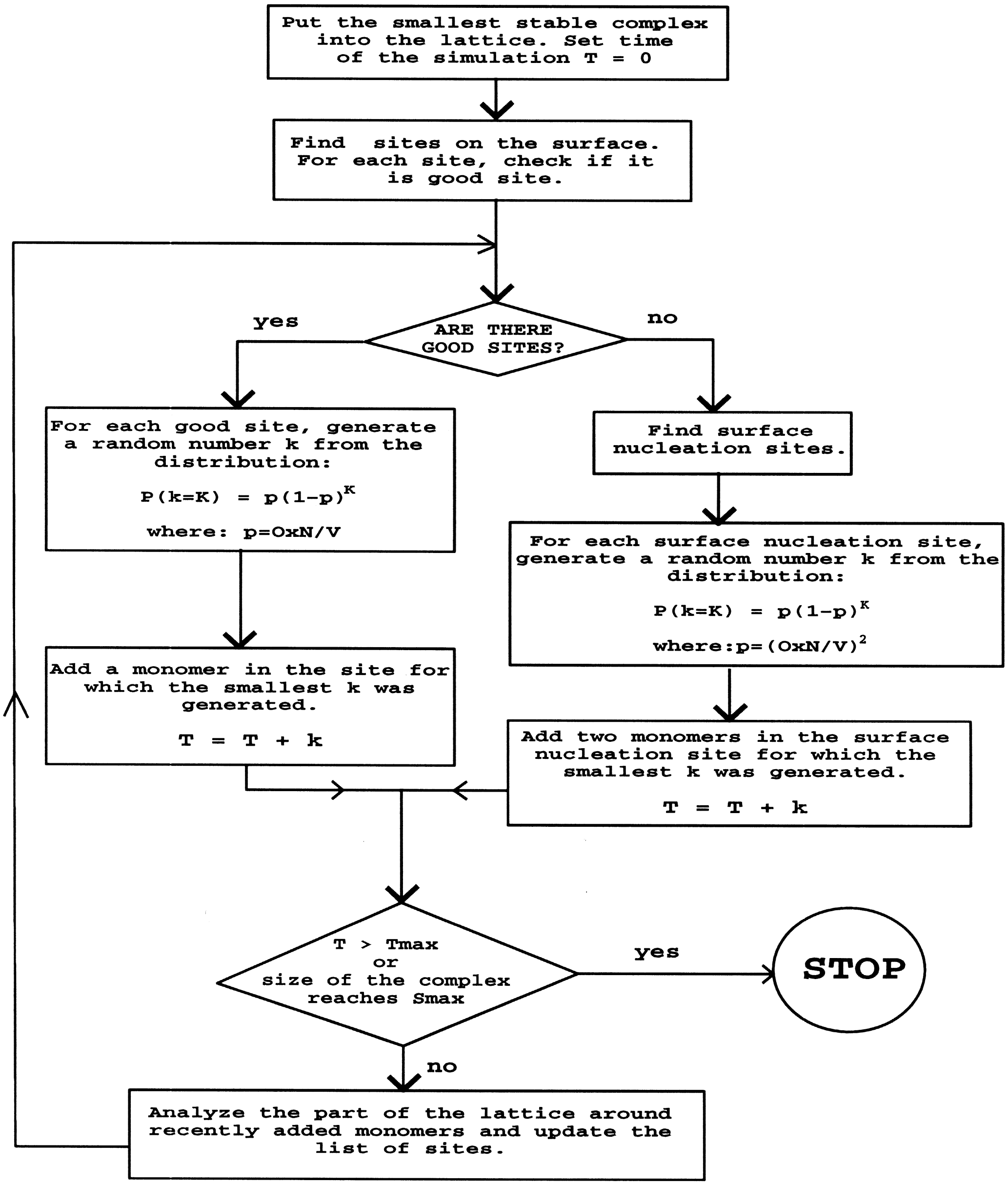

Fig. 3. Flow chart of simulation steps.

2.5. Test of the new algorithm

The new algorithm has been shown to give equivalent results to the random walk simulation by the following test. The tetragonal lysozyme lattice was build with the size of 15 × 15 × 15 unit cells. Then, 21 growth trajectories starting from the smallest stable complex (the tetramer) were calculated using the random walk algorithm. The number of monomers in the lattice was 1000 (volume fraction 0.037) and the orientational probability was set to 0.01. Interaction energies were taken from Table 1 and the entropic penalty was set to 12 kcal mol^{-1}. Whenever the monomer aggregated with the growing crystal, a new one was added into the lattice in order to keep concentration of monomers constant. Periodic boundary conditions were simply set by 'wrapping the lattice around'. Each random walk simulation was stopped when the crystal reached the size of 1000 monomers and the number of executed random walk steps was recorded. Thus 21 numbers of random walk steps required for growth from the smallest stable complex to a complex of size 1000 were recorded. Then, the 21 growth trajectories, started from the smallest stable complex and ended when the complex reached the size 1000 monomers, were generated using the new algorithm with the same orientational probability and volume fraction as in the random walk simulation. The times of growth were recorded. The mean time of growth in the random walk simulation was $1.11 \times 10^7 \pm 4.39 \times 10^6$. The mean time of growth obtained in simulations with the new algorithm was $1.31 \times 10^7 \pm 4.19 \times 10^6$. One can see that the times obtained by both algorithms have the same order of magnitude.

The crystallisation process spans many orders of magnitude in both time and size scale. From this perspective, the slight discrepancy between the results of both algorithms is not significant. Therefore, we conclude that the new algorithm is equivalent to the random walk one for the simulation of the growth of the single smallest stable complex.

2.6. Size distribution of the microcrystals

We assume that for the tetragonal lysozyme crystal [31], which is the model system in our studies, each smallest stable complex grows independently of others, i.e. there is no aggregation and competition for the monomers between them. This assumption is justified by the results of our simulations which are described in the following section of the paper. Therefore, one can calculate the size distribution of the microcrystals as the function of time for crystal growth in the crystallisation batch with constant volume V and number of monomers equal to N using the following protocol:

1. Set the time T_1 equal to 0.
2. Using Eqs. (3)–(5) and (7) calculate the probability $P(O,V)$ of the occurrence of the smallest stable complex in the crystallisation batch with volume V, volume fraction of monomers N/V, and orientational probability O.
3. Generate the waiting time for the appearance of the smallest stable complex as the random number from the geometric distribution: $P(t = T) = P(O,V)[1 - P(O,V)]^T$. Add this time to T_1.
4. Start the simulation of the growth trajectory with the time of the simulation set to T_1. End the growth trajectory if the time of the simulation reaches $T_{\max}$ or the size of the growing crystal reaches the size of $S_{\max}$ monomers (computer memory limit).
5. Repeat points 3 and 4 until T_1 reaches the value of $T_{\max}$.

The statistics from all the trajectories obtained in point 4 were collected for the $i_{\max}$ time intervals. For each time interval $i \in \{0,..,i_{\max}\}$ and each size of the complex $s \in \{0,...,S_{\max}\}$ the number $K(i,s)$ of growth trajectories in which the complex of size s was present in the time interval i was calculated. $K(i,s)$ is the size distribution of microcrystals as the function of time.

3. Results

All simulations were carried out for the tetragonal lysozyme crystal [31]. The interaction energies, listed in Table 1, value of the entropic penalty equal to 12 kcal mol^{-1} and orientational probability 10^{-5} were adopted from our previous paper — where their validity has been analysed and justified. The volume fraction was set to 0.037 in all calculations. This value was found by Georgalis et al. [6] to be optimal for lysozyme crystallization in 0.64 M NaCl and 0.1 M Na-acetate buffer (pH 4.2). The time step of the simulations was 28 ns.

3.1. Probability of the occurrence of the smallest stable complex

As shown in our previous paper, the smallest stable complex in the lattice of the tetragonal lysozyme crystal is the particular tetramer (Fig. 1). In order to calculate the probability of its appearance we have generated 10^6 random configurations of 1000 monomers in the lattice with a size $15 \times 15 \times 15$ unit cells (volume fraction 0.037). Among these configurations, 6287 were found to contain at least one tetramer. Using Eqs. (3)–(5) and (7) we have calculated the probability of the tetramer's occurrence in a lattice with 13.1×10^{16} nodes, corresponding to a volume of 5 μl. The result was 3.05×10^{-5}. This value was used in the calculations of size distribution. The corresponding mean waiting time for the appearance of the tetramer is 920 μs. Thus, after 8 h we expect 3.1×10^7 tetramers to be present in the solution. The concentration of the tetramers is therefore 5.9×10^{-7} mg ml^{-1} (1.02×10^{-5} μM). It is clear that under these extremely low concentrations the aggregation of tetramers cannot play a significant role in the crystallisation process during the time scale considered in our simulation. This provides supporting evidence to our assumption that each stable complex grows independently.

3.2. Size distribution of microcrystals

We have performed calculations of the size distribution for a volume of 5 μl and the time $T_{\max} = 8$ h. As the calculations of size distribution are time consuming, and their time grows with the volume, we have chosen the smallest volume which, to the best of our knowledge, was used in experiments on lysozyme crystallisation [27].

The calculated size distributions of microcrystals after 2, 4, 6 and 8 h are shown in Fig. 4. As one can see, the most populated complexes are tetramers. Only approximately 13% of all smallest stable complexes generated increased their size during 8 h. This is due to the fact that the tetramer does not have good sites on its surface and thus requires surface nucleation in order to grow. As the mean time of surface nucleation is approximately 56 h (7.3×10^{12} time steps) only the tetramers on which surfaces 'unusually fast' surface nucleation happens, were able to grow.

From all 3.1×10^7 growth trajectories started during the calculations, only eight achieved the size of 10 000 molecules which was set as the maximal size ($S_{\max}$) in this simulation. These trajectories are shown in Fig. 5. One can see that during the first few hours the growth of the microcrystal is very slow. Then, after exceeding the size of approximately 100 monomers (the average from 8 trajectories is 105) the growth trajectory reaches the size of 10 000 monomers in the time scale of minutes. For the fastest growth trajectory, the time of growth from the tetramer to a complex with the size of 10 000 monomers was 3.1 h.

Many complexes with a size larger than four molecules were also found to lack good sites on their surfaces. That is why such an extremely small number of trajectories reached the size of 10 000 monomers. Only those trajectories which have the small number of complexes without good sites and accounted, by chance, very fast surface nucleations grow to the critical size of approximately 100 monomers. Complexes without good sites and of a size larger than the critical one appear very rarely. Even if this happens, large complexes have a higher number of surface nucleation sites and this increases the chance of surface nucleation events. For these reasons, the growth rate of microcrystals increases so dramatically after their size exceeds 100 monomers.

The number of monomers incorporated into the microcrystals during the simulation was 1.4×10^8. The number of monomers in the 5 μl of solution with a volume fraction 0.037 is 4.8×10^{15} — seven orders of magnitude greater. Thus, one can see that microcrystals appearing in solution during 8 h did not cause a significant depletion of monomers in the solution. This excludes the possibility that the growing microcrystals compete for monomers and thus supports our assumption of their independent growth in the time scale of the calculations.

4. Discussion

4.1. Scenario of nucleation and early growth stages of the tetragonal lysozyme crystal

At the start of the discussion, let us summarise the scenario of the early stages of tetragonal

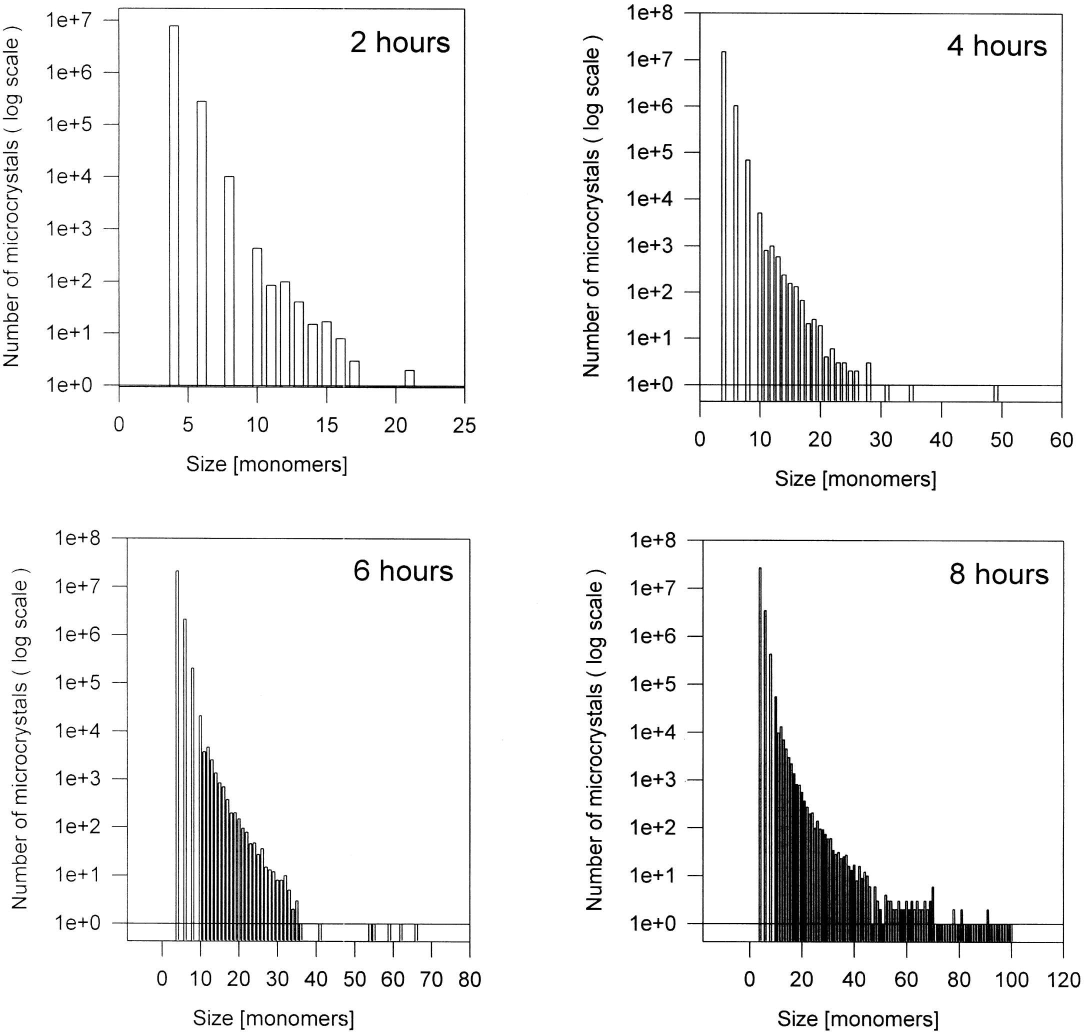

Fig. 4. Size distribution of microcrystals after 2, 4, 6 and 8 h. Only microcrystals of the size below 100 monomers are shown. As explained in the text, there is a very small number of larger complexes.

lysozyme crystal growth as implied by the results of our calculations.

The first critical event in the process of crystallisation is the formation of the smallest stable complex having crystal order. Our calculations predict that the first stable complex is a particular tetramer. The calculations of the mean expected time of the appearance of the first stable complex show that the concentration of tetramers is too small to consider that their further aggregation plays a significant role in the formation of larger complexes. Additionally, the result of the size distribution simulations shows that the decrease of monomer concentration in the first hours of crystal growth is too small to assume competition between growing microcrystals. Therefore, our conclusion is that each of the stable complexes appearing in the solution grows independently of others in the timescale of our calculations.

The first stable complex does not have good sites on its surface and its further growth requires the surface nucleation of two monomers. For an orientational probability of 10^{-5} and a volume fraction of 0.037, the mean time of this event is

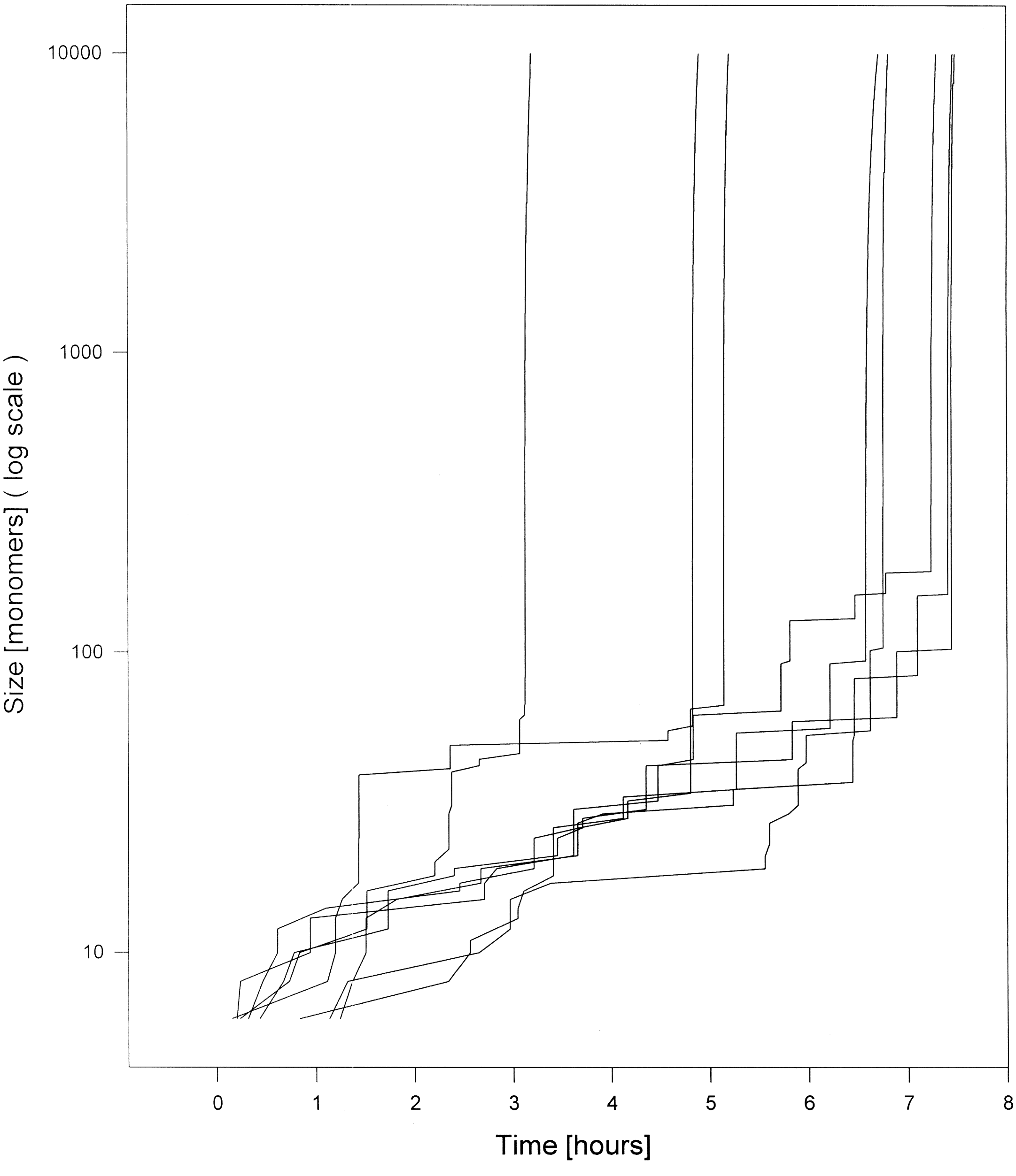

Fig. 5. Eight trajectories, exceeding the second critical size, obtained in the simulation within the timescale of 8 h.

approximately 56 h. This is why the first necessary surface nucleation was observed for only 13% of the tetramers within 8 h. As larger microcrystals without good sites are also frequently found, only eight trajectories were found to reach the *second critical size* of approximately 100 monomers. This size appeared to be critical in the sense that almost all larger complexes do have good sites and they grow to the size of 10 000 monomers in the timescale of minutes. We postulate that only those growth trajectories which exceed the second critical size lead to the formation of macroscopic crystals. Therefore, our model predicts that eight macroscopic crystals appear as the result of 8 h of nucleation. As discussed later, this number is in reasonable agreement with the experimental data provided by arrested nucleation experiments.

4.2. Two critical radii and crystallisation pathways

One should notice that the appearance of a microcrystal which lacks good sites and has the size between the two critical ones strongly decreases the probability that a given growth trajectory reaches the second critical size. This means that what we define as the growth trajectory can be viewed as the kinetic *crystallisation pathway* which can be slow or fast depending on the number complexes without good sites it contains.

The difference between the entropic penalty and interaction energies depends on the surface properties of the monomers and the crystal order. The order of magnitude of the orientational probability does not vary among protein crystals. Therefore, both critical sizes and the number of fast crystallisation pathways depends on the surface properties of the monomers and the geometry of the crystal. This suggests that various crystals of various proteins, or different crystal forms of the same protein can have different kinetics of growth. This, in turn, should be taken into account in searching for quality factors for monitoring the protein crystallisation process.

4.3. Comparison with experimental data

The rates of nucleation for the tetragonal lysozyme crystal have been experimentally measured by Kam and Feher [26] and Blow et al. [27]. In Figure 12 of Kam and Feher's work, one can see that the nucleation rate is approximately 20 nuclei/30 min in 50 μl of solution. Assuming a linear scaling with volume, this would correspond to 2 nuclei h^{-1} in a volume of 5 μl. The arrested nucleation experiments of Blow et al. [27] give a rate between 2 and 7 nuclei h^{-1} depending on the protein concentration in 5 μl of solution. The value predicted by the model presented here is roughly 1 nucleus h^{-1}. The agreement of the order of magnitude can be considered reasonable, bearing in mind that the probability of individual tetramer formation is of the order of 10^{-15} due to geometric constraints, i.e. orientational probability.

The calorimetric studies of Georgalis et al. [9] for lysozyme show the appearance of heat power peaks between 2 and 13 h, depending on protein and salt concentration. This heat effect can be explained, in the framework of our model, by the appearance of quickly growing microcrystals in the time scale of hours.

As shown in our previous paper, the model correctly predicts the speed of movement of the step on 110 face of the crystal if the orientational probability is set to 10^{-5}. For the same value, the model predicts a reasonable nucleation rate. At the same time, the value of orientational probability predicted by our model corresponds well with experimental data concerning formation of the barnase–barstar complex. Schreiber and Fersht [32] measured the influence of the ionic strength on the association rate constant of that complex. Their data, when extrapolated to very high salt concentrations allowed the determination of the basal association rate constant which appeared to be four to five orders of magnitude smaller than the diffusion limit for collisions given by the Smoluchowski–Einstein equation. Recently, Janin [33] analysed the data of Schreiber and Fersht obtaining an orientational probability of 1.5×10^{-5}.

One should note that the present model does not account for dissociation of monomers from microcrystals which certainly appears in the time scale of hours. Reasonable agreement with the arrested nucleation experiments shows that this process does not significantly influence very quickly growing microcrystals. This can be understood if one takes into account energies of the good sites (energies of the monomers attached to the good sites). In the simulations, any microcrystal larger than 100 monomers contains good sites with energies lower than -9 kcal mol^{-1}. Microcrystals larger than 200 monomers have good sites with energies lower than -15 kcal mol^{-1}. Equilibrium constants at 298 K corresponding to these energies are 4×10^6 and 10^{11}, respectively. Thus, the growth of the complexes larger than second critical size is practically irreversible. On the other hand, dissociation of monomers can decrease the number of small complexes present in the solvent in the time scale of hours. Therefore, the numbers presented in Fig. 4 should be treated as the upper bounds of numbers of microcrystals present in the solutions.

We conclude that our model is supported by three kinds of independent experimental evidence: (i) it reproduces the speed of the movement of the dislocation on the 110 face of the crystal; (ii) the nucleation rate predicted by the model is of the order of magnitude of experimentally measured ones; and (iii) the orientational probability has the same order of magnitude as that determined for the barnase–barstar complex.

5. Conclusions

It was shown that there are two critical aggregate sizes in the process of formation of the tetragonal lysozyme crystal. The first one corresponds to the smallest stable complex of crystal order. The second is the size at which there is an abrupt increase in the crystal growth rate due to the fact that all complexes have sites at their surfaces capable of incorporating a single molecule. It was also shown that the growth of the complexes with a size between the two critical ones proceeds through the kinetic pathways which are determined by the presence or absence of good sites on the surfaces of intermediate aggregates. The above scenario is justified by the agreement, within an order of magnitude of the nucleation rate with the experimentally observed one.

It is important to note that the kinetic pathways of the early stages of tetragonal lysozyme crystal growth are determined by the geometry of the surface of the molecule and by the geometry of the growing crystal. The implication of this fact is that other crystal forms of other protein molecules can have different kinetics of early growth stages.

The application of this model to the crystallisation processes of other model systems will allow us to explore the possible diversity of kinetic scenarios of growth suggested by this study. This can be of practical importance as it can show if the quality factors for monitoring protein crystallisation, determined according to the study of one model protein, would work in the case of others.

Acknowledgements

We are grateful to Dr Y Georgalis for critical comments on the manuscript. A.K. acknowledges financial support of the Foundation for Polish Science.

References

[1] C. Carter, Methods Enzymol. 276 (1997) 74.
[2] I. Rayment, Methods Enzymol. 276 (1997) 171.
[3] S.D. Durbin, G. Feher, J. Mol. Biol. 212 (1990) 763.
[4] J.H. Konnert, P.D. Antonio, K.B. Ward, Acta Cryst. D50 (1994) 603.
[5] W. Eberstein, Y. Georgalis, W. Saenger, J. Cryst. Growth 143 (1994) 71.
[6] Y. Georgalis, J. Schuler, J. Frank, D.M. Soumpasis, W. Saenger, Adv. Colloid Interface Sci. 58 (1995) 57.
[7] M. Muschol, F. Rosenberger, J. Chem. Phys. 103 (1995) 1004.
[8] N. Niimura, Y. Minezaki, M. Ataka, T. Katsura, J. Cryst. Growth 154 (1995) 136.
[9] Y. Georgalis, P. Umbach, A. Zielenkiewicz, E. Utzig, W. Zielenkiewicz, P. Zielenkiewicz, W. Saenger, J. Am. Chem. Soc. 119 (1997) 11959.
[10] M.D. Soumpasis, Y. Georgalis, Biophys. J. 72 (1997) 2770.
[11] V. Mikol, R. Giege, The physical chemistry of protein crystallization, in: A. Ducruix, R. Giege (Eds.), Crystallization of Nucleic Acids and Proteins. A Practical Approach, IRL Press/Oxford University Press, Oxford, 1992, p. 219.
[12] Y. Bessho, M. Ataka, M. Asai, T. Katsura, Biophys. J. 66 (1994) 310.
[13] P.R. Wolde, D. Frenkel, Science 277 (1977) 1975.
[14] C. Chothia, J. Janin, Nature 256 (1975) 705.
[15] S.H. Northrup, H.P. Erickson, Proc. Natl. Acad. Sci. USA 89 (1992) 3338.
[16] S.D. Durbin, G. Feher, J. Cryst. Growth 110 (1991) 41.
[17] J. Janin, Proteins 28 (1997) 153.
[18] A.M. Kierzek, W.M. Wolf, P. Zielenkiewicz, Biophys. J. 73 (1997) 571.
[19] J. Janin, F. Rodier, Proteins 23 (1995) 580.
[20] S. Dasgupta, H.I. Ganesh, S.H. Bryant, C.E. Lawrence, J.A. Bell, Proteins 28 (1997) 495.
[21] O. Carugo, P. Argos, Prot. Sci. 6 (1997) 2261.
[22] S.Y. Patro, T.M. Przybycien, Biophys. J. 70 (1996) 2888.

[23] M. Pellegrini, W.S. Wukovitz, T.O. Yeates, Proteins 28 (1997) 515.

[24] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer, Jr., M.D. Brice, J. Mol. Biol. 112 (1977) 535.

[25] J.T.W.M. Tissen, J. Drenth, H.J.C. Berendsen, J.G.E.M. Fraaije, Biophys. J. 67 (1994) 1801.

[26] Z. Kam, G. Feher, Methods Enzymol. 114 (1985) 77.

[27] D.M. Blow, N.E. Chayen, L.F. Lloyd, E. Saridakis, Prot. Sci. 3 (1994) 1638.

[28] W. Feller, An Introduction to Probability Theory and its Applications, Vol. 1, 2nd ed., John Wiley and Sons, New York, London, 1961.

[29] G. Marsaglia, A. Zaman, W.W. Tsang, Stat. Prob. Lett. 8 (1990) 35.

[30] G. Fishman, Monte Carlo: Concepts, Algorithms and Applications, Springer Verlag, New York, 1996.

[31] R. Diamond, D.C. Phillips, C.C.F. Blake, A.C.T. North, J. Mol. Biol. 82 (1974) 371.

[32] G. Schreiber, A. Fersht, Nat. Struct. Biol. 3 (1996) 427.

[33] J. Janin, Proteins 28 (1997) 153.