ELSEVIER

Biophysical Chemistry 91 (2001) 1–20

Biophysical Chemistry

www.elsevier.nl/locate/bpc

Review

Models of protein crystal growth

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

[a] *Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawinskiego 5a, 02-106 Warsaw, Poland*
[b] *Institute of Experimental Plant Biology, Warsaw University, Pawinskiego 5a, 02-106 Warsaw, Poland*

Received 26 February 2001; received in revised form 26 March 2001; accepted 26 March 2001

Abstract

The growth of large and well ordered protein crystals remains the major obstacle in protein structure determination by means of X-ray crystallography. One of the reasons is that the physico-chemical aspect of protein crystallization process is not understood. This article reviews efforts towards formulation of models that could become theoretical frameworks for the interpretation of voluminous experimental data collected on protein crystal growth. Special attention is devoted to microscopic models that recognize the role of the shape of protein molecules in crystal formation.

Keywords: Protein crystallization; Computer simulations; Microscopic models; Protein–protein interactions

1. Introduction

Most biochemical reactions essential for cell function involve specific associations between proteins, small molecules and nucleic acids. As predicted by Emil Fisher, more than 100 years ago, the basis of the specificity is the complementary shape of the interacting molecules. Therefore, in order to investigate the elementary mechanisms of basic biochemical reactions one needs to study the structures of the interacting molecules at the atomic level. For this reason, the study of atomic resolution structures of proteins has developed into an important field of modern molecular biology.

Three-dimensional, atomic resolution structures of the protein molecules can be determined by X-ray or neutron crystallography and nuclear magnetic resonance (NMR) spectroscopy. Crystallography is usually the method of choice as most systems of interest are too large to be studied by

* Corresponding author. Tel.: +48-22-658-47-03; fax: +48-39-12-16-23.
E-mail address: piotr@ibbrain.ibb.waw.pl (P. Zielenkiewicz).

0301-4622/01/$ - see front matter
PII: S0301-4622(01)00157-0

NMR. With X-ray crystallography it is possible to study macromolecules as large as a nucleosome [1], multienzyme complexes such as pyruvate dehydrogenase [2] and viruses [3]. Crystallography also gives a much better resolution, reaching to 1 Å. Due to these advantages more than 80% of known protein structures have been solved by X-ray crystallography (statistics are on the WWW (www.rcsb.org)).

Although the calculation of a protein structure from the diffraction pattern is not a trivial task, the major obstacle on the way to study protein structures by means of crystallography is protein crystallization [4]. The reason for this is that each protein crystallizes under unique set of conditions that cannot be predicted from easily measurable physico-chemical properties, such as the molecular mass or isoelectric point. Therefore, crystallization conditions must be empirically established for each protein to be crystallized. There is also a biological reason of difficulties with protein crystallization. In the cell, proteins function under a crowded environment — the total concentration of molecules is higher than protein concentrations used in crystallization set-ups [5]. Thus, protein molecules must be optimized during evolution in such a way that their crystallization is difficult. Otherwise, they would crystallize or precipitate inside the cell.

Protein crystallization is carried out in supersaturated protein solutions. Unfortunately, it is not enough to bring the solution to supersaturation to obtain well ordered protein crystals-supersaturated protein solutions more frequently produce amorphous precipitates than crystals. Crystallizing solutions contain a number of precipitating agents in addition to protein and water. These agents include various inorganic salts, polyethylene glycol (PEG) and small organic molecules. Detergents, preservatives and other small molecules are often necessary auxiliary additives. Methods also differ in the techniques used to bring the solution to supersaturation. The simplest approach, which is known as the batch method, is to directly mix components in the proper concentrations. Usually, more sophisticated methods are used in which additives undergo diffusion through either a liquid (e.g. dialysis, free interface diffusion) or vapor (e.g. vapor diffusion, evaporation). The practice of protein crystallization is described in the books of Ducruix and Giege [6] and McPherson [7].

As one can see, many parameters such as the concentrations of additives, temperature, pH and crystallization set-up need to be taken into account during the search for the crystallization conditions of a given protein. Protein crystallization is usually sensitive even to small changes in these parameters. Taking into account that the above conditions cannot be predicted, one can understand why protein crystallization is known as a very time consuming and tedious trial and error process. The necessity for the examination of large number of possible crystallization conditions made the structure determination of some proteins take several years.

There is an ongoing interest in the development of systematic approaches to protein crystallization. Statistical methods known as sparse matrix or incomplete factorial methods [8] were applied to gain more information from a small number of trials. Arrays of prepared crystallization set-ups and various robotic systems are used to decrease the time spent on the preparation of solutions. Advances in protein chemistry and genetic engineering are extensively used to explore the possibility of speeding up crystallization by various modifications of the molecules. It has been shown that the removal of several charged groups, either by chemical modification [9] or site directed mutagenesis, enhances the crystallization process. Large proteins which are resistant to crystallization attempts are frequently cut into smaller domains. The active control of crystallization conditions by the reduction of supersaturation after a nucleation period has been applied to obtain larger and better ordered crystals [6]. During several space missions, protein crystallization attempts in microgravity have been carried out [10].

Regardless of the method used to prepare crystallization set-ups, the samples must be assessed for the occurrence of crystals. This is usually done by optical microscopy. As an analytical tool opti-

cal microscopy has two disadvantages: it is subjective and the growth of crystals suitable for detection can take several days. The long time of detection of crystals makes a large number of crystallization attempts very time consuming. Subjectivity of crystal detection precludes the full automatization of the crystallization process. For this reason, there is an interest in the design of better diagnostic methods which will be able to test samples for crystallization in shorter time scales (minutes) and by means of objective quality factors [11]. The obvious candidate for such a method is light scattering. This method is able to monitor the development of microstructures in colloidal suspensions and, in contrast to X-ray and neutron scattering, can be easily set-up in any laboratory. It was shown that light scattering can quickly detect abrupt and polydisperse amorphous aggregation of a protein under certain conditions [6]. This makes it possible to quickly reject large numbers of crystallization set-ups that lead to amorphous precipitation. Recently, portable and computerized light scattering devices became available for testing solution monodispersity.

As follows from the above brief description of protein crystallization practice, protein crystallization lacks a good physico-chemical treatment. There is a need for the theory which would give a quantitative description of the nature of molecular interactions which are essential for crystal formation and the kinetics of the crystal growth. Such a theory would give guidelines for the optimization of the crystallization process by varying the solution conditions and for the development of diagnostic methods. For that reason there is an ongoing interest in the study of the mechanisms of the protein crystal growth [12].

The investigation of protein crystallization is also an interesting field in its own right. In order to explain the mechanisms of protein crystal growth one needs to answer several questions concerning the physical chemistry of concentrated protein solutions. These questions are important as the living cells contain proteins at very high concentrations. Therefore, all the biochemical processes in vivo are subject to macromolecular crowding effects [5] which are similar to the phenomena occurring in supersaturated protein solutions. Protein crystal growth is also an interesting process for physicists. The data collected on supersaturated protein solutions can be used to study various phenomena of phase transitions (e.g. spinodal demixing; [13]), nucleation and behaviour of colloidal suspensions.

The first systematic studies of protein crystal growth were reported in the late 1970s. In volume 114 of *Methods in Enzymology* (1984) Kam and Feher [14] summarized the results of early observations of the protein crystal growth with the use of light scattering, arrested nucleation assays and ultraviolet light transmission. They developed an analytical, kinetic model for the interpretation of light scattering data. Since this time many state-of-the-art experimental techniques have been applied to collect observables on protein crystal growth. The growth of crystal faces has been studied by electron microscopy and atomic force microscopy [15–17]. The computerization of light scattering equipment made it possible to collect important data concerning the early stages of the crystallization process [18–20,13]. For the same task, X-ray and neutron scattering techniques were also applied [21,22,11]. Calorimetric studies have been also carried out [23].

Many efforts have been dedicated to the study of the influence of various physical factors on protein crystallization. The effects of temperature, pressure, electric and magnetic fields has been studied [24,4,25]). Protein crystallization has also been studied in microgravity conditions during several space missions [10].

As the amount of experimental data on protein crystallization increased many authors attempted to develop theoretical models for their interpretation. Authors used a plenitude of approaches starting from fitting parameters of classical nucleation theories to the protein concentration measurements [14,26] and ending up on attempts to describe effective interactions in crystallizing protein solutions by liquid state theories [27]. The purpose of this review is to give brief account of the major ideas used for the development of the theoretical models used to study protein crystal formation.

2. Macroscopic theories of protein crystal formation

2.1. Basic concepts

Many authors attempted to establish the physico-chemical principles of protein crystallization using concepts of thermodynamics and the experimental kinetics of small molecule crystal growth [6,28]. From this perspective, the driving force for crystallization is the difference of the chemical potential ($\Delta\mu$) of the protein molecule in the solution and the crystal:

$$\Delta\mu = -kT\ln(c/s) \quad (1)$$

where c is the protein concentration in the solution, s is the solubility of the protein, k is the Boltzmann's constant and T the absolute temperature. The solubility is defined as the concentration of the protein for which the solution is in dynamic equilibrium with the crystal. If the concentration of protein molecules equals the solubility the crystal neither grows nor dissolves because, on average, the same number of molecules attach to the crystal and dissociate from its surface. The ratio c/s is called supersaturation. Protein crystallization requires high supersaturations in the range 3–30 which is larger than required for the growth of conventional inorganic crystals.

Although supersaturation is required for crystal growth, it does not guarantee the appearance of crystals. In supersaturated protein solutions the solid phase more frequently forms an amorphous precipitate than a crystal. Therefore, it is not enough to find the conditions under which the solution becomes supersaturated. Rather, it is necessary to find conditions under which the solid phase appears in the form of crystal. In recent years it was convincingly shown that, at least for some proteins, an amorphous precipitate also exists under perfect crystallization conditions [11]. Protein crystal growth is therefore the result of the interplay between free monomers, crystals and amorphous precipitates.

Three phases of protein crystal growth can be distinguished: nucleation, post-nucleational growth and cessation of growth [6]. During the nucleation phase ordered nuclei are formed in the solution as the result of fluctuations of matter. Nuclei which exceed the critical size enter the second phase and grow to macroscopic crystals. Protein crystals cease to grow when they reach a certain size. It is not possible to further increase the size of the crystal even if perfect conditions for crystal growth are restored. The phenomenon of the cessation of growth is unexplained. Most authors suggest that it is the result of accumulation of lattice defects and impurities [4,14].

The outcome of the screening for crystallization conditions is frequently visualized on plot of protein concentration vs. precipitating agent concentration (Fig. 1). These plots are frequently referred to as phase diagrams although they are usually not phase diagrams in a strict, physical sense (they do not show the equilibrium states of the system but rather states which are kinetically accessible in the particular experiment). Only free protein monomers are present in solution in the stable region of the phase plot. In the labile region, stable nuclei can form and grow, while in the metastable region nucleation cannot occur but crystals can grow. It is an important observation that there is competition between nucleation and the crystal growth. Crystallization of the protein in the labile region of the phase plot usually causes the appearance of showers of microcrystals. In an ideal crystallization set-up, the process is moved into the metastable region after the time required for nucleation. In this case, a few large crystals appear.

2.2. Nucleation

A nucleation process in which the solid phase is formed from molecules which are free in solution is called homogenous nucleation. If the molecules nucleate on the walls of the container, or on the surfaces of seeds added to the solution, then the process is called heterogenous nucleation. In the following sections only theories describing homogenous nucleation will be reviewed as there is no good theoretical treatment for the heterogenous case.

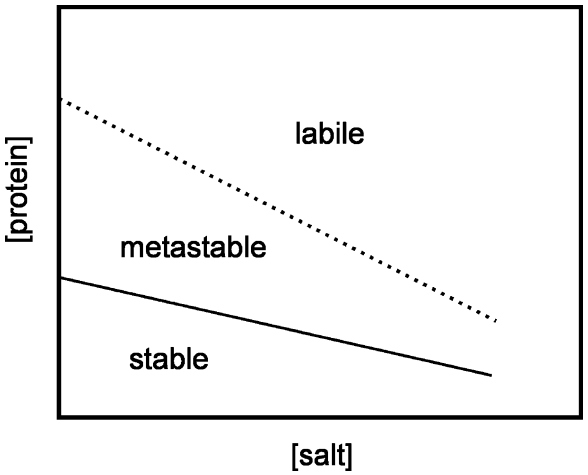

Fig. 1. Typical phase plot. Labile, metastable and stable regions are marked.

2.2.1. Free energy barrier of nucleation

The classical nucleation theory predicts that the free energy of nuclei depends on two terms. The negative term results from bonds formed by molecules in the crystal. It is proportional to the volume of the crystal. The positive term corresponds to the unsatisfied bonds present on the surface of the crystal. It is proportional to the surface of the growing nuclei. The change in standard free energy ΔG_j^o for the aggregation of j monomers can be expressed as [14]:

$$\Delta G_j^o = v_j G_B + \beta j^\gamma G_S \quad (2)$$

where G_B is the bulk free energy per unit volume and G_S is the free energy per unit area of the surface; v is the volume of growth unit. The coefficients β and γ depend on the shape of the nuclei. Under crystallization conditions G_B is negative as it corresponds to the favourable interactions formed in the crystal. G_S is positive as it results from unsatisfied bonds. The shape of the free energy vs. size of the nuclei curve is shown in Fig. 2.

The energetic cost of the addition of a new monomer to the nucleus is given by the following formula:

$$\Delta G_{j+1} - \Delta G_j = d\Delta G_j / d_j = vG_B + \gamma\beta j^{\gamma-1} G_S \quad (3)$$

At the beginning of nucleation the positive, surface term prevails. The addition of each new growth unit causes an increase in the free energy. As the volume grows faster than the surface with the number of monomers, the favourable volume term starts to dominate the free energy expression when the nuclei exceed a certain critical size. The free energy of the nuclei with a critical size is the activation free energy of nucleation. The nuclei which are smaller than the critical size tend to dissolve. They can cross the free energy barrier due to thermal fluctuations in the system.

The energetics of amorphous precipitation differs from that of crystal formation. This can be clearly explained if one uses a linear aggregate as the 'ideal case' of amorphous aggregation. In the linear aggregate, the free energy of a single bond is greater than the unfavorable surface term. Therefore, the cost for the addition of a new monomer is constant. This results in the lack of a free energy barrier for nucleation. Real amorphous precipitates are not linear but they are also not as compact as crystals. Their shape can be described using fractal dimension. For such aggregates there is a free energy barrier — but it is much lower than for the formation of compact structure.

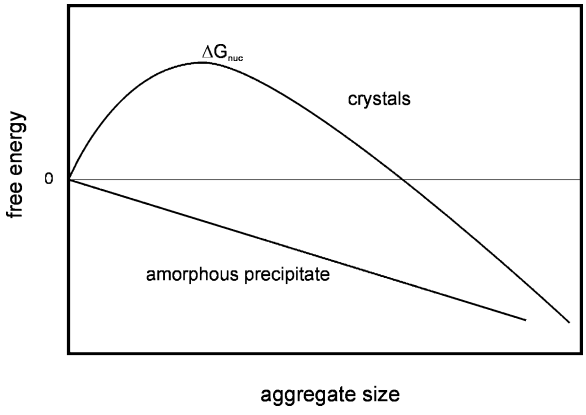

Fig. 2. Shape of the free energy plot as a function of aggregate size, as predicted by Eq. (2). In contrast to amorphous aggregates, ordered nuclei must cross a free energy barrier ΔG_{nuc}.

2.2.2. Kinetics of nucleation

There are many works in which authors attempt to formulate equations describing nucleation in terms of reaction rates. These works are so diverse that it is not possible to extract from them a 'classical theory' of protein crystal nucleation kinetics. This is in contrast with the theories of nucleation thermodynamics which agree with the concepts reviewed above.

In 1984, Kam and Feher [14] presented their kinetic model of protein crystal growth. They had assumed that a protein crystal grows by the addition of monomers to the crystal. The addition of monomers had been treated as a second order reaction described by forward reaction rate; dissociation had been considered as a first order reaction with backward reaction rate. The ratio of the forward:backward reaction rates is proportional to the surface:volume ratio. This theory predicts that the crystal growth process is very slow when nuclei are smaller than the critical size. On the crystal growth curve predicted by this theory, one can see characteristic lag-period during nucleation followed by an abrupt increase in the rate of growth.

A much simpler nucleation treatment was applied by Bessho et al. [26] to describe the crystallization kinetics of lysozyme. They used a differential equation developed by Nielsen [29]:

$$\mathrm{d}m/\mathrm{d}t = kc^{\mathrm{i}} \tag{4}$$

where m is the number concentration of nuclei, c is the concentration of monomers and k the rate constant for nucleation. The exponent i is the number of monomers in the critical nucleus and is independent of concentration. The authors calculated the rate constant of nucleation and critical size by fitting Eq. (4) to the experimental data on concentration changes during lysozyme crystallization.

In these two above examples of protein crystal nucleation theory, the authors assume that the nuclei are formed by the addition of single monomers. Some authors questioned this assumption. Li et al. [30] developed a model in which the initial stages of the protein crystal formation are described by the monomer ↔ dimer ↔ tetramer ↔ 16-mer pathway.

2.3. Kinetics of post-nucleational growth

The kinetics of protein crystal growth have been approached from the perspective of various theories, similar to the kinetics of nucleation. Some of them describe growth as the continuous mass transfer from a solution to a surface of spherical nuclei similar to the classical work of Lifshitz and Slyozov [31]. Others make assumptions about the mechanism governing the growth of the crystal faces. It is impossible to review all these works here as almost each of them has its own and formulation of the model. Below two examples of models are given.

Kam and Feher 1984 [14] formulated an equation for the rate of deposition of the protein mass M at the crystal surface by applying the Fick's law of diffusion:

$$\mathrm{d}M/\mathrm{d}t = k(c-s)S; \quad k = D(\mathrm{d}c/\mathrm{d}x)/(c-s) \tag{5}$$

where k is the rate constant for growth, c is the protein concentration in the vicinity of the crystal surface, s is the protein solubility, S is the area of crystal surface and D is the protein diffusion constant. The concentration of the protein near the crystal surface has been determined by means of analysis of UV photographs.

Bessho et al. [26] applied a polynuclear growth model:

$$\mathrm{d}r/\mathrm{d}t = kc^{\mathrm{p}} \tag{6}$$

with $c = c_0 - (\omega/\nu)\int(\mathrm{d}m/\mathrm{d}\tau)r_{t,\tau}^3\mathrm{d}\tau$ where $r_{\mathrm{t},\tau}$ is the radius of the crystal nucleated at time τ and grow until time t and k is the rate constant for growth. The exponent p represents the order of growth kinetics; c_0 is initial concentration, ω is the shape factor of the crystal and ν is the volume per unit mass of the crystal. The authors also used a differential equation describing the details of the growth mechanisms. They assumed that the crystal grows by two-dimensional nucleation on the crystal surface and a spread of the

islands. Equations for the rate of two-dimensional nucleation and the linear growth rate of the nuclei were used.

3. Protein–protein association

There have been many attempts to produce a computer simulation of crystallization process. In all these models it has been essential to use or formulate a reasonable model of interaction among growth units under crystallization conditions. Thus, we believe that it is reasonable to start reviewing attempts of modelling the protein crystallization process from a brief account of the knowledge concerning thermodynamics and kinetics of protein–protein association. From this perspective assumptions of the various models of protein crystal formation will be more clear to the reader.

Protein–protein associations are essential for most cellular processes. The highly specific association of proteins occurs during signal transduction, assembly of multienzyme complexes and protein complexes which regulate gene expression. It is not surprising therefore, that a large scientific effort has been dedicated to the understanding of the physical principles governing protein association processes. Crystal structures of protein–protein complexes now allow study at the atomic level. Our works cited in this article [32–34] argue that protein crystallization should also be described as a kind of protein–protein recognition process. The difference between protein crystal growth and protein–protein interaction leading to the formation of functional complexes is that protein–protein interfaces in the crystals are smaller than these found in native oligomeric proteins. However, in the crystal environment, a single molecule is surrounded by several symmetry related molecules. Interactions with many neighbours are strong enough to fix the molecule in the lattice.

3.1. Free energy of protein–protein association

One of the challenging problems of modern molecular biophysics is the calculation of association/dissociation free energy of the protein–protein complex given its atomic resolution structure. Although a strict physical solution of the problem is still missing, reasonable empirical models have already been developed.

3.1.1. Main contributions to the free energy of protein association

The protein–protein association reaction can be described by the thermodynamic cycle drawn in Fig. 3 [35]. According to the cycle, the free energy of protein–protein association, ΔG_{assoc}, is first estimated in the gas phase and then solvation term is added to it:

$$\Delta G_{\text{assoc}} = \Delta G^{\text{gas}} + \Delta G^{\text{wat}} \tag{7}$$

where ΔG^{gas}, ΔG^{wat} are the association free energy in the gas phase and solvation term respectively. ΔG^{wat} is the difference between the solvation free energies evaluated independently for the complex and its isolated components:

$$\Delta G^{\text{wat}} = \Delta G_{\text{A}}^{\text{wat}} + \Delta G_{\text{B}}^{\text{wat}} - \Delta G_{\text{AB}}^{\text{wat}} \tag{8}$$

Any conformational change is assumed to occur identically in solution and in the gas phase. The vertical lines on Fig. 3 represent the transfer of the complex and its free components from the gas phase to solution. They account for the protein–solvent and solvent–solvent interactions. The interactions between macromolecules are evalu-

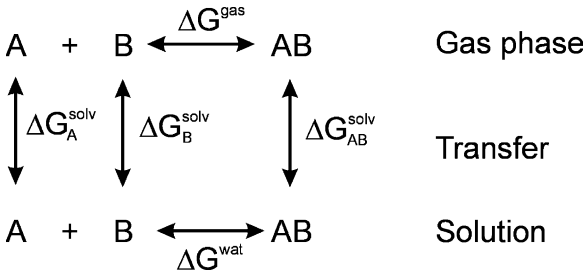

Fig. 3. Thermodynamic cycle of protein association reaction.

ated in the gas phase. The reaction in the gas phase involves four conceptual steps:

1. making one molecule from two;
2. establishing van der Waals and electrostatic interactions;
3. adjusting or changing conformation; and
4. immobilizing side chain and main chain atoms.

For several protein–protein complexes, the structures of both the complex and separated components are available. Analysis of these examples shows that the structure of the protein monomer does not significantly change upon aggregation [36]. Therefore, the association of small globular proteins can be treated as the association of rigid bodies. In this so-called rigid body approximation one considers the free energy contributions of steps 3 and 4 to be negligible.

3.1.2. Entropic penalty for aggregation

Formally, only steps 1 and 4 change the system entropy. Although entropy is difficult to calculate, several attempts have been made.

The entropy change of side chain immobilization has been estimated according to the statistics from rotamer libraries [37,38]. It has been found to be small as there are only a few side chains that are really free to rotate in a dissociated complex.

The entropy change of making one molecule from two has two contributions: (i) entropy change associated with losing rotational and translational degrees of freedom; and (ii) the entropy change as a result of the decrease of frequency of motions in the complex. The entropic penalty for loosing rotational and translational degrees of freedom has been estimated by the Sackur–Tetrode equation [35]. Several authors argued that this equation is not applicable for proteins and proposed their own estimates [39,40]. The vibrational component of the entropy change has been estimated from molecular mechanics [41]. Advances in the calculation of entropic contribution to protein folding and binding are summarized in the review of Brady and Sharp [42].

The values of the positive free energy term resulting from entropy changes which are reported by different authors vary from 0 to 30 kcal/mol. This shows a difficulty in the quantitative analysis of entropic effects in protein aggregation. Tamura and Privalov [43] performed calorimetric experiments aimed at the measuring the entropic penalty for aggregation. Surprisingly, their results show that the entropic penalty is very small, i.e. approximately 1 kcal/mol. This result provoked discussion about the contribution of unfolded states of the protein and thus the relevance of the experiment to the problem of folded molecules aggregation [44,45]. The problems with separating various enthalpic and entropic contributions in interpretation of calorimetric data on biomolecular interactions have been reviewed by Cooper [46].

3.1.3. Interactions in the gas phase

In the gas phase, the proteins interact through van der Waals interactions, electrostatic forces and hydrogen bonds. The energies of these interactions can be calculated using existing forcefields such as GROMOS, AMBER, CHARMM or CVFF. In some of the these models (e.g. CVFF), hydrogen bonds are not explicitly treated but their energies are calculated as the sum of electrostatic and van der Waals contributions. The adjustment of the side chain conformation during aggregation can also be considered using forcefields as interaction models and proper molecular mechanics approaches to find conformations which are in local energy minima.

3.1.4. Solvation free energy

The solvation free energy ΔG^{solv} is the free energy of transfer of the molecule from a vacuum to water. Most of the authors break ΔG^{solv} into following components:

$$\Delta G^{\mathrm{solv}} = \Delta G_{\mathrm{vdW}} + \Delta G_{\mathrm{cav}} + \Delta G_{\mathrm{E}} \quad (9)$$

where ΔG_{vdW} is the free energy of van der Waals interactions with the solvent; ΔG_{cav} is the free energy of cavity formation; ΔG_{E} is the free energy of electrostatic interactions.

The van der Waals interaction energy change during solvation is negative. As the van der Waals interactions are short-ranged, most of the ΔG_{vdW}

term is contributed by the first solvation shell. Therefore, ΔG_{vdW} is roughly proportional to the surface area of the solute which is accessible to solvent molecules.

The free energy of cavity formation is the sum of the entropic penalty for reorganizing the solvent molecules around a solute and the work done against the solvent pressure to create a cavity in the solvent (ΔG_{cav} is positive). The entropic effects of solvent reorganization mainly concern the molecules in the first solvation shell. The number of these molecules is proportional to the surface area of the solute accessible to the solvent.

The sum of ΔG_{vdW} and ΔG_{cav} can be reasonably estimated as terms proportional to the accessible surface area, S:

$$\Delta G_N = \Delta G_{vdW} + \Delta G_{cav} = (a + b)S \quad (10)$$

The accessible surface area (ASA) can be calculated in a variety of ways. In the Conolly method [47] ASA is approximated as the surface of the contact of the sphere with the radius 1.4 Å rolling over the molecular surface. In the Richmond method [48] it is analytically calculated as the surface accessible to the centre of the same probe sphere. The proportionality constants (a and b) in Eq. (10) can be extracted from the experimental data on the solvation fee energy of small alkanes. For these compounds, the electrostatic contribution is negligible in comparison to the non-polar component.

The electrostatic component of the solvation free energy can be estimated using the Poisson–Boltzmann equation [49]. The solvent is assumed to be a continuous medium characterized by the dielectric constant. The interior of the molecule is considered as another continuous medium with a different dielectric constant. The accessible surface area is used as the boundary between the molecule and the solvent. The charge distribution inside the molecule is determined by the partial charges of the atoms. The exterior charge density is the result of the redistribution of ions in response to the electric potential in solution and is modelled by a Boltzmann distribution. For a set of discrete grid points in space, the electrostatic potential is calculated by a numerical solution of the Poisson–Boltzmann equation. From these results one can calculate the total electrostatic energy of the system. The difference between the total electrostatic energy of the molecule in the medium with dielectric constant of 1.0 (vacuum) and a medium with dielectric constant of the solvent is the electrostatic component of the solvation free energy.

3.1.5. Problems in free energy calculations

Although the main factors contributing to the free energy of protein association are known their quantitative calculation still remains a problem. The consideration of all the free energy contributions mentioned above leads to very complex calculations involving many models — the parameters of which are not accurately determined.

The most difficult quantity to calculate is entropic penalty for aggregation. As was mentioned above the results of different models vary between 0 and 30 kcal/mol. The considerations of the atomic motions by molecular mechanics methods give results which strongly depend on the forcefield used. It is not clear which of the many forcefields available is the most appropriate for this application.

Electrostatic calculations are determined by values of dielectric constants which do not have physical sense for systems studied at the atomic level. Solutions of the Poisson–Boltzmann equation are also sensitive to side chain positions. This means that the results are influenced by the choice of forcefield applied to optimize side chain positions. Another problem with applying the Poisson–Boltzmann equation is that it is only correct for low ionic strength solutions. There are attempts to modify Poisson–Boltzmann theory in order to increase the range of ionic strength for which calculations are valid [50].

In spite of these efforts it is very unlikely that approaches which have Debye–Huckel theory in their roots will accurately describe electrostatic interactions under crystallization conditions. Typical buffers used for protein crystallization contain salts at high concentrations. It has been shown both in theoretical and experimental papers that salts such an sodium chloride or am-

monium sulfate are able to form clusters under these conditions. The clusters have been detected by light scattering in undersaturated solutions of sodium chloride, ammonium sulfate and sodium citrate [51]. Several authors studied formation of these clusters by Monte Carlo and Molecular Dynamics simulations [52–55]. These works show that concentrated solutions used to enhance protein crystallization poses complex and poorly understood atomic-level structures that make application of continuous solvation models unreliable.

Difficulties in the calculations of the free energy of protein–protein association were discussed by Novotny et al. [56]. The work shows the results of blind-tests of association free energy calculations. The authors attempted to predict the association free energy differences between several mutants of a lysozyme–antibody complex. The experimentally determined values of the energies were not known to authors before the calculations were performed. In spite of the fact that the entropic penalty cancels out for the calculations of free energy difference between mutants, the authors obtained very poor correlation between calculated and experimental values. This shows that the application of very complex calculations which attempt to take into account all the contributions of the association free energy does not guarantee accurate results.

3.1.6. A practical method of association free energy calculation

In 1959, Kauzmann [57] suggested that a major factor in the formation of protein–protein complexes is the hydrophobic effect. Chothia and Janin [58] found that the hydrophobic effect of protein association can be modelled by a free energy term proportional to the accessible surface area change during complex formation. They showed that the change of accessible surface area multiplied by an empirical constant gives reasonable estimates of the association free energy. The empirical constant was found to be a proportionality coefficient between the solvation free energy of small alkanes and their accessible surface areas. Eisenberg and McLachlan [59] pointed out that it is an oversimplification to calculate the energy of association on the basis of surface area alone because the polarity and charge must also be taken into account. They introduced five proportionality constants for five atom types to account for the character of each atom. Therefore, the interaction free energy of the protein–protein complex calculated according to their theory can be written as:

$$\Delta G_{\text{inter}} = \sum_i \delta_i (\text{ASA}_{\text{d},i} - \text{ASA}_{\text{m},i}) \tag{11}$$

where $\text{ASA}_{\text{d},i}$ and $\text{ASA}_{\text{m},i}$ are the accessible surface areas of the i-th atom in the dimer and monomer respectively, δ_{i} is the atomic solvation parameter (proportionality constant) of the i-th atom. The atomic solvation parameters in Eq. (11) have been fitted to reproduce experimental data of transfer free energies of small molecules from octanol to water. In order to calculate the association free energy one should add an estimate of the entropic penalty of aggregation to the interaction free energy calculated by Eq. (11).

In 1992, Horton and Lewis [60] further improved the approach of Eisenberg and McLachlan [59]. They introduced an additional two proportionality constants to account for the fact that some polar atoms are hydrogen bonded. The entropic penalty for aggregation was determined empirically by fitting the results of the calculations to the experimentally determined association free energies of 15 protein–protein complexes. Horton and Lewis [60] achieved a 96% correlation of the calculated and experimentally measured association free energies.

Vallone et al. [61] calculated the association free energy of hemoglobin mutants using the Eisenberg and McLachlan [59] approach. They obtained a very good agreement with experimentally measured values.

The above examples show that the free energy of protein–protein association can be reasonably estimated by the application of Eq. (11). This means that the changes in the first solvation shell (which are proportional to ΔASA) provide the main contribution to the association free energy. From a practical point of view, calculations with the Eisenberg and McLachlan [59] or Horton and Lewis [60] methods are much more computatio-

nally simple than explicit considerations of all free energy contributions shown on the Fig. 3. The later requires the application of many complex models and frequently proper 'empirical weighting' of their results. The approach of Horton and Lewis [60] requires the application of a single, computationally cheap model. Therefore, the practical conclusion from the above chapter is that the best approach to protein association free energy calculation is still an empirical model in which the association free energy is calculated as proportional to the surface area buried in the protein–protein interface.

3.2. Kinetics of protein–protein association

From the kinetic point of view, protein–protein association is described by a second order reaction rate constant. Therefore, research efforts have been dedicated to the calculations of second order reaction rate constant based on the knowledge of protein structure.

As mentioned in Section 3.1.1 association of globular proteins can be treated as association of rigid bodies. This means that the protein–protein interface is formed if the 'reactive patches' of the protein surfaces are brought together by translational and rotational diffusion, and that conformational changes of the protein molecules do not play significant role in this process. In other words, the association of two protein molecules is considered as a diffusion limited process.

3.2.1. Collisions, encounters and anisotropy of globular proteins

For two hard spheres of equal size, the second order reaction rate constant is given by the Smoluchowski–Einstein equation [62]:

$$k_{\mathrm{coll}} = 8RT/3\eta = 6.6 \times 10^{9}\ \mathrm{M^{-1}s^{-1}}\ (\text{at } 300\ \mathrm{K}) \tag{12}$$

where R is the gas constant, T the absolute temperature and η viscosity of water. Eq. (12) does not depend on the size of the spheres as an increase in target size is compensated by the decrease in the diffusion constant. The collisions of protein molecules in water solution cannot occur with a rate higher than k_{coll}. For this reason k_{coll} is frequently referred to as the diffusion-limited rate.

Protein molecules are highly anisotropic. In order to form a native complex they must form a particular interface between the proper fragments of the surfaces of both molecules. Atoms in the interface must be aligned with approximately 2 Å precision. Requirement for such a precise orientational alignment causes a decrease in the second order reaction rate by many orders of magnitude, with respect to k_{coll}.

The second order reaction rate of a typical protein–protein complex is in the order of magnitude of $10^{6}\ \mathrm{M^{-1}\ s^{-1}}$. Northrup and Erickson [63] pointed out that it is surprisingly high. They showed, from geometrical considerations, that two protein molecules which collided in space form the proper interface with a probability of the order of 10^{-7}. Thus, the collision rate multiplied by the orientational probability is $10^{2}\ \mathrm{M^{-1}\ s^{-1}}$ — four orders of magnitude lower than experimentally measured one. Northrup and Erickson [63] performed Brownian dynamics simulations to explain this data. They found that after the first collision of the molecules, there was an increased probability that molecules collided again. They defined the full set of interactions between the molecules, from the first collision to final separation as an encounter. Brownian dynamics simulations showed that there are, on average, nine collisions per encounter. Molecules were considered to form a collision complex whenever their surfaces were found to have 2–4 Å separation. Due to the fact that there are many collisions per encounter, the chance of finding the proper orientation between molecules is significantly increased. One should notice that this effect results from the statistical properties of the diffusion process, but not from the interactions between the molecules.

Another effect which can increase the rate of the formation of a complex is stabilization of the partially formed interface. The probability that molecules separate is decreased, even if only part of the atomic interactions in the interface is formed. In this intermediate complex, molecules

change their orientation and finally form the proper interface.

Northrup and Erickson [63] claim that the effects of encounters and interactions in partially formed interfaces account for the difference between the diffusion-limited rate, the rate resulting from geometrical considerations and experimentally measured values.

3.2.2. Electrostatic and hydrodynamic steering

In spite of the anisotropy of the protein molecules, some of them form complexes with the very high rates close to the diffusion limited ones (10^9 M^{-1} s^{-1}). This is achieved by long ranged electrostatic forces acting between the molecules. Probably many complex-forming protein subunits evolved in such a way that there are opposite charges exposed on the surfaces forming their interface. Therefore, when one molecule approaches the other it is properly oriented by electrostatic forces before collision takes place.

Schreiber and Fersht [64] measured the magnitude of electrostatic steering in the barnase–barstar complex and their mutants. The association rates were measured for different concentrations of NaCl. The authors showed that there is a linear dependence between the logarithm of the association constant and logarithm of electrostatic contribution to the activity of NaCl. This relationship allowed them to estimate the association rate constant in the absence of electrostatic forces by extrapolation to high salt concentrations where electrostatic forces are screened. Electrostatic interactions were found to increase the association rate constant by four to five orders of magnitude.

Long ranged forces in the formation of several protein–protein complexes have also been studied by Brownian Dynamics simulations [65,66]. In these simulations, the electrostatic potentials around protein molecules were calculated by the Poisson–Boltzmann equation. This allowed the authors to study electrostatic steering at the atomic level. Brownian dynamics simulations also showed that protein–protein association is influenced by hydrodynamic forces resulting from the interactions between layers of water molecules moved by protein molecules. In some cases, they cause a pre-orientation of the approaching protein molecules. Simulations show that the effects of hydrodynamic forces are much smaller than the effects of electrostatic interactions. Hydrodynamic effects do not change the order of magnitude of the association rate constant [67].

3.2.3. Orientational probability

As is clear from the above review, the main factor which influences the association rate constant of globular proteins is their anisotropy. Electrostatic steering increases the rate of association only in the few cases of proteins for which biological function very fast rate of complex formation is essential.

It is convenient to express the anisotropy of the protein molecule as the orientational probability. The orientational probability is the probability that two molecules which would be brought close to one another by translational diffusion would form the proper interface. In principle, if the structure of the complex is known, one can calculate the orientational probability as the fraction of proper complexes among large number of randomly generated ones. In this way, the value of 10^{-7} reported Northrup and Erickson [63] was estimated. The serious problem with the calculation of the orientational probability from the geometrical constraints is that arbitrary criteria are used to decide which complex, from large number, has the proper interface. These criteria are formulated in terms of interatomic distances or angles describing the relative positions of the monomers. The way to avoid arbitrary criteria in the determination of orientational probability is to calculate its value from kinetic data.

In 1997, Janin [68] analysed the data of Schreiber and Fersht [64]. He calculated the orientational probability as the factor by which the diffusion limited rate must be multiplied in order to obtain the basal rate constant (association rate constant when the electrostatic interaction are screened). He obtained a value of 1.5×10^{-5}. A value of the order of 10^{-5} has also been independently reported by Kierzek et al. [32]. Earlier Durbin and Feher [69] reported a value of 10^{-6}.

4. Microscopic models of the crystallization process

Many phenomena in the crystallization process can only be understood if single growth units and interactions among them are explicitly taken into consideration. Molecular level structures, such as the two dimensional nuclei, spiral dislocations and various defects, are essential for the crystal growth and quality. For these reasons microscopic considerations of crystallization process appeared as early as in the late 1940s. Bragg and Nye [70] observed the behaviour of specially prepared rafts of soap bubbles. The bubbles formed ordered structures with defects. Many of the effects observed by Bragg and Nye have been later observed in real crystals.

As soon as computers become available to scientific community computer simulations have been applied to study crystallization at the molecular level. There are numerous examples of the computer simulation protocols formulated to study general physical principles governing crystallization process. The classical example is the so-called solid-on-solid model [71–74].

As follows from Section 3, protein molecules have complex shapes and surface properties that determine their aggregation behaviour. It is therefore very difficult to include their molecular level interactions into the simple, analytical formulas of the kind presented in Section 1. As will be shown below such interactions can be, to much larger extent, considered if the computer simulation protocols are used.

In the following part of this work the microscopic models of protein crystallization will be divided into three sections. Section 4.1 reviews attempts of using detailed, atomic-level representation of protein structure to study various aspects of crystallization process. Section 4.2 presents applications of computer simulation protocols that are rigorously derived from the principles of statistical mechanics. Complex phase behaviour of protein solutions have been addressed in these works. Section 4.3 presents computer simulation protocols in which lattice discretization is used to decrease the computational complexity of the problem in order to extend spatio-temporal scale of the calculations.

4.1. Models with detailed representation of protein structure

Atomic level representation of protein structure was used by Tissen at al. [75] to study the diffusion of protein molecules under crystallization conditions. In this work protein molecules were represented as rigid bodies defined by the surface elements calculated from surface triangulation of experimentally determined structure of the molecule under investigation. These authors argued that under crystallization conditions the long range electrostatic interactions are screened. Therefore, the only long range forces influencing the protein crystallization process are hydrodynamic interactions. The simulation protocol presented by authors was designed to properly describe a shape-dependent hydrodynamic steering effect (Section 3.2.2). Thus, the motion of protein molecules was represented by Langevin's equation of motion and the Stokes equations described motion of the solvent. The total force acting on the molecule was the sum of the systemic, hydrodynamic and random forces. The goal of this approach called 'microhydrodynamics' was to study the kinetics of crystal nucleation by simulation of the motion of the few monomers forming first ordered complexes. The method could be also applied to the simulation of the single growth unit attachment to the site on the surface of the growing crystal. Unfortunately the authors were able to apply the method only to the 'static' calculation of the diffusion coefficients of various protein molecules (good agreement with experimental data has been achieved). No results concerning simulations of aggregating monomers have been presented so far. In spite of this fact the microhydrodynamics approach will probably become valuable tool in the future. The method has sound physical basis and offers the prospect of studying molecular-level, dynamic models of the initial events in the protein crystallization. The reason why it has not been done so far is clearly the computational cost of the calculations.

As the computational power of the hardware grows quickly application of this method, armed with sophisticated short-range interaction potentials should be possible in the near future.

Second virial coefficient of the dilute protein solution has been reported to correlate with the solution conditions conductive to crystallization [76,77]. To study molecular basis of these phenomena Neal at al. [78,79] attempted to compute second virial coefficient of the protein on the basis of the knowledge of its atomic level structure. Second virial coefficient has been calculated as the five dimensional integral of the potential of mean force (PMF) over variables describing angular orientation and intermolecular separation of two protein molecules. During configurational sampling by Monte Carlo procedure the PMF has been computed using continuous solvent approximation and protein–protein interaction energies computed according to atomic level structures. For the computation of intermolecular interaction energy the authors developed complex, semi-empirical model involving following contributions. Long ranged electrostatic interactions have been computed using Poisson–Boltzmann equation (Section 3.1.4). To avoid prohibitive computational costs during this calculation the structure was represented as a sphere with molecular volume of the protein and angular distribution of charges the same as in the explicit structure. Short range interactions were evaluated using atomic level description of the structure. For the pairs of atoms that are not separated by water molecules authors used Lennard–Jones potential with OPLS forcefield parameters. For the atom pairs separated by the solvent a Lifshitz–Hamaker potential was applied. Second virial coefficients values calculated for chymotrypsinogen A were in good agreement with static light scattering measurements. Moreover, the relationship between second virial coefficient and the ionic strength was correctly reproduced. This strongly supports the model as in the case of chymotrypsinogen A the relationship between second virial coefficient and ionic strength is complex and counterintuitive — at approximately pH 7 a second virial coefficient increases with increasing ionic strength. Authors have shown that the anisotropy of protein molecules has pronounced effect on calculations. From a large number of configurations sampled during Monte Carlo integration a few had contributed disproportionately strongly to the result.

Recently, Asthagiri et al. [80] used a protein–protein interaction model described above to study variations in the morphology of subtilisin BPN′ variant s88 crystals as the function of the ionic strength. The authors have shown that simple computations of the interaction strength of crystal contacts cannot explain observed trends in growth rates. Instead, intensive configurational sampling must be performed in order to account for the presence of the interactions incommensurate with the crystal order. Protein molecules bound to other ones or the faces of the growing crystal in these configurations are shown by the authors to be competitive inhibitors of the crystal growth. Results of the docking of the protein molecule to the other one or to the clusters of molecules representing sites on the growing crystal faces were able to explain observed morphology changes of the crystal under investigation. The predominant directions of crystal growth were found to correspond to contacts whose energies were well separated from the energies of random configurations. Other crystal contacts had energies of the order of those found for random contacts. These results are in agreement with the ‘random energy model’ used by Janin [81] to study specificity of biologically relevant protein–protein interactions.

4.2. Applications of statistical mechanics

Statistical mechanics of liquid systems is the field in which computer simulations have found one of its best known applications. The excellent book of Allen and Tildesley [82] presents the large numbers of computer simulation algorithms rigorously derived from the first principles of statistical mechanics in order to study various aspects of the liquid state matter. One of the most intensively studied subjects in this field is the determination of phase plots and free energy barriers for the systems undergoing phase transitions. Some state of the art algorithms have been applied by Wolde and Frenkel [83] for the molec-

ular level interpretation of the phase plots characteristic for crystallizing proteins.

In their work protein molecules were modelled as spheres interacting with short range attractive potential. The interaction potential has the form of Lennard–Jones, so the strength and the range of the interaction depends on two adjustable parameters. This representation of the systems allows calculation of the phase plots by the simulation protocols rigorously derived for the calculations of phase behaviour of Lennard–Jones fluid [84]. As shown in these works the range of the intermolecular attraction has a drastic effect on the phase diagram. Thus authors adjusted the range of attraction in their model potential in such a way that the phase plot obtained resembled the one experimentally determined for the protein, i.e. the fluid–fluid coexistance curve was located in the metastable region $\sim 20\%$ below the equilibrium crystallization curve. Then the authors used the umbrella sampling Monte Carlo simulation protocol of Torrie and Valleau [85] in order to determine free energy barriers on the path from the metastable fluid to the critical crystal nucleus. They have found that the presence of the fluid–fluid critical point decreases the free energy barrier of nucleation. The system evolving along the lowest free energy pathway first forms the liquid colloid droplets. Density fluctuations in these droplets lead to the rearrangement of protein molecules to the ordered crystal nucleus. The authors argue that their results are relevant to the design of crystallization protocols as the location of the metastable critical point can be controlled by the composition of the crystallization buffers.

Protein crystallization in aqueous-electrolytic environment can be treated as the process driven by the effective solvent mediated interactions described in terms of PMF. Soumpasis and Georgalis [27] tried to obtain first-order quantitative description of these PMFs using liquid state theory of three-component system of charged species. The model system in this study was lysozyme crystallizing in NaCl water solution. Three species (lysozyme, Na and Cl ions) were modelled as charged hard spheres with the radius 19.2 Å for the protein and 4.9 Å for ions. The charge of the lysozyme molecule has been set to $+7e$. Water has been treated as continuous dielectric medium. Using the Orstein–Zernike equation with hypernetted chain closure relation all pair-wise potentials of mean force have been computed. Results have been compared with quasi-stationary hydrodynamic radius of the lysozyme aggregates, experimentally determined by light scattering. For the protein and salt concentrations for which little aggregation has been indicated by light scattering protein–protein PMF has been purely repulsive. In the case of high salt concentrations, in the conditions for which protein aggregation has been shown experimentally, the PMF exhibited short range attractive well and moderately repulsive long-range barrier. Therefore, theoretically predicted trends were in a very good agreement with experimental data which indicate that charge screening and excluded volume effects alone have an important contribution to the phase behaviour of the protein solutions. The method described above has been subsequently used by Georgalis et al. 1998 [86] to rationalize effective protein–protein interactions in the lysozyme-NaCl system undergoing spinodal decomposition.

4.3. Lattice simulations

The works presented in Sections 4.1 and 4.2 show that microscopic simulations of the protein crystal growth are enormously demanding computationally. If one attempts to represent protein molecule at atomic level and rigorously compute its motion in the solution it is hardly possible to go beyond the simulation of the diffusion of single protein molecule. Docking or configurational sampling of the molecules at crystal contacts and lattice sites is possible but it does not allow to study directly kinetics of the process. Wolde and Frenkel [83] explicitly state in their work that even if the proteins are approximated by the spheres the kinetics of the crystallization cannot be directly studied by molecular dynamics simulation.

Lattice simulation is the approach commonly used to build approximate, discrete model of the system that cannot be simulated if all the degrees of freedom are accounted for. In the case of

protein crystallization problem several authors attempted to formulate simulation protocols in which protein molecules were approximated as points occupying nodes of the three-dimensional lattice. Diffusion of the monomers was simulated as the random walk of the points on the lattice. Lattices are defined in such a way that the edges correspond to main intermolecular bonds (crystal contacts) in the crystal under investigation. Therefore, for any cluster of molecules, interaction energy can be computed by adding values corresponding to the strength of pair-wise interactions represented by lattice edges. These values can be treated as free parameters of the model and fitted, so that the model reproduces experimental data or estimated using approaches reviewed Section 3.

In 1991, Durbin and Feher [69] studied the growth of the (101) and (110) faces of the tetragonal lysozyme crystal. In their study, the energies of the three main crystal contacts had been explicitly taken into account as free parameters of the model. They also introduced the orientational probability, i.e. the probability that a molecule arriving at the crystal site is properly oriented to form the contacts. From the number of successful attachments in Monte Carlo simulation protocol of Gilmer [71] they calculated the rates of crystal growth. The free parameters of the model were fitted so that the simulations reproduced geometry of the faces observed by electron microscopy and experimentally measured growth rates. The value of orientational probability was found to be 10^{-6}. The authors achieved a satisfactory agreement with experimental data. The faces of the crystal, like actual ones, grow by the nucleation and growth of the islands. The islands on the (110) face were two molecules high which was also observed by electron microscopy. The authors were able to explain the observed growth rates. In this work there has been no attempt to use the model to study the early stages of crystal formation.

Kierzek et al. [32–34] formulated the lattice simulation protocol that allows to study early stages of the protein crystal formation at the time and size scales ranging from the assembly of initial ordered complexes up to the microcrystals of the size 10^5 monomers. The authors used the lattice in which the edges correspond to all the contacts in the crystal environment of the molecule under investigation (all the intermolecular contacts in the crystal in which at least one interatomic distance is smaller than 5 Å). Diffusion of the protein monomers was simulated as the random walk of the points representing protein monomers in the lattice. The timestep of the random walk simulation was calculated as the time required by the molecule to move a distance equal to its diameter, using the equation for mean squared displacement of Brownian particle. Attachment of the molecule to the site on the surface of the cluster depended on translational diffusion simulated by random walk and rotational diffusion modelled by orientational probability. The rate of monomer dissociation was determined by interaction free energy of the monomer in the given lattice site. Interaction free energy corresponding to the crystal contact was calculated as proportional to accessible surface area buried during formation of this contact [Eq. (11)]. Atomic solvation parameters of Eisenberg and McLachlan [59] (see Section 3.1.5) were used. Therefore, the above simulation protocol allowed application of the interaction model based on the atomic level structure of the system under investigation and at the same time allowed computation of the 'growth trajectory' of the microcrystal large enough to see the crystal faces already formed. The model has been first tested in the simulation of the step growth on the (110) face of the tetragonal lysozyme crystal. Results have been compared with molecular level, time-resolved atomic force microscopy observations [16]. Good agreement with experimental data has been obtained if the orientational probability has been set to 10^{-5}. A similar value of orientational probability (1.5×10^{-5}) for protein–protein association process has been independently obtained by Janin [68] on the basis of measurements of barnase-barstar aggregation kinetics [64] (see Section 3.2.3).

The algorithm of Kierzek et al. allowed simulation of the growth of the tetragonal lysozyme crystal from free monomers to complexes having 10^5 molecules. These microcrystals had the faces already formed and the geometry of these faces

was in good agreement with atomic force and electron microscopy observations. Also the morphology of the crystal was correctly reproduced. General results concerning the existence of the kinetic traps on the crystallization pathways of protein crystals have been presented. Authors also discussed relation between the first ordered complexes and critical complexes of classical nucleation theory (see Section 2.2.1).

Nadarajah et al. [87,88] argued that the growth of (110) and (101) faces of tetragonal lysozyme crystal can be explained only if the growth unit larger than monomer is assumed. They studied geometry of the crystal contacts and postulated tetramer and octamer growth units. With this assumption, they applied analytical kinetic models which gave results in agreement with several measurements of the growth rate concentration dependence. This work is contradictory to the work of Durbin and Feher and Kierzek et al. [32,69] which show that experimental data concerning the growth of tetragonal lysozyme crystal faces can be explained using models assuming monomer to be the growth unit. For a detailed discussion of this discrepancy see Kierzek et al. [34].

Lattice simulations presented above allowed calculations concerning the particular model systems, i.e. the lattices were built according to geometries of the crystal contacts in the model systems and interaction potentials were estimated according to analysis of these contacts. Some authors attempted to study general properties of the aggregation and crystallization of the molecules with variable surface patches using models that did not reflect structures of any particular molecule. In 1996, Patro and Przybycien [89] performed two-dimensional 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. Pellegrini et al. [90] were able to explain non-uniform distributions of space groups found in the Brookhaven Protein Data Bank by Monte Carlo simulations in which nuclei were assembled from three dimensional rigid bodies with different surface patches.

5. Conclusions

This article clearly shows that there are still no satisfactory models of protein crystallization process. The reason for this is the enormous complexity of the problem. One has to calculate elementary interactions among protein monomers in the concentrated salt solutions the properties of which are still the matter of research [51–55]. When the interactions are estimated by approximate approaches there is a problem of the time and size scales. The timestep of any reasonable microscopic simulation should be of the order of tens of nanoseconds while macroscopic crystals appear after hours or days. The size of the growth unit is of the order of nanometers while protein crystals suitable for structure determination reach the size of millimetres. Therefore, the crystallization process spans many orders of magnitude on both time and size scales which is prohibitive for most of the computer simulation approaches.

Due to these difficulties the field of crystallogenesis is at the stage when many approaches based on extremely variable scientific backgrounds are tested. Further development of experimental methods and theoretical approaches will probably lead to some unification. On the way, studies on the mechanisms of protein crystal growth formation provoked a lot of valuable research in the fields of protein–protein interaction and physical chemistry of solutions.

References

[1] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 A resolution Nature 389 (6648) (1997) 251–260.

[2] A. Mattevi, G. Obmolova, E. Schulze et al., Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex Science 255 (5051) (1992) 1544–1550.

[3] S.S. Kim, T.J. Smith, M.S. Chapman et al., Crystal structure of human rhinovirus serotype 1A (HRV1A) J. Mol. Biol. 210 (1) (1989) 91–111.

[4] S.D. Durbin, G. Feher, Protein crystallization Annu. Rev. Phys. Chem. 47 (1996) 171–204.

[5] S.B. Zimmerman, A.P. Minton, Macromolecular crowding: biochemical, biophysical, and physiological consequences Annu. Rev. Biophys. Biomol. Struct. 22 (1993) 27–65.

[6] A. Ducruix, R. Giege (Eds.), Crystallization of Nucleic Acids and Proteins. A Practical Approach, IRL Press/Oxford University Press, Oxford, 1992.
[7] A. McPherson, Crystallization of Biological Macromolecules, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1999.
[8] C. Carter, Response surface methods for optimising and improving reproducibility of crystal growth Methods Enzymol. 276 (1997) 74–99.
[9] I. Rayment, Reductive alkylation of lysine residues to alter crystallization properties of proteins Methods Enzymol. 276 (1997) 171–179.
[10] S. Koszelak, J. Day, C. Leja, R. Cudney, A. McPherson, Protein and virus crystal growth on international microgravity laboratory — 2 Biophys. J. 69 (1995) 13–19.
[11] Y. Georgalis, J. Schuler, J. Frank, D.M. Soumpasis, W. Saenger, Protein crystallization screening through scattering techniques Adv. Colloid Interface Sci. 58 (1995) 57–75.
[12] R. Giege, B. Lorber, A. Theobald-Dietrich, Crystallogenesis of biological macromolecules: facts and perspectives Acta Cryst. D 50 (1994) 339–351.
[13] P. Umbach, Y. Georgalis, W. Saenger, Time-resolved small angle static light scattering on lysozyme during nucleation and growth J. Am. Chem. Soc. 120 (1998) 2382–2390.
[14] Z. Kam, G. Feher, Nucleation and growth of protein crystals: general principles and assays Methods Enzymol. 114 (1984) 77–112.
[15] S.D. Durbin, G. Feher, Studies of crystal growth mechanism of proteins by electron microscopy J. Mol. Biol. 212 (1990) 763–774.
[16] J.H. Konnert, P.D. Antonio, K.B. Ward, Observation of growth steps, spiral dislocations and molecular packing on the surface of lysozyme crystals with the atomic force microscope Acta Cryst. D 50 (1994) 603–613.
[17] A.J. Malkin, Yu.G. Kuznetsov, T.A. Land, J.J. DeYoreo, A. McPherson, Mechanisms of growth for protein and virus crystals Nat. Struct. Biol. 2 (1995) 956–959.
[18] W. Eberstein, Y. Georgalis, W. Saenger, Molecular interactions in crystallizing lysozyme solutions studied by photon correlation spectroscopy J. Cryst. Growth. 143 (1994) 71–78.
[19] A.J. Malkin, A. McPherson, Light-scattering investigations of nucleation processes and kinetics of crystallization in macromolecular systems Acta Cryst. D 50 (1994) 385–395.
[20] A. Schaper, Y. Georgalis, P. Umbach, J. Raptis, W. Saenger, Precrystallization structures in supersaturated lysozyme solutions studied by dynamic light scattering and scanning force microscopy J. Chem. Phys. 106 (1997) 8587–8594.
[21] O.D. Velev, E.W. Kaler, A.M. Lenhoff, Protein interactions in solution characterized by light and neutron scattering. Comparison of lysozyme and chymotrypsinogen Biophys. J. 75 (1998) 2682–2697.
[22] N. Niimura, Y. Minezaki, M. Ataka, T. Katsura, Aggregation in supersaturated lysozyme solution studied by time-resolved small-angle neutron scattering J. Cryst. Growth 154 (1995) 136–142.
[23] Y. Georgalis, P. Umbach, A. Zielenkiewicz et al., Microcalorimetric and small-angle light scattering studies on nucleating lysozyme solutions J. Am. Chem. Soc. 119 (1997) 11959–11965.
[24] D.M. Blow, N.E. Chayen, L.F. Lloyd, E. Saridakis, Control of nucleation of protein crystals Prot. Sci. 3 (1994) 1638–1643.
[25] N.I. Wakayama, Quantitative study of crystallization kinetics of hen egg-white lysozyme using magnetic orientation J. Cryst. Growth 191 (1998) 199–205.
[26] Y. Bessho, M. Ataka, M. Asai, T. Katsura, Analysis of the crystallization kinetics of lysozyme using a model with polynuclear growth mechanism Biophys. J. 66 (1994) 310–313.
[27] M.D. Soumpasis, Y. Georgalis, Potential of mean force treatment of salt mediated protein crystallization Biophys. J. 72 (1997) 2770–2774.
[28] A.A. Chernov, Protein versus conventional crystals: creation of defects J. Cryst. Growth. 174 (1997) 354–361.
[29] A.E. Nielsen, Kinetics of Precipitation, Pergamon Press, Oxford, 1964.
[30] M. Li, A. Nadarajah, M.L. Pusey, Modeling the growth rates of tetragonal lysozyme crystals J. Cryst. Growth. 156 (1995) 121–132.
[31] I.M. Lifshitz, V.V. Slyozov, The kinetics of precipitation from supersaturated solid solutions J. Phys. Chem. Solids 19 (1961) 35–50.
[32] A.M. Kierzek, W.M. Wolf, P. Zielenkiewicz, Simulations of nucleation and early growth stages of protein crystals Biophys. J. 73 (1997) 571–580.
[33] A.M. Kierzek, P. Pokarowski, P. Zielenkiewicz, Lattice simulations of protein crystal formation Biophys. Chem. 77 (1999) 123–137.
[34] A.M. Kierzek, P. Pokarowski, P. Zielenkiewicz, Microscopic model of protein crystal growth Biophys. Chem. 87 (2000) 43–61.
[35] J. Janin, Elusive affinities Proteins 21 (1995) 30–39.
[36] R. Huber, D. Kukla, W. Bode et al., Structure of the complex formed by bovine trypsin and bovine pancreatic inhibitor. II. Crystallographic refinement at 1.9 A resolution J. Mol. Biol. 89 (1) (1974) 73–101.
[37] T.P. Creamer, G.D. Rose, Side chain entropy opposes alpha helix formation but rationalizes experimentally determined helix-forming propensities Proc. Natl. Acad. Sci. USA 89 (1992) 5937–5941.
[38] M.J.E. Sternberg, J.S. Chickos, Protein side chain conformational entropy derived from fusion data comparison with other empirical scales Prot. Eng. 7 (1994) 149–155.
[39] K.P. Murphy, D. Xie, K.C. Garcia, L.M. Amzel, E. Freire, Structural energetics of peptide recognition: Angiotensin II/antibody binding Proteins 15 (2) (1993) 113–120.

[40] L.M. Amzel, Loss of translational entropy in binding, folding and catalysis Proteins 28 (1997) 144–149.
[41] B.R. Brooks, D. Janezic, M. Karplus, Harmonic-analysis of large systems. 1. Methodology J. Comp. Chem. 16 (1995) 1522–1542.
[42] G.P. Brady, K.A. Sharp, Entropy in protein folding and in protein–protein interactions Curr. Opin. Struct. Biol. 7 (1997) 215–221.
[43] A. Tamura, P.L. Privalov, The entropy loss of protein association J. Mol. Biol. 273 (1997) 1048–1060.
[44] J. Janin, M. Karplus, Comment on: the entropy cost of protein association Prot. Eng. 12 (3) (1999) 185–186.
[45] P.L. Privalov, A. Tamura, Comments on the comments Prot. Eng. 12 (3) (1999) 187.
[46] A. Cooper, Thermodynamic analysis of biomolecular interactions Curr. Opin. Chem. Biol. 3 (5) (1999) 557–563.
[47] M.L. Conolly, Solvent accessible surfaces of proteins and nucleic acids Science 221 (4612) (1983) 709–713.
[48] T.J. Richmond, Solvent accessible surface area and excluded volume in proteins J. Mol. Biol. 178 (1984) 63–89.
[49] K.A. Sharp, Electrostatic interactions in macromolecules Curr. Opin. Struct. Biol. 4 (1994) 234–239.
[50] S. Gavryushov, P. Zielenkiewicz, Electrostatic potential and free energy of proteins. A comparison of the Poisson–Boltzmann and the Bogolyubov–Born–Green–Yvon equations J. Phys. Chem. 101 (50) (1997) 10903–10909.
[51] Y. Georgalis, A.M. Kierzek, W. Saenger, Cluster formation in aqueous electrolyte solutions observed by dynamic light scattering J. Phys. Chem. 104 (2000) 3405–3406.
[52] L. Degreve, F.L.B. da Silva, Detailed microscopic study of 1 M aqueous NaCl solution by computer simulations J. Mol. Liquids 87 (2000) 217–232.
[53] L. Degreve, F.L.B. da Silva, Large ionic clusters in concentrated aqueous NaCl solution J. Chem. Phys. 111 (1999) 5150–5156.
[54] L. Degreve, F.L.B. da Silva, Structure of concentrated aqueous NaCl solution: A Monte Carlo study J. Chem. Phys. 110 (1999) 3070–3078.
[55] P.E. Smith, Computer simulation of cosolvent effects on hydrophobic hydration J. Phys. Chem. B 103 (1999) 525–534.
[56] J. Novotny, R.E. Bruccoleri, M. Davis, K.A. Sharp, Empirical free energy calculations: a blind test and further improvements to the method J. Mol. Biol. 268 (1997) 401–411.
[57] W. Kauzmann, Some factors in the interpretation of protein denaturation Adv. Protein. Chem. 14 (1959) 1–63.
[58] C. Chothia, J. Janin, Principles of protein–protein recognition Nature 256 (1975) 705–708.
[59] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding Nature 319 (1986) 199–203.
[60] N. Horton, M. Lewis, Calculation of the free energy of association for protein complexes Prot. Sci. 1 (1992) 169–181.
[61] B. Vallone, A.E. Miele, P. Vecchini, E. Chiancone, M. Brunori, Free energy of burying hydrophobic residues in the interface between protein subunits Proc. Natl. Acad. Sci. USA 95 (11) (1998) 6103–6107.
[62] M.V. Smoluchowski, Versuch einer mathematischen theorie der koagulationskinetik kolloider losungen Z. Phys. Chem. 92 (1917) 129–168.
[63] S.H. Northrup, H.P. Erickson, Kinetics of protein–protein association explained by Brownian dynamics computer simulations Proc. Natl. Acad. Sci. USA 89 (1992) 3338–3342.
[64] G. Schreiber, A. Fersht, Rapid, electrostatically assisted association of proteins Nat. Struct. Biol. 3 (1996) 427–431.
[65] J.D. Madura, J.M. Briggs, R.C. Wade et al., Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program Comp. Phys. Commun. 91 (1995) 57–95.
[66] R.R. Gabdoulline, R.C. Wade, Simulation of the diffusional association of barnase and barstar Bipohys J. 72 (1997) 1917–1929.
[67] J. Antosiewicz, J.M. Briggs, J.A. McCammon, Orientational steering in enzyme-substrate association: ionic strength dependence of hydrodynamic torque effects Eur. Biophys. J. 24 (3) (1996) 137–141.
[68] J. Janin, The kinetics of protein–protein recognition Proteins 28 (1997) 153–161.
[69] S.D. Durbin, G. Feher, Simulation of lysozyme crystal growth by the Monte Carlo method J. Cryst. Growth 110 (1991) 41–51.
[70] L. Bragg, J.F. Nye, Dynamic model of crystal structure Proc. R. Soc. Lond. 190 (1947) 474–481.
[71] G.H. Gilmer, Computer models of crystal growth Science 208 (1980) 355–363.
[72] G.H. Gilmer, P. Bennema, Simulation of crystal growth with surface diffusion, J. Appl. Phys. 43 (4) (1972) 1347-1360.
[73] R. Xiao, J.I.D. Alexander, F. Rosenberger, Growth morphologies of crystal surfaces Phys. Rev. A 43 (6) (1991) 2977–2993.
[74] R. Xiao, J.I.D. Alexander, F. Rosenberger, Morphological evolution of growing crystals: a Monte Carlo simulation Phys. Rev. A 38 (5) (1988) 2447–2456.
[75] J.T.W.M. Tissen, J. Drenth, H.J.C. Berendsen, J.G.E.M. Fraaije, Simulation of protein crystallization. I. Static calculation Biophys. J. 67 (1994) 1801–1805.
[76] A. George, W.W. Wilson, Predicting protein crystallization from a dilute solution property Acta Cryst. D 50 (1994) 361.
[77] A. George, Y. Chiang, B. Guo, A. Arabashi, Z. Cai, W.W. Wilson, Second virial coefficient as predictor in protein crystal growth: macromolecular crystallography Methods Enzymol. A 276 (1997) 100.

[78] B.L. Neal, D. Asthagiri, A.M. Lenhoff, Molecular origins of osmotic second virial coefficients of proteins Biophys. J. 75 (5) (1998) 2469–2477.
[79] B.L. Neal, D. Asthagiri, O.D. Velev, A.M. Lenhoff, E.W. Kaler, Why is the osmotic second virial coefficient related to protein crystallization? J. Cryst. Growth. 196 (1999) 377–387.
[80] D. Asthagiri, A.M. Lenhoff, D.T. Gallagher, Role of competitive interactions in growth rate trends of subtilisin s88 crystals J. Cryst. Growth. 212 (2000) 543–554.
[81] J. Janin, Quantifying biological specificity: the statistical mechanics of molecular recognition Proteins 25 (1996) 438–445.
[82] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Clarendon Press, Oxford, 1989.
[83] P.R. Wolde, D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations Science 277 (1997) 1975–1978.
[84] M.H.J. Haagen, D. Frenkel, Determination of phase diagrams for the hard-core attractive Yukawa systems J. Chem. Phys. 101 (1994) 4097.
[85] G.M. Torrie, J.P. Valleau, Monte Carlo free energy estimates using non-Boltzmann sampling: application to the sub-critical Lennard-Jones fluid Chem. Phys. Lett. 28 (1974) 578–581.
[86] Y. Georgalis, P. Umbach, D.M. Soumpasis, W. Saenger, Dynamics of microstructure formation during nucleation of lysozyme solutions J. Am. Chem. Soc. 120 (1998) 5539–5548.
[87] A. Nadarajah, L. Meirong, M.L. Pusey, Growth mechanism of the (110) face of the tetragonal lysozyme crystals Acta Cryst. D 53 (1997) 524–534.
[88] A. Nadarajah, E.L. Forsythe, M.L. Pusey, Growth of (101) faces of tetragonal lysozyme crystals: measured growth-rate trends Acta Cryst. D 55 (1999) 1005–1011.
[89] S.Y. Patro, T.M. Przybycien, Simulations of reversible protein aggregate and crystal structure Biophys. J. 70 (1996) 2888–2902.
[90] M. Pellegrini, W.S. Wukovitz, T.O. Yeates, Simulation of protein crystal nucleation Proteins 28 (1997) 515–521.