Effect of Molecular Anisotropy on the Nucleation of Lysozyme

Jan Drenth,* Klaas Dijkstra, and Cor Haas

Laboratory for Biophysical Chemistry, Nyenborgh 4, 9747 AG Groningen, The Netherlands

Jörg Leppert and Oliver Ohlenschläger

Abteilung Molekulare Biophysik/NMR Spektroskopie, Institut für Molekulare Biotechnologie, Postfach 100813, 07708 Jena, Germany

Received: November 20, 2002

The growth of protein crystals is preceded by an induction period during which the protein solution prepares itself for crystallization. We have measured the length of the induction period for lysozyme as a function of the temperature for a solution of 16.9 mg/mL at pH 4.5 with 5% NaCl as the precipitating agent. The results are described with classical nucleation theory. The number of molecules in the critical nucleus n^* turns out to be quite small: $n^* = 3-9$. Such a small aggregate can hardly be considered as crystalline, as is assumed in the classical theory. By taking into account the anisotropy of the interaction between lysozyme molecules, we show that small nuclei, with $n < 6$, will have a linear, fiberlike structure. Larger nuclei have a more compact structure. The energy of small, linear nuclei smoothly joins the values for larger nuclei and large crystalline aggregates. The data measured for the induction time can be grouped into two regions, one with short and one with long induction times. The transition point is near 22 °C. This implies that the best starting point for crystallization of lysozyme under the conditions we used is at or just below this temperature. The transition point corresponds approximately to the change from linear to more compact nuclei.

1. Introduction

Protein crystals are, like all crystals, born in a nucleation process. Classical nucleation theory starts with Gibbs, around 1900.[1] In this theory, spherical crystalline embryos are expected to be formed spontaneously in the supersaturated solution by a thermally activated process. The theory has been successfully applied to small molecule crystallization. In the beginning of the 20th century it was not yet known what protein molecules looked like and it took half a century before X-ray crystallography supplied a reliable picture.[2] Interest in the study of protein crystallization was fuelled by its essential role in protein X-ray crystallography. In this article, we present data on the induction time for lysozyme crystallization as derived from NMR spectral information. The solution contained 16.9 mg/mL lysozyme with 5% NaCl and had a pH of 4.5. The data were evaluated with classical nucleation theory resulting in information on the size of the critical nucleus and on the surface tension of this nucleus. The low values we find are in agreement with previously published data.[3] However, they disagree with the assumptions in the classical theory: spherical nuclei with a crystalline arrangement of the molecules. An explanation of this disagreement will be given, based on the anisotropy of the lysozyme molecule. An important feature for crystallization experiments is the rather sudden change in nucleation time near 22 °C, corresponding to a supersaturation $S^* = 6.4$.

2. Materials and Methods

For the experiments presented in this paper, a three times crystallized, dialyzed, and lyophilized lysozyme preparation of Sigma (lot 20K0956) was used without further purification. The protein content, checked by absorption spectroscopy, was 92.4%. Solubility data were taken from Cacioppo and Pusey.[4] The protein was dissolved in 0.1 M Na-acetate buffer pH 4.5 to an effective concentration of 33.8 mg/mL. This solution contained 7% D_2O for locking the magnetic field during the experiment. Addition of D_2O does not affect seriously the behavior of lysozyme in solution. It has only a minor effect on the value of the critical temperature T_c.[5−7] To achieve supersaturation, an equal volume of 10% NaCl in the same Na-acetate buffer and also containing 7% D_2O was added to the protein solution while violently stirring with a vortex mixer. If this is not done, the addition of the 10% NaCl solution causes locally high salt concentrations with a bad effect on the results. The final protein concentration was 16.9 mg/mL. After centrifugation at 15 000 rpm for 5 min, the solution was transferred to an NMR tube and inserted into the NMR spectrometer, a 600 MHz Varian $^{\text{Unity}}$*INOVA* instrument. Spectra were collected every 15 min. The integrated intensity of a peak at 0.8 ppm was followed as a function of time. We have selected the NMR peak at 0.8 ppm because it is rather sharp and has a low noise level. It can be attributed to methyl groups. The line width stays the same during the entire NMR experiment indicating the presence of monomers in the solution. We further point out that the decrease in the NMR spectral intensity is true for the entire spectrum. The spectrum does not change shape. The intensity decreases because monomeric lysozyme molecules have left the solution to be incorporated in aggregates/crystals. The intensity change is connected to the overall properties of the protein molecule, not to any specific proton interactions. One peak of the acetate spectrum was chosen for checking the homogeneity of the magnetic field during the experiment. The NMR curve starts as a horizontal line and bends downward at the moment the first crystals start to grow. This signals the end of the induction period. Previously, a large number of similar experiments have

* To whom correspondence should be addressed. Phone: 31-50-3634382. Fax: 31-50-3634800. E-mail: j.drenth@chem.rug.nl.

10.1021/jp022452z CCC: $25.00
Published on Web 04/08/2003

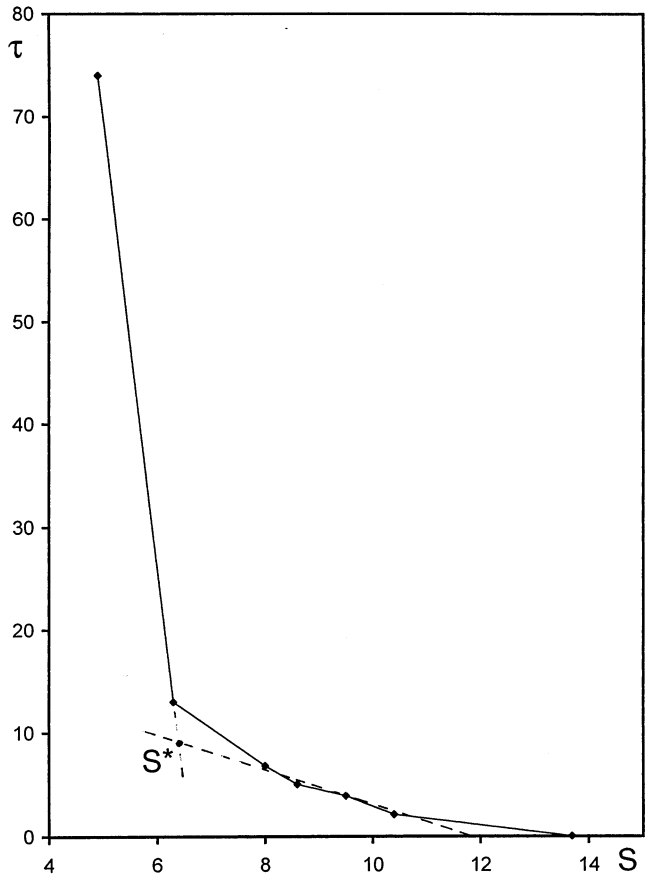

Figure 1. Induction time τ (in hours) plotted as a function of the supersaturation S. The point S^* is taken as the transition point between the regions with short and with long induction times.

TABLE 1: Induction Time τ as a Function of Temperature and Supersaturation S[a]

°C	14	17	18	19	20	22.5	25
τ (hours)	<0.2	2.1	3.9	5.0	6.8	13	74
S	13.7	10.4	9.5	8.6	8.0	6.3	4.9

[a] $S = (16.9\ \text{mg/mL})/\varphi_s$; φ_s is the solubility reported by Cacioppo and Pusey (ref. 4).

been performed with different lysozyme preparations, under different experimental conditions, and in different NMR spectrometers. It turned out that the protein samples did not behave equally. It seems that their solubility increases with age, even if kept in the refrigerator. This could be due to deamidation of Asn or Gln residues. Aswad[8] lists five positions in lysozyme at which deamidation is likely but not proven.

3. Results

With the NMR technique, we have measured the induction time for lysozyme crystallization at different temperatures. The concentration of the lysozyme in the solutions was 16.9 mg/mL and the precipitating agent 5% NaCl. We observe a strong increase in induction time above 22 °C as a function of the supersaturation (Table 1 and Figure 1). The data were processed with classical nucleation theory. In this theory, it is assumed that the nucleus is a crystalline spherical particle with radius r and volume $V = 4\pi r^3/3$. The number of molecules in the nucleus $n = V/\upsilon_o$, where υ_o is the volume occupied by one molecule in the crystal. The Gibbs free energy of a nucleus is given by

$$\Delta G = -n\Delta\mu + \Delta G' \qquad (1)$$

where $\Delta\mu$ is the difference in thermodynamic potential between a molecule in the bulk of a crystal and a molecule in the solution and $\Delta G'$ is the difference between the Gibbs free energy of the nucleus and the Gibbs free energy of n molecules in the bulk of a crystal. For a spherical crystalline nucleus with radius r, $\Delta G'$ is the surface energy of the nucleus and is $4\pi r^2\gamma$, where γ is the surface energy per cm^2 of the crystal in the solution.

It is easily shown that $\Delta\mu = kT \ln S$, where $S = \varphi/\varphi_s$ is the supersaturation. φ is the concentration of the protein, and φ_s is the solubility.

Nucleation is the thermally activated process of the formation of a critical nucleus with radius r^* containing n^* molecules and having a Gibbs free energy ΔG^*. The critical nucleus will spontaneously grow in the solution. The values of ΔG^*, r^*, and n^* are obtained by the condition $d\Delta G/dr = 0$ and are given by

$$\Delta G^* = \frac{16\pi\upsilon_o^{\,2}\gamma^3}{3(kT)^2(\ln S)^2}, \quad r^* = \frac{2\upsilon_o\gamma}{kT\ln S}, \quad n^* = \frac{4\pi(r^*)^3}{3\upsilon_o} \qquad (2)$$

The rate of nucleation is proportional to $\exp[-\Delta G^*/kT]$. The corresponding induction time is usually taken as

$$\tau = \tau_o \exp[\Delta G^*/kT] \qquad (3)$$

This is not correct because, for $\Delta G^* = 0$, τ should be 0. The more accurate expression is (see Appendix 1)

$$\tau = \tau_o \exp\left[\frac{\Delta G^*}{kT}\right] - \tau_o \qquad (4)$$

Because the effect of $-\tau_o$ in eq 4 is small, we used the simpler eq 3 for deriving γ

$$\ln\tau = \ln\tau_o + \frac{16\pi\upsilon_o^{\,2}\gamma^3}{3(kT)^3(\ln S)^2} \qquad (5)$$

To test eq 5, we plotted $\ln\tau$ as a function of $1/[T^3(\ln S)^2]$ (Figure 2); from the slope, we can derive γ. The graph consists of two straight lines with the upper one shifted to the right with respect to the lower one. Although it is unclear whether this shift is real or an experimental artifact, the slopes of the two lines are approximately equal and result in a γ value of 0.44 erg/cm^2. This value is used to calculate the critical radius and the number of molecules in the critical nucleus (Table 2).

4. Discussion

A. Induction Time. The induction time is the period between the moment supersaturation has been established and the appearance of the first aggregates as measured by NMR. It consists of three phases (Walton, D. in ref 1). In the first phase, a redistribution of thermal fluctuations in the protein concentration takes place until a steady state of embryo distribution is attained. In the second phase, nuclei are formed by a thermally activated process, and in the third phase, these nuclei grow to crystals. The data of Figure 2 show that the observed induction time follows closely the strong exponential dependence on $1/[T^3(\ln S)^2]$. This indicates that the main contribution to the induction time comes from phase 2, the actual nucleus formation. The induction time contributions of the other phases also depend on S but not in the strong exponential way as for phase 2.

B. Surface Energy. From the data, we deduce a surface energy of $\gamma = 0.44$ erg/cm^2. This value is reasonably close to the value of $\gamma = 0.51-0.64$ erg/cm^2 reported by Galkin and

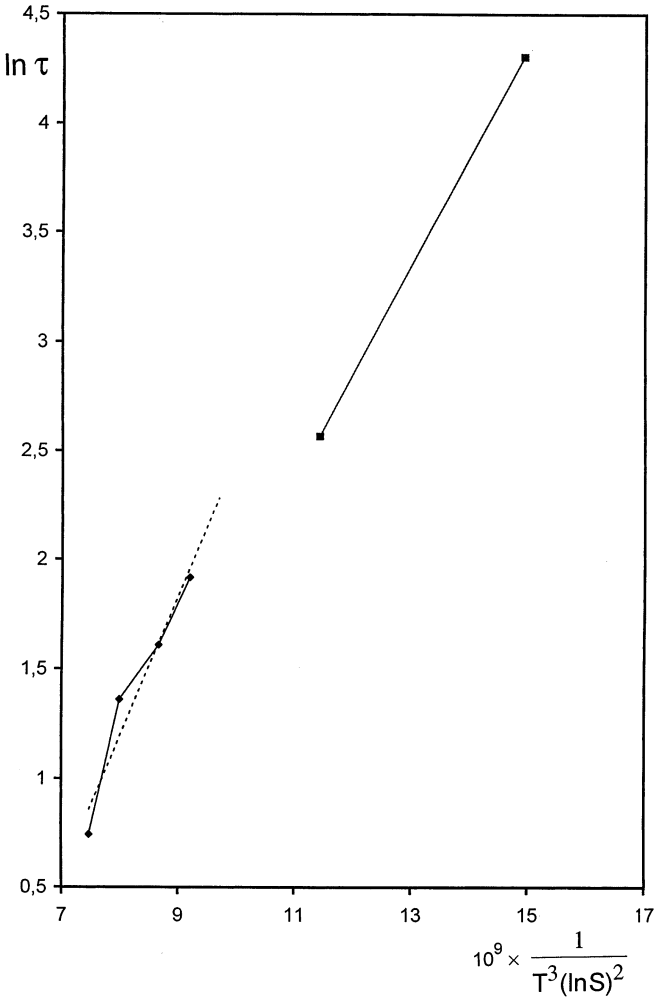

Figure 2. ln τ plotted as a function of $1/T^3(\ln S)^2$. Note the shift to the right for the upper two data points. The slope of the line connecting these two points is approximately the same as the slope of the trendline for the lower data points.

TABLE 2: Radius *r and Number of Molecules *n** in the Critical Nucleus Calculated from the NMR Data of Table 1**

S	*r** (Å)	*n**	*S*	*r** (Å)	*n**	*S*	*r** (Å)	*n**
4.9	39.9	9	8.0	31.0	4	9.5	28.9	3
6.3	34.8	6	8.6	30.1	4	10.4	27.9	3

Vekilov[3]. We can show that this value is also in good agreement with the value expected for a crystal in the aqueous solution.

To calculate the surface energy, we must take into account the anisotropy of the interactions between a lysozyme molecule and its neighbors. Durbin and Feher[9] derived for lysozyme from data published by Blake et al.[10] the existence of four contact sites: Y_1, Y_2, and two times X. We have used the lysozyme PDB file 193L (http://www.rcsb.org/pdb/) and found with the program WHAT IF[11] that in the crystal each lysozyme molecule is surrounded by two molecules at short distance and by two other molecules at a somewhat larger distance (Appendix 2). This corresponds to two strong interactions X and two weak interactions Y_1 and Y_2 between a molecule and its neighbors. In the crystal, the energy per molecule is $g_c = {}^1/_2(2X + Y_1 + Y_2)$. During growth a crystal will be formed bounded by facets of the lowest energy. For lysozyme these are (110) and (101) surfaces.[12] We have selected the (110) surfaces for our calculation. Each of the molecules at the (110) surfaces has one broken weak bond Y_1 or Y_2. Therefore, the surface energy of the crystal will be $\gamma_o = Y/2A$, where Y is the average value of the two weak bonds: $Y = {}^1/_2(Y_1 + Y_2)$. $A = 10.6 \times 10^{-14}$ cm^2 is the surface per molecule on the (110) plane. The value of g_c can be calculated from the solubility: $g_c = kT \ln(\varphi_s/m)$ with $m = 790$ for lysozyme.[13] From the molecular interactions, we can estimate the ratio of the strength of the strong and weak interactions: $X/Y = 2.4$ (Appendix 2). From this ratio and the value of $g_c = -51.6 \times 10^{-14}$ erg at 294 K,[14] we find $X = 36 \times 10^{-14}$ erg and $Y = 15 \times 10^{-14}$ erg and a surface energy $\gamma_0 = 0.72$ erg /cm^2 at 294 °K.

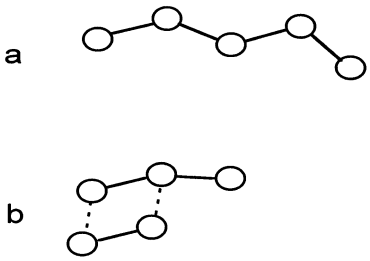

Figure 3. a. Linear aggregate of five molecules with four strong bonds: energy $G_{linear} = 4X$. b. Compact aggregate of five molecules with three strong and two weak bonds; energy $G_{compact} = 3X + 2Y$.

We have previously shown that a thin film with a high protein concentration surrounds a protein crystal. This is particularly true near and below the critical temperature T_c for liquid/liquid-phase separation.[15] The presence of such a film reduces the surface energy of a protein crystal but also of the small aggregates that form in the nucleation process. It is estimated to be $\Delta\gamma = 0.2 - 0.3$ erg/cm^2. We expect therefore that the surface energy γ of a crystal in an aqueous solution at or near T_c will be $\gamma = \gamma_o - \Delta\gamma \cong 0.4-0.5$ erg/cm^2, in good agreement with the value of 0.44 erg/cm^2 deduced from the NMR experiments.

C. Nature of the Critical Nucleus. A remarkable feature of the analysis presented above is that the critical nuclei consist of only a small number of molecules; *n** has values between 3 and 9 (Table 2). Such small nuclei can hardly be considered as the spherical crystalline particles assumed in classical nucleation theory. Moreover the anisotropy of the interactions between the protein molecules prevents the formation of compact, crystalline aggregates; instead, fiber-type nuclei are expected to be formed because they have a lower energy than compact nuclei with the same number of molecules. This can be easily shown with a simple example (see Figure 3). A fiber-type aggregate of five molecules will have four strong bonds and an energy $G_a = 4X$, whereas a compact aggregate of five molecules has three strong and two weak bonds, and an energy $G_b = 3X + 2Y$. Thus, for an aggregate of five molecules, the fiber-type is more stable than the compact aggregate if $X > 2Y$, and this is presumably the case for lysozyme (Appendix 2). The linear aggregates may well have a helical arrangement of the molecules.[16,17] One can easily show that for an anisotropy of the bonds $X/Y = 2.4$, as is expected for lysozyme, the linear aggregates are more stable than the compact aggregates for $n^* < 6$.

Electron microscopy investigations on the nucleation process of lysozyme by Michinomae et al. show that in the initial state short linear aggregates are formed with a thickness of one molecule and a length corresponding to 4 molecules on the average.[18] Georgalis et al.[19] deduced a fractal character of the nuclei from light scattering experiments, indicating a noncrystalline structure of the nuclei with fractal dimension smaller than three. These data are in good agreement with the considerations based on the anisotropy of the interactions between the protein molecules.

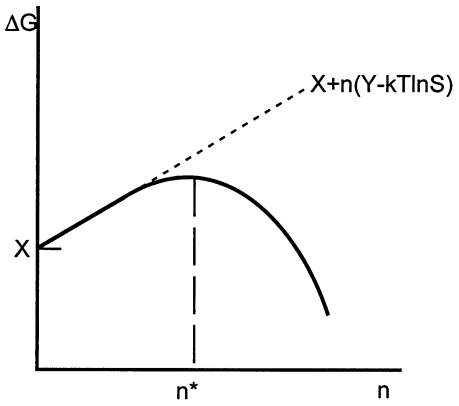

Figure 4. ΔG for the smooth transition from linear to more compact aggregates with n molecules in the aggregate. The free energy for linear aggregates, $\Delta G = X + n(Y - kT \ln S)$, is a linear function of n and is represented by the dotted line.

D. Applicability of Classical Nucleation. In the preceding sections, we showed that the critical nucleus for lysozyme contains only a small number of molecules and that for large supersaturation ($S > 6$) the nucleus is expected to have a linear and not a compact structure. These small nuclei cannot be considered as spherical crystalline particles. Therefore, it is quite surprising that classical nucleation theory describes quite accurately the nucleation data with a value of the surface energy close to what is expected for a crystal in the aqueous solution. In this section, we explain this somewhat surprising result.

It can be shown that the energy of the aggregates smoothly adapts from the smaller to the larger particles. The small nuclei are linear aggregates. For n lysozyme molecules in a linear arrangement, the energy contribution $\Delta G'$ in eq 1 is

$$\Delta G' = ng_c - (n-1)X = X + nY \tag{6}$$

The total energy ΔG is

$$\Delta G = -nkT \ln S + X + nY = X + n(Y - kT \ln S) \tag{7}$$

For large values of n, compact clusters become more stable than the linear aggregates. As a consequence, the ΔG curve in Figure 4 will gradually bend downward with a maximum at n^*. This indicates that in the entire range of aggregates the ΔG curve has a smooth shape expected for classical nucleation, even if the smaller aggregates do not have a crystalline arrangement.

An interesting effect is expected for large supersaturation. For $\ln S > Y/kT$, the straight line for linear aggregates in Figure 4 has a negative slope, and therefore, the ΔG vs n curve does not have a maximum. In that case, nucleation will start with single molecules, the size of the critical nucleus being $n^* = 1$. Galkin and Vekilov[3] observed in lysozyme solutions with 4% NaCl a sudden change of the size of the critical nucleus from a value of $n^* = 4.7$ at low supersaturation to a value of $n^* = 1$ or 2 at supersaturation values larger than $S = 22$. This would, in our interpretation, correspond to a value $Y = kT \ln S = 12.8 \times 10^{-14}$ erg which is not far from the estimated value of $Y = 15.8 \times 10^{-14}$ erg for a lysozyme solution with 5% NaCl. This sudden change of n^* is a direct consequence of the strong anisotropy of the interactions between the lysozyme molecules.

E. Induction Time vs Supersaturation. The curve in Figure 1, showing the induction time τ as a function of supersaturation, has a sharp increase of the induction time below $S^* = 6.4$. The value of S^* is important for a practical reason. It is generally accepted that the growth of protein crystals should ideally start in conditions with a relatively high nucleation speed. Subsequent growth should then occur in conditions in which the appearance of more nuclei is limited. These requirements are automatically fulfilled if nucleation starts at or just above S^* in the region of short induction times (or fast nucleation). The relatively high nucleation speed drops considerably when a small amount of protein has been transferred to the crystalline state, favoring continued growth of the existing crystals.

The sharp increase of the induction time τ for lysozyme nucleation above $S^* = 6.4$ is just at the point where a change in the nature of the nuclei from linear to more compact aggregates is expected. Another interesting feature is that $S^* = 6.4$ corresponds to a temperature of 22.4 °C. This is very close to the critical temperature $T_c = 22.5$ °C for liquid/liquid-phase separation as found by Muschol and Rosenberger[20] for lysozyme and 5% NaCl at pH 4.5. Whether this is a coincidence or not will be the subject of further study.

5. Conclusion

We have measured the induction time in lysozyme crystallization experiments by means of NMR. The data can be explained with classical nucleation theory. This theory assumes the formation of nuclei with a crystalline arrangement of the molecules. In lysozyme nucleation, this is not applicable, because of the anisotropic character of protein molecules. Nuclei consisting of a small number of molecules prefer to be linear. Nevertheless, the data can be processed according to classical nucleation because there is a smooth transition between the regions without and with a crystalline arrangement. The anisotropic molecular shape distinguishes proteins from colloids. Association of protein molecules results in the formation of species in which the strongest interactions play a dominant role. This expresses itself also in the final crystal structures, which are far more complicated than colloidal crystal structures. The values measured for the induction time as a function of supersaturation can be grouped into two regions, one with short and one with long induction times. We have pointed out that the most favorable conditions for starting a crystallization experiment are near this transition point. It is remarkable that the transition point between the two regions corresponds to the critical temperature for liquid/liquid-phase separation at 5% NaCl. Moreover, at this point, the character of the nuclei changes from linear to more compact.

Acknowledgment. This work was supported by the Commission of the European Community under Contract BIO4-CT98-0086-European Biocrystallogenesis Initiative, BIOTECH, and Life Sciences and Technologies, Biotechnologie Programme (1994-1998), and under Contract No. FMGE-CT98-0121, IMB Large Scale Facility, Center for Design and Structure in Biology (CDSB), Jena, under Access to Large-Scale Facilities Activity of the Specific Program for RTD, in the Field of the Training and Mobility of Researchers. J.D. thanks professor B.W. Dijkstra for the hospitality in his laboratory. He is also especially grateful for the hospitality at the NMR group of the IMB in Jena. We thank professor J.-M. Garcia Ruiz for critical information on classical nucleation, professor Vriend for assistance with WHAT IF, and Dr. R. M. Scheek for NMR advice.

Appendix 1

The growth speed of the nucleus is a thermally activated process. For a nucleus with radius r, the growth speed is proportional to $\exp[-\Delta G(r)/kT]$ where $\Delta G(r)$ is the energy barrier for the incorporation of an additional molecule in the nucleus. During the induction time τ, the energy barrier $\Delta G(r)$

increases from 0 to $\Delta G(r^*)$, the value for the critical nucleus. The time required for the energy barrier $\Delta G(r)$ to increase by the amount $d\Delta G(r)$ is

$$dt = d\Delta G(r)/(A \exp[-\Delta G(r)/kT]) = \frac{1}{A} \exp[\Delta G(r)/kT] d\Delta G(r)$$

The induction time τ is given by

$$\tau = \int_0^\tau dt = 1/A \int_{\Delta G(r)=0}^{\Delta G(r^*)} \exp[\Delta G(r)/kT]\, d\Delta G(r) = \tau_0 \exp[\Delta G(r^*)/kT] - \tau_0 \text{ with } \tau_0 = kT/A$$

Appendix 2

With the program WHAT IF,[11] lysozyme file 193L in the PDB was evaluated for contacts between partner molecules. Contacts were listed if the distance between their van der Waals spheres is less than 0.5 Å. H bonds were assigned a relative strength between 0.1 and 1.The majority of the contacts is of the van der Waals type. All of them were given a strength of 0.4. From the analysis, we obtained for the relative strength of the interactions $X:Y_1:Y_2 = 24.8:11.8:8.4$.

Durbin and Feher[9] derived for lysozyme from data published by Blake et al.[10] the existence of four contact sites: Y_1, Y_2, and two times X. In Monte Carlo model simulations, the main features of observed crystal growth rate and surface structure were calculated. The interaction energies X, Y_1, and Y_2 were introduced as parameters and were varied to obtain the best fit between observed and calculated growth rate and surface morphology. The best fit gave $X:Y_1:Y_2 = 5:4:2$. With an average strength $Y = {}^1/_2(Y_1 + Y_2) = 3$ for the weaker bonds and a strength of 5 for the two stronger bonds. Their ratio of the strength of the strong to weak interactions is 1.7, much lower than our estimate of 2.4. We prefer to use our value based on direct calculation of distances.

References and Notes

(1) Zettlemoyer, A. C. *Nucleation*; Marcel Dekker: New York, 1969; p 13.
(2) Perutz, M. F. *Endeavour* **1958**, *17*, 190.
(3) Galkin, O.; Vekilov, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 156.
(4) Cacioppo, E.; Pusey, M. L. *J. Crystal Growth* **1991**, *114*, 286.
(5) Boué, F.; Lefaucheux, F.; Robert, M. C.; Rosenman, I. *J. Crystal Growth* **1993**, *133*, 246.
(6) Niimura, N.; Minezaki, Y.; Ataka, M.; Katsura, T. *J. Crystal Growth* **1995**, *154*, 136.
(7) Gripon, C.; Legrand, L.; Rosenman, I.; Vidal, O.; Robert, M. C.; Boué, F. *J. Crystal Growth* **1997**, *177*, 238.
(8) Aswad, D. W. *Deamidation and Aspartate Formation in Peptides and Proteins*; CRC Press: Boca Raton, FL, 1995; p 232.
(9) Durbin, S. D.; Feher, G. *J. Crystal Growth* **1991**, *110*, 41.
(10) Blake, C. C. F.; Mair, G. A.; North, A. C. T.; Philips, D. C.; Sarma,V. R. *Proc. R. Soc. (London)* **1967**, *B167*, 365.
(11) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52.
(12) Grimsbergen, R. F. P.; Boek, E. S.; Meekes, H.; Bennema, P. *J. Crystal Growth* **1999**, *207*, 112.
(13) Haas, C.; Drenth, J. *J. Crystal Growth* **1995**, *154*, 126.
(14) Haas, C.; Drenth, J. *J. Phys. Chem. B* **1998**, *102*, 4226.
(15) Haas, C.; Drenth, J. *J. Phys. Chem. B* **2000**, *104*, 368.
(16) Nadarajah, A.; Pusey, M. L. *Acta Crystallogr.* **1996**, *D52*, 983.
(17) Strom, C. S.; Bennema, P. *J. Crystal Growth* **1997**, *173*, 150 and 159.
(18) Michinomae, M, Mochizuki, M.; Ataka, M. *J. Crystal Growth* **1999**, *197*, 257.
(19) Georgalis, Y.; Umbach, P.; Saenger, W.; Ihmels, B.; Soumpasis, D. M. *J. Am. Chem. Soc.* **1999**, *121*, 1627.
(20) Muschol, M.; Rosenberger, F. *J. Chem. Phys.* **1997**, *107*, 1953.