PROTEINS: Structure, Function, and Genetics 28:515–521 (1997)

Simulation of Protein Crystal Nucleation

Matteo Pellegrini,[1] Stephanie W. Wukovitz,[2] and Todd O. Yeates[1,3]*
[1]*Molecular Biology Institute, University of California, Los Angeles, California 90095*
[2]*Department of Biomathematics, University of California, Los Angeles, California 90095*
[3]*Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095*

ABSTRACT **To attempt to understand the physical principles underlying protein crystallization, an algorithm is described for simulating the crystal nucleation event computationally. The validity of the approach is supported by its ability to reproduce closely the well-known preference of proteins for particular space group symmetries. The success of the algorithm supports a recent argument that protein crystallization is limited primarily by the entropic effects of geometric restrictions imposed during nucleation, rather than particular energetic factors. These simulations provide a new tool for attacking the problem of protein crystallization by allowing quantitative evaluation of new ideas such as the use of racemic protein mixtures. Proteins 28:515–521, 1997.** © 1997 Wiley-Liss, Inc.

Key words: protein crystallization; space group symmetry; protein-protein contacts; aggregation; critical nuclei

INTRODUCTION

The growing capacity of computers is allowing scientists to model complex nonequilibrium systems that would otherwise be difficult to investigate by analytical techniques. The nucleation and growth of protein crystals is such a system. Although considerable attention has been given to the protein crystallization process, theoretical understanding of the phenomenon is still in its early stages.[1,2] For instance, the connection between space group symmetry and the probability of nucleation has not yet been studied; we will attempt to demonstrate that nuclei preferentially adopt certain space groups symmetries.

To study this phenomenon, we present a simulation to construct the critical nuclei of protein crystals. One of the principal questions we ask is whether the symmetries of the simulated nuclei match the strikingly nonuniform distribution of space group symmetries found in the Brookhaven Protein Data Bank (PDB). Several attempts have already been made to understand the highly nonuniform occurrence of space groups for crystals of both small molecules[3–5] and proteins.[6,7] Previous work of two of the authors successfully related space group frequencies to the number of rigid-body degrees of freedom inherent in each symmetry.[6] Unlike the above work, however, the technique presented here has the capacity of generating quantitative predictions of this distribution for crystal nuclei.

Our approach attempts to mimic the physical reality of protein crystallization to the extent that it is feasible with current computational power. We present four hypotheses that underlie the algorithm and attempt to justify the assumptions on physical grounds.

We assume that crystallization is a nucleated phenomenon; there is a positive free-energy activation barrier to the formation of a nucleus of a few hundred molecules. Once the nucleus has formed, the growth process is thermodynamically favored. It is still unknown experimentally whether the nucleus is a crystallographically ordered or disordered aggregate of proteins. Because the primary purpose of this simulation is to study the effect of space group symmetries on nucleation and not to model the thermodynamics of crystal growth, we assume that the nucleus is crystallographically ordered from the start. If, in fact, protein crystal nuclei are disordered, our conclusion would not be significantly affected because our ordered aggregate must, at some step, appear within the disordered nucleus. In this case, we shall refer to the emergence of crystalline order as nucleation and not to the initial formation of a disordered cluster.

Next, we suggest that for a crystal to form, the set of all possible contacts between two identical monomers must be characterized by a few low-energy contacts that are separated from all others by a positive energy gap; in the case in which the lowest energy contacts are greatly degenerate or the energy gap approaches zero, it would not be possible to establish long-range order in an aggregate. Similar ideas have been invoked in connection with protein folding.[8] Second, the few lowest-energy contacts must correspond to molecular orientations close to the crystallographically allowed values (0°, 60°, 90°, 120°, or 180°) for the aggregate to be well ordered.

*Correspondence to: Todd O. Yeates, Molecular Biology Institute, University of California, Los Angeles, CA 90095.
Received 5 November 1996; Accepted 27 February 1997

Finally, we postulate that attachment of each new molecule must initially occur reversibly for it to sample many contacts until it binds in accordance with one of the low-energy contacts. Our simulation attempts to mimic these three conditions by allowing monomers to contact each other in a small number of ways, all of which are consistent with allowed crystallographic angles. Inherent in the algorithm is the restriction that the crystals are formed out of monomers, with only one molecule per asymmetric unit.

Applying the concepts of the first hypothesis, we initially construct clusters consisting of four molecules to define a limited number of low-energy contacts. We choose a cluster of size four based on a recent result[6] regarding the minimum number of distinct contact types (or unrelated spatial transformations) required to achieve connectivity between molecules in the various space group symmetries. (We use the term *contact* to refer to both the actual touching of two molecules and the symmetry operation or spatial transformation that defines the position of one with respect to the other.) The minimum number of distinct contacts ranges from two to five, but space groups for which this number is greater than three are almost never observed. The disposition of four molecules can define as many as three unrelated transformations: from molecule 1 to 2, 1 to 3, and 1 to 4. Other possible pairwise relationships (e.g., 2 to 3) are inferred by combinations of the original three transformations. Therefore, clusters of four molecules are necessary and sufficient to represent most of the space groups (52 of 65 total, including all that are observed with any regularity in the database).

To apply the second hypothesis, we restrict our attention to "crystal nucleation space," the space of molecular arrangements that may lead to a crystal. Because crystal packings of molecules allow only two-, three-, four- or sixfold rotational symmetries, as well as translations, we must restrict the rotation angles between molecules to allowed values (0°, 60°, 90°, 120°, or 180°). The rotation angle is evaluated by computing the 3×3 matrix that brings the coordinates of the two molecules, with respect to their centers of mass, onto each other. The angle is then given by the equation

$$\text{Trace}(R) = 1 + 2\cos(\theta) \qquad (1)$$

where R is the rotation matrix and θ the rotation angle. To arrive at ideal intramolecular rotation angles during the construction of the four-molecule cluster, we retain only pairs of molecules whose relative orientation lies within 10° of an allowed crystallographic value. A Monte Carlo move selection algorithm is then used to further minimize the differences between the actual and ideal values of the angles. This process inevitably raises the intermolecular energy. However, once the ideal angle is achieved, we further relax the contact, with the angle fixed, to find the local energy minimum. Limiting our attention to "crystal nucleation space" should not affect the distribution of crystalline nuclei; the effect is simply to screen out noncrystalline arrangements rapidly.

Finally, applying the principles of our third hypothesis, we must ensure that new molecules added to our original cluster are attached only by using contacts randomly selected from the list defined above. Accordingly, each new molecule is attached by randomly selecting one of the operations derived from the four molecule prenucleus and applying it to one of the molecules already in the cluster. This is repeated, if it is possible to do so without having molecular collisions, until the cluster contains 200 molecules. The number 200 is chosen because it is the approximate size of experimentally measured critical nuclei for proteins, nuclei that self-propagate into a macroscopic crystal.[9] The pairwise transformations between the 200 monomers are sufficiently numerous to permit us to evaluate the degree to which a given nucleus is ordered, as defined below, and thus whether it is crystalline or not. In the case of a crystalline cluster, we establish its space group symmetry.

METHODS

The molecules we use in our simulation are designed to be roughly globular with surfaces that contain "sticky" patches and thus share some basic properties with real proteins. Otherwise, we have not made any attempt to make our molecules resemble proteins in detail, and hence the results of our simulations should apply to all molecules with these characteristics. The molecules are generated as follows: Each one consists of ten overlapping spherical particles (of radius 5–7 Å) whose centers are randomly distributed on the surface of a sphere of radius 15 Å. The molecule is kept rigid throughout the simulation and hence has only three translational and three rotational degrees of freedom. The particles of different molecules interact via pairwise potentials. The potentials are square wells with a hard core repulsive component for interparticle distances <10 Å to 14 Å and an attractive component extending about 2 Å beyond that. The well's origin, width, and depth (from 0 to -1.0 in arbitrary energy units) are randomized during each run, as are the positions of the particles that comprise the molecule.

A flowchart describing the algorithm is shown in Figure 1. The first two molecules are assembled using a Monte Carlo algorithm, with the second molecule initially oriented randomly with respect to the first and moved 50 Å from the first in a random direction. The two molecules are brought into contact using a Metropolis sampling algorithm[10] in which the total energy, the sum of the pairwise energies, is minimized for 100 steps. The two mol-

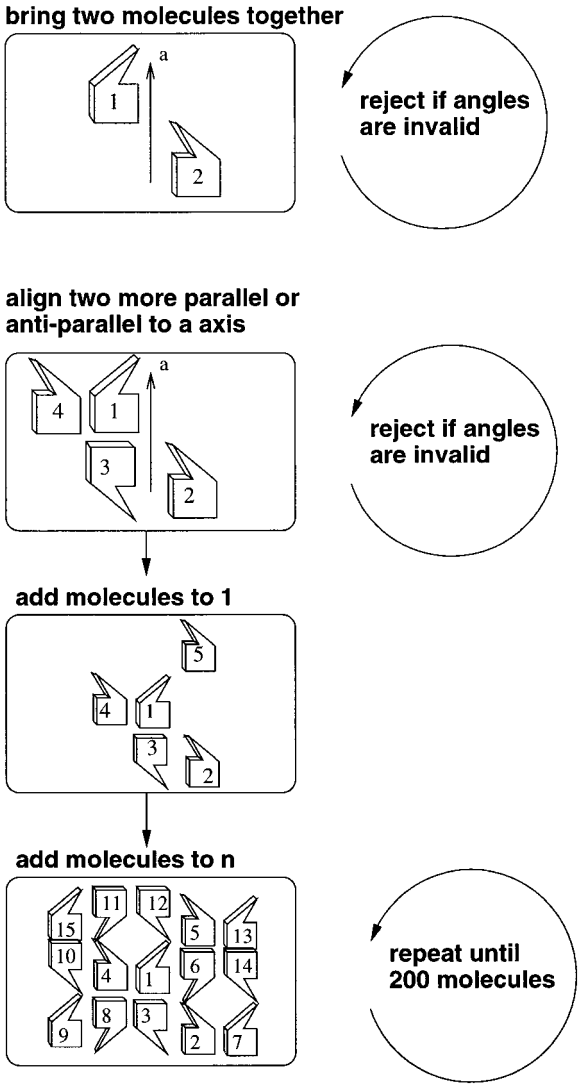

Fig. 1. A flowchart representing schematically the principal steps of the algorithm for assembling protein crystal nuclei. In the chart, we demonstrate the construction of a two-dimensional layer group while in the article we construct three-dimensional clusters. A discussion of the individual steps in the algorithm is found in the Methods section of the text.

ecules are only allowed to diffuse along the vector connecting their centers of mass. This fact greatly restricts their degrees of freedom and hence allows them to come together in few steps. This process does not emulate real diffusion; it is simply used to generate random starting conformations.

The final arrangement is rejected if the rotation angle between the two molecules is not within 10° of a crystallographically allowed value (0°, 60°, 90°, 120°, 180°). The rotation axis relating the first two molecules is defined to lie along the *a* axis of the crystal lattice, which is predestined to be the axis of highest rotational symmetry in the crystal; if the molecules are related by a translation, we continue to attach other molecules until one that is related by a rotation allows the *a* axis to be defined.

The third and fourth molecules are attached as follows: An internal vector, *v*, of the first molecule is defined parallel to the *a* axis. Because the second molecule is simply rotated about the *a* axis, it must also have its *v* vector parallel to it. For all noncubic space groups, the subsequent molecules must have their *v* vectors either parallel or antiparallel to the *a* axis. This reduces the rotational degrees of freedom of the third and subsequent molecules from three to one and is essential in rendering the procedure presented here computationally feasible. To determine the disposition of the third molecule relative to the first two, its *v* vector is chosen randomly to lie either parallel or antiparallel to the *a* axis. As before, we randomly select a vector to displace this molecule's center of mass away from that of the first molecule by 50 Å, randomly rotate it about the *a* axis (once again discarding the molecule if its orientation relative to other molecules is inconsistent with crystallographic symmetry), and bring the molecule into contact with the others by the same procedure as before. Once the molecule's contact energy is minimized, the rotation angle is adjusted, as explained above, to lie within 1° of its allowed value. A fourth molecule is attached in the same manner. This cluster of four molecules defines as many as three unrelated contacts or transformations between molecules.

Starting from the first molecule, we randomly choose a contact from a set containing the original three contacts and their inverses (if they are distinct) and attempt to add a fifth molecule to the cluster corresponding to this transformation. If the new molecule overlaps any of the existing ones, then it is rejected. Otherwise, it is added to the cluster. We continue to add molecules to the first protein until all attempts to use one of the known contacts leads to the addition of a molecule that closely overlaps one of the preexisting ones.

At this point, no more molecules can be attached to the first, so the procedure is continued with the addition of as many nonoverlapping molecules as possible to each successive protein molecule. The process is repeated until a cluster of 200 molecules is obtained but terminates if it becomes impossible to reach a cluster of this size. The fact that the protein attachments are saturated sequentially is also an artifact of our simulation that does not reflect the random attachments during real crystal growth. However, we believe that this fact does not significantly affect the final distribution of nuclei.

One limitation of the present algorithm is that it does not allow us to generate crystals in some cubic space groups. However, this limitation in the simulation is not too severe, because monomeric proteins very rarely crystallize in cubic space groups anyway.

In the final step, we must decide whether our aggregate is ordered. A first test for crystalline order consists of analyzing the distribution of the three contacts and their inverses. In a true crystal, each protein exists in the same environment, hence the three possible contacts must be approximately equally distributed (the surface of the crystal disrupts the perfectly uniform distribution). Because both the forward and inverse transformations are counted together, it must be taken into account that contacts with pure twofold symmetries occur only half as often as other types. Aggregates with highly nonuniform contact distributions are rejected. Our loose criteria for uniformity consist of accepting all distributions in which each of the three unique contacts occurs at least 15% of the time.

If the contacts are sufficiently uniformly distributed, we evaluate the periodicity of the cluster. Translation vectors are examined between all pairs of molecules that have the same orientation (a rotation angle $<1°$ between them). From this list, we extract the three shortest, linearly independent vectors. We then calculate the percentage of vectors from the complete list that may be expressed as a linear combination, with integer coefficients, of the three basis vectors. We accept all coefficients that are within 5% of the nearest integer value. Finally, we keep only clusters for which at least 75% of the vectors may be expressed as integral multiples of the three basis vectors.

Finally, for the nuclei that are ordered we determine the space group symmetry. To uniquely determine a space group, one must know both the underlying lattice of the crystal and the symmetry operation along each axis. A primitive lattice can be constructed from the three basis vectors described above. Linear combinations—with integer coefficients—of these basis vectors are examined to see if a centered cell of higher symmetry can be generated.

Having identified the crystal lattice for a cluster, we next determine the order of any rotational symmetries along the crystal axes, distinguishing between pure rotations and screw rotations. To completely specify the space group, we also need to consider the handedness of the rotations, although in this work, results are reported only for paired enantiomorphic space groups.

RESULTS AND DISCUSSION

The results of the simulations are reported in Table I. Approximately 2,000,000 nuclei of 200 molecules were assembled, of which 2,340 were judged to be crystalline. Two examples of simulated nuclei are shown in Figure 2. The entire simulation required approximately three days on ten Digital Alpha workstations. The results are shown in Figure 3, where the percentage occurrence of each space group found in the simulation is compared with the percentage occurrence found in the PDB, based on 244 dissimilar monomeric proteins from the data bank.[6] From these two distributions, we obtain a linear correlation,

$$C = \frac{\Sigma_i x_i y_i}{(\Sigma_i x_i^2 \Sigma_j y_j^2)^{1/2}} \quad (2)$$

equal to 0.966. Especially striking is the agreement with the well-known observation that space group $P2_12_12_1$ is by far the most popular symmetry for protein crystals.

We also performed a χ^2 test to calculate the probability that the simulated and PDB frequencies were drawn from an identical distribution. The χ^2 value was calculated as follows:[11]

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{17} \frac{(n_{ij} - n_{ij}^e)^2}{n_{ij}^e} \quad (3)$$

In this formulation, $i = 1$ indicates the simulated distribution and $i = 2$ the PDB distribution. Rarely observed space groups were combined, giving a total of 17 categories indicated by the index j, each with a number of occurrences in the data bank greater than five. The expected number of occurrences is given by $n_{ij}^e = (n_{1j} + n_{2j})\, n_i^{tot}/n^{tot}$, in which n_i^{tot} is the sum of all occurrences in i and n^{tot} is the the sum of all occurrences in $i = 1$ and $i = 2$.

We found $\chi^2 = 34$ (with 16 degrees of freedom), implying that 1% of the time, two datasets drawn from the same distribution would be at least as dissimilar as the observed and simulated distributions reported here. Therefore, it appears that the simulated and PDB frequencies are drawn from very similar, but nonidentical, distributions. The simulations capture the vast majority of the information in the observed space group frequencies, but other factors not considered here may contribute to the minor differences. The most notable discrepancy is that the simulations predict the space group $P2$ to be relatively abundant (ninth most seen), although it is not found in the database generated during our earlier study.[6] This discrepancy may be due to the limited number of entries in the database; we have repeated a search on the most recent PDB and found two instances of space group $P2$.

These results are only slightly affected by most of the parameters used in our algorithm. The only parameter that significantly affects the final distribution is the acceptance window of the initial rotation angle between two molecules before it is adjusted to one of its ideal values. As mentioned above, we reject configurations of the initial two molecules with orientations that differ by $>10°$ of one of the permitted angles. Varying the acceptance window has the effect of changing the ratio of the distribution of molecules related by a 0° rotation to those related by 60°, 90°, 120°, and 180° rotations. This is because,

TABLE I. Simulated vs. Observed Space Group Frequencies for Monomeric Proteins

Space group	Sim. %	PDB %	PDB #*	Space group	Sim. %	PDB %	PDB #
$P2_12_12_1$	31.8	36.1	88	$I4_1$	0.3	0.0	0
$P2_1$	7.6	11.1	27	$P6_2$, $P6_4$	0.9	0.0	0
$P4_32_12$, $P4_12_12$†	7.9	7.8	19	$P2_13$	0.0	0.0	0
$P3_12_1$, $P3_22_1$	2.9	7.8	19	$I2_13$	0.0	0.0	0
$C2$	8.7	6.1	15	$P4_122$, $P4_322$	0.0	0.0	0
$P6_122$, $P6_522$	9.5	5.4	13	$P432$	0.0	0.0	0
$C222_1$	3.0	3.7	9	$P4_132$, $P4_332$	0.0	0.0	0
$P2_12_12$	5.3	3.7	9	$P4_232$	0.0	0.0	0
$P1$	4.3	2.9	7	$I4_132$	0.0	0.0	0
$P6_1$, $P6_5$	2.6	2.0	6	$C222$	0.0	0.0	0
$I4$	0.1	2.0	5	$P222_1$	0.0	0.0	0
$R3$	0.3	2.0	5	$P2$	3.6	0.0	0
$I222$	0.3	1.5	4	$F432$	0.0	0.0	0
$P4_22_12$	1.3	1.5	4	$P3$	0.1	0.0	0
$P4_1$, $P4_3$	1.0	1.0	4	$P23$	0.0	0.0	0
$P3_1$, $P3_2$	0.3	1.0	3	$P4$	0.0	0.0	0
$P6$	0.0	0.5	1	$P312$	0.0	0.0	0
$P6_3$	0.1	0.5	1	$P422$	0.1	0.0	0
$I422$	0.3	0.5	1	$P4_222$	0.1	0.0	0
$P3_112$, $P3_212$	0.1	0.5	1	$P222$	0.0	0.0	0
$R32$	1.6	0.5	1	$P622$	0.0	0.0	0
$P4_2$	0.0	0.5	1	$F23$	0.0	0.0	0
$P6_222$, $P6_422$	0.6	0.0	1	$P42_12$	0.2	0.0	0
$I2_12_12_1$	1.0	0.0	0	$F4_132$	0.0	0.0	0
$I4_122$, $I4_322$	0.5	0.0	0	$P6_322$	1.0	0.0	0
$P321$	0.2	0.0	0	$I23$	0.0	0.0	0
$P4_132$	0.0	0.0	0	$I432$	0.0	0.0	0
$F222$	0.0	0.0	0				

*PDB # refers to the number of occurrences in a set of 244 dissimilar monomeric proteins in the Brookhaven Data Bank.

†For enantiomorphic pairs of space groups, the frequencies refer to the sum of the contributions from the two symmetries.

starting from random rotations, it is very improbable to obtain two molecules whose orientations differ by only a few degrees. However, this probability increases rapidly as a function of the acceptance window. The final effect of broadening the window is to increase the ratio of the space group $P1$ (for which all the molecular orientations are the same) to that of all others. In fact, we chose the value of 10° to give us a frequency for $P1$ close to that found in the PDB. However, this parameter does not significantly affect the relative frequency of other space groups.

We therefore conclude that our final distribution represents a general geometric result of the relative probability of generating nuclei belonging to the 65 biologic space groups (although we cannot conclude anything about the cubic groups). We find that by using our crude square well energy model, the average calculated energies of nuclei do not vary consistently as a function of space group; no correlation is observed between the space group and the packing energy. Therefore, one may interpret the probability of a space group as a product of two probabilities determined purely by the geometry of the contacts: the entropic accessibility of a set of contacts that are consistent with a particular space group, and the likelihood that when these contacts are repeated, they lead (without molecular collisions) to a crystalline nucleus. The observation that space group frequencies are a function of the former probability is related to our previous analysis of space group frequencies based on their available degrees of freedom.[6] In other words, certain symmetries are more prevalent simply because they impose fewer constraints on the ways that the molecules can be arranged in a crystal.

CONCLUSIONS

It is important to emphasize that in the above analysis we are comparing the distribution of simulated nuclei to that of observed macroscopic crystals and not observed nuclei. The fact that these two are very nearly the same implies that the statistically averaged probability that a critical nucleus grows into a macroscopic crystal does not depend on its underlying space group.

If we accept the physical validity of the assumptions adopted in the simulations, then we may conclude that the relative probabilities of occurrence

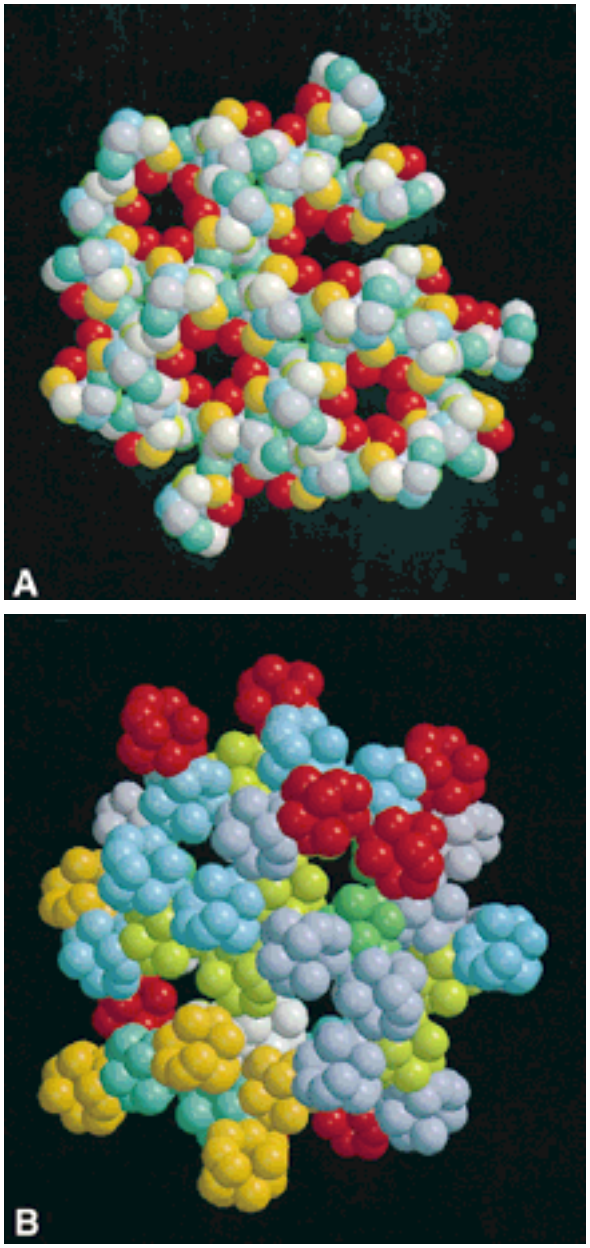

Fig. 2. Space filling images of two of the simulated crystal nuclei. Each cluster contains 200 molecules. **A:** Every molecule is identically colored, with each of the ten particles that compose it in a different tone. A view down the sixfold axis of the nucleus is shown. **B:** All the atoms of a molecule are colored the same way, and the molecules are displayed in one of ten colors. A view down the fourfold axis is shown.

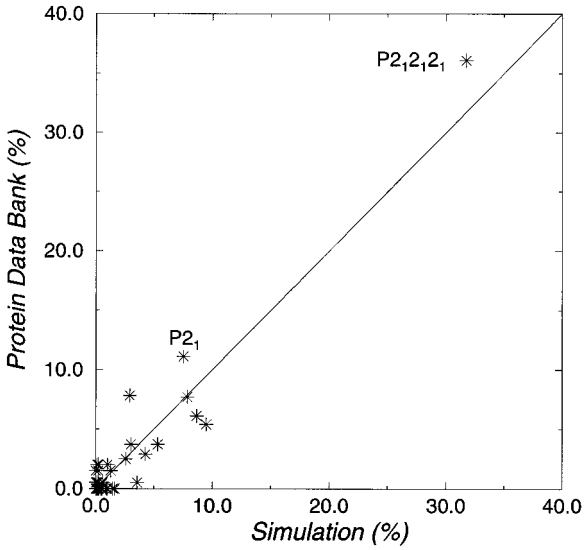

Fig. 3. The percentage occurrence of the 65 space groups obtained from our simulation versus that obtained from the Protein Data Bank. The list of PDB structures is obtained from ref. 6 and contains 244 entries. It includes only monomeric proteins without noncrystallographic symmetry determined to at least 2.5 Å resolution and excludes multiple observations of closely related proteins in the same crystal form. The straight line represents the case in which the frequencies are equal. The correlation coefficient is 0.966.

of space group symmetries in protein crystal nuclei are not influenced by complex physical interactions or the need to pack proteins in a specific manner. Instead, the strong preference for certain symmetries emerges from simple geometric, or entropic, considerations. The entropic accessibility of the required combinations of intermolecular contacts is vastly different for the various space groups and is responsible for the observed and simulated distributions.

This conclusion should not be confused with the fact that solvent and buffer conditions are critically important for the growth of a protein crystal in a laboratory. By varying the solvent mixture containing the pure protein, it is possible to vary the strengths and types of intermolecular protein contacts. Only under certain, usually rare, conditions will these contacts lead to the formation of a protein crystal. Our conclusions do not suggest that protein-protein contacts are unimportant for crystal growth, for they are crucial, but rather that these contacts do not favor the formation of one space group over another on average. In other words, in a large ensemble of crystals the energy of intermolecular contacts does not vary as a function of space group.

One of the implications of this simulation is that the nucleation of a crystalline aggregate is primarily limited by the conformational constraints that are imposed by the space group symmetry. The discussion was limited to the 65 biologic space groups (those lacking mirror and inversion symmetries) because those are the ones accessible to natural proteins. It is our contention, however, that certain nonbiologic space groups ($P\bar{1}$ in particular) impose fewer constraints on the molecular conformations and should therefore be more likely to form crystalline nuclei.[6] The supporting conclusions drawn from our simulation, that entropic and not energetic con-

siderations determine the likelihood of nucleation, should motivate further experiments on crystallizing macromolecules from racemic mixtures.[12]

Our analysis of the crystalline symmetry of simulated nuclei represents a first computational step toward a more complete understanding of the nucleation and growth of protein crystals. We intend to extend the approach to investigate the distribution of nuclei for symmetric oligomers and racemic mixtures of proteins as well as to probe disorder in nuclei and larger crystalline aggregates.

ACKNOWLEDGMENTS

The authors thank David Eisenberg for useful comments on the original manuscript. M. Pellegrini is supported by the Sloan Foundation, Department of Energy Postdoctoral Fellowship. S. Wukovitz is supported by the Program in Mathematics and Molecular Biology graduate fellowship.

REFERENCES

1. McPherson, A., Malkin, A.J., Kuznetsov, Y.G. The science of macromolecular crystallization. Structure 3:759–768, 1995.
2. Patro, S.Y., Przybycien, T.M. Simulations of reversible protein aggregate and crystal structure. Biophy. J. 70:2888–2902, 1996.
3. Brock, C.P., Dunitz, J.D. Towards a grammar of crystal packing. Chem. Materials 6:1118–1127, 1994.
4. Kitaigorodskii, A.I. "Organic Chemical Crystallography." New York: Consultants Bureau, 1961. (Originally published in Russian by Press of the Academy of Sciences of the USSR, Moscow, 1955.)
5. Wilson, A.J.C. Kitajgorodskij's categories. Acta Crystallogr. A49:210–212, 1993.
6. Wukovitz, S.W., Yeates, T.O. Why protein crystals favor some space-groups over others. Nature Struct. Biol. 2:1062–1067, 1995.
7. Padmaja, N., Ramakurar, S., Viswamitra, M.A. Space-group frequencies of proteins and of organic compounds with more than one formula unit in the asymmetric unit. Acta Crystallogr. A46:725–730, 1990.
8. Sali, A., Shakhnovich, E., Karplus, M. How does a protein fold? Nature 369:248–251, 1994.
9. Malkin, A.J., McPherson, A. Light-scattering investigations of nucleation processes and kinetics of crystallization in macromolecular systems. Acta Crystallogr. D50:385–395, 1994.
10. Metropolis, N., Rosenbluth, M.N., Rosenbluth, A., Teller, H., Teller, E. Equations of state calculations by fast computing machines. J. Chem. Phys. 21:1087, 1953.
11. Crow, E.L., Davis, F.A., Maxfield, M.W. "Statistics Manual." New York: Dover Publications, 1960.
12. Berg, J.M., Zawadzke, L.E. Racemic macromolecules for use in X-ray crystallography. Curr Opin. Biotechnol. 5:343–345, 1994.