© 1995 Nature Publishing Group http://www.nature.com/nsmb

Why protein crystals favour some space-groups over others

Stephanie W. Wukovitz[1] and Todd O. Yeates[2]

One of the most puzzling observations in protein crystallography is that the various space-group symmetries occur with striking non-uniformity. Molecular close-packing has been invoked to explain similar observations for crystals of small organic compounds, but does not appear to be the dominant factor for proteins. Instead, we find that the observed frequencies for both two- and three-dimensional crystals can be explained by an entropic model. Under a requirement for connectivity, the favoured space groups are simply less restrictive than others in that they allow the molecules more rigid-body degrees of freedom and can therefore be realized in a greater number of ways. This result underscores the importance of the nucleation event in crystallization and leads to specific ideas for crystallizing water-soluble and membrane proteins.

[1]Department of Biomathematics University of California, Los Angeles Box 951766 Los Angeles, California 90095-1766, USA
[2]Department of Chemistry and Biochemistry, University of California, Los Angeles, Box 951569 Los Angeles, California 90095-1569, USA

Correspondence should be addressed to T.O.Y.

In order to determine a molecular structure by X-ray crystallography, a well-ordered crystal must first be grown. Although three-dimensional structures of proteins from more than 400 distinct crystal forms have been determined over the past 40 years, protein crystallization is still a major obstacle to X-ray diffraction work and persists as perhaps the most ill-defined protocol in molecular biology. Although there has been considerable progress on practical and theoretical aspects of the problem[1–5], fundamental questions remain. Perhaps the most enigmatic observation is the non-uniformity with which the different space-group symmetries occur. In particular, although 65 space groups are available to biological (chiral) macromolecules, a single space-group, $P2_12_12_1$, occurs roughly one-third of the time for monomeric proteins, more than three times as frequently as any other symmetry. About half of the other symmetry groups have never been observed for crystals of monomeric proteins.

The question of why certain symmetries are adopted by individual molecules[6] or collections of molecules is a fundamental one in chemistry. With regard to crystalline symmetries[7], the non-uniform occurrence of the 230 space groups for crystals of small organic compounds was first noted by Nowacki[8] in the 1940's and has since been examined carefully by others[9–12]. In the 1950's, Kitaigorodskii[13] developed an explanation for the preference of certain space groups based on elaborate considerations of molecular close-packing. This argument, revised by Wilson[14], has been widely accepted. The existence of a satisfactory explanation of this phenomenon in the case of organic molecules and some similarities between the observed frequencies for small organic and protein crystals has resulted in the protein problem receiving less attention than it deserves. However, a comparison of the observed frequencies for asymmetric organic molecules in the 65 'chiral' space groups (those that do not contain operations of the second kind—inversions or reflections) with the observed frequencies for monomeric proteins reveals that—apart from the preference for $P2_12_12_1$ and $P2_1$—there are marked differences between the two distributions (Fig. 1). In fact, Kitaigorodskii's theory forbids the tetragonal space groups, several of which are common for proteins. Since these observed preferences for certain space groups are significantly different, the same explanation cannot apply to both patterns. Furthermore, the close-packing argument does not seem to apply to proteins; while crystals of organic compounds are usually exceptionally well-packed, leaving very little free space, protein crystals contain on average nearly 50% solvent by volume. Also, the number of neighbouring molecules with which a protein makes direct van der Waals contact—the 'coordination number'—in a crystal is only about 7.5 on average (based on a 4.5 Å distance cut-off between non-hydrogen atoms). The coordination number for organic molecules usually ranges between ten and fourteen[13]; twelve contacts are achieved by the cubic close-packing of spheres. Finally, an examination of average protein crystal packing densities for each space group (data not shown) shows that proteins in the most common space group, $P2_12_12_1$, do not achieve an average packing density that is any higher than the average for proteins in other space groups. In this analysis we provide an explanation for the tendency for proteins to crystallize in a small number of preferred space groups. Macromolecules that form symmetric oligomers often crystallize in space groups which can partially or completely accommodate their

© 1995 Nature Publishing Group http://www.nature.com/nsmb

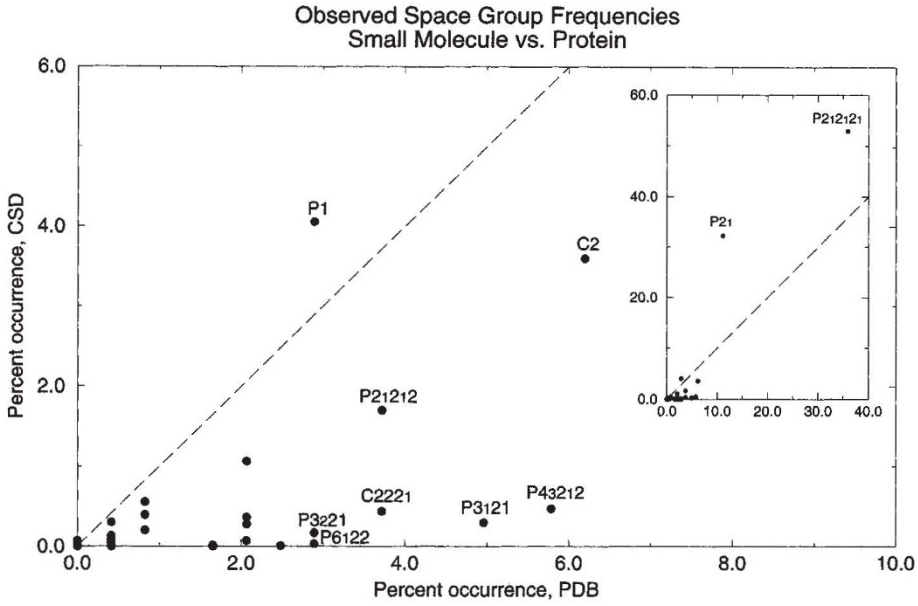

Fig. 1 Number of occurrences of the 65 chiral space-groups for asymmetric organic molecules (CSD) and monomeric proteins (PDB), expressed as percentages. A dashed line is drawn where the frequencies for the two cases would be equal. Several common symmetry groups are labelled. The two most common groups, $P2_12_12_1$ and $P2_1$ are shown in the inset. The space-group frequencies for proteins are based on monomeric proteins having resolutions better than 2.5 Å in release #70 of the Brookhaven Protein Databank[26]. When multiple related structures with the same space group and unit cell were reported, only one instance was retained, giving a total of 245 unique observations. The 245 proteins are not all unrelated, however, since some identical or similar proteins are reported in multiple crystal forms. The space-group frequencies for crystals of organic compounds are based on a survey of the Cambridge Structural Database, excluding entries where a symmetric molecule falls on a symmetry element.

natural internal symmetries as part of the crystal symmetry. We therefore restrict our attention here to proteins that are monomeric, and consequently lack internal symmetry, in order to avoid such complications.

Rigid-body degrees of freedom

The absence of obvious energetic differences between protein crystals in different space groups suggests that there are innate properties of the symmetries themselves, independent of molecular properties, which make some space groups more probable than others; the conditions imposed by some crystallographic symmetries could simply be more restrictive than those imposed by others. In support of this, we note that for a particular space group, only a certain number of rigid-body degrees of freedom are available for assembling the first few molecules before the internal structure of the crystal is completely defined; this number of rigid-body degrees of freedom depends entirely on the space-group symmetry. Space groups for which this number is high should be the most common ones for purely statistical, rather than energetic, reasons.

The total number of rigid-body degrees of freedom available to a set of molecules that must satisfy a particular space-group symmetry is determined by three quantities: i), the number of meaningful rigid-body degrees of freedom for orienting and positioning the first molecule in space (relative to any rotational symmetry elements); ii) the number of independent unit-cell parameters—each such parameter provides one translational degree of freedom for subsequent molecules related by translation; and iii), the number of distinct intermolecular contacts required to produce a connected network of molecules in the crystal—each such contact provides a single constraint on the position of one molecule relative to another.

A one-dimensional example (with no rotational symmetry) of the number of rigid-body degrees of freedom is illustrated in Fig.2. For placing the first molecule, there is one rotational degree of freedom and no translational degrees of freedom; the position along the line is arbitrary due to the absence of any rotational symmetry elements. There is one free unit-cell parameter (the repeat distance), and only one contact type is required for connectivity. As illustrated, only one degree of freedom is available for creating this hypothetical one-dimensional crystal; once the orientation of the first molecule has been chosen, the orientations and positions of the remaining molecules are fixed.

The general rule for calculating the total number of rigid-body degrees of freedom, D, for a particular space group is

$$D=S+L-C$$

where S is the number of meaningful degrees of freedom for orienting and positioning a single molecule in the unit cell, L is the number of independent parameters for describing the unit cell, and C (a number which we report here for each space group) is the minimum number of unique contacts required to make the set of symmetry-related molecules into a connected three-dimensional network. D may also be interpreted as the dimensionality of the rigid-body space that gives rise to a connected crystal, but should not be confused with the 'degrees of freedom' given by Brock and Dunitz[12], which is essentially S+L. The terms S, L, C, and D are all positive integers that depend only on the space-group symmetry and not on any properties of the molecule in question. It is worth emphasizing here that S, L, and C are not adjustable parameters in this model, but are instead fixed terms that determine the value of a physically meaningful quantity, D.

S contains the rotational and relevant translational degrees of freedom for the first molecule in the crystal. In all space groups, three degrees of freedom are available for orienting a single molecule. In space-groups with dihedral or higher symmetry (such as $P2_12_12$ and I23), there are also three translational degrees of freedom, giving S=6. In space groups with only one rotational symmetry axis (such as $P2_1$ and R3), the

© 1995 Nature Publishing Group http://www.nature.com/nsmb

position of the molecule along the direction of the axis is arbitrary, so S=5. In space group P1, the position of the first molecule is arbitrary in all three directions, so S=3+0=3. The value of S is determined entirely by the underlying point-group symmetry of the crystal.

L, the number of free parameters for choosing the unit cell, is the number of lengths and angles of the unit cell that are not predetermined by specifying the space-group; this number depends on the underlying crystal lattice. For example, for space-group P1, which falls on a triclinic lattice, we may choose all three axial lengths and all three angles of the unit cell, so L=6. On the other hand, L=1 for cubic space groups, since all three angles must equal 90° and all three cell lengths must be equal in those groups.

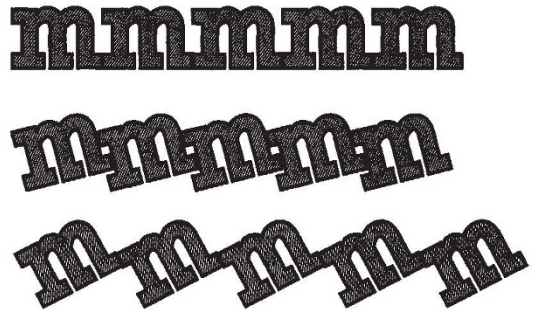

Fig. 2 A hypothetical one-dimensional example of the total number of rigid-body degrees of freedom available to a collection of molecules (represented by the letter 'm'), subject to crystallographic symmetry. In the example shown, there is only one degree of freedom since the internal structure of the crystal can be specified by the choice of a single parameter, such as the orientation of one molecule. The freedom for positioning additional molecules is negated by the requirement that the molecules be in contact. Three different configurations are shown.

The minimum contact number

C, the minimum number of unique contact types required to establish connectivity in a particular space-group symmetry, is the smallest number of non-equivalent contacts that make it possible to trace a path of van der Waals contacts from any one molecule to any other molecule. This condition must be satisfied in order for the crystal to be a solid. Two contacts are considered to be equivalent if both corresponding pairs of molecules are related by the same symmetry operation. Fig. 3 illustrates the meaning of C in two different plane group symmetries, p2 and p4.

The value C may be obtained for each three-dimensional space group. The problem of finding the minimum contact number is equivalent to the problem of identifying the minimal set of symmetry elements that is sufficient to generate a particular space group. C was obtained for each three-dimensional space group G, as follows: a set T of elements of the group G generates G if all elements of G can be expressed as products of the elements of T and of their inverses. There is a one-to-one correspondence between the sets of contacts which connect a crystal and the sets of generators for G. Therefore, C can be determined by finding a minimal set of generators for G. Beginning with a large but finite subset of G, elements that were powers of other elements in the set were first removed, then products of reasonable length of two remaining elements, then of three, and so on. The size of the remaining set is an upper bound for C. The procedure was repeated, beginning with finite factor groups of G, groups whose elements consist of equivalence classes of elements of G, which established a lower bound for C. These

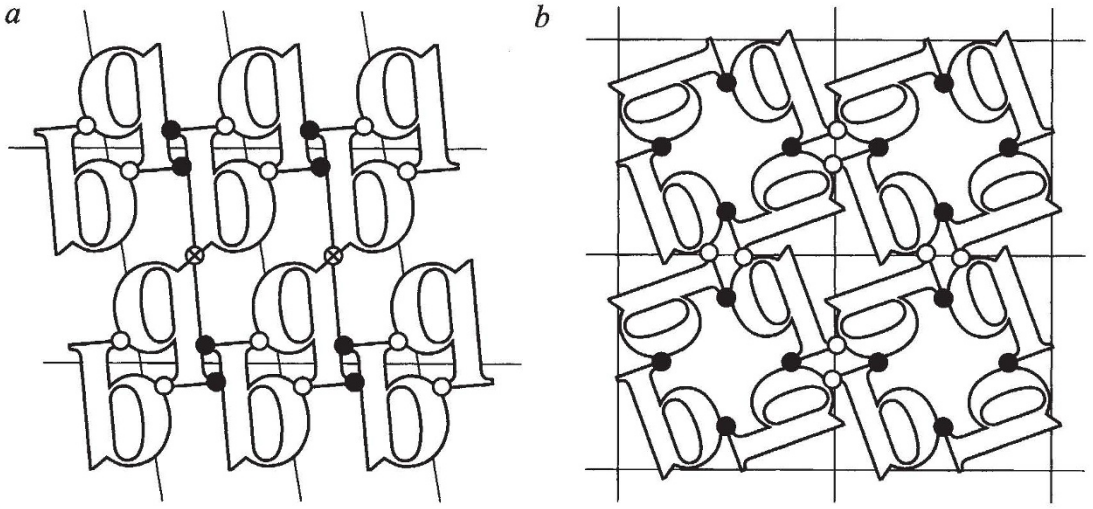

Fig. 3 Two-dimensional illustrations of the minimum number of contact types required for connectivity. Individual molecules are represented by the letter 'q'. The underlying lattices for the two figures are shown, but the symmetry elements are omitted for clarity. *a*, In plane group p2, at least three distinct contact types (denoted by open, filled and crossed circles) are always required for connectivity. A smaller number is never sufficient in this group, regardless of molecular shape, as long as the molecule lacks internal symmetry. *b*, In plane group p4, only two unique contact types are required (open and filled circles). The minimum contact number, C, is a factor in determining the number of rigid-body degrees of freedom, or the dimensionality, D, of the rigid-body space that gives rise to a connected network of molecules in a given crystallographic symmetry.

© 1995 Nature Publishing Group http://www.nature.com/nsmb

Table 1 The number of rigid-body degrees of freedom, D, and the observed occurrence of monomeric protein crystals among the 65 crystallographic space groups

Symmetry group	S	L	C	D[1]	Freq	%[2]
$P2_12_12_1$	6	3	2	7	88	36.1
$P2_1$	5	4	3	6	27	11.1
C2	5	4	3	6	15	6.1
$P4_32_12$	6	2	2	6	14	5.7
$P3_121$	6	2	2	6	12	4.9
$C222_1$, $P2_12_12$	6	3	3	6	9	3.7
$P3_221$,$P6_122$	6	2	2	6	7	2.9
P1	3	6	3	6	7	2.9
$P6_522$	6	2	2	6	6	2.5
$P4_12_12$	6	2	2	6	5	—
I222	6	3	3	6	4	—
$I2_12_12_1$	6	3	3	6	0	—
I4,$P6_1$,R3	5	2	2	5	5	—
$P4_22_12$	6	2	3	5	4	—
$P3_1$,$P4_1$,$P4_3$	5	2	2	5	2	—
$P3_2$,P6,$P6_3$,$P6_5$	5	2	2	5	1	—
I422,$P3_112$ $P6_422$,R32	6	2	3	5	1	
$P42_12$,$P4_122$,$P4_322$, $I4_122$,P321,$P3_212$, P622,$P6_222$,$P6_322$	6	2	3	5	0	
$I4_1$,$P6_2$,$P6_4$	5	2	2	5	0	
$P2_13$,I23,$I2_13$, $P4_132$,$P4_232$,$P4_332$ P432,I432,$I4_132$ F432,$F4_132$	6	1	2	5	0	
$P222_1$,C222,F222	6	3	4	5	0	
P2	5	4	4	5	0	
$P4_2$	5	2	3	4	1	
P4,P3	5	2	3	4	0	
P23,F23	6	1	3	4	0	
P312,P422,$P4_222$	6	2	4	4	0	
P222	6	3	5	4	0	

[1]Sorted according to D, then frequency. Notice the tendency of a high occurrence for space groups having large D. D=S+L−C (see text for definitions)
[2]Percentages are reported only for space groups that were observed more than five times.
The protein crystal database is described in the legend to Fig. 1.

bounds coincided with each other for all groups. We have computed the minimal number of generators for each of the 65 chiral space-groups. To our knowledge, these values have not been previously reported.

The value of C ranges from a maximum of 5 (for P222) to a minimum of 2. Contrary to intuition, connectivity in three dimensions can be achieved with just two contact types in several space groups (Table 1), including $P2_12_12_1$. Space groups with screw axes of symmetry often have lower values of C than corresponding groups with pure rotation axes (for example, $P2_1$ versus P2), since a contact between molecules related by a screw axis gives connectivity in one direction (that is, a chain of molecules).

Agreement with observed data

The values we found for D, the total number of degrees of freedom, are correlated with the frequencies of protein crystal space groups observed in our database (Table 1). D ranges from 4-7, and effectively divides the 65 chiral space-groups into four categories (D=4,5,6 or 7). The most notable observation is that $P2_12_12_1$ is the only space group with seven rigid-body degrees of freedom, and is therefore singled out by our analysis as the space-group that least restricts the possible orientations and positions of the molecules in the crystal. This appears to explain why $P2_12_12_1$ occurs so much more frequently than any other group. Of the 13 space-groups with D=6, each occurs 9.4 times on average in our survey, with a standard deviation of 6.4. The 42 groups with D=5 occur 0.8 times on average (max=5) and the nine groups with D=4 occur only 0.1 times on average (max=1). Histograms illustrate the number of space groups with a given number of occurrences for each value of D (Fig. 4). The frequency distributions for the four categories are nearly non-overlapping so that a division of the 65 space groups according to D produces nearly distinct frequency classes, which we designate 'optimal' (D=7), 'favourable' (D=6), 'unfavourable' (D=5), and 'forbidden' (D=4). Space group $I2_12_12_1$ provides a single exception: it has D=6 but crystals having this space-group were not observed in our survey. The only protein in our database that was found in any 'forbidden' space-group is proteinase A (lsgc) in $P4_2$.

We estimate that the probability of achieving equally good agreement between predicted and observed frequencies by random partitioning of the space groups is less than 10^{-6}. The 65 space groups for chiral molecules were randomly partitioned into four categories two million times and the goodness-of-fit between each partitioning and the observed space-group frequencies was evaluated by two criteria: the uniformity of the

© 1995 Nature Publishing Group http://www.nature.com/nsmb

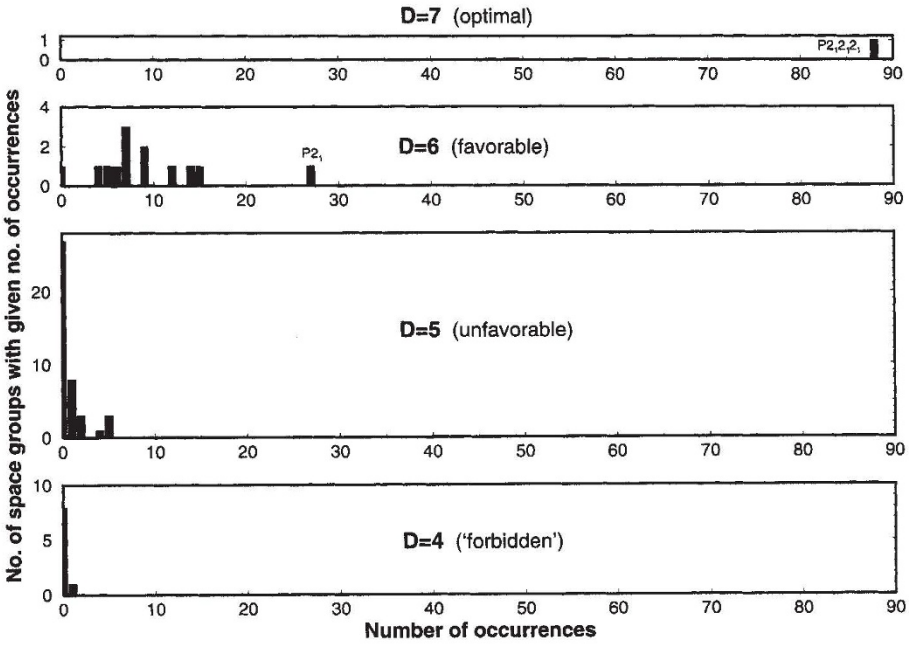

Fig. 4 Histograms for each degree of freedom category (D) showing the number of space-groups with given numbers of occurrences. The ordinate axes of the plots are not on identical scales. The four degrees of freedom categories show distinct distributions for space-group occurrence, and are named accordingly.

frequencies in each category, and the degree of frequency overlap between symmetry groups from different categories. The partitioning based on D was better than all random partitionings, according to either measure of fit.

The explanation presented here for space-group populations is statistical in nature and assumes nothing about molecular shape or interactions. Its predictive success suggests that the preference for certain space groups may be a statistical rather than thermodynamic effect, arising at the nucleation stage of crystal growth, during which time the molecular rigid-body degrees of freedom are consumed. This model also supports the argument that protein crystallization may be limited primarily by the ability to form a nucleation site consistent with crystallographic symmetry and the solid state. Of course, thermodynamics also plays a role, and a given protein often gives rise to multiple crystal forms, depending upon experimental conditions. The importance of nucleation events in dictating which of several crystal forms is realized has not been studied systematically.

Our analysis does not explain the relatively wide range of observed frequencies for the 'favourable' space groups (those with D=6), such as the preference for $P2_1$ or the absence of $I2_12_12_1$. A complete explanation would presumably require analysis of other factors, including energetics, packing efficiency and molecular shape. In this regard, it should be pointed out that there may be combinations of space group and molecular shape for which it is impossible to make a set of connections that is minimal in the sense described here, without having collisions between molecules. $I2_12_12_1$ is noteworthy, since the multiple, non-intersecting, pure two-fold symmetry axes are likely to obstruct formation of contacts between molecules related by the screw axes, as would be required in arrangements with the minimum number of contact types.

It is also worth noting that the actual number of unique contacts formed in a protein crystal is usually greater than the minimum number required for connectivity, C. The average number of unique contacts (based on a 4.5 Å cut-off for the distance between atom centres) is approximately 4.5, which is about two more than the average value of C. This observation is not at odds with the hypothesis presented here; it simply suggests that some of the rigid-body degrees of freedom available to the molecules are expended in order to make additional contacts.

Our discussion so far has been neutral toward the formation of contacts between molecules in the crystal, stating only that a certain number are required. Crystallizing integral membrane proteins presents a special problem in this respect, due to the presence of a relatively fluid lipid or detergent layer around much of the protein[15]. Of the few membrane proteins whose structures have been determined by X-ray diffraction, all have fewer unique contacts than the average for water-soluble proteins. There are two unique contact types in both crystal forms of porin from *Rhodobacter capsulatus*[16] (PDB codes 2por and 3por), two in the bacterial photosynthetic reaction centre from *Rhodobacter sphaeroides*[17,18] (2rcr, 4rcr), and three in the reaction centre from *Rhodopseudomonas viridis*[19] (1prc). The structure of porin from *Rb. capsulatus* is especially interesting. One of the two unique contact types

Table 2 Rigid-body degrees of freedom, D, for the two-dimensional layer-groups

symmetry group	S[1]	L	C	D[2]
$p22_12_1$	4	2	2	4
p2	3	3	3	3
$p12_1$, c12	3	2	2	3
$p222_1$, c222	4	2	3	3
$p422_1$, p321, p622	4	1	2	3
p1	1	3	2	2
p3, p4, p6	3	1	2	2
p12	3	2	3	2
p222	4	2	4	2
p422, p312	4	1	3	2

[1]We assume here that the molecules are membrane proteins confined to a lipid bilayer, so that they have only one rotational degree of freedom (about the normal to the bilayer).
[2]D=S+L−C (see text for definitions).

© 1995 Nature Publishing Group http://www.nature.com/nsmb

between protein chains is between pairs of monomers of the natural porin trimer, which sits on a crystallographic three-fold axis of symmetry in the R3 space-group. Although C=2 for R3, the porin crystal takes advantage of the built-in trimer contact so that only one fortuitous contact between trimers is required to generate connectivity. This suggests a possible advantage for crystallizing membrane proteins that are symmetrical either by nature or by design. Space-groups that can be formed by a single contact type between symmetric oligomers include R3, R32, I4, I422, and the cubic groups.

An analysis of the 165 'non-biological' space-groups, those that contain operations of the second kind and consequently require either achiral molecules or racemic mixtures, is not yet complete. It is clear, however, that the maximum number of degrees of freedom, D=8, is achieved only in space-group $P\bar{1}$. Therefore, we predict that $P\bar{1}$ will be the most frequently observed space-group for racemic protein mixtures. In contrast, small organic compounds exhibit a preference for the non-biological space-group $P2_1/c$, which occurs about twice as often as $P\bar{1}$. To date, only one protein structure has been determined from a racemic crystal, a process which requires that the non-biological enantiomer be chemically synthesized entirely from D-amino acids. Zawadzke and Berg[20] determined the structure of rubredoxin, which crystallized in space-group $P\bar{1}$ from a synthetic racemic mixture, emphasizing that the phase problem is simplified in centrosymmetric groups. Perhaps more importantly, since we expect $P\bar{1}$ to be significantly more probable than other symmetries, we predict that racemic protein mixtures will crystallize more readily than samples consisting only of the biological enantiomer. Unfortunately, the cost of chemical synthesis makes this route efficacious only for small or extraordinarily important proteins. Furthermore, the accessibility to space-group $P\bar{1}$ which is gained by racemic mixtures is not likely to be an advantage for membrane proteins, since a relatively large number of unique contact types is required for connectivity in $P\bar{1}$ (C=4). Nonetheless, the prediction that $P\bar{1}$ will predominate for racemic mixtures of aqueous proteins is an ultimately testable prediction of our theory.

Finally, we consider two-dimensional protein crystals; so far, electron microscopy and diffraction of such specimens has yielded three-dimensional atomic structures of two integral membrane proteins[21,22]. As with the space groups, we can calculate the number of degrees of freedom for the 17 chiral two-dimensional layer groups[23]. For the layer-groups, D ranges from 2 to 4 (Table 2). Symmetry $p22_12_1$ is the only group with D=4 and thus should be favoured. The nature of the crystallization experiment may restrict which of the 17 layer groups can be formed, since twelve of the groups (all but p1, p2, p3, p4, and p6) require the protein molecules to be oriented half-up and half-down in the bilayer. Nonetheless, we find that $p22_12_1$ is indeed observed significantly more frequently than any other group[24], as has been pointed out by others[25].

In summary, a statistical effect arising from differences in the number of rigid-body degrees of freedom appears to govern space (and layer) group frequencies for two and three-dimensional protein crystals. While small organic molecules prefer to crystallize in space groups in which it is easiest to fill space, proteins crystallize primarily in space-groups in which it is easiest to achieve connectivity. This realization may be useful in developing new crystallization strategies, including the use of racemic samples for water-soluble proteins and artificial symmetrization for transmembrane proteins.

Acknowledgements
The authors thank David Eisenberg and Scott LeGrand for helpful discussions. This work was supported by the National Institutes of Health and the National Science Foundation (T.O.Y.) and a training grant from the National Institutes of Health (S.W.W.). This paper is dedicated to the memory of Dr. Jairo Arevalo.

1. Diffraction Methods for Biological Macromolecules (eds Wyckoff, H.W., Hirs, C.H.W. & Timasheff S.N.) *Meths. Enzymol.* **114** (Academic Press, Orlando, 1985).
2. Ducruix, A. & Giege, R. *Crystallization of Nucleic Acids and Proteins: A Practical Approach* (IRL Press, Oxford, 1992).
3. Carter, C.W. Jr. & Carter, C.W. Protein crystallization using incomplete factorial experiments. *J. biol. Chem.* **254**, 12219–12223 (1979).
4. Zulauf, M. & D'Arcy, A. Light scattering of proteins as a criterion for crystallization. *J. cryst. Growth* **122**, 102–106 (1992).
5. Velikov, P. G., Ataka, M. & Katsura, T. Laser Michelson interferometry investigation of protein crystal growth. *J. cryst. Growth* **130**, 317–320 (1993).
6. Mezey, P.G. A global approach to molecular symmetry: theorems on symmetry relations between ground and excited-state configurations. *J. Am. Chem. Soc.* **112**, 3791–3802 (1990).
7. International Tables for Crystallography. Vol. A. 2nd ed.(ed. Hahn, T.) Space Group Symmetry. (Kluwer Academic Publishers, Dordrecht, 1989).
8. Nowacki, W. Symmetrie und physikalisch-chemische eigenschaften krystallisierter verbindungen. II. Die allgemeiner bauprinzipien organischer verbindungen. *Helv. chim. Acta* **26**, 459–462 (1943).
9. Mighell, A.D. & Himes, V.L. Space-group frequencies for organic compounds. *Acta crystallogr.* A **39**, 737–740 (1983).
10. Donohue, J. Revised space–group frequencies for organic compounds. *Acta crystallogr.* A **41**, 203–204 (1985).
11. 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.* A**46**, 725–730 (1990).
12. Brock, C.P. & Dunitz, J.D. Towards a grammar of crystal packing. *Chem. Mat.* **6**, 1118–1127 (1994).
13. Kitaigorodskii, A.I. Organic Chemical Crystallography (Consultants Bureau: New York, 1961) (Originally published in Russian by Press of the Academy of Sciences of the USSR, Moscow, 1955).
14. Wilson, A.J.C. Kitajgorodskij's categories. *Acta crystallogr.* A**49**, 210–212 (1993).
15. Michel, H. Crystallization of membrane proteins. *Trends biochem. Sci.* **8**, 56–59 (1983).
16. Weiss, M.S., Abele, U., Weckesser, J., Welte, W., Schiltz, E. & Schultz, G.E. Molecular architecture and electrostatic properties of a bacterial porin. *Science* **254**, 1627–1630 (1991).
17. Allen, J.P., Feher, G., Yeates, T.O., Komiya, H. & Rees, D.C. Structure of the reaction center from Rhodobacter sphaeroides R-26: The protein subunits. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6162–6166 (1987).
18. Chang, C.-H., El-Kabbani, O., Tiede, D., Norris, J. & Schiffer, M. Structure of the membrane-bound protein photosynthetic reaction center from *Rhodobacter sphaeroides*. *Biochemistry* **30**, 5352–5360 (1991).
19. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. X-ray structure analysis of a membrane protein complex. *J. molec. Biol.* **180**, 385–398 (1984).
20. Zawadzke, L.E. & Berg, J.M. The structure of a centrosymmetric protein crystal. *Proteins* **16**, 301–305 (1993).
21. Henderson, R. *et al.* Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. molec. Biol.* **213**, 899–929 (1990).
22. Kühlbrandt, W., Wang, D.N. & Fujiyoshi, Y. Atomic model of plant light-harvesting complex by electron crystallography. *Nature* **367**, 614–621 (1994).
23. Holser, W.T. Point groups and plane groups in a two-sided plane and their subgroups. *Z. Kristallogr.* **110**, 266–281 (1958).
24. Kühlbrandt, W. Two-dimensional crystallization of membrane proteins. *Q. Rev. Biophys.* **25**, 1–49 (1992).
25. Dekker, J.P., Betts, S.D., Yokum, C.F. & Boekema, E.J. Characterization by electron microscopy of isolated particles and two-dimensional crystals of the CP47-D1-D2-cytochrome b-559 complex of photosystem II. *Biochem.* **29**, 3220–3225 (1990).
26. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F. Jr.& Brice, M.D. The protein data bank: a computer-based archival file for macromolecular structures. *J. molec. Biol.* **112**, 535–542 (1977).