Why biomolecules prefer only a few crystal structures

Yu. E. Kitaev,* A. G. Panfilov,* V. P. Smirnov,† and P. Tronc
Laboratoire d'Optique Physique, Ecole Superieure de Physique et Chimie Industrielles, 10 rue Vauquelin, 75005 Paris, France
(Received 26 July 2002; published 22 January 2003)

We have shown that, in determining the biomolecule-crystal symmetry, the occupation of low-site-symmetry Wyckoff positions is crucial, which contrasts with the overwhelming majority of nonmolecular, inorganic crystals where atoms mainly reside in high-symmetry Wyckoff positions. We consider the general relation between the symmetry of an isolated molecule and the possible symmetries of biomolecular crystals it can generate. We reveal that the improper symmetry operations (inversion and mirror symmetries) are prohibited in the chirally pure biomolecular crystals. Next, we show that the low (C_1) symmetry of large biological molecules substantially decreases the space in a crystal where the molecules can reside. The space ''forbidden'' for molecule centers is found to be in the R vicinity of the higher-symmetry Wyckoff positions on symmetry lines, where R is the molecule characteristic size. The remaining free space and hence the probability for the structure to exist are shown to be drastically increased when replacing any rotation axis by a screw one. Based on the proposed model, we have explained the peculiar distribution of biomolecular crystals over the space groups, which can be obtained from biomolecule-crystal databases.

DOI: 10.1103/PhysRevE.67.011907 PACS number(s): 87.15.Nn, 61.50.Ah, 87.10.+e, 81.10.Aj

I. INTRODUCTION

There is a continuously growing interest in the study of structure and symmetry of biological objects. Progress in biotechnology enabled the growth of rather perfect crystals made of biological macromolecules (proteins, nucleic acids, complexes, viruses, etc.) and the compilation of large databanks on the structure of thousands of biomolecular crystals, such as the biological macromolecule crystallization database (BMCD) [1] and the protein data base (PDB) [2]. These data offer a vast field for a comprehensive theoretical analysis.

One can reveal several important features of biomolecular crystals which are connected with the symmetry of biomolecules themselves. First, the overwhelming majority of biomolecular crystals contains neither an inversion center nor mirror symmetry. To the best of our knowledge, there is only one example of synthetically prepared protein which has been crystallized in a centrosymmetric form [3].

Second, the crystal distribution over the remaining space groups has another intriguing peculiarity: the biomolecules tend to crystallize mostly in space groups with screw axes. Moreover, the number of observed structures generally increases with the number of screw axes in the group [1,2]. Indeed, according to Ref. [1], among orthorhombic crystals with a simple lattice, there are four crystals belonging to the $P222$ space group (three rotation axes, no screw axis), 18 with the $P222_1$ group (two rotation axes, one screw axis), 150 with $P2_12_12$ group (one rotation axis, two screw axes), and 801 with the $P2_12_12_1$ group (no rotation axis, three screw axes). Thus, replacing each rotation axis by a screw one increases the number of observed crystals by a factor 5 to 8. The alternative database [2] gives similar ratios for the crystals in question (for details, see Table I below). The same tendency is evident for $P2$ and $P2_1$ groups, $P3$, $P3_1$, and $P3_2$ groups, etc.

The fact that crystals prefer structures belonging to only a few space groups was discussed many years ago [4]. However, to explain this trend, no model has been proposed there, apart from some general speculations about close-packed configurations only. This was not surprising since the crystals compiled together in Ref. [4] were dissimilar, i.e., inorganic and organic ones, including molecular crystals made of molecules having various point symmetries.

In Sec. II, we investigate general relations between the symmetries of an isolated molecule and its crystalline forms. Then, based on these group-theory considerations and taking into account the packing restrictions as well, we introduce a model for biomolecule-crystal formation (Sec. III). In Sec. IV, we discuss a comparison of our theoretical predictions with the experimental results.

II. GENERAL RELATIONSHIPS BETWEEN THE SYMMETRY GROUPS OF A MOLECULE AND OF THE MOLECULAR CRYSTAL IT GENERATES

A. Notations

Let the space group of a crystal be G consisting of the elements $g=(h|\mathbf{v}_h+\mathbf{a_n})$, where the orthogonal operation h is followed by the improper translation $\mathbf{v}_h$ and lattice translation $\mathbf{a_n}$.

$$\mathbf{a_n}=\sum_{i=1}^{3} n_i\mathbf{a}_i, \tag{1}$$

$\mathbf{a}_i$ being primitive translation vectors. The vectors $\mathbf{a_n}$ form the invariant subgroup T_a of the space group G ($T_a<G$). The point group $G_{\rm cr}$ of the $n_{\rm cr}$ orthogonal operations h describes the symmetry of directions in the crystal and is called

*Permanent address: Ioffe Physico-Technical Institute, Politekhnicheskaya 26, 194021, St. Petersburg, Russia.
†Permanent address: Institute of Fine Mechanics and Optics, Sablinskaya 14, 197101 St. Petersburg, Russia.

the *crystalline class* or point symmetry group of the crystal. The orthogonal operations h can be *proper* [rotations, $\det(h)=1$] or *improper* ones [inversion, mirror plane, mirror rotation, $\det(h)=-1$].

The *site symmetry group* $G_{\mathbf{q}}$ of the point $\mathbf{q}$ in the space of the crystal is formed by the $n_{\mathbf{q}}$ operations $g_{\mathbf{q}} \in G$ leaving invariant the site $\mathbf{q}$ of the crystal ($G_{\mathbf{q}} \subset G$) [5]. The site group $G_{\mathbf{q}}$ is a point group. The number of its orthogonal operations cannot be superior to that of the crystal class G_{cr}. Therefore,

$$G_{\mathbf{q}} \subseteq G_{\mathrm{cr}}. \tag{2}$$

Let us write the decomposition of G into cosets with respect to its site subgroup $G_{\mathbf{q}}$,

$$G = \sum_{n,j=1}^{t_q} (h_j|\mathbf{v}_j + \mathbf{a}_{\mathbf{n}}) G_{\mathbf{q}}, \quad t_{\mathbf{q}} = n_{\mathrm{cr}}/n_{\mathbf{q}}. \tag{3}$$

An array of points $\mathbf{q}_{j,\mathbf{n}} = (h_j|\mathbf{v}_j + \mathbf{a}_{\mathbf{n}})\mathbf{q}$ is called a *crystallographic orbit* (or right system of points) [5]. It has $t_{\mathbf{q}}$ points per primitive unit cell, the number $t_{\mathbf{q}} = n_{\mathrm{cr}}/n_{\mathbf{q}}$ being the *multiplicity* of the Wyckoff position. The points $\mathbf{q}_{j,\mathbf{n}}$ of the orbit have the isomorphic site symmetry groups $G_{j,\mathbf{n}} = (h_j|\mathbf{v}_j + \mathbf{a}_{\mathbf{n}})\, G_{\mathbf{q}} (h_j|\mathbf{v}_j + \mathbf{a}_{\mathbf{n}})^{-1}$. The orbit is characterized by any of its points $\mathbf{q}$ taken within the primitive unit cell. In the following, international notation [6] is used for point and space groups.

Crystallographic orbits are partitioned according to the so-called *Wyckoff positions*. All the crystallographic orbits with the same site symmetry group are related to the same Wyckoff position, the latter being unambiguously characterized by the group $G_{\mathbf{q}}$ and denoted by small Roman letters. Among the Wyckoff positions, there are [5,6] (a) isolated symmetry points; (b) the points along a rotation axis $\{G_{\mathbf{q}} = C_n(n)$, C_{nv} (nm or nmm), $n \neq 1$, one-parameter array$\}$; (c) the points on a symmetry plane $\{G_{\mathbf{q}} = C_s(m)$, two-parameter array$\}$; and (d) the points of the general position $\{G_{\mathbf{q}} = C_1(1)$, three-parameter array$\}$.

B. Molecular and crystal symmetries

In the molecular crystals, not atoms but molecules are elementary building blocks, from which the crystal is constructed. Let a molecule with the symmetry described by the group G_{mol} occupy the Wyckoff position $\mathbf{q}$, i.e., let its center of gravity reside at this point with the site symmetry group $G_{\mathbf{q}}$. The symmetry operations $g_{\mathbf{q}}$ ($g_{\mathbf{q}} \in G_{\mathbf{q}}$) have to be also the symmetry operations of the molecule which occupies the position $\mathbf{q}$ in the crystal, i.e., $g_{\mathbf{q}} \in G_{\mathrm{mol}}$, and then

$$G_{\mathbf{q}} \subseteq G_{\mathrm{mol}}. \tag{4}$$

In the particular case when $G_{\mathbf{q}} = G_{\mathrm{cr}}$ (symmorphic space group; the crystallographic orbit has only one representative in the primitive unit cell),

$$G_{\mathrm{cr}} \subseteq G_{\mathrm{mol}}. \tag{5}$$

When $G_{\mathbf{q}} \subset G_{\mathrm{cr}}$ ($G_{\mathbf{q}} \neq G_{\mathrm{cr}}$), the groups G_{cr} and G_{mol} are not necessarily connected by a group-subgroup relation.

In the molecular crystals, the binding between the atoms of the molecule being in general much stronger than the binding between the molecules, the influence of the crystalline surrounding on the relative positions of the atoms in the molecule is very weak. Nevertheless, strictly speaking, the crystalline field reduces the symmetry of the molecule from G_{mol} for the free molecule to $\tilde{G}_{\mathrm{mol}} = G_{\mathbf{q}}$ for the molecule in the crystal. This reduction of symmetry of the molecule appears, for example, in the splitting of the electron energy levels. If one does not take into account the small modifications of the molecular structure caused by the crystalline neighborhood, one can attribute to the molecule in the crystal the symmetry G_{mol} of the free molecule.

The atoms of the molecule have a symmetry inferior or equal to that of the molecule itself (free or in the crystalline surrounding). By definition, the site symmetry group of an atom in the molecule is a subgroup of the symmetry group of the molecule ($G_{\mathrm{at}} \subseteq G_{\mathrm{mol}}$ for the free molecule or $\tilde{G}_{\mathrm{at}} \subseteq \tilde{G}_{\mathrm{mol}}$ for the molecule in the crystal). The relation $G_{\mathrm{at}} = G_{\mathrm{mol}}$ ($\tilde{G}_{\mathrm{at}} = \tilde{G}_{\mathrm{mol}}$) is possible for the atom at the symmetry center of the molecule, or at the symmetry line for $G_{\mathrm{mol}} = C_r, C_{nv}$ ($\tilde{G}_{\mathrm{mol}} = C_n, C_{nv}$), or at the symmetry plane for $G_{\mathrm{mol}} = C_s$ ($\tilde{G}_{\mathrm{mol}} = C_s$). In a molecule without symmetry ($G_{\mathrm{mol}} = \tilde{G}_{\mathrm{mol}} = C_1$), also the atoms have no symmetry ($G_{\mathrm{at}} = \tilde{G}_{\mathrm{at}} = C_1$).

C. The case of chiral molecules

If the symmetry group G_{mol} of the molecule contains improper operations, it is called symmetric. Any orthogonal operation g ($g \in G_{\mathrm{mol}}$ or $g \notin G_{\mathrm{mol}}$) leaves the symmetric molecule invariable except (if $g \notin G_{\mathrm{mol}}$) for its position in space. The symmetric molecule coincides with its mirror-image counterpart. If the symmetry group G_{mol} of the molecule consists of proper operations only, it is called chiral. The improper operations transform a chiral molecule into its mirror-image counterpart. In particular, a molecule without any symmetry, i.e., a molecule with the symmetry group $G_{\mathrm{mol}} = C_1$, is chiral. Its mirror-image counterpart cannot be superposed with the molecule itself.

The crystal class G_{cr} of order n_{cr} can consist of proper operations only (of n_{cr} proper rotations) or of $n_{\mathrm{cr}}/2$ proper and $n_{\mathrm{cr}}/2$ improper operations. If the crystal class contains proper operations only, there could be the following Wyckoff positions: (i) the points on symmetry lines (($G_{\mathbf{q}} = C_n$, $n = 2,3,4,6$); (ii) symmetry-line intersections, called isolated symmetry points $\{G_{\mathbf{q}} = D_n$, $n = 2,3,4,6$; $T,O)$. The Wyckoff positions in these "proper" crystals (or, in other words, the crystallographic orbits) can be occupied by either symmetric or chiral molecules.

If there are improper operations in the crystal class, the Wyckoff positions (or the crystallographic orbits) can be occupied by symmetric molecules. As to chiral molecules in this case, they can occupy only half of the points in the primitive unit cell of any crystallographic orbit, the other half of them being occupied by their mirror-image counterparts. The existence of a molecular crystal related to a crystalline class with improper operations is forbidden for chirally pure crystals (formed by identical molecules, without mirror-image counterparts).

III. THE "FREE-SPACE" MODEL

A. Symmetry requirements

The large biomolecules are chiral, the overwhelming majority of them having low symmetry, $G_{\mathrm{mol}}=C_1$. Any point within such a molecule has C_1 symmetry too. In particular, the constituting atoms can occupy only general positions, i.e., Wyckoff positions with $G_{\mathbf{q}}=C_1$ ($\tilde{G}_{\mathrm{at}}\subseteq\tilde{G}_{\mathrm{mol}}$, see Sec. II B). So, the higher-symmetry points and lines in a crystal form a ''forbidden space'' for the biomolecules. Note that this space is one-dimensional for the crystals in question since all the symmetry points lie at symmetry lines.

What is more, the finite size of biomolecules induces additional restrictions. To avoid the molecule overlap with its rotational image, the centers of molecules cannot be nearer than some distance R from the symmetry axes of the crystal, R being a characteristic size of the molecule. If, for simplicity, one implies that molecules have a globular shape with a diameter D, then $R=D/2$ for the C_2 rotation axis, $R=D/\sqrt{3}$ for the C_3 one, $R=D/\sqrt{2}$ for the C_4 one, and $R=D$ for the C_6 axis. Thus, the higher the order n of an n-fold rotational axis is, the larger is the forbidden space. This consideration could be generalized not only for ellipsoidal molecules but also for an arbitrary molecular shape. Note that for a screw axis, the distance R depends on the lattice parameter B along this screw axis, namely $R=\sqrt{(D^2/4-B^2/16)}$ for the twofold $(C_2|0,0,1/2)$ screw axis, $R=\sqrt{(D^2/3-B^2/27)}$ for the threefold $(C_3|0,0,1/3)$ one, $R=\sqrt{(D^2/2-B^2/32)}$ for the fourfold $(C_4|0,0,1/4)$ screw axis, and $R=\sqrt{(D^2-B^2/36)}$ for the sixfold $(C_6|0,0,1/6)$ one. However, the forbidden space induced by a screw axis appears only at lattice parameter B smaller than a critical value [e.g., $B<2D$ for the $(C_2|0,0,1/2)$ screw axis]. Moreover, screw axes often coincide with rotational axes generating larger forbidden spaces. For the $P4_2$ space group taken as an example, the $(C_4|0,0,1/4)$ screw axis coincides with a C_2 rotational axis. So, the restrictions imposed by the screw axes are weaker than those imposed by the rotational axes.

Summing up, the low symmetry of biomolecules limits the space in a crystal unit cell in which the molecules can reside. There is a forbidden space around any symmetry axis, the space *free* for molecule centers to reside being out of the R-radius cylinders around the axes. This free space can be a discrete set of points or a continuum (one-, two-, or three-dimensional). Since the contributions of rotational axes to the forbidden space are larger than those of screw axes, one should focus attention on the number of rotational axes (symmetry lines) per unit cell.

It is reasonable to assume that the existence of forbidden space hampers crystal formation. Thus, the largest probability for molecules to crystallize is for space groups having no symmetry restrictions, i.e., for structures with C_1 Wyckoff positions only. The addition of each symmetry line within the unit cell decreases the free space and thus this probability. The existence of the line intersections (higher-symmetry points) increases comparatively the free-space volume due to the partial overlap of forbidden-space cylinders around the lines.

We therefore suggest that there is the free space V_{free} in the unit cell where the molecule centers are allowed to reside by symmetry considerations and that a probability for a molecule to crystallize with a particular space-group symmetry is somehow proportional to this free space.

B. Packing requirements

The relative free-space volume $V_{\mathrm{free}}/V_{\mathrm{cell}}$ depends on the ratio between molecule and unit-cell volumes, i.e., on the packing coefficient

$$p=N_{\mathrm{mol}}V_{\mathrm{mol}}/V_{\mathrm{cell}}, \tag{6}$$

where V_{cell} is the crystallographic unit-cell volume, V_{mol} is the molecule volume, and N_{mol} is the number of biomolecules per crystallographic unit cell. The smaller the packing coefficient is, the larger is the relative free-space volume.

Note that, for the body-centered (I) and base-centered (C) lattices, the crystallographic unit-cell volumes and the numbers of molecules per cell are two times larger than those for the related primitive (P) lattices. For the face-centered (F) lattices, the volumes and numbers are four times larger than for the P lattices. Correspondingly,

$$N_{\mathrm{mol}}=k\sum_{j=1}^{s} m_j t_{q_j}, \tag{7}$$

where k is the ratio between the crystallographic and primitive cell volumes, that is, 1, 2, 2, or 4 for the primitive, base-centered, body-centered, and face-centered lattices, respectively, s is the number of occupied Wyckoff positions, m_j is the number of occupied crystallographic orbits for the Wyckoff position $\mathbf{q}_j$, and t_{q_j} is the number of points in the crystallographic orbit related to the $\mathbf{q}_j$ Wyckoff position.

In our case of biomolecules without symmetry, only the lowest-symmetry Wyckoff position is occupied, that is, $s=1, n_{\mathbf{q}}=1$. Then the position multiplicities are $t_{\mathbf{q}}=n_{\mathrm{cr}}$ and the minimum possible ($m=1$) numbers of biomolecules per primitive and crystallographic unit cells equal n_{cr} and $N_{\mathrm{cr}}=k\ n_{\mathrm{cr}}$, respectively.

On the other hand, crystals in nature tend to form possibly the most compact structure, thus increasing the packing coefficient. However, such a structure should be compatible with the symmetry requirements. For low-symmetry molecules, the close packing is unattainable for some crystal systems. For instance, the so-called closest-cubic packing, $p=\pi/3\sqrt{2}\approx 0.741$, is impossible for any group of the cubic O class for any number of crystallographic orbits, possible for the $P2$ group beginning with two orbits, and possible for the $P1$ group for any number of orbits.

Note that the relative free space versus packing coefficient is a monotonically decreasing function for any space group. The numerical calculations (see below) show that, for a given packing coefficient, among two crystal structures with different space groups, the structure with the largest *attainable* packing coefficient usually ensures the largest free space for molecules.

Let us summarize the main points of the free-space model. First, the existence of forbidden space around symmetry axes hinders the crystal formation. The smaller the number of symmetry lines per unit cell is, the less is the forbidden space, and thus the larger is the free space. Second, the crystal structure tends to be the most compact but compatible with the symmetry requirements. As a result of the two opposite tendencies (enlarging either the free space V_{free} or the packing coefficient $p \sim 1/V_{\text{free}}$), the crystal is formed. The probability for molecules to form a crystal with a particular space group is assumed to be proportional to the free space in the unit cell.

IV. RESULTS AND DISCUSSION

A. Molecule chirality and possible crystal structures

Nature is dominated by chemical isomers of one-handedness rather than the others. For example, L-aminoacids predominate over D-aminoacids in most living organisms. (Note, some of the L-compounds are not levorotatory but dextrorotatory.) The chirality is intrinsic not only for all biopolymers (proteins, nucleic acids, carbohydrates, lipids, etc.), but also for some simpler compounds in living cells. Except for glycine, which is symmetric, the canonic imino/aminoacids are chiral. Thus nearly all molecules are synthesized by living organisms.

The above group-theory analysis (see Sec. II C) forbids the existence of a chirally pure biomolecular crystal (formed by identical molecules, without mirror-image counterparts) related to a crystalline class with improper operations (inversion, mirror plane, mirror rotation). Such molecules can generate chirally pure crystals with only 66 of the 230 space groups consisting of the proper operations only. These groups, which are presented in Table I, are related to the 11 crystal classes: $C_1(1)$, $C_2(2)$, $D_2(222)$, $C_3(3)$, $D_3(32)$, $C_4(4)$, $D_4(422)$, $C_6(6)$, $D_6(622)$, $T(23)$, and $O(432)$, where the numbers in parentheses refer to international notations.

The above statement is fully confirmed by the analysis of the space groups of thousands of observed biomolecular crystals [1,2] provided that the biological macromolecules have low C_1 symmetry. The synthetically prepared protein crystal with space group $P\bar{1}$ (centrosymmetric form) [3] is not an exception since half of the Wyckoff positions are occupied by levorotatory molecules and the other half by their mirror-image dextrorotatory counterparts.

B. Distribution of biomolecule crystals over the space groups

According to our symmetry consideration, the largest probability for molecules to crystallize exists in space groups having no high-symmetry Wyckoff positions except for C_1, e.g., $P1$ of the triclinic system, $P2_1$ of the monoclinic one, and $P2_12_12_1$ of the orthorhombic D_2 system. Indeed, these 3 of the 66 possible groups provide nearly half of all biomolecular crystals (see Table I).

The addition of each next symmetry line within the unit cell decreases the free space and thus this probability, whereas the existence of the line intersections increases them comparatively. Indeed, within the row of the orthorhombic crystals with a simple lattice, $P2_12_12_1$-$P2_12_12$-$P222_1$-$P222$, the addition of each successive rotation axis decreases the number of observed crystals by a factor 5 to 8.

Moreover, any increase in forbidden space is accompanied by a decrease in the number of crystals. Proving this, Table I shows a relationship between the numbers of high-symmetry lines/points that correlates with the forbidden space and the crystal-distribution statistics (BMCD-PDB-averaged) over the space groups of 11 allowed crystal classes.

It is worthwhile to recall that the higher the order of a rotational axis is, the larger is the forbidden space. As a result, the most compact structures are formed for the least symmetric classes only (triclinic crystals without symmetry axes, and monoclinic and orthorhombic crystals with the twofold axes), and the number of the trigonal, tetragonal, hexagonal, and cubic crystals (with the threefold, fourfold, and sixfold axes) drastically decreases.

One can see that an increase in forbidden space is regularly (without any exception) followed by a decrease in the number of crystals within any crystal class. This regular trend proves the model qualitatively. The quantitative analysis is cumbersome even without respect to all features of biomolecules (irregular shape, intermolecular and biomolecule-water interactions in a crystal, etc.). However, to illustrate the possibility of a still approximate, averaged description, we give an example of such a quantitative analysis for one of the simplest systems, namely the monoclinic one, where there are three space groups $P2$, $P2_1$, and $C2$. A less detailed analysis is carried out for the other crystal classes.

1. Monoclinic crystals

In the space group $P2$, there are four symmetry lines $(0,y,0)$, $(0,y,\frac{1}{2})$, $(\frac{1}{2},y,0)$, and $(\frac{1}{2},y,\frac{1}{2})$ and two molecules per unit cell [5,6]. As noted above, the molecule centers cannot lie in the forbidden space limited by D-diameter cylinders around the symmetry lines. Therefore, to calculate the free volume, one should subtract the forbidden volume from the unit-cell volume and take into account possible overlaps of cylinders around symmetry lines. Far from the overlap region, for the unit cell with the A,B,C lattice constants and the angle β between **A** and **C** translation vectors, we obtain $V_{\text{free}}=V_{\text{cell}}-4V_{\text{cyl}}=ABC\sin\beta-4B\pi D^2/4$. Then the relative free volume is $V_{\text{free}}/V_{\text{cell}}=1-\pi D^2/AC\sin\beta$. When the packing coefficient p increases, the free space monotonously decreases as $1-\xi p^{2/3}$. For $D>\min(A,C,\sqrt{(A^2+C^2\pm 2AC\cos\beta)})$, the cylinder overlap takes place. In order not to overload the paper, we give only the formula for an oversimplified case, $A=C$ and $\beta=\pi/2$ (the monoclinic symmetry can be kept since the symmetry of the Bravais lattice of molecular crystals can be higher than required by the point symmetry),

$$V_{\text{free}}/V_{\text{cell}}=1-(2D/A)^2\{\pi/4-\arccos(A/2D)+(A/2D)\sqrt{[1-(A/2D)^2]}\}. \quad (8)$$

TABLE I. Distribution of crystals made of biological macromolecules over the space groups according to the BMCD [1] and PDB [2] databases. For each group, the minimum possible numbers of biomolecules per primitive and crystallographic unit cell (n_{cr} and N_{cr}) and the numbers of symmetry lines, N_L, and symmetry points, N_P (at the line intersections, in parentheses), per crystallographic unit cell are given.

Crystal system and crystal class	Space group	Group No.	n_{cr}	N_{cr}	N_L ($-N_P$)	Number of crystals (ppm)		
						BMCD	PDB	Average
Triclinic C_1	$P1$	1	1	1	0	32	199	115
Monoclinic C_2	$P2$	3	2	2	4 (−0)	5.5	1	3
	$P2_1$	4	2	2	0	144	112	128
	$C2$[a]	5	2	4	4 (−0)	75.5	74.5	75
Orthorhombic D_2	$P222$	16	4	4	12 (−8)	1	0	1
	$P222_1$	17	4	4	8 (−0)	5	0.5	3
	$P2_12_12$	18	4	4	4 (−0)	42	48	45
	$P2_12_12_1$	19	4	4	0	223	201	212
	$C222_1$	20	4	8	8 (−0)	38.5	38.5	39
	$C222$	21	4	8	16 (−8)	5	1	3
	$F222$	22	4	16	24 (−16)	1	1	1
	$I222$	23	4	8	12 (−8)	17.5	17	17
	$I2_12_12_1$	24	4	8	12 (−0)	8	0.5	4
Tetragonal C_4	$P4$	75	4	4	4 (−0)	1.5	0.5	1
	$P4_1$	76	4	4	0	7	6	7
	$P4_2$	77	4	4	4 (−0)	1	0.5	1
	$P4_3$	78	4	4	0	6.5	5	6
	$I4$	79	4	8	4 (−0)	4.5	3	4
	$I4_1$	80	4	8	4 (−0)	0.5	2	1
Tetragonal D_4	$P422$	89	8	8	16 (−8)	1.5	0.5	1
	$P42_12$	90	8	8	8 (−4)	5	3.5	4
	$P4_122$	91	8	8	12 (−0)	10.5	1.5	6
	$P4_12_12$	92	8	8	4 (−0)	36.5	26	31
	$P4_222$	93	8	8	16 (−12)	1.5	0.5	1
	$P4_22_12$	94	8	8	8 (−4)	9.5	7	8
	$P4_322$	95	8	8	12 (−0)	8	2	5
	$P4_32_12$	96	8	8	4 (−0)	50	39.5	45
	$I422$	97	8	16	20 (−12)	5.5	6	6
	$I4_122$	98	8	16	20 (−8)	2	5.5	4
Trigonal C_3 (rhombohedral)	$R3$	146	3	3	1 (−0)	10.5	11	11
Trigonal C_3 (hexagonal)	$P3$	143	3	3	3 (−0)	1	2	1
	$P3_1$	144	3	3	0	4	2.5	3
	$P3_2$	145	3	3	0	6	3.5	5
Trigonal D_3 (rhombohedral)	$R32$	155	6	6	7 (−2)	11	11	11
Trigonal D_3 (hexagonal)	$P312$	149	6	6	13 (−6)	0	0	0
	$P321$	150	6	6	9 (−2)	5	6	5
	$P3_112$	151	6	6	10 (−0)	4.5	0.5	3
	$P3_121$	152	6	6	6 (−0)	39	32.5	36
	$P3_212$	153	6	6	10 (−0)	4	1	3
	$P3_221$	154	6	6	6 (−0)	39.5	45.5	42
Hexagonal C_6	$P6$	168	6	6	6 (−0)	2.5	5.5	4
	$P6_1$	169	6	6	0	11.5	9	10

TABLE I. (*Continued.*)

Crystal system and crystal class	Space group	Group No.	n_{cr}	N_{cr}	N_L $(-N_P)$	Number of crystals (ppm)		
						BMCD	PDB	Average
	$P6_5$	170	6	6	0	9.5	5.5	8
	$P6_2$	171	6	6	4 (−0)	3.5	1	2
	$P6_4$	172	6	6	4 (−0)	2	1.5	2
	$P6_3$	173	6	6	3 (−0)	7	5.5	6
Hexagonal D_6	$P622$	177	12	12	22 (−12)	3	1	2
	$P6_122$	178	12	12	16 (−0)	19.5	16	18
	$P6_522$	179	12	12	16 (−0)	18	8	13
	$P6_222$	180	12	12	20 (−12)	8	2.5	5
	$P6_422$	181	12	12	20 (−12)	8	2.5	5
	$P6_322$	182	12	12	19 (−8)	8.5	4.5	7
Cubic T	$P23$	195	12	12	16 (−8)	1	0	0
	$F23$	196	12	48	40 (−16)	0.5	0	0
	$I23$	197	12	24	19 (−8)	4.5	3	4
	$P2_13$	198	12	12	4 (−0)	6.5	6.5	6
	$I2_13$	199	12	24	19 (−0)	3	4	3
Cubic O	$P432$	207	24	24	28 (−8)	0	0	0
	$P4_232$	208	24	24	31 (−28)	1	0.5	1
	$F432$	209	24	96	82 (−40)	2	1.5	2
	$F4_132$	210	24	96	104 (−48)	1	0.5	1
	$I432$	211	24	48	43 (−28)	2	2	2
	$P4_332$	212	24	24	20 (−8)	1	0.5	1
	$P4_132$	213	24	24	20 (−8)	2	0.5	1
	$I4_132$	214	24	48	49 (−40)	1	1	1

[a]As for some other groups, here equivalent notations are included: $A2$, $B2$, $B112$, $C121$, $C2_1$, $C12_11$, $I12_11$, $I121$, $I2$, and $I2_1$.

One can see that analytical calculations are too complicated. The results of numerical calculations by the Monte Carlo method are given in Fig. 1 both for the oversimplified case (bold lines) and for different crystal-lattice parameters. In the latter case, the results are averaged over a set of the *A,B,C* lattice parameters with their ratios varying from 1/3 to 3/1, which is broader than the real biomolecule-crystal lattice-parameter distribution. It can be seen that the difference between the two cases is not significant.

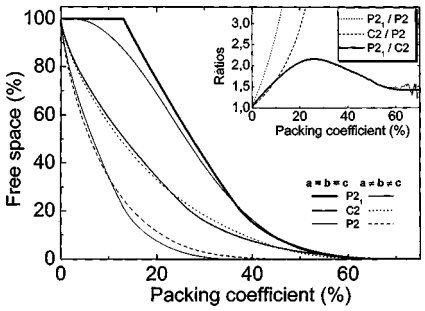

FIG. 1. The dependences of relative free-space volume on the packing coefficient p for three groups of monoclinic crystals. Thick lines are calculated within an oversimplified model, thin lines within a more realistic one (see text). Inset shows the corresponding ratios of these free-space volumes averaged for the same packing coefficient.

In the $P2_1$ space group, the rotation axis is substituted by a screw one. As a result, the symmetry of Wyckoff positions along the lines is reduced to general ones. The symmetry lines disappear. When $B \geqslant 2D$, there is no forbidden space, the free-space volume coinciding with the unit-cell volume. For smaller B, forbidden space appears too, related to the screw axis [also a cylinder around each axis, but with the diameter $d=\sqrt{(D^2-B^2/4)}$]. The relative free volume drops abruptly from the maximal value 100%, as shown in Fig. 1. Nevertheless, the free space for this group remains incomparably larger than that for the $P2$ space group, as one can see from the inset in the figure, where free-space ratios are presented.

In the space group $C2$, the symmetry lines in the base-centered crystallographic unit cell coincide with those in the primitive $P2$ unit cell (equivalent to the crystallographic one for the primitive lattices). However, there are four molecules per crystallographic unit cell instead of two molecules for $P2$, and the unit-cell volume V_{cell} is two times larger for the same packing coefficient. As a result, this group stands in between $P2$ and $P2_1$. According to the numerical calculations (see Fig. 1), the relative free-space volume V_{free}/V_{cell} is 1.5 to 2 times less here than that for the $P2_1$ group, being much larger than for the $P2$ group.

Comparing our model results with the available experimental data, we see a good agreement: the averaged (over

both databases, see Table I) ratio of numbers of crystals for the $P2_1$ and $C2$ groups is about 1.7, the number of reported crystals with the $P2$ group being much smaller than with the other two groups.

2. Other crystal classes

One can make detailed calculations for any crystal system. In the above model, the probability for molecules to form a crystal with a particular space group is assumed to be proportional to the relative free space $V_{\text{free}}/V_{\text{cell}}$ in the unit cell assuring the largest packing coefficient compatible with the symmetry restrictions. Since the forbidden space is created by symmetry lines, even the number of symmetry lines provides information on the free space.

This is clearly seen from the above comparison of the numerical and qualitative results for groups $P2$, $P2_1$, and $C2$. So, in predicting the biomolecule-crystal distribution over space groups, it is reasonable for the sake of simplicity to restrict ourselves to a comparison of the numbers of symmetry lines (taking into account their intersections, which increases V_{free}), thus implying the comparison of the relative free-space volumes.

The case of the orthorhombic D_2 class is exemplary. In the symmorphic orthorhombic space group $P222$, there are 12 symmetry lines (four lines along each direction *x*, *y*, and *z*) and eight symmetry points lying at the intersections of these lines. For the nonsymmorphic $P222_1$ group, all the symmetry points and symmetry lines along the screw axis direction disappear, the number of symmetry lines reduces to eight, and there is no line intersection. As a result, the free space for a molecule increases. For the $P2_12_12$ group, the number of symmetry lines reduces to four and the free space increases even more, reaching the maximum for the group $P2_12_12_1$ containing only general Wyckoff positions with the C_1 symmetry. And this is immediately reflected in the number of observed crystals. Indeed, for the orthorhombic crystals with a simple lattice, replacing each rotation axis by a screw one increases the number of observed crystals by one order of magnitude.

For the base-centered orthorhombic lattices, the replacement of rotation axis in $C222$ with a screw one in the $C222_1$ group reduces twice the number of symmetry lines (from 16 to 8), thus increasing the free volume. However, line intersections disappear, thus partly compensating for the effect. Nevertheless, the ratio of distribution numbers is about 10. For both groups with the orthorhombic body-centered lattices, $I222$ and $I2_12_12_1$, there are 12 symmetry lines per crystallographic unit cell. Nevertheless, in the $I222$ group they intersect, which increases the free volume, whereas in the $I2_12_12_1$ group the symmetry lines do not intersect and the free volume is less. Correspondingly, the number of crystals is also smaller (about four times).

To compare the orthorhombic subclasses with the primitive (*P*) and centered (*C* and *I*) lattices, one should take into account that the centering of the unit cell increases both V_{free} and V_{cell}. For example, in going from $P2_12_12$ to $C222_1$, the number of rotational axes and, consequently, the forbidden-space volume double. However, the unit-cell volume also doubles, keeping the relative free space and the number of crystals almost the same. Furthermore, in going from $C222_1$ to $I2_12_12_1$, the unit-cell volume is not changed, whereas the number of axes is increased from 8 to 12 (without intersections in all cases). As a result, the number of crystals also decreases.

In the tetragonal D_4 class, the minimum number of symmetry lines (four) appears for $P4_32_12$ and $P4_12_12$ space groups (with a slight difference between them), which corresponds to the maximum free C_1 space. Not surprisingly, two-thirds of the tetragonal D_4 crystals are from these groups. On the contrary, the $P422$ and $P4_222$ groups (16 symmetry lines) are practically not present. One can also compare the D_4 tetragonal body-centered groups ($I422$ and $I4_122$), where the decrease in the number of line intersections results in a decrease in free space and is accompanied by a decrease in the number of crystals. A quite similar but less clear picture can be seen for space groups of the other tetragonal class, C_4. For example, the case of C_4 tetragonal body-centered groups (four nonintersecting lines per crystallographic unit cell in both $I4$ and $I4_1$) can be made transparent with direct numerical calculations only. Note, however, that two-thirds of the C_4 crystals are from $P4_3$ and $P4_1$ space groups, as can be expected.

Within the trigonal system, a few subsystems and subclasses can be found. For example, within the trigonal (hexagonal) 321 subclass, equal maxima of distribution numbers are obviously expected (and occur) for $P3_121$ and $P3_221$ groups, where six high-symmetry lines are present. The two groups comprise more than half of all trigonal crystals.

For the hexagonal D_6 class, our model predicts the largest (and nearly equal) distribution numbers for the $P6_122$ and $P6_522$ space groups (16 high-symmetry lines), smaller (and also nearly equal) numbers for the $P6_222$ and $P6_422$ groups (20 lines), and the minimum number of crystals with the $P622$ group (22 high-symmetry lines), everything being observed in reality. Indeed, two-thirds of the crystals are from the two former space groups, whereas the latter provide less than a quarter. A very similar picture takes place for the hexagonal C_6 class, namely equal maxima for $P6_1$ and $P6_5$ (no high-symmetry positions in the crystal), parity, and minima for $P6_2$, $P6_4$, and $P6$, but statistical error makes it not as clear as in the previous case.

Even for the most symmetric cubic system, where the statistical material is insufficient, the largest number of crystals (about 100 in the PDB) belongs to the $P2_13$ space group with the minimum number of high-symmetry lines per unit cell, which also supports our model.

C. Perspectives for future studies

As the biomolecular-crystal databases increase, it will become possible to test the presented ''free-space'' model analyzing the distribution over the space groups for the various crystals based on the same biomolecule. This will allow us to reveal fine details of biomolecule crystallization.

Indeed, the low symmetry of the biomolecules dictates that the atoms have no symmetry in any biomolecular crystal but C_1. Incidentally, this means that any atom does not change its position symmetry during any crystal-phase tran-

TABLE II. Distribution of crystals made of biological macromolecules over the crystal systems according to the BMCD database [1].

Number of crystals	Triclinic	Monoclinic	Orthorhombic	Tetragonal	Trigonal	Hexagonal	Cubic
All (hundreds)	1	8	12	5.5	4.5	3.5	1
Lysozyme	1	6	15	13	10	4	

sition, and there is no jump of the atom energy (cohesive, potential, etc.), the changes of the structure (lattice parameters, angles) being continuous as well. Since an energy profit for a phase transition is not large, any protein could form any lattice under certain conditions (molecular density in the solution or number of surrounding H_2O molecules, acidity, or pH, etc.). What is more, the probability for a certain protein to have a certain lattice (space group) should be the same as for any other protein (if there are no shape peculiarities). That is to say, the averaged distribution statistics for all proteins and that for a certain protein should be the same.

There are already moderate statistics of distributions of the various crystals made of the identical biomolecules. Among thousands, the lysozyme protein is one of the most studied, 49 lysozyme crystals being presented in the BMCD [1], belonging to 16 space groups (see Table II). One can see the pronounced similarity of distribution over crystal systems for this protein and the others in total. Of course, it is difficult to expect an accurate coincidence. First, obtaining the higher-symmetry phases is more advanced for many further studies, e.g., structural ones. Second, if the shape of a biomolecule is very special, statistics could be different as it takes place for the hemoglobin crystals. Unfortunately, because of small statistics, an analogous search cannot be done for distributions over the space groups for crystals made of other biomolecules. However, the BMCD [1] and PDB [2] databases, which collect slightly different subsets of the protein crystals, show nevertheless that the percentage statistics are very similar. That is why, to prove our *free-space* model, we refer to the combined BMCD-PDB statistics, all averaged over the data on many crystals.

V. CONCLUSION

We have carried out a detailed group-theory analysis of the interrelation between the symmetry of an isolated macromolecule and the possible symmetries of the crystals it can generate. We link rigorously the absence of symmetry in biomolecules with the chirality of the biomolecular crystals.

We have found some restrictions for the molecule disposition in the crystal structure, and suggest a model taking into account both these restrictions and trends of the biomolecule packing. According to our model, the largest probability for molecules to crystallize exists in space groups having no high-symmetry Wyckoff positions.

Within this ''free-space'' model, we performed an analysis of biomolecule crystal distribution for the possible 66 space groups with proper symmetry operations only. The predictions of our model are confirmed by the protein database statistics. So, without resorting to thermodynamics, in a framework of a symmetry approach, we have managed to describe important peculiarities of the biomolecule crystallization.

ACKNOWLEDGMENTS

We thank Dr. M. Ries for very fruitful discussions and acknowledge the support of Mairie de Paris and Ministere de la Recherche.

[1] G. L. Gilliland, M. Tung, D. M. Blakeslee, and J. Ladner, Acta Crystallogr., Sect. D: Biol. Crystallogr. **D50**, 408 (1994).

[2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res. **28**, 235 (2000).

[3] W. R. Patterson, D. H. Anderson, W. F. DeGrado, D. Cascio, and D. Eisenberg, Protein Sci. **8**, 1410 (1999).

[4] B. K. Vainshtein, *Modern Crystallography* (Springer, Berlin-Heidelberg, 1994), Vol. 2.

[5] R. A. Evarestov and V. P. Smirnov, *Site Symmetry in Crystals. Theory and Applications*, 2nd ed., Springer Series in Solid-State Sciences Vol. 108, edited by M. Cardona (Springer, Berlin, 1997).

[6] T. Hahn, *International Tables for Crystallography. Vol. A: Space-Group Symmetry* (Kluwer Academic, Dordrecht, 1983).