



Mathematical aspects of molecular replacement. III. Properties of space groups preferred by proteins in the Protein Data Bank

G. Chirikjian,^{a*} S. Sajjadi,^a D. Toptygin^b and Y. Yan^a

Received 30 July 2014

Accepted 5 November 2014

Keywords: Bieberbach groups; Sohncke groups; protein crystals; normal subgroups.

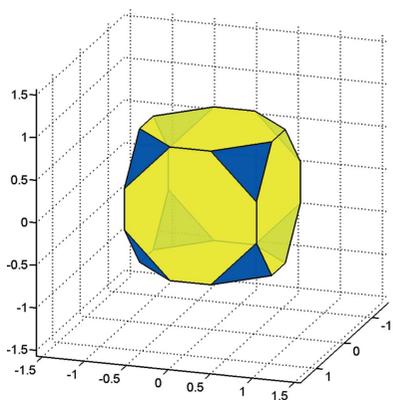
^aDepartment of Mechanical Engineering, Whiting School of Engineering, Johns Hopkins University, Hackerman Hall 116, Baltimore, USA, and ^bDepartment of Biology, Krieger School of Arts and Sciences, Johns Hopkins University, 144 Mudd Hall, 3400 N. Charles St., Baltimore, MD 21218, USA. *Correspondence e-mail: gregc@jhu.edu

The main goal of molecular replacement in macromolecular crystallography is to find the appropriate rigid-body transformations that situate identical copies of model proteins in the crystallographic unit cell. The search for such transformations can be thought of as taking place in the coset space $\Gamma \backslash G$ where Γ is the Sohncke group of the macromolecular crystal and G is the continuous group of rigid-body motions in Euclidean space. This paper, the third in a series, is concerned with viewing nonsymmorphic Γ in a new way. These space groups, rather than symmorphic ones, are the most common ones for protein crystals. Moreover, their properties impact the structure of the space $\Gamma \backslash G$. In particular, nonsymmorphic space groups contain both Bieberbach subgroups and symmorphic subgroups. A number of new theorems focusing on these subgroups are proven, and it is shown that these concepts are related to the preferences that proteins have for crystallizing in different space groups, as observed in the Protein Data Bank.

1. Introduction

This series of papers is concerned with the mathematical properties of the coset space $\Gamma \backslash G$ where $G = SE(3)$ is the continuous ‘special Euclidean group’ consisting of all rigid-body motions in three dimensions that preserve handedness, and $\Gamma < G$ is isomorphic to any of the 65 Sohncke¹ three-dimensional space groups.² The first two papers in this series (Chirikjian, 2011; Chirikjian & Yan, 2012) explored the algebraic and geometric properties of this manifold in general, and explored choices for the associated fundamental domain $F_{\Gamma \backslash G} \subset G$ in the symmorphic case. In those papers it was also shown that $F_{\Gamma \backslash G}$ is the space in which rigid-body searches in molecular replacement (Rossmann, 2001) are performed. [While traditional molecular-replacement methods break the problem up into a three-dimensional rotation search followed by a three-dimensional translation one, our approach characterizes the six-dimensional search space, as does that of Kissinger *et al.* (1999).] In this third paper in the series we focus on the case where Γ is nonsymmorphic and analyze which space groups are more common amongst the roughly 80 000 proteins crystallized to date.

Though G is the (outer) semidirect product of the continuous translation group with the continuous rotation group



¹ There are 65 such handedness-preserving space groups. We referred to these as ‘chiral space groups’ in the previous papers in this series. Here we have updated the terminology and now call them ‘Sohncke groups’ to reflect current IUCr nomenclature.

² Γ is not normal in G and so we do not write G/Γ or G/Γ , whereas in cases when a subgroup is normal, the coset space can be written equivalently in all three of these ways.

[written as $G = X \rtimes \mathcal{R}$ where $X \cong (\mathbb{R}^3, +)$ and $\mathcal{R} \cong \text{SO}(3)$], Γ can only be written as a semidirect product of a discrete translation group and finite rotation group for the 24 symmorphic Sohncke groups. In the second paper in this series, it was explained that for a symmorphic Sohncke group $\Gamma_p = T \rtimes \mathbb{P}$, where T is the finest group of lattice translations³ and $\mathbb{P} \cong \frac{\Gamma}{T}$ is a chiral point group with $|\mathbb{P}| = [\Gamma : T]$ (the index of T in Γ), searching in $F_{\Gamma \setminus G}$ is equivalent to a search in $F_{\Gamma \setminus X} \times \mathcal{R}$ or $F_{\mathcal{T} \setminus X} \times F_{\mathbb{P} \setminus \mathcal{R}}$. And if $\mathbb{S} \triangleleft \mathbb{P}$, then intermediate choices exist to perform a search over a translational domain of the form $F_{\Gamma_s \setminus X}$ where $\Gamma_s = T \rtimes \mathbb{S}$. The domain $F_{\Gamma_s \setminus X}$ is larger in terms of volume than $F_{\Gamma \setminus X}$ and smaller than $F_{\mathcal{T} \setminus X}$. And if such an intermediate choice is made, then the corresponding searches over rotation domains are smaller than \mathcal{R} and larger than $F_{\mathbb{P} \setminus \mathcal{R}}$ in a way that depends on $\mathbb{S} \setminus \mathbb{P}$.

In the current paper, we establish the foundations for similar decompositions of Γ for the 41 nonsymmorphic Sohncke groups. This builds on the concept of a Bieberbach group (or torsion-free space group). A Bieberbach group can be thought of as the opposite extreme from a symmorphic group, in that none of its elements (other than the identity) form a point subgroup. With Bieberbach groups at one extreme and symmorphic groups at the other, all of the remaining space groups lie somewhere in the middle. We show that inside of every nonsymmorphic Sohncke group are both Bieberbach and symmorphic subgroups. We define the concepts of minimal-index Bieberbach and minimal-index symmorphic subgroups, and show how the coset decomposition of space groups by Bieberbach subgroups relates to the relative frequency of occurrence of the various space groups in the Protein Data Bank (PDB). The topics of characterizing search spaces in molecular replacement and the statistical preferences of proteins are treated together in the same paper for a reason: the number of possible arrangements of physical proteins in an environment with crystallographic symmetry is related to the difficulty of the computational problem of searching for these arrangements. In a sense, these problems are dual to each other in that the more possibilities a protein has, the more likely it should be for an energetically favorable arrangement to exist, whereas the more space one needs to search, the higher the computational cost will be. And the issue of how much free space a protein has to roam in the configuration space $\Gamma \setminus G$ is intimately related to the structure of Γ studied here.

Our motivation for this study fits the theme of this series for the following reason. One can view an ideal macromolecular crystal as either an infinite number of identical bodies in X , or $[\Gamma : T]$ bodies in $\mathcal{T} \setminus X$, or a single body in $\Gamma \setminus X$. In all cases, the smallest nonredundant space of motions of these bodies is not G , but rather $\Gamma \setminus G$. When it is possible to write $\Gamma = \Gamma_B \times \mathbb{S}$ where Γ_B is a subgroup consisting only of translations and screw motions, and \mathbb{S} is a point subgroup of the abstract point

group $\mathbb{P} \cong \Gamma/T$, then since as a group $G = X \rtimes \mathcal{R}$ (or equivalently $G = \mathcal{R} \ltimes X$), and as a set $G = X \times \mathcal{R}$, the fundamental domain corresponding to $\Gamma \setminus G$ can be decomposed naturally as

$$F_{\Gamma \setminus G} = F_{\Gamma_B \setminus X} \times F_{\mathbb{S} \setminus \mathcal{R}}. \quad (1)$$

The significance of this is that $\Gamma_B \setminus X$ is a flat manifold (also called a Euclidean space form) and $\mathbb{S} \setminus \mathcal{R}$ is a spherical space form. And both of these objects can be sampled in an efficient way.

The remainder of this paper is structured as follows. In §2 definitions and theorems related to the structure of space groups containing Bieberbach subgroups are provided. §3 addresses the concept of a minimal-index symmorphic subgroup of a space group and that of a minimal-index Bieberbach subgroup. §4 presents a summary of all normal Bieberbach and normal symmorphic subgroups in tabular form. In §5 the preferences of all proteins in the PDB (Berman *et al.*, 2000, 2002; Berman, 2008; Bernstein *et al.*, 1977) are tabulated and analyzed in light of the definitions and results of §2.

2. Bieberbach subgroups: related definitions and new theorems

As a matter of convention in crystallography, the position of a point in a unit cell is given by values of x, y, z each ranging from 0 to 1, regardless of whether the unit cell is primitive or not. Therefore, for a space group in the primitive setting, $P1$ means a translation of the primitive lattice, and $P1 = T_p$. In contrast, for a space group in a nonprimitive (conventional centered) setting, $T_p < T$ (a proper subgroup), and T_p is the translation group of a sublattice of the one defined by T . Since in protein crystallography roughly 20% of the most common space groups encountered are nonprimitive, it makes sense to consider $T_p \leq T$ where equality holds only in the case of P -type space groups.

Let Γ_N denote a space group of nonprimitive ($N \in \{C, I, F, R\}$) type. In this case let $T = T_N$ denote the corresponding minimal-index translation subgroup. It is Γ_N/T_N that is isomorphic to the point group, \mathbb{P} , and not the quotient Γ_N/T_p , since the latter cosets will contain representatives that are pure (fractional) translations. Moreover, it is a fundamental domain of $T_N \setminus X$ that can always be identified with the primitive unit cell whereas when discussing nonprimitive space groups a fundamental domain of $T_p \setminus X$ corresponds to the conventional nonprimitive cell, and

$$\frac{\text{Vol}(T_p \setminus X)}{\text{Vol}(T_N \setminus X)} = \left| \frac{T_N}{T_p} \right| \doteq n, \quad (2)$$

where $n = 2, 2, 4, 3$ depending on the type C, I, F or R , respectively.

The translation group of a primitive lattice, T_N , is related to the translation group $T_p = P1$ of the sublattice corresponding to the nonprimitive (centered) conventional unit cells by conjugation with respect to affine transformations of the form $\alpha_N = (A, \mathbf{0})$ where A is the inverse of the centering matrices

³ This is the lattice such that the unit cell, identified with the fundamental domain $F_{\mathcal{T} \setminus X}$, is primitive, *i.e.*, the smallest one. Every other lattice translation group is a subgroup of T .

given in Hahn (2002) and Wondratschek & Müller (2010). That is, $T_N = (T_p)^{\alpha_N} \doteq \alpha_N T_p \alpha_N^{-1}$.

As a matter of convention, we view the point group $\mathbb{P} \cong \Gamma/T$ as a subgroup of the rotation group, $\mathbb{P} < \mathcal{R}$. Strictly speaking then, \mathbb{P} is not a subgroup of Γ , not even in the symmorphic case. But if Γ_p is symmorphic, then when each $R \in \mathbb{P}$ is appended with a zero translation vector, then $(R, \mathbf{0}) \in \Gamma_p$. Similarly, if $\mathbb{S} \ltimes T < \Gamma$ then $\mathbb{S} \ltimes \{\mathbf{0}\} < \Gamma$ also. Since $\mathbb{S} \ltimes \{\mathbf{0}\} \cong \mathbb{S}$ the distinction between these two objects is often blurred in the literature, and both are called point subgroups. When this distinction is important in our formulation we will indicate so.

Definition 2.1. A Bieberbach group, Γ_B , is a space group that has no torsion elements. That is, there is no $\gamma \in \Gamma_B$ such that for any given $\mathbf{x} \in X$ the equality $\gamma \cdot \mathbf{x} = \mathbf{x}$ holds other than $\gamma = e$. In other words, a Bieberbach group cannot contain a nontrivial point group of the form $\mathbb{S} \ltimes \{\mathbf{0}\}$ as a subgroup. And if a Bieberbach group is also a Sohncke group, then the generators are either pure screw motions or pure translations.

Definition 2.2. A Bieberbach group, Γ_B , and a symmorphic space group, Γ_S , are called *nontrivial* if they are not pure translation groups.

As with any space group, two Bieberbach groups are considered to be equivalent if they are equal under conjugation by an affine transformation that preserves handedness. Recall that the set of all such transformations is

$$\text{Aff}^+(3) = \{(A, \mathbf{t}) | A \in GL^+(3), \mathbf{t} \in \mathbb{R}^3\} = GL^+(3) \ltimes \mathbb{R}^3$$

with the semidirect-product group law $(A_1, \mathbf{t}_1)(A_2, \mathbf{t}_2) = (A_1 A_2, A_1 \mathbf{t}_2 + \mathbf{t}_1)$ where $GL^+(3)$ denotes the group of invertible 3×3 matrices with real entries and positive determinant.⁴

Explicitly, the conjugation action of $\text{Aff}^+(3)$ on Γ can be defined easily when representing elements of both groups as 4×4 matrices. After fixing an origin and choosing a basis for \mathbb{R}^3 , every space group can be represented as a 4×4 matrix of the form

$$\gamma = \begin{pmatrix} R_\gamma & \mathbf{t}_\gamma + \mathbf{v}(R_\gamma) \\ \mathbf{0}^T & 1 \end{pmatrix},$$

where $\mathbf{0}^T = [0, 0, 0]$ and $\mathbf{v}(R_\gamma)$ is the translational part of a screw displacement. Similarly, every affine transformation can be represented as

$$\alpha = \begin{pmatrix} A & \mathbf{a} \\ \mathbf{0}^T & 1 \end{pmatrix},$$

where $A \in GL^+(3)$ and $\mathbf{a} \in \mathbb{R}^3$. Conjugating Γ by α is then equivalent to the matrix operation $\alpha \gamma \alpha^{-1}$ for all $\gamma \in \Gamma$.

When referring to Bieberbach groups in the literature, what is often meant is classes of Bieberbach groups. The classes of Bieberbach groups in three dimensions are important in

⁴ If the + sign were dropped the result would be the full set of affine transformations, $\text{Aff}(3)$, which will not be used here.

differential geometry, see e.g. Szczepański (2012) or Wolf (2010), because each quotient of Euclidean space by a Bieberbach group, $\Gamma_B \backslash X$, produces a flat compact manifold (with fundamental group isomorphic to Γ_B), and there are no flat compact manifolds other than those produced in this way. In contrast, the quotient of Euclidean space by a non-Bieberbach space group is not a manifold, but rather is an orbifold (Thurston, 1997; Montesinos, 1987).⁵ In particular, we are concerned with Sohncke Bieberbach space groups, and these result in quotients $\Gamma_B \backslash X$ that are not only flat, but also orientable manifolds.

Definition 2.3. The classes of Sohncke Bieberbach groups in three dimensions have representatives in the set

$$\mathcal{F} \doteq \{P1, P2_1, P3_1, P3_2, P4_1, P4_3, P6_1, P6_5, P2_1 2_1 2_1\}$$

where the enantiomorphic pairs $(P3_1, P3_2)$, $(P4_1, P4_3)$, $(P6_1, P6_5)$ are equivalent under affine transformations but inequivalent under affine transformations that preserve handedness. This set is called the full family of Sohncke–Bieberbach-group representatives, and an individual representative is denoted as $\Gamma_B \in \mathcal{F}$. If $\mathcal{F}' \subset \mathcal{F}$ then \mathcal{F}' is called a subfamily of Sohncke–Bieberbach-group representatives. In particular, $\{P1\}$ is called the trivial subfamily.

Definition 2.4. Every translation subgroup of the form $\alpha(P1)\alpha^{-1} \in \Gamma$ where $\alpha \in \text{Aff}^+(3)$ is called a trivial Bieberbach subgroup.

Let $\Gamma_B^\alpha \doteq \alpha \Gamma_B \alpha^{-1}$. Suppose that it is possible to find $\alpha, \alpha' \in \text{Aff}^+(3)$ such that $\Gamma_B^\alpha, \Gamma_B^{\alpha'} \leq \Gamma$. The notation

$$[\Gamma : \Gamma_B^\alpha] \doteq |\Gamma_B^\alpha \backslash \Gamma| = |\Gamma / \Gamma_B^\alpha|$$

is used for the index of Γ_B^α in Γ (without presupposing normality). If for some $\alpha \in \text{Aff}^+(3)$ and $\Gamma_B \in \mathcal{F}'$ it is the case that

$$[\Gamma : \Gamma_B^\alpha] \leq [\Gamma : \Gamma_B^{\alpha'}] \tag{3}$$

for all $\Gamma_B' \in \mathcal{F}'$ and all $\alpha' \in \text{Aff}^+(3)$, then Γ_B^α is called a minimal-index Bieberbach subgroup of Γ represented by the subfamily \mathcal{F}' . If \mathcal{F}' is replaced with the full family \mathcal{F} , then Γ_B^α is simply called a minimal-index Bieberbach subgroup of $\text{elb} \backslash \Gamma$ and is denoted as $\Gamma_{B_M}^\alpha$. By this definition, there can be multiple minimal-index Bieberbach subgroups. Here the term ‘minimal-index’ is used to denote the the largest amongst the Bieberbach subgroups in the sense of equation (3). But these are not always ‘maximal subgroups’ in the classical sense.

Theorem 2.1. The minimal-index translation subgroup, $T \leq \Gamma$, is the unique minimal-index Bieberbach subgroup of Γ represented by the trivial subfamily.

Proof. Every member of the trivial family has elements of the form $t = (I, B\mathbf{z})$ where $B\mathbf{z} = z_1 \mathbf{b}_1 + z_2 \mathbf{b}_2 + z_3 \mathbf{b}_3 \in \mathbb{L}$ with $z_i \in \mathbb{Z}$ and $\{\mathbf{b}_i\}$ being a set of lattice vectors. For the finest

⁵ However, the coset space $\Gamma \backslash G$ is always a manifold and usually is not flat.

Table 1
The symmorphic case.

International No.	Γ	T	$[\Gamma : T]$	Γ_{B_M}	$[\Gamma : \Gamma_{B_M}^\alpha]$
1	$P1$	T_P	1	$P1$	1
3	$P2$	T_P	2	$P2_1$	2
5	$C2$	T_C	2	$P2_1$	2
16	$P222$	T_P	4	$P2_1$	4
21	$C222$	T_C	4	$P2_1, P2_1, 2_1$	4
22	$F222$	T_F	4	$P2_1, P2_1, 2_1$	4
23	$I222$	T_I	4	$P2_1, P2_1, 2_1$	4
75	$P4$	T_P	4	$P2_1, P4_1, P4_3$	4
79	$I4$	T_I	4	$P2_1, P4_1, P4_3$	4
89	$P422$	T_P	8	$P2_1, P4_1, P4_3, P2_1, 2_1$	8
97	$I422$	T_I	8	$P2_1, P4_1, P4_3, P2_1, 2_1$	8
143	$P3$	T_P	3	$P3_1, P3_2$	3
146	$R3$	T_R	3	$P3_1, P3_2$	3
149	$P312$	T_P	6	$P2_1, P3_1, P3_2$	6
150	$P321$	T_P	6	$P2_1, P3_1, P3_2$	6
155	$R32$	T_R	6	$P2_1, P3_1, P3_2$	6
168	$P6$	T_P	6	$P2_1, P3_1, P3_2, P6_1, P6_5$	6
177	$P622$	T_P	12	$P2_1, P3_1, P3_2, P6_1, P6_5, P2_1, 2_1$	12
195	$P23$	T_P	12	$P2_1, P3_1, P3_2$	12
196	$F23$	T_F	12	$P2_1, P3_1, P3_2, P2_1, 2_1$	12
197	$I23$	T_I	12	$P2_1, P3_1, P3_2, P2_1, 2_1$	12
207	$P432$	T_P	24	$P2_1, P3_1, P3_2, P2_1, 2_1, P4_1, P4_3$	24
209	$F432$	T_F	24	$P2_1, P3_1, P3_2, P4_1, P4_3, P2_1, 2_1$	24
211	$I432$	T_I	24	$P2_1, P3_1, P3_2, P4_1, P4_3, P2_1, 2_1$	24

lattice \mathbb{L} of a given crystal, by definition these vectors have the shortest length possible. The conjugation by an affine transformation $\alpha^{-1} = (A^{-1}, -A^{-1}\mathbf{a})$ gives $\alpha^{-1} \circ t \circ \alpha = (I, A^{-1}\mathbf{Bz})$. The constraint that $A^{-1}\mathbf{Bz} \in \mathbb{L}$ for all possible $\mathbf{z} \in \mathbb{Z}^3$ forces A^{-1} to have integral coefficients, and all of the eigenvalues of $A^{-T}A^{-1}$ have the property $\lambda(A^{-T}A^{-1}) \geq 1$. Then, from the elementary properties of the Euclidean norm,

$$\|A^{-1}\mathbf{Bz}\| \geq [\lambda_{\min}(A^{-T}A^{-1})]^{1/2} \|\mathbf{Bz}\| \geq \|\mathbf{Bz}\|,$$

indicating that affine transformation cannot produce a finer lattice. \square

If a nontrivial Bieberbach group ($\Gamma_B^\alpha \neq T$) exists with $\alpha \in \text{Aff}^+(3)$ and is minimal-index according to equation (3) amongst all $\Gamma_{B'} \in \mathcal{F}$, then in addition to being denoted as $\Gamma_{B_M}^\alpha$, it is called a minimal-index nontrivial Bieberbach subgroup of Γ .

In Table 1 the indices of minimal-index nontrivial Bieberbach subgroups of symmorphic Sohncke groups are computed and compared with the minimal-index trivial ones, $[\Gamma : T] = |\mathbb{P}|$. This is broken down into the symmorphic and nonsymmorphic cases. Note that from Lagrange's theorem and one of the isomorphism theorems, in the case of nonprimitive settings $[\Gamma_N : T_N] = [\Gamma_N : T_P]/[T_N : T_P]$.

The affine transformations α that can cause $\alpha\Gamma_{B_M}^\alpha$ to be contained in Γ can be found easily by using the 'MAXSUB' function (Aroyo *et al.*, 2006) in the Bilbao Crystallographic Server (Kroumova *et al.*, 2003).

Table 2
The nonsymmorphic case.

Note that in this table we did not consider the normality of subgroups, but only the relative size of quotients.

International No.	Γ	T	$[\Gamma : T]$	Γ_{B_M}	$[\Gamma : (\Gamma_{B_M}^\alpha)^\alpha]$
4	$P2_1$	T_P	2	$P2_1$	1
17	$P222_1$	T_P	4	$P2_1$	2
18	$P2_1, 2_1, 2_1$	T_P	4	$P2_1, P2_1, 2_1, 2_1$	2
19	$P2_1, 2_1, 2_1$	T_P	4	$P2_1, 2_1, 2_1$	1
20	$C222_1$	T_C	4	$P2_1, P2_1, 2_1, 2_1$	2
24	$I2_1, 2_1, 2_1$	T_I	4	$P2_1, 2_1, 2_1$	2
76	$P4_1$	T_P	4	$P4_1$	1
77	$P4_2$	T_P	4	$P4_1, P4_3$	2
78	$P4_3$	T_P	4	$P4_3$	1
80	$I4_1$	T_I	4	$P4_1, P4_3$	2
90	$P4_2, 2_1$	T_P	8	$P2_1, P2_1, 2_1, 2_1$	4
91	$P4_1, 2_1, 2_1$	T_P	8	$P4_1$	2
92	$P4_1, 2_1, 2_1$	T_P	8	$P4_1, P2_1, 2_1, 2_1$	2
93	$P4_2, 2_1, 2_1$	T_P	8	$P4_1, P4_3$	4
94	$P4_2, 2_1, 2_1$	T_P	8	$P2_1, P4_1, P4_3, P2_1, 2_1, 2_1$	4
95	$P4_3, 2_1, 2_1$	T_P	8	$P4_3$	2
96	$P4_3, 2_1, 2_1$	T_P	8	$P4_3, P2_1, 2_1, 2_1$	2
98	$I4_1, 2_1, 2_1$	T_I	8	$P4_1, P4_3, P2_1, 2_1, 2_1$	4
144	$P3_1$	T_P	3	$P3_1$	1
145	$P3_2$	T_P	3	$P3_2$	1
151	$P3_1, 1_2$	T_P	6	$P3_1$	2
152	$P3_1, 2_1$	T_P	6	$P3_1$	2
153	$P3_2, 1_2$	T_P	6	$P3_2$	2
154	$P3_2, 2_1$	T_P	6	$P3_2$	2
169	$P6_1$	T_P	6	$P6_1$	1
170	$P6_5$	T_P	6	$P6_5$	1
171	$P6_2$	T_P	6	$P3_2, P6_1$	2
172	$P6_4$	T_P	6	$P3_1, P6_5$	2
173	$P6_3$	T_P	6	$P2_1, P6_1, P6_5$	3
178	$P6_1, 2_1, 2_1$	T_P	12	$P6_1$	2
179	$P6_5, 2_1, 2_1$	T_P	12	$P6_5$	2
180	$P6_2, 2_1, 2_1$	T_P	12	$P3_2, P6_1$	4
181	$P6_1, 2_1, 2_1$	T_P	12	$P3_1, P6_5$	4
182	$P6_3, 2_1, 2_1$	T_P	12	$P2_1, P6_1, P6_5$	6
198	$P2_1, 3$	T_P	12	$P2_1, 2_1, 2_1$	3
199	$I2_1, 3$	T_I	24	$P2_1, 2_1, 2_1$	6
208	$P4_2, 3_2$	T_P	24	$P4_1, P4_3$	12
210	$F4_1, 3_2$	T_F	24	$P4_1, P4_3, P2_1, 2_1, 2_1$	12
212	$P4_3, 3_2$	T_P	24	$P4_3, P2_1, 2_1, 2_1$	6
213	$P4_1, 3_2$	T_P	24	$P4_1, P2_1, 2_1, 2_1$	6
214	$I4_1, 3_2$	T_I	24	$P4_1, P4_3, P2_1, 2_1, 2_1$	12

From Table 1, we conclude the following:

Remark 2.2. For every symmorphic space group, T is a minimal-index Bieberbach subgroup (*i.e.*, not only minimal-index in the trivial subfamily $\{P1\}$ but also in the entire family \mathcal{F}). This follows by observing that $[\Gamma : T] \leq [\Gamma : \Gamma_{B_M}^\alpha]$ in every column of Table 1.

This situation is different for the nonsymmorphic Sohncke groups, as can be observed in Table 2, and the theorem that results from observing this table. Here it is possible for the minimal-index Bieberbach subgroup not to have T as a subgroup.

Theorem 2.3. In every nonsymmorphic space group a minimal-index Bieberbach group is never the trivial one.

Proof. $[\Gamma : \Gamma_{B_M}^\alpha] < [\Gamma : T]$ in each entry in Table 2. \square

3. Minimal-index symmorphic subgroups

Let T be the minimal-index translation subgroup of the space group Γ and let Γ/T denote the factor group with \mathbb{P} denoting the corresponding abstract point group. We can construct a fundamental domain $F_{\Gamma/T} \subset \Gamma$ with $|\mathbb{P}|$ distinguished coset representatives chosen with the smallest possible translational part. Recall that $\Gamma/T \cong \mathbb{P}$, but in general $F_{\Gamma/T} \neq \mathbb{P}$, with equality only possible in the symmorphic case, and this only after removing the zero translation vectors from the representatives in $F_{\Gamma/T}$.

Corresponding to each space group Γ with point group \mathbb{P} , there is another space group $\Gamma_p = \mathbb{P} \ltimes T$ that will only be equal to Γ when Γ is symmorphic. Γ_p contains the proper subgroup $\mathbb{P} \times \{\mathbf{0}\} = \mathbb{P} \ltimes \{\mathbf{0}\}$, which in crystallography is identified with \mathbb{P} itself, and so one writes $\mathbb{P} \ltimes \{\mathbf{0}\} < \Gamma_p$. Though this is not true for nonsymmorphic Γ , nevertheless $\Gamma < G < \Gamma_p$, and since the intersection of subgroups is a subgroup, $\Gamma \cap \Gamma_p < G$.

Definition 3.1. Let Γ_{S_M} be a symmorphic space group such that for some $\beta \in \text{Aff}^+(3)$ it holds that for the nonsymmorphic space group Γ we have $\Gamma_{S_M}^\beta < \Gamma$ and $[\Gamma : \Gamma_{S_M}^\beta]$ is minimized over all such Γ_{S_M} . Then $\Gamma_{S_M}^\beta$ is called a minimal-index symmorphic subgroup of type Γ_{S_M} .

$\Gamma_{S_M}^\beta$ can be obtained easily using the Bilbao Server or *International Tables for Crystallography* and checking the index of each symmorphic subgroup. For every Sohncke group we have tabulated the minimal-index symmorphic subgroup (Table 3). For the symmorphic space groups, the minimal-index symmorphic subgroup is the group itself. For nonsymmorphic space groups, the minimal-index symmorphic subgroups are given. Note that minimality of the index does not imply uniqueness or normality.

For each entry in Table 3, there exists some $\beta \in \text{Aff}^+(3)$ such that Γ_{S_M} is a symmorphic space-group type with smallest index $[\Gamma : (\Gamma_{S_M})^\beta]$. But there is no *a priori* guarantee that T is a subgroup of $(\Gamma_{S_M})^\beta$.

4. Normal Bieberbach and symmorphic subgroups

This section characterizes normal Bieberbach and symmorphic subgroups of space groups. Those that are normal are not always minimal-index, and those that are minimal-index are not always normal. We use heavily the following general results.

Lemma 4.1. (Senechal, 1985) Let Γ be a space group with space subgroups N and H , such that $N \trianglelefteq \Gamma$ and $N \leq H \leq \Gamma$, then

$$\frac{H}{N} \trianglelefteq \frac{\Gamma}{N} \iff H \trianglelefteq \Gamma.$$

This can be used to prove the following theorem.

⁶ We use ‘a’ rather than ‘the’ here, because in some cases it is possible to find other symmorphic subgroups with the same minimal value of index.

Table 3

Minimal-index symmorphic subgroups for nonsymmorphic space groups.

International No.	Γ	T	$[\Gamma : T]$	Γ_{S_M}	$[\Gamma : (\Gamma_{S_M})^\beta]$
4	$P2_1$	T_P	2	$P1$	2
17	$P222_1$	T_P	4	$P2$	2
18	$P2_12_12$	T_P	4	$P2$	2
19	$P2_12_12_1$	T_P	4	$P1$	4
20	$C222_1$	T_C	4	$C2$	2
24	$I2_12_12_1$	T_I	4	$C2$	2
76	$P4_1$	T_P	4	$P1$	4
77	$P4_2$	T_P	4	$P2$	2
78	$P4_3$	T_P	4	$P1$	4
80	$I4_1$	T_I	4	$C2$	2
90	$P42_12$	T_P	8	$P4, C222$	2
91	$P4_122$	T_P	8	$P2, C2$	4
92	$P4_12_12$	T_P	8	$C2$	4
93	$P4_222$	T_P	8	$P222, C222$	2
94	$P4_22_12$	T_P	8	$C222$	2
95	$P4_322$	T_P	8	$P2, C2$	4
96	$P4_32_12$	T_P	8	$C2$	4
98	$I4_122$	T_I	8	$F222$	2
144	$P3_1$	T_P	3	$P1$	3
145	$P3_2$	T_P	3	$P1$	3
151	$P3_112$	T_P	6	$C2$	3
152	$P3_121$	T_P	6	$C2$	3
153	$P3_112$	T_P	6	$C2$	3
154	$P3_212$	T_P	6	$C2$	3
169	$P6_1$	T_P	6	$P1$	6
170	$P6_5$	T_P	6	$P1$	6
171	$P6_2$	T_P	6	$P2$	3
172	$P6_4$	T_P	6	$P2$	3
173	$P6_3$	T_P	6	$P3$	2
178	$P6_122$	T_P	12	$C2$	6
179	$P6_222$	T_P	12	$C2$	6
180	$P6_22_12$	T_P	12	$C222$	3
181	$P6_122$	T_P	12	$C222$	3
182	$P6_322$	T_P	12	$P321, P312$	2
198	$P2_13$	T_P	12	$R3$	4
199	$I2_13$	T_I	24	$R3$	4
208	$P4_232$	T_P	24	$P23$	2
210	$F4_132$	T_F	24	$F23$	2
212	$P4_332$	T_P	24	$R32$	4
213	$P4_132$	T_P	24	$R32$	4
214	$I4_132$	T_I	24	$R32$	4

Theorem 4.2. Let Γ be a space group with point group \mathbb{P} and let Γ_{S_M} be a minimal-index symmorphic subgroup of Γ . Then, if $T < \Gamma_{S_M}$ and if $\mathbb{S} \leq \mathbb{S}_M \leq \mathbb{P} \cong \Gamma/T$, then

$$\mathbb{S} \trianglelefteq \Gamma/T \iff \mathbb{S} \ltimes T \trianglelefteq \Gamma$$

regardless of whether Γ is symmorphic or nonsymmorphic.

Proof. It follows directly from the definition of $\mathbb{S}_M \ltimes \mathbf{0} = F_{\Gamma/T} \cap (\mathbb{P} \ltimes \mathbf{0})$ and Lemma 4.1. \square

We now provide some examples to illustrate these concepts.

Example 4.1. For any symmorphic space group, $\mathbb{S}_M = \mathbb{P}$, setting $\mathbb{S} = \mathbb{S}_M = \mathbb{P}$ satisfies the theorem trivially, and in this case $\Gamma = \mathbb{P} \ltimes T$.

Example 4.2. If $\Gamma_S = \mathbb{S} \ltimes T < \Gamma$ and $[\Gamma : \Gamma_S] = 2$, then $\Gamma_S \trianglelefteq \Gamma$.

This is a very common case, but not every normal symmorphic subgroup falls into the categories of Examples 4.1 and 4.2. The next two examples illustrate other cases.

Example 4.3. Consider the nonsymmorphic space group No. 171, $P6_2$. When $\Gamma = P6_2$ and $T = P1$, then

$$\frac{\Gamma}{T} = \left\{ \begin{array}{ll} (x, y, z); & (-y, x - y, z + 2/3); \\ (-x + y, -x, z + 1/3); & (-x, -y, z); \\ (y, -x + y, z + 2/3); & (x - y, x, z + 1/3) \end{array} \right\}.$$

The subgroup

$$\mathbb{S} = \{(x, y, z); (-x, -y, z)\},$$

which is isomorphic to $P2/P1$, is normal in Γ/T , and so $P2 \triangleleft P6_2$. In this case $P2$ is both the minimal-index symmorphic subgroup and it is normal, but $[P6_2 : P2] = 3$.

Example 4.4. Consider the nonsymmorphic space group No. 180, $P6_22$. When $\Gamma = P6_22$ and $T = P1$, then

$$\frac{\Gamma}{T} = \left\{ \begin{array}{ll} (x, y, z); & (-y, x - y, z + 2/3); \\ (-x + y, -x, z + 1/3); & (-x, -y, z); \\ (y, -x + y, z + 2/3); & (x - y, x, z + 1/3); \\ (y, x, -z + 2/3); & (x - y, -y, -z); \\ (-x, -x + y, -z + 1/3); & (-y, -x, -z + 2/3); \\ (-x + y, y, -z); & (x, x - y, -z + 1/3) \end{array} \right\}.$$

Using the Bilbao Server, we find that an Aff^+ -conjugated version of $C222$ is the minimal-index symmorphic subgroup of $P6_22$, and has index 3. But this is not normal. However,

$$\mathbb{S} = \{(x, y, z); (-x, -y, z)\},$$

which is isomorphic to $P2/P1$, is normal in Γ/T and so $P2 \triangleleft P6_22$ even though $[P6_22 : P2] = 6$.

It is also worth mentioning that there are nonsymmorphic groups with relatively large symmorphic subgroups which are not normal. For example, $P4_322$ has an Aff^+ -conjugated version of $R32$ as a minimal-index symmorphic subgroup of index 4, but this is not normal. In fact, the only symmorphic subgroup of $P4_322$ that is normal is $P1$.

Similar examples can be constructed for Bieberbach subgroups. In the PDB it is often the case that a protein crystallizes in a Bieberbach group. And when $\Gamma \neq \Gamma_B^\alpha$ it is often the case that $[\Gamma : \Gamma_{B_M}^\alpha] = 2$, indicating both minimal-index and normality. But there are cases for which $[\Gamma : \Gamma_{B_M}^\alpha] \neq 2$ and yet $\Gamma_{B_M}^\alpha \triangleleft \Gamma$. One such case is given below.

Example 4.5. Again consider the nonsymmorphic space group No. 180, $P6_22$. Representatives of Γ/T were given in Example 4.4. Here $\Gamma_B = P3_2$ and $T = P1$, then

$$\frac{\Gamma_B}{T} = \{(x, y, z); (-y, x - y, z + 2/3); (-x + y, -x, z + 1/3)\}.$$

(In this example no conjugation by α is required.) It is not difficult to show that $\frac{\Gamma_B}{T} \triangleleft \frac{\Gamma}{T}$, and hence by Lemma 4.1 in this case $P3_2 \triangleleft P6_22$ even though $[P6_22 : P3_2] = 4$.

Finally, there are cases in which a nonsymmorphic Sohncke group has a nontrivial minimal-index Bieberbach subgroup which is not normal. For example, $P4_232$, which has $[P4_232 : P1] = 24$, has affine-conjugated versions of $P4_1$ and $P4_3$ as subgroups with index 12. But the only Bieberbach subgroup of $P4_232$ which is normal is $P1$.

The following theorem summarizes Table 4 regarding the normality of both Bieberbach and symmorphic subgroups of Sohncke groups.

Theorem 4.3. Every nontrivial Sohncke group, $\Gamma > T$, either contains a normal nontrivial Bieberbach subgroup ($\Gamma_B^\alpha \trianglelefteq \Gamma$), in which case $\Gamma/\Gamma_B^\alpha \cong \mathbb{S} < \mathbb{P}$, or it contains a normal nontrivial symmorphic subgroup ($\Gamma_S^\beta \trianglelefteq \Gamma$), where $\Gamma_S^\beta \leq \mathbb{S} \ltimes T$ and $\mathbb{S} \ltimes \mathbf{0} \leq (\mathbb{P} \ltimes \mathbf{0}) \cap F_{\Gamma/T}$, in which case $\mathbb{S} \trianglelefteq \mathbb{P}$ and $\Gamma/\Gamma_S^\beta \cong \Gamma_B^\alpha/T$. And these conditions are not mutually exclusive. Moreover, if $|\mathbb{P}|$ is not prime and if $\Gamma_B \neq \Gamma \neq \Gamma_S$ then \trianglelefteq can be replaced with \triangleleft .

Proof. By direct evaluation of each representative of the 65 classes of Sohncke groups we arrive at Table 4, which contains the information summarized in the above theorem. \square

In Table 4 only nontrivial Γ_B are listed. Γ' in place of Γ indicates that a nonrigid affine transformation was required to place it as a subgroup. Parentheses indicate that, though normal, the subgroup is either not minimal-index or not proper. Improper ones are listed on the left side of each column, and those that are not minimal-index are listed on the right.

From Table 4 we see that for each space group we can find either a nontrivial normal Γ_B or a nontrivial normal Γ_S . And in some cases we have both. And comparing this table with the tables for minimal-index symmorphic and minimal-index Bieberbach subgroups, we see that:

Theorem 4.4. Every nonsymmorphic space group except for No. 208, $P4_232$ and No. 210, $F4_132$ has a nontrivial minimal-index normal Bieberbach subgroup.

The indices of these normal subgroups are listed in Table 4 and can be compared with those of the minimal-index subgroups of each type given in the previous tables.

5. Macromolecular crystallography statistics

We tabulated the space groups of all 80 083 protein structures obtained from X-ray crystallography reported in the PDB as of 14 January 2014 and report these results in Table 5. The first column is the number of proteins in the PDB with a particular symmetry, followed by the percentage and the international name of the space group. The third column asks if the space group contains as normal subgroups conjugated versions of $P2_1$ or $P222$. The next two columns ask if the group is nonsymmorphic or Bieberbach. The final three columns compute the indices of three factor groups: the quotient of Γ by the minimal-index normal Bieberbach subgroup, Γ by the minimal-index normal symmorphic subgroup, and finally the index of the quotient by the largest translation subgroup.

Table 4
Nontrivial normal Bieberbach and symmorphic subgroups.

Group	$\alpha\Gamma_B\alpha^{-1} \trianglelefteq \Gamma$	$\beta\Gamma_S\beta^{-1} \trianglelefteq \Gamma$
(1, P_1)	(P_1)	(P_1)
(3, P_2)	P_1	(P_2)
(4, P_{2_1})	(P_{2_1})	P_1
(5, C_2)	T_C	(C_2)
(16, $P222$)	P_1	($P222$), P_2 , $P'222$, $C'222$, $F'222$
(17, $P222_1$)	P_{2_1}	P_2
(18, $P_{2_1}2_12_1$)	P_{2_1} , $P'2_12_12_1$	P_2
(19, $P_{2_1}2_12_1$)	($P_{2_1}2_12_1$), P_{2_1}	P_1
(20, $C222_1$)	$P'2_1$, $P_{2_1}2_12_1$	C_2
(21, $C222$)	T_C	($C222$), $C'222$, C_2 , $I'222$, $P222$, $P'2$
(22, $F222$)	T_F	($F222$), $C222$, $C'2$
(23, $I222$)	T_I	($I222$), $P222$, $C'2$
(24, $I_{2_1}2_12_1$)	$P_{2_1}2_12_1$	$C'2$
(75, P_4)	P_1	(P_4), P_2 , $P'4$, $I'4$
(76, P_{4_1})	(P_{4_1}), P_{2_1} , $P'4_1$	P_1
(77, P_{4_2})	$P'4_1$, $P'4_3$	P_2
(78, P_{4_3})	(P_{4_3}), P_{2_1} , $P'4_3$	P_1
(79, I_4)	T_I	(I_4), P_4 , $C'2$
(80, I_{4_1})	P_{4_1} , P_{4_3}	$C'2$
(89, $P422$)	P_1	($P422$), P_4 , $P222$, $C'222$, $P'422$, $I'422$, (P_2 , $I'4$, $C'222$, $P'222$, $P'4$, $P'4222$)
(90, $P_{4_2}2_1$)	P_{2_1}	P_4 , $C'222$, (P_2 , $P222$, $C'222$, $I'222$, $P'4$)
(91, $P_{4_1}2_12_1$)	P_{4_1} , (P_{2_1} , $P'2_12_12_1$, $P'4_1$)	P_1
(92, $P_{4_1}2_12_1$)	P_{4_1} , $P_{2_1}2_12_1$, (P_{2_1} , $P'2_12_12_1$)	P_1
(93, $P_{4_2}2_12_1$)	$P'4_1$, $P'4_3$	$P222$, $C'222$, (P_2 , $F'222$)
(94, $P_{4_2}2_12_1$)	$P'4_1$, $P'4_3$, $P'2_12_12_1$	$C'222$, (P_2 , $P'222$)
(95, $P_{4_3}2_12_1$)	P_{4_3} , (P_{2_1} , $P'2_12_12_1$, $P'4_3$)	P_1
(96, $P_{4_3}2_12_1$)	P_{4_3} , $P_{2_1}2_12_1$, (P_{2_1} , $P'2_12_12_1$)	P_1
(97, $I422$)	T_I	($I422$), $P422$, I_4 , $I222$, $F'222$, (P_4 , $P222$, $C'222$, $C'2$)
(98, $I_42_12_1$)	$P_{2_1}2_12_1$, P_{4_1} , P_{4_3}	$F'222$, ($C'2$)
(143, P_3)	P_1	(P_3)
(144, P_{3_1})	(P_{3_1})	P_1
(145, P_{3_2})	(P_{3_2})	P_1
(146, R_3)	T_R	(R_3)
(149, $P312$)	P_1	($P312$), P_3 , $P'312$
(150, $P321$)	P_1	($P321$), P_3 , $P'321$
(151, $P_{3_1}12$)	P_{3_1}	P_1
(152, $P_{3_1}21$)	P_{3_1}	P_1
(153, $P_{3_2}12$)	P_{3_2}	P_1
(154, $P_{3_2}21$)	P_{3_2}	P_1
(155, $R32$)	T_R	($R32$), R_3 , $R'32$, ($P321$)
(168, P_6)	P_1	(P_6), P_3 , $P'6$, (P_2)
(169, P_{6_1})	(P_{6_1}), P_{3_1} , (P_{2_1})	P_1
(170, P_{6_5})	(P_{6_5}), P_{3_2} , (P_{2_1})	P_1
(171, P_{6_2})	P_{3_2} , $P'6_1$	P_2
(172, P_{6_4})	P_{3_1} , $P'6_5$	P_2
(173, P_{6_3})	P_{2_1} , $P'6_1$, $P'6_5$	P_3
(177, $P622$)	P_1	($P622$), P_6 , $P312$, $P321$, $P'622$, (P_2 , P_3)
(178, $P_{6_1}2_12_1$)	P_{6_1} , (P_{3_1} , P_{2_1})	P_1
(179, $P_{6_5}2_12_1$)	P_{6_5} , (P_{3_2} , P_{2_1})	P_1
(180, $P_{6_2}2_12_1$)	P_{3_2} , $P'6_1$	P_2
(181, $P_{6_4}2_12_1$)	P_{3_1} , $P'6_5$	P_2
(182, $P_{6_3}2_12_1$)	$P'6_1$, $P'6_5$, P_{2_1}	$P312$, $P321$, (P_3)
(195, P_{23})	P_1	(P_{23}), $F'23$, ($P222$)
(196, F_{23})	T_F	(F_{23}), $F222$, (P_{23})
(197, I_{23})	T_I	(I_{23}), P_{23} , ($I222$, $F'23$, $P222$)
(198, $P_{2_1}2_13$)	$P_{2_1}2_12_1$	P_1
(199, $I_{2_1}2_13$)	$P_{2_1}2_12_1$	T_I
(207, $P432$)	P_1	($P432$), P_{23} , $P'432$, $F'432$, ($P222$, $F'23$, $F'222$, $I'222$)
(208, $P_{4_3}3_2$)	P_1	P_{23} , ($F'23$, $F'222$, $P222$, $C'222$, $I'23$)
(209, F_{432})	T_F	(F_{432}), F_{23} , ($F222$)
(210, $F_{4_3}3_2$)	T_F	F_{23} , ($F222$)
(211, I_{432})	T_I	(I_{432}), P_{432} , I_{23} , ($I222$, P_{23} , $P222$, F_{23})
(212, $P_{4_3}3_2$)	$P_{2_1}2_12_1$	P_1
(213, $P_{4_1}3_2$)	$P_{2_1}2_12_1$	P_1
(214, $I_{4_1}3_2$)	$P_{2_1}2_12_1$	T_I

Nucleic acid structures were not included in this analysis. From Table 5, we make the following observations:

Observation 5.1. Although Bieberbach groups comprise less than 14% of the Sohncke groups (9 of 65), 47.18% of proteins crystallize with Bieberbach-group symmetry (including the trivial P_1 Bieberbach group).

Observation 5.2. More than 99.83% of proteins crystallize in a space group containing a nontrivial normal Bieberbach subgroup.

Observation 5.3. Space groups containing P_{2_1} subgroups are favorable, constituting 64.66% of the total, and those that contain normal $P222$ subgroups are generally unfavorable, constituting 5.59% of the total.

Fifteen of 80 083 proteins with nonstandard names ‘ $I -4_1$ ’, ‘ $I -4C2$ ’, ‘ $P 12_1/N1$ ’, ‘ $P 12_1/C1$ ’, ‘ $I 4_1/A$ ’ and ‘ $P -1$ ’ were excluded from this analysis. Additional proteins reported in the PDB with nonstandard names were renamed and counted according to Table 6.

Note also that if each of the 11 enantiomorphic pairs (P_{3_1} , P_{3_2}), (P_{4_1} , P_{4_3}), (P_{6_1} , P_{6_5}), ($P_{3_2}12$, $P_{3_1}12$), ($P_{3_2}21$, $P_{3_1}21$), ($P_{4_3}22$, $P_{4_1}22$), ($P_{4_3}2_12$, $P_{4_1}2_12$), ($P_{4_3}3_2$, $P_{4_1}3_2$), ($P_{6_1}22$, $P_{6_5}22$), (P_{6_2} , P_{6_4}), ($P_{6_2}22$, $P_{6_4}22$) are considered equivalent and their numbers are pooled, then this will also alter the statistics toward an even greater preference of proteins toward Bieberbach groups and nonsymmorphic groups. In other words, pooling gives Table 7.

Here the top 16 classes are chosen, which cumulatively cover 89.56% of all proteins in the PDB. These percentages are consistent with those reported earlier on smaller data sets in Wukovitz & Yeates (1995) and Rupp (2010).

A natural question to ask is ‘Why do proteins in the PDB appear to favor Bieberbach groups so heavily?’ Different variants of this question have been asked previously using both the full PDB (as we have done), and with specialized subsets of nonredundant structures (Allen, 2002; Pidcock *et al.*, 2003; Filippini & Gavezzotti, 1992; Padmaja *et al.*, 1990). In all cases, space-group frequencies show a strong preference for $P_{2_1}2_12_1$ and P_{2_1} as we have observed.

Three kinds of arguments have been put forth in the past in an attempt to explain these. The first is concerned with the solvent content in a crystal, which is a topic first analyzed in Matthews (1968). See for example Chruszcz *et al.* (2008) and Weichenberger & Rupp (2014) for a discussion of these arguments. The second kind of argument is concerned with the

Table 5
Space groups of all 80 083 protein structures in the PDB as of 14 January 2014.

No.	%	Name	International No.	P_{21}, P_{222}	NS	B	$[\Gamma : (\Gamma_B)^\alpha]$	$[\Gamma : (\Gamma_S)^\beta]$	$ \frac{\Gamma}{\Gamma} $
12	0.01	P_{222}	16	N,Y	N	N	4	1	4
14	0.02	P_{312}	149	N,N	N	N	6	1	6
22	0.03	P_{4_32}	208	N,Y	Y	N	24	2	24
35	0.04	P_{422}	89	N,Y	N	N	8	1	8
39	0.05	P_{432}	207	N,Y	N	N	24	1	24
46	0.06	P_{23}	195	N,Y	N	N	12	1	12
50	0.06	P_{4_22}	93	N,Y	Y	N	4	2	8
63	0.08	F_{4_32}	210	N,N	Y	N	12	2	24
73	0.09	P_{3_12}	151	N,N	Y	N	2	6	6
74	0.09	P_{622}	177	N,Y	N	N	12	1	12
75	0.09	F_{23}	196	N,N	N	N	12	1	12
76	0.09	P_{3_212}	153	N,N	Y	N	2	6	6
77	0.10	P_{4_2}	77	N,N	Y	N	2	4	4
79	0.10	I_{4_32}	214	N,N	Y	N	12	24	24
79	0.10	P_{4_32}	212	Y,N	Y	N	6	24	24
88	0.11	P_4	75	N,N	Y	N	4	1	4
100	0.12	I_{432}	211	N,Y	N	N	24	1	24
102	0.13	F_{222}	22	N,Y	N	N	4	1	4
114	0.14	P_{222_1}	17	Y,N	Y	N	2	2	4
120	0.15	P_{4_32}	213	Y,N	Y	N	6	24	24
151	0.19	P_{6_4}	172	N,N	Y	N	2	3	6
152	0.19	P_3	143	N,N	N	N	3	1	3
164	0.20	C_{222}	21	N,Y	N	N	4	1	4
171	0.21	P_{6_2}	172	N,N	Y	N	2	3	6
172	0.21	F_{432}	209	N,Y	N	N	24	1	24
175	0.22	P_{4_22}	91	Y,N	Y	N	2	8	8
187	0.23	$I_{2_12_12_1}$	24	N,N	Y	N	2	2	4
190	0.24	P_2	3	N,N	N	N	2	1	2
206	0.26	P_6	168	N,N	N	N	6	1	6
216	0.27	P_{4_322}	95	Y,N	Y	N	2	8	8
230	0.29	I_{2_13}	199	N,N	Y	N	6	12	12
242	0.30	I_4	80	N,N	Y	N	2	2	4
255	0.32	P_{6_22}	181	N,Y	Y	N	4	6	12
291	0.36	P_{6_22}	180	N,Y	Y	N	4	6	12
325	0.41	P_{321}	150	N,N	N	N	6	1	6
326	0.41	I_{23}	197	N,Y	N	N	12	1	12
342	0.43	P_{42_12}	90	Y,N	Y	N	4	2	8
454	0.57	I_4	79	N,N	N	N	4	1	4
467	0.58	P_{2_13}	198	Y,N	Y	N	3	12	12
477	0.60	P_{6_322}	182	Y,N	Y	N	6	2	12
495	0.62	P_{4_3}	78	Y,N	Y	Y	1	4	4
498	0.62	P_{3_1}	144	N,N	Y	Y	1	3	3
501	0.63	I_{4_22}	98	N,N	Y	N	4	2	8
507	0.63	P_{3_2}	145	N,N	Y	Y	1	3	3
521	0.65	$P_{4_22_12}$	94	Y,Y	Y	N	4	4	8
526	0.66	I_{422}	97	N,Y	N	N	8	1	8
603	0.75	P_{6_3}	173	Y,N	Y	N	3	2	6
623	0.78	P_{4_1}	76	Y,N	Y	Y	1	4	4
843	1.05	P_{6_5}	170	Y,N	Y	Y	1	6	6
863	1.08	P_{6_1}	169	Y,N	Y	Y	1	6	6
986	1.23	P_{6_522}	179	Y,N	Y	N	2	12	12
1152	1.44	R_{32}	155	N,N	N	N	6	1	6
1193	1.49	R_3	146	N,N	N	N	3	1	3
1361	1.70	P_{6_22}	178	Y,N	Y	N	2	12	12
1740	2.17	I_{222}	23	N,Y	N	N	4	1	4
2521	3.15	$P_{4_22_12}$	92	Y,N	Y	N	2	8	8
2605	3.25	P_{3_121}	152	N,N	Y	N	2	6	6
2799	3.50	P_{3_221}	154	N,N	Y	N	2	6	6
3013	3.76	$P_{4_32_12}$	96	Y,N	Y	N	2	8	8
3058	3.82	P_1	1	N,N	N	Y	1	1	1
3883	4.85	C_{222_1}	20	Y,N	Y	N	2	2	4
4216	5.26	$P_{2_12_12}$	18	Y,N	Y	N	2	2	4
7876	9.83	C_2	5	N,N	N	N	2	1	2
12600	15.73	P_{2_1}	4	Y,N	Y	Y	1	2	2
18297	22.85	$P_{2_12_12_1}$	19	Y,N	Y	Y	1	4	4
80068 of 80083	99.98	Total							

interplay of molecular shape and the number of stabilizing contacts that form between proteins, as articulated in Wukovitz & Yeates (1995) and Andersson & Hovmöller (2000). The third is a purely geometric argument, one version of which can be found in Kitaev *et al.* (2003) and a more detailed version of which is given below. This argument centers around the fact that high-symmetry Wyckoff positions correspond to locations where bodies cannot be placed without colliding with other bodies.

Stated mathematically, the reason why Bieberbach groups are favored is that the quotients $\Gamma_B \backslash X$ are manifolds because the action of Γ_B on X has no fixed points. It is shown in Chirikjian & Shiffman (2014) in the planar case that for objects moving in concert with wallpaper symmetry groups $p2$ or $p3$ large regions of the space $p2 \backslash X$ or $p3 \backslash X$ (where here $X = \mathbb{R}^2$) correspond to bodies being in collision. These ‘collision zones’ are centered on the special positions of the wallpaper groups $p2$ and $p3$. That paper shows that even for symmetry-related bodies that are very small in comparison to lattice distances, large percentages of the space $\Gamma \backslash X$ (viewed as the translation part of $\Gamma \backslash G$) are inaccessible as they correspond to bodies in collision.

Since Bieberbach groups do not have fixed points (and hence no special positions), the collision zones corresponding to non-Bieberbach space groups with fixed points are nonexistent for Bieberbach space groups. This means that all other things being equal, there is ‘more room to move’ for a body in $\Gamma_B \backslash X$ (under the quasigroup action of $\Gamma_B \backslash G$) than there is in a general $\Gamma \backslash X$ (under the quasigroup action of $\Gamma \backslash G$). This is why we said in the introduction that the problem of protein preferences and the cost of a molecular-replacement search are dual to each other in the sense that the bigger the unhindered search space is, the more likely it is that bodies can find lower-energy packing arrangements, and simultaneously, a molecular-replacement search must cover more space that cannot be eliminated by simply removing from consideration collision zones.

This mathematical observation relates to both the arguments of Rupp (2010), since more room to move means more

Table 6
Mapping of nonstandard names to standard names.

No.	Nonstandard name	Standard name
1	A1	P1
1	B112	C2
1	C21	C2
2	A2	C2
5	C12 ₁ 1	C2
5	I12 ₁ 1	C2
9	I2 ₁	C2
94	I121	C2
7759	C121	C2
35	B2	P2
155	P121	P2
11	P112 ₁	P2 ₁
12589	P12 ₁ 1	P2 ₁
1145	H32	R32
1185	H3	R3
4	P2 ₁ 2 ₁ 2A	P2 ₁ 2 ₁ 2 ₁

entropy is allowed, and to Wukovitz & Yeates (1995), who argued that the preferred space groups are those that maximize the number of inter-body contacts. Our argument, which is akin to that of Kitaev *et al.* (2003), is simply that if as a crystal forms the first few proteins manage to dance around each other with Bieberbach symmetry, they will have more freedom to find stabilizing contacts that lead to forming a crystal. This also means that in the molecular-replacement problem, there will be more space that must be searched.

6. Conclusions

A new view of nonsymmorphic Sohncke groups as being a hybrid of symmorphic and Bieberbach subgroups is introduced here. It is shown from the relative frequency of space groups observed in the Protein Data Bank that protein crystals have a strong preference for space groups containing nontrivial normal Bieberbach subgroups, and a strong dislike for space groups with large symmorphic subgroups. In the next paper in this series we will examine how the decomposition of nonsymmorphic space groups into symmorphic and Bieberbach subgroups influences the choice of fundamental domain for $\Gamma \backslash G$ in the molecular-replacement problem.

Acknowledgements

We thank Mr Joshua Davis and Mr Qianli Ma for checking Table 5 in §5, and Professor Massimo Nespolo for his assistance in relating nonstandard space-group names to standard ones.

References

Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
 Andersson, K. M. & Hovmöller, S. (2000). *Acta Cryst.* **D56**, 789–790.
 Aroyo, M., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Z. Kristallogr.* **221**, 15–27.
 Berman, H. M. (2008). *Acta Cryst.* **A64**, 88–95.
 Berman, H. M. *et al.* (2002). *Acta Cryst.* **D58**, 899–907.
 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
 Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice,

Table 7
Most-common space groups with enantiomorphic pairs grouped.

No.	%	Name
526	0.66	I422
1005	1.25	P3 ₁ , P3 ₂
1152	1.44	R32
1193	1.49	R3
1188	1.48	P4 ₁ , P4 ₃
1706	2.13	P6 ₁ , P6 ₅
1740	2.17	I222
2346	2.93	P6 ₁ 22, P6 ₅ 22
3058	3.82	P1
3883	4.85	C222 ₁
4216	5.26	P2 ₁ 2 ₁ 2
5404	6.75	P3 ₂ 21, P3 ₁ 21
5534	6.91	P4 ₃ 2 ₁ 2, P4 ₁ 2 ₁ 2
7876	9.83	C2
12600	15.73	P2 ₁
18297	22.85	P2 ₁ 2 ₁ 2 ₁
71724	89.56	Total

M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *Eur. J. Biochem.* **80**, 319–324.
 Chirikjian, G. S. (2011). *Acta Cryst.* **A67**, 435–446.
 Chirikjian, G. S. & Shiffman, B. (2014). *Collision-Free Configuration-Spaces in Macromolecular Crystals, Workshop on Robotics Methods for Structural and Dynamic Modeling of Molecular Systems*, Robotics Science and Systems Conference, Berkeley, California, July 12, 2014. <https://cs.unm.edu/amprg/rss14workshop/PAPERS/Chirikjian.pdf>.
 Chirikjian, G. S. & Yan, Y. (2012). *Acta Cryst.* **A68**, 208–221.
 Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P. & Minor, W. (2008). *Protein Sci.* **17**, 623–632.
 Filippini, G. & Gavezzotti, A. (1992). *Acta Cryst.* **B48**, 230–234.
 Hahn, Th. (2002). Editor. *International Tables for Crystallography, Vol. A, Space-Group Symmetry: Brief Teaching Edition*. Dordrecht: Kluwer.
 Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* **D55**, 484–491.
 Kitaev, Yu. E., Panfilov, A. G., Smirnov, V. P. & Tronc, P. (2003). *Phys. Rev. E*, **67**, 011907.
 Kroumova, E., Capillas, C., Aroyo, M. I., Perez-Mato, J. M., Ivantchev, S., Madariaga, G., Kirov, A., Wondratschek, H., Stokes, H. & Hatch, D. (2003). *IOP Conf. Ser.* **173**, 383–386.
 Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
 Montesinos, J. M. (1987). *Classical Tessellations and Three-Manifolds*. Berlin: Springer-Verlag.
 Padmaja, N., Ramakumar, S. & Viswamitra, M. A. (1990). *Acta Cryst.* **A46**, 725–730.
 Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2003). *Acta Cryst.* **B59**, 634–640.
 Rossmann, M. G. (2001). *Acta Cryst.* **D57**, 1360–1366.
 Rupp, B. (2010). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. New York: Garland Science, Taylor and Francis Group.
 Senechal, M. (1985). *J. Math. Phys.* **26**, 219–228.
 Szczepański, A. (2012). *Geometry of Crystallographic Groups*. Singapore: World Scientific Publishing Company.
 Thurston, W. P. (1997). *Three-Dimensional Geometry and Topology*, edited by S. Levy. Princeton University Press.
 Weichenberger, C. X. & Rupp, B. (2014). *Acta Cryst.* **D70**, 1579–1588.
 Wolf, J. A. (2010). *Spaces of Constant Curvature*, 6th ed. Providence, RI: AMS Chelsea Publishing.
 Wondratschek, H. & Müller, U. (2010). Editors. *International Tables for Crystallography, Vol. A1, Symmetry Relations Between Space Groups*, 2nd ed. Chichester: Wiley.
 Wukovitz, S. W. & Yeates, T. O. (1995). *Nat. Struct. Biol.* **2**, 1062–1067.