FOUNDATIONS
ADVANCES

ISSN: 2053-2733
journals.iucr.org/a

# Mathematical aspects of molecular replacement. V. Isolating feasible regions in motion spaces

**Bernard Shiffman, Shengnan Lyu and Gregory S. Chirikjian**

*Acta Cryst.* (2020). A**76**, 145–162

Acta Cryst A FOUNDATIONS ADVANCES

ISSN 2053-2733

# Mathematical aspects of molecular replacement. V. Isolating feasible regions in motion spaces

**Bernard Shiffman,[a] Shengnan Lyu[b,c]* and Gregory S. Chirikjian[d,c]**

[a]Department of Mathematics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA, [b]School of Mechanical Engineering and Automation, Beihang University, Peope's Republic of China, [c]Department of Mechanical Engineering, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA, and [d]Department of Mechanical Engineering, National University of Singapore, Singapore. *Correspondence e-mail: lvshengnan5@gmail.com

Received 31 December 2018
Accepted 1 November 2019

Edited by H. Schenk, University of Amsterdam, The Netherlands

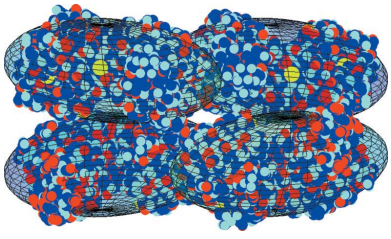
**Keywords:** molecular replacement; crystallographic symmetry; coset space; crystal topology; Minkowski sum; protein crystals; Laue groups; Euclidean normalizers; molecular coordination number; fundamental domain.

This paper mathematically characterizes the tiny feasible regions within the vast 6D rotation–translation space in a full molecular replacement (MR) search. The capability to *a priori* isolate such regions is potentially important for enhancing robustness and efficiency in computational phasing in macromolecular crystallography (MX). The previous four papers in this series have concentrated on the properties of the full configuration space of rigid bodies that move relative to each other with crystallographic symmetry constraints. In particular, it was shown that the configuration space of interest in this problem is the right-coset space $\Gamma\backslash G$, where $\Gamma$ is the space group of the chiral macromolecular crystal and $G$ is the group of rigid-body motions, and that fundamental domains $F_{\Gamma\backslash G}$ can be realized in many ways that have interesting algebraic and geometric properties. The cost function in MR methods can be viewed as a function on these fundamental domains. This, the fifth and final paper in this series, articulates the constraints that bodies packed with crystallographic symmetry must obey. It is shown that these constraints define a thin feasible set inside a motion space and that they fall into two categories: (i) the bodies must not interpenetrate, thereby excluding so-called 'collision zones' from consideration in MR searches; (ii) the bodies must be in contact with a sufficient number of neighbors so as to form a rigid network leading to a physically realizable crystal. In this paper, these constraints are applied using ellipsoidal proxies for proteins to bound the feasible regions. It is shown that the volume of these feasible regions is small relative to the total volume of the motion space, which justifies the use of ellipsoids as proxies for complex proteins in MR searches, and this is demonstrated with $P1$ (the simplest space group) and with $P2_12_12_1$ (the most common space group in MX).

## 1. Introduction and brief review of molecular replacement

Molecular replacement (MR) is an important computational method for addressing the phase problem in macromolecular crystallography. The method originated in the early 1960s (Rossmann & Blow, 1962), and has been refined over the years with many successful software packages including *AMORE* (Navaza, 1994), *Phaser* (McCoy *et al.*, 2007) *etc*. For a more complete review of the history of MR see Rossmann (2001). In these methods the search for the correct orientation and position of a candidate molecular shape is separated into a rotation search followed by a translation search. This separation works when one molecule at a time is placed in the crystallographic unit cell.

The previous papers in this series developed the concept of a motion space as the coupled 6D translation–rotation search space in MR by simultaneously placing one copy of a molecule



per asymmetric unit (Chirikjian, 2011; Chirikjian & Yan, 2012; Chirikjian, Sajjadi *et al.*, 2015; Chirikjian *et al.*, 2017). The present paper is the final one in this series, and studies the spaces of motions relevant for MR searches. These so-called 'motion spaces' can be viewed as regions (or fundamental domains) in the group $G = \mathrm{SE}(3)$ of rigid-body motions, or as the orbit space corresponding to the left action of a Sohncke space group $\Gamma$ on $G$. This fundamental domain contains one point drawn from each right coset $\Gamma g$, where $\Gamma$ is the space group of the macromolecular crystal and $g \in G$. From the perspective of an MR search, the closed fundamental domain, denoted here as $F_{\Gamma\backslash G}$, therefore is equivalent (up to an inconsequential set of measure zero resulting from taking the closure) to the right-coset space $\Gamma\backslash G$. In some contexts it is useful to think of $F_{\Gamma\backslash G}$ (rather than $\Gamma\backslash G$) as the motion space.

Previous papers in this series have characterized the algebraic (Chirikjian, 2011), geometric (Chirikjian & Yan, 2012) and measure-theoretic (Chirikjian *et al.*, 2017) properties of these spaces. These properties are related to the underlying structure of the crystallographic symmetry group in which the macromolecule of interest crystallizes. As such, new ways to characterize and decompose space groups have been undertaken in relation to this work (Chirikjian, Ratnayake *et al.*, 2015). Also relevant to the present discussion is the observation that macromolecular crystals predominantly prefer Bieberbach (torsion-free) space groups (Chirikjian, Sajjadi *et al.*, 2015), with almost one quarter of all protein crystals having $P2_12_12_1$ symmetry. Other papers not in the series also shed light on the problems addressed here, including Chirikjian & Shiffman (2016) and Chirikjian (2015).

This paper makes the following contributions:

(i) The portion of the joint rotation–translation search space that results in collision when placing arbitrary bodies with crystallographic symmetry is characterized mathematically, building on our previous work in Chirikjian & Shiffman (2016).

(ii) The condition that a macromolecular crystal is a solid object is formulated in terms of semigroup generators, and is used to further reduce the size of the feasible search space in which collisions are avoided.

(iii) It is observed that the critical factor is not the exact shape of the body (or macromolecule) $B$ itself, but rather the Minkowski sums [see equation (27)] of $B$ with its reflected rotated versions $-R \cdot B, R \in \mathrm{SO}(3)$. And these Minkowski sums can be approximated using the Minkowski sums of equivalent-moment-of-inertia ellipsoids derived from the original body.

(iv) It is shown that, when taken in combination, the above factors lead to small feasible regions in rotation–translation space where computational resources can be focused. Using ellipsoidal proxies approximating the bodies, one obtains descriptions of the approximate feasible regions in terms of Minkowski sums of ellipsoids, which are easily parameterized, thereby making high-dimensional searches more tractable.

(v) Because of roto-reflection symmetries of ellipsoids and of the Laue groups of the diffraction patterns, only a fraction of the feasible regions need to be explored.

(vi) Examples are provided to demonstrate this theory in the planar case with ellipses in $p2$ symmetry and in 3D with triaxial ellipsoids in the case of $P2_12_12_1$ symmetry.

The discussion in this paper concerns only the case where there is one protein molecule per asymmetric unit in the crystal. However, the general approach used here can be applied to the articulated case and multi-protein-per-asymmetric-unit case, where the dimension of the motion space is higher, but the percentage of feasible motions is much smaller.

### 1.1. Mathematics review

Let $\Gamma < G \doteq \mathrm{SE}(3)$ denote the orientation-preserving (or Sohncke) space group of a macromolecular crystal, and let $T \triangleleft \Gamma$ denote the translational symmetry group of the finest translational lattice $\mathbb{L}$. The group $G$ of continuous rigid-body motions consists of rotation–translation pairs of the form $g = (R, \mathbf{t}) \in \mathrm{SO}(3) \ltimes \mathbb{R}^3 \equiv G$ acting on points $\mathbf{x} \in X \doteq \mathbb{R}^3$ by $g \cdot \mathbf{x} = R\mathbf{x} + \mathbf{t}$, and with the group law[1]

$$g_1 g_2 = (R_1, \mathbf{t}_1)(R_2, \mathbf{t}_2) = (R_1 R_2, R_1 \mathbf{t}_2 + \mathbf{t}_1).$$

The translation group $T$ of the primitive lattice, $\mathbb{L}$, acts on Euclidean space $X$ without fixed points, and the resulting orbit space $T\backslash X$ is the 3-torus. The realizations of the crystallographic unit cell and asymmetric units are the fundamental domains $F_{T\backslash X}$ and $F_{\Gamma\backslash X}$, respectively. Previous papers in this series examined the relationships between $F_{\Gamma\backslash G}$, $F_{T\backslash X}$ and $F_{\Gamma\backslash X}$.

The configuration space of all possible ways to place a rigid body with symmetry $\Gamma$ is the right-coset space $\Gamma\backslash G$. Within this context, the main goal of MR can be stated as the search for the rigid-body motion $g \in F_{\Gamma\backslash G} \subset G$ that places a model of a macromolecular structure in the asymmetric unit at the same position and orientation as the true molecule in the crystal. The search for these placements is based on information obtained directly from the diffraction pattern.

If $f(\mathbf{x})$ is the density of an individual macromolecule, then the density $\rho(\mathbf{x}; g)$ in the crystal with the molecule placed at its true position and orientation $g$ is given by

$$\rho(\mathbf{x}; g) = \sum_{\gamma \in \Gamma} f((\gamma g)^{-1} \cdot \mathbf{x}) \qquad (1)$$

where

$$\gamma \cdot \mathbf{x} = R_\gamma \mathbf{x} + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma).$$

Here $R_\gamma \in \mathrm{SO}(3)$ is an element of the concrete point group $\mathbb{P}$ of $\Gamma$, and $\mathbf{t}_\gamma \in \mathbb{L}$ [whereas $(\mathbb{I}, \mathbf{t}_\gamma) \in T$]. For symmorphic space groups the function $\mathbf{v} : \mathbb{P} \to F_{T\backslash X}$ can always be set to $\mathbf{v}(R_\gamma) = \mathbf{0}$ for all $R_\gamma \in \mathbb{P}$ by an appropriate choice of origin, whereas for nonsymmorphic space groups $\mathbf{v}(R_\gamma) \neq \mathbf{0}$ for at least one $R_\gamma \in \mathbb{P}$ no matter how the origin is chosen. Using the terminology from Chirikjian, Ratnayake *et al.* (2015), $T\backslash\Gamma$ is the *abstract* point group of a space group $\Gamma$, and a non-unique finite set of representatives of this quotient group can be constructed as

[1] We use the symbol $\doteq$ to denote a definition. In previous papers in this series, the product $g_1 g_2$ was denoted as $g_1 \circ g_2$.

$$F_{T\backslash\Gamma} \doteq \{(R_\gamma, \mathbf{v}(R_\gamma)) | R_\gamma \in \mathbb{P}\}.$$

Then $\rho(\mathbf{x}; g)$ can be expressed as a Fourier series of the form

$$\rho(\mathbf{x}; g) = \sum_{\mathbf{k}\in\mathbb{L}^*} \hat{\rho}(\mathbf{k}; g) \exp(2\pi i \mathbf{k}\cdot\mathbf{x})$$

where $\mathbb{L}^*$ is the reciprocal (or dual) lattice of $\mathbb{L}$ and the Fourier coefficients of $\rho$ are of the form

$$\hat{\rho}(\mathbf{k}; g) = \int_X \sum_{\gamma\in F_{T\Gamma}} f((\gamma g)^{-1}\cdot\mathbf{x}) \exp(-2\pi i \mathbf{k}\cdot\mathbf{x})\, d\mathbf{x}. \quad (2)$$

Since $f(\mathbf{x})$ takes values of zero outside the macromolecule, it is compactly supported and so the integrand in the above integral only has contributions from a bounded domain. This domain may not fit in the standard unit cell, but there is a choice of unit cell which contains it.

The diffraction pattern consists of intensities of the form

$$I(\mathbf{k}; g) = |\hat{\rho}(\mathbf{k}; g)|^2 = \hat{\mathcal{P}}(\mathbf{k}; g),$$

where $\mathcal{P}(\mathbf{x}; g)$ is the Patterson correlation function given by

$$\begin{aligned}\mathcal{P}(\mathbf{x}; g) &\doteq \int_{F_{T\backslash X}} \rho(\mathbf{y}; g)\rho(\mathbf{y}+\mathbf{x}; g)\, d\mathbf{y} \\ &= \sum_{\gamma\in F_{T\Gamma}} \int_X f\big((\gamma g)^{-1}\cdot\mathbf{y} + R_g^{\mathrm{T}} R_\gamma^{\mathrm{T}}\mathbf{x}\big)\rho(\mathbf{y}; g)\, d\mathbf{y}. \quad (3)\end{aligned}$$

It follows that

$$\mathcal{P}(\pm R_\gamma\mathbf{x}; g) = \mathcal{P}(\mathbf{x}; g)\ \forall R_\gamma \in \mathbb{P}. \quad (4)$$

Let $\mathbb{P}^*$ denote the group of rotational symmetries of the reciprocal lattice, consisting of elements of the form $R_\gamma^*$. Then $I(\mathbf{k}, g) = \hat{\mathcal{P}}(\mathbf{k}; g)$ is invariant under the Laue group $\mathbb{P}^* \cup (-\mathbb{I})\mathbb{P}^* < \mathrm{O}(3)$; *i.e.*

$$I(\pm R_\gamma^*\mathbf{k}; g) = I(\mathbf{k}; g)\ \forall R_\gamma^* \in \mathbb{P}^*.$$

We shall use this Laue class symmetry to reduce the size of the search space in Section 4.2.

A 'strong' MR cost function for determining correct placements is of the form

$$\mathrm{Cost}\,(g) = \sum_{\mathbf{k}\in\mathbb{L}^*} \left|I_{\mathrm{model}}(\mathbf{k}; g) - I_{\mathrm{actual}}(\mathbf{k}; g_{\mathrm{actual}})\right|^2, \quad (5)$$

wherein the model diffraction intensity mimics the actual one. The computational drawback of this in comparison with classical MR is that it requires a search over the 6D space $F_{\Gamma\backslash G}$. For this reason, in classical MR, a 'weaker' cost function is used in which $I_{\mathrm{model}}(\mathbf{k}; g)$ in equation (5) is replaced by $|\hat{f}_{\mathrm{model}}(\mathbf{k}; g)|^2$, where

$$\begin{aligned}\hat{f}_{\mathrm{model}}(\mathbf{k}; g) &= \int_X f(g^{-1}\cdot\mathbf{x}) \exp(-2\pi i \mathbf{k}\cdot\mathbf{x})\, d\mathbf{x} \\ &= \exp(-2\pi i \mathbf{k}\cdot\mathbf{t}) \int_X f(R^{\mathrm{T}}\mathbf{x}) \exp(-2\pi i \mathbf{k}\cdot\mathbf{x})\, d\mathbf{x}, \quad (6)\end{aligned}$$

where $g = (R, \mathbf{t})$. The reason for doing so is that $|\exp(-2\pi i \mathbf{k}\cdot\mathbf{t})| = 1$ and hence the resulting cost function is independent of $\mathbf{t}$ and depends only on $R \in \mathrm{SO}(3)$. This is not the case for equation (5). Hence, in the classical MR method, a pure rotation search is conducted first, followed by a translational search. This separation of the dimensions leads to reduced computational burden, but the search has less specificity than when using the stronger cost function.

However, in our method, the cost function (5) will have a value of zero when an exact proxy for the unknown molecule is placed at the correct position and orientation, and it would be reasonable to expect that small changes in the shape of the proxy would not lead to radically different values of the optimal pose. In contrast, it is well known that when using classical MR, even when using a good proxy, the highest peak in the rotation function may not be the one to use, and sometimes many peaks (on the order of 100 or more) may need to be investigated. This is because the rotation function is designed to examine the correlation between the diffraction pattern of a single proxy molecule (in a unit cell) against the actual diffraction pattern. The latter has built in it the 'cross talk' between symmetry mates, which are often thought of as 'noise' in the context of classical MR. That is, unlike in our cost function, global optimality of the rotation function need not correspond to the correct rotation. And if the correct rotation is not obtained in the first stage of a classical MR search, then there is no way to obtain the correct translation in the second stage.

Building on developments in previous papers in this series, this paper provides the mathematical underpinnings for efficient searches in the joint rotational–translational space. The key features that make our approach efficient rely on two observations: (i) large parts of the search space are infeasible because they correspond to symmetry mates being placed in collision; (ii) since a crystal is a solid physical object, symmetry mates must be in contact in such a way that there are no isolated islands. With these constraints, MR searches in principle could be limited to boundaries of collision zones, thereby reducing the dimensionality of the search. But in practice these boundaries are difficult to parameterize exactly due to the complicated geometry of proteins. Instead we apply these constraints using ellipsoidal proxies for proteins to bound the feasible regions, and we see that the volume of these feasible regions can be quite small relative to the total volume of the motion space. Constraint (i) has been previously addressed by the authors in the case of planar motions of circular bodies and chiral wallpaper group symmetry (Chirikjian & Shiffman, 2016), and the theory is extended here to general shapes and 3D space groups. Constraint (ii) is related to topological crystallography, the literature of which is reviewed below.

### 1.2. Literature review

The literature on packing bodies subject to crystallographic symmetry is immense. At one extreme of this literature lie works on tilings with wallpaper symmetry in the plane or with crystallographic symmetry in 3D (Grünbaum & Shephard, 1987). Closed tiles that fill space under the action of a crystallographic group can all be viewed as realizations of the fundamental domain $F_{\Gamma\backslash X}$, *i.e.* the closure of the crystallographic asymmetric unit. At another extreme lie 'crystal nets' of sparsely separated points (Wells, 1977). In the middle lies the study of dense sphere packings (which in the 3D case

has been studied at least from the time of Kepler, and in higher dimensions has important implications for coding theory) (Conway & Sloane, 1999; Thompson, 1983). Dense irregular packings have also been studied extensively in recent years. A much smaller literature exists on packing ellipses and ellipsoids subject to symmetry constraints (Matsumoto & Nowacki, 1966; Matsumoto, 1968).

A less-studied area is that of packing irregular bodies (such as biological macromolecules) subject to crystallographic symmetry constraints. Such works are usually motivated by the chemical physics behind these macromolecular interactions (Kitaigorodskii, 1973; Brock & Dunitz, 1994; Dunitz, 1996). However, our motivation in studying packing constraints in the context of MR is somewhat different because the exact structure of the macromolecules under consideration is not known *a priori*. Rather, we have only an estimate of the shape obtained from a homolog, and seek to limit searches in MR using both the fact that collisions between symmetry mates must not occur and surface contacts between them must occur.

Works in the area of so-called 'topological crystallography' include Johnson *et al.* (1996), Thimm (2009), Peresypkina & Blatov (2000), O'Keeffe (1995), Sunada (2013), and provide some tools that are useful for our problem. Namely, quotient graph structures convert the discussion of a given infinite graph (or net) of vertices with crystallographic symmetry to that of a finite graph structure in the space of orbits of vertices. In recent work, Eon (2016*a*,*b*) used quotient graph methodologies to study crystallographic topology. Such works presuppose a given graph. In our work, the graph is not known *a priori* and we express all possible graphs as an enumeration problem using the language of semigroup generators.

## 2. The feasible space

We first give a general description of the feasible space for a crystal formed by a body $B$ and its symmetry mates under a chiral (or Sohncke) space group $\Gamma < \mathrm{SE}(3)$ [and, more generally, an orientation-preserving crystallographic group $\Gamma < \mathrm{SE}(n)$].

We shall state our main results in arbitrary dimension $n$, which allows us to illustrate some of our concepts in dimension 2. So we now let $X \doteq \mathbb{R}^n$ and $G \doteq \mathrm{SE}(n) \equiv \mathrm{SO}(n) \ltimes X$. To simplify notation, we make the identification $\mathrm{SO}(n) \equiv \mathrm{SO}(n) \times \{\mathbf{0}\} < G$.

By a *body* $B$ we mean a bounded, connected, open subset of $X$, which in the case $n = 3$ can represent one of the molecules forming a crystal. We denote the closure of $B$ by $\overline{B}$, which we call a *closed body*, and we let

$$C \doteq g \cdot \overline{B}$$

denote the closed body as it is placed in the crystal by a motion $g \in G$. We say that the conglomeration of closed bodies[2]

$$\Gamma \cdot C \doteq \bigcup_{\gamma \in \Gamma} \gamma \cdot C = \bigcup_{\gamma \in \Gamma} \gamma g \cdot \overline{B}$$

is *connected* if every symmetry mate $\gamma \cdot C$ is connected to $C$ by a finite chain of symmetry mates; *i.e.* there is a sequence $\gamma_0 = e, \gamma_1, \gamma_2 \ldots, \gamma_k = \gamma$ of elements of $\Gamma$, such that $\gamma_{j-1} \cdot C \cap \gamma_j \cdot C \neq \emptyset$ for $1 \leq j \leq k$.[3] A necessary condition for a body $g \cdot B$ and its $\Gamma$-symmetry mates to form a crystal is that $g \cdot B$ does not intersect any of its symmetry mates $\gamma g \cdot B$ and that $\Gamma \cdot C$ is connected. The first condition guarantees that there are no steric clashes between the bodies, since if $\gamma_1 g \cdot B \cap \gamma_2 g \cdot B \neq \emptyset$, then $g \cdot B \cap \gamma_1^{-1}\gamma_2 g \cdot B \neq \emptyset$. The second condition is necessary for the crystal to be solid, since otherwise one chain of symmetry mates would be free to move relative to the others.

We fix a fundamental domain $F_{\Gamma\backslash G}$ for $\Gamma\backslash G$ and we let $\mathcal{A}(\Gamma, B) \subset F_{\Gamma\backslash G}$ denote the set of motions $g \in F_{\Gamma\backslash G}$ such that the sets $\{\gamma g \cdot B : \gamma \in \Gamma\}$ are pairwise disjoint and the closure of their union, $\Gamma \cdot C$, is connected. We call $\mathcal{A}(\Gamma, B)$ the *feasible space* for $B$. When $g \in \mathcal{A}(\Gamma, B)$, the bodies $\gamma g \cdot B$ (for all $\gamma \in \Gamma$) form a solid crystal without steric clashes. As discussed in Section 1, the MR method involves the evaluation of a cost function for each candidate pose $g \cdot B$ of the body $B$. One way to make this search efficient is to restrict the search to a small region in the search space $F_{\Gamma\backslash G}$ containing the feasible space $\mathcal{A}(\Gamma, B)$. If $g \in \mathcal{A}(\Gamma, B)$, then each symmetry mate $\gamma g \cdot \overline{B}$ of the closed body $g \cdot \overline{B}$ is either disjoint from $g \cdot \overline{B}$ or has only boundary points in common with $g \cdot \overline{B}$. When the intersection of two closed bodies $\overline{B}_1, \overline{B}_2$ contains only common boundary points of the bodies and is non-empty, then we say that $B_1$ and $B_2$ *kiss*. Thus the feasible space consists of those $g \in \mathcal{A}(\Gamma, B)$ such that the sets $\gamma g \cdot B$ are pairwise disjoint and enough of them kiss $g \cdot B$ to form a crystal. The precise mathematical description of $\mathcal{A}(\Gamma, B)$ is given by equation (12) below [and equivalently by equations (13) or (14)] and is based on an algebraic criterion for a macromolecular crystal to be solid (Lemma 2.1).

For a subset $Y \subset X$, we write

$$\begin{aligned} Z(\gamma, Y) &\doteq \left\{ g \in F_{\Gamma\backslash G} : (g \cdot Y) \cap (\gamma g \cdot Y) \neq \emptyset \right\} \\ &= \left\{ g \in F_{\Gamma\backslash G} : Y \cap (g^{-1}\gamma g \cdot Y) \neq \emptyset \right\}, \end{aligned} \tag{7}$$

for $\gamma \in \Gamma \setminus \{e\}$. We note that $Z(\gamma^{-1}, Y) = Z(\gamma, Y)$ since $g \cdot Y$ collides with $\gamma g \cdot Y$ if and only if $\gamma^{-1} g \cdot Y$ collides with $g \cdot Y$. In the case where $Y$ is the open body $B$, then $Z(\gamma, B)$ is the space consisting of motions $g$ which cause $g \cdot B$ to intersect its symmetry mate $\gamma g \cdot B$ and is called a *collision zone*.

Our first condition on the feasible space of a body $B$ states that we need to avoid those motions $g \in G = \mathrm{SE}(n)$ which cause symmetry mates $\{\gamma g \cdot B : \gamma \in \Gamma\}$ to intersect. In a previous paper (Chirikjian & Shiffman, 2016), we described the space of motions where $g \cdot B$ intersects at least one of its symmetry mates. This space is the *collision space* $S(\Gamma, B)$ given by

[2] We write $(\gamma g) \cdot \overline{B}$ as $\gamma g \cdot \overline{B}$, and correspondingly for similar expressions.

[3] Since $C$ is a compact connected set and $\Gamma$ is a properly discontinuous action on $\mathbb{R}^n$, this condition is equivalent to connectedness of $\Gamma \cdot C$ in the topological sense. See Section 3.1 of Chirikjian *et al.* (2017) for a review of properly discontinuous actions.

$$S(\Gamma, B) \doteq \bigcup_{\gamma\in\Gamma\setminus\{e\}} Z(\gamma, B)$$
$$= \left\{ g \in F_{\Gamma\backslash G} : \left[ B \cap \bigcup_{\gamma\in\Gamma\setminus\{e\}} (g^{-1}\gamma g \cdot B) \right] \neq \emptyset \right\}. \quad (8)$$

We then have our first outer estimate for the feasible region:

$$\mathcal{A}(\Gamma, B) \subset F_{\Gamma\backslash G} \setminus S(\Gamma, B)$$
$$= \bigcap_{\gamma\in\Gamma\setminus\{e\}} \{g \in F_{\Gamma\backslash G} : B \cap g^{-1}\gamma g \cdot B = \emptyset\}. \quad (9)$$

The MR search can be limited to the region given by the right-hand side of equation (9), as was discussed by Chirikjian & Shiffman (2016). However, the search can be limited to a much smaller region by applying our second condition, which we shall describe more precisely. To do this, we begin with some mathematical preliminaries.

We recall that a semigroup is a set closed under a binary associative operation. We say that a subset $\Lambda$ of a semigroup $\Pi$ *generates the semigroup* $\Pi$ if every element of $\Pi$ is a finite product of elements of $\Lambda$. Every group is a semigroup and, in particular, $\Gamma$ is a semigroup as well as a group. For example, if $\Gamma < \mathrm{SE}(3)$ is a space group of type $P1$ and $\gamma_1, \gamma_2, \gamma_3 \in \Gamma$ are translations such that $\gamma_1(\mathbf{0}), \gamma_2(\mathbf{0}), \gamma_3(\mathbf{0})$ generate the lattice in $\mathbb{R}^3$, then $\{\gamma_1, \gamma_2, \gamma_3\}$ generates the group $\Gamma$, but does not generate the semigroup $\Gamma$. Instead $\{\gamma_1, \gamma_2, \gamma_3, \gamma_1^{-1}, \gamma_2^{-1}, \gamma_3^{-1}\}$ generates the semigroup $\Gamma$. In general, if $\Lambda$ generates the group $\Gamma$, then $\Lambda \cup \Lambda^{-1}$ generates the semigroup $\Gamma$.

*Lemma 2.1*. Let $C$ be a bounded connected closed subset of $X$, and let $\Lambda$ be the set of $\gamma \in \Gamma$ such that $C \cap \gamma \cdot C \neq \emptyset$. Then $\Gamma \cdot C \doteq \bigcup_{\gamma\in\Gamma} \gamma \cdot C$ is connected if and only if $\Lambda$ generates the semigroup $\Gamma$.

*Proof*. Suppose that $\Gamma \cdot C$ is connected, and let $\beta \in \Gamma$ be arbitrary. Then there exists a finite sequence $\gamma_0 = e, \gamma_1, \ldots, \gamma_{k-1}, \gamma_k = \beta$ in $\Gamma$ such that $\gamma_{j-1} \cdot C \cap \gamma_j \cdot C \neq \emptyset$, for $1 \leq j \leq k$. Therefore $C \cap \lambda_j \cdot C \neq \emptyset$ where $\lambda_j = \gamma_{j-1}^{-1}\gamma_j$ and hence $\lambda_j \in \Lambda$ for $1 \leq j \leq k$, and $\beta = \lambda_1 \ldots \lambda_k$. Since $\beta \in \Gamma$ is arbitrary, $\Lambda$ is a set of semigroup generators of $\Gamma$.

Conversely, suppose that $\Lambda$ generates the semigroup $\Gamma$, and let $\beta \in \Gamma$ be arbitrary. Then $\beta = \lambda_1 \ldots \lambda_k$ with the $\lambda_j \in \Lambda$. Then $C \cap \lambda_1 \cdot C \neq \emptyset$, $\lambda_1 \cdot C \cap \lambda_1\lambda_2 \cdot C = \lambda_1 \cdot (C \cap \lambda_2 \cdot C) \neq \emptyset$, and it follows by induction that

$$\lambda_1 \ldots \lambda_{j-1} \cdot C \cap \lambda_1 \ldots \lambda_j \cdot C \neq \emptyset, \quad \text{for} \quad j = 1, \ldots, k.$$

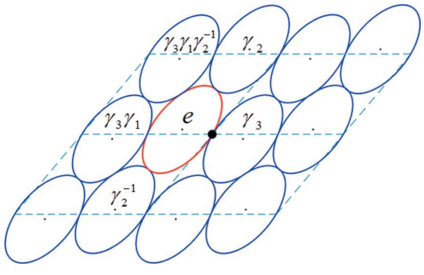



**Figure 1**
One ellipse kissing five neighbors with $p2$ symmetry.

**Table 1**
Values of $\lambda_j$ in Example 1.

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|---|---|---|---|---|
| $\gamma_2$ | $\gamma_3$ | $\gamma_3\gamma_1\gamma_2^{-1}$ | $\gamma_3\gamma_1$ | $\gamma_2^{-1}$ |

Therefore, $C$ is connected to $\beta \cdot C = \lambda_1 \ldots \lambda_k \cdot C$ by a chain of symmetry mates. □

In a molecular crystal, the number of symmetry mates $\gamma g \cdot B$ kissing $g \cdot B$ is called the molecular coordination number, or ligancy (Peresypkina & Blatov, 2000). Since the molecular coordination number of a crystal composed of symmetry mates of a molecule $B$ with position $g \cdot B$ is the cardinality of the set of $\gamma \in \Gamma\setminus\{e\}$ such that $g \cdot \overline{B} \cap \gamma g \cdot \overline{B} \neq \emptyset$, we obtain the following consequence of Lemma 2.1:

*Theorem 2.2*. The molecular coordination number in a crystal is equal to or greater than the minimum number of semigroup generators of $\Gamma$.

In the following, two examples are shown to illustrate Theorem 2.2 in dimension 2. In both of the examples, $\Gamma$ is in the class $p2$, which is made up of both translations and 180° rotations. (The 180° rotation is referred to as a half-turn.) The unit cell is a parallelogram.

In our first example illustrated in Fig. 1, each body is kissing exactly five of its neighbors, so its 2D molecular coordination number is 5. The figure shows four unit cells outlined by dotted lines, and the origin is indicated by a dot. As before, we let $C \doteq g \cdot \overline{B}$ denote the reference body; $C$ is indicated in the figures below as bounded by a red ellipse, and its symmetry mates are blue.

We denote the translation along the horizontal axis of the lattice as $\gamma_1$ and the translation along the other axis as $\gamma_2$. The 180° rotation about the origin is described as $\gamma_3$. The five symmetry mates kissing $C$ can be expressed by $\lambda_j \cdot C$ $(j = 1, \ldots, 5)$, where the values of the $\lambda_j$ are given in Table 1 and appear inside the corresponding ellipses in the figure.

We note that the set $\Lambda \doteq \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$ generates the semigroup $\Gamma$ as Lemma 2.1 tells us, but it is not a minimal generating set since, for example, $\{\lambda_1, \lambda_2, \lambda_4\}$ is a (minimal) generating set ($\lambda_3 = \lambda_4\lambda_2\lambda_1\lambda_2$, $\lambda_5 = \lambda_2\lambda_1\lambda_2$).

In our second example with molecular coordination number 3, each elliptical body kisses exactly three of its neighboring symmetry mates, as we can see in Fig. 2.

The three symmetry mates kissing $C$ in Fig. 2 can be expressed by $\lambda_1 \cdot C$, $\lambda_2 \cdot C$, $\lambda_3 \cdot C$, where the $\lambda_j$ are given in Table 2.

Here, $\{\lambda_1, \lambda_2, \lambda_3\}$ is a minimal generating set for the semigroup $\Gamma$. Since no set of two elements can generate $\Gamma$, the minimal molecular coordination number for 2D crystals with $p2$ symmetry is 3.

We now give some mathematical descriptions of the feasible space $\mathcal{A}(\Gamma, B)$. First, we define the 'collision-causing symmetries' $\Lambda(Y) \subset \Gamma$ associated with a set $Y \subset X$ given by

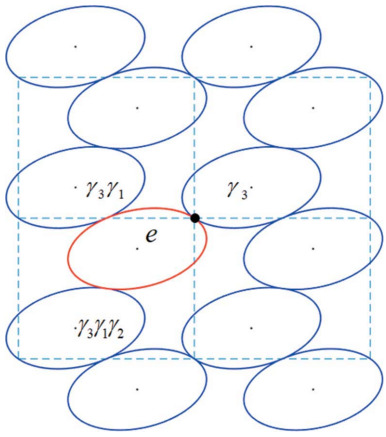



**Figure 2**
One ellipse kissing three neighbors with $p2$ symmetry.

$$\Lambda(Y) \doteq \{\gamma \in \Gamma \setminus \{e\} : Y \cap \gamma \cdot Y \neq \emptyset\}, \tag{10}$$

so that

$$\gamma \in \Lambda(g \cdot Y) \iff g \in Z(\gamma, Y). \tag{11}$$

We note that $\gamma \in \Lambda(Y)$ if and only if $\gamma^{-1} \in \Lambda(Y)$, since the condition that $Y$ intersects $\gamma \cdot Y$ is equivalent to the condition that $\gamma^{-1} \cdot Y$ intersects $Y$. The following algebraic description of the feasible space is a consequence of Lemma 2.1.

*Theorem 2.3*. Let $\Gamma < G$ be a crystallographic group, and let $B$ be a bounded connected open set in $X$. Then

$$\begin{aligned}\mathcal{A}(\Gamma, B) &= \{g \in F_{\Gamma\backslash G} : \Lambda(g \cdot B) = \emptyset,\ \ \Lambda(g \cdot \overline{B}) \text{ generates } \Gamma\} \\ &= \{g \in F_{\Gamma\backslash G} : \Lambda(g \cdot B) = \emptyset,\ \ \Lambda(g \cdot \partial B) \text{ generates } \Gamma\}.\end{aligned} \tag{12}$$

*Proof*. Since $\Lambda(Y) = \Lambda(Y)^{-1}$ for any set $Y$, $\Lambda(g \cdot \overline{B})$ generates the *group* $\Gamma$ if and only if it generates the *semigroup* $\Gamma$. Equation (12) then follows from Lemma 2.1 with $C = g \cdot \overline{B}$ and from the fact that $\Lambda(g \cdot \overline{B}) = \Lambda(g \cdot \partial B)$. □

Equation (12) says that a motion $g \in F_{\Gamma\backslash G}$ is in the feasible space $\mathcal{A}(\Gamma, B)$ if and only if $g \notin S(\Gamma, B)$ and there exist $\gamma_1, \ldots, \gamma_m$ generating $\Gamma$ such that $g^{-1}\gamma_j g \cdot \partial B$ intersects $\partial B$ for $1 \leq j \leq m$.

Let $\mathcal{M}$ denote the collection of minimal generating subsets of the group $\Gamma$; *i.e.* a generating subset $\Lambda$ is in $\mathcal{M}$ if it does not properly contain another subset that generates the group $\Gamma$. Using equation (11), we have an equivalent formulation of equation (12) in terms of intersections and unions of collision sets:

$$\mathcal{A}(\Gamma, B) = \bigcup_{\Lambda \in \mathcal{M}} \left[ \bigcap_{\gamma \in \Lambda} Z(\gamma, \overline{B}) \right] \setminus \bigcup_{\gamma \in \Gamma \setminus \{e\}} Z(\gamma, B). \tag{13}$$

Equivalently,

**Table 2**
Values of $\lambda_j$ in Example 2.

| $\lambda_1$ | $\lambda_2$ | $\lambda_3$ |
|---|---|---|
| $\gamma_3$ | $\gamma_3\gamma_1$ | $\gamma_3\gamma_1\gamma_2$ |

$$\mathcal{A}(\Gamma, B) = \bigcup_{\Lambda \in \mathcal{M}} \left[ \bigcap_{\gamma \in \Lambda} Z(\gamma, \partial B) \right] \setminus \bigcup_{\gamma \in \Gamma \setminus \{e\}} Z(\gamma, B). \tag{14}$$

## 3. Use of ellipsoidal models of proteins

A protein consists of many atoms which are bonded together and can be modeled as an articulated multi-rigid-body. Here we consider the simplest model of a single rigid-body protein. Packing of proteins indicates that the position and orientation of each protein, in a specific lattice under certain symmetry, are fixed. This results in the constraint [equation (12)] that the surface of each protein should be in contact with (or 'kiss') certain of its neighbors. Since proteins have irregular shapes, either checking or setting kissing constraints between each pair of proteins is considerably difficult. A reasonable method is to use an envelope with regular shape to approximate and replace the protein for further calculation.

The use of ellipsoidal models of proteins has a long history. Richards (1977), Lesk & Rose (1981), Zehfus *et al.* (1985), Thornton *et al.* (1986), Prabhakaran & Ponnuswamy (1982) all used equivalent inertia ellipsoids to analyze the surface properties of proteins. In works spanning decades, Taylor and coworkers (Taylor *et al.*, 1983, 2001; Taylor & Aszódi, 2005) made arguments for the efficacy of ellipsoidal models in various contexts in the study of globular proteins. Ellipsoidal models have long been used to approximate the hydrodynamic effects in rotational diffusion experiments (Ryabov *et al.*, 2006). As such, it should be noted that when a homolog is not available for use in MR, an ellipsoidal approximation nevertheless may be obtained from the protein of interest itself from anisotropy spectroscopy experiments. Ellipsoidal models have also been used previously in the context of crystallographic and other crowded environments (Müller & Schrauber, 1992; Qin & Zhou, 2013).

The novelty of our analysis lies in the observation that even when an ellipsoidal model does not fit the protein so well, it is nevertheless the case that the Minkowski sum of a protein with rotated reflected copies of itself matches well with the Minkowski sum of ellipsoids with equivalent moments of inertia. We then utilize a recent finding that the boundary of the Minkowski sum of two ellipsoidal bodies can be parameterized in closed form (Yan & Chirikjian, 2015).

By an *ellipsoidal body* we mean an open (solid) ellipsoid, *i.e.* a set of the form

$$\{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}^{\mathrm{T}} M \mathbf{x} < 1\},$$

where $M$ is a positive definite symmetric $n \times n$ matrix.

Now, let us suppose that we have inscribed and circumscribed ellipsoidal bodies $E_1 \subseteq B \subseteq E_2$. Because ellipses/ellipsoids are relatively simple compared with polygons/poly-

hedra, they are widely applied in collision checking in robotics. Ellipsoids can serve as the boundary for humanoids in motion planning and are also used as the bounding volume for links in serial manipulators. Those applications support ellipsoids as good candidates for the bounding bodies $E_1$, $E_2$, as well as for the replacement of a real protein in packing calculations.

Let $E_1 \subseteq B \subseteq E_2$. (Although our motivation involves the use of ellipsoids, the bodies $E_1$, $E_2$ can be arbitrary bounded connected open sets in the following discussion.) We note that $S(\Gamma, E_1) \subseteq S(\Gamma, B)$ when $E_1 \subseteq B$, since if a motion $g$ causes $g \cdot E_1$ to collide with $\gamma g \cdot E_1$, then the larger region $g \cdot B$ will collide with $\gamma g \cdot B$. Thus we have, by equation (9),

$$\begin{aligned}\mathcal{A}(\Gamma, B) &\subset F_{\Gamma\backslash G} \setminus S(\Gamma, B) \subset F_{\Gamma\backslash G} \setminus S(\Gamma, E_1) \\ &= \bigcap_{\gamma \in \Gamma \setminus \{e\}} Z(\gamma, E_1)^{\mathrm{C}}, \end{aligned} \tag{15}$$

where

$$Z(\gamma, E_1)^{\mathrm{C}} \doteq F_{\Gamma\backslash G} \setminus Z(\gamma, E_1) = \{g \in F_{\Gamma\backslash G} : E_1 \cap g^{-1}\gamma g \cdot E_1 = \emptyset\} \tag{16}$$

denotes the complement of $Z(\gamma, E_1)$ in the fundamental domain $F_{\Gamma\backslash G}$.

Furthermore, for every feasible $g$, the closed body $g \cdot \overline{E}_2$ must collide with at least one of its $\Gamma$-symmetry mates. Thus for bodies $E_1 \subseteq B \subseteq E_2$, we also have

$$\mathcal{A}(\Gamma, B) \subset \bigcup_{\gamma \in \Gamma \setminus \{e\}} \{g \in F_{\Gamma\backslash G} : E_1 \cap g^{-1}\gamma g \cdot E_1 = \emptyset, \overline{E}_2 \cap g^{-1}\gamma g \cdot \overline{E}_2 \neq \emptyset\}. \tag{17}$$

Hence the search can be restricted to the space given by the right side of equation (17). A much smaller search space, as illustrated in Fig. 4 below, is the space

$$\mathcal{I}(\Gamma, E_1, E_2) \doteq \left\{g \in F_{\Gamma\backslash G} : \Lambda(g \cdot E_1) = \emptyset, \Lambda(g \cdot \overline{E}_2) \text{ generates } \Gamma\right\}. \tag{18}$$

Note that $\mathcal{I}(\Gamma, E_1, E_2)$ can be described in terms of intersections and unions as

$$\mathcal{I}(\Gamma, E_1, E_2) = \bigcup_{\Lambda \in \mathcal{M}} \left[\bigcap_{\gamma \in \Lambda} Z(\gamma, \overline{E}_2)\right] \setminus \bigcup_{\gamma \in \Gamma \setminus \{e\}} Z(\gamma, E_1). \tag{19}$$

Thus, a motion $g \in F_{\Gamma\backslash G}$ is in $\mathcal{I}(\Gamma, E_1, E_2)$ if and only if $g \notin S(\Gamma, E_1)$ and there exist $\gamma_1, \ldots, \gamma_m$ generating $\Gamma$ such that $g^{-1}\gamma_j g \cdot \overline{E}_2$ intersects $\overline{E}_2$ for $1 \leq j \leq m$.

It follows from equation (12) that

$$\mathcal{A}(\Gamma, B) = \mathcal{I}(\Gamma, B, B) \tag{20}$$

and

$$E_1 \subseteq B \subseteq E_2 \Rightarrow \mathcal{A}(\Gamma, B) \subseteq \mathcal{I}(\Gamma, E_1, E_2). \tag{21}$$

We call $\mathcal{I}(\Gamma, E_1, E_2)$ the *informed search space* for bodies $E_1 \subseteq B \subseteq E_2$ since the resulting search is 'informed' by the lack of collisions of symmetry mates of $g \cdot E_1$ and required collisions of symmetry mates of $g \cdot \overline{E}_2$. Furthermore, if $E_1 \subseteq B \subseteq E_2$ is a tight encapsulation, then $\mathcal{I}(\Gamma, E_1, E_2)$ will be a good approximation to the feasible space $\mathcal{A}(\Gamma, B)$ and thus may provide an efficient search space for MR.

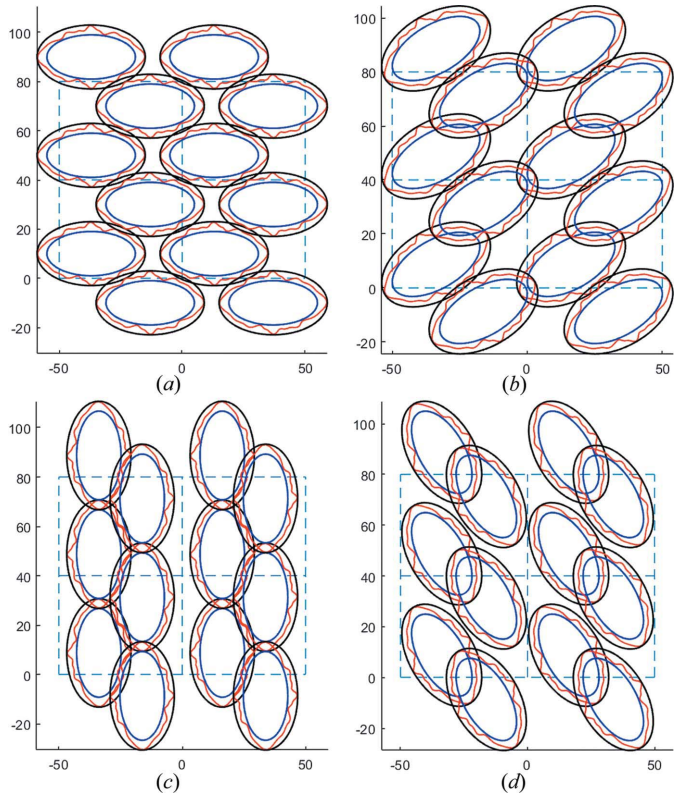



**Figure 3**
Configurations with $p2$ symmetry. ($a$) The orientation angle $\theta = 0$, ($b$) $\theta = \pi/6$, ($c$) $\theta = \pi/2$, ($d$) $\theta = 2\pi/3$.

Figs. 3–4 illustrate the informed search space in 2D for the case of $p2$ symmetry. Fig. 3 shows four configurations of a 2D irregular body $B$ (in red) with inscribed and circumscribed concentric ellipses $E_1, E_2$ together with their $p2$ symmetry mates; only Fig. 3($a$) is a configuration in the informed search space $\mathcal{I}(\Gamma, E_1, E_2)$.

The translational parts of the search spaces of equation (17) for four rotations of the body (corresponding to the orientations in Fig. 3) are unions of concentric rings, each ring representing one of the sets in the union in equation (17), as illustrated in Fig. 4. The translational parts of the corresponding informed search space $\mathcal{I}(\Gamma, E_1, E_2)$ consist of the points in Fig. 4 where three or more of the rings intersect; these points represent motions $g$ where $g \cdot \overline{E}_2$ intersects at least three symmetry mates. (Recall that the minimal molecular coordination number for 2D crystals with $p2$ symmetry is 3.) The regions consisting of these points are dark green in the figure. Note that the informed search space in Fig. 4($c$) is empty. In Fig. 4($a$), the points in the central part of the 'X' correspond to translations where $E_2$ intersects four of its symmetry mates, as shown in Fig. 3($a$). The circled black dots[4] in Figs. 4($a$)–4($d$) correspond to the respective configurations

[4] These dots are more visible when the figures are enlarged. The dot in Fig. 4($b$) is slightly outside of the dark green region since inner ellipses in Fig. 3($b$) intersect.

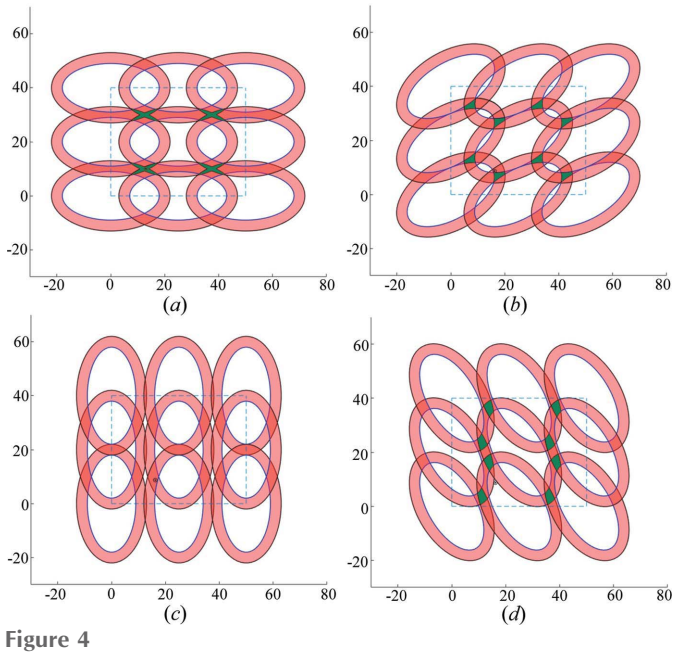



**Figure 4**
Search space for $p2$ symmetry: only the small green regions are in the informed search space. (*a*) $\theta = 0$, (*b*) $\theta = \pi/6$, (*c*) $\theta = \pi/2$, (*d*) $\theta = 2\pi/3$.

in Figs. 3(*a*)–3(*d*). The symmetries in Fig. 4 enable a reduction of the search space, as explained in Section 4.2.

Although $\Gamma$ and $\mathcal{M}$ are infinite sets, one only needs to check a finite number of conditions in equation (18) [or equivalently in equation (13)] since a bounded body $Y$ will not intersect $\gamma \cdot Y$ if the translational part of $\gamma$ is large. To quantify this observation, for a bounded set $D \subset X$ we let

$$r_D \doteq \sup\{|\mathbf{x}| : \mathbf{x} \in D\}. \tag{22}$$

For a fundamental domain for $\Gamma\backslash G$ of the form $F_{\Gamma\backslash G} = \mathrm{SO}(n) \times F_{\Gamma\backslash X}$, we have

$$\sup\big\{|g \cdot \mathbf{x}| : g \in F_{\Gamma\backslash G}, \mathbf{x} \in B\big\} = r_B + r_{F_{\Gamma\backslash X}}. \tag{23}$$

By translating $B$, we can choose $r_B$ to equal the circumradius of $B$. Then Jung's theorem tells us that $r_B \leq (\frac{n}{2n+2})^{1/2}\mathrm{diam}(B)$; if $B = -B$ (*e.g.* if $B$ is a centered ellipsoid), then $r_B = \frac{1}{2}\mathrm{diam}(B)$.

Recall that

$$\gamma = (R_\gamma, \mathbf{t}_\gamma + \mathbf{v}(R_\gamma)) \in \Gamma,$$

where $\mathbf{t}_\gamma$ is an element of the maximal lattice $\mathbb{L}$ corresponding to translations in $\Gamma$, and $\mathbf{v}(R_\gamma)$ is a fraction of an element of $\mathbb{L}$. Since

$$\begin{aligned} &g \cdot \overline{B} \cap \gamma g \cdot \overline{B} = g \cdot \overline{B} \cap ((R_\gamma g \cdot \overline{B} + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma)) \neq \emptyset \\ &\implies \mathbf{t}_\gamma + \mathbf{v}(R_\gamma) \in g \cdot \overline{B} + (-R_\gamma g \cdot \overline{B}), \end{aligned}$$

it follows from equation (23) that

$$\begin{aligned} &\bigcup_{g \in F_{\Gamma\backslash G}} \Lambda(g \cdot \overline{B}) \subseteq \Phi(\Gamma, B) \\ &\doteq \{(R_\gamma, \mathbf{t}_\gamma + \mathbf{v}(R_\gamma)) \in \Gamma \setminus \{e\} : |\mathbf{t}_\gamma + \mathbf{v}(R_\gamma)| \leq 2(r_B + r_{F_{\Gamma\backslash X}})\}, \end{aligned} \tag{24}$$

and hence to determine which symmetries $\gamma \in \Gamma$ are in $\Lambda(g \cdot E_1)$ and $\Lambda(g \cdot \overline{E}_2)$ in equation (18), one needs only to consider symmetries in the finite sets $\Phi(\Gamma, E_1)$ and $\Phi(\Gamma, E_2)$, respectively. Correspondingly, by defining

$$\mathcal{M}(\Gamma, B) \doteq \{\Lambda \in \mathcal{M} : \Lambda \subseteq \Phi(\Gamma, B)\}, \tag{25}$$

equation (13) can be written using only finite unions and intersections as

$$\begin{aligned} &\mathcal{I}(\Gamma, E_1, E_2) \\ &= \bigcup_{\Lambda \in \mathcal{M}(\Gamma, E_2)} \left[\bigcap_{\gamma \in \Lambda} Z(\gamma, \overline{E}_2)\right] \cap \left[\bigcap_{\gamma \in \Phi(\Gamma, E_1)} Z(\gamma, E_1)^{\mathrm{C}}\right], \end{aligned} \tag{26}$$

where $Z(\gamma, E_1)^{\mathrm{C}}$ denotes the complement of $Z(\gamma, E_1)$ [as in equation (16)].

Thus the informed search space $\mathcal{I}(\Gamma, E_1, E_2)$ is given by a finite number of conditions. In the case where $E_1$ and $E_2$ are ellipsoids, it is given by a finite number of polynomial inequalities;[5] *i.e.* it is the intersection of $F_{\Gamma\backslash G}$ with a semi-algebraic variety in $X$. If $E_1 = B = E_2$, then equation (20) implies that $\mathcal{A}(\Gamma, B) = \mathcal{I}(\Gamma, E_1, E_2)$ has zero volume in $F_{\Gamma\backslash G}$. (This follows from the fact that, in a crystal, some symmetry mates must kiss, which lowers the dimension of the search space.) Thus if we can find good approximating ellipsoids, we can restrict the search of motion space to regions of small volume in $F_{\Gamma\backslash G}$ that are amenable to computation.

## 4. General description of the collision space

We now review the general formulas from Chirikjian & Shiffman (2016) for the collision space $S(\Gamma, B)$ and give specific details involving cylinders over slices of Minkowski sums in the 3D case.

We let $\mathbb{P} < \mathrm{SO}(n)$ denote the 'concrete' point group corresponding to $\Gamma$, and we let $T \triangleleft \Gamma$ denote the maximal lattice translation (normal) subgroup of $\Gamma$, *i.e.* the one for which the index $[\Gamma : T]$ is minimal. Then $\mathbb{P} \cong T\backslash\Gamma$ (the abstract point group), and we let

$$F_{T\backslash\Gamma} \doteq \{(R_\gamma, \mathbf{v}(R_\gamma)) | R_\gamma \in \mathbb{P}\} \subset \Gamma.$$

We let $\mathbb{L} = T \cdot \mathbf{0} \subset X$ denote the lattice of rank $n$ in $X = \mathbb{R}^n$, which we can identify with the translation subgroup $T$. As above, we consider fundamental domains for $\Gamma\backslash G$ of the form $F_{\Gamma\backslash G} = \mathrm{SO}(n) \times F_{\Gamma\backslash X}$. For descriptions of some other fundamental domains, see Chirikjian *et al.* (2017). The formulas for the collision space from Chirikjian & Shiffman (2016) involve the *Minkowski sum*,

$$B + B' \doteq \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in B,\ \ y \in B'\}, \tag{27}$$

of two bodies $B, B' \in X$. Fig. 5 gives two simple examples of the Minkowski sum of two elliptical 2D bodies. Minkowski sums of sample protein bodies are illustrated in Section 6.2.

For a general body $B$, we write

[5] The condition that two ellipsoids intersect can be given by polynomial inequalities in the coefficients of the defining equations of the ellipsoids; see Theorem 3.10 in Jia *et al.* (2011).

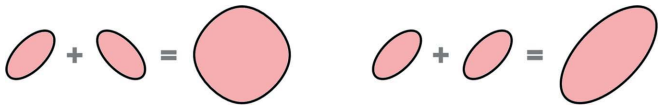


**Figure 5**
Minkowski sums of elliptical bodies.

$$B_R \doteq R \cdot B = \{R\mathbf{x} : \mathbf{x} \in B\}, \quad \text{for} \quad R \in \mathrm{SO}(3).$$

We also write $-B \doteq (-\mathbb{I}) \cdot B$, where $\mathbb{I}$ denotes the identity matrix in SO(3), and thus $-B_R = (-R) \cdot B$. We recall from Proposition 2.3 of Chirikjian & Shiffman (2016) that

$$Z(\gamma, B) = \big\{(R, \mathbf{t}) \in \mathrm{SO}(n) \times F_{\Gamma\backslash X} : (\mathbb{I} - R_\gamma)\mathbf{t} \in (R_\gamma \cdot B_R) + (-B_R) + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma)\big\}. \tag{28}$$

To describe the collision zones $Z(\gamma, B)$, we let $Z_R(\gamma, B)$ denote the set of translations $\mathbf{t} \in F_{\Gamma\backslash X}$ for which $(R, \mathbf{t}) \in Z(\gamma, B)$ for a fixed rotation $R \in \mathrm{SO}(n)$:

$$Z_R(\gamma, B) \doteq \{\mathbf{t} \in F_{\Gamma\backslash X} : (R, \mathbf{t}) \in Z(\gamma, B)\}. \tag{29}$$

Note that it follows from equation (28) that

$$Z_R(\gamma, B) = Z_{\mathbb{I}}(\gamma, B_R), \tag{30}$$

and thus $Z_R(\gamma, B)$ depends only on $\gamma$ and $B_R$.

In cases where $\gamma$ is a translation, *i.e.* when $R_\gamma = \mathbb{I}$, we have

$$\begin{aligned} &Z\big((\mathbb{I}, \mathbf{t}_\gamma), B\big) \\ &= \{(R, \mathbf{t}) \in \mathrm{SO}(n) \times F_{\Gamma\backslash X} : \mathbf{0} \in B_R + (-B_R) + \mathbf{t}_\gamma\} \\ &= \{R \in \mathrm{SO}(n) : \mathbf{t}_\gamma \in R \cdot [B + (-B)]\} \times F_{\Gamma\backslash X}, \end{aligned} \tag{31}$$

and thus

$$Z_R\big((\mathbb{I}, \mathbf{t}_\gamma), B\big) = \begin{cases} F_{\Gamma\backslash X} & \text{if} \quad \mathbf{t}_\gamma \in B_R + (-B_R) \\ \emptyset & \text{if} \quad \mathbf{t}_\gamma \notin B_R + (-B_R) \end{cases}. \tag{32}$$

### 4.1. The collision space in 3D

We now restrict our discussion to our case of interest, $n = 3$. As noted in equation (32) above, if $R_\gamma = \mathbb{I}$, then $Z_R(\gamma, B)$ is either empty or all of $F_{\Gamma\backslash X}$, for each $R \in \mathrm{SO}(3)$. We now describe $Z_R(\gamma, B)$ for symmetries $\gamma$ with $R_\gamma \neq \mathbb{I}$. Suppose that $R_\gamma$ is the (counterclockwise) rotation about the unit vector $\mathbf{n}_\gamma = [n_1, n_2, n_3]^{\mathrm{T}} \in \mathbb{S}^2$ by the angle $\theta_\gamma \in (0, \pi]$. Then

$$R_\gamma = \exp\big(\theta_\gamma N_\gamma\big), \quad N_\gamma = \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}. \tag{33}$$

The skew-symmetric matrix $N_\gamma$ satisfies $N_\gamma \mathbf{a} = \mathbf{n}_\gamma \times \mathbf{a}$ for all $\mathbf{a} \in \mathbb{R}^3$. The 3D case is fundamentally different from the planar case studied by Chirikjian & Shiffman (2016), since in the 3D case $(\mathbb{I} - R_\gamma)$ is never invertible.

To illustrate our approach, suppose that $\mathbf{n}_\gamma = \mathbf{e}_1$. (In fact, one can always choose coordinates so that this is the case.) Then

$$R_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_\gamma & -\sin\theta_\gamma \\ 0 & \sin\theta_\gamma & \cos\theta_\gamma \end{pmatrix}.$$

A key ingredient is the (translated) Minkowski sum from equation (28):

$$M(\gamma, B_R) \doteq (R_\gamma \cdot B_R) + (-B_R) + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma). \tag{34}$$

By equations (28)–(29),

$$\begin{aligned} Z_R(\gamma, B) &= \big\{\mathbf{x} \in F_{\Gamma\backslash X} \ : \ (\mathbb{I} - R_\gamma)\mathbf{x} \in M(\gamma, B_R)\big\} \\ &= \left\{\mathbf{x} \in F_{\Gamma\backslash X} : \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1-\cos\theta_\gamma & \sin\theta_\gamma \\ 0 & -\sin\theta_\gamma & 1-\cos\theta_\gamma \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in M(\gamma, B_R)\right\} \\ &= \left\{\mathbf{x} \in F_{\Gamma\backslash X} : \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\cos\theta_\gamma & \sin\theta_\gamma \\ 0 & -\sin\theta_\gamma & 1-\cos\theta_\gamma \end{pmatrix} \begin{bmatrix} 0 \\ x_2 \\ x_3 \end{bmatrix} \in M(\gamma, B_R)\right\}. \end{aligned} \tag{35}$$

To give a geometric description of $Z_R(\gamma, B)$, we let

$$\Pi_1 \doteq \{\mathbf{x} \in X : x_1 = 0\} = \{0\} \times \mathbb{R}^2$$

denote the $x_2x_3$ plane in $\mathbb{R}^3$. Thus $Z_R(\gamma, B) = D \cap F_{\Gamma\backslash X}$, where $D$ is the cylinder in $\mathbb{R}^3$ with axis $= \mathbf{n}_\gamma = \mathbf{e}_1$ and with base[6]

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{-\sin\theta_\gamma}{2-2\cos\theta_\gamma} \\ 0 & \frac{\sin\theta_\gamma}{2-2\cos\theta_\gamma} & \frac{1}{2} \end{pmatrix} \cdot \big[M(\gamma, B_R) \cap \Pi_1\big]. \tag{36}$$

We see from equation (34) that the base equation (36) of the cylinder $D$ is a planar slice of a Minkowski sum.

A coordinate-free approach to equations (35)–(36) is as follows: let $P_\gamma \doteq \mathbf{n}_\gamma \mathbf{n}_\gamma^{\mathrm{T}}$ denote the orthogonal projection to the line generated by $\mathbf{n}_\gamma$, and let $\Pi_\gamma \doteq \ker P_\gamma$ denote the planar subspace of $\mathbb{R}^3$ orthogonal to $\mathbf{n}_\gamma$. Then $P_\gamma^\perp \doteq \mathbb{I} - P_\gamma$ is the orthogonal projection to $\Pi_\gamma$.

Now suppose that $\mathbf{x} \in Z_R(\gamma, B)$. Since $(\mathbb{I} - R_\gamma)P_\gamma \mathbf{x} = 0$, we have

$$\begin{aligned} (\mathbb{I} - R_\gamma)\mathbf{x} &= (\mathbb{I} - R_\gamma)P_\gamma^\perp \mathbf{x} \\ &= (P_\gamma + \mathbb{I} - R_\gamma)P_\gamma^\perp \mathbf{x} \in M(\gamma, B_R) \cap \Pi_\gamma. \end{aligned}$$

Hence,

$$P_\gamma^\perp \mathbf{x} \in (P_\gamma + \mathbb{I} - R_\gamma)^{-1} \cdot [M(\gamma, B_R) \cap \Pi_\gamma], \tag{37}$$

and $P_\gamma \mathbf{x}$ is arbitrary. In fact, $\mathbf{x} \in Z_R(\gamma, B)$ if and only if $\mathbf{x} \in F_{\Gamma\backslash X}$ and equation (37) holds; *i.e.* $Z_R(\gamma, B) = D_\gamma \cap F_{\Gamma\backslash X}$, where $D_\gamma$ is the cylinder with axis $\mathbf{n}_\gamma$ and base given by equation (37). Note that equation (37) reduces to equation (36) when $\mathbf{n}_\gamma = \mathbf{e}_1$.

[6] The first columns of the matrices in the last line of equation (35) and in equation (36) are arbitrary; the choice $[1\ 0\ 0]^{\mathrm{T}}$ is to simplify the inverse in equation (36) of the matrix in equation (35).

It then follows from equation (17) that the translational part of the feasible space for a given rotation $R$ is contained in a finite union of cylindrical shells. Using equation (26), we can write the translational part (for a rotation $R$) of the informed search space $\mathcal{I}(\Gamma, E_1, E_2)$ as a finite union of finite intersections of cylinders.

### 4.2. Reducing the search space

When $B \subset \mathbb{R}^3$ is described well as an ellipsoid (centered and aligned with the coordinate axes), we have the symmetries $B = k \cdot B$ where $k$ is any of the proper transformations taking $e \cdot \mathbf{x} = \mathbf{x} = (x_1, x_2, x_3)$ into $k_1 \cdot \mathbf{x} \doteq (x_1, -x_2, -x_3)$, $k_2 \cdot \mathbf{x} \doteq (-x_1, x2, -x_3)$, $k_3 \cdot \mathbf{x} \doteq (-x_1, -x_2, x_3)$, which compose the Klein-four group,

$$K \doteq \{e, k_1, k_2, k_3\} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2.$$

Consequently, the collision zones have the symmetries

$$Z(\gamma, B) = Z(\gamma, k_i \cdot B) = Z(\gamma, B)k_i,$$

so the size of the search space $F_{\Gamma\backslash G}$ can be reduced by a factor of four because of this effect alone; that is, the search space can be reduced to $F_{\Gamma\backslash G/K}$.

The search space can be reduced further, even if $B$ does not have rotational symmetries (as is usually the case with biomolecular structures). For example, if $\Gamma$ is of type $P1$, then the feasible space is invariant under translation as can be seen from equation (31), and the intensity function, being the squared magnitude of $\hat{\rho}$, is also translation invariant. Thus in the $P1$ case, the feasible space and intensity depend only on the rotational part $R \in \mathrm{SO}(3)$ of $g = (R, \mathbf{t}) \in G$, and hence the search space is 3D, as is well known. Also, for a space group with cyclic point group (*i.e.* all rotations and screw motions are about the same unit vector $\mathbf{n}$), which is the case for approximately 38% of all protein crystals [see Table 5 in Chirikjian, Sajjadi *et al.* (2015)], the feasible space and intensity are independent of translations along $\mathbf{n}$, and hence the search space is 5D in this case. Another way that the dimension of the search space can be reduced is if $B$ is axisymmetric. Combining these two effects gives a 4D search space.

**4.2.1. Reducing the search space using discrete symmetries.** For any Sohncke space group $\Gamma$, we show in equation (45) below that [for $F_{\Gamma\backslash G}$ as in equation (39)]

$$\hat{\gamma} Z(\gamma, B) = Z(\gamma', B) \quad \gamma' = \hat{\gamma}\gamma\hat{\gamma}^{-1},$$

for all orientation-preserving $\hat{\gamma}$ in the Euclidean normalizer $\hat{\Gamma}$ of $\Gamma$, which leads to invariance of the search space under left multiplication by $\hat{\Gamma} \cap G$. If the point group of $\Gamma$ is not cyclic, then $\hat{\Gamma}$ is a (not necessarily Sohncke) crystallographic group, which is described by Hirshfeld (1968) and Koch *et al.* (2002) and in the Bilbao server (Aroyo *et al.*, 2006), and this symmetry leads to a potential reduction in the size of the search space by a factor of $[\hat{\Gamma} \cap G : \Gamma]$. If $B$ is an ellipsoid, by combining these two factors, one can thus reduce the search space to $F_{(\hat{\Gamma}\cap G)\backslash G/K}$.

Moreover, an ellipsoidal body $B$ has the additional reflection symmetries $q_1 \cdot \mathbf{x} = (-x_1, x_2, x_3)$, $q_2 \cdot \mathbf{x} = (x_1, -x_2, x_3)$, $q_3 \cdot \mathbf{x} = (x_1, x_2, -x_3)$, $q_4 \cdot \mathbf{x} = (-x_1, -x_2, -x_3)$, and so the symmetry group of $B$ is the full orthorhombic point group

$$K \cup \{q_1, q_2, q_3, q_4\} = K \cup (-\mathbb{I})K < \mathrm{O}(3) \tag{38}$$

of order eight, assuming that the semiaxes of $B$ are distinct. (Though macromolecular structures are chiral and hence do not have these reflection symmetries, if they can be bounded by ellipsoids, then these reflection symmetries become applicable to the bounding bodies.) If $\hat{\Gamma}$ is not Sohncke, the search space can then be reduced by an additional factor of two due to symmetries of the form $\hat{\gamma} Z(\gamma, B) q_i = Z(\gamma', B)$ for orientation-reversing $\hat{\gamma} \in \hat{\Gamma}$. In fact, it is even possible for $B$ to have a symmetry group with more symmetry elements. For example, $B$ could be a hexagonal ring structure consisting of additional dihedral symmetry operations. Or it might be a viral capsid with icosahedral symmetry. Then the above reductions occur with equation (38) replaced by the symmetry group of $B$.

For example, if $\Gamma$ is in the most common biomolecular crystal class $P2_12_12_1$, then $[\hat{\Gamma} : \Gamma] = 16$ (see Section 7.2), which results in a reduction by a factor of $[\hat{\Gamma} \cap G : \Gamma] = 8$. If in addition the body $B$ is ellipsoidal, then the combination of these effects leads to a reduction in the size of the search space by a factor of 64. The reason for the factor of 64 rather than $32 = |K|[\hat{\Gamma} \cap G : \Gamma]$ is the additional twofold symmetry of $G$ under conjugation by $(-\mathbb{I}, \mathbf{0})$.

We state our reduction result for crystals of any dimension:

*Theorem 4.1.* Let $\Gamma < G = \mathrm{SE}(n)$ be a crystallographic group, and let $B$ be a bounded connected open set in $X = \mathbb{R}^n$. Suppose there exist a finite group $Q < \mathrm{O}(n)$ and a crystallographic group $\hat{\Gamma} < \mathrm{E}(n)$ such that

(i) $q \cdot B = B \quad \forall q \in Q$;

(ii) $\Gamma \triangleleft \hat{\Gamma}$.

Let $F_{\Gamma\backslash\hat{\Gamma}}$ be a fundamental set for $\Gamma\backslash\hat{\Gamma}$. Then there exists a set $F^0 \subset G$ such that

$$F_{\Gamma\backslash G} \doteq \bigcup \{\hat{\gamma} F^0 q : q \in Q, \hat{\gamma} \in F_{\Gamma\backslash\hat{\Gamma}}, \hat{\gamma} q \in G\} \tag{39}$$

is a fundamental domain for $\Gamma\backslash G$,

$$\mathcal{A}(\Gamma, B) = \bigcup \{\hat{\gamma}[F^0 \cap \mathcal{A}(\Gamma, B)]q : q \in Q, \hat{\gamma} \in F_{\Gamma\backslash\hat{\Gamma}}, \hat{\gamma} q \in G\}, \tag{40}$$

and

$$\frac{\mathrm{vol}(F_{\Gamma\backslash G})}{\mathrm{vol}(F^0)} = \begin{cases} |Q|[\hat{\Gamma} : \Gamma], & \text{if} \quad \hat{\Gamma} < G, Q < \mathrm{SO}(n) \\ \frac{1}{2}|Q|[\hat{\Gamma} : \Gamma], & \text{if} \quad \hat{\Gamma} \not< G \text{ or } Q \not< \mathrm{SO}(n) \end{cases} \tag{41}$$

where 'vol' denotes invariant volume in G.

Equation (40) tells us that to determine the points of the feasible space $\mathcal{A}(\Gamma, B)$, one needs only to search $F^0$. Thus by equation (41), the volume of the motion space to be searched to find the points of $\mathcal{A}(\Gamma, B)$ is reduced by a factor of $|Q|[\hat{\Gamma} : \Gamma]$ if both $Q$ and $\hat{\Gamma}$ preserve orientation, and by half this factor otherwise.

We define $F^0$ in the proof below as a fundamental domain of an action on $\mathrm{E}(n)$ [or as the component in $\mathrm{SE}(n)$ of the fundamental domain]. These fundamental domains are easy to

construct in 2D and 3D; an example for the case of a crystallographic group of type $P2_12_12_1$ is given in Section 7.2. If both $Q$ and $\hat{\Gamma}$ contain orientation-reversing elements, then allowing for these elements provides an additional reduction by a factor of two, as can be seen from equation (41). [The condition $\hat{\gamma}q \in G$ in equations (39)–(40) is satisfied when $\hat{\gamma}$ and $g$ both preserve orientation or both reverse orientation.]

The symmetry of equation (40) in the case of a wallpaper group $\Gamma < \mathrm{SE}(2)$ of type $p2$ with a rectangular unit cell can be seen in Fig. 4. Let $\mathbb{L}$ denote the (orthogonal) lattice of $\Gamma$. One can easily check that its normalizer $\hat{\Gamma}$ is of type $pmm$ with the finer lattice $\frac{1}{2}\mathbb{L}$, and thus $[\hat{\Gamma} : \Gamma] = 8$. Since the symmetry of an ellipse is the dihedral group $D_2 \cong K$,

$$\mathrm{vol}(F_{\Gamma\backslash \mathrm{SE}(2)})/\mathrm{vol}(F^0) = \tfrac{1}{2}|D_2|[\hat{\Gamma} : \Gamma] = 16.$$

But the only left–right $(\hat{\gamma}, q)$ actions on $g = (R, \mathbf{t}) \in \mathrm{SE}(2)$ that leave $R$ invariant are left actions by $\hat{\Gamma} \cap \mathrm{SE}(2)$. Since $[\hat{\Gamma} \cap \mathrm{SE}(2) : \Gamma] = [\frac{1}{2}\mathbb{L} : \mathbb{L}] = 4$, we obtain the fourfold translation symmetry in the unit cell visible in Fig. 4.[7]

The proof of Theorem 4.1 makes use of the elementary general lemma below. We recall that when a discrete group $\Delta$ acts properly discontinuously on a manifold $\mathbf{M}$ such that the fixed point set of each action $g \in \Delta \setminus \{e\}$ has measure zero,[8] we can find a fundamental domain whose images under the action of the group cover $\mathbf{M}$ with overlaps of measure zero [see, for example, Section 3.1 of Chirikjian *et al.* (2017)].

*Lemma 4.2*. Let $\hat{\Delta}$ be a discrete group acting properly discontinuously on a manifold $\mathbf{M}$ with fixed point sets of measure zero. Let $\Delta < \hat{\Delta}$ and let $A \doteq F_{\Delta\backslash\hat{\Delta}}$ be a fundamental set for $\Delta\backslash\hat{\Delta}$. If $F_{\hat{\Delta}\backslash\mathbf{M}}$ is a fundamental domain for $\hat{\Delta}$ acting on $\mathbf{M}$, then the sets $\{aF_{\hat{\Delta}\backslash\mathbf{M}} : a \in A\}$ have pairwise intersections of measure zero, and their union is a fundamental domain for $\Delta$ acting on $\mathbf{M}$.

*Proof*. Since each element $\hat{\delta} \in \hat{\Delta}$ has a unique decomposition as a product of $\delta \in \Delta$ and $a \in A$, we have

$$\mathbf{M} = \bigcup_{\hat{\delta}\in\hat{\Delta}} \hat{\delta} F_{\hat{\Delta}\backslash\mathbf{M}} = \bigcup_{\delta\in\Delta} \delta\left(\bigcup_{a\in A} aF_{\hat{\Delta}\backslash\mathbf{M}}\right),$$

where the above unions are pairwise disjoint outside a set of measure zero. □

*Proof of Theorem 4.1*. We shall apply Lemma 4.2 with $\mathbf{M} \doteq \mathrm{E}(n)$ and

$$\Delta \doteq \Gamma \times \{\mathbb{I}\} \ \triangleleft\ \hat{\Delta} \doteq \hat{\Gamma} \times Q$$

acting on $g \in \mathrm{E}(n)$ by

$$(\hat{\gamma}, q) \bullet g \doteq \hat{\gamma} g q^{-1}, \quad \hat{\gamma} \in \hat{\Gamma}, \quad q \in Q.$$

Writing $g = (R, \mathbf{t})$, we have

$$(\hat{\gamma}, q) \bullet g = \big((R_{\hat{\gamma}}, \mathbf{t}_{\hat{\gamma}}), (q, 0)\big) \bullet (R, \mathbf{t}) = (R_{\hat{\gamma}} R q^{-1}, R_{\hat{\gamma}}\mathbf{t} + \mathbf{t}_{\hat{\gamma}}). \tag{42}$$

It follows from equation (42) that the $\hat{\Delta}$ action is properly discontinuous with fixed point sets of measure 0. Hence we can choose a fundamental domain $F_{\hat{\Delta}\backslash\mathrm{E}(n)} = F_{\hat{\Gamma}\backslash\mathrm{E}(n)/Q}$ for $\hat{\Delta}$ acting on $\mathrm{E}(n)$. If $Q \not< \mathrm{SO}(n)$ or $\hat{\Gamma} \not< G$, we can choose $F_{\hat{\Delta}\backslash\mathrm{E}(n)} \subset G$ (since in this case every orbit intersects $G$) and we let $F^0 = F_{\hat{\Delta}\backslash\mathrm{E}(n)}$. On the other hand, if $\hat{\Gamma} < G$ and $Q < \mathrm{SO}(n)$, we let $F^0 = F_{\hat{\Delta}\backslash\mathrm{E}(n)} \cap G$ so that $F_{\hat{\Delta}\backslash\mathrm{E}(n)} = F^0 \cup (-\mathbb{I}) \cdot F^0$.

We choose a fundamental set for $\Delta\backslash\hat{\Delta}$ of the form $F_{\Delta\backslash\hat{\Delta}} \doteq F_{\Gamma\backslash\hat{\Gamma}} \times Q$. Then, by Lemma 4.2,

$$F_{\Delta\backslash\mathrm{E}(n)} \doteq \bigcup\left\{\hat{\gamma} F_{\hat{\Delta}\backslash\mathrm{E}(n)} q^{-1} : (\hat{\gamma}, q) \in F_{\Gamma\backslash\hat{\Gamma}} \times Q\right\} \tag{43}$$

is a fundamental domain for $\Delta\backslash\mathrm{E}(n)$ (where the above union is disjoint outside a set of measure zero). Then

$$F_{\Gamma\backslash G} \doteq F_{\Gamma\backslash\mathrm{E}(n)} \cap G = F_{\Delta\backslash\mathrm{E}(n)} \cap G$$

is a fundamental domain for $\Gamma\backslash G$ and satisfies equation (39). Since $F_{\Delta\backslash\mathrm{E}(n)} = F_{\Gamma\backslash G} \cup (-\mathbb{I}) \cdot F_{\Gamma\backslash G}$, we have

$$\mathrm{vol}(F_{\Gamma\backslash G}) = \tfrac{1}{2}\mathrm{vol}(F_{\Delta\backslash\mathrm{E}(n)}).$$

By equation (43),

$$\frac{\mathrm{vol}(F_{\Delta\backslash\mathrm{E}(n)})}{\mathrm{vol}(F_{\hat{\Delta}\backslash\mathrm{E}(n)})} = [\hat{\Delta} : \Delta] = [\hat{\Gamma} : \Gamma]|Q|. \tag{44}$$

Formula (41) follows from equation (44), since $\mathrm{vol}(F^0) = \frac{1}{2}\mathrm{vol}(F_{\hat{\Delta}\backslash\mathrm{E}(n)})$ if $\hat{\Gamma} < G$, $Q < \mathrm{SO}(n)$, and $F^0 = F_{\hat{\Delta}\backslash\mathrm{E}(n)}$ otherwise.

To complete the proof, we must verify equation (40). Let $\hat{\gamma} \in \hat{\Gamma}, q \in Q$ such that $\hat{\gamma}q \in G$. Since $q^{-1} \cdot B = B$ and $\Gamma$ is normal in $\hat{\Gamma}$,

$$\begin{aligned} g \in Z(\gamma, B) &\iff g \cdot B \cap \gamma g \cdot B \neq \emptyset \\ &\iff \hat{\gamma} g q^{-1} \cdot B \cap \hat{\gamma}\gamma\hat{\gamma}^{-1}\hat{\gamma} g q^{-1} \cdot B \neq \emptyset \\ &\iff \hat{\gamma} g q^{-1} \in Z(\gamma', B), \gamma' = \hat{\gamma}\gamma\hat{\gamma}^{-1} \in \Gamma. \end{aligned}$$

Therefore,

$$\hat{\gamma} Z(\gamma, B) q^{-1} \in Z(\gamma', B), \quad \gamma' = \hat{\gamma}\gamma\hat{\gamma}^{-1} \in \Gamma. \tag{45}$$

It follows from equation (45) with $B$ replaced by $\overline{B}$ that

$$\hat{\gamma}\Lambda(g \cdot \overline{B})\hat{\gamma}^{-1} = \Lambda(g' \cdot \overline{B}), \quad g' = (\hat{\gamma}, q) \bullet g. \tag{46}$$

Let

$$\widehat{\mathcal{A}}(\Gamma, B) = \big\{g \in G : \Lambda(g \cdot B) = \emptyset, \Lambda(g \cdot \overline{B}) \text{ generates } \Gamma\big\}, \tag{47}$$

so that

$$\mathcal{A}(\Gamma, B) = \widehat{\mathcal{A}}(\Gamma, B) \cap F_{\Gamma\backslash G}.$$

Since $\Lambda(g \cdot \overline{B})$ generates $\Gamma$ if and only if $\hat{\gamma}\Lambda(g \cdot \overline{B})\hat{\gamma}^{-1}$ generates $\Gamma$, it follows from equations (18), (20) and (46) that $\widehat{\mathcal{A}}(\Gamma, B)$ is invariant under the $(\hat{\gamma}, q)$ action. Equation (40) then follows from equation (39) and the invariance of $\widehat{\mathcal{A}}(\Gamma, B)$. □

[7] Figs. 4(*a*) and 4(*c*) exhibit additional mirror symmetries (resulting in a 16-fold symmetry in the unit cell), since the ellipses are positioned in these figures to be invariant under the mirror reflections.

[8] This is always the case when the action is effective and real-analytic.

**4.2.2. Search space reduction and Laue symmetry.** Theorem 4.1 together with the Laue symmetry of the diffraction pattern (4) shows that the volume of the space needed to compute the cost function [equation (5)] can likewise be reduced, as a consequence of the following result:[9]

*Theorem 4.3.* Let $\Gamma \lhd \hat{\Gamma}$ be crystallographic groups, and let $\mathbb{P}$ denote the point group of $\Gamma$. Suppose that the point group of $\hat{\Gamma}$ is either $\mathbb{P}$ or $\mathbb{P} \cup (-\mathbb{P})$. Then the Patterson function and intensity function of a crystal with symmetry group $\Gamma$ satisfy the group invariance properties

$$\mathcal{P}(\mathbf{x}; \hat{\gamma} g q) = \mathcal{P}(\mathbf{x}; g), I(\mathbf{k}; \hat{\gamma} g q) = I(\mathbf{k}; g), \quad \forall\, q \in Q, \hat{\gamma} \in \hat{\Gamma}. \tag{48}$$

*Proof.* Let $\hat{\gamma} \in \hat{\Gamma}$ be given and write $\hat{\gamma}^{-1} \alpha \hat{\gamma} = \gamma' \in \Gamma$ for $\alpha \in \Gamma$. Recalling equation (1), we have

$$\begin{aligned} \rho(\mathbf{x}; \hat{\gamma} g) &= \sum_{\gamma \in \Gamma} f\big((\gamma \hat{\gamma} g)^{-1} \cdot \mathbf{x}\big) \\ &= \sum_{\gamma' \in \Gamma} f\big((\gamma' g)^{-1} \cdot (\hat{\gamma}^{-1} \cdot \mathbf{x})\big) = \rho(\hat{\gamma}^{-1} \cdot \mathbf{x}; g). \end{aligned} \tag{49}$$

Therefore,

$$\begin{aligned} \mathcal{P}(\mathbf{x}; \hat{\gamma} g) &= \int_{F_{T\backslash X}} \rho(\hat{\gamma}^{-1} \cdot \mathbf{y}; g) \rho(\hat{\gamma}^{-1} \cdot \mathbf{y} + R_{\hat{\gamma}}^{-1} \cdot \mathbf{x}; g)\, \mathrm{d}\mathbf{y} \\ &= \int_{F'_{T\backslash X}} \rho(\mathbf{z}; g) \rho(\mathbf{z} + R_{\hat{\gamma}}^{-1} \cdot \mathbf{x}; g)\, \mathrm{d}\mathbf{z} \\ &= \mathcal{P}(R_{\hat{\gamma}}^{-1} \cdot \mathbf{x}; g) = \mathcal{P}(\mathbf{x}; g), \end{aligned}$$

where $F'_{T\backslash X} = \hat{\gamma}^{-1} \cdot F_{T\backslash X}$. The final equality follows from equation (4) and the hypothesis that $R_{\hat{\gamma}} \in \mathbb{P} \cup (-\mathbb{P})$.

Invariance of the Patterson function under right multiplication by $q \in Q$ is an immediate consequence of the $Q$ symmetry of $B$. By taking the Fourier transform of the Patterson function, one then obtains the stated invariance of the intensity. □

When equations (40) of Theorem 4.1 and (48) of Theorem 4.3 hold, one only needs to consider motions $g$ in the domain $F^0$ in order to compute the cost function [equation (5)] as well as to determine the feasible space [equation (12)]. When this occurs, we say that *the motion space search can be reduced by a factor of* $\mathrm{vol}(F_{\Gamma\backslash G})/\mathrm{vol}(F^0)$.

The five most common space groups for biomolecular crystals are $P2_12_12_1$, $P2_1$, $C2$, $P2_12_12$ and $C222_1$ (Chirikjian, Sajjadi *et al.*, 2015). For these groups, Theorem 4.1 yields substantial reductions. For groups $\Gamma$ in the classes $P2_12_12_1$, $P2_12_12$ and $C222_1$, we take $\hat{\Gamma}$ to be the normalizer of $\Gamma$ in E(3). For $\Gamma$ of type $P2_12_12_1$ or $P2_12_12$ with ellipsoidal bodies, $[\hat{\Gamma} : \Gamma] = 16$ and the motion space search can be reduced by a factor of 64; for $\Gamma$ of type $C222_1$, the reduction factor is 32. The 5D search spaces for $P2_1$ and $C2$ can be reduced by a factor of 32 and 16, respectively (for ellipsoidal bodies).

[9] See Hirshfeld (1968) and Koch *et al.* (2002). The Euclidean normalizer $\hat{\Gamma}$ is also called the *Cheshire group* of $\Gamma$.

We note that the argument in Theorem 4.1 also yields the following result:

*Theorem 4.4.* Suppose that $E_1 \subset B \subset E_2$. Assume the hypotheses of Theorem 4.1 with (i) replaced by

$$q \cdot E_1 = E_1, q \cdot E_2 = E_2 \quad \forall q \in Q.$$

Then the conclusions of Theorem 4.1 hold with $\mathcal{A}(\Gamma, B)$ replaced by $\mathcal{I}(\Gamma, E_1, E_2)$.

Thus, if an irregular molecule $B$ is sandwiched between ellipsoids $E_1$ and $E_2$, the informed search space $\mathcal{I}(\Gamma, E_1, E_2)$ will have the (right) $Q$ symmetries as well as the (left) $\hat{\Gamma}$ symmetries, and the cost function for the molecule $B$ will have $\hat{\Gamma}$ symmetries. Both symmetries will allow reductions of the computations in the 6D search.

## 5. The crystallographic group *P*1

We now consider the case where $\Gamma$ is in the class $P1$, which is the symmetry group made up only of translations by elements of the lattice $\mathbb{L}$.

In this case, $\Gamma = T$, $\mathbb{P} = \{\mathbb{I}\}$, $\mathbf{v}(\mathbb{I}) = \mathbf{0}$ and $F_{\Gamma\backslash X} = F_{T\backslash X}$ is a unit cell. Then

$$M((\mathbb{I}, \mathbf{t}_\gamma), B_R) = B_R + (-B_R) + \mathbf{t}_\gamma, \tag{50}$$

$$\begin{aligned} \Lambda(B_R) &= \{(\mathbb{I}, \mathbf{t}_\gamma) \in T \backslash \{e\} : B_R \cap (B_R + \mathbf{t}_\gamma) \neq \emptyset\} \\ &= \{\mathbb{I}\} \times \{\mathbf{t}_\gamma \in \mathbb{L} \backslash \{\mathbf{0}\} : \mathbf{t}_\gamma \in B_R + (-B_R)\}, \end{aligned} \tag{51}$$

$$\begin{aligned} Z[(\mathbb{I}, \mathbf{t}_\gamma), B] &= \{R \in \mathrm{SO}(n) : \mathbf{0} \in M((\mathbb{I}, \mathbf{t}_\gamma), B_R)\} \times F_{\Gamma\backslash X} \\ &= \{R \in \mathrm{SO}(n) : \mathbf{t}_\gamma \in R \cdot [B + (-B)]\} \times F_{\Gamma\backslash X}. \end{aligned} \tag{52}$$

Thus the collision space is given by

$$\begin{aligned} S(\Gamma, B) &= \bigcup_{\mathbf{t}_\gamma \in \mathbb{L} \backslash \{\mathbf{0}\}} Z((\mathbb{I}, \mathbf{t}_\gamma), B) \\ &= \big\{R \in \mathrm{SO}(3) : \big[\mathbb{L} \backslash \{\mathbf{0}\}\big] \cap R \cdot [B + (-B)] \neq \emptyset\big\} \\ &\quad \times F_{\Gamma\backslash X}. \end{aligned}$$

The condition above means that for collision at least one nonzero lattice point must be inside $B_R + (-B_R)$.

The informed search space for approximating sets $E_1 \subseteq B \subseteq E_2$ (where $E_1, E_2$ are ellipsoids, for example) is given by

$$\begin{aligned} \mathcal{I}(\Gamma, E_1, E_2) = \Big\{ &R \in \mathrm{SO}(3) : [\mathbb{L} \backslash \{\mathbf{0}\}] \cap R \cdot [E_1 + (-E_1)] = \emptyset, \\ &\mathbb{L} \cap R \cdot [\overline{E}_2 + (-\overline{E}_2)] \text{ generates } \mathbb{L} \Big\} \times F_{T\backslash X}. \end{aligned} \tag{53}$$

It follows from equation (51) that

$$\Lambda(g \cdot \overline{B}) \subseteq \widetilde{\Phi}(\Gamma, B) \doteq \{\mathbf{t}_\gamma \in \Gamma \backslash \{e\} : |\mathbf{t}_\gamma| \leq 2 r_B\}, \tag{54}$$

which is sharper than equation (24) [in fact, we can replace $2r_B$ with $r_{B+(-B)}$]. Thus, to verify equation (53), one only needs to check lattice points $\mathbf{t}_\gamma$ of norm $|\mathbf{t}_\gamma| < 2r_{E_1}$ for the first condition and $|\mathbf{t}_\gamma| \leq 2r_{E_2}$ for the second condition.

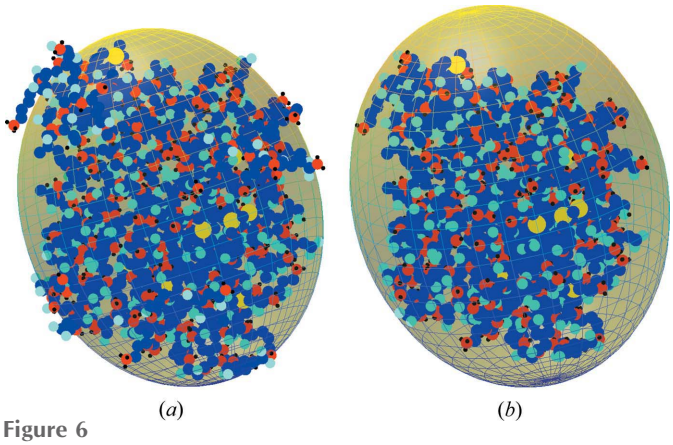


Figure 6
Ellipsoids for the PDB entry 1aky. (*a*) Moment-of-inertia ellipsoid; (*b*) minimum-volume ellipsoid.

We note that if $E_1$ and $E_2$ are (centered) ellipsoids, then $E_1 + (-E_1) = 2E_1$ and $\overline{E}_2 + (-\overline{E}_2) = 2\overline{E}_2$. Furthermore, one sees from equation (53) that equation (21) holds in the $P1$ case if $E_1$ and $E_2$ are ellipsoids that satisfy the weaker Minkowski-sum inclusion relations:

$$2E_1 \subseteq B+(-B) \subseteq 2E_2. \tag{55}$$

This observation allows us to choose better approximating ellipsoids, as illustrated in the next section.

## 6. Detailed examples with *P*1 symmetry

In Section 3, we discussed how ellipsoids can be used to replace protein molecules in the feasible region analysis. In this section we give examples of proteins that crystallize with $P1$ symmetry and their approximating ellipsoids. In these examples, ellipsoids are seen to give better approximations for the Minkowski sum $B + (-B)$ in Section 5 than for the protein.

### 6.1. Ellipsoidal proxies

Here we consider two proxy ellipsoids: the *moment-of-inertia ellipsoid*, which has the same moment of inertia as the protein,[10] and the *minimum-volume ellipsoid* (Todd, 2016), which is the ellipsoid of least volume that contains all the atoms of the protein.

Fig. 6(*a*) shows the atoms of the Protein Data Bank (PDB, http://www.rcsb.org/pdb/) entry 1aky (Abele & Schulz, 1995) (an adenylate kinase), which crystallizes with $P1$ symmetry, and its moment-of-inertia ellipsoid. As can be seen from the figure, the moment-of-inertia ellipsoid coincides with the protein in a high proportion, though there is empty space inside the ellipsoid and some atoms are outside of the ellipsoid. The minimum-volume ellipsoid of 1aky is shown in Fig. 6(*b*). It can be seen that the minimum ellipsoid is significantly larger than the moment-of-inertia ellipsoid since it must contain all atoms of the protein.

[10] The moment-of-inertia ellipsoid is assumed to have the same total mass as the protein and to have uniform mass density.

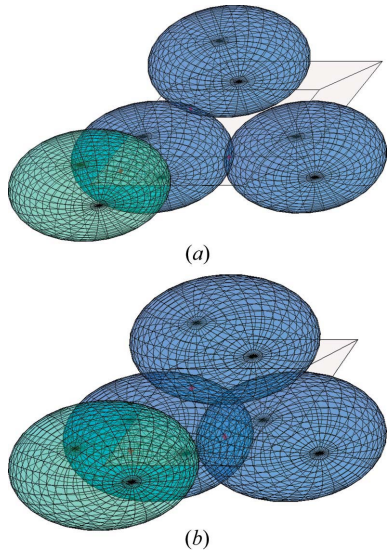


Figure 7
Placement of ellipsoids for the PDB entry 1aky. (*a*) Moment-of-inertia ellipsoid; (*b*) minimum-volume ellipsoid.

Figs. 7(*a*) and 7(*b*) show the placements of the moment-of-inertia ellipsoids and the minimum-volume ellipsoids, respectively, of four symmetry mates of 1aky intersecting a unit cell (outlined). It can be seen from Fig. 7(*a*) that the moment-of-inertia ellipsoids can be placed with only minor collisions with their neighbors in the lattice, whereas the larger minimum-volume ellipsoids in Fig. 7(*b*) collide more. Therefore, compared with the minimum-volume ellipsoid, the moment-of-inertia ellipsoid may provide a better choice for replacing the protein in further analysis.

### 6.2. Minkowski sums

The Minkowski sum determines the collision zone of two bodies. Comparisons between three proteins with $P1$ symmetry [1aky (Abele & Schulz, 1995), 1tt8 (Smith *et al.*, 2006), 5m0e (Keune *et al.*, 2017)] and their moment-of-inertia ellipsoids are given in Table 3. Similar comparisons are given in Table 4 for the corresponding Minkowski sums $B + (-B)$, where $B \subset \mathbb{R}^3$ represents the protein body. One sees from Tables 3–4 that the ellipsoidal approximation of the Minkowski sum $B + (-B)$ is better than the ellipsoidal approximation of the protein body $B$ in these cases.

To describe the notation in the tables, we let $E$ denote the moment-of-inertia ellipsoid, and we note that the Minkowski sum $E + (-E) = 2E$.

$P_a$ – the fraction of atoms of the protein that are outside of the ellipsoid $E$.

$P_{in}$ – the fraction of the volume of $E$ occupied by protein atoms (the total volume of 'boxes' occupied by atoms, the edge dimensions of the box being the average distance between adjacent alpha carbons) inside $E$.

**Table 3**
Sample proteins in $P1$.

| Protein | $P_a$ | $P_{in}$ | $P_{out}$ |
|---|---|---|---|
| 1aky | 0.1038 | 0.6490 | 0.0733 |
| 1tt8 | 0.0933 | 0.8253 | 0.1049 |
| 5m0e | 0.1212 | 0.4651 | 0.0670 |

**Table 4**
Minkowski sums for sample proteins in $P1$.

| Protein | $M_a$ | $M_{in}$ | $M_{out}$ |
|---|---|---|---|
| 1aky | 0.0023 | 0.9212 | 0.0911 |
| 1tt8 | 0.0019 | 0.9676 | 0.1262 |
| 5m0e | 0.0083 | 0.9384 | 0.1633 |

$P_{out}$ – the volume of the space occupied by protein atoms outside $E$, divided by the volume of $E$.

$M_a$ – the fraction of the volume of the Minkowski sum $B+(-B)$ that is outside of $2E$.

$M_{in}$ – the volume of the portion of the Minkowski sum $B+(-B)$ inside $2E$, divided by the volume of $2E$.

$M_{out}$ – the volume of the portion of the Minkowski sum $B+(-B)$ outside of $2E$, divided by the volume of $2E$.

The Minkowski sum $B+(-B)$ for the body $B$ corresponding to the PDB entry 1aky is displayed in Fig. 8. The ellipsoids in the figure are the ellipsoid $2E$ (where $E$ is the moment-of-inertia ellipsoid of $B$) and its further scaling by a factor of 1.25.

Slices of the Minkowski sum of Fig. 8 are shown in Fig. 9. The dotted ellipses, the outer ellipses and the inner ellipses delineate the slices of $2E$ and its further scaling by factors of 1.25 and 0.75, respectively.

Fig. 10 shows the atoms of the PDB entry 5m0e and its moment-of-inertia ellipsoid. The corresponding Minkowski sum, ellipsoids and slices for 5m0e are displayed in Figs. 11–12.

**Figure 8**
Minkowski sum for the PDB entry 1aky.

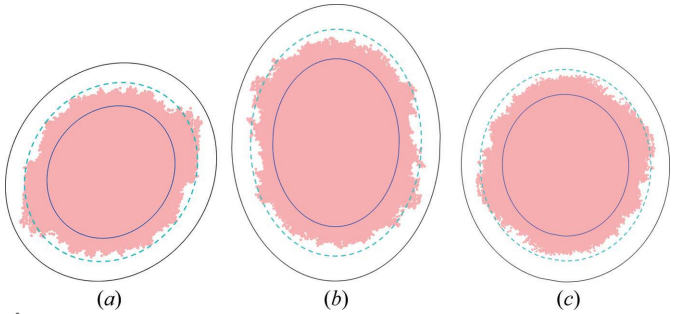


**Figure 9**
Slices of the Minkowski sum for the PDB entry 1aky. (*a*) Slice with the $x_1x_2$ plane; (*b*) slice with the $x_1x_3$ plane; (*c*) slice with the $x_2x_3$ plane.

## 7. The crystallographic group $P2_12_12_1$

We now consider the case where $\Gamma$ is in the class $P2_12_12_1$, which is the most common type of crystallographic group formed by proteins. In this case, we can choose coordinates so that the lattice $\mathbb{L}$ is generated by $(a_1, 0, 0)$, $(0, a_2, 0)$, $(0, 0, a_3)$, where we make the identification

$$(x_1, x_2, x_3) \equiv (x_1, x_2, x_3)^{\mathrm{T}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3.$$

We choose the (centered) unit cell

$$F_{T\backslash X} \doteq \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : -\frac{a_i}{2} \leq x_i < \frac{a_i}{2}, 1 \leq i \leq 3\right\}$$

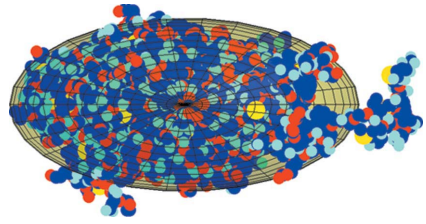


**Figure 10**
The PDB entry 5m0e and its moment-of-inertia ellipsoid.

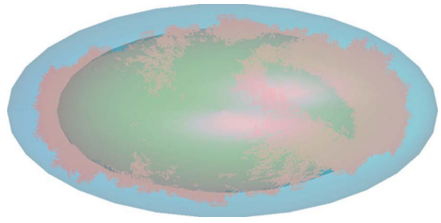


**Figure 11**
Minkowski sum for the PDB entry 5m0e.

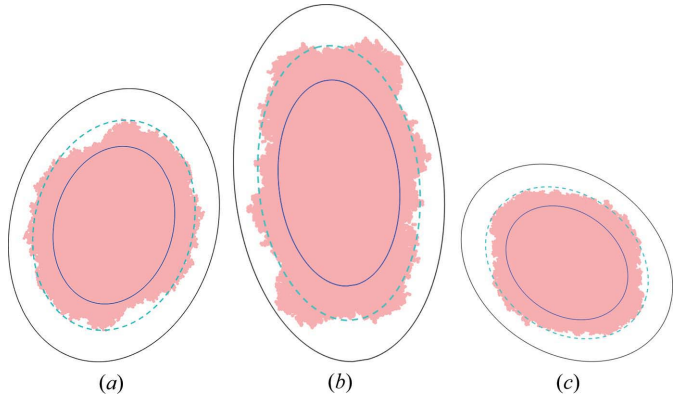


**Figure 12**
Slices of the Minkowski sum for the PDB entry 5m0e. (*a*) Slice with the $x_1x_2$ plane; (*b*) slice with the $x_1x_3$ plane; (*c*) slice with the $x_2x_3$ plane.

and fundamental domain $F_{\Gamma\backslash G} = \mathrm{SO}(3) \times F_{\Gamma\backslash X}$, with (centered) asymmetric unit

$$F_{\Gamma\backslash X} \doteq \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : \right.$$
$$\left. -\frac{a_1}{2} \le x_1 < \frac{a_1}{2}, -\frac{a_2}{4} \le x_2 < \frac{a_2}{4}, -\frac{a_3}{4} \le x_3 < \frac{a_3}{4}\right\}. \tag{56}$$

Then the point group $T\backslash\Gamma$ is isomorphic to $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ (the Klein group), and we can write $F_{T\backslash\Gamma} = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$ where each element of $F_{T\backslash\Gamma}$ acts on $X = \mathbb{R}^3$ as $\gamma_0 \cdot (x_1, x_2, x_3) = (x_1, x_2, x_3)$; $\gamma_1 \cdot (x_1, x_2, x_3) = (x_1 + a_1/2, -x_2 + a_2/2, -x_3)$; $\gamma_2 \cdot (x_1, x_2, x_3) = (-x_1, x_2 + a_2/2, -x_3 + a_3/2)$; $\gamma_3 \cdot (x_1, x_2, x_3) = (-x_1 + a_1/2, -x_2, x_3 + a_3/2)$.

In terms of homogeneous transformation matrices, these operators are

$$\gamma_0 = e = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \ \gamma_1 = \begin{pmatrix} 1 & 0 & 0 & \frac{a_1}{2} \\ 0 & -1 & 0 & \frac{a_2}{2} \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$
$$\gamma_2 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{a_2}{2} \\ 0 & 0 & -1 & \frac{a_3}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}; \ \gamma_3 = \begin{pmatrix} -1 & 0 & 0 & \frac{a_1}{2} \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & \frac{a_3}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

In this case $\mathbf{n}_{\gamma_i} = \mathbf{e}_i$, for $i = 1, 2, 3$, where $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is the standard orthonormal basis for $\mathbb{R}^3$. Furthermore

$$\mathbf{v}(R_{\gamma_1}) = \frac{a_1}{2}\mathbf{e}_1 + \frac{a_2}{2}\mathbf{e}_2, \quad \mathbf{v}(R_{\gamma_2}) = \frac{a_2}{2}\mathbf{e}_2 + \frac{a_3}{2}\mathbf{e}_3,$$
$$\mathbf{v}(R_{\gamma_3}) = \frac{a_3}{2}\mathbf{e}_3 + \frac{a_1}{2}\mathbf{e}_1.$$

Let $\tau_{\mathbf{k}} \in T \triangleleft \Gamma$ denote translation by the lattice point $(k_1a_1, k_2a_2, k_3a_3)^{\mathrm{T}}$, where $\mathbf{k} = (k_1, k_2, k_3) \in \mathbb{Z}^3$. Then

$$\Gamma = \{\tau_{\mathbf{k}}\gamma_i : \mathbf{k} \in \mathbb{Z}^3, \ i = 0, 1, 2, 3\}.$$

We now construct the (translational) collision zones $Z_R(\tau_{\mathbf{k}}\gamma_i, B)$. For $i = 0$, we have $\tau_{\mathbf{k}}\gamma_0 = \tau_{\mathbf{k}}$, and thus by equation (52), we have

$$Z_R(\tau_{\mathbf{k}}, B) = \begin{cases} F_{\Gamma\backslash X} & \text{for } (k_1a_1, k_2a_2, k_3a_3)^{\mathrm{T}} \in B_R + (-B_R) \\ \emptyset & \text{for } (k_1a_1, k_2a_2, k_3a_3)^{\mathrm{T}} \notin B_R + (-B_R) \end{cases}. \tag{57}$$

Next we let $i = 1$; in this case, we have $\mathbf{n}_{\gamma_1} = \mathbf{e}_1$ and

$$R_{\gamma_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \qquad \mathbb{I} - R_{\gamma_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

We let

$$M_1(B_R) \doteq R_{\gamma_1} \cdot B_R + (-B_R) = \left\{ \begin{bmatrix} x_1 - y_1 \\ -x_2 - y_2 \\ -x_3 - y_3 \end{bmatrix} : \mathbf{x}, \mathbf{y} \in B_R \right\}. \tag{58}$$

Recalling equation (34), we have

$$M(\tau_{\mathbf{k}}\gamma_1, B_R) = M_1(B_R) + \begin{bmatrix} (k_1 + \frac{1}{2})a_1 \\ (k_2 + \frac{1}{2})a_2 \\ k_3a_3 \end{bmatrix}. \tag{59}$$

By equation (35),

$$Z_R(\tau_{\mathbf{k}}\gamma_1, B) = \left\{ \mathbf{x} \in F_{\Gamma\backslash X} : \begin{bmatrix} 0 \\ 2x_2 \\ 2x_3 \end{bmatrix} \in M(\tau_{\mathbf{k}}\gamma_1, B_R) \right\}$$
$$= \left\{ \mathbf{x} \in F_{\Gamma\backslash X} : \begin{bmatrix} -(\frac{k_1}{2} + \frac{1}{4})a_1 \\ x_2 - (\frac{k_2}{2} + \frac{1}{4})a_2 \\ x_3 - \frac{k_3}{2}a_3 \end{bmatrix} \in \frac{1}{2}M_1(B_R) \right\}. \tag{60}$$

Thus the translational collision zones $Z_R(\tau_{\mathbf{k}}\gamma_1, B)$ are planar slices of translates of the Minkowski sum $M_1(B_R)$, given as follows. We define the planes

$$\Pi_1(k) \doteq \left\{ \mathbf{x} \in \mathbb{R}^3 : x_1 = -\left(\frac{k}{2} + \frac{1}{4}\right)a_1 \right\}, k \in \mathbb{Z}. \tag{61}$$

Then $Z_R(\tau_{\mathbf{k}}\gamma_1, B)$ is the cylinder in $F_{\Gamma\backslash X}$ with axis in the $x_1$ direction and base given by the planar slice

$$\mathcal{S}_R \doteq \left( \frac{1}{2}M_1(B_R) + \begin{bmatrix} 0 \\ (\frac{k_2}{2} + \frac{1}{4})a_2 \\ \frac{k_3}{2}a_3 \end{bmatrix} \right) \cap \Pi_1(k_1). \tag{62}$$

Similarly we let

$$M_2(B_R) \doteq R_{\gamma_2} \cdot B_R + (-B_R) = \left\{ \begin{bmatrix} -x_1 - y_1 \\ x_2 - y_2 \\ -x_3 - y_3 \end{bmatrix} : \mathbf{x}, \mathbf{y} \in B_R \right\}, \tag{63}$$

and we obtain

$$Z_R(\tau_{\mathbf{k}}\gamma_2, B) = \left\{ \mathbf{x} \in F_{\Gamma\backslash X} : \begin{bmatrix} x_1 - \frac{k_1}{2}a_1 \\ -(\frac{k_2}{2} + \frac{1}{4})a_2 \\ x_3 - (\frac{k_3}{2} + \frac{1}{4})a_3 \end{bmatrix} \in \frac{1}{2}M_2(B_R) \right\}, \tag{64}$$

which is a cylinder in $F_{\Gamma\backslash X}$ with axis in the $x_2$ direction and base obtained by slicing a translate of $\frac{1}{2}M_2(B_R)$.

Finally, let

$$M_3(B_R) \doteq R_{\gamma_3} \cdot B_R + (-B_R) = \left\{ \begin{bmatrix} -x_1 - y_1 \\ -x_2 - y_2 \\ x_3 - y_3 \end{bmatrix} : \mathbf{x}, \mathbf{y} \in B_R \right\}. \tag{65}$$

Then

$$Z_R(\tau_{\mathbf{k}}\gamma_3, B) = \left\{ \mathbf{x} \in F_{\Gamma\backslash X} : \begin{bmatrix} x_1 - (\frac{k_1}{2} + \frac{1}{4})a_1 \\ x_2 - \frac{k_2}{2}a_2 \\ -(\frac{k_3}{2} + \frac{1}{4})a_3 \end{bmatrix} \in \frac{1}{2}M_3(B_R) \right\}, \tag{66}$$

which is similarly a cylinder in $F_{\Gamma\backslash X}$ with axis in the $x_3$ direction and base obtained by slicing a translate of $\frac{1}{2}M_3(B_R)$. Note that

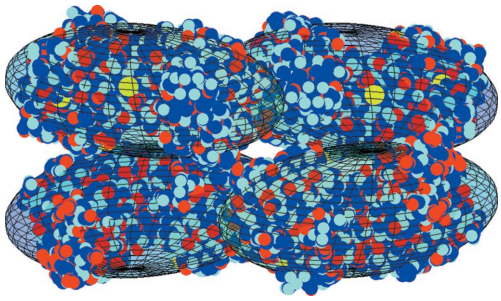


**Figure 13**
Symmetry mates for 1lfg and their moment-of-inertia ellipsoids in one unit cell.

equations (63)/(64) and (65)/(66) can be obtained from equations (58)/(60) by cyclic permutations of the formulas for the coordinates.

### 7.1. Example with $P2_12_12_1$ symmetry

Here we illustrate protein packing in a $P2_12_12_1$ crystal. In each unit cell of a biomolecular crystal with $P2_12_12_1$ symmetry, there are four asymmetric units, each corresponding to a symmetry mate. Fig. 13 shows symmetry mates for the PDB entry 1lfg (Haridas *et al.*, 1995) and their moment-of-inertia ellipsoids in a unit cell.

Fig. 14 shows the Minkowski-sum slice $\mathcal{S}_R$ given by equation (62) for the PDB entry 1lfg (and $k_1 = -1, k_2 = k_3 = 0$), with the dotted line delineating the slice $\mathcal{S}_R^{\rm e}$ obtained by replacing the protein body $B$ by its moment-of-inertia ellipsoid in equation (62). The outer and inner lines delineate the slices obtained by scaling the moment-of-inertia ellipsoid by factors of 1.25 and 0.75, respectively. The area occupied by the Minkowski-sum slice $\mathcal{S}_R$ outside of the slice $\mathcal{S}_R^{\rm e}$ is 6.12% of the area of the $\mathcal{S}_R^{\rm e}$. And 92.66% of the area of $\mathcal{S}_R$ is contained in $\mathcal{S}_R^{\rm e}$.

### 7.2. Determining the collision-causing symmetry sets

In this section we apply equations (57), (60), (64) and (66) to improve the upper estimate [equation (24)] for the collision-causing symmetry sets $\Lambda(g\cdot\overline{E}_2)$.

Suppose that we have ellipsoids $E_1$, $E_2$ with $E_1 \subseteq B \subseteq E_2$. We order the unit-cell dimensions so that $a_1 \le a_2 \le a_3$, and we let

$$\rho_i = r_{E_2}/a_j, \ \text{ for } j = 1, 2, 3.$$

We write $\mathbf{k}^i = (k_{i1}, k_{i2}, k_{i3}) \in \mathbb{Z}^3$, for $0 \le i \le 3$. First, suppose that $\tau_{\mathbf{k}^0} \in \Lambda(g\cdot E_2)$. Then by equation (57) with $B$ replaced by $\overline{E_2}$ [and noting that $\tau_{\mathbf{0}} = e \notin \Lambda(g\cdot\overline{E}_2)$ by definition], we have

$$0 < \sum k_{0j}^2 a_j^2 \le 4r_{E_2}^2.$$

Therefore

$$\rho_1^{-2}k_{01}^2 + \rho_2^{-2}k_{02}^2 + \rho_3^{-2}k_{03}^2 \le 4. \tag{67}$$

Next, let $\tau_{\mathbf{k}^1}\gamma_1 \in \Lambda(g\cdot\overline{E}_2)$. Then by equation (60) with $B$ replaced by $\overline{E_2}$,

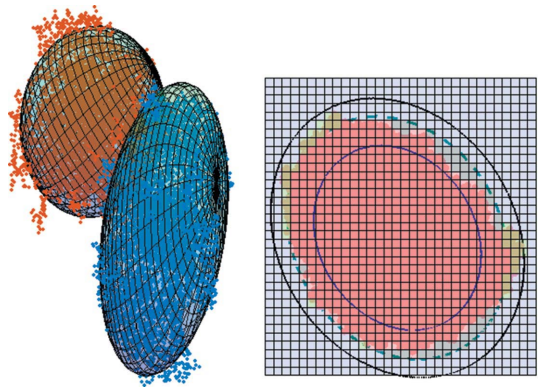


**Figure 14**
A slice of the Minkowski sum for the PDB entry 1lfg.

$$\begin{bmatrix} |k_{11}+\frac{1}{2}|a_1 \\ |k_{12}+\frac{1}{2}|a_2 \\ |k_{13}|a_3 \end{bmatrix} \in \left\{\mathbf{y}+\mathbf{z}\in\mathbb{R}^3 : |\mathbf{y}| \le 2r_{E_2},\, z_1 = 0,\, |z_2| \le \frac{a_2}{2},\, |z_3| \le \frac{a_3}{2}\right\},$$

which consists of points of distance less than $2r_{E_2}$ from the rectangle $\{0\}\times\left[-\frac{a_2}{2},\frac{a_2}{2}\right]\times\left[-\frac{a_3}{2},\frac{a_3}{2}\right]$. Therefore,

$$\rho_1^{-2}(k_{11}+\tfrac{1}{2})^2 + \rho_2^{-2}(|k_{12}+\tfrac{1}{2}|-\tfrac{1}{2})^2 + \rho_3^{-2}[(|k_{13}|-\tfrac{1}{2})_+]^2 \le 4, \tag{68}$$

where we use the notation $a_+ = \max(0, a)$.

Similarly, if $\tau_{\mathbf{k}^2}\gamma_2 \in \Lambda(g\cdot E_2)$, then by equation (64)

$$\begin{bmatrix} |k_{21}|a_1 \\ |k_{22}+\frac{1}{2}|a_2 \\ |k_{23}+\frac{1}{2}|a_3 \end{bmatrix} \in \left\{\mathbf{y}+\mathbf{z}\in\mathbb{R}^3 : |\mathbf{y}| \le 2r_{E_2},\, |z_1| \le a_1,\, z_2 = 0,\, |z_3| \le \frac{a_3}{2}\right\},$$

and therefore

$$\rho_1^{-2}[(|k_{21}|-1)_+]^2 + \rho_2^{-2}(k_{22}+\tfrac{1}{2})^2 + \rho_3^{-2}(|k_{23}+\tfrac{1}{2}|-\tfrac{1}{2})^2 \le 4. \tag{69}$$

If $\tau_{\mathbf{k}^3}\gamma_3 \in \Lambda(g\cdot E_2)$, then by equation (66)

$$\rho_1^{-2}[(|k_{31}+\tfrac{1}{2}|-1)_+]^2 + \rho_2^{-2}[(|k_{32}|-\tfrac{1}{2})_+]^2 + \rho_3^{-2}(k_{33}+\tfrac{1}{2})^2 \le 4. \tag{70}$$

In the $P1$ case discussed in Section 5, we observed that one only needs to check which symmetries $(\mathbb{I}, \mathbf{t}) \in \Gamma$ with $|\mathbf{t}| \le 2r_{E_2}$ cause symmetry mates of $g\cdot\overline{E}_2$ to collide, since collisions cannot occur if $|\mathbf{t}| > 2r_{E_2}$. We now give a description of the symmetries that we need to evaluate for collisions in the $P2_12_12_1$ case. Let $\Phi^*(\rho_1,\rho_2,\rho_3)$ denote the set of symmetries $\mathbf{k}^i\gamma_i \in \Gamma$ satisfying equations (67)–(70), so that $\Lambda(g\cdot\overline{E}_2) \subseteq \Phi^*(\rho_1,\rho_2,\rho_3)$. We then let $\widetilde{\Phi}(\rho_1,\rho_2,\rho_3)$ be a set consisting of one element from each pair $\{\gamma,\gamma^{-1}\}$ in $\Phi^*(\rho_1,\rho_2,\rho_3)$, so that

$$\Lambda(g\cdot\overline{E}_2) \subseteq \widetilde{\Phi}(\rho_1,\rho_2,\rho_3) \cup \widetilde{\Phi}(\rho_1,\rho_2,\rho_3)^{-1}. \tag{71}$$

Thus, to determine whether a motion $g \in F_{\Gamma\backslash G}$ satisfies equation (18), it is sufficient to check which symmetries

$\gamma \in \widetilde{\Phi}(\rho_1, \rho_2, \rho_3)$ cause collisions of $g \cdot \overline{E}_2$ and which cause collisions of $g \cdot E_1$.

From an analysis of over 7000 protein molecules in the PDB that crystallize with $P2_12_12_1$ symmetry (with four copies in each unit cell), we observed that over 96% of these molecules have circumradius

$$r_B < \min\{1.2a_1,\ 0.9a_2,\ 0.6a_3\}. \tag{72}$$

For these molecules satisfying equation (72), one needs only to consider symmetries in the finite set $\widetilde{\Phi}(1.2, 0.9, 0.6)$ in order to determine which symmetries $\gamma \in \Gamma$ are in $\Lambda(g \cdot E_1)$ and $\Lambda(g \cdot \overline{E}_2)$ in equation (18).

We now list the elements of $\widetilde{\Phi}(1.2, 0.9, 0.6)$. Let us write

$$\widetilde{\Phi}(1.2, 0.9, 0.6) = \bigcup_{i=0}^{3} \{\tau_{\mathbf{k}^i}\gamma_i : \mathbf{k}^i \in A_i\},$$

where $A_i \in \mathbb{Z}^3$ for $0 \le i \le 3$. Substituting $\rho_1 = 1.2$, $\rho_2 = 0.9$, $\rho_3 = 0.6$ in equations (67)–(70), we find that

$$A_0 = \{(0, 0, 1), (0, 1, 0), (1, -1, 0), (1, 0, -1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (2, 0, 0)\}.$$

The points of $A_1, A_2, A_3$ are given in Figs. 15, 16 and 17, respectively. Altogether, $\widetilde{\Phi}(1.2, 0.9, 0.6)$ contains 92 elements (the sum of the cardinalities of the sets $A_0, A_1, A_2, A_3$) of the crystallographic group, although for proteins with a bound on the circumradius that is sharper than equation (72), the number could be smaller.

We now construct a reduced search space $F^0$ of Theorem 4.1 for the group $\Gamma$ of type $P2_12_12_1$ and a centered ellipsoidal body $B$ with axes aligned with the unit-cell directions: we recall that the symmetry group of $B$ is the group $K \cup (-\mathbb{I})K$ of order 8 given in equation (38); the normalizer $\hat{\Gamma} < \mathrm{E}(3)$ of $\Gamma$ is the symmorphic group of type *Pmmm* with lattice $\hat{\mathbb{L}}$ generated by $\{(\frac{a_1}{2}, 0, 0), (0, \frac{a_2}{2}, 0), (0, 0, \frac{a_3}{2})\}$ [see Hirshfeld (1968), Koch *et al.* (2002) or the Bilbao server (Aroyo *et al.*, 2006)]. Then $\hat{\Gamma} \cap G$ is of type $P222$ with lattice $\hat{\mathbb{L}}$. We now choose the reduced search space

$$F^0 \doteq F_{\mathrm{SO}(3)/K} \times F^+_{(\hat{\Gamma}\cap G)\backslash X},$$

where

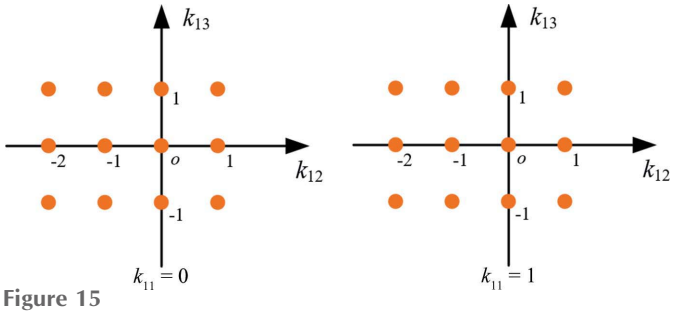



**Figure 15**
Points of $A_1$.

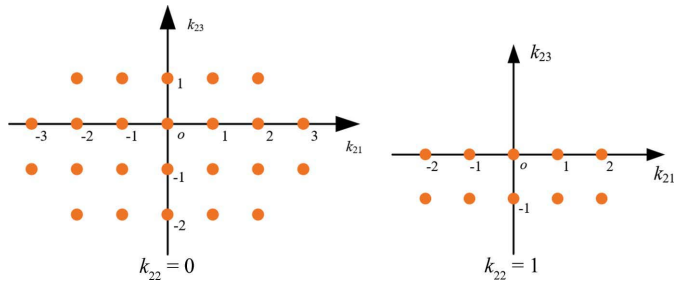



**Figure 16**
Points of $A_2$.

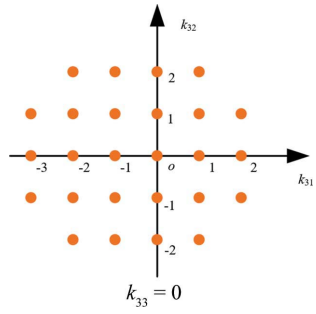



**Figure 17**
Points of $A_3$.

$$F^+_{(\hat{\Gamma}\cap G)\backslash X} \doteq \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \le x_i < \frac{a_i}{4},\ 1 \le i \le 3\right\},$$
$$F_{\mathrm{SO}(3)/K} \doteq \left\{R \in \mathrm{SO}(3) : R_{11} \ge 0, R_{21} \ge 0\right\}.$$

We note that $\mathrm{vol}(F^0) = \frac{1}{64}\mathrm{vol}(\mathrm{SO}(3) \times F_{\Gamma\backslash X})$, where $F_{\Gamma\backslash X}$ is given by equation (56), consistent with Theorem 4.1.

## 8. Conclusions

In this paper we derive conditions for reducing the volume of a full rotational–translational MR search with strong cost function by using two opposing constraints: (i) symmetrically arranged macromolecules in a crystal cannot interpenetrate each other; and (ii) macromolecules in a crystal must contact a sufficient number of neighbors to form a physical crystal. These concepts are formalized mathematically and demonstrated using ellipsoidal models. It is observed that the Minkowski sum of proteins generally is more similar to the Minkowski sum of ellipsoidal proxies of proteins than the similarity of the protein and proxy. Since it is the Minkowski sum of symmetry mates that enters our formulation rather than the detailed shape properties of individual molecules, the use of ellipsoidal proxies is justified. Consequently, the fraction of volume of the motion space $F_{\Gamma\backslash G}$ that actually needs to be searched in a MR problem can be extremely small. Moreover, it is plausible (particularly for proteins that are very prolate) that large regions in rotation space will not have any feasible translations based on our packing formalism, further reducing the computational effort. Additional efficiencies can be gained using symmetry properties of ellipsoids and of the crystallographic group itself. Future algorithmic work that builds on the mathematical foundations presented

in this series of papers will focus on explicit computer codes for use by the macromolecular crystallography community.

## Acknowledgements

Yuqing Pan assisted with the preparation of Figs. 8 and 11. The authors would also like to thank the referee for providing many helpful suggestions.

## Funding information

This work was supported by National Science Foundation grant NSF CCF-1640970.

## References

Abele, U. & Schulz, G. E. (1995). *Protein Sci.* **4**, 1262–1271.
Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Z. Kristallogr.* **221**, 15–27.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Chirikjian, G. S. (2011). *Acta Cryst.* A**67**, 435–446.
Chirikjian, G. S. (2015). *ASME J. Comput. Inf. Sci. Eng.* **15**, 1–7.
Chirikjian, G. S., Ratnayake, K. & Sajjadi, S. (2015). *Z. Kristallogr.* **230**, 719–741.
Chirikjian, G. S., Sajjadi, S., Shiffman, B. & Zucker, S. M. (2017). *Acta Cryst.* A**73**, 387–402.
Chirikjian, G., Sajjadi, S., Toptygin, D. & Yan, Y. (2015). *Acta Cryst.* A**71**, 186–194.
Chirikjian, G. S. & Shiffman, B. (2016). *Robotica*, **34**, 1679–1704.
Chirikjian, G. S. & Yan, Y. (2012). *Acta Cryst.* A**68**, 208–221.
Conway, J. H. & Sloane, N. J. A. (1999). *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer.
Dunitz, J. D. (1996). *X-ray Analysis and Structure of Organic Molecules*, 2nd ed. New York: Wiley.
Eon, J.-G. (2016*a*). *Acta Cryst.* A**72**, 268–293.
Eon, J.-G. (2016*b*). *Acta Cryst.* A**72**, 376–384.
Grünbaum, B. & Shephard, G. C. (1987). *Tilings and Patterns*. New York: Freeman.
Haridas, M., Anderson, B. F. & Baker, E. N. (1995). *Acta Cryst.* D**51**, 629–646.
Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.
Jia, X., Choi, Y. K., Mourrain, B. & Wang, W. (2011). *Comput. Aided Geom. Des.* **28**, 164–176.
Johnson, C. K., Burnett, M. N. & Dunbar, W. D. (1996). *Crystallographic Computing*, Vol. 7, *Macromolecular Crystallographic Data*, edited by P. E. Bourne & K. D. Watenpaugh. Oxford University Press.
Keune, W. J., Potjewyd, F., Heidebrecht, T., Salgado-Polo, F., Macdonald, S. J., Chelvarajan, L., Abdel Latif, A., Soman, S., Morris, A. J., Watson, A. J., Jamieson, C. & Perrakis, A. (2017). *J. Med. Chem.* **60**, 2006–2017.
Kitaigorodskii, A. I. (1973). *Molecular Crystals and Molecules.* New York: Academic Press.
Koch, E., Fischer, W. & Müller, U. (2016). *International Tables for Crystallography*, Vol. A, *Space-group symmetry*, 2nd online ed., edited by M. I. Aroyo, pp. 826–851. Chester: International Union of Crystallography.
Lesk, A. M. & Rose, G. D. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4304–4308.
Matsumoto, T. (1968). *Z. Kristallogr.* **126**, 170–174.
Matsumoto, T. & Nowacki, W. (1966). *Z. Kristallogr.* **123**, 401–421.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Müller, J. J. & Schrauber, H. (1992). *J. Appl. Cryst.* **25**, 181–191.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
O'Keeffe, M. (1995). *Z. Kristallogr.* **210**, 905–908.
Peresypkina, E. V. & Blatov, V. A. (2000). *Acta Cryst.* B**56**, 1035–1045.
Prabhakaran, M. & Ponnuswamy, P. K. (1982). *Macromolecules*, **15**, 314–320.
Qin, S. & Zhou, H. X. (2013). *J. Chem. Theory Comput.* **9**, 4633–4643.
Richards, F. M. (1977). *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
Rossmann, M. G. (2001). *Acta Cryst.* D**57**, 1360–1366.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Ryabov, Y. E., Geraghty, C., Varshney, A. & Fushman, D. (2006). *J. Am. Chem. Soc.* **128**, 15432–15444.
Smith, N., Roitberg, A. E., Rivera, E., Howard, A., Holden, M. J., Mayhew, M., Kaistha, S. & Gallagher, D. T. (2006). *Arch. Biochem. Biophys.* **445**, 72–80.
Sunada, T. (2013). *Topological Crystallography: With a View Towards Discrete Geometric Analysis.* Tokyo: Springer.
Taylor, W. R. & Aszódi, A. (2005). *Protein Geometry, Classification, Topology and Symmetry: a Computational Analysis of Structure.* New York: Taylor & Francis.
Taylor, W. R., May, A. C., Brown, N. P. & Aszódi, A. (2001). *Rep. Prog. Phys.* **64**, 517–590.
Taylor, W. R., Thornton, J. M. & Turnell, W. G. (1983). *J. Mol. Graph.* **1**, 30–38.
Thimm, G. (2009). *Acta Cryst.* A**65**, 213–226.
Thompson, T. M. (1983). *From Error-Correcting Codes through Sphere Packings to Simple Groups.* The Carus Mathematical Monographs, No. 21. Mathematical Association of America.
Thornton, J. M., Edwards, M. S., Taylor, W. R. & Barlow, D. J. (1986). *EMBO J.* **5**, 409–413.
Todd, M. J. (2016). *Minimum-Volume Ellipsoids. Theory and Algorithms*, Vol. 23. SIAM.
Wells, A. F. (1977). *Three-Dimensional Nets and Polyhedra.* New York: John Wiley and Sons.
Yan, Y. & Chirikjian, G. S. (2015). *Geom. Dedicata*, **177**, 103–128.
Zehfus, M. H., Seltzer, J. P. & Rose, G. D. (1985). *Biopolymers*, **24**, 2511–2519.