Collision-Free Configuration-Spaces in Macromolecular Crystals

Gregory S. Chirikjian
Department of Mechanical Engineering
Johns Hopkins University
Baltimore, MD 21218
Email: gregc@jhu.edu

Bernard Shiffman
Department of Mathematics
Johns Hopkins University
Baltimore, MD 21218
Email: shiffman@math.jhu.edu

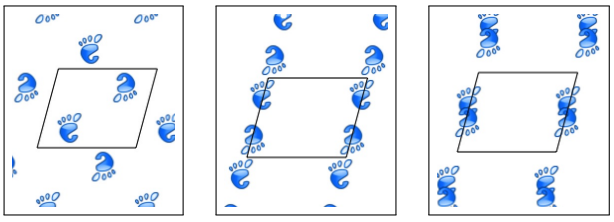
***Abstract*—Molecular replacement (MR) is a well-established computational method for phasing in macromolecular crystallography. In MR searches, spaces of motions are explored for determining the appropriate placement of rigid single-body (or articulated multi-rigid-body) models of macromolecules in crystallographic asymmetric units. This paper investigates which portion of the motion space is physically realizable given that packing of protein molecules in a crystal are subject to the constraint that they cannot interpenetrate. This imposes severe restrictions on which points in the motion space are accessible. By determining a priori which portions of motion space correspond to symmetry mates in collision, it becomes feasible to construct more efficient MR techniques which avoid searching in those regions of motion space.**

I. Introduction

Imagine an ideal crystal composed of an infinite number of identical copies of a solid body. Although a true crystal is always finite in its extent, it consists of such a large number of copies that the idealization does not introduce too much discrepancy from the reality. With knowledge of the shape and pose (position and orientation) of one of the bodies in the crystal, and knowledge of the crystallographic symmetry group, it is possible to reconstruct the whole crystal. This is essentially the situation that arises in macromolecular crystallography (MX), where the solid body may be a protein, nucleic acid, or complex composed of both. And information about the crystal symmetry is provided by an x-ray diffraction experiment.

This symmetry information, together with prior information about the shape of the bodies obtained from the Protein Data Bank (PDB) [1], can be used to assist in solving the so-called "phase problem" in x-ray crystallography in a purely computational way using a modified version of the method of molecular replacement (MR) [14, 15]. In so doing, it becomes feasible to phase large biomolecular structures that currently require expensive experimental methods. The one piece of information that is missing is how to find viable candidate poses efficiently. This is the subject of this paper. By excluding at the beginning of an MR search those poses corresponding to configurations that put the bodies in nonphysical collisions/overlaps, MR searches can be made more efficient. This reduces MR searches to a problem of characterizing c-space (configuration space) obstacles corresponding to bodies colliding with their symmetry mates (i.e., bodies that move in a concerted way subject to symmetry constraints), and sampling only in the collision-free regions of c-space. This is somewhat different than traditional robot motion planning problems in which there are static obstacles and moving robots. The robot analogy in the MR problem would be more akin to the coordination of multiple robots moving in formation, with c-space obstacles describing when members of the formation collide with each other.

The infinite number of bodies that fill Euclidean space in an ideal crystal form a highly symmetrical pattern described by a crystallographic space group. For simple (inorganic) molecules, the number of allowable space groups is quite large. But for macromolecules, which are chiral (i.e., they have handedness), far fewer symmetries are allowed, since mirror reflections and glide planes are not possible. For this reason, there are only five planar crystallographic groups for which patterns composed of solid planar bodies can exist (out of a total of 17 planar wallpaper groups), and only 65 are allowable in the spatial case (out of 230). Figure 1 shows three configurations of a planar footprint pattern generated using the free iphone app "Eschermobile" developed by [13]. These figures all have the same wallpaper group symmetry,

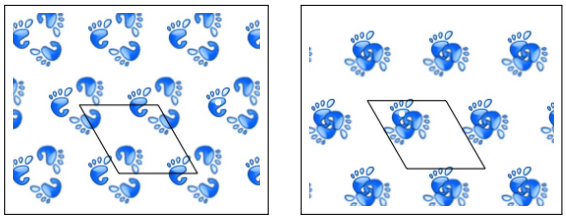


(a) (b) (c)

Fig. 1. Three configurations of solid bodies with $p2$ symmetry

called $p2$, but differ by translational degrees of freedom. It is also possible to construct configurations with $p2$ symmetry for which each solid body rotates in concert with each other. In Figures 1(a) and (b) the bodies are not in collision, but in Figure 1(c) the bodies are in collision. Obviously, if these bodies were physical, such collisions would not be allowed. Figure 2 shows two configurations of the footprint pattern with $p3$ symmetry. In Figure 2(a) the bodies are not in collision,

while in Figure 2(b) the bodies are in collision. In all of these



(a) (b)

Fig. 2. Two configurations of solid bodies with $p3$ symmetry

figures, the parallelogram in the middle is the crystallographic unit cell, which can be viewed as a tile that reproduces the whole infinite crystal under the action of a discrete lattice translation group.

This paper characterizes the subset of motions that will cause bodies to collide with their symmetry mates. In so doing, the complement of this space, the free space, is naturally characterized. These results are directly applicable to accelerate MR searches by eliminating collision states. Section II formulates the general problem in n-dimensional space and shows how the collision space for an infinite crystal corresponds to when several coordinated bodies moving in a torus overlap.

Section III works out the details for the planar case, and it is shown that the resulting c-space obstacles can be characterized as Minkowski sums in the torus. Equations (11)–(12) describe the collision space in the crystallographic unit cell; alternatively, equation (13) describes the collision space in the torus. Equation (13) is applied to give specific formulas for the collision space for crystals with $p2$ symmetry and with $p3$ symmetry, and the results are demonstrated with elliptical bodies in Section IV. The $p4$ and $p6$ cases, as well as some 3D cases will be given in a future paper.

In the remainder of this section, the mathematical aspects of the MR problem are reviewed. The glossary of symbols below describes the terminology used throughout the paper.

Glossary:

X - n-dimensional Euclidean space, $\mathbb{R}^n$.

$R \in SO(n)$ - an $n \times n$ orthogonal matrix with determinant 1, i.e., R is a rotation matrix.

$R(\theta) \in SO(2)$ - rotation in $\mathbb{R}^2$ by an angle θ.

$\mathbf{t}, \mathbf{x} \in X$ - an n-dimensional translation vector or position on which this translation acts by vector addition.

$g = (R, \mathbf{t}) \in SE(n)$ - a special-Euclidean transformation (i.e., a proper rigid-body motion). Mathematically, $SE(n) = \mathbb{R}^n \rtimes SO(n)$ (a semi-direct product) which acts on positions as $g \cdot \mathbf{x} = R\mathbf{x} + \mathbf{t}$.

G - shorthand for $SE(n)$.

$\gamma = (R_\gamma, \mathbf{t}_\gamma) \in \Gamma$ - an element of a chiral crystallographic space group (discrete subgroup of G that contains a lattice of translations of rank n).

$\mathbb{L}$ - a lattice in X of rank n.

T - the lattice translation subgroup of Γ, consisting of elements identified with the set $\mathbb{L}$; T is the maximal normal Abelian subgroup of Γ.

$T\backslash X$ - an n-dimensional torus.

$F_{T\backslash X}$ - the crystallographic unit cell.

$\mathbb{P} = \{R_\gamma \in SO(n) : \exists\ \mathbf{t}_\gamma \in X$ such that $(R_\gamma, \mathbf{t}_\gamma) \in \Gamma\}$ (a discrete rotation group, called the point group; $\mathbb{P} \cong T\backslash\Gamma$; in the symmorphic case, $\mathbb{P} = \{R \in SO(n) : (R, \mathbf{0}) \in \Gamma\}$.

$\mathbf{t}_\gamma \in T$ - a lattice translation such that $\mathbf{t}_\gamma \cdot \mathbb{L} = \mathbb{L}$.

$\mathbf{v}(R_\gamma)$ - a translation along a screw axis by a fraction of a unit cell dimension defined for $R_\gamma \in \mathbb{P}$ as an element $\mathbf{v}(R_\gamma) \in X$ (uniquely given modulo T) such that $(R_\gamma, \mathbf{v}(R_\gamma)) \in \Gamma$.

(Mathematically, the map $\mathbf{v} : \mathbb{P} \to T\backslash X$ yields an element $[\mathbf{v}]$ of the cohomology group $H^1(\mathbb{P}, T\backslash X)$, and the correspondence $\Gamma \mapsto [\mathbf{v}]$ is a bijection between translation-conjugacy classes of crystallographic groups with point group $\mathbb{P}$. (See [10])

$F_{\Gamma\backslash X}$ - the crystallographic asymmetric unit (a fundamental domain for Γ acting on X, i.e., a tile which, under the action of Γ, fills Euclidean space, X, without gaps or overlaps of positive measure, as computed in [12] in the planar case and in [8] in 3D); The space $\Gamma\backslash X$ itself is called a Euclidean orbifold [7].

$F_{\Gamma\backslash G}$ - the smallest finite-volume space of rotations and translations in which molecular replacement searches need to be performed; This is a "tile" which fills G without gaps or overlaps of positive measure under the action of Γ.

$B_1 + B_2$ - the Minkowski sum of two bodies B_1,B_2 in X.

$\widetilde{B}_1 \boxplus \widetilde{B}_2$ - the Minkowski sum of $\widetilde{B}_1$, $\widetilde{B}_2$ in $T\backslash X$ regarded as an Abelian group.

$V_n(B)$ - the volume of an n-dimensional body $B \subset X$.

$\rho_X = \rho_X^B : X \to \{0, 1\}$ - the indicator function for a body $B \subset X$ (which takes a value of 1 on the body, and zero otherwise).

$\rho_{\Gamma\backslash X} = \rho_{\Gamma\backslash X}^B : \Gamma\backslash X \to \mathbb{Z}_{\geq 0}$ - the density function for a crystal composed of bodies $\gamma \cdot B$ with density $\sum_{\gamma\in\Gamma} \rho_X \circ \gamma^{-1}$. (The density $\rho_{\Gamma\backslash X}$ is an indicator function if the bodies do not collide.)

MR is a computational method to phase macromolecular crystals that has been in use for more than half a century [15]. We briefly review the mathematical formulation of this method.

Let $X = \mathbb{R}^n$, n-dimensional Euclidean space, where of course the most practical case is $n = 3$. The inputs to MR computations are then: (1) the electron density, $\rho_X(\mathbf{x})$, (normalized to take the value 1 on the body) of a known rigid macromolecule (or fragment thereof) called the reference molecule; and (2) the symmetry group of the crystal, Γ, which

is a discrete subgroup of $G = SE(n)$, the (continuous) Lie group of proper motions of rigid-bodies in n-dimensional Euclidean space. The group operation for G and Γ is denoted as "$\circ$", and their action on Euclidean space is denoted as "$\cdot$".

The reference molecule should be similar in structure to the one to be determined in order for the MR method to work. Such knowledge for proteins may come from prior knowledge of the similarity of the amino acid sequences of the reference and actual molecules, and the many tens of thousands of existing structures in the Protein Data Bank (PDB) [1].

In MR, the goal is to position and orient copies of the electron densities of the reference molecule in the crystallographic unit cell by some $g \in G$ to form a model density of the form

$$\rho_{\Gamma\backslash X}(\mathbf{x}; g) = \sum_{\gamma\in\Gamma} \rho_X\left((\gamma \circ g)^{-1} \cdot \mathbf{x}\right). \tag{1}$$

(This can be viewed as a function of $\mathbf{x}$ indexed by g.) The density function $\rho_X(\mathbf{x})$, takes a nonnegative value on the reference molecule and a zero value away from it.

Suppose that all of the dimensions of the reference molecule are smaller than all of the dimensions of the fundamental domain $F_{\Gamma\backslash X} \subset X$, which is identified with the asymmetric unit of the crystal. Then if the reference frame in which $\rho_X(\mathbf{x})$ is defined is centered at the origin of X, and $F_{\Gamma\backslash X}$ is defined to have its origin at the origin of X, and if g is a small motion, the body will still be fully contained in $F_{\Gamma\backslash X}$. In such a circumstance, the sum in (1) will only have nonzero contribution from $\gamma = e$.

For each fixed $g \in G$, $\rho_{\Gamma\backslash X}(\mathbf{x}; g)$ can be viewed either as a function on the asymmetric unit $F_{\Gamma\backslash X}$, or as function on the unit cell $F_{T\backslash X}$. In the latter case, the function will have symmetry within the unit cell described by the finite factor group $\mathbb{F} = F_{T\backslash\Gamma} = F_{\Gamma/T}$. Here and throughout this paper, $F_{A\backslash B} \subset B$ denotes a fundamental domain from which the space B can be tiled or reconstructed by the left action of the group A on $F_{A\backslash B}$.

As established in [5], all candidate positions and orientations (called "poses"), $g \in G$, can be chosen without loss of generality to be of the form $g \in F_{\Gamma\backslash G}$, or equivalently, each g can be viewed as a distinguished representative of the coset $\Gamma g \in \Gamma\backslash G$. These candidate poses can be evaluated and ranked according to the value of a cost function such as

$$C(g) = \sum_{\mathbf{k}\in\mathcal{F}(\mathcal{T}\backslash\mathcal{X})} \left(|\hat{\rho}_{\Gamma\backslash X}(\mathbf{k}; g)| \,-\, \hat{P}(\mathbf{k})\right)^2 \tag{2}$$

where $\hat{\rho}$ is the Fourier transform of ρ, and $\mathcal{F}(T\backslash X) \cong \mathbb{Z}^n$ is the unitary dual of (Fourier space corresponding to) $T\backslash X$.

The main goal of molecular replacement is to obtain a list of physically realizable candidate poses $\{g\}$ rank-ordered by the value of $C(g)$. To this end, this paper seeks to characterize the set of all $g \in F_{\Gamma\backslash G}$ when reference molecules and their symmetry mates are in collision, in order that such regions of motion space can be circumvented in MR searches, thereby speeding up computations.

II. Room-to-Move for Bodies in Crystallographic Environments

Let $\mathbb{L} \subset X = \mathbb{R}^n$ denote a lattice. The corresponding lattice translation group is $T = \mathbb{L} \rtimes \{\mathbb{I}\}$ where $\mathbb{I} \in SO(n)$ is the identity rotation. The distinction between T and $(\mathbb{L}, +)$ is often blurred, but it is T defined in this way which forms a normal subgroup of any crystallographic space group, Γ, with Bravais lattice $\mathbb{L}$.

Let dg denote the bi-invariant integration measure for $G = SE(n)$, the group of rigid-body motions in n-dimensional Euclidean space. This provides a proper way to integrate on the motion space $F_{\Gamma\backslash G} \subset G$.

As discussed in [6], we can take

$$F_{\Gamma\backslash G} = F_{\Gamma\backslash X} \times SO(n).$$

The action of any $g \in F_{\Gamma\backslash G}$ on any $\mathbf{x} \in F_{\Gamma\backslash X}$ will in general not produce a result that is in $F_{\Gamma\backslash X}$. But, as discussed in [5], it is always possible to find a $\gamma \in \Gamma$ such that $\gamma \circ g \cdot \mathbf{x} \in F_{\Gamma\backslash X}$ and this "equivalent point in $F_{\Gamma\backslash X}$ can be thought of as a "quasigroup action."

Let $B \subset F_{\Gamma\backslash X}$ denote an open set (i.e., a "body without boundary") that can be fully contained in the chosen crystallographic asymmetric unit. In particular, let

$$\delta = \min\{\, \|\mathbf{t}_\gamma\| \,:\, \mathbf{t}_\gamma \in \mathbb{L} \smallsetminus \{\mathbf{0}\}\,\}. \tag{3}$$

(Here "$\smallsetminus$" denotes the difference of sets.) We shall restrict the discussion to bodies with sizes limited by the condition

$$\text{diam}(B) \doteq \max_{\mathbf{x}_1,\mathbf{x}_2\in B} \|\mathbf{x}_1 - \mathbf{x}_2\| \,\le\, c_\Gamma\, \delta \tag{4}$$

where $0 < c_\Gamma < 1$ is a fraction depending only on Γ.

Let

$$S = \left\{ g \in F_{\Gamma\backslash G} \,:\, (g\cdot B) \,\cap \bigcup_{\gamma\in\Gamma\smallsetminus\{e\}} (\gamma\circ g)\cdot B \;\neq\; \emptyset \right\}, \tag{5}$$

where e denotes the identity in Γ. This set corresponds to all motions $g = (R, \mathbf{t}) \in F_{\Gamma\backslash G}$ of the body B so that the resulting body $g \cdot B$ will collide with (intersect) at least one of its symmetry mates $(\gamma \circ g) \cdot B$. As we shall see below, we only need to consider finitely many of the γ in (5), as the rest are redundant. This condition is equivalent to the motion g causing a collision between any two symmetry mates, since $(\gamma_1 \circ g)\cdot B \cap (\gamma_2 \circ g)\cdot B = \gamma_1 \cdot \big[(g\cdot B) \cap (\gamma_1^{-1} \circ \gamma_2 \circ g)\cdot B\big]$.

Our goal is to compute the collision probability

$$Prob_c(B) \doteq \frac{vol(S)}{vol(F_{\Gamma\backslash G})}, \tag{6}$$

where vol denotes invariant volume on G. We shall instead consider the crystallographic unit cell (fundamental domain) $F_{T\backslash X}$ in X and the corresponding fundamental domain $F_{T\backslash G} = F_{T\backslash X} \times SO(n)$ in G, and we let

$$S^* = \left\{ g \in F_{T\backslash G} \,:\, (g\cdot B) \,\cap \bigcup_{\gamma\in\Gamma\smallsetminus\{e\}} (\gamma\circ g)\cdot B \;\neq\; \emptyset \right\}.$$

Since the unit cell, $F_{T\backslash X}$, consists of k copies of the asymmetric unit, $F_{\Gamma\backslash X}$, where k is the order of the finite group $T\backslash\Gamma$, and correspondingly, S^* consists of k copies of S, we have

$$Prob_c(B) = \frac{vol(S^*)}{vol(F_{T\backslash G})}\,. \tag{7}$$

It appears that the problem of computing $Prob_c(B)$ in (7) has not been studied either in the mathematics or the crystallography literature. The closest work to this that we are aware of originates in the field of integral geometry and geometric probability. In this field, the so-called Principal Kinematic Formula computes the volume in G corresponding to collisions of one moving body and one fixed body (both convex) [2, 4, 11, 16, 17, 18]. Some work has generalized this to the case when the bodies are in a homogenous space other than Euclidean space [3] or when instead of a single fixed body, a periodic array of fixed bodies is present [19]. But neither of these works apply directly to our problem.

Fix the orientation $R \in SO(n)$ of the moving version of body B, and define

$$B_R \doteq R\cdot B = \{R\mathbf{x} \,:\, \mathbf{x}\in B\}.$$

Then let

$$S_R^* \,=\, \{\mathbf{t}\in F_{T\backslash X} \,:\, (R,\mathbf{t})\in S^*\}.$$

Note that any $\gamma\in\Gamma$ can be written as $\gamma = (R_\gamma, \mathbf{t}_\gamma + \mathbf{v}(R_\gamma))$ where $R_\gamma \in \mathbb{P}$, $\mathbf{t}_\gamma \in T$; only in the symmorphic case is $\mathbf{v}(R_\gamma) = \mathbf{0}$ for all values of R_γ. Suppose that $\mathbf{t}\in F_{T\backslash X}$. Then

$$\mathbf{t}\in S_R^* \iff \exists\,\mathbf{a},\mathbf{b}\in B_R,\ \gamma\in\Gamma\smallsetminus\{e\}$$

such that

$$\mathbf{a}+\mathbf{t} = R_\gamma(\mathbf{b}+\mathbf{t}) + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma). \tag{8}$$

This can be rearranged as

$$(\mathbb{I}-R_\gamma)\mathbf{t} = R_\gamma\mathbf{b} - \mathbf{a} + \mathbf{t}_\gamma + \mathbf{v}(R_\gamma). \tag{9}$$

In particular, in (5) we only need to consider finitely many $\gamma\in\Gamma$; e.g., those γ satisfying the bound

$$\begin{aligned}\|\mathbf{t}_\gamma\| &\le \|\mathbf{t}\| + \|R_\gamma\mathbf{t}\| + \|R_\gamma\mathbf{b}\| + \|\mathbf{a}\| + \|\mathbf{v}(R_\gamma)\| \\ &\le 5\,\sup_{\mathbf{t}\in F_{T\backslash X}}\|\mathbf{t}\|\,.\end{aligned}$$

III. Characterization of Collision Space for the Planar Case

In the planar, $n=2$, case, $R_\gamma\neq\mathbb{I}$ never has unit eigenvalues because the identity point-group rotation was chosen for B_R, and all other symmetry mates that can possibly intersect B_R have nontrivial point rotations as a result of the condition in (4). And since all chiral space groups in the plane are symmorphic, it is possible to write (9) as

$$\mathbf{t} = (\mathbb{I}-R_\gamma)^{-1}(R_\gamma\mathbf{b} - \mathbf{a} + \mathbf{t}_\gamma). \tag{10}$$

Suppose that $\mathrm{diam}(B) < \delta$. We let

$$B'_\gamma = (\mathbb{I}-R_\gamma)^{-1}\left[(R_\gamma B) + (-B)\right]. \tag{11}$$

where B_1+B_2 denotes the Minkowski sum of bodies B_1 and B_2, $R_\gamma B = \{R_\gamma\,\mathbf{b} : \mathbf{b}\in B\}$ and $-B = (-1)B$. Then

$$S_R^* = F_{T\backslash X}\cap \bigcup_{R_\gamma\in\mathbb{P}^*}\left\{(\mathbb{I}-R_\gamma)^{-1}\mathbb{L} + R\,B'_\gamma\right\} \tag{12}$$

where $R = R(\theta)$ is an arbitrary planar rotation.

We let $p : X\to T\backslash X$ denote the (universal) covering map given by

$$p(\mathbf{x}) \doteq T\cdot\mathbf{x} = \mathbb{L}+\mathbf{x}\in T\backslash X\,.$$

(Here, $T\cdot\mathbf{x}$ denotes the T-orbit of $\mathbf{x}$, which is identical to the coset $\mathbb{L}+\mathbf{x}$ in the Abelian group X.) Since the point group leaves $\mathbb{L}$ invariant, we have $R_\gamma\mathbb{L} = \mathbb{L}$, and therefore for all $\mathbf{t}_\gamma\in\mathbb{L}$,

$$(\mathbb{I}-R_\gamma)^{-1}\mathbb{L}+\mathbf{t}_\gamma = (\mathbb{I}-R_\gamma)^{-1}[\mathbb{L}+\mathbf{t}_\gamma - R_\gamma\mathbf{t}_\gamma] = (\mathbb{I}-R_\gamma)^{-1}\mathbb{L}.$$

Therefore the sets $(\mathbb{I}-R_\gamma)^{-1}\mathbb{L} + R\,B'_\gamma$ are invariant under the action of T, and it follows that we can identify S_R^* with its image

$$\widetilde{S_R^*} \doteq \bigcup_{R_\gamma\in\mathbb{P}\smallsetminus\{\mathbb{I}\}}\left\{p([\mathbb{I}-R_\gamma]^{-1}\mathbb{L})\boxplus p(R\,B'_\gamma)\right\}\subset T\backslash X\,, \tag{13}$$

where the $\boxplus$ denotes the Minkowski sum of two sets $Q_1,\,Q_2\subset T\backslash X$ is given by

$$Q_1\boxplus Q_2 = \{q_1+q_2 \bmod T : q_1\in Q_1,\ q_2\in Q_2\}.$$

In particular,

$$V_2(S_R^*) = V_2(\widetilde{S_R^*})\,. \tag{14}$$

For each $R_\gamma\in\mathbb{P}^*$, we let n_γ denote the number of distinct points of the form $p([\mathbb{I}-R_\gamma]^{-1}\mathbf{t}_\gamma)\in T\backslash X$, where $\mathbf{t}_\gamma$ runs through the points of the lattice $\mathbb{L}$, i.e.,

$$n_\gamma \doteq \#\left\{p([\mathbb{I}-R_\gamma]^{-1}\mathbb{L})\right\}.$$

Thus the collision probabilities $Prob_c(B)$ depend on the areas of B'_γ and the cardinalities n_γ. We shall determine the sets $p([\mathbb{I}-R_\gamma]^{-1}\mathbb{L})$ and their cardinalities n_γ explicitly and compute $Prob_c(B)$ for the four chiral cases separately.

Remark: The numbers n_γ can also be obtained from the Lefschetz fixed point formula, as follows: We note that $p([\mathbb{I}-R_\gamma]^{-1}\mathbb{L})$ is the fixed point set of R_γ acting on $X\backslash T$, and each fixed point has index 1 (since R_γ is a rotation about the fixed point). Identifying $\mathbb{L}$ with $H_1(T\backslash X,\mathbb{Z})$ via the covering map $X\to T\backslash X$, we see that the action of R_γ on $H_1(T\backslash X,\mathbb{Z})$ is just the action of R_γ on $\mathbb{L}$. Therefore,

$$\begin{aligned}n_\gamma &= 1-\mathrm{Trace}\big[R_{\gamma_*} : H_1(T\backslash X,\mathbb{Z})\to H_1(T\backslash X,\mathbb{Z})\big]+1 \\ &= 2-\mathrm{Trace}(R_\gamma) = 2-2\,\cos\alpha\,,\end{aligned}$$

where R_γ is a rotation by the angle α. Note that α is always a multiple of either $\pi/2$ or $\pi/3$, and thus n_γ takes the values 1, 2, 3, or 4, depending on the symmetry.

A. $p2$ symmetry.

In this case, $\mathbb{P} = \{\mathbb{I}, -\mathbb{I}\}$ and $\mathbb{L}$ is arbitrary. If the diameter of B is less than the minimum distance δ between points of $\mathbb{L}$, then B cannot collide with its translates, and we need only to consider symmetries γ with $R_\gamma = -\mathbb{I}$. Then (13) becomes

$$\widetilde{S_R^*} = p\left(\tfrac{1}{2}\,\mathbb{L}\right) \boxplus p\left(R\,B'_\gamma\right), \tag{15}$$

where

$$B'_\gamma = -\tfrac{1}{2}(B + B).$$

Furthermore, $p(\frac{1}{2}\mathbb{L})$ consists of exactly 4 points:

$$\alpha_1 \doteq p(\mathbf{0}), \quad \alpha_2 \doteq p(\tfrac{1}{2}\mathbf{v}_1),$$
$$\alpha_3 \doteq p(\tfrac{1}{2}\mathbf{v}_2), \quad \alpha_4 \doteq p(\tfrac{1}{2}\mathbf{v}_1 + \tfrac{1}{2}\mathbf{v}_2),$$

where $\{\mathbf{v}_1, \mathbf{v}_2\}$ generates $\mathbb{L}$.

Now suppose that $\mathrm{diam}(B) \le \frac{1}{2}\delta$. Then for distinct points $\mathbf{t}_\gamma, \mathbf{t}_{\gamma'} \in \mathbb{L}$,

$$\mathrm{diam}(R\,B'_\gamma) = \mathrm{diam}\left[\tfrac{1}{2}(B + B)\right] = \mathrm{diam}(B) \le \left\|\tfrac{1}{2}\mathbf{t}_\gamma - \tfrac{1}{2}\mathbf{t}_{\gamma'}\right\|, \tag{16}$$

and the bodies $R\,B'_\gamma + \frac{1}{2}\mathbf{t}_\gamma$ and $R\,B'_\gamma + \frac{1}{2}\mathbf{t}_{\gamma'}$ are disjoint. Thus

$$\widetilde{S_R^*} = \bigsqcup_{j=1}^{4} \left[p(R\,B'_\gamma) \boxplus \alpha_j\right], \tag{17}$$

where $\bigsqcup$ denotes a disjoint union.

Therefore

$$V_2(S_R^*) = V_2\left(\widetilde{S_R^*}\right) = 4\,V_2\left(B'_\gamma\right) = V_2(B + B). \tag{18}$$

Since (18) is independent of the rotation R, it then follows from (7) that the collision probability for $p2$ symmetry is given by

$$P_c(B) = \frac{V_2(B + B)}{V_2(F_{T\setminus X})}, \quad \text{for} \;\; \mathrm{diam}(B) \le \tfrac{1}{2}\delta. \tag{19}$$

Note that if B is convex, then $B + B = 2B$ and $P_c(B) = 4\,V_2(B)/V_2(F_{T\setminus X})$.

B. $p3$ symmetry.

In this case, $\mathbb{P} \equiv \{\mathbb{I}, R(\varphi), R(2\varphi)\}$, where $\varphi = 2\pi/3$, and we can assume that the lattice $\mathbb{L}$ is generated by $\{\delta\,\mathbf{u}_1,\, \delta\,\mathbf{u}_2\}$, where $\mathbf{u}_1 = (1, 0)$ and $\mathbf{u}_2 = (-\frac{1}{2}, \frac{\sqrt{3}}{2})$, for some $\delta > 0$. Recalling (3), we again suppose that $\mathrm{diam}(B) \le \delta$, so that B does not intersect its translates.

We note that we only need to consider symmetries $\gamma = (R_\gamma, \mathbf{t}_\gamma)$ with $R_\gamma = R(\varphi)$ in (5). Indeed, suppose that $\gamma' = (R(2\phi), \mathbf{t}_{\gamma'})$. If $g\cdot B$ intersects $(\gamma' \circ g)\cdot B = R(2\varphi)(g\cdot B) + \mathbf{t}_{\gamma'}$, then $R(\varphi)(g\cdot B)$ intersects $g\cdot B + R(\varphi)\mathbf{t}_{\gamma'}$, and hence $g\cdot B$ intersects $R(\varphi)B_R + \mathbf{t}_\gamma$, where $\mathbf{t}_\gamma = -R(\varphi)\mathbf{t}_{\gamma'}$. Here, we used the fact that $R(\varphi)\mathbb{L} = \mathbb{L}$. Therefore, we only need to consider $R_\gamma = R(\varphi)$ in (13), which becomes

$$\widetilde{S_R^*} = p([\mathbb{I} - R(\varphi)]^{-1}\mathbb{L}) \boxplus p(R\,B'_\gamma). \tag{20}$$

We easily see that $n_R(\varphi) = 3$ and

$$\begin{aligned} p([\mathbb{I} - R(\varphi)]^{-1}\mathbb{L}) &= \{p(\mathbf{0}), p(\mathbf{w}_1), p(\mathbf{w}_2)\} \\ &= \{p(\mathbf{0}), p(\mathbf{w}_1), p(-\mathbf{w}_1)\}, \end{aligned}$$

where $\mathbf{w}_1 = (0, \frac{\sqrt{3}}{3}\delta)$, $\mathbf{w}_2 = (\frac{1}{2}\delta, \frac{\sqrt{3}}{6}\delta)$.

We note that

$$B'_\gamma = \left[R\left(\varphi\right) - \mathbb{I}\right]^{-1}\left[B + R\left(-\frac{\pi}{3}\right)B\right],$$

and therefore

$$\mathrm{diam}(R\,B'_\gamma) = \frac{1}{\sqrt{3}}\,\mathrm{diam}\left[B + R\left(\frac{\pi}{3}\right)B\right] \le \frac{2}{\sqrt{3}}\delta,$$

and for all distinct points $\mathbf{t}_\gamma \neq \mathbf{t}_{\gamma'} \in \mathbb{L}$,

$$\|(\mathbb{I} - R(\varphi))^{-1}\mathbf{t}_\gamma - (\mathbb{I} - R(\varphi))^{-1}\mathbf{t}_{\gamma'}\| \ge \frac{\delta}{\sqrt{3}}.$$

It follows as before that if $\mathrm{diam}(B) \le \delta/2$, then

$$\widetilde{S_R^*} = \bigsqcup_{j=0}^{2} \left[p(R\,B'_\gamma) \boxplus \mathbf{w}_j\right], \tag{21}$$

where $\mathbf{w}_0 = \mathbf{0}$, and $\mathbf{w}_1, \mathbf{w}_2$ are as above. Therefore

$$V_2\left(\widetilde{S_R^*}\right) = 3\,V_2\left(B'_\gamma\right) = V_2\left[B + R\left(\frac{\pi}{3}\right)B\right]. \tag{22}$$

Since (22) is independent of the rotation $R = R(\theta)$, we conclude as before that the collision probability for $p3$ symmetry is given by

$$P_c(B) = \frac{V_2\left[B + R\left(\frac{\pi}{3}\right)B\right]}{V_2(F_{T\setminus X})}, \quad \text{for} \;\; \mathrm{diam}(B) \le \frac{1}{2}\,\delta. \tag{23}$$

IV. Demonstration with Elliptical Bodies

In this section we demonstrate formulas (19) and (23) for the case of elliptical bodies that satisfy the size restrictions imposed in the derivations leading to these equations. Consider a p2 unit cell that is a unit square. Then $V_2(F_{T\setminus X}) = 1$. Then, for bodies satisfying the diameter condition $\mathrm{diam}(B) \le 1/2$, formula (19) will hold. In particular, if B is an elliptical disk with semi-axis lengths $a, b \le 1/4$, then $V_2(B + B) = 4\,V_2(B) = 4\pi ab$. In the limiting case when $a = b = 1/4$, this means that $\pi/4$ of the unit volume, or more than 78 percent of the MR search space, need not be searched because we know a priori that it will lead to collision. Moreover, we know the shape of the regions that contribute to this volume and their location, and so search algorithms can be constructed to avoid them.

In the case of p3, let the unit cell be the one with the distance between neighboring lattice points be unity. Then the volume of the unit cell will be equal to that of two equilateral triangles with side lengths equal to unity. Then $V_2(F_{T\setminus X}) = \sqrt{3}/2$. The diameter condition under which (23) holds is $\mathrm{diam}(B) \le 1/2$. If we choose B to be an ellipse with semi-axis lengths $a, b \le 1/4$, then (23) can be used together with the results in [20] to compute in closed form the volume of the resulting Minkowski sum, as well as parametric descriptions of the curves that bound these regions. In the limiting case when $a = b = 1/4$, (23) gives

$$P_c(B) = \frac{\pi/4}{\sqrt{3}/2} = \frac{\pi}{2\sqrt{3}},$$

which means that more than 90% of the MR search space need not be searched.

V. Conclusions

The subset of a motion space corresponding to collision-free motions of rigid models of macromolecular structures has been analyzed. Methods from integral geometry in Euclidean space have been combined with new formulas specific to motion spaces to both parameterize free space, as well as to quantify the amount of free space to translate when a test molecule is oriented in a specific way. The details for two of the five planar crystallographic (wallpaper) groups were worked out. Other planar cases and the 65 spatial cases will be worked out in the future. By partitioning a motion space into feasible and infeasible regions, MR searches can be limited to only the feasible regions, the volume of which can be small in comparison to the volume of the whole motion space.

Acknowledgements: Chirikjian's research is supported by NSF grant IIS-1162095; Shiffman's research is supported by NSF grant DMS-1201372.

References

[1] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., "The protein data bank," *Nucleic acids research* 28, no. 1 (2000): 235-242.

[2] Blaschke, W., *Vorlesungen über Integralgeometrie*, Berlin, Deutscher Verlag der Wissenschaften, 1955.

[3] Brothers, J.E., "Integral Geometry in Homogeneous Spaces," *Trans. Amer. Math. Soc.*, 124: 408-517, 1966.

[4] Chern, S.-S., "On the Kinematic Formula in the Euclidean Space of N Dimensions," *American Journal of Mathematics* 74, no. 1 (1952), 227-236

[5] Chirikjian, G.S., "Mathematical Aspects of Molecular Replacement: I. Algebraic Properties of Motion Spaces," *Acta Cryst. A* (2011). A67, 435-446.

[6] Chirikjian, G.S., Yan, Y., "Mathematical Aspects of Molecular Replacement: II. Geometry of Motion Spaces," *Acta Cryst. A* (2012).

[7] Dunbar, W.D., *Fibered Orbifolds and Crystallographic Groups*, PhD Dissertation, Department of Mathematics, Princeton University, June, 1981.

[8] Grosse-Kunstleve, R. W., Wong, B., Mustyakimov, M., Adams, P. D., "Exact direct-space asymmetric units for the 230 crystallographic space groups," *Acta Cryst. A* 67(3), 269-275, 2011.

[9] Hahn, T., ed., *International Tables for Crystallography, Space Group Symmetry: Brief Teaching Edition*, 2002.

[10] Iversen, B., *Lectures on crystallographic groups*, Lecture Notes Series 60, Aarhus Universitet, Matematisk Institut, Aarhus, 1990.

[11] Klain, D.A., Rota, G.-C., *Introduction to Geometric Probability*, Cambridge University Press, 1997.

[12] Lučić, Z., and E. Molnár. "Fundamental domains for planar discontinuous groups and uniform tilings," *Geometriae Dedicata* 40(2):125-143, 1991.

[13] Orlov, I., Schoeni, N., Chapuis, G., "Crystallography on mobile phones," *J. Appl. Cryst.* 39, 595-597, 2006. http://escher.epfl.ch/mobile/

[14] Rossmann, M.G., "Molecular replacement - historical background," *Acta Cryst. D*57:1360-1366 (2001).

[15] Rossmann, M.G., Blow, D.M., "The Detection of Sub-Units within the Crystallographic Asymmetric Unit," *Acta Cryst.* 15:24-31 (1962).

[16] Santaló, L., *Integral Geometry and Geometric Probability*, Cambridge University Press, 2004 (originally published in 1976 by Addison-Wesley)

[17] Schneider, R., "Kinematic measures for sets of colliding convex bodies," *Mathematika* 25, 1-12, 1978.

[18] Schneider, R., Weil, W., *Stochastic and Integral Geometry*, Springer-Verlag, Berlin, 2008.

[19] Trandafir, R., "Problems of integral geometry of lattices in an Euclidean space E_3," *Bollettino dell'Unione Matematica Italiana*, 22(2), 1967.

[20] Yan, Y., Chirikjian, G.S., "Closed-form characterization of the Minkowski sum and difference of two ellipsoids," *Geometriae Dedicata* (to appear)