Proceedings of the ASME 2011 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference
IDETC/CIE 2011
August 28-31, 2011, Washington, DC, USA

DETC2011-48583

MOLECULAR REPLACEMENT FOR MULTI-DOMAIN STRUCTURES USING PACKING MODELS

Yan Yan

Department of Mechanical Engineering
Johns Hopkins University
Baltimore, MD 21218
Email: yyan9@jhu.edu

Gregory S. Chirikjian *

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

ABSTRACT

Molecular replacement (MR) is frequently used to obtain phase information for a unit cell packed with a macromolecule of unknown structure. The goal of MR searches is to place a homologous/similar molecule in the unit cell so as to maximize the correlation with x-ray diffraction data. MR software packages typically perform rotation and translation searches separately. This works quite well for single-domain proteins. However, for multi-domain structures and complexes, computational requirements can become prohibitive and the desired peaks can become hidden in a noisy landscape. The main contribution of our approach is that computationally expensive MR searches in continuous configuration space are replaced by a search on a relatively small discrete set of candidate packing arrangements of a multi-rigid-body model. These candidate arrangements are generated by collision detections on a coarse grid in the configuration space first. The list of feasible arrangements is short because packing constraints together with unit cell symmetry and geometry impose strong constraints. After computing Patterson correlations of the collision-free arrangements, an even shorter list can be obtained using the 10 candidates with highest correlations. In numerical trials, we found that a candidate from the feasible set is usually similar to the arrangement of the target structure within the unit cell. To further improve the accuracy, a Rapidly-exploring Random Tree (RRT) can be applied in the neighborhood of this packing arrangement. Our approach is demonstrated with multi-domain models in silico for 3D, with ellipsoids representing both the domains of the model and target structures. Configurations are defined by sets of angles between the ellipsoids. Our results show that an approximate configuration can be found with mean absolute error (MAE) less than 5 degrees.

INTRODUCTION

The field of structural biology is concerned with characterizing the shape, composition, flexibility, and motion of biological macromolecules and the complexes that they form. An ultimate goal of this field is to link these properties with the function of macromolecular structures, in the hope of better understanding biological phenomena and designing new drugs.

Here we review some of the issues involved in translating experimental data into 3D structures in the context of protein crystallography. Macromolecular X-ray crystallography (MX) has been the most used method for determining protein structures and associated complexes. It works very well for simple proteins that can be described as single rigid-bodies (called domains). This is because information about the shape of more than 70,000 previously solved structures in the Protein Data Bank (many of which are single-domain structures) can be used to augment new MX experimental information to gain a complete picture.

However, a challenge to MX arises in interpreting x-ray diffraction patterns for crystals composed of multi-domain sys-

*Address all correspondence to this author.

tems. This is because even when a multi-domain structure has been solved previously, its overall shape may vary widely from a new version of the structure with, for example, a bound drug. In this case, a widely used computational method called the molecular replacement method (MR) that is highly successful for single-domain proteins becomes combinatorially intractable due to the large number of degrees of freedom in multi-domain systems. We present a new method for phasing based on geometric packing that can serve as an alternative to MR.

The remainder of this paper is structured as follows. The mathematical aspects of the MR method for single-domain proteins is reviewed first. Then the multi-domain phase problem is formulated. Finally, we present out initial findings that diffraction patterns for multi-domain systems can be phased using our new "phasing by packing" method.

Essentials of Macromolecular X-Ray Crystallography (MX)

A biological macromolecule is a large collection of atomic nuclei that are stabilized through a combination of covalent bonds, hydrogen bonds, and hydrophobicity. A traditional goal in structural biology is to obtain the Cartesian coordinates of all atoms in a rigid single-domain protein.

Let $\mathbf{x}_i = (x_i, y_i, z_i)$ denote the Cartesian coordinates of the i^{th} of n atoms in a single-domain protein structure, and let $\rho_i(\mathbf{x})$ be the electron density of that atom in a reference frame centered on it. The density of the whole structure is then of the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} \rho_i(\mathbf{x} - \mathbf{x}_i). \tag{1}$$

The coordinates $\{\mathbf{x}_i\}$ are typically given either in a reference frame attached to a crystallographic unit cell, or to the center of mass of the protein.

MX does not provide $f(\mathbf{x})$ directly. Rather, it provides partial information about $f(\mathbf{x})$. The goal is then to computationally obtain $f(\mathbf{x})$ and fit an atomic model to it, thereby extracting the coordinates $\{\mathbf{x}_i\}$. But what information does an MX experiment provide? Well, a macromolecular crystal is composed of *unit cells* that have a discrete symmetry group, Γ. This symmetry group divides $\mathbb{R}^3$ into unit cells, and also describes how copies of the density $f(\mathbf{x})$ are located within the unit cell. The whole group Γ can be generated by translating unit cells and moving within the unit cell using generators $\{\gamma_1, ..., \gamma_m\}$. These form a subgroup of Γ, which is in turn a subgroup of the group of rigid-body motions, $SE(3)$, which will be denoted as G.

The result of an MX experiment for a single-domain protein is a diffraction pattern. This is the magnitude of the Fourier transform of the full contents of the crystallographic unit cell. Mathematically, this is written as

$$I(\mathbf{k}) = \left| \mathscr{F}\left(\sum_{j=1}^{m} f((\gamma_j \circ g)^{-1} \cdot \mathbf{x}) \right) \right|, \tag{2}$$

where $|\cdot|$ denotes the modulus of a complex number, $c = a + ib = |c|e^{i\phi}$. Here $g \in G$ and $\circ$ is the group operation for both G and Γ. In particular, it is well-known in robotics that each rigid-body motion consists of a rotation-translation pair $g = (R, \mathbf{t})$, and the composition of any two rigid-body motions g_1 and g_2 defines the operation $\circ$:

$$g_1 \circ g_2 = (R_1, \mathbf{t}_1) \circ (R_2, \mathbf{t}_2) = (R_1 R_2, R_1 \mathbf{t}_2 + \mathbf{t}_1). \tag{3}$$

Given that $g_i = (R_i, \mathbf{t}_i) \in G$ is a rotation-translation pair, its action on $\mathbb{R}^3$ is defined by

$$g_i^{-1} \cdot \mathbf{x} = R_i^T(\mathbf{x} - \mathbf{t}_i). \tag{4}$$

Then the density of a collection of single-domain proteins in the unit cell for $i = 1, ..., m$ will be $\sum_{i=1}^{m} f((\gamma_i \circ g)^{-1} \cdot \mathbf{x})$.

The difficulty in extracting $f(\mathbf{x})$ from the MX data is that this measurement folds in both information about $f(\mathbf{x})$ and the symmetry group Γ, and kills the phase information, $\phi(\mathbf{k})$, without which $f(\mathbf{x})$ cannot be recovered by inverse Fourier transform. Moreover, there is an unknown $g \in G$ that describes how each symmetry-related copy of $f(\mathbf{x})$ sits in the unit cell. This is what single-domain MR is mostly about - finding that unknown g.

The crystallographic space groups have been cataloged in great detail in the crystallography literature. For example, summaries can be found in [1, 2, 5, 18, 19, 24, 27, 28, 34] as well as in various online resources. Treatments of space group symmetry from the perspective of pure mathematicians can be found in [9, 13, 20, 22, 36, 41, 48].

Of the 230 possible space groups, only 65 are possible for biological macromolecular crystals (i.e., the chiral/proper ones). The reason for this is that biological macromolecules such as proteins and nucleic acids are composed of constituent parts that have handedness and directionality (e.g., amino acids and nucleic acids respectively have $C - N$ and $5' - 3'$ directionality). This is discussed in greater detail in [31, 35, 43, 46]. Of these 65, some occur much more frequently than others. And these are typically nonsymmorphic space groups (i.e., those that possess screw symmetry operations, and which cannot be described as a simple semi-direct product). For example, more than a quarter of all proteins crystallized to date have $P_{2_1 2_1 2_1}$ symmetry, and the three most commonly occurring symmetry groups represent approximately half of all macromolecular crystals [46, 53].

The number of proteins in a unit cell, the space group, Γ, and aspect ratios of the unit cell can be taken as known inputs in MR computations, since they are all provided by experimental observation. And from homology modeling, it is often possible to have reliable estimates of the shape of each domain in a multi-domain protein. What remains unknown are the relative positions and orientations of theses domains and the overall position and orientation of the symmetry-related copies of the proteins within the unit cell.

Once these are known, a model of the unit cell can be constructed and used as an initial phasing model that can be combined with the x-ray diffraction data. This is, in essence, the molecular replacement approach that is now more than half a century old [21,29,44,45]. Many powerful software packages for molecular replacement include those described in [6, 10, 38, 52]. Typically these perform rotation searches first, followed by translation searches.

Recently full 6 degree-of-freedom rigid-body searches and 6N DOF multi-rigid body searches have been investigated [23, 25, 26, 49] where N is the number of domains in each molecule or complex. These methods have the appeal that the false peaks and "noise" that results when searching the rotation and translation functions separately can be reduced. This paper analyzes the mathematical structure of these search spaces and examines what happens when rigid-body motions in crystallographic environments are concatenated. It is shown that unlike the symmetry operations of the crystal lattice, or rigid-body motions in Euclidean space, the set of motions of a domain (or collection of domains) within a crystallographic unit cell (or asymmetric unit) with faces "glued" in an appropriate way *does not* form a group. Rather, it has a quasigroup structure lacking the associative property.

The Multi-Domain Molecular Replacement Method (MMR)

The molecular replacement (MR) method, originally developed in the 1960s [11, 12, 29, 44] is a computational method for phasing x-ray diffraction data for biomolecular structures. It has been integrated into crystallographic structure determination codes [4, 10, 15, 39, 40, 50, 51]. For recent reviews and novel molecular replacement methods see [6, 14, 47, 52]. Though MR has been wildly successful for single-domain proteins, significant issues arise when using MR for multi-domain proteins and complexes.

Currently two major computational paradigms exist for phasing of x-ray diffraction patterns of multi-domain proteins: (1) use existing software packages to obtain candidate peaks in the rotation function for individual domains separately, then solve for the translation function [3, 7, 30, 37, 49]; (2) attempt to morph multi-domain candidate models that contain their full "6N" degrees of freedom and iteratively refine those models [16, 17, 23, 33, 42]. Both methods suffer from different aspects of the "curse of dimensionality," which we seek to circumvent using a combination of our initial results reported in [25] and new approaches based on advanced mathematical concepts that are new to the crystallography community.

Consider a multi-domain protein or complex consisting of N rigid bodies. If $f_i(\mathbf{x})$ denotes the density of the i^{th} body, then the density of the whole complex will be of the form $f(\mathbf{x}) = \sum_{i=1}^{N} f_i(g_i^{-1} \cdot \mathbf{x})$ where $g_i = (R_i, \mathbf{t}_i)$ is a rigid-body motion consisting of a rotation-translation pair and $g_i^{-1} \cdot \mathbf{x} = R_i^T(\mathbf{x} - \mathbf{t}_i)$. These motions are the unknowns in our problem.

If m identical copies of this complex are arranged symmetrically in a unit cell by symmetry operators $\gamma_j = (Q_j, \mathbf{a}_j) \in \Gamma$ (which is the group consisting of discrete rigid-body motions that are known a priori from the crystal symmetry and geometry), an x-ray diffraction experiment provides the magnitude (without phase) of the Fourier transform of $\sum_{j=1}^{m} f(\gamma_j^{-1} \cdot \mathbf{x})$. In contrast, the model density for a single domain and its symmetry mates is $\sum_{j=1}^{m} f_i(h_i^{-1} \circ \gamma_j^{-1} \cdot \mathbf{x})$ where h_i is the candidate position and orientation. The Fourier transform of the Patterson functions $P(g_1, ..., g_N; \mathbf{x})$ and $p_i(h_i; \mathbf{x})$ that correspond to these densities and their correlation are respectively

$$\hat{P}(g_1, ..., g_N; \mathbf{k}) = \left| \sum_{j=1}^{m} \mathscr{F}[f(\gamma_j^{-1} \cdot \mathbf{x})] \right|, \tag{5}$$

$$\hat{p}_i(h_i; \mathbf{k}) = \left| \sum_{j=1}^{m} \mathscr{F}[f_i(h_i^{-1} \circ \gamma_j^{-1} \cdot \mathbf{x})] \right|, \tag{6}$$

$$c(h_i) = \int_{\mathbf{x} \in \mathscr{C}} P(g_1, ..., g_N; \mathbf{x}) p_i(h_i; \mathbf{x}) d\mathbf{x} \tag{7}$$

where the Fourier transform $\mathscr{F}$ converts a function of spatial position, $\mathbf{x}$, into a function of spatial frequency, $\mathbf{k}$. Here $\mathscr{C}$ is the unit cell and in MR searches the hope is that peaks in the function $c(\cdot)$ correspond to $h_i = g_i$. The difficulty is that, unlike the single domain case, in the multi-domain case P depends on many g_j's that all interact with each other. Therefore, peaks in this rotational correlation function do not necessarily correspond to good overall matches.

Phasing By Packing

Instead of running traditional MR searches on domain orientation or full conformation, we propose to construct packing models for the multi-domain systems of interest. This will generate candidate sets of motions $\{h_1, ..., h_N\}$ that can then be used to construct a *model* of $P(h_1, ..., h_N; \mathbf{x})$ *rather than* $p_i(h_i; \mathbf{x})$. If $P(h_1, ..., h_N; \mathbf{x})$ and $P(g_1, ..., g_N; \mathbf{x})$ match well to each other, then that is a much stronger indication that $h_i = g_i$ than having high correlations between $p_i(h_i; \mathbf{x})$ and $P(g_1, ..., g_N; \mathbf{x})$.

But in order for our proposed approach to work, the fraction of the total $6N$-dimensional search space that we search must be very small. Otherwise it will be computationally expensive. In other words, we must rapidly determine "where not to look." Preliminary results along these lines are very encouraging. We hypothesize that the combination of crystal packing constraints and limitations on the range of motion between domains imposed by known motion constraints (in the case of multi-domain proteins consisting of covalently bonded rigid domains) severely restricts the allowable motions. And more restrictive symmetry groups than P1 will disallow relative motion even for smaller packing ratios. This leads us to believe that it will be possible to rapidly eliminate vast portions of high-dimensional configuration space based on their incompatibility with constraints, and that the enumeration of packing geometries can be performed in a computationally tractable manner.

In this paper, ellipsoids are used to represent different domains of protein structures. The reason is that the ellipsoid or the combination of ellipsoids can be used to describe a large variety of shapes and also be expressed in simple closed-form equations. To illustrate our approach, we construct a multi-ellipsoid-shaped "rabbit" with one "face" and two "ears" as a packing model for a 3-domain structure in P1 crystal symmetry. The rabbit has 7 degrees of freedom (DOF)—roll (α_1), pitch(β_1) and yaw(γ_1) of the face and rolls ($\alpha_2,\ \alpha_3$) and pitches ($\beta_2,\ \beta_3$) of the two axisymmetric ears (Fig.1). The most important constraint of the motion is that the rabbits cannot collide with (or insert into) each other. With 50% volume ratio between the packing model and the unit cell (see the dimensions in Table 1), there is not much free room to move for the packing model. So the rabbits have to be "smartly" close packed in the configuration space to avoid collision, as most protein molecules are in real crystals. Fig.2 shows examples of packing configurations with and without collisions using our packing model in P1 symmetry, and the yellow part in (a) shows the collision areas. Also, some constraints on the motion between domains are imposed(see the ranges of motion for each DOF in Table 1).

The main procedures of finding phase information using packing models can be described by a flowchart in Fig. 3. In the first step, we discretize the configuration space by a coarse grid (in 10-degree increments in this paper), and detect collisions for the packing configurations on this grid. The collision detection function of an arbitrary point to an ellipsoid is defined as

$$H(\mathbf{x}) = (\mathbf{x}-\mathbf{x_0})^{\mathbf{T}}\mathbf{R}^{\mathbf{T}}\mathbf{A}\mathbf{R}(\mathbf{x}-\mathbf{x_0}), \tag{8}$$

where $\mathbf{x_0} = [x_0, y_0, z_0]^T$ represents the origin of the ellipsoid in Cartesian coordinates. $A = \mathrm{diag}[1/(r_x^2), 1/(r_y^2), 1/(r_z^2)]$, where r_x, r_y and r_z are semi-axis lengths in the x-, y- and z- axes, respectively. R is the rotation matrix that describes the rotation of the ellipsoid relative to the space-fixed frame. If $H(\mathbf{x})$ is less than 1, the point is inside the ellipsoid, and vice versa. For our rabbit packing model, we only need to check $H(\mathbf{x})$ for surface points on the center copy with other surrounding copies. With a closed form, evaluating $H(\mathbf{x})$ is much less computationally expensive compared to calculating $c(h_i)$ in traditional MR searches (see Eqn.(7)). After the collision detection, we reduce the whole configuration space to a much shorter list. In this paper, only 150 collision-free configurations are found out of 944784 ($9^3 \cdot 6^4$) on a 10-degree-grid search.

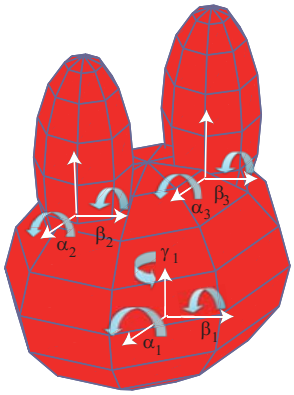

FIGURE 1. ILLUSTATION OF 7 DEGREES OF FREEDOM IN THE PACKING MODEL.

In the next step, we use a Fourier-based cost function (FCF), where

$$\begin{aligned}&\mathrm{FCF}(h_1,...,h_N) \\ &= \left[\int_{\mathbf{k}\in\Omega} \left(\hat{P}(g_1,...,g_N;\mathbf{k}) - \hat{P}(h_1,...,h_N;\mathbf{k})\right)^2 d\mathbf{k}\right]^{\frac{1}{2}},\end{aligned} \tag{9}$$

to sort these collision-free configurations from low to high. Minimizing $\mathrm{FCF}(h_1,...,h_N)$ is similar to finding peaks in $c(h_i)$ except that we use a multi-domain model rather than a single-domain one. After the sorting, we keep the 10 configurations with lowest FCF as a candidate list. These candidates indicate high correlations with the target structure. The FCF has the rugged surface of the configuration space, so to further improve the accuracy, a stochastic sampling method—Rapidly-exploring random tree (RRT) algorithm [32] is used to minimize the FCF around the "best candidate". The best candidate can be first chosen as the one with the lowest FCF in the set. If its FCF cannot be reduced below a threshold C after running the RRT, we switch the best candidate to the one with the next lowest FCF.

Numerical Results

In the numerical experiments, the same packing model is used each time to construct target structures. All of the angu-

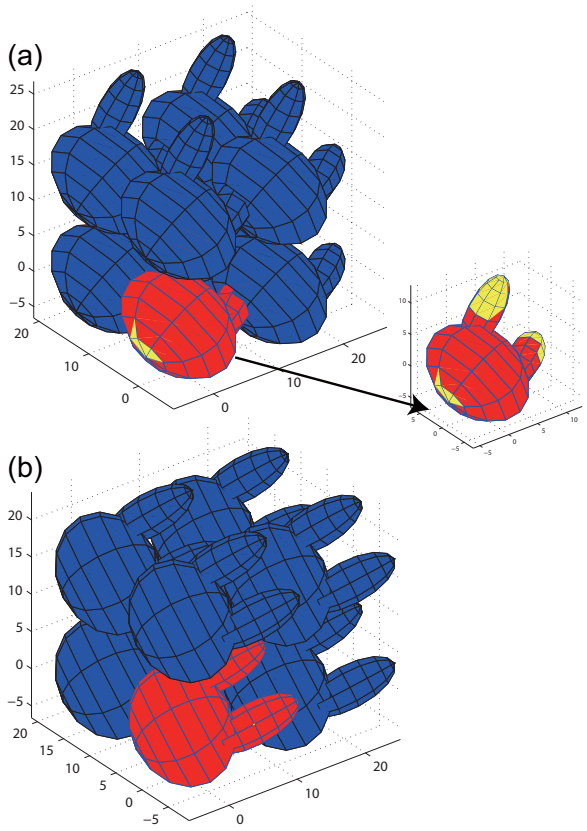

FIGURE 2. EXAMPLES OF PACKING CONFIGURATIONS (a) WITH COLLISIONS VS (b) WITHOUT COLLISIONS.

lar parameters of the target structure are treated as being unknown, and the only priori information that we have is the magnitude of the Fourier transform of the electron density function $\hat{P}(g_1,...,g_N;\mathbf{k})$. Our goal is to find the closest model configuration $\{h_1,...,h_N\}$ with respect to the target structure $\{g_1,...,g_N\}$. To evaluate the packing results, three different errors—E_h, MAE, and E_{max} are defined as,

$$E_h = \sum_i^N ||g_i - h_i||_W, \tag{10}$$

$$E_{max} = \max\{\Delta\alpha_1, \Delta\beta_1, \Delta\gamma_1, \Delta\alpha_2, \Delta\beta_2, \Delta\alpha_3, \Delta\beta_3\},$$

$$\text{MAE} = \text{mean}\{|\Delta\alpha_1|, |\Delta\beta_1|, |\Delta\gamma_1|, |\Delta\alpha_2|, |\Delta\beta_2|, |\Delta\alpha_3|, |\Delta\beta_3|\},$$

where E_h is the error metric of motion $\{h_1,...,h_N\}$ relative to $(g_1,...,g_N)$ and W is the weight matrix $\begin{pmatrix} J & \mathbf{0} \\ \mathbf{0^T} & M \end{pmatrix}$ as in [8]. Since the examples in this paper have symmetry group Γ=P1, and since there is no translation involved in P1 symmetry, g_i and h_i reduce to pure translations and W reduces to $J = \int_V \mathbf{x}\mathbf{x}^T \rho(\mathbf{x}) dV$. For ellipsoids, $J = \text{diag}[Mr_x^2/5, Mr_y^2/5, Mr_z^2/5]$, where M is the mass

TABLE 1. THE DIMENSIONS AND RANGES OF MOTION OF THE RABBIT PACKING MODEL.

	f	
Dimensions	size of the unit cell	$14 \times 14 \times 14$
	semi-axis lengths of the face	8; 6; 6
	semi-axis lengths of the ears	2.5; 2.5; 6
Face	range of roll (deg)	$0 \sim 90$
	range of pitch (deg)	$0 \sim 90$
	range of yaw (deg)	$0 \sim 90$
Ears	range of roll (deg)	-30 $\sim$ 30
	range of pitch (deg)	-30 $\sim$ 30

and r_x, r_y and r_z are the semi-axis lengths. Here $\rho(x)$ is taken to be 1 and semi-axis lengths are reported in Table 1. We note that the absolute value of E_h depends on the mass of the model. Also, E_{max} and MAE are maximum error and mean absolute error of the angle parameters, respectively.

To demonstrate the proposed approach, the angular parameters of target structures are generated in two ways: 1) chosen from the grid; 2) randomly sampled in the configuration space. We note that all the target structures should be collision free due to the physical constraints in the real world. In case 1 (see the example in Fig.4), the best candidate in the set is identical to the target structure, with three zero errors and zero FCF. When the target structure is randomly generated in the configuration space, as in case 2 (see the examples in Fig.5), we can see that the set of candidates (Table 2) show similar conformations as the target structure and the best candidate in the set (Cand.1) has only 3.9 degrees of MAE and 7.1 degrees of E_{max}. After running the RRT stochastic search for 30 steps, MAE is further reduced to 2.5 degrees and E_h is also decreased by 50%. Fig.6 shows the trends of errors before and after applying the RRT. The plot is generated by the results of 20 trials. In the figure, we can see both E_h and MAE go down as the RRT is running. We note that the reason for the slight increase in E_{max} may be caused by the different importance of the face and ears in FCF. The RRT places more weight on putting the face on the right position while the ears could be off alignment to a small extent.

Conclusions

Macromolecular crystallography has been the traditional workhorse for determining structural models in the field of biophysics. Within macromolecular crystallography, the molecular replacement method has been a highly successful method for providing phasing models to combine with experimental infor-

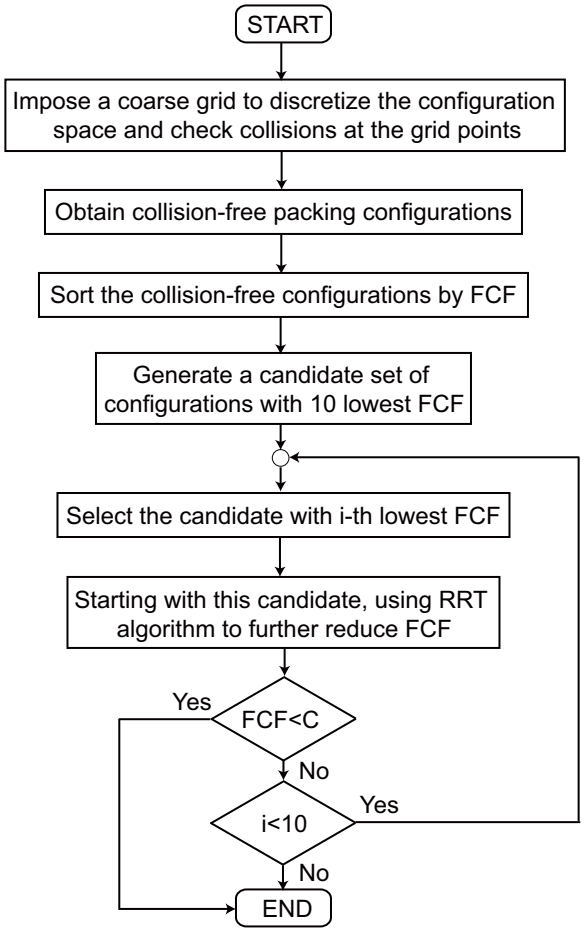

FIGURE 3. FLOWCHART OF PROCEDURES FOR GENERATING CANDIDATE PHASING MODELS BY PACKING.

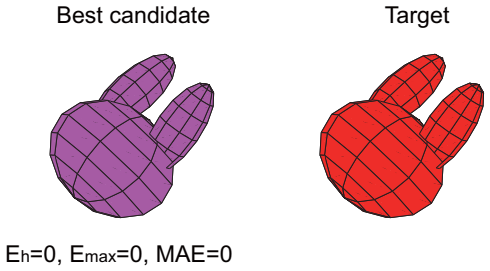

FIGURE 4. AN EXAMPLE OF PACKING RESULTS WITH THE ANGULAR PARAMETERS OF THE TARGET STRUCTURE CHOSEN FROM THE GRID.

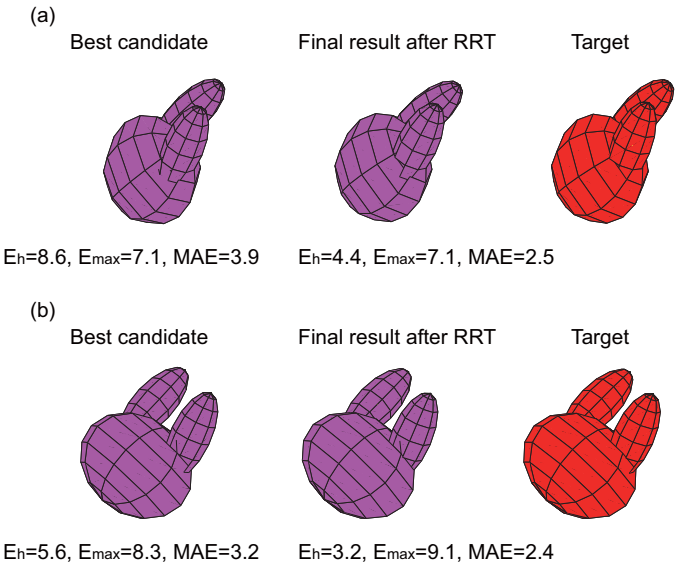

FIGURE 5. TWO EXAMPLES OF PACKING RESULTS WITH THE ANGULAR PARAMETERS OF THE TARGET STRUCTURES RANDOMLY SAMPLED IN THE CONFIGURATION SPACE.

mation to obtain 3D models. In this paper we demonstrate that an alternative to molecular replacement, called "phasing by packing" is promising for multi-rigid-domain structures. Numerical results illustrate the potential of this method.

ACKNOWLEDGMENT

We acknowledge NIH Grant R01GM075310 for the support of this work and Dr. E. Lattman for useful discussions.

REFERENCES

[1] Aroyo, M.I., et al, "Representations of Crystallographic Space Groups," Commission on Mathematical and Theoretical Crystallography, Nancy, France, June 28-July 2, 2010.

[2] Bradley, C., Cracknell, A., *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups*, Oxford University Press, 1972/2010.

[3] Brünger, A.T., "Extension of molecular replacement: a new search strategy based on Patterson correlation refinement," Acta Crysta, A46, 46-57, 1990

[4] Brünger, A.T., X-PLORE. Version 3.1. A System for X-ray Crystallography and NMR. Yale University Press, New Haven Connecticut, USA 1992.

[5] Burns, G., Glazer, A.M., *Space Groups for Solid State Scientists*, 2nd ed., Academic Press, Boston, 1990.

[6] Caliandro, R., Carrozzini, B., Cascarano, G.L., Giacovazzo, C., Mazzone, A., Siliqi, D., "Molecular replacement: the probabilistic approach of the program REMO09 and its applications," *Acta Cryst. A*65:512527 (2009).

[7] Castellano, E.E., Oliva, G., and Navaza, J. (1992). Fast rigid-body refinement for molecular-replacement tech-

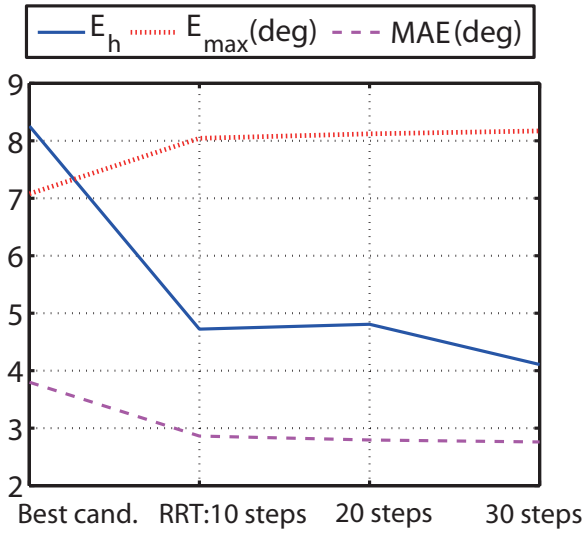

FIGURE 6. THE TREND OF ERRORS BEFORE AND AFTER APPLYING THE RRT ALGORITHM.

TABLE 2. THE ANGLE PARAMETERS OF THE TARGET STRUCUTRE AND THE SET OF CANDIDATES IN FIG.5 (a).

	α_1	β_1	γ_1	α_2	β_2	α_3	β_3
Target	46.3	50.2	11.0	-29.6	-29.9	-0.09	-12.1
Cand.1	45	45	15	-25	-25	-5	-5
Cand.2	45	45	15	-25	-25	5	-5
Cand.3	45	45	15	-25	-25	15	-5
Cand.4	55	55	5	-25	-25	-15	-25
Cand.5	55	55	5	-25	-25	-25	-25
Cand.6	35	45	15	-25	-25	-5	-5
Cand.7	35	45	15	-25	-25	5	-5
Cand.8	35	45	15	-25	-25	15	-5
Cand.9	35	45	15	-25	-25	25	-5
Cand.10	35	45	15	-25	-25	-15	-5

niques. *Journal of Applied Crystallography* **25:** Part 2 APR 1, 281-284.

[8] Chirikjian, G.S., Zhou, S., "Metrics on Motion and Deformation of Solid Models," *ASME J. Mechanical Design*, Vol. 120, No. 2, June, 1998, pp. 252-261.

[9] Conway, J.H., Delgado Friedrichs, O., Huson, D.H., Thurston, W.P., "On Three-Dimensional Space Groups," *Beiträge zur Algebra und Geometrie* Contributions to Algebra and Geometry Volume 42(2):475-507 (2001).

[10] Collaborative Computational Project Number 4, "The CCP4 suite: programs for protein crystallography," *Acta Cryst. D*50:760-766 (1994) http://www.ccp4.ac.uk/

[11] Crowther, R.A., and Blow, D.M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallographica* **23**, 544.

[12] Crowther, R.A. (1972). The fast rotation function. In *The Molecular Replacement Method*, M.G. Rossmann, ed. New York: Gordon and Breach Science Publishers, 173-178.

[13] Engel, P., *Geometric Crystallography: An Axiomatic Introduction to Crystallography*, D. Reidel Publishing Co., Kluwer, Boston, 1986.

[14] Evans, P., McCoy, A., "An introduction to molecular replacement," *Acta Cryst.* (2008). D64, 1-10

[15] Fitzgerald, P. M. D., "MERLOT, an integrated package of computer programs for the determination of crystal structures by molecular replacement," J. Appl. Cryst. (1988). 21, 273-278

[16] Glykos, N.M., Kokkinidis,"Multidimensional molecular replacement," Acta Crystallographica Section D - Biological Crystallography, D57, 1462-1473, 2001.

[17] Glykos, N.M., Kokkinidis, M. , "Structure determination of a small protein through a 23-dimensional molecular replacement search", Acta Crystallogr., D59, 709-718, 2003.

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

[19] Hammond, C., *The Basics of Crystallography and Diffraction*, Oxford University Press, Oxford, England, 1997.

[20] Hilton, H., Mathematical Crystallography and the Theory of Groups of Movements, 1903, (Dover Publications Inc. 1963.)

[21] Hirshfeld, F. L., "Symmetry in the Generation of Trial Structures," *Acta Cryst. A*24: 301-311 (1968).

[22] Iversen, B., *Lectures on Crystallographic Groups*, Aarhus Universitet Matematisk Institut Lecture Series 1990/91 No. 60, Fall 1990.

[23] Jamrog, D.C., Zhang, Y., Phillips Jr., G.N., "SOMoRe: a multi-dimensional search and optimization approach to molecular replacement," *Acta Cryst. D*59:304-314 (2003).

[24] Jannsen, T., *Crystallographic Groups*, North Holland/Elsevier, New York, 1973.

[25] Jeong, J., Lattman, E., Chirikjian, G.S., "A Method for Finding Candidate Conformations for Molecular Replacement Using Relative Rotation Between Domains of a Known Structure," *Acta Crstallographica D* D62, pp. 398-409, 2006.

[26] Jogl, G.,Tao, X., Xu, Y., Tong, L., "COMO: a program for combined molecular replacement," *Acta Cryst. D*57:1127-1134 (2001).

[27] Julian, M.M., *Foundations of Crystallography with Com-*

puter Applications, CRC Press, Taylor and Francis Group, 2008.
[28] Ladd, M.F.C., *Symmetry in Molecules and Crystals*, Ellis Horwood Limited/John Wiley and Sons, New York, 1989.
[29] Lattman, E.E., Love, W.E., "A Rotational Search Procedure for Detecting a Known Molecule in a Crystal," *Acta Cryst. B*26:1854-1857 (1970).
[30] Lattman, E.E., "Use of the Rotation and Translation Functions," Methods Enzymol. 115, 55-77, 1985
[31] Lattman, E.E., Loll, P.J., *Protein Crystallography: A Concise Guide*, The Johns Hopkins University Press, Baltimore, Maryland, 2008.
[32] LaValle, S. M. *Planning Algorithms*, Cambridge University Press, Cambridge, U.K., 2006.
[33] Liu, Q., Weaver, A.J., Xiang, T., Thiel, D.J., Hao, Q., " Low-dimensional molecular replacement using a six-dimensional search," *Crystallographica Section D - Biological Crystallography*, D59, 1016-1019, 2003
[34] Lockwood, E.H., MacMillan, R.H., *Geometric Symmetry*, Cambridge University Press, Cambridge England, 1978.
[35] McPherson, A., *Introduction to Macromolecular Crystallography*, John Wiley and Sons, Hoboken, NJ, 2003.
[36] Miller, W. Jr., *Symmetry Groups and Their Applications*, Academic Press, New York, 1972.
[37] Navaza, J. (1987). On the fast rotation function. *Acta Crystallographica Section A* **43**: Part 5, 645-653.
[38] Navaza, J., "AMoRe: an Automated Package for Molecular Replacement," *Acta Cryst. A*50:157-163 (1994).
[39] Navaza, J., and Saludjian, P. (1997). AMoRe: An automated molecular replacement program package. *Methods in Enzymology* **276**, 581-594.
[40] Navaza, J., "Implementation of molecular replacement in AMoRe," Acta Crystallographica Section D - Biological Crystallography, D57, 1367-1372, 2001.
[41] Nespolo, M., "Does mathematical crystallography still have a role in the XXI century?" *Acta Cryst. A*64:96111 (2008).
[42] Rabinovich, D., Shakked, Z., "A new approach to structure determination of large molecules by multi-dimensional search methods," Acta Cryst. A40, 195-200, 1984
[43] Rhodes, G., *Crystallography Made Crystal Clear*, 2^{nd} ed., Academic Press, San Diego, CA, 2000.
[44] Rossmann, M.G., Blow, D.M., "The Detection of Sub-Units within the Crystallographic Asymmetric Unit," *Acta Cryst.* 15:24-31 (1962).
[45] Rossmann, M.G., "Molecular replacement - historical background," *Acta Cryst. D*57:1360-1366 (2001).
[46] Rupp, B., *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, Taylor and Francis Group, New York, 2010.
[47] Sayre, D., "X-Ray Crystallography: The Past and Present of the Phase Problem," *Structural Chemistry*, Vol. 13, No. 1, February 2002
[48] Senechal, M., "A Simple Characterization of the Subgroups of Space Groups," *Acta Cryst. A*36:845-850 (1980).
[49] Sheriff, S., Klei, H.E., Davis, M.E., "Implementation of a six-dimensional search using the AMoRe translation function for difficult molecular-replacement problems," *J. Appl. Cryst.* 32:98-101 (1999).
[50] Tong, L., "REPLACE, a suite of computer programs for molecular-replacement calculations," J. Appl. Cryst. 26, 748-751, 1993.
[51] Vagin, A., Teplyakov, A.,"MOLREP: an Automated Program for Molecular Replacement," J. Applied Cryst., 30, 1022-1025 1997
[52] Vagin, A., Teplyakov, A., "Molecular replacement with MOLREP," *Acta Cryst. D*66:2225 (2010).
[53] Wukovitz, S.W., Yeates, T.O., "Why protein crystals favour some space-groups over others," *Nature Structural Biology*, 2(12):1062-1067, Dec., 1995.