Packing Models for Multi-Domain Biomolecular Structures in Crystals with $P2_12_12_1$ Space-Group Symmetry

Yan Yan and **Gregory S. Chirikjian**
Department of Mechanical Engineering
Johns Hopkins University
Baltimore, MD 21218, USA
gregc@jhu.edu

Abstract

In the context of X-ray crystallography, the molecular replacement (MR) method is frequently used to obtain phase information for a crystallographic unit cell packed with a macromolecule of unknown conformation. This is important because an X-ray diffraction experiment on its own does not provide full structural information. The shape and symmetry of the unit cell is determined by the space group symmetry. The most common space group for biological macromolecules is $P2_12_12_1$. 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, and then to use the model to add the unknown phase information to the experimental data. MR software packages typically perform rotation and translation searches separately. This works quite well for single-domain proteins that can be treated as rigid bodies. 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. First, candidate arrangements are generated by collision detections on a coarse grid in the configuration space. This list of feasible arrangements is short because collision-free packing requirement 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. Our results show that an approximate phase can be found with mean absolute error less than 5 degrees and search speed is effectively improved.

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 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 $\sim$ 80,000 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 systems. This is because even when a multi-domain structure has been solved previously, its overall shape may very 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), which has been 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. Decades ago, the concept of building models of crystallographic unit cells to phase crystallographic data was explored in the context of small molecules (Hendrickson and Ward 1976; Williams 1965; Damiani et al. 1967). But to our knowledge, this approach has not been pursued and is virtually unknown in the context of multi-domain macromolecular crystallography, and "phasing by packing" therefore represents a very different way of approaching the problem than 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 our 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. Due to thermal motions, the electron density of each of these atomic nuclei can be treated as a Gaussian distribution. The density of the whole structure is then of the form

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

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 $\{\mathbf{x}_i\}$. A macromolecular crystal is composed of *unit cells* that have a discrete symmetry group, Γ, called the crystallographic space group. This symmetry group divides $\mathbb{R}^3$ into unit cells, $U \cong \Gamma\backslash\mathbb{R}^3$ 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 here as G. The group $\Gamma = \mathrm{P2_12_12_1}$ is of particular importance because roughly one third of all biological macromolecules that have been crystallized to date have this symmetry.

The result of an MX experiment 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 for a single-domain protein as

$$\hat{P}(g; \mathbf{k}) = \left| \mathcal{F} \left(\sum_{j=0}^{m-1} f((\gamma_j \circ g)^{-1} \cdot \mathbf{x} \right) \right|, \quad (2)$$

where $|\cdot|$ denotes the modulus of a complex number, $c = a + ib = |c|e^{i\phi}$. Our reason for using the notation $\hat{P}(g; \mathbf{k})$ will be explained shortly. Here $g \in G$ is the unknown pose of the protein that is sought, 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 $\circ$:

$$g_1 \circ g_2 = (R_1, \mathbf{t}_1) \circ (R_2, \mathbf{t}_2) = (R_1R_2, R_1\mathbf{t}_2 + \mathbf{t}_1). \quad (3)$$

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

$$g \cdot \mathbf{x} = R\mathbf{x} + \mathbf{t}. \quad (4)$$

Then the density of a collection of single-domain proteins in the unit cell for $j = 0, ..., m-1$ will be $\sum_{i=0}^{m-1} f((\gamma_i \circ g)^{-1} \cdot \mathbf{x})$.

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. Single-domain MR is mostly about finding the unknown g, and most commonly this is done by dividing the search into rotational and translational parts.

The crystallographic space groups have been studied in great detail in the crystallography literature. For example, summaries can be found in (Bradley and Cracknell 1972; Aroyo and et al 2010) as well as in various online resources. Treatments of space group symmetry from the perspective of pure mathematicians can be found in (Miller 1973).

Of the 230 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 (McPherson 2011; Rhodes 2010; Rupp 2009). 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 these symmetry groups cannot be described as a simple semi-direct product). For example, more than a third of all proteins crystallized to date have $P2_12_12_1$ symmetry, and the three most commonly occurring symmetry groups represent approximately half of all macromolecular crystals (Rupp 2009; Wukovitz and Yeates 1995).

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 (Rossmann and Blow 1962; Rossmann 2001). Many powerful software packages for molecular replacement include those described in (Navaza 1994; Bailey 1993). Typically these perform rotation searches first, followed by translation searches.

Recently full 6 degree-of-freedom (DOF) rigid-body searches and 6N DOF multi-rigid body searches have been investigated (Jamrog, Zhang, and Phillips Jr 2003; Jeong, Lattman, and Chirikjian 2006) 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.

The Multi-Domain Molecular Replacement Method (MMR)

The molecular replacement (MR) method, originally developed in the 1960s (Rossmann and Blow 1962; Crowther and Blow 1967; Crowther 1972; Lattman and Love 1970) is a computational method for phasing X-ray diffraction data for biomolecular structures. It has been integrated into crystallographic structure determination codes (Navaza 1994; Vagin and Teplyakov 1997). 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 (Lattman 1985); (2) attempt to morph multi-domain candidate models that contain their full "6N" degrees of freedom and iteratively refine those models (Jamrog, Zhang, and Phillips Jr 2003). 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 (Jeong, Lattman, and Chirikjian 2006) 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 n 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 $\sum_{j=0}^{m-1} f(\gamma_j^{-1} \cdot \mathbf{x})$. In contrast, the model density for a single domain and its symmetry mates is $\sum_{j=0}^{m-1} f_i(h_i^{-1} \circ \gamma_j^{-1} \cdot \mathbf{x})$ where h_i is the candidate position and orientation. In traditional MR, the Fourier transform of the Patterson functions, $\hat{P}(g_1, ..., g_N; \mathbf{k}) = \mathcal{F}[P(g_1, ..., g_N; \mathbf{x})]$ and $\hat{p}_i(h_i; \mathbf{k}) = \mathcal{F}[p_i(h_i; \mathbf{x})]$, that correspond to these densities and their correlation are respectively

$$\hat{P}(g_1, ..., g_N; \mathbf{k}) = \left| \sum_{j=0}^{m-1} \mathcal{F}[f(\gamma_j^{-1} \cdot \mathbf{x})] \right|, \quad (5)$$

$$\hat{p}_i(h_i; \mathbf{k}) = \left| \sum_{j=0}^{m-1} \mathcal{F}[f_i(h_i^{-1} \circ \gamma_j^{-1} \cdot \mathbf{x})] \right|, \quad (6)$$

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

where the Fourier transform $\mathcal{F}$ converts a function of spatial position, $\mathbf{x}$, into a function of spatial frequency, $\mathbf{k}$. The real-space Pattersons themselves are obtained by applying the inverse Fourier transform. Of the quantities in (5)-(7), $\hat{P}(g_1, ..., g_N; \mathbf{k})$ comes from the experiment (this is the multi-domain version of (2)), and $\hat{p}_i(h_i; \mathbf{k})$ and $c(h_i)$ are computed. Here $\mathcal{C}$ 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. 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 $P2_12_12_1$ crystal symmetry. The rabbit has 10 DOF—roll

(α_1), pitch (β_1) and yaw (γ_1) of the face, rolls ($\alpha_2,\ \alpha_3$) and pitches ($\beta_2,\ \beta_3$) of the two axis-symmetric ears and translation in x-, y- and z- directions (P_x, P_y, and P_z). (see Fig. 1). The most important constraint of the motion is that the rabbits cannot collide with (or insert into) each other. With $\sim 60\%$ volume ratio between the packing model and the unit cell (see the dimensions in Table 1), there is not much free space to move for the packing model. So the rabbits have to be "smartly" close packed in 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 $P2_12_12_1$ symmetry, and the yellow part in (a) shows 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 as follows. 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. In our packing model, we only need to check the collisions of the surface points on the center copy with other surrounding copies. With a closed-form equation, evaluating the collisions between ellipsoids is much less computationally expensive compared to calculating $c(h_i)$ in traditional MR searches (see (7)). After the collision detection, we reduce the whole configuration space to a much shorter list.

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

$$\begin{aligned} &\text{FCF}(h_1, ..., h_N) \qquad (8)\\ &= \left[\int_{\mathbf{k}\in\Omega} \left(\hat{P}(g_1, ..., g_N; \mathbf{k}) - \hat{P}(h_1, ..., h_N; \mathbf{k})\right)^2 d\mathbf{k}\right]^{\frac{1}{2}}, \end{aligned}$$

to sort these collision-free configurations from low to high. Minimizing 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 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 (LaValle 2006) 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 value C after running the RRT, we switch the best candidate to the one with the next lowest FCF.

Numerical Results

In the experiment, the same packing model is used to construct target structures. We pretend that the phase information of the target structure is unknown, and the only information 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 errors—E_h, MAE$_1$, and MAE$_2$, are

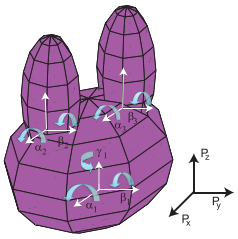

Figure 1: Illustration of 10 Degrees of Freedom in the Packing Model.

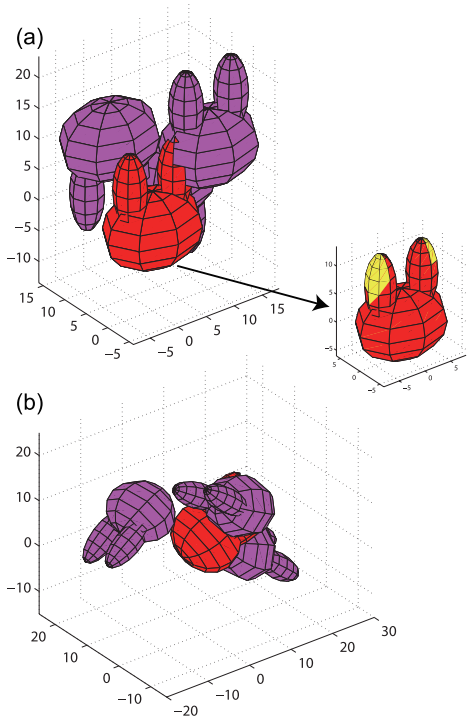

Figure 2: The Examples of Packing Configurations (a) with Collisions and (b) without Collisions.

defined as,

$$\begin{aligned} E_h &= \sum_{i=1}^{3} ||g_i - h_i||_W, \qquad (9)\\ \text{MAE}_1 &= \text{mean}\{|\Delta\alpha_1|, |\Delta\beta_1|, |\Delta\gamma_1|, |\Delta\alpha_2|, |\Delta\beta_2|, |\Delta\alpha_3|, |\Delta\beta_3|\},\\ \text{MAE}_2 &= \text{mean}\{|P_x|, |P_y|, |P_z|\}. \end{aligned}$$

where E_h is the error in metric of motion $\{h_1, ..., h_N\}$ relative to $(g_1, ..., g_N)$ and W is the weight matrix $\begin{pmatrix} J & \mathbf{0} \\ \mathbf{0^T} & M \end{pmatrix}$ with $J = \int_V \mathbf{x}\mathbf{x}^T\rho(\mathbf{x})dV$ and $M = \int_V \rho(\mathbf{x})dV$ (Chirikjian and Zhou 1998). Here, $W_1 = \text{diag}[1544.2, 868.6, 868.6, 120.63]$ for the face and $W_2 = W_3 = \text{diag}[19.6, 19.6, 113.1, 15.71]$ for the two ears. MAE$_1$, and MAE$_2$ are defined as the mean absolute errors for rotations and translations, respectively.

To demonstrate the proposed approach, target structures are generated in two ways: 1) strictly on the grid; 2) randomly sampled in the space. We note that all the target struc-

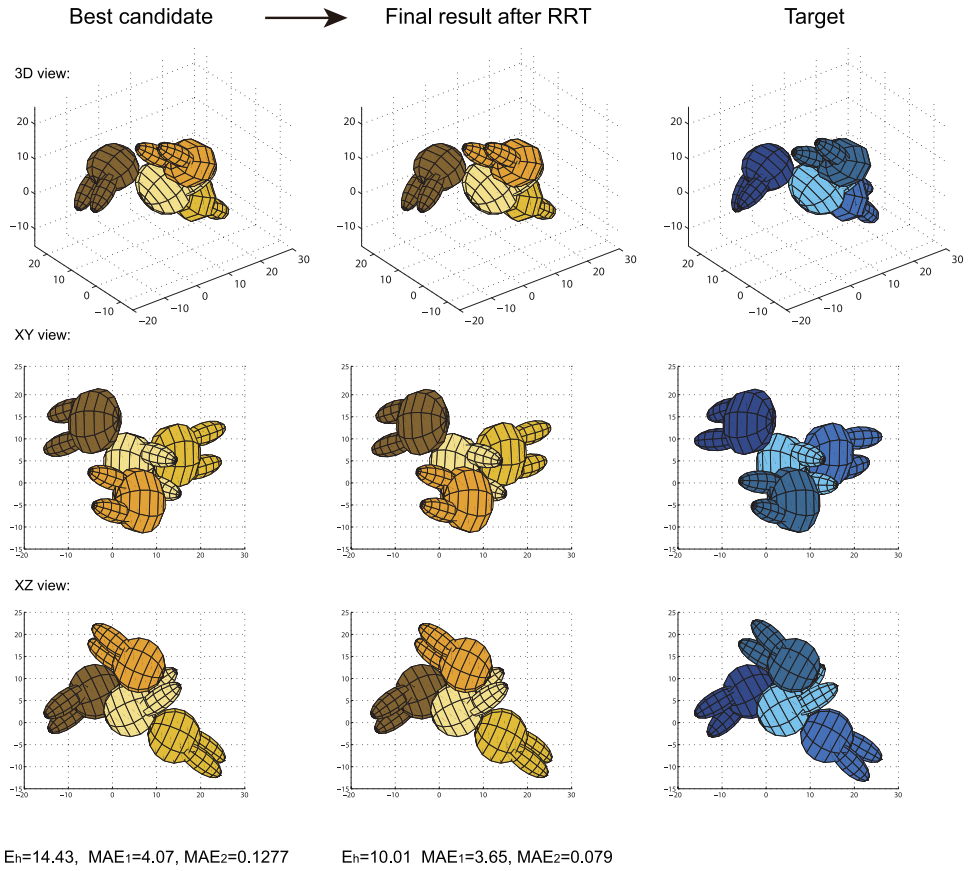

Figure 5: An Example of Packing Results with the Target Structure Randomly Sampled in the Space.

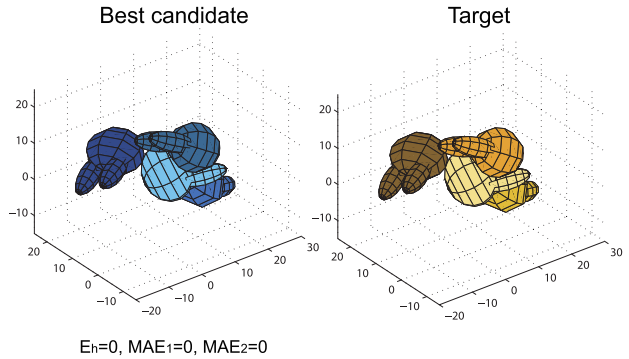

Figure 3: An Example of Packing Results with the Target Structure on the Grid.

tures should be collision free due to the physical constraints in the real world. In case 1 (see the example in Fig. 3), 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 space, as in case 2 (see the example in Fig. 5), we can see that the set of candidates show similar conformations as the target structure and the best candidate in the set (Cand.1) has only 4.07 degrees of MAE_1 and 0.1277 of MAE_2. After running the RRT stochastic search for 20 steps, MAE_1 and MAE_2 are further reduced to 3.65 degrees and 0.079, respectively and E_h is also decreased by 30%. Fig. 4 shows the trends of errors before and after applying the RRT. The plot is generated by the results of 10 trials.

Table 1: The Dimensions and Ranges of Motion of the Packing Model.

Dimensions	size of the unit cell	$20 \times 20 \times 20$
	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
Translation	x-axis	$0 \sim 4$
	y-axis	$0 \sim 4$
	z-axis	$0 \sim 3$

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 information 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 in the context of $P2_12_12_1$ space-group symmetry.

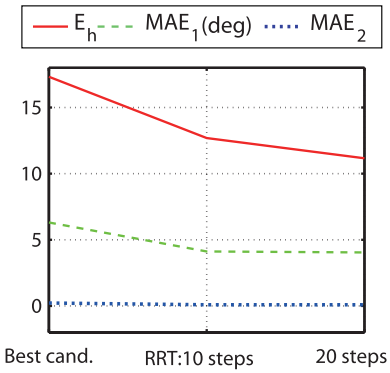

Figure 4: The Trend of Errors Before and After Applying the RRT.

Acknowledgements

We acknowledge Dr. E. Lattman for useful discussions.

References

Aroyo, M., and et al. 2010. Representations of crystallographic space groups, commission on mathematical and theoretical crystallography (online report).

Bailey, S. 1993. *The CCP4 suite: programs for protein crystallography*. Daresbury Laboratory.

Bradley, C. J., and Cracknell, A. P. 1972. *The mathematical theory of symmetry in solids: representation theory for point groups and space groups*. Clarendon Press Oxford.

Chirikjian, G., and Zhou, S. 1998. Metrics on motion and deformation of solid models. *Journal of Mechanical Design* 120:252.

Crowther, R. t., and Blow, D. 1967. A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallographica* 23(4):544–548.

Crowther, R. 1972. The fast rotation function. *Int. Sci. Rev. Ser.* 13:173–178.

Damiani, A.; Giglio, E.; Liquori, A. M.; and Mazzarell, L. 1967. Calculation of crystal packing: A novel approach to the phase problem. *Nature* 215:1161–1162.

Hendrickson, W. A., and Ward, K. B. 1976. A packing function for delimiting the allowable locations of crystallized macromolecules. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* 32(5):778–780.

Jamrog, D. C.; Zhang, Y.; and Phillips Jr, G. N. 2003. Somore: a multi-dimensional search and optimization approach to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography* 59(2):304–314.

Jeong, J. I.; Lattman, E. E.; and Chirikjian, G. S. 2006. A method for finding candidate conformations for molecular replacement using relative rotation between domains of a known structure. *Acta Crystallographica Section D: Biological Crystallography* 62(4):398–409.

Lattman, E. E., and Love, W. E. 1970. A rotational search procedure for detecting a known molecule in a crystal. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* 26(11):1854–1857.

Lattman, E. 1985. Use of the rotation and translation functions. *Methods in Enzymology* 115:55–77.

LaValle, S. M. 2006. *Planning algorithms*. Cambridge university press.

McPherson, A. 2011. *Introduction to macromolecular crystallography*. Wiley-Blackwell.

Miller, W. 1973. *Symmetry groups and their applications*, volume 50. Academic Press.

Navaza, J. 1994. Amore: an automated package for molecular replacement. *Acta Crystallographica Section A: Foundations of Crystallography* 50(2):157–163.

Rhodes, G. 2010. *Crystallography made crystal clear: a guide for users of macromolecular models*. Academic press.

Rossmann, M. G., and Blow, D. M. 1962. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica* 15(1):24–31.

Rossmann, M. G. 2001. Molecular replacement-historical background. *Acta Crystallographica Section D: Biological Crystallography* 57(10):1360–1366.

Rupp, B. 2009. *Biomolecular crystallography*. Garland Science.

Vagin, A., and Teplyakov, A. 1997. Molrep: an automated program for molecular replacement. *Journal of Applied Crystallography* 30(6):1022–1025.

Williams, D. E. 1965. Crystal packing of molecules. *Science* 147(3658):605–606.

Wukovitz, S. W., and Yeates, T. O. 1995. Why protein crystals favour some space-groups over others. *Nature Structural & Molecular Biology* 2(12):1062–1067.