Multiresolution phase extension of a trypsin inhibitor structure from 5 Å to 2 Å based on diffraction amplitudes alone

Xiangan Liu and Wu-Pei Su*
Department of Physics and Texas Center for Superconductivity, University of Houston, Houston, Texas 77204, USA
(Received 22 May 2009; published 22 October 2009)

For a specific structure of trypsin inhibitor, starting from a 5 Å phase set, about 80% of the 2 Å phases are correctly determined within an error of 18° by applying a multiresolution refinement procedure. The refinement proceeds both in real and reciprocal spaces. In extending the structure from 5 to 2 Å, the amplitudes of the reflections are the only requirements for this procedure. In contrast to the conventional phase extension scheme, the amplitudes are used not only in the Fourier synthesis but also in the real-space density modifications.

DOI: 10.1103/PhysRevE.80.047701 PACS number(s): 02.70.Tt, 61.05.cp, 42.30.Rx

I. INTRODUCTION

Refinement of protein structures is a common practice in macromolecular x-ray crystallography. Many new structures are determined initially only up to a certain resolution which is not accurate enough for chain tracing, hence a refinement procedure is needed for obtaining an atomic model. If the initial structure derives from a physical method such as anomalous scattering, chances are that the phases of some higher resolution reflections are available, at least approximately. In that case; the refinement consists basically of the improvement of the phases. On the other hand, if the initial low resolution structure is obtained through cryo-electron microscopy [1,2], which should be able to provide the initial phases of the low resolution reflections of the x-ray scattering, a phase refinement procedure is needed to extend the low resolution phases to high resolution ones. In this work, we report a trial calculation of the latter category.

We consider a synthetic structure of lima bean trypsin inhibitor. With the assumption that the phases are known up to 5 Å, using our multiresolution procedure, we are able to gradually extend the phases up to 2 Å. For each step, the refinement is first done under a certain resolution in real space. Many optimized substructures are generated using a biased Monte Carlo sampling algorithm [3]. An average of those substructures constitutes a new density map. This new density map is improved by a Fourier synthesis with modified amplitudes of the Fourier coefficients. A better resolved map is expected after those steps. This refinement procedure is done iteratively with higher resolution data being added in each iteration. This multiresolution refinement procedure requires only additional diffraction amplitudes at each step of the phase extension. The final 2 Å map tracks the atomic model very well. The trial calculation thus demonstrates that direct phase extension is possible starting from a medium resolution structure.

In the following, we first give the trial structure. The details of the refinement cycles are then described. Implications of the trial calculation are discussed in the conclusion.

*Author to whom correspondence should be addressed; wpsu@uh.edu

II. TRIAL STRUCTURE

For our calculations, we have picked a lima bean trypsin inhibitor structure (PDB ID: 1H34). The molecule contains 424 nonhydrogen protein atoms and 78 solvent atoms. The space group is $I2_13$ with cubic cell dimension 109.16 Å. There are 24 copies of the molecule in the unit cell. We use the atomic coordinates of the native structure to fabricate ideal reflection data. The reflections are indexed according to the original reflection data. Up to 5 Å there are 866 independent reflections. At 2.04 Å, there are 13 638 reflections altogether.

III. ALGORITHM

We start the iteration by generating an initial electron density model (Figs. 1 and 2) from the assumed ideal phases up to 5 Å. To represent the initial model in the following data processing, about ninety thousand point scatterers are picked with a probability proportional to the initial density. Normalized structure factors $E_{low-res}(hkl)$ are calculated for the above configuration of point scatterers up to 4 Å resolution (1788 reflections). The ensuing refinement procedure is executed with a biased simulated annealing algorithm [3]. To prepare for the biased Monte Carlo sampling, a smooth density map is obtained by blurring the point scatterers to a certain resolution. A mask [4] within the unit cell is chosen to cover regions where the smooth density is higher than a certain threshold. The mask is used to guide the Monte Carlo sampling to be described next.

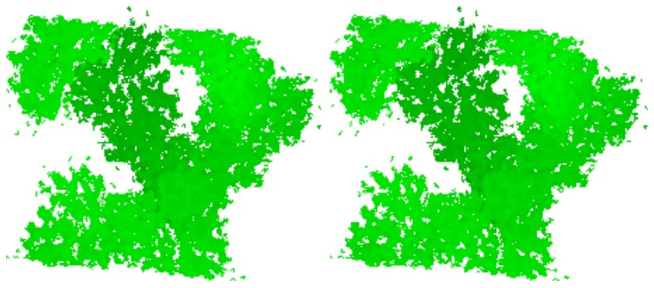

FIG. 1. (Color online) Stereoscopic view of a 5 Å density map of the lima bean trypsin inhibitor structure (PDB ID: 1H34) covering the entire unit cell.

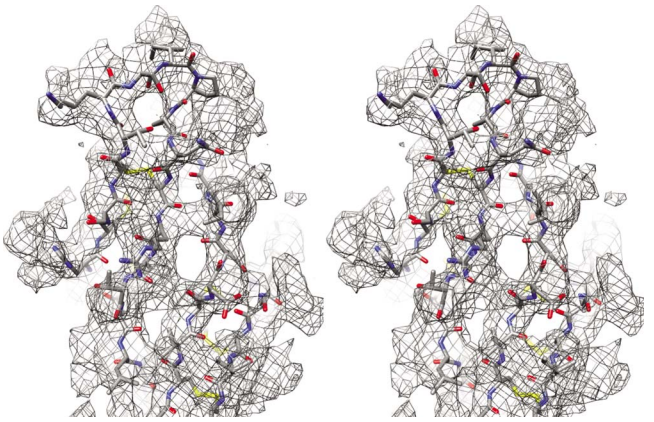

FIG. 2. (Color online) A zoomed-in stereoscopic view of the initial 5 Å density map superimposed on the corresponding part of the native model.

Apart from the ninety thousand point scatterers (referred to as set A), a small fraction (about 100 points in the asymmetric unit, 2400 points in the unit cell) of that (referred to as set B) is chosen to represent a difference density between the 4 Å map and the 5 Å map. A normalized structure factor $E_{cal}(hkl)$ of set B is added to the corresponding normalized structure factor $E_{low-res}(hkl)$ of set A to yield an averaged normalized structure factor $E_{avg-cal}(hkl)$. Let R be the residual of $|E_{avg-cal}(hkl)|$ with respect to the observed $|E_{obs}(hkl)|$ including all reflections hkl up to 4 Å,

$$R = \sum_{hkl} (|E_{avg-cal}(hkl)| - |E_{obs}(hkl)|)^2. \quad (1)$$

Set A is held fixed, whereas the positions of points in set B are varied to minimize the residual R in a simulated annealing scheme [5]. In updating the positions, only new positions within the mask are considered. That prevents the points from moving into low density regions. The employment of two densities A and B is crucial. Density A is used to

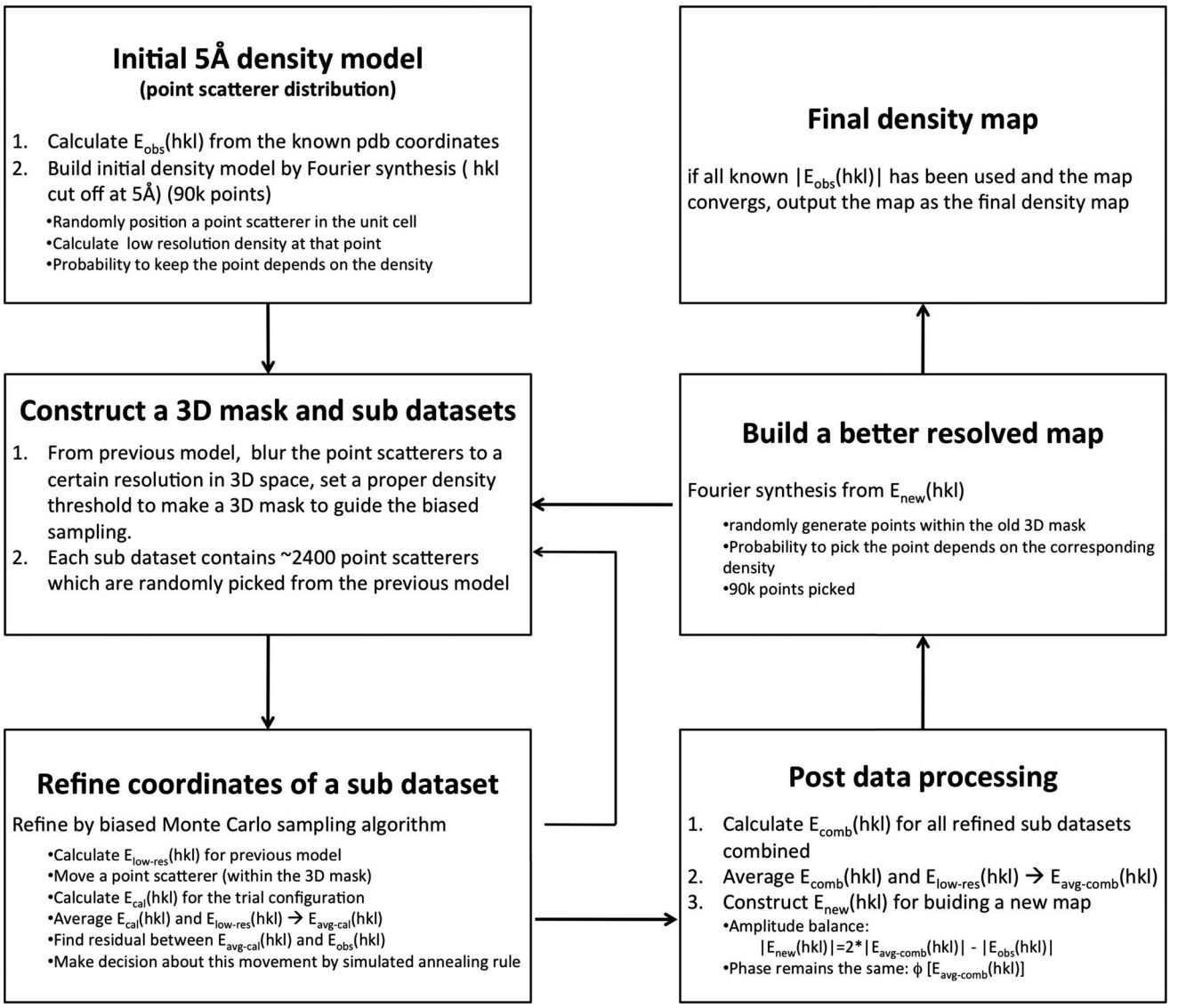

FIG. 3. A flow chart of the multiresolution refinement algorithm.

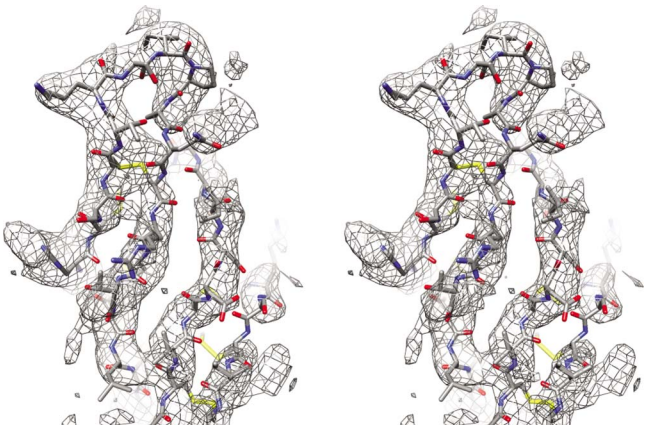

FIG. 4. (Color online) A zoomed-in stereoscopic view of the calculated 3.5 Å density map superimposed on the corresponding part of the native model.

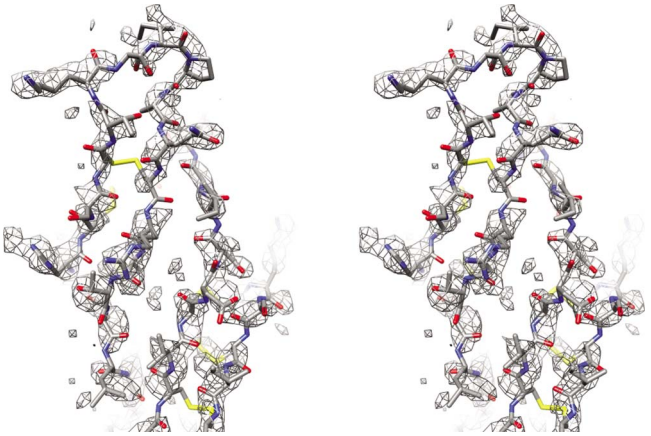

FIG. 5. (Color online) A stereoscopic view of the final calculated 2 Å density map superimposed on the native structure.

retain features of the original density map, whereas density B allows new features at a higher resolution to be introduced without disrupting the original features. Thus A together with B permits a stepwise increase in resolution consistent with the general idea of a multiresolution analysis [6].

The minimization of R does not lead to a unique configuration of set B. A combination of many (about 100) such configurations together with set A is used to evaluate the final normalized structure factor $E_{avg-comb}(hkl)$, the phase of which is combined with a new modified amplitude $2|E_{avg-comb}(hkl)|-|E_{obs}(hkl)|$ in a Fourier synthesis to produce a new density map of 4 Å resolution which forms the starting point of the next round of iteration. The simulated annealing calculation requires significant amount of computer time; that is why B is much smaller than A. A flow chart of the procedure above is given in Fig. 3.

After the 4 Å refinement, the resolution of the reflection data is increased to 3.5 Å (2700 reflections), 3 Å (4339 reflections), and 2.5 Å (7507 reflections) consecutively before reaching the final 2 Å resolution.

IV. RESULT

To gain some insight into the stepwise increase in the resolution of the electron density, the density maps at various resolution levels are depicted in Figs. 2, 4, and 5 in stereoscopic views. Figure 2 is a zoomed-in stereogram of the initial 5 Å density model superimposed on the native structure (only part of the molecule is shown). As the refinement proceeds, the map becomes more resolved. Figure 4 is a 3.5 Å calculated map compared with the native structure. The map's quality is significantly improved compared to the initial 5 Å density map (Fig. 2). The final 2 Å map is shown in Fig. 5. Figure 6 offers another more expanded view. 80% of the phases are less than 18° away from the true phases. From the stereograms, we can see that the map's quality is good enough to build an all-atom model.

V. DISCUSSION

Refinement of electron density map is a common procedure in the solution of new protein structures. Usually approximate values of most phases are known at the beginning, and phase refinement is done in conjunction with real-space density modifications such as solvent flattening, histogram matching and Sayre equation [7]. A Fourier synthesis using known Fourier amplitudes is carried out after that to produce a new density map.

We have followed a distinct approach in this work. From the known 5 Å phases of a trypsin inhibitor structure, we are able to gradually determine the higher resolution phases directly from the diffraction data alone [8]. At the end, we have reached an almost atomic resolution map. In contrast to the conventional density modification schemes, the diffraction amplitudes alone are employed to guide the modification. Conceivably, our procedure could be combined with histogram matching and Sayre equation to make it more powerful. Conversely, our methodology could be beneficially incorporated in the conventional refinement procedure.

An essential feature of our methodology is the stepwise increment in resolution range of the data used. During each

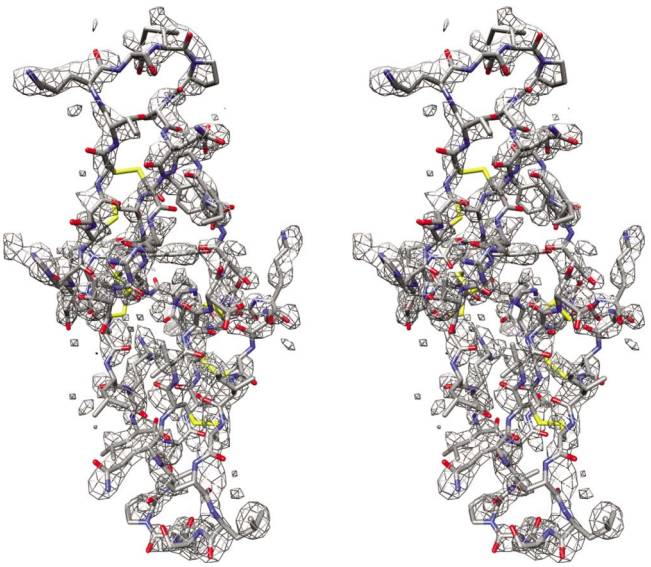

FIG. 6. (Color online) Another stereoscopic view of the final calculated 2 Å density map superimposed on the native structure.

step the density is modified first in real space and then in reciprocal space. The real-space density modification is guided by the diffraction amplitudes and a reference density (mask) at a lower resolution (see the bottom left box of Fig. 3). The mask greatly facilitates the search for optimal real-space configurations by narrowing down the search scope [3,9]. Further density modification is achieved by multiconfiguration average and Fourier amplitude correction as described in the bottom right box of Fig. 3. Without the division of the resolution range, it would be nearly impossible to achieve a 2 Å map directly from the 5 Å map simply because there are so many reflections that it is extremely difficult to modify the real-space density to optimize the residual.

As an extension of the present work, it is important to repeat the calculation on real diffraction data. It is also desirable to start the calculation with a lower resolution of the known phases. There are probably other ways to implement a multiresolution procedure; we have explored only one of them here. It is worthwhile to consider alternates [10,11].

ACKNOWLEDGMENTS

This work was partially supported by the Texas Center for Superconductivity and the Robert A. Welch Foundation (E-1070).

[1] Z. H. Zhou, M. L. Baker, W. Jiang, M. Dougherty, J. Jakana, G. Dong, G. Lu, and W. Chiu, Nat. Struct. Biol. **8**, 868 (2001).
[2] S. J. Ludtke, D. H. Chen, J. L. Song, D. T. Chuang, and W. Chiu, Structure **12**, 1129 (2004).
[3] X. Liu and W. P. Su, Phys. Rev. E **66**, 066703 (2002).
[4] P. Chacón, F. Morán, J. F. Díaz, E. Pantos, and J. M. Andrew, Biophys. J. **74**, 2760 (1998).
[5] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Science **220**, 671 (1983).
[6] W. P. Su, Physica A **221**, 193 (1995).
[7] M. Woolfson and H. F. Fan, *Physical and Non-physical Methods of Solving Crystal Structures* (Cambridge University Press, Cambridge, England, 1995).
[8] D. Sayre, Acta Crystallogr., Sect. A: Found. Crystallogr. **30**, 180 (1974).
[9] C. I. Chou and T. K. Lee, Acta Crystallogr., Sect. A: Found. Crystallogr. **58**, 42 (2002).
[10] S. T. Miller, J. M. Hogle, and D. J. Filman, J. Mol. Biol. **307**, 499 (2001).
[11] J. Wilson and P. Main, Acta Crystallogr., Sect. D: Biol. Crystallogr. **56**, 625 (2000).