

Article

# Trial Direct Phasing Calculation of A Thyroid Hormone Receptor Alpha Structure (4LNW)

Mengchao Jiang, Hongxing He  and Wu-Pei Su \*

Department of Physics and Texas Center for Superconductivity, University of Houston, Houston, TX 77204, USA; mch.jiang.nju@gmail.com (M.J.); hellohehongxing@gmail.com (H.H.)

\* Correspondence: wpsu@uh.edu

Received: 11 September 2019; Accepted: 14 October 2019; Published: 16 October 2019



**Abstract:** A thyroid receptor alpha structure (PDB ID: 4LNW) was studied for ab initio phasing. With the diffraction intensity data, protein sequence, and ligand structure as the only input, a high-resolution structure was successfully reconstructed by using an iterative projective method based on the hybrid input–output (HIO) algorithm.

**Keywords:** ab initio phasing; iterative projection algorithm; protein crystallography

## 1. Introduction

We previously applied a new iterative projective method based on the hybrid input–output (HIO) algorithm to the direct phasing of a series of structures with a solvent fraction typically exceeding 70% [1,2]. In this paper, we focus on a specific structure, the thyroid hormone receptor alpha (4LNW) [3] with a ligand molecule T3 and a solvent content less than 70%. In all of the trial calculations carried out so far, the known structures (the phases) were used to monitor the progress of the iteration in the calculation. To better simulate the process of solving an unknown structure, only the sequence, diffraction data, and ligand structure were given at the beginning of the calculation described in this paper. Only at the end of the calculation, the reconstructed map was compared with the known structure, thus removing all possible biases.

After a brief description of the 4LNW test structure in Section 2, a reference histogram is discussed in Section 3. Section 4 is devoted to methodology. Calculated density and models are presented in Sections 5 and 6.

## 2. Test Structure

From the 4LNW diffraction data, the obtained space group is  $P2_12_12_1$  with a solvent fraction of 69%. The cell dimensions are  $a = 59.78$ ,  $b = 80.79$ , and  $c = 102.56$  Å. The high-resolution end of the diffraction data is 1.9 Å, and the low one is 38 Å. There is a ligand bound to the protein; it is a T3 molecule  $C_{15}H_{12}I_3NO_4$ .

The above experimental data together with the protein sequence were downloaded from the protein data bank by one of the authors (W.-P.S.) and given to the two other authors without mentioning the PDB ID: 4LNW. M.J. and H.H. took the data of this unknown structure (to them) and carried out all of the following calculations.

## 3. Reference Histogram

Before the phasing calculation, a reference density histogram needs to be prepared. This describes the distribution of electron density inside a region that contains a protein together with some solvent region, i.e., a loose mask of the protein. It is well known that the histogram is independent of the details of the protein structure to a certain extent.

Generally a known structure from the same space group with a similar number of amino acids and similar solvent content is selected to compute the reference histogram. In this blind test, the structure 3RD5 was chosen to compute the protein histogram. The cell dimensions of 3RD5 are  $a = 64.920 \text{ \AA}$ ,  $b = 78.100 \text{ \AA}$ , and  $c = 86.220 \text{ \AA}$ . The solvent content is 65%. The crystal diffracts from 1.5 to 49.92  $\text{\AA}$ . Since the density histogram of a structure greatly depends on the temperature factor, we must rescale the B-factor of 3RD5 to match that of the test structure. The average B-factor of the test structure can be obtained directly from the diffraction data. It is worth mentioning that 3RD5 is a test structure with the lowest solvent content that we have successfully applied the direct method to.

#### 4. Methodology

With the above input data, M.J. and H.H. proceeded to apply an iterative projective algorithm. The tasks performed at each iteration are described below.

At the beginning of each iteration, the calculated electron density is Fourier transformed to yield both the magnitude and phase of each Fourier component. The calculated phases are combined with the observed magnitudes of the structure factors to construct a new density; this constitutes a Fourier refinement step. An averaged density based on this density is employed to compute an approximate protein mask.

The density inside the mask is modified according to the reference histogram mentioned before [4]. The density outside of the mask, inside the solvent region, is taken to be a certain combination of the old density (before Fourier transform) and new density (after the Fourier transform), hence the name “hybrid input-output (HIO)”. This combination is devised to drive the density gradually to zero, i.e., it is a softening of solvent flattening. This hybrid algorithm was found to be critical in avoiding stagnation of the iteration scheme, which can lead to incorrect phases.

The density modified both inside and outside of the mask is then fed into the next iteration cycle, which repeats itself until the correct density emerges after thousands of iterations. The HIO modification is replaced by strict solvent flattening toward the end of the calculation to obtain more accurate density.

The fractional volume of the protein mask is varied from calculation to calculation. To optimize the success rate, the optimal mask volume is 34%. The radius of the Gaussian used in the average density varies from iteration to iteration, it can go from 8  $\text{\AA}$  at the beginning of a calculation to 2  $\text{\AA}$  at the end.

The description of an iteration step in the phasing would seem to suggest that the algorithm works only on an approximately correct density function, otherwise the boundary would be ill-defined, for example. The amazing fact is that it works even with a completely random density at the beginning of the calculation, i.e., it can gradually modify a random density map into a correct density map.

Since the very low-resolution reflections contain many measurement errors due to very small reflection angles, we set a low-resolution cutoff at 18  $\text{\AA}$ . The measured reflections with a resolution less than 18  $\text{\AA}$  are removed from the work data set. All missing reflections, removed reflections, and the free reflections in the test data set are substituted by the calculated ones. For more details on the iterative transform algorithms, please refer to earlier publications [1,2,5].

#### 5. Results

Typically, 200 trial calculations were run for a given parameter set (solvent fraction,...) starting from random phases.  $R_{free}$  and  $R_{work}$  were calculated to monitor the progress of the iteration. Based on our experience, a sudden drop in R as a function of iteration number is a good signal of a successful run, as opposed to the smooth evolution of an unsuccessful one. Examples are shown in Figure 1. This criterion of success is corroborated by the distribution chart of the final Rs in Figure 2. The 200 runs clearly fall into two clusters on the diagram, with the successful runs corresponding to a smaller  $R_{free}$  and  $R_{work}$ .

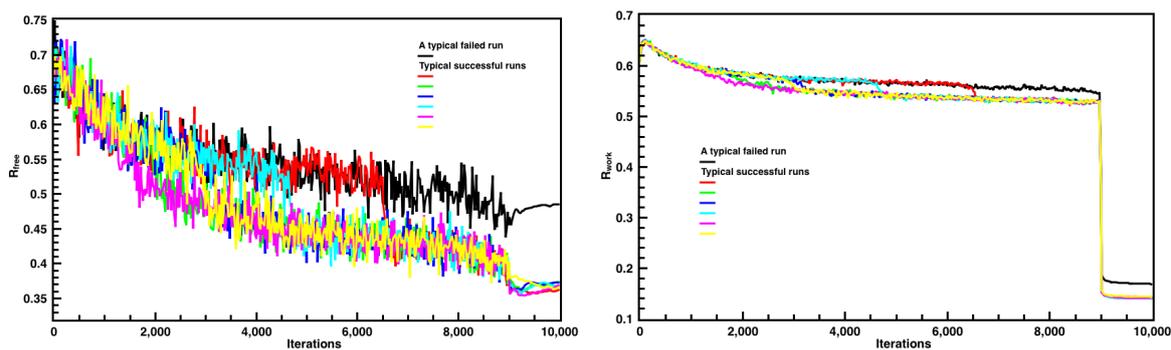


Figure 1. The evolution of  $R_{free}$  and  $R_{work}$ .

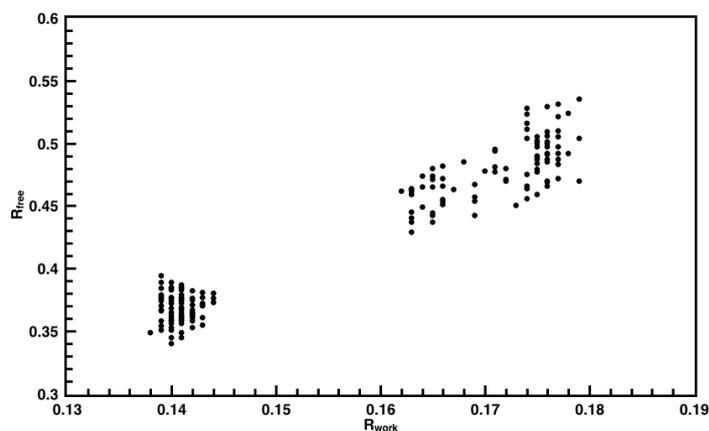


Figure 2. The distribution of the final  $R_{free}$  and  $R_{work}$ .

A close look at the density maps of the successful runs reveals small differences; therefore, 30 successful runs with the lowest Rs were averaged over [6] for model building with the CCP4 [7] software. The reconstructed model and the calculated protein boundary are shown in Figure 3. To examine how well the model fits the calculated density, we single out an alpha helix in Figure 4.

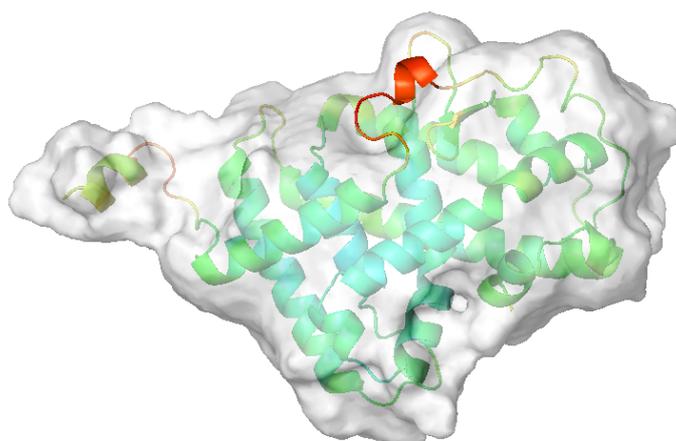
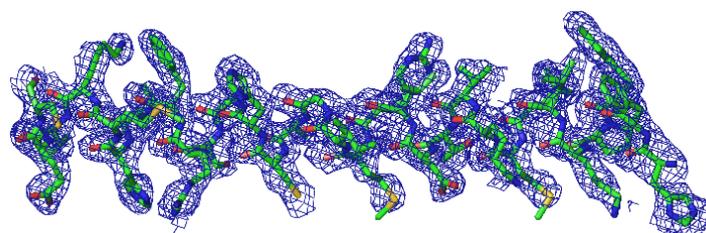


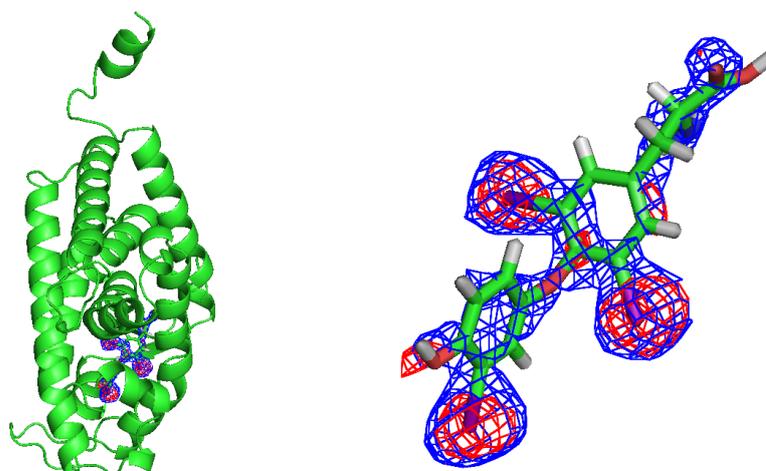
Figure 3. The reconstructed model and the calculated protein boundary of the test structure.



**Figure 4.** Comparison of calculated electron density with the reconstructed model on an  $\alpha$ -helix separated from the rest of the test structure.

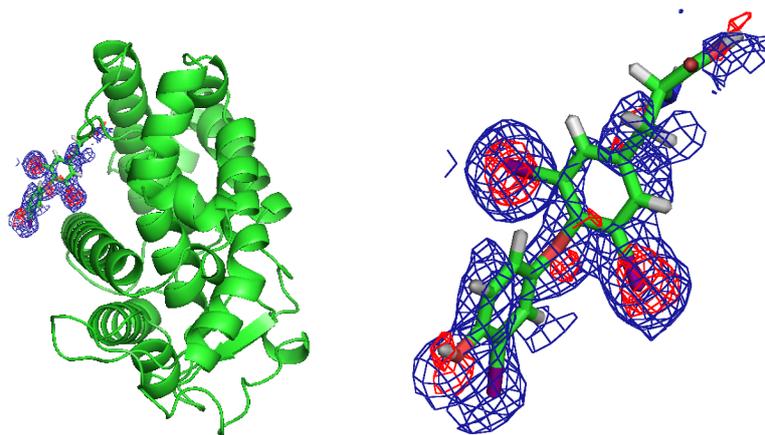
The ligand T3 molecule contains three heavy iodine atoms; they should lead to strong peaks in the calculated density maps. The COOT software [8] was used to calculate the difference map between the reconstructed model and the calculated density map. The density map of the rebuilt model without ligands diverges from the calculated density map. The two ligands can be located at regions with the largest divergence. As expected, the positions of iodine atoms can be visually detected on the difference map. After locating those iodine atoms, the two ligands were placed and adjusted by hand.

One ligand was found to be located inside the protein, and the other is attached to the surface of the protein. Figures 5 and 6 show the two ligands placed into the reconstructed model. The reconstructed model is displayed in cartoons. The two ligands are shown as sticks covered by the blue calculated density map. The difference map is presented as red mesh, which helps us to locate the six iodine atoms and the two ligands.

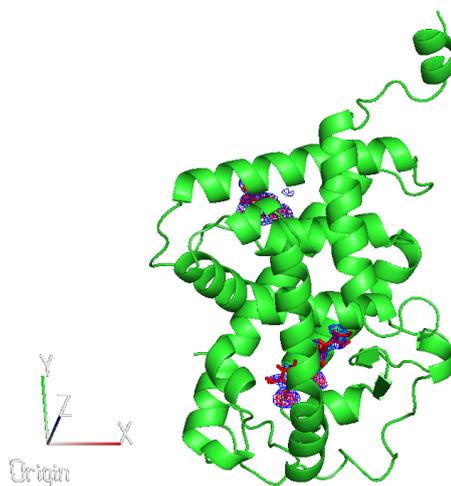


**Figure 5.** One ligand is buried inside the reconstructed model.

Figure 7 shows the reconstructed model with two ligands placed by hand according to the difference density map. As a consequence, a refinement is necessary to position the two ligands accurately.

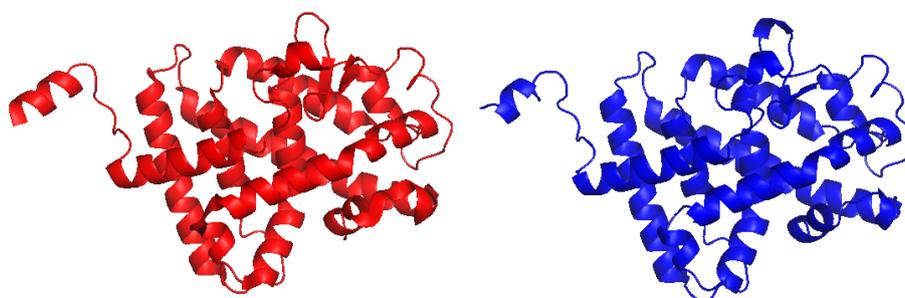


**Figure 6.** The other ligand is attached to the surface of the reconstructed model.

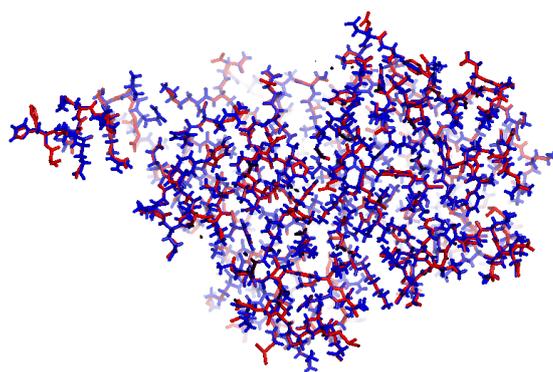


**Figure 7.** The two ligands and the reconstructed model of the test structure.

After locating the two ligands, M.J. and H.H. were informed that the test structure has the PDB code 4LNW. The calculated model was then compared to the deposited PDB model in Figure 8. All alpha helices were basically recovered correctly and most of the loops were correct too. A comparison in stick diagrams is shown in Figure 9.



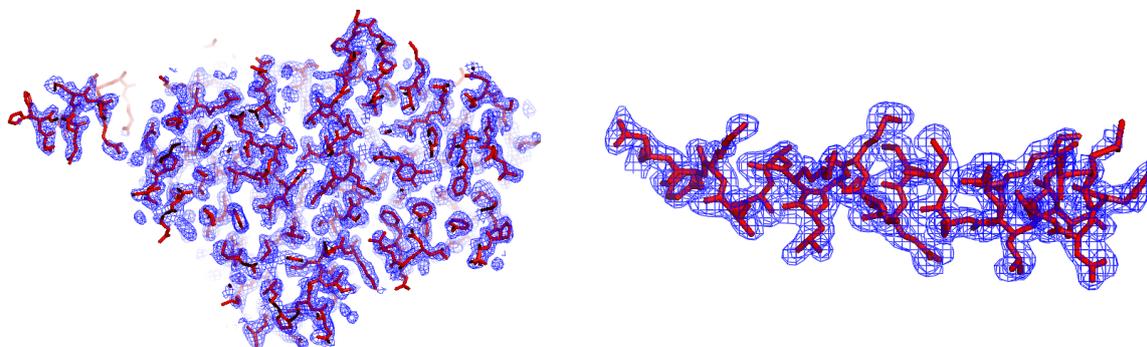
**Figure 8.** Comparison of the reconstructed model (**right**) with the PDB model (**left**).



**Figure 9.** Comparison of the reconstructed model (blue) with the PDB model (red) of the test structure.

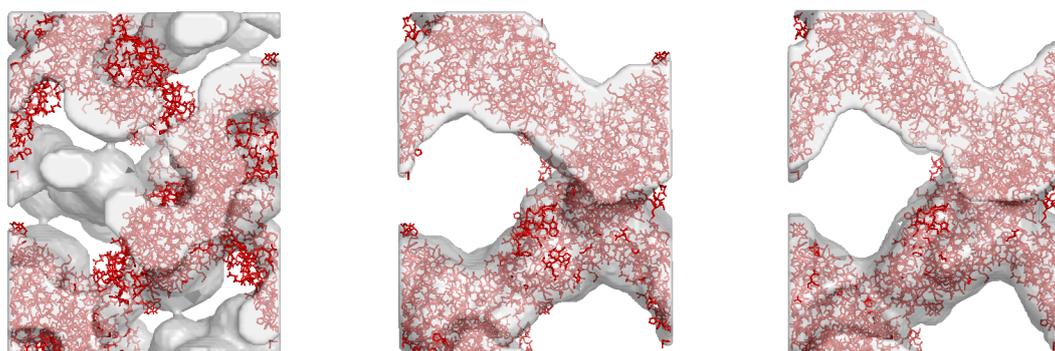
## 6. Further Comparisons with the True Structure

For a more detailed comparison, the true model structure is superimposed on sections of the calculated density in Figure 10. A very good match can be clearly seen.



**Figure 10.** The calculated density map superimposed on the PDB model.

The evolution of the calculated protein boundary is also compared with the PDB model in Figure 11, at various iteration numbers. A good agreement is reached at the 3000th iteration.



**Figure 11.** The reconstructed protein boundary at the 10th, 3000th, and 4000th iterations superimposed on the PDB model.

## 7. Discussion

Starting from the diffraction data, we were able to deduce a very good high-resolution density map of a thyroid hormone receptor alpha structure. A quite accurate model was built to fit the calculated density. We believe that with further refinements, the final PDB model could be completely recovered.

Since the 4LNW test structure was picked randomly from the PDB data bank, the trial calculation suggests that direct phasing is possible for some unknown structure with a high solvent fraction.

**Author Contributions:** M.J., and H.H. carried out the calculations; W.-P.S. designed the simulation and wrote the paper.

**Funding:** This research was funded by the Robert A. Welch Foundation under grant number E-1070.

**Acknowledgments:** This work was partially supported by the Texas Center for Superconductivity. The authors acknowledge the use of the Maxwell/Opuntia Cluster and the advanced support from the Center for Advanced Computing and Data Systems at the University of Houston to carry out the research presented here.

**Conflicts of Interest:** The authors declare no conflict of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## References

1. He, H.; Su, W.-P. Direct phasing of protein crystals with high solvent content. *Acta Cryst. A* **2015**, *71*, 92–98. [[CrossRef](#)] [[PubMed](#)]
2. He, H.; Su, W.-P. Improving the convergence rate of a hybrid input-output phasing algorithm by varying the reflection data weight. *Acta Cryst. A* **2018**, *74*, 36–43. [[CrossRef](#)] [[PubMed](#)]
3. Souza, P.C.; Puhl, A.C.; Martinez, L.; Aparicio, R.; Nascimento, A.S.; Figueira, A.C.; Nguyen, P.; Webb, P.; Skaf, M.S.; Polikarpov, I. Identification of a new hormone-binding site on the surface of thyroid hormone receptor. *Mol. Endocrinol.* **2014**, *28*, 534–545. [[CrossRef](#)] [[PubMed](#)]
4. Zhang, K.Y.J.; Main, P. Histogram matching as a new density modification technique for phase refinement and extension of protein molecules. *Acta Cryst. A* **1990**, *46*, 41–46. [[CrossRef](#)]
5. He, H.; Fang, H.; Miller, M.D.; Phillips, G.N., Jr.; Su, W.-P. Improving the efficiency of molecular replacement by utilizing a new iterative transform phasing algorithm. *Acta Cryst. A* **2016**, *72*, 539–547. [[CrossRef](#)] [[PubMed](#)]
6. Jiang, M.; He, H.; Cheng, Y.; Su, W.-P. Resolution dependence of an *ab initio* phasing method in protein X-ray crystallography. *Crystals* **2018**, *8*, 156. [[CrossRef](#)]
7. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta. Cryst. D* **2011**, *67*, 235–242. [[CrossRef](#)] [[PubMed](#)]
8. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and Development of Coot. *Acta Cryst. D* **2010**, *66*, 486–501. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).