

# Using RAPTOR to Find Homologous Protein Structures for Molecular Replacement Phasing of X-ray Crystallography

Libo Yu

Bioinformatics Solutions

## Introduction

X-ray crystallography is the most commonly used method in chemistry and biochemistry for determination of protein structure. We show here that RAPTOR software can be used to identify known structures for molecular replacement (MR) phasing of a target protein.

## Application

Solving a protein structure by X-ray crystallography involves three basic steps. The first step is to produce diffraction-quality crystals of the target protein. In the second step, an X-ray diffraction data set is collected from the crystal using an intense X-ray source. In the third step, the collected intensity data amplitudes are combined with phases to recover to the electron density of the crystal, from which the structure model of the protein can be generated.

The spots are related to the density of electrons in the crystal through Fourier transform. Therefore, if the Fourier transform is known, the electron distribution can be recovered from the diffraction pattern, from which the structure model can be built. By recording the intensity of the spots, the magnitude of the Fourier transform can be determined. The phase of the Fourier transform cannot be obtained directly from the diffraction pattern and must be computed in some other way. This is called the phase problem.

Molecular Replacement (MR) is a method for solving the phase problem. It is based on using a previously solved protein structure that is homologous to the target protein from which the diffraction data are obtained. In MR, by rotating and translating a solved structure into the unit cell of the target protein crystal,

initial phases for structure solution are obtained. The more homologous the solved protein structure is to the unknown target structure, the easier it is for the crystallographer to find the correct orientation and position of the target protein in the unit cell by MR.

RAPTOR software can be used to identify known structures for solving a target protein structure by the MR method, which can potentially shorten the process. RAPTOR has a structure template library derived from the PDB database. To identify the template that is most homologous to the target protein, RAPTOR threads the target protein's sequence to each structure in its library. Both sequence homology and structure homology have been used in the scoring function of RAPTOR to do threading. Frequently, the protein of interest cannot be readily associated with known 3-D structures due to the marginal sequence homology. The unique integer programming algorithm used in RAPTOR treats the pair-wise contact potential rigorously, which makes RAPTOR most effective for sequences with low homology and highly efficient in pinpointing structural homologues for the target.

RAPTOR's superior performance has been demonstrated in recent Critical Assessment of Structure Prediction (CASP), which are organized by NIH. The following table shows the performance of RAPTOR in CASP5

<b>Servers</b>	<b>Sum MaxSub Score</b>	<b># correct</b>
<i>3ds5 robeta</i>	5.17-5.25	15-17
<i>pmod 3ds3 pmode3</i>	4.21-4.36	13-14
<b>RAPTOR</b>	<b>3.98</b>	<b>13</b>
Shgu	3.93	13
<i>3dsn orfeus</i>	3.64-3.90	12-13
<i>pcons3</i>	3.75	12
<i>fugu3 orf_c</i>	3.38-3.67	11-12
...	...	...
pdbblast	0.00	0
...	...	...

blast	0.00	0
-------	------	---

RAPTOR is the No. 1 ranked individual server and correctly predicted 13 out of 20 FR (hard) targets. Meta-servers are listed in italics.

(<http://www.cs.bgu.ac.il/~dfischer/CAFASP3>, released on Dec., 2002.)

### **Case Study: Prediction of the cleavage site of the dystroglycan**

Post-translational modifications of the extracellular matrix receptor dystroglycan (DG) determine its functional state, and defects in these modifications are linked to muscular dystrophies and cancers. A prominent feature of DG biosynthesis is a precursor cleavage that segregates the ligand-binding and transmembrane domains into noncovalently attached  $\alpha$ - and  $\beta$ - subunits. Sequence and mutational analyses reveals that the cleavage occurs within a full SEA (sea urchin, enterokinase, agrin) domain with traits matching those ascribed to autoproteolysis. Researchers from the Lawrence Berkeley National Lab used server protein 3D structure prediction software to predict the cleavage domain in the DG. The predictions are shown in Figure 1. The structural models are validated by experimental analysis.

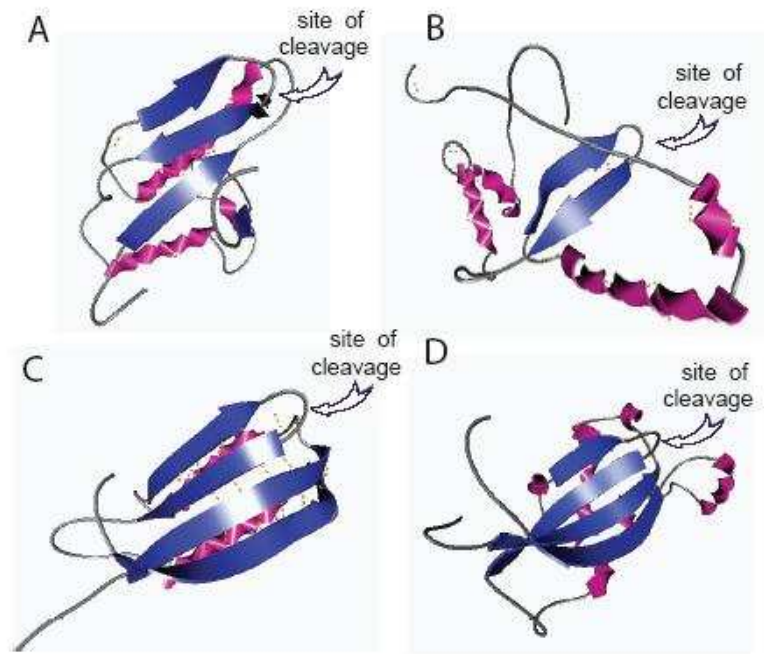


Figure 1: Modeling of the cleaved DG domain, obtained by A) I-TASSER, B) Robetta C) RAPTOR D) SAM-T06 algorithm

## Conclusion

We have shown that RAPTOR software can be an efficient tool in finding homologous structures. RAPTOR's high accuracy prediction can aid the molecular replacement process by identifying homologous known structures for phasing the diffraction data from the crystal of a target protein.

## References

1. Ch10 in "Principles of Protein X-ray Crystallography", by Jan Drenth (2nd Edn.) Springer, 1999
2. Jinbo Xu. Protein Fold Recognition by Predicted Alignment Accuracy. ACM/IEEE Transactions on Computational Biology and Bioinformatics, 2(2):157-165. 2005.
3. Jinbo Xu, Feng Jiao, Bonnie Berger. A Tree-Decomposition Approach to Protein Structure Prediction. CSB 2005, Stanford, USA.
4. Jinbo Xu, Ying Xu, Ming Li. Protein Threading by Linear Programming: Theoretical Analysis and Computational Results, Journal of Combinatorial Optimization, In Press 2004.
5. Jinbo Xu, Ming Li, Dongsup Kim, Ying Xu. RAPTOR: optimal protein threading by linear programming. Journal of Bioinformatics and Computational Biology 1:1(2003) 95-117.

6. Jinbo Xu and Ming Li. Assessment of RAPTOR's linear programming approach in CAFASP3. *Proteins: Structure, Function, and Genetics*, 53(S6): 579--584, Oct. 2003. Invited paper for CASP5, voted by peers as the "most innovative method in CASP5".
7. Akhavan, A. Crivelli, S. Singh, M. Lingappa, V.R. Muschler J.L. SEA domain proteolysis determines the functional composition of dystroglycan