

X-RAY SCATTERING TECHNIQUE REPLACES X-RAY CRYSTALLOGRAPHY IN PROTEIN MODELING: APPLICATION TO ZEBRAFISH ARBIN

I.M. KHATER

Biophysics Department, Faculty of Science, Cairo University, Giza, 12613, Egypt,
e-mail: ibrahim.khater@hotmail.com

Abstract. In this work, the native state of zebrafish arbin protein was modeled from protein sequence and was refined using solution X-ray scattering. The sequence of zebrafish arbin was converted to 3D structure using Iterative Threading Assembly Refinement Program (I_TASSER). The experimental X-ray scattering profile was fitted to theoretical X-ray profile of zebrafish arbin model by CRY SOL Program, and the chi error was 8.517. Zebrafish arbin model was superimposed to its average shape by SUPREF Program and the distance between them was 4.715. The *Phi* and *Psi* angles of amino acids of zebrafish arbin model were altered using *Phi* and *Psi* Change Tool of Swiss-pdbViewer. The chi and distance of the zebrafish arbin model were decreased to 3.421 and 2.255, respectively. 3D structure of many proteins can be solved by this method without the need of protein crystallization and complicated crystallographic technique.

Key words: Zebrafish arbin, protein structure, X-ray scattering.

INTRODUCTION

Solution small angle X-ray scattering (SAXS) is a simple technique that has been used to characterize tissue [4], ascertain the overall shape of a protein without the limits of a crystal array and identify native protein models from a large set of protein models [12, 21]. It has been also used in protein structure prediction algorithms to reconstruct proteins [8, 9, 16]; the combination of SAXS experimental data with protein structure prediction algorithms providing a method to predict structures closer to the native state [1, 3, 5]. For this purpose, some computational tools may be used. ATSAS package [10] is a free protein X-ray scattering program package. CRY SOL program [15], one of the ATSAS components, predicts the X-ray scattering profile starting from the protein 3D structure and fits it to the experimental one and computing the chi error value. Chi value is produced from a chi-squared test. The chi-squared test is used to determine

Received: August 2017;
in final form November 2017.

whether there is a significant difference between the expected values and the observed values and has no unit. DAMMIN [14] and GASBOR [17] are other ATSAS components that may predict the average shape of the protein starting from its experimental X-ray scattering data. SUPREF [7] is a program performing superposition refinement between any two 3D structures and computing the final distance between them. The program represents each input structure as an ensemble of points, then minimizes a normalized spatial discrepancy, or distance (NSD) to find the best alignment of two models. Distance is a measure of quantitative similarity between sets of three-dimensional points and is calculated in the following way: if two three-dimensional models are represented as a set of points, for every point in the first set (model 1), the minimum value among the distances between this point and all points in the second set (model 2) is found, and the same is done for the points in the second set. These distances are added and normalized against the average distances between the neighboring points for the two sets. Distance has no unit.

These programs can estimate if a certain protein conformation is close or far from the experimental X-ray scattering data and hence a native structure.

Theoretical modeling, such as homology, threading and *ab initio* modeling, is a developed method used to predict protein 3D structures, but the accuracy of these models could be low. The 3D structure of any protein backbone is determined by the *Phi* and *Psi* angles describing rotation around the backbone $C\alpha-N$ and $C\alpha-C$ bonds (Fig. 1).

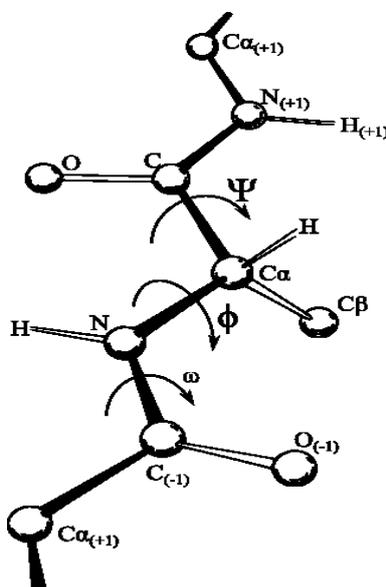


Fig. 1. *Phi* and *Psi* angles of amino acids [12].

Altering *Phi* and *Psi* angles of any amino acid may lead to a new conformation of protein which may be closer to, or far from its native state. X-ray scattering data can determine if this change is diverged to or converged from the native state of protein. Altering *Phi* and *Psi* angles of a low accuracy model and using X-ray scattering data can light the way to the native state of protein.

MATERIALS AND METHODS

MODELING OF ZEBRAFISH ARBIN

The sequence of zebrafish arbin was obtained from UNIPROT Database [2] with the code entry Q1LWJ6 (ARPIN_DANRE):

```
MSRIYDNTALLNKPVHNEKLSFTWDPIVHQSGHGVILEGTVVDFSRH  
AITDVKNRKERYNVLYIKPSRVHRRKYDSKGNEIEPNFSDTKKVNTGFLMS  
SYKVEAKGETDCLDERQLREIVNKEQLVKVTIKHCPREAFAFWISEAEMDK  
TELEPGQEVRLKTKGDGPFIFSFALDSTVTKCNFAGDENAGASWTEKIM  
ANKSNQENTGKSAAQEGEGADDDEWDD
```

It was converted to 3D structure using I_TASSER [20] Program. Five models were produced by I_TASSR Program.

X-RAY SCATTERING DATA OF ZEBRAFISH ARBIN

The experimental X-ray scattering profile of zebrafish arbin proteins was obtained from SASBDB [18] with entry code SASDBV2. The average shape of zebrafish arbin protein was constructed with DAMMIN and GASBOR Programs.

BEST MODEL SELECTION

Crysol program evaluating the solution scattering from five models of zebrafish arbin produced by I-TASSER and fitting them to experimental scattering curve from X-ray Scattering calculate the Chi between them. SUPREF Program was used to superimpose the average shape of zebrafish arbin to five models produced by I_TASSER and the distances between them were calculated. The best model (lowest chi and distance) was chosen as the starting point for the next step.

REFINEMENT OF BEST MODEL USING X-RAY SCATTERING DATA

The *Phi* and *Psi* angles of amino acids of the best zebrafish arbin model were altered using *Phi* and *Psi* Change Tool of Swiss-pdbViewer [6]. The minor change

of *Phi* and *Psi* was done by increasing and decreasing them by 20 degrees in four separate steps. The major change occurred by cycling each amino acid between the allowed regions of Ramachandran plot [11] shown in Figure 2.

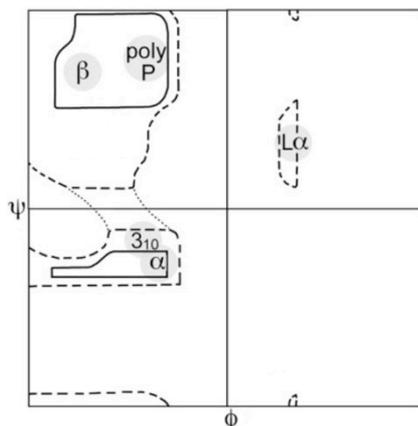


Fig. 2. Ramachandran plot.

Each minor or major change was followed by energy minimization using GROMOS 96 force field [19] implemented in Swiss Deep-Viewer. The experimental X-ray scattering profile was fitted to theoretical X-ray profile of altered zebrafish arbin model predicted by CRY SOL Program and the chi error was estimated. The altered zebrafish arbin model and the average shape of zebrafish arbin were superimposed using SUPREF Program and the distance between them was calculated. The previous steps were repeated for all amino acids of the best model, chi and distance were measured for every change. *Phi* and *Psi* change of an amino acid leads to new protein conformation. If this new conformation results in simultaneous decrease of chi and distance, it is the starting point for further processing and so on (something like iteration). If this new conformation did not result in simultaneous decrease of chi and distance, it will be ignored and the starting point of further processing is the previous conformation. The lowest chi and distance values represent the native state of zebrafish arbin.

RESULTS

The average shape of zebrafish arbin protein was constructed with DAMMIN and GASBOR Programs and is shown in Figure 3.

The chi and distance values for the five models of zebrafish arbin, modeled by I_TASSER Program (as described in Introduction), are shown in Table 1.

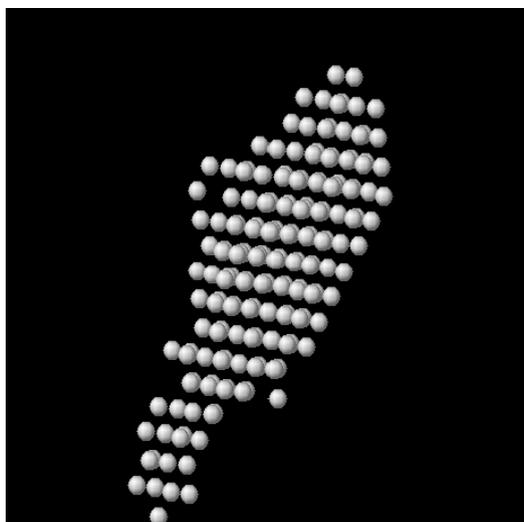


Fig. 3. Zebrafish arbin shape constructed by DAMMIN and GASBOR Programs.

Table 1

Chi and distance values for the five models of zebrafish arbin, modeled by I_TASSER Program

Model number	Chi	Distance
1	8.517	4.715
2	10.416	4.626
3	14.162	5.378
4	16.093	5.152
5	15.523	4.421

The model number 1 was chosen as the best one and is shown in Figure 4.

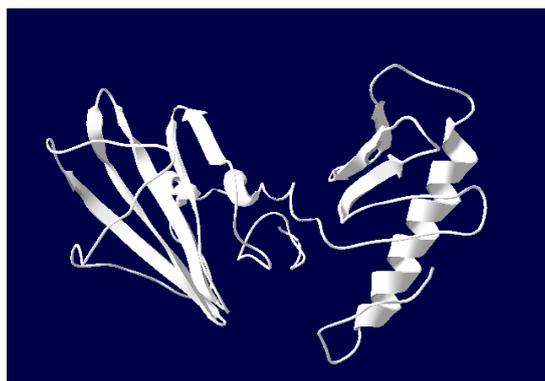


Fig. 4. Zebrafish arbin model 1; modeled by I_TASSER Program, was chosen as the best one.

The fitting of experimental X-ray scattering profile to theoretical X-ray profile of the model 1, predicted by CRY SOL Program, is shown in Fig 5.

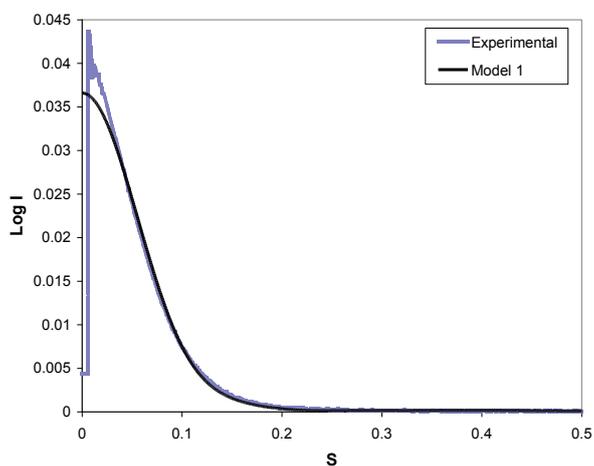


Fig. 5. The fitting of experimental X-ray scattering profile to theoretical X-ray profile of model 1, predicted by CRY SOL Program.

Models that resulted from alteration of *Phi* and *Psi* angles of amino acids of the backbone of zebrafish arbin model 1 are shown in Figs. 6(a), 6(b), 6(c), 6(d). The model in Fig. 6(d) has the lowest chi from experimental X-ray scattering data and the lowest distance from the average shape of zebrafish arbin constructed using X-ray scattering data. The fitting of experimental X-ray scattering profile to theoretical X-ray profile of model 6d predicted by CRY SOL Program was shown in Figure 7.

It could be said with high confidence that this model corresponds to the native state of zebrafish arbin.

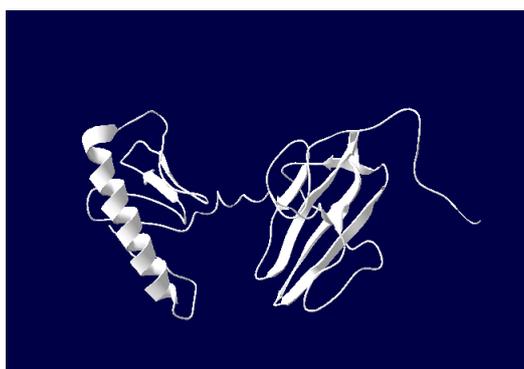


Fig. 6(a). chi = 4.76 and distance = 3.77.

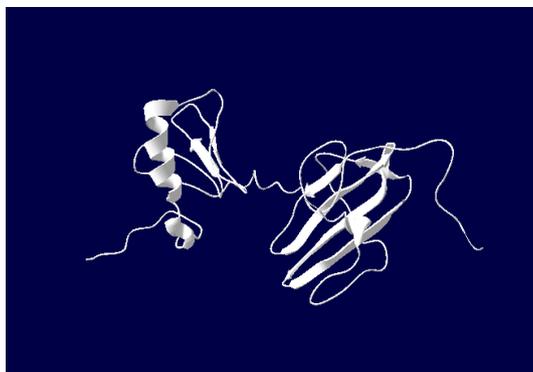


Fig. 6(b). $\chi = 4.367$ and distance = 2.582.

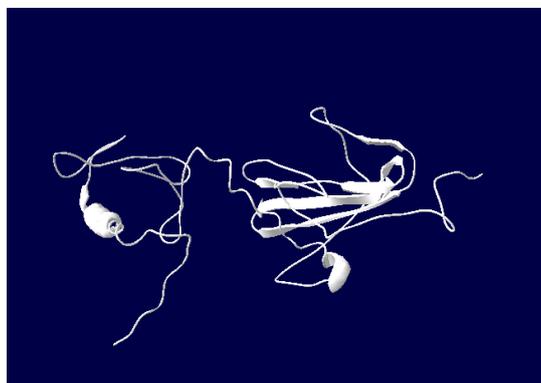


Fig. 6(c). $\chi = 3.591$ and distance = 2.314.

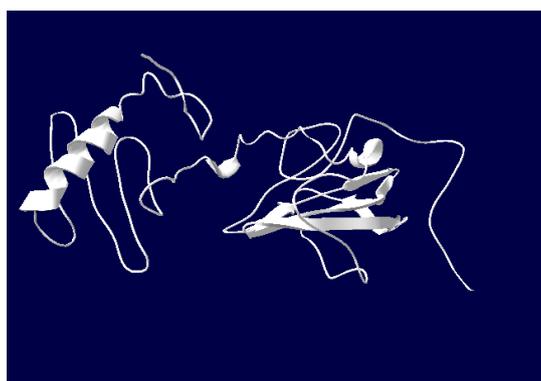


Fig. 6(d). $\chi = 3.421$ and distance = 2.255.

Fig. 6. Models that resulted from alteration of *Phi* and *Psi* angles of amino acids of zebrafish arbin, model 1.

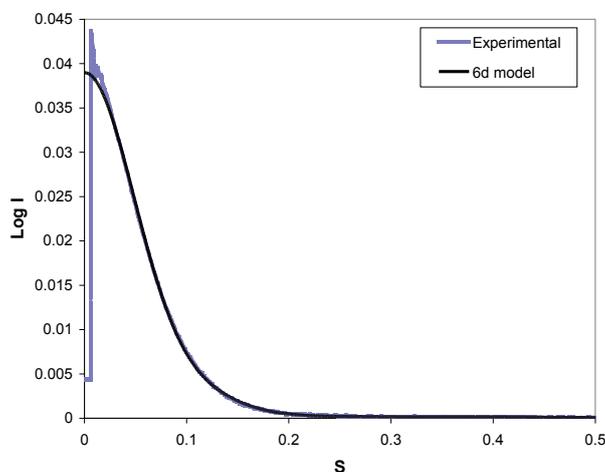


Fig. 7. The fitting of experimental X-ray scattering profile to theoretical X-ray profile of model 6(d), predicted by CRY SOL Program.

CONCLUSION

It is concluded that a high resolution model of any protein can be built, using the new method introduced in this paper, from protein sequence and X-ray scattering data which may represent the native state of the protein. This method opens the door to know the 3D structure of many proteins from their X-ray scattering data without the need of protein crystallization and complicated crystallographic technique.

REFERENCES

1. BERNADO, P., L. BLANCHARD, P. TIMMINS, D. MARION, R.W. RUIGROK, M. BLACKLEDGE, A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, *Proc. Natl. Acad. Sci. U S A*, 2005, **102**, 17002-17007.
2. BOECKMANN, B., A. BAIROCH, R. APWEILER, The SWISS-PROT protein knowledge base and its supplement TrEMBL, *Nucleic Acids Res.*, 2003, **31**, 354-370.
3. CHEN, B., X. ZUO, Y.-X. WANG, T.K. DAYIE, Multiple conformations of SAM-II riboswitch detected with SAXS and NMR spectroscopy, *Nucleic Acids Res.*, 2012, **40**(7), 3117-3130.
4. ELSHEMEY, W.M., S.M. FAYROUZ, M.K. IBRAHIM, X-ray scattering for the characterization of lyophilized breast tissue samples, *Radiat. Phys. Chem.*, 2013, **90**, 67-72.
5. GABEL, F., B. SIMON, M. NILGES, M. PETOUKHOV, D. SVERGUN, M. SATTler, A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints, *J. Biomol. NMR*, 2008, **41**(4), 199-208.
6. GUEX, N., M.C. PEITCH, Swiss model and the Swiss-pdb viewer: An environment for comparative protein modeling, *Electrophoresis*, 1997, **18**, 2714-2723.

7. KOZIN, M.B., D.I. SVERGUN, A software system for automated and interactive rigid body modeling of solution scattering data, *J. Appl. Crystallogr.*, 2000, **33**, 775–777.
8. MERTENS, H.D., D.I. SVERGUN, Structural characterization of proteins and complexes using small-angle X-ray solution scattering, *J. Struct. Biol.*, 2010, **172**(1), 128–141.
9. PETOUKHOV, M.V., D. FRANKE, A.V. SHKUMATOV, G. TRIA, A.G. KIKHNEY, M. GAJDA, C. GORBA, H.D.T. MERTENS, P.V. KONAREV, D.I. SVERGUN, New developments in the ATSAS program package for small-angle scattering data analysis, *J. Appl. Cryst.*, 2012, **45**, 342–350.
10. PETOUKHOV, M.V., D.I. SVERGUN, Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophys. J.*, 2005, **89**(2), 1237–1250.
11. RAMACHANDRAN, G.N., C. RAMAKRISHNAN, V. SASISEKHARAN, Stereochemistry of polypeptide chain configurations, *J. Mol. Biol.*, 1963, **7**, 95–99.
12. RICHARDSON, J.S., Anatomy and taxonomy of protein structures, *Advances in Protein Chemistry*, 1981, **34**, 167–339.
13. SONDERMANN, H., B. NAGAR, D. BAR-SAGI, J. KURIYAN, Computational docking and solution X-ray scattering predict a membrane-interacting role for the histone domain of the Ras activator son of sevenless, *Proc. Natl. Acad. Sci. U S A*, 2005, **102**(46), 16632–16632.
14. SVERGUN, D.I., Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing, *Biophys. J.*, 1999, **76**, 2879–2886.
15. SVERGUN, D.I., C. BARBERATO, M.H.J. KOCH, CRY SOL – A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, *J. Appl. Cryst.*, 1995, **28**, 768–773.
16. SVERGUN, D.I., M.H.J. KOCH, Small-angle scattering studies of biological macromolecules in solution, *Rep. Prog. Phys.*, 2003, **66**(10), 1735–1782.
17. SVERGUN, D.I., M.V. PETOUKHOV, M.H.J. KOCH, Determination of domain structure of proteins from X-ray solution scattering, *Biophys. J.*, 2001, **80**, 2946–2953.
18. VALENTINI, E., A.G. KIKHNEY, G. PREVITALI, C.M. JEFFRIES, D.I. SVERGUN, SASBDB, a repository for biological small-angle scattering data, *Nucleic Acids Res.*, 2015, **43**, D357–363.
19. VAN GUNSTEREN, W.F., S.R. BILLETER, A.A. EISING, P.H. HÜNENBERGER, P. KRÜGER, A.E. MARK, W.R.P. SCOTT, I.G. TIRONI, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996, pp. 1–1042.
20. ZHANG, Y., I-TASSER server for protein 3D structure prediction BMC, *Bioinformatics*, 2008, **9**, 40–47.
21. ZHENG, W., S. DONIACH, Fold recognition aided by constraints from small angle X-ray scattering data, *Protein Eng. Des. Sel.*, 2005, **18**(5), 209–219.