

P03.07.26*Acta Cryst.* (2008). A64, C226**Prediction of secondary structure and dihedral angles in proteins**

Tuping Zhou, Nanjiang Shu, Hovmoller Sven

Stockholm University, Structural Chemistry, Arrhenius Laboratory, Svante Arrhenius vag 12, Frescati, Stockholm, Stockholm, 10691, Sweden, E-mail: zhou@struc.su.se

A method for simultaneous prediction of secondary structure and dihedral angles of the polypeptide backbone in proteins is presented here. Based on a ten-fold cross-validation on a non-redundant set of 2670 protein chains with $\leq 25\%$ sequence identity, the three-state accuracy (Q3) is 81-82%. Every doubling of the number of non-redundant protein chains used in the training set results in 1% better prediction of secondary structure. With the dihedral angles discretized as 8 or 3 states on the Ramachandran plot, the accuracies for shape symbol prediction are 68.4% and 82.1% respectively. Thus, we show here for the first time that the conformations of all amino acids in proteins can be as accurately predicted as the secondary structure. Out of the residues predicted to be random coils with accuracy of 76.5%, 69.2% of corresponding shape symbols is predicted correctly in 3-state shape classification.

Keywords: structure prediction, secondary structure, dihedral angles

P03.11.27*Acta Cryst.* (2008). A64, C226**Crystal structure prediction of flexible molecules with genetic algorithms and standard force field**Julio C Facelli¹, Seonah Kim¹, Anita M Orendt¹, Marta B Ferraro², Ian Pimienta¹, Victor Bazterra¹

¹University of Utah, Biomedical Informatics and Center for High Performance Computing, 405 INSCC, University of Utah, SALT LAKE CITY, Utah, 84112-0190, USA, ²Departamento de Fisica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina, E-mail: julio.facelli@utah.edu

In this presentation we describe our distributed computing framework for crystal structure prediction, MGAC (Modified Genetic Algorithms for Crystal and Cluster Prediction) and its application to predict the structure of flexible molecules using CHARMM with the Generalized Amber Force Field (GAFF). MGAC is capable of performing crystal structure searches for flexible molecules within any space group and with an arbitrary number of molecules in the asymmetric unit. The distributed computing framework includes a series of tightly integrated computer programs for generating the molecule's force field, sampling possible crystal structures using a distributed parallel genetic algorithm, locally minimization of the structures and classifying, sorting and archiving the most relevant ones. Our results indicate the method can consistently find the experimentally known structures of a set of flexible molecules when GAFF reproduces the torsional energetics of the molecule, but unfortunately in some cases GAFF exhibit serious errors in describing this energetics. For instance the match between the experimental and predicted structures of norephedrine (racemic 2-amino-1-phenyl-1-propanol) has an RMS (root-mean-square) of 0.315 Å over a cluster of 15 molecules using the COMPACK, (Chisholm, J. A. and S. Motherwell, 2005, J. of App. Crystall. 38, 228) method. Both the experimental and predicted structure belong to the P21/c symmetry group, with the predicted cell parameters of (experimental values in parenthesis), $a = 12.447$ Å (12.507 Å), $b =$

8.293 Å (8.771 Å), $c = 7.808$ Å (8.130 Å) and $\beta = 104.63^\circ$ (106.20°).

Keywords: crystal structure prediction, crystal structure software, parallel algorithms

P03.11.28*Acta Cryst.* (2008). A64, C226**Hybrid genetic algorithm for a full-profile analysis of XRD powder patterns**Yaroslav Yakimov¹, Eugeny Semekin², Igor Yakimov³

¹Siberian Federal University, Krasnoyarskii rabochii 95, Krasnoyarsk, Krasnoyarsk, 660025, Russia, ²Siberian Federal University, Krasnoyarskii rabochii 95, Krasnoyarsk, Krasnoyarsk, 660025, Russia, ³Siberian Federal University, Krasnoyarskii rabochii 95, Krasnoyarsk, Krasnoyarsk, 660025, Russia, E-mail: I-S-Yakimov@yandex.ru

The full-profile fitting of powder patterns by Rietveld method is widespread tool for refinement of structural models as well as for quantitative phase analysis. The full-profile fitting is based on non-linear least-square method (NLSM) and is executed for different groups of refinement parameters consecutively. NLSM requires a good initial approximation of refinement parameters. Over the past few years, a genetic algorithm (GA) has been used to determine structural models of powders successfully. The focus of this work is to extend GA to the Rietveld method including full-profile fitting and refinement of crystal structure models and phase content. A hybrid two-level evolutionary genetic algorithm has been developed for this purpose. The hybrid algorithm is based on composition of a conventional GA with a NLSM designed as follows. First-level GA chromosomes comprise values of profile and structure parameters like that are used in the Rietveld method. Second-level GA chromosome is a bit string containing one bit per each parameter. Unit bits define the group of parameters to be refined with the NLSM on a current iteration. GA fitness function is the usual weighted profile R-factor (Rwp). The first-level GA determines initial parameter values of acceptable Rwp. The second-level GA manages NLSM full-profile fitting with found initial parameter values. A number of the parametric masks are used for reduction of dimensionality of the problem. The algorithm was implemented as a shell over the full-profile analysis program DDM [1] and was tested on some powder patterns of single and multi-phases samples with known stable crystal structures.

1. Solovyov L. A. // J. Appl. Cryst. 2004. 37. 743-749.

Keywords: evolutionary algorithm, full-profile analysis, Rietveld method

P03.11.29*Acta Cryst.* (2008). A64, C226-227**Consistency of particle shape determination from small-angle scattering data: Computer modeling**

Vladimir V. Volkov, Sergej V. Amarantov, Eleonora V. Shtykova

Institute of Crystallography Russian Academy of Sciences, Laboratory of Small-Angle Scattering, Leninsky prospekt, 59, Moscow, Moscow region, 119333, Russia, E-mail: vvo@ns.crys.ras.ru

Ab initio determination of shape of homogeneous particles from small angle scattering data obtained from monodisperse samples is considered. It was analytically shown that in the case of expansion of shape of homogeneous body in a limited orthogonal series of spherical harmonics the solution ambiguity consists in obtaining

enantiomorphous shapes. In the case of infinite series the possible solutions may not be enantiomorphous. An example of two different pyramidal bodies is considered. Another type of ambiguity is associated with numerical instability of solution. The dependence of solution dispersion from the maximum scattering angle is considered. Several numerical examples which demonstrate an optimum data angular range existence are given. The main conclusion that prior to experimental data interpretation one should to perform modelling of the expected structure to estimate the degree of ambiguity and to choose both data weighting functions and angular range is shown by the results of model calculations.

Keywords: *ab-initio* structure determination, nonlinear optimization, numerical methods and simulation techniques

P03.10.30

Acta Cryst. (2008). A64, C227

Development of a scoring method for predicting protein complex structures

Yuko Tsuchiya¹, Eiji Kanamori², Daron M Standley³, Haruki Nakamura³, Kengo Kinoshita¹

¹Institute of Medical Science, the University of Tokyo, Human Genome center, 4-6-1 Shirokane-dai, Minato-ku, Tokyo, 108-8639, Japan, ²Biomedical Information Research Center, 2-41-6 Aomi, Koto-ku, Tokyo, 135-0064, Japan, ³Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka, 565-0871, Japan, E-mail: yukoo@hgc.jp

The information about protein-protein interactions increases much more rapidly than the increase of the number of the tertiary structures of those protein complexes. Therefore, precise prediction of protein complex structures by protein-protein docking simulations is required. When the protein complex is re-built from its component protomers which derive from experimentally determined complex structure (native structure) by docking, the complex models with rmsd < 10Å from the native structure (near-native model) could be obtained, along with a great number of false positives (decoy). The separation of near-native models from many decoys is therefore needed in the prediction of complex structures by docking. In this study, we developed the method for scoring docking models so that the near-native models were higher in rank than decoys, based on the assumption that the interfaces of near-native models are more complementary in terms of surface properties and shapes compared to those of decoys. We used 125 non-redundant hetero-dimers (native structures) as targets. For each target, maximum 500 complex models were generated by our docking method. We also observed these targets in terms of the shape of the interfaces of their native structures. As a result, we found that these targets could be classified into two groups according to their interface shapes, and moreover, that this classification correlated with another classification which was based on the number of models with high docking score, namely, the difficulty in the separation of near-native models. We therefore only focused on 75 targets classified as difficult targets which need the separation. So far our method could separate the near-native models from the decoys in 70% of these targets.

Keywords: complex structure prediction, protein-protein docking, analysis of protein-protein interfaces

P03.10.31

Acta Cryst. (2008). A64, C227

3D homology structure model for a pyrazinamide susceptibility test in *Mycobacterium tuberculosis*

Luis R Castillo

Universidad Peruana Cayetano Heredia, Biochemistry, Molecular Biology and Pharmacology/Unidad de Bioinformatica, Av. Honorio Delgado 430, Urb. Ingeniería, San Martin de Porres, Lima, Lima 31, Peru, E-mail: eloluis87@yahoo.com

Pyrazinamide (PZA) constitutes one of the First-Line Drugs for Tuberculosis treatment and appears to be the most important drug killing the latent *M. tuberculosis*. It is hydrolyzed by *Mycobacterium tuberculosis* pyrazinamidase / nicotinamidase (PZAse) to pyrazinoic acid (POA), the bactericidal agent. X-Ray fluorescence Spectroscopy analysis shows that PZAse is a metalloenzyme that contains Zn⁺², and metal depletion of PZAse using EDTA decreases significantly its activity. Analysis of all PZA resistant strains reported in Peru and Worldwide shows that 90% of resistance are due to single point missense mutations on the PZAse sequence which are not homogeneously distributed along PZAse sequence, but clustered in the region near to the hypothetical PZAse Metal Coordination Site (MCS). Further analysis evaluating distinct amino acid physicochemical parameters and relative distances among the mutated residues in the PZAse 3D Homology Model shows that, those strains which have the highest PZA resistance (>1200 Minimum Inhibitory Concentration (MIC)), have mutations on residues which side chains aim at the MCS (p<0.02). These data indicate that PZAse coordinates, at least, a metal ion (Zn⁺²), which is required for its enzymatic function, and the presented PZA Resistance Prediction Model, which is based on the analysis of the PZAse MCS, could be used in Low-Income countries for rapid and low-cost PZA susceptibility tests.

Keywords: pyrazinamidase (PZA), *Mycobacterium tuberculosis*, homology modelling of proteins

P03.10.32

Acta Cryst. (2008). A64, C227-228

Homology modeling of *Arabidopsis thaliana* glycolipid transfer protein

Lenita Viitanen, Tiina A Salminen

Abo Akademi University, Biochemistry, Tykistokatu 6, Turku, Turku, 20520, Finland, E-mail: lenita.viitanen@abo.fi

Glycolipid transfer protein (GLTP) is a 24 kDa cytosolic protein that can transfer glycolipids between membranes *in vitro*, but its function *in vivo* is still unknown. GLTPs have been found in various organisms and recently also in plants. Three different forms of GLTP are expressed in *Arabidopsis thaliana*: AtGLTP_1, AtGLTP_2 and AtGLTP_3. Our collaborators have tested the lipid transfer preferences of the AtGLTPs with a transfer assay. Using the known crystal structures of human GLTP [1, 2] as templates, I have constructed homology models of the AtGLTPs in complex with glycolipid ligands. I have studied and compared the models and templates in order to find structural characteristics that explain the differences in lipid transfer preference of GLTPs. The project is done in collaboration with the research groups of Dr. Johan Edqvist (Linköping University) and Dr. Peter Mattjus (Åbo Akademi University).

[1] Malinina et al. (2004) Nature, 430, 1048-1053

[2] Malinina et al. (2006) PLoS Biol, 11, e362