MS02.06.04 AB INITIO PHASE DETERMINATION FOR VIRUSES; THE USE OF NON-CRYSTALLOGRAPHIC SYMMETRY FOR PHASE REFINEMENT. Chapman, M.†‡, Tsao, J.*‡, Rossmann, M.‡, Munshi, S.¶‡, Johnson, J.§‡, †Dept. Chem. & Inst. Molec. Biophys., Florida State Univ., Tallahassee, FL 32306; *Dept. Microbiol., Univ. Alabama, Birmingham, AL 35394; ‡Dept. Biol. Sci., Purdue Univ., W. Lafayette, IN 47907; ¶MRL, WP144B-122; Merck & Co. Inc., W. Point, PA 19486; §Dept. Molec. Biol., Scripps Res. Inst., La Jolla, CA 92037

The atomic structure of Canine Parvovirus (CPV) was determined with the use of phases calculated *ab initio* [Tsao *et al.*, 1992, *Acta Cryst.*. **B48**:75]. *Ab initio* phases were refined and extended to 8 Å and used to solve the positions of a heavy atoms for a single isomorphous replacement (SIR) structure solution. In retrospect, it was shown [Tsao *et al.*, 1992, *Acta Cryst.* **A48**:293] that the phases could be extended from 8 to 3.25 Å and that the virus structure could have been determined *ab initio* without SIR. Similar approaches have since been used to determine the structures of ϕX174 [McKenna, *et al.*, 1992, *Acta Cryst.* **B48**:499] and Nωv.

With the advantage of phase refinement and extension using non-crystallographic symmetry of high redundancy (60-fold for CPV), crude centric approximations can suffice for the initial starting phases at low resolution. Approximating the virus by a spherical shell of uniform densities for protein and nucleic acid, the sign of the molecular transform oscillates with the positions of nodes highly dependent on the choice of shell radii. *Post mortem* examination of successful and failed attempts show that it is critical to have a self consistent set of phases in the resolution shell from which extension will take place. This means that the radii must be estimated to within ~1% of their true values (or be similarly consistent with an alternative Babinet solution). Although trial-and-error may eventually give a suitable initial phasing model, it is also possible to refine the radii and position of the spherical shell model through optimization against low resolution amplitudes from crsytallography or solution scattering [Chapman et al., 1992, *Acta Cryst.* **A48**:301].

MS02.06.05 GENETIC ALGORITHMS AND MACROMOLECULAR PHASING M. Fujinaga & M. N. G. James MRC of Canada, Group in Protein Structure and Function Department of Biochemistry, University of Alberta Edmonton Alberta, Canada T6G 2H7

A program has been developed for exploring the use of genetic algorithms for macromolecular phasing. Genetic algorithms are powerful optimization techniques that borrow ideas from natural evolution. Unlike normal optimization techniques, it deals with a set (population) of possible solutions and these are improved by combining (mating) pairs of solutions from the population. The choice of the pair of solutions to mate is governed by the function value that one is trying to optimize (fitness function). The implementation of a genetic algorithm for the crystallographic problem has been done by representing a set of carbon atoms on a three-dimensional grid. The fitness function includes contributions from the agreement between observed and calculated structure factors as well as conformity to the expected distribution of atoms in a protein. One of the advantages of genetic algorithm is that it does not rely on gradients so that functions without derivatives can be included in the fitness function. The mating is done by a continuous crossover method that involves complete mixing and random separation of two sets of atoms. The number of atoms in common is monitored and the offspring are penalized according to the degree of inbreeding. The entire process is rather computationally expensive and the program has been parallelized to run on 20 IBM RS6000 workstations linked together using PVM (parallel virtual machine). The method is being tested on the complex of TEM-1 beta-lactamase and beta-lactamase inhibitory protein (BLIP). The enzyme part of the structure (62%) had been solved by molecular replacement but it was not possible the locate the inhibitor. Density modification techniques have failed to improve the phases sufficiently to solve the remaining 38% of the structure. The structure was eventually resolved by molecular replacement and refined so that a set of 'correct' phases exists to monitor the progress of the genetic algorithm.

This work is supported by the Medical Research Council of Canada.

MS02.06.06 PHASING OF PEPTIDE AND PROTEIN DATA SETS BY AB INITIO MODELLING. Piet Gros, Dept. of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 18, 3584 CH Utrecht, The Netherlands (gros@chem.ruu.nl)

A method is presented for ab initio phasing of protein and peptide data sets at medium resolution; examples of phasing peptide data sets up to low resolution (ca. 4.5 Ang) will be presented. The optimization method consists primarily of solvent flattening and structure refinement (using Molecular Dynamics and Energy Minimization). To make structure refinement feasible for random starting models a force field is introduced that is applicable to loose and all equal atoms. Pseudo-energy potential functions (potentials of mean force) in this force field are derived from radial-distribution functions of all non-hydrogen atoms as is observed in refined structures. Thus, the configuration of atoms is restrained towards atomic distributions of protein structures. Solvent flattening is applied to estimate the bulk solvent contribution to the structure factors. This procedure has been tested for a few peptide data sets. The data was generously supplied by Dr Isabella Karle. In the tests the data were truncated to 2.5 Ang resolution, because data to atomic resolution is rarely available when a protein structure is being determined. The resulting models from these optimization show a correspondence at low resolution to the known structures. Analysis of the phase differences and map correlation coefficients (w.r.t. the known answers) indicates that phase information up to approx. 4.5 Ang is obtained.

MS02.06.07 THE *AB INITIO* STRUCTURE AND REFINEMENT OF A SCORPION PROTEIN TOXIN. G.D. Smith[1,2], R.H. Blessing[1], S.E. Ealick[3], J.C. FontecillaCamps[4], H.A. Hauptman[1], D. Housset[4], D.A. Langs[1], R. Miller[1,5]. [1]Hauptman-Woodward Med. Res. Inst., 73 High St., Buffalo, NY 14203 USA; [2]Roswell Park Cancer Institute, Elm & Carlton St., Buffalo, NY 14263 USA; [3]Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853 USA; [4]Laboratoire de Cristallographie et Cristallisation des Proteines, Institut de Biologie Structurale Jean Pierre Ebel-CEA, CNRS, 41, Avenue des Martyrs, 38027 Grenoble Cedex 1 France; [5]Dept. of Computer Science, State Univ. of New York at Buffalo, Buffalo, NY, 14260 USA.

The structure of a 64 residue scorpion toxin, which crystallizes in space group $P2_12_12_1$, has been determined *ab initio* at 0.96Å using the *SnB* program. A total of 50,000 triples were generated from 5,000 phases and the value of R_{min} clearly revealed a single solution amongst the 1619 trials, using 255 *Shake-and-Bake* cycles per randomly generated trial structure. The number of residues in the protein and the amino acid sequence were deliberately withheld from the Buffalo group; the only known information was that the protein was composed of approximately 500 atoms and contained 4 disulfide bonds. A very conservative initial starting model consisted of five fragments varying in length from 5 to 18 residues for a total of 241 atoms. The entire structure was revealed following multiple cycles of Xplor refinement and Fourier

maps, and only four residues were incorrectly identified; three of these residues were subsequently found to be seriously disordered while the fourth suffered from high thermal motion. The structure has been refined to a residual of 0.160 for 512 protein atoms, 112 water molecules, and 418 protein hydrogen atoms. A mean phase error of 19.1° was calculated for the difference between the *SnB* phases and the final refined phases. It is estimated that given the sequence, 88% of the backbone atoms and 30% of the side chain atoms could have been extracted from the initial *SnB* phase set. Research supported by NIH grant GM-46733 and NSF grant IRI9412415.

MS02.06.08 THE AB INITIO SOLUTION OF THE HALORHODOPSIN AND OMP F PORIN MEMBRANE PROTEINS AT 6Å FROM ELECTRON DIFFRACTION PROJECTION DATA USING THE MAXIMUM ENTROPY-LIKELIHOOD METHOD IN THE MICE COMPUTER PROGRAM. C.J.Gilmore, W.N.Nicholson Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, Scotland U.K. and D.L.Dorset, Electron Diffraction Department, Hauptman-Woodward Institute, 73 High Street, Buffalo, New York 14203, U.S.A.

Using the combination of maximum entropy and likelihood (Bricogne (1984) Acta Cryst. A40, 410-445) in the MICE computer program (Gilmore, Bricogne & Bannister (1990) Acta Cryst. A46, 297-308), an ab initio phase determination was carried out at low resolution (6Å) for two dissimilar membrane proteins, the Omp F porin from the outer membrane of E. coli (which is largely beta-sheet) and halorhodopsin (which is largely alpha-helix). Surprisingly for a structure of this complexity and at this resolution, accurate phase information was found for the most likely solutions which enabled potential maps to be calculated that contained most of the essential structural details at 10Å resolution of these macromolecules in projection without the need for any image derived phases. The mean phase errors for the porin structure were less than 10°, whilst those from halorhodopsin were less than 20°. The calculations were remarkably easy using the MICE program more or less in default mode. A comparison with the use of the Sayre equation and phase annealing as an *ab initio* phasing procedure is made. (Dorset (1995). Proc. Natl. Acad. Sci. USA 92, 10074-10078, Dorset, Kopp, Fryer & Tivol (1995) Ultramicroscopy **57**, 59-89.) Both methods have their strengths and weaknesses which will be discussed.

Research funded in part by grants from the National Institute of General Medical Sciences (GM-46733) and the Human Frontier Science Program.

PS02.06.09 APPLICATION OF THE SAS TANGENT FORMULA TO MULTIPLE SITE PROBLEMS: A FEASIBILITY STUDY. Chun-Shi Chang, Charles M. Weeks, Debashis Ghosh, Herbert A. Hauptman, Hauptman-Woodward MRI, 73 High Street, Buffalo, New York 14203-1196, USA

The SAS tangent formula [Hauptman, *Acta Cryst.* A, in press] provides the basis for an *ab initio* multisolution (or multitrial) phasing procedure which utilizes invariant values estimated from SAS data [Hauptman (1982) *Acta Cryst.* **A38**, 632-641]. This procedure has been tested by application to 2Å single-wavelength, error-free data for a fabricated 10 Se analog of cucumber basic protein [Guss *et al.* (1988) *Science*, **241**, 806-811]. Se isotropic thermal parameters in the range 6.7-25.5Å^2 were based on the average side-chain values in the real structure, and Se anomalous scattering factors (f', f'') of (-7. 623, 5.083) at 0.9789Å were used.

Trial sets of random phases were refined iteratively until the corresponding values of the SAS maximal function converged, and solutions were identified on the basis of mean phase error. Nonsolutions consisting of part of the protein structure having the correct hand combined with enantiomorphic positions of the anomalous scatterers occurred frequently and had SAS maximal function values similar to solutions. In addition, under some circumstances trials yielding no locations for anomalous scatterers had even higher maximal function values. In either case, solutions could be detected automatically by picking the top 10 peaks on a Bijvoet difference Fourier and then computing protein phases based on these positions and comparing them to the corresponding original set of SAS tangent phases. Solutions had the smallest mean phase differences and, in this way, the proper absolute configuration could be selected. The SAS tangent solutions compared favorably to maps phased by the 10 Se placed at their true positions and given precisely correct thermal parameters. This research was supported by GM-46733 (NIH).

PS02.06.10 LOW RESOLUTION AB INITIO ENVELOPES OF YEAST RNA POLYMERASE II. David, P. R., Bushnell, D. A., Leuther, K. K., Subbiah, S., Bellamy, H. and Kornberg, R. D, Structural Biology Dept. Stanford University Medical Center MS 5400 Stanford, CA, 94305-5400 USA

Low resolution ab initio envelopes for yeast RNA polymerase II, an enzyme complex of 550,000MW, have been computed, using the condensing protocol methods. These envelopes show the packing within the unit cell and a strong similarity to the existing envelopes from two dimensional electron microscopy. The special methods used to obtain the lowest resolution reflections for such a large protein and assymmetric unit will be discussed. Despite the large size of the protein, envelopes can be calculated easily in a matter of minutes.

PS02.06.11 DIRECT-METHOD STRUCTURE DETERMINATION OF THE NATIVE AZURIN II PROTEIN USING ONE-WAVELENGTH ANOMALOUS SCATTERING DATA. Q. Hao, Department of Applied Physics, De Montfort University, Leicester LE1 9BH, England; Zheng Xiao-feng & Fan Hai-fu, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China; F. E. Dodd & S. S. Hasnain, CCLRC Daresbury Laboratory, Warrington WA4 4AD, England.

The one-wavelength anomalous scattering (OAS) X-ray diffraction data of azurin II, a copper-containing protein from *Alcaligenes xylosoxidans* were collected at the Photon Factory at a "routine" wavelength of 0.97Å. The structure had been originally solved by the molecular replacement method (Dodd, Hasnain, Abraham, Eady & Smith (1995) *Acta Cryst.*, **D51**, 1052-1064). As a technique of *ab initio* structure determination, the direct method (Fan, Hao, Gu, Qian & Zheng (1990). *Acta Cryst.* **A46**, 935-939) was attempted to break the phase ambiguity intrinsic to one-wavelength anomalous scattering data. The phases were then improved using the solvent-flattening method. The final electron-density map clearly shows most Cα positions and many side chains and it is traceable without prior knowledge of the structure. It is concluded that the direct method is capable of phasing anomalous scattering data collected at one wavelength from moderate sized native proteins (Mw ~ 20kDa) which contain copper or atoms with a similar scattering power.