



A Direct Approach to MAD Phasing

Ashley Deacon - *Stanford University*

About the author:

Ashley Deacon came to Cornell after completing his Ph.D. in the laboratory of John Helliwell at the University of Manchester, England. He has recently moved to become a staff scientist at the Stanford Synchrotron Radiation Laboratory. His recent work is cited in a news section article of *Science* magazine entitled “New Math Speeds the Search for Protein Structures” by D. Kestenbaum (vol. 282, 2 October 1998, pages 30-31).

A direct approach to MAD phasing

Introduction

During a standard protein crystallography experiment the intensities of the diffracted X-ray beams are recorded. Unfortunately, their relative phases, crucial for reconstructing an image of the molecule, are lost and have to be determined indirectly, either by making additional measurements or by exploiting some prior structural knowledge. Multi-wavelength anomalous dispersion (MAD) is a method of macromolecular phase determination that exploits the tunable energy capability of a synchrotron X-ray source [1]. By using a highly monochromatized X-ray beam the experimenter is able to probe the anomalous scattering from some heavy atoms, incorporated into the target molecule. The location of the atoms in this, often simple, anomalous scattering sub-structure can then be used as a stepping stone in solving the complete molecular structure. The potential of this technique is now being recognized by many structural biologists and it is fast becoming the method of choice for *de novo* macromolecular structure determination efforts. It can relieve the experimenter of the time-consuming search for viable heavy atom derivatives and can lead very quickly from an initial crystal, via a single multi-wavelength diffraction experiment, to a readily interpretable electron density map of the molecule under investigation. The use of selenomethionine substitution as a method of readily incorporating anomalous scatterers has allowed the technique to become broadly applicable to a large variety of macromolecular systems [2]. However, more complex sub-structures, consisting of multiple selenium atoms, are required to provide adequate signal for the study of large macromolecules. As a result, the determination of the anomalous scattering sub-structure itself can become a significant hurdle in the MAD phasing process [3].

MAD developments at CHESS

The growth in popularity of the MAD technique over the last few years has spurred a complete redesign of the CHESS F-2 station and the installation of a new oscillation

camera [4]. It is now operated as a dedicated MAD phasing facility. The structure determination of the interferon- receptor complex, with just six selenium atoms in 120kDa of protein, clearly showed the potential of this facility and also demonstrated that weak signals can be effectively used to solve large structures [5]. Recently, the addition of a large area Quantum-4 CCD detector has further enhanced the capability of the station, allowing it to handle larger structures in a more efficient manner.

This new station configuration has given me the opportunity to address another important question relating to the scope and applicability of the MAD phasing technique. Namely, how viable is MAD phasing for solving very large structures and how are current methods able to deal with locating a large number of anomalous scatterers?

A challenging MAD problem

The project I have been focusing on arose from a collaboration with Dr. W. Coleman Jr. from the NIDDK laboratory of the National Institutes of Health. Dr. Coleman has been studying the enzyme ADP-L-glycero-D-mannoheptose 6-epimerase for a number of years. He recognized that his efforts, aimed at targeting the enzyme for drug design, would be greatly enhanced if he could gain a glimpse of the molecular structure and some insight into its catalytic mechanism. The enzyme is involved in the biosynthesis of lipopolysaccharide, a key component in the outer cell membrane of gram-negative bacteria. These bacteria are less viable and more susceptible to antibiotic treatments if this biosynthetic pathway is disrupted.

I was keen to try and use MAD phasing to solve the structure in the shortest possible time. After obtaining some initial crystals from the native enzyme I was able to collect a dataset on the CHESS F-1 station and soon the magnitude of the structure determination task became apparent. The dataset revealed that the enzyme was pentameric and additionally it showed that two crystallographically independent pentamers, amounting to 370kDa of protein (more than 25,000 atoms),

were located in the asymmetric unit (Figure 1). Since each individual molecule of the enzyme contained seven methionine residues (one of the 20 amino acid residues, that are the fundamental building blocks of proteins), this meant that a total of 70 selenium atoms would have to be located if I continued to pursue the structure determination by MAD phasing. I also had some additional concerns because the crystals formed in a low symmetry (monoclinic) space group and they diffracted anisotropically, with data extending to around 2.0Å on images collected down the B axis, but only to 2.8Å down the A and C axes.

However, at the same time, I was encouraged by the results of Turner, Howell and co-workers, who had successfully located and used 30 selenium atoms in the structure determination of AdoHcy hydrolase [6].

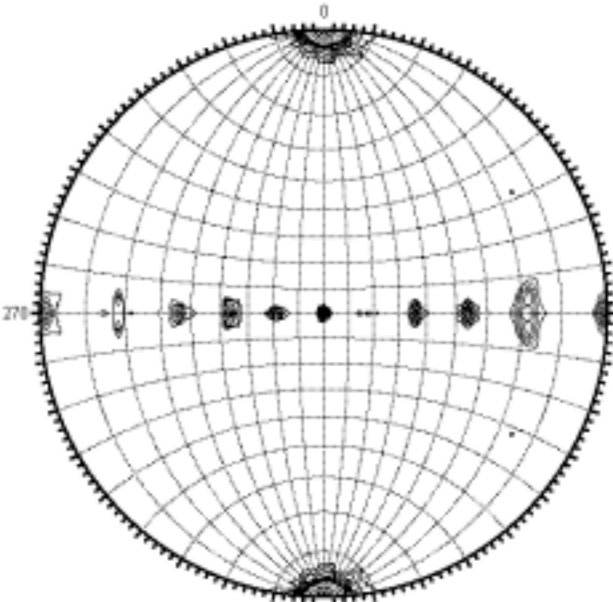
A night on F-2

Shortly after I had collected and analyzed the data from the native protein, Dr. Coleman sent me samples of the freshly prepared selenomethionine enzyme. With some minor tweaking of the native crystallization conditions I soon obtained some small, but sharp-faced crystals. In what seemed like a very short space of time I was ready to conduct a MAD experiment. Unfortunately, the fast progress in the wet lab left me facing an already packed F-2 schedule, so I was forced to freeze the crystals and hold out for a future data collection

Figure 1: 180 degree self-rotation search indicating, two pentamers aligned along the B axis, as evidenced by the two-fold peaks falling every 18 degrees around the equatorial plane.

I had a crystal in place and an initial image from which I could calculate a data collection strategy. A quick data processing run identified the best starting orientation and minimum sweep of images needed for a complete dataset. Despite the fact that the crystal diffracted further, I positioned the detector to collect a 3.0Å resolution MAD dataset, which allowed me to keep the exposure times down to 25 seconds. I now had 15 hours of data collection ahead of me.

Optimistically, I had ten fills of the CHESS machine available, and if all went smoothly, a single 140° sweep would require two fills to yield a dataset of decent completeness (>95%). The all important peak anomalous wavelength (f'' maximum) was collected first, with an inverse beam sweep (ie. a sweep of data collected with the crystal rotated by 180° with respect to the initial sweep) to guarantee the measurement of all the Friedel pairs. Four fills passed and things remained on track. Next came the remote wavelength; here I decided to sacrifice some potential phasing information by collecting data at the low energy side of the edge. The critical advantage was that an inverse beam sweep was not necessary, because of the lack of anomalous signal below the Se K absorption edge. Two fills later the remote wavelength was making its way onto the backup tape. However, because I had underestimated the time between CHESS fills, time was now running short. The third wavelength was set at the Se K-edge inflection point (f'' minimum). Some further fine-tuning of the data collection strategy allowed a slightly truncated third wavelength to be collected with an inverse beam sweep during the next three fills. The last images streamed off the CCD detector just in



surprised one an immediate er group had 1 augured well ion. However, ext users were thorough and w dimension. 70 selenium MAD phasing, ery tight time I had to wake e French roast HESS. As the mulate a plan.

It was immediately apparent that I would have to cut a few corners in order to complete the task within the allocated time period. I began by eliminating the possibility of following the commonly used MAD protocol, where the complete dataset is assembled by collecting consecutive small wedges at each of the chosen wavelengths. It was clear that I did not have the additional time required for such repetitive wavelength changes. Instead I chose to collect each wavelength individually from start to finish. Within an hour from the initial phone call

the nick of time, before the end of the run. Despite my excitement at completing the MAD dataset it was time for sleep, before continuing with the next steps in the structure determination process.

Searching for seventy selenium sites

As the first step in the data analysis I decided to take a look at an anomalous difference Patterson map for the peak wavelength. I obviously had no hope of trying to interpret it manually. However, I wanted to quickly convince myself that there was some significant signal in the data. Even though the Harker section ($B=1/2$) seemed hopelessly crowded, it did at least show some clear peaks above the background noise (Figure 2).

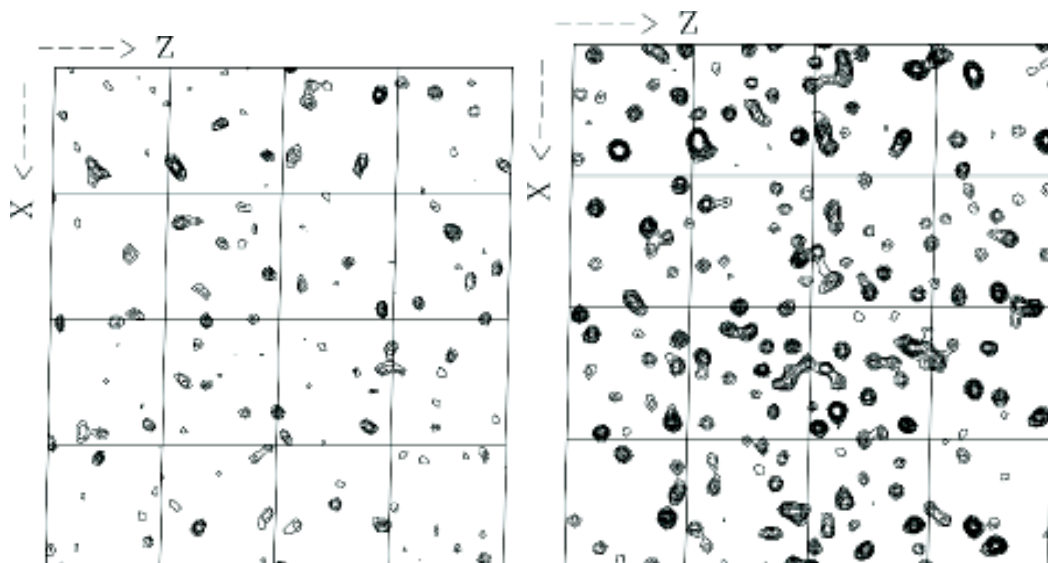


Figure 2: Patterson maps ($y=0.5$ Harker section) for the 70 selenium atom sub-structure of ADP-L-glycero-D-mannoheptose 6-epimerase. a) Experimental anomalous difference data. b) Theoretical Patterson from the known selenium coordinates. Reproduced with permission from Figure 5 of *J. Synchrotron Rad.* **6**, 822-833 (1999)

My structure determination efforts now had to become a bit more serious. For the selenium sub-structure determination, I decided to turn to the direct methods program SnB v2.0 [7], which had recently been released. The program is based around the dual space Shake-and-Bake algorithm, which alternates reciprocal space phase refinement with peak picking in real space. My previous experience with SnB in the *ab initio* solution of triclinic lysozyme gave me confidence in using the program [8]. In addition, it had recently been used to solve a few smaller selenium sub-structures from peak wavelength anomalous difference data.

SnB is known as a multi-trial direct methods program. In essence, many attempts at solving the structure are carried out. Each attempt (or trial) starts from a random set of atoms, which are then subjected to a number of cycles of phase refinement. A criterion, known as the minimal function, is then used to assess each trial and identify possible solutions. The algorithm is computer intensive, because of the Fourier transforms that are required to continually switch between real and reciprocal space. When I set my first job running I soon realized the daunting consequences of the $100\text{\AA} \times 110\text{\AA} \times 180\text{\AA}$ cell dimensions, as each SnB trial required 20 minutes to run on the fastest computer I had available. Additionally, I had no idea how many trials would be needed to find a promising solution. Fortunately, the SnB algorithm is

inherently parallel. It is perfectly feasible to set up jobs on several computers, since each job can be started from a different random number seed. The results from all the trials can then be assembled together and analyzed for potential solutions. I wasted no time in setting up jobs on all the computer platforms I had access to. I placed my jobs at the lowest priority level so that they wouldn't disturb other computer users, while hopefully running at full speed during the nighttime hours. As a result, after less than 24 hours, there appeared to be a promising SnB solution – a single trial had converged to a significantly lower minimal function, compared with all the other trials that had been processed (Figure 3). I was soon able to confirm that the sites were correct by locating the local five-fold symmetry from the output selenium positions. Shortly thereafter I had a second independent SnB solution and I could compare the two individual solutions for common sites. Over the next few days, I proceeded with the phasing calculations and eventually arrived at an interpretable electron density map, into which I could build the ADP-L-glycero-D-mannoheptose 6-epimerase model (Figure 4). The efficient way in which the SnB program managed to locate this large number of selenium sites, using anisotropic data of only modest resolution, should encourage others to pursue even larger structures by MAD phasing and push the limits of the F-2 station even further.

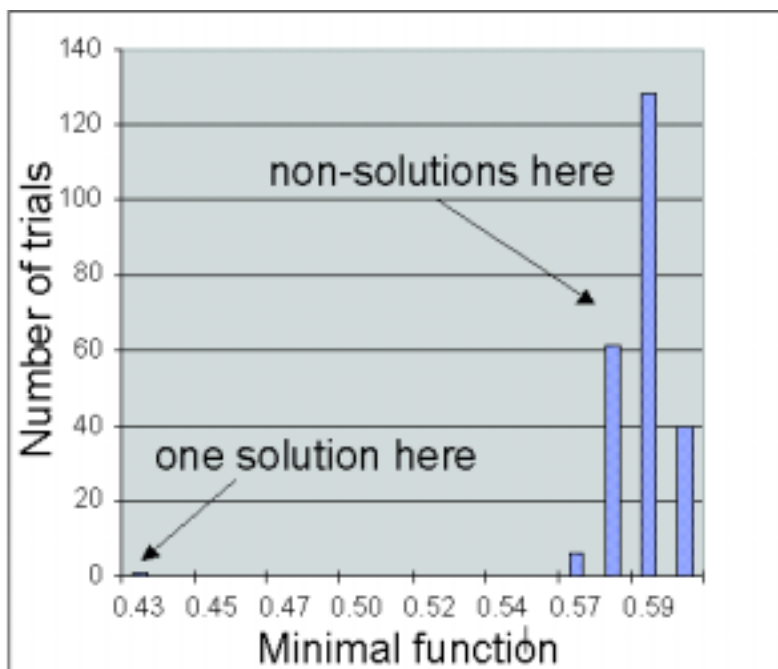


Figure 3: Histogram of minimal function values for 236 SnB trials. The correct solution is evidenced by the bi-modal distribution with a single trial at a significantly lower minimal function value.

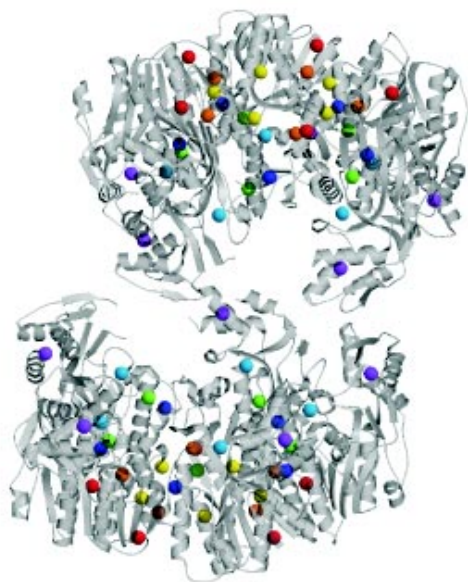


Figure 4: The structure of ADP-L-glycero-D-mannoheptose 6-epimerase in ribbon representation with the 70 selenium atoms superimposed in colour. Equivalent selenium atoms from molecule to molecule are coloured the same.

Acknowledgments

This work was conducted while I was a postdoctoral research associate in the laboratory of Steven Ealick at Cornell University. I am grateful to Steve for his support and also to the other members of his lab for their advice and encouragement. I would also like to acknowledge our collaborators, William Coleman Jr. and Yisheng Ni at NIDDK, Bethesda, Maryland.

References

- [1] Hendrickson, W.A. Determination of Macromolecular Structures from Anomalous Diffraction of Synchrotron Radiation. *Science* **254**, 51-58 (1991)
- [2] Doublé, S. Preparation of Selenomethionyl Proteins for Phase Determination. *Methods Enzymol* **276**, 523-530 (1997)
- [3] Deacon, A.M. & Ealick, S. E. Se-based MAD Phasing: Setting the Sites on Larger Structures. *Structure*, R161-R166 (1999)
- [4] Thiel, D. J., Blank, B. and LaIuppa, J. A New Macromolecular Oscillation Camera at CHESS. *J. Synchrotron Rad.* **5**, 914-916 (1998)
- [5] Thiel, D.J., et al. Observation of an Unexpected Third Receptor Molecule in the Crystal Structure of Human Interferon-gamma Receptor Complex *Structure* **8**, 927-936 (2000)
- [6] Turner, M.A., Yuan C.S., Borchardt, R.T., Hershfield, M.S., Smith, G.D. and Howell, P.L. Structure Determination of Selenomethionyl S-Adenosylhomocysteine Hydrolase using Data at a Single Wavelength. *Nat. Struct. Biol.* **5**, 369-376 (1998)
- [7] Weeks, C. M. and Miller, R. The Design and Implementation of SnB Version 2.0. *J. Appl. Cryst.* **32**, 120-124 (1999)
- [8] Deacon A. M., Weeks C. M., Miller R. and Ealick S. E. The Shake and Bake Structure Determination of Triclinic Lysozyme. *Proc. Natl. Acad. Sci.*, **95**, 9284-9289 (1998)