From a diffraction image to the electron density map: A case study

Ana González
SSRL

July 8, 2002

Abstract

This lecture will illustrate how to carry out a crystal diffraction data collection using the example of a real Multiwavelength Anomalous Dispersion (MAD) experiment. The main emphasis will be on the tools used and the criteria followed to make decisions about the optimal strategy for the particular example.

1 Introduction

Although in theory it is possible to distinguish clearly between what is good and bad practice in macromolecular crystallography data collection and give general detailed guidelines for collecting good data, in the practice we have to deal with limitations imposed both by the sample (sometimes it is hard or impossible to make good crystals) and by the experiment set-up (properties of the X-ray source, detector, limited time, etc). In addition, the data collection conditions have to be consistent with the aims of the experiment. These three factors (experiment purpose, sample characteristics and x-ray source or beamline) are interconnected which means that there is often no single answer to the question about how to collect data to maximize the chances of success. Experience, and detailed knowledge of the sample are very useful. Continuous monitoring of the data collection and data analysis on the fly are required in order to detect and correct problems with the initial set-up.

2 MAD data collection on a selenomethionine protein

This experiment is part of the Joint Center for Structural Genomics (JSCG) project. The aim of the data collection is the structure solution of the product of a gene of the bacterium *Thermotoga maritima*. A native data set had already been carried out and the Se-methionine substituted protein was crystallized to be able to solve the structure by MAD. The maximum resolution of the native data is better than 1.45 Å and the space group I422. The molecule has 364 residues in the asymetric unit, 7 of which are methionines. This results in a higher than

average ratio of Se atoms to total number of atoms, which results in a high anomalous signal and less sensitiveness of the data to errors.

2.1 fluorescence scan and wavelength selection

The first step in a MAD experiment (also useful whenever there in an anomalous scatterer present in a phasing experiment) is the measurement of a fluorescence scan. This is required, because both the position of the absorption edge and the anomalous scattering factor values are affected by the chemical environment of the anomalous scatterer.

A good scan should extend about 200eV around the absorption edge energy. This allows easier estimation of the anomalous scattering factors f' and f" for the near edge wavelengths with as program like Chooch (1)[1]. These values, unlike the theoretical ones, most often do not require refinement during the phasing. This makes it easier to refine other correlated parameters (such as the anomalous scatterer occupancy and temperature factor). ON the other hand, for wavelengths away from the absorption edge there is not much difference between the theoretical and experimental values of the anomalous scattering factors and the former ones can be used without fitting to a scan.

MAD experiments usually consist in data collection at three wavelengths. The data at the "peak" or maximum f" wavelength has the largest anomalous differences, the data at the inflection and remote provide the largest dispersive differences. It is also possible to do a two-wavelength MAD experiment (2) however there is not much extra advantage in collecting four wavelengths, with extra data redundancy achieving a similar improvement of the experimental phases.

For the Se experiment we select the remote wavelength at 13500 eV (wavelength 0.91A). At this wavelength the value of f' is sufficiently small to obtain large dispersive differences, while f" is large enough to allow the measurement of a significant anomalous signal.
The recommended order of data collection in the general case is 1) largest f" wavelength (peak) 2) remote and 3) inflection. The reasoning for collecting the peak wavelength first is that the high anomalous signal at this wavelength makes it easier to find the position of the Se atoms. Finding the heavy atoms is a valuable predictor of the sucess of the experiment. However, In the prresent example case, because of the large Se/MW ratio, we expect to measure a strong anomalous signal at all the wavelengths, so we can leave the peak wavelength last, avoiding the increased absorption which may be associated to faster radiation damage.

2.2 Test diffraction images

The next step is to estimate the quality of the crystal and determine from one or two diffraction images 90 degrees apart the exposure time, oscillation angle (delta phi) and range:

[1] http://babinet.globalphasing.com/people/gwyndaf/Chooch.html

1. Select an adequate collimator size and beamstop to sample distance.

2. Collect a diffraction image and inspect it for oddly shaped diffraction spots, overlapping reflections, overloads, etc. An image 90 degrees apart should also be collected before starting the data collection, to ensure that the crystal is good in all directions.

3. Autoindex the image. If the space group and cell are not previously known, select the lowest symmetry consistent with the indexing. The choice of the lattice can be confirmed by scaling the data.

4. Run a strategy program (like the strategy option on mosflm or rotgen[2] to determine the data collection angular range and delta phi (3,4) In the case of MAD, or whenever a heavy atom is present in the sample, it is useful to collect a total phi range such that Friedel related reflections are collected. If not, the Friedel pairs must be measured by collecting the same phi range with the crystal rotated 180 degrees (inverse beam geometry). Often it is possible to collect a complete Friedel related pairs set with far fewer data than required by the inverse geometry, shortening the experiment without a significant reduction in the quality of the phases

5. Integrate the image and examine the I/sigma of the reflections at the highest resolution shell. While a I/sigma over 2 can always be worth collecting, it is unlikely to measure significant anomalous of dispersive differences at that resolution and the experimental phase figures of merit will be poor for the higher resolution data. Try increasing the exposure time to obtain a higher I/sigma (without overloads!). If high resolution phases are needed but there are overloads at low resolution, you need to collect the data in two passes [3]. For MAD phasing, collect the low resolution pass first. The high resolution data need only be collected at one wavelength.

6. Check the exposure time at the remote wavelength: the diffraction intensity, detector sensitivity and beamline intensity are different at different wavelengths. If the crystal has an irregular shape, check different crystal orientations: as different volumes of the crystal get illuminated, the diffraction intensity will be different.

7. Select a detector to sample distance so that the maximum useful resolution data are at the edge of the detector. This separates the diffraction spots and improves I/sigma. Note that the optimal detector distance for the remote wavelength will be further away from the sample than for the edge wavelengths (The distance was 175mm for the near edge wavelengths and 189mm for the remote)

2.3 Monitoring the experiment

The data should be analyzed as the data collection is proceeding. The data scaling statistics and completeness of the data need to be monitored. If the preliminary phasing does not give

[2]http://www.ccp4.ac.uk/dist/x-windows/Rotgen/doc/rg_top_source.html
[3]http://smb.slac.stanford.edu/public/datacollect/ultrah.html

good results or the maps look uninterpretable, extra redundant data collected over a wider phi range may help. Here is a non exhaustive list of indicators for the quality of the data:

1. Spot refinement, integration: Good profiles, stable refinement, accurate prediction, few rejections per image.

2. Scaling, merging: Predicted completeness, few outliers, reasonable I/sigma and r-factors, relative anomalous signal consistent with f" at different wavelengths, good agreement between wavelengths

3. Phasing: Solved heavy atom substructure, good phasing statistics, interpretable maps.

3 References

1. G. Evans and R. F. Pettifer (2001) J. Appl. Cryst. **34**, 82-86.

2. A. González, J.-D. Pédelacq, M. Sola, F.X. Gomis-Rüth, M. Coll, J.-P. Samama and S. Benini (1999). *Acta Cryst.*, D**55**, 1449.

3. Z. Dauter (1997) in Methods in Enzymology, vol 276 part A, eds. C.W. Carter and R. M. Sweet, Academic Press.

4. J. W. Pflugrath (1999) *Acta Cryst.*, D**55**, 1718.