# Protein crystallography with spallation neutrons

**Benno P. Schoenborn* and Paul Langan**

*Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. E-mail: schoenborn@lanl.gov*

Spallation neutrons are ideal for diffraction studies of proteins and oriented molecular complexes. With spallation neutrons and their time dependent wavelength structure, one can select data with an optimal wavelength band and cover the whole Laue spectrum as time (wavelength) resolved diffraction data. This optimises data quality with best peak to background ratios and provides spatial and energy resolution to eliminate peak overlaps. Such a Protein Crystallography Station (PCS) has been built and tested at Los Alamos Neutron Science Centre. A partially coupled moderator is used to increase flux and data are collected by a cylindrical $He^3$ detector covering $120^o$ with 200mm height. The PCS is described along with some examples of data collected from proteins.

**Keywords: neutron diffraction; enzyme mechanism; protein crystallography station; PCS.**

### 1. Introduction

The PCS was built by Los Alamos National Laboratory's Bioscience Division at the Manual Lujan Neutron Scattering Center, run by Los Alamos Neutron Science Center (LANSCE). Funded by the U.S. Department of Energy Office of Biological and Environmental Research, this station is the first PCS in the world to be built at a spallation neutron source, the first to use Time-of-Flight (TOF) methods, and the only neutron protein crystallography resource in North America. TOF methods have however been applied to small molecule crystallography for many years at a number of spallation sources (SXD at ISIS, SCD at LANSCE, SCD at IPNS and TriCS at PSI). The PCS has been run as a user facility for the structural Biology community since August 2002. It provides an important proof-of-principle for future stations at advanced spallation sources being planned and built worldwide.

Neutron diffraction is a powerful technique for locating hydrogen atoms and water molecules. Hydrogen atom positions and the coordination of water molecules cannot be directly determined using X-ray diffraction at resolutions typical for most protein crystals. Small degrees of mobility of hydrogen atoms can kill their x-ray diffraction signal but not their (deuterium) neutron scattering signal (Habash *et al*, 2000). The majority of user experiments carried out on the PCS have been aimed at locating functionally important hydrogen atoms in order to determine detailed enzyme mechanism. However experiments aimed at investigating hydrogen bonding and hydration have also been carried out. In this report we give an overview of the PCS and then describe some of the preliminary data collected during the first year of operation as a user facility.

### 2. The Protein Crystallography Station

The PCS is located on Flight Path (FP) 15. Neutrons are produced by protons striking a W target in microsecond pulses with a frequency of 20Hz. These high-energy neutrons are moderated in a reflector and then reduced to thermal energies in a moderator specifically tailored for protein crystallography (Schoenborn *et al*, 1999). The moderated neutrons are then extracted down a beam-pipe. A schematic representation of the beam layout on FP15 is shown in Figure 1. From the moderator neutrons travel a total flight path length of 28m down vacuum pipes with collimation inserts that taper the neutrons to produce a fine, almost parallel beam with a divergence of $0.12^o$ that matches the mosaic of a typical protein crystal.

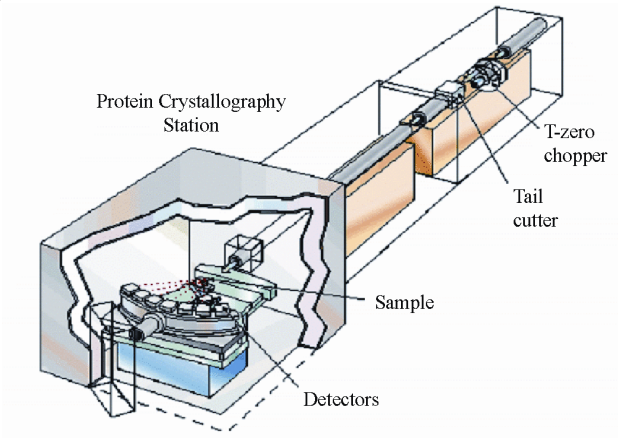


**Figure 1**
A schematic layout of the PCS.

At 9.5m from the moderator the vacuum pipe is interrupted by two beam shaping devices; a composite T0/T1 chopper and a proposed tail-cutting device (Langan *et al*, 2001). These devices remove unwanted high- and low-energy neutrons in each pulse, thus optimising the neutron beam for high counting rates and low backgrounds at reasonable instrument resolutions. The vacuum pipe is tightly surrounded by heavy shielding until it reaches the sample position, where the shielding opens up to a large cave.

In the cave, a kappa-circle goniometer made by Huber moves the crystal through a number of orientations, recording a Laue pattern at each setting. A complete data set typically requires between 6-30 crystal settings depending on the symmetry of the crystal. An optional 8T super-conducting magnet can be mounted on the goniometer to orient membrane and fibre samples. A 700 series nitrogen cryostream cooler made by Oxford Cryosystems can be used to control the temperature of the sample down to ~100K. A large cylindrical PSD built by the Instrumentation Division of Brookhaven National Laboratory (BNL) collects as many of the spots as possible at each crystal setting without having to reposition the detector.

The positions of the detected neutrons are calculated by advanced electronics associated with the detector and designed by BNL, before being passed to a VXI-based data-acquisition system designed by LANSCE where they are stored in memory along with their TOF. After a sufficient number of neutrons have been collected at a particular crystal and detector setting the data and the parameters that describe the instrument setting are written to a data archive in the form of a NeXus file.

The whole data collection process is coordinated from a Personal Computer (PC) with a Windows system (Microsoft) that runs a Java-based graphical user interface (GUI). A real time display of the data being collected on the detector is achieved on another PC that runs a LabVIEW (National Instruments Corporation) GUI to the detector electronics. On a third PC, a customized version of the commercial software package d*TREK (Molecular Structure Corporation), designated PCS-d*TREK is used to display and process the archived NeXus files and ultimately to produce a measured intensity and an index for each diffraction spot. The intensities are then wavelength normalized using LAUENORM (Helliwell *et al*, 1989). The intensities and indices provide the input required by standard protein-structure-refinement software and graphics programs that ultimately produce an image of the protein crystal's atomic structure.

  

### 3. Preliminary results

#### 3.1. Porcine insulin

The structure of porcine insulin was recently determined by neutron diffraction using the BIX-3 protein crystallography station at the Japan Atomic Energy Research Institute (JAERI) (Maeda *et al*, 2003). Porcine insulin can crystallize in cubic space group, $I2_13$, which has 12 symmetry equivalent reflections. It is therefore an ideal system with which to test the corrections applied to symmetry equivalent reflections recorded at different wavelengths on the PCS. An experimental team from JAERI and Bioscience Division collected data for 16 hours at a single stationary setting as summarized in Table 1.

**Table 1**

Data collection parameters for porcine insulin.

| | |
|---|---|
| Space Group | $I2_13$ |
| Unit cell dimensions (A)[a] | a = 78.9 |
| Wavelength range (A) | 0.6 - 5 |
| Station | PCS |
| Temperature | 295 |
| Crystal settings | 1 |
| Exposure time | 16 hours |
| Observed reflections | 4222 |
| Observed reflections d>1.8A | 3742 |
| Unique reflections d>1.8A | 2581 |

[a] Unit cell parameters are from X-ray diffraction.

The data were analysed using d*TREK, Figure 2. The value of $R_{merge}$, when calculated as a function of I/σ, rises rapidly for values of I/σ <2, suggesting that below this threshold the intensities are not accurately measured. When $R_{merge}$ is tabulated as a function of resolution, after applying a I/σ>1.5 threshold, Table 2, significant data can be found out to ~1.8A resolution. The completeness of the data is 27.4% at 2.1A resolution (41.7% overall). These results indicate that in order to collect a data set to 2.1A resolution, at least 5-6 crystal settings equivalent to 3 days of beam time would be required. The fact that a significant number of reflections are found out to resolutions of around 1.8A indicates that by extending the collection time at each crystal setting, a higher resolution data set could be obtained over a longer period of time.

**Table 2**

$R_{merge}$ (I>1.5σ), completeness (%Comp), and <I/σ> verses resolution for porcine insulin. The cumulative $R_{merge}$ and completeness are also given ($CumR_{merge}$, Cum%Comp). The completeness for all reflections without an I/σ threshold, %AllComp is given. The $\sigma_a$ corresponds to the standard deviation in merging reflections as opposed to σ, the error involved in measuring individual intensities.

| Resolution | Num. Refs. | %Comp | Cum% Comp | $R_{merge}$ | $CumR_{merge}$ | $<I/\sigma_a>$ |
|---|---|---|---|---|---|---|
| ∞ - 5.69 | 241 | 62.7 | 62.7 | 0.161 | 0.161 | 13.7 |
| - 4.02 | 411 | 61.5 | 61.9 | 0.124 | 0.142 | 6.4 |
| - 3.29 | 451 | 52.2 | 57.5 | 0.138 | 0.141 | 4.7 |
| - 2.85 | 467 | 45.5 | 53.3 | 0.231 | 0.153 | 2.6 |
| - 2.55 | 445 | 39.2 | 49.3 | 0.259 | 0.159 | 2.0 |
| - 2.32 | 410 | 33.5 | 45.5 | 0.185 | 0.160 | 1.6 |
| - 2.15 | 381 | 27.4 | 41.7 | 0.260 | 0.161 | 1.3 |
| - 2.01 | 331 | 23.6 | 38.5 | 0.154 | 0.161 | 1.1 |
| - 1.90 | 326 | 21.2 | 35.7 | 0.259 | 0.162 | 1.0 |
| - 1.80 | 279 | 17.3 | 33.0 | 0.222 | 0.162 | 0.9 |

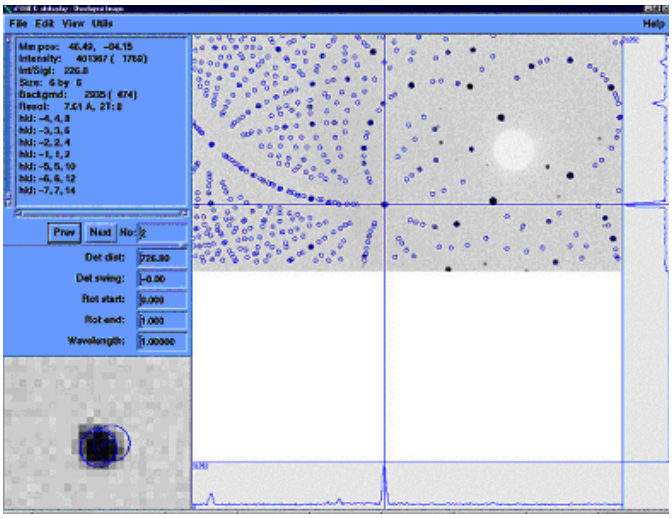


**Figure 2**

Analysis of data collected from insulin using PCS-d*TREK. The main display area shows the data as a Laue projection. The zoom display area in the bottom-left corner shows a cursor-selected reflection with superimposed circles representing starting and refined predicted positions. Horizontal and vertical line sections through the cursor position are shown along the borders of the main display area. The top-left display area contains statistical details about the cursor-selected reflection (Courtesy of Ichiro Tanaka and Nobuo Niimura, JAERI).

#### 3.2. D-Xylose isomerase

The protein D-xylose isomerase is an enzyme that catalyzes the conversion of D-xylose to D-xylulose and glucose to fructose by hydrogen atom transfer. This protein plays an important economic role in the food industry and is also the subject of intense research in other areas. One possible mechanism for the enzyme is metal ion-mediated ionisation of water, the metal being magnesium. The catalytic motif of a magnesium cation, water molecule and enzyme carboxylate group, is common to many magnesium containing enzymes that are important in cancer research. By determining the exact catalytic mechanism important insights will be gained into a whole class of enzymes, with a view to designing better anticancer drugs.

An experimental team from Oak Ridge National Laboratory (ORNL), Fox Chase Cancer Centre, and Bioscience Division collected data at 23 crystal settings for times ranging from 6 hours to 48 hours at each setting, as time became available, from a crystal of about 5-6$mm^3$ in volume (Hanson *et al*, 2003). The average collection time at each setting was ~20 hours, as shown in Table 3. An example of the data collected at one crystal setting is shown in Figure 3. From Table 4, which shows the values of $R_{merge}$, completeness and <I/σ> at a threshold of I/σ >1.5 as a function of resolution, the data set has a resolution of ~2.1A-2.0A. Significant intensities were observed out to 1.5A resolution and suggest that a higher resolution data set could be obtained by recording for longer.

From a preliminary 7A-2.1A neutron (2Fo-Fc) Sigma-A density map, shown in Figure 4, it is already possible to see that deuterium atoms have replaced labile hydrogen atoms. A detailed analysis of this data is underway at ORNL.

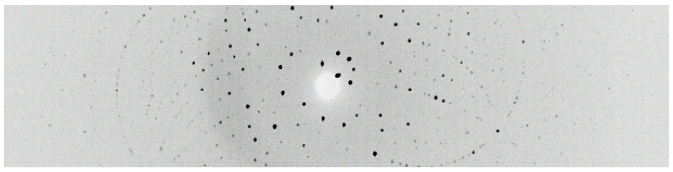

**Figure 3**

A Laue projection of data collected at one crystal setting from D-xylose isomerase. A total of 181,797 reflections were recorded, over 23 crystal settings (Courtesy of Gerry Bunick and Leif Hanson, ORNL).

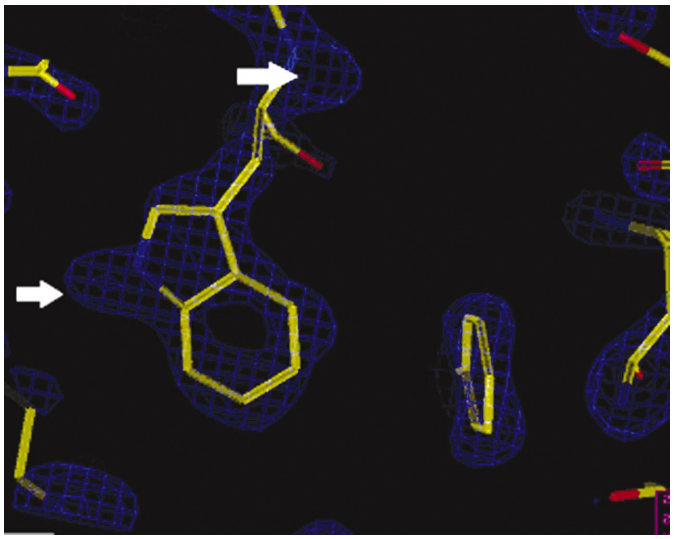

**Figure 4**

A slice of neutron scattering length density covering a tryptophan residue in D-xylose isomerase in a preliminary neutron (2Fo-Fc) Sigma-A map using data from 7A-2.1A resolution (Courtesy of Gerry Bunick and Leif Hanson, ORNL). Labile H atoms that have been replaced by D atoms are indicated by arrows.

**Table 3**

Data collection parameters for D-xylose isomerase.

| | |
|---|---|
| Space Groups | I222 |
| Unit cell dimensions (A)[a] | a=93.78, b=99.53, c=102.90 |
| Wavelength range (A) | 0.6 - 5 |
| Station | PCS |
| Temperature | 295 |
| Crystal settings | 23 |
| Average time per setting | 20 hours |
| Observed reflections | 181,797 |
| Observed reflections d>1.8A | 146,996 |
| Unique reflections d>1.8A | 38,029 |

[a] Unit cell parameters are from X-ray diffraction.

**Table 4**

$R_{merge}$, completeness (%Comp), and <I/σ> verses resolution for D-xylose isomerase. The cumulative $R_{merge}$ and completeness are also given ($CumR_{merge}$, Cum%Comp). Only reflections with I/σ >1.5 are included in the above terms. The completeness for all reflections without an I/σ threshold, %AllComp is given. The $\sigma_a$ corresponds to the standard deviation in merging reflections as opposed to σ, the error involved in measuring individual intensities.

| Resolution | Num. Refs. | %Comp | Cum %Comp | All %Comp | $R_{merge}$ | Cum $R_{merge}$ | $<I/\sigma_a>$ |
|---|---|---|---|---|---|---|---|
| ∞ - 5.69 | 9263 | 96.6 | 96.6 | 99.6 | 0.172 | 0.172 | 19.1 |
| - 4.02 | 16480 | 94.0 | 94.9 | 99.1 | 0.233 | 0.208 | 10.0 |
| - 3.29 | 17800 | 88.3 | 91.9 | 96.9 | 0.260 | 0.226 | 5.8 |
| - 2.85 | 17665 | 74.0 | 85.6 | 94.0 | 0.312 | 0.239 | 2.9 |
| - 2.55 | 17013 | 57.3 | 77.5 | 91.3 | 0.323 | 0.245 | 1.9 |
| - 2.32 | 15998 | 42.2 | 69.1 | 88.1 | 0.316 | 0.248 | 1.5 |
| - 2.15 | 15082 | 31.1 | 61.2 | 84.6 | 0.302 | 0.249 | 1.3 |
| - 2.01 | 13884 | 23.5 | 54.4 | 80.8 | 0.307 | 0.250 | 1.2 |
| - 1.90 | 12520 | 17.6 | 48.4 | 76.1 | 0.256 | 0.250 | 1.1 |
| - 1.80 | 11291 | 13.4 | 43.3 | 70.8 | 0.263 | 0.250 | 1.1 |

## Discussion

The results presented here are among the first to be produced from this new tool for structural biology. The D-xylose isomerase result is significant in that, with a unit cell volume of ~1$MA^3$, it is by far the largest enzyme system to be studied to high resolutions with neutrons. Data were collected from a number of proteins during the first user cycle and are currently being analysed. The PCS has the potential to play an important auxiliary role in high throughput proteomics programs by providing unique information about the catalytic mechanism of newly discovered enzymes.

The PCS is funded by the Office of Science and the Office of Biological and Environmental Research of the U.S. Department of Energy.

## References

Evans, P., *"SCALA"*, MRC Laboratory of Molecular Biology, Cambridge, UK.
Habash, J., Raftery, J., Nuttall, R., Price, H.J., Wilkinson, C., Kalb, A.J. & Helliwell, J.R. (2000), *Acta Cryst. D* **56**, 541-550.
Hanson, B.L., Langan, P., Katz, A., Li, X., Harp, J., Glusker, J., Bunick, G. & Schoenborn, B.P. (2003) "A Preliminary Time-of-Flight Neutron Diffraction Study of *S. rubiginosus* D-Xylose Isomerase [EC5.3.1.5]" *Acta Cryst. D* (submitted).
Helliwell, J.R., Habash, J., Cruickshank, D.W.J., Harding, M.M., Greenhough, T.J., Campbell, J.W., Clifton, I.J., Elder, M., Machin, P.A., Papiz, M.Z. *et. al.*. (1989), *J.App. Cryst*. **22**, 483-497
Langan, P., Schoenborn, B.P. & Daemon, L.L. (2001) In *Neutron Optics*, James L. Wood & Ian Anderson, Eds, *Proceedings of SPIE* **4509**, 66-72.
Maeda, M, Chatake, T., Tanaka, I., Osterman, A. & Niimura, N. (2003) "Crystallization of large single crystals of cubic insulin for neutron protein crystallography studies", submitted to *J. Synch.Rad*., to be published as proceedings of ISDB 2003, Tsukuba, Japan.
Schoenborn, B.P., Court, J.D., Larson, A.C. & Ferguson, P.J. (1999), *Neutrons Research* **7**, 89-106.