

# Victor Lamzin - Research Report 2000

## Protein Crystallography

Group leader: V.S. Lamzin

Post-doctoral fellow: A. Razeto \*

Pre-doctoral fellows: R. Meijers, R.J. Morris\*, P.H. Zwart\*

Collaborators and short term visitors:

E. Cedergren-Zeppezauer, H.-W. Adolph, U. Heinz, I. Hubatsch, M. Kiefer, I. P. Kuranoa, V. R. Samygina, S. V. Antonyuk, S. Kochhar, M. Delley, H. Hottinger, J.-E. Germond, B. Galunsky, V. Kasche, K. Wilson, A. Kornelyuk, A. Dubrovsky, J. Sevcik, C. Jelsch, V. Pichon-Pesme, C. Lecomte, M. M. Teeter, R. H. Blessing

The structural investigation of macromolecules provides a deeper understanding of the properties and function of individual proteins, of their complexes and of even more complicated biological systems. The vast majority of macromolecular structures is determined by X-ray crystallography. In particular, crystal structures determined at atomic resolution (1.2 or higher) provide invaluable information. This resolution is essential to provide the electronic chemical details of biological processes. Availability of extremely accurate structural templates allow the modelling of features that, until very recently, were considered not to be identifiable in macromolecules.

The traditional approach of crystallographic structure determination involves a tedious and time demanding stage of manual building of a molecular model that is often subjective and relies heavily on user experience. It is the unified process of the building and the refinement of a crystallographic

macromolecular model that the ARP/wARP software automates, resulting in a technique that is faster, more objective and reliable than traditional methods. In the light of the forthcoming Genomics projects and to enable more biologists to carry out structural work without being experts in crystallographic techniques, it is essential to automate this process of model building and refinement.

Brief reports on the group's activity in these areas are given below.

### **Valence electron distribution in a small protein crambin**

*With C. Jelsch, V. Pichon-Pesme, C. Lecomte (Nancy University), M.M. Teeter (Boston College) and R.H. Blessing (Medical Research Institute Buffalo)*

The electronic charge density distribution of a molecule carries information that determines its

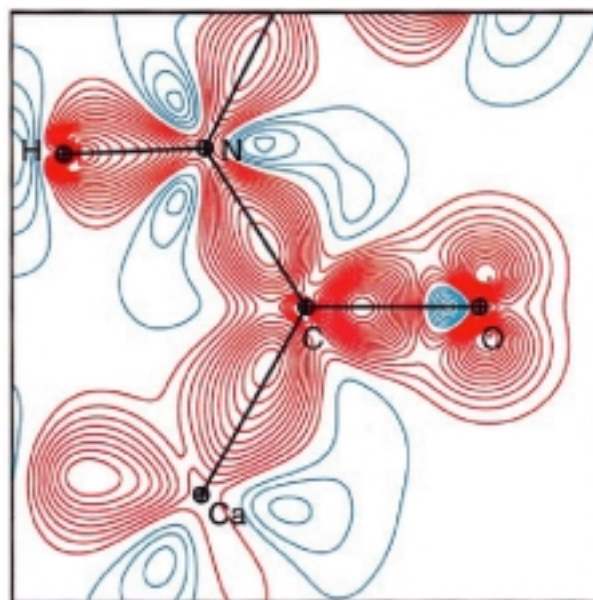
intermolecular interactions. The electrostatic potential and electric moments derivable from the charge density provide maps that can guide the design of molecules for specified interactions. Furthermore, powerful insights into the nature and strength of hydrogen bonding and ionic interactions result from analysis of the electron density gradient and Laplacian. Extension of such analyses to proteins would permit unique understanding of the driving forces between biological macromolecules as well as the subtleties of enzymatic reactions.

Here we report on the crystallographic charge density refinement of a 46 residues protein, crambin, which is present in seeds of *Crambe abyssinica* and homologous to membrane-active plant toxins. X-ray diffraction data from a single crambin crystal were measured at 100 K to 0.54 Å resolution at the BW7A beamline using synchrotron radiation with a wavelength of 0.54 Å and an imaging plate detector. The data were processed using the DENZO (Otwinowski & Minor, 1997) and DREAR (Blessing, 1997) program suites. The structure was refined using the MOPRO least-squares computer program for multipolar charged atom modeling, which is a version of MOLLY (Hansen & Coppens, 1978) extensively modified for protein applications.

The structure was first refined classically by using a spherical, neutral atom model. The residual electron density maps show systematic bonding density features and also contained a significant amount of random noise. Taking advantage of the repetition of the same chemical motif along the polypeptide main chain, the signal to noise ratio of the crambin residual map was increased by averaging over the 34 non-disordered peptide groups. After transfer of the statistically significant multipoles from the database, the residual features were greatly reduced. The average statistically significant multipole and partial net charge parameters of the peptide main-chain atoms thus were allowed to vary in

the refinement. The resultant peptide-averaged residual density map was much clearer, which is a convincing evidence of real improvement in the modelling.

The resolution and the quality of the diffraction data permitted the refinement of the average multipole parameters for the polypeptide backbone, Figure 1, but not yet for the individual atoms. However the work described here already opens the way to numerous electron density studies for proteins. The determination of the charges and the electronic distribution for the atoms in the active site of enzymes will provide new information and enable a better understanding of their function. An important application of charge density to structural biology would be the determination of electronic properties and oxidation states of reactive metallic centres in redox and electron transfer metalloproteins. Ultra-high resolution crystallographic studies performed in parallel on metalloproteins in combination with quantum mechanical calculations will yield new insights into redox processes in biology.



**Figure 1.** Experimental static map obtained by using atomic charges and multipole parameters refined against the crambin diffraction data.

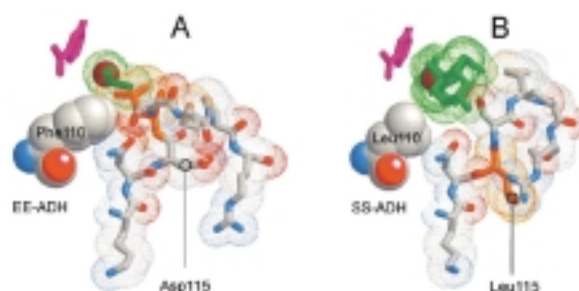
## Substrate specificity of horse liver alcohol dehydrogenase isozymes

With H.-W. Adolph, I. Hubatsch, M. Kiefer (Universitaet des Saarlandes, Saarbruecken) and Eila Cedergren-Zeppezauer (Lund University)

In horse liver, ethanol (EE-ADH) and steroid (SS-ADH) metabolising isozymes of alcohol dehydrogenase, ADH, are found together with the hybrid dimer, ES-ADH (Pietruszko *et al.*, 1966). There are a total of 10 amino acid differences between the E- and S-chains, including the deletion of AspEE115 located in the substrate-binding region. Substrate specificity of ADH is governed by the design of the substrate channel. It was proved by site-directed mutagenesis that deletion of AspEE115 rendered the mutated enzyme active toward steroid substrates. Only the S-subunit can convert 3 $\beta$ -hydroxysteroids, which are large and bulky substrates, whereas ethanol is a substrate for all three isozymes of ADH. Here we present the first X-ray structure of SS-ADH in complex with NAD/NADH and cholic acid. Cholic acid is a potent inhibitor of SS-ADH in the presence of NAD and differs only in the position of the 3-hydroxyl group from a 3 $\beta$ -hydroxy steroid substrate.

X-ray data to 1.5 Å resolution at 100 K were collected at the BW7B beamline using an imaging plate scanner. The data were processed using the HKL suite (Otwinowski & Minor, 1997), the structure solved with molecular replacement and refined to an R factor of 15 %. Most of the amino acid differences between the EE- and SS-enzymes influence various features: (i) the size of the substrate channel, (ii) the hydrophobicity of the substrate binding area, (iii) the charge distribution of the protein, and (iv) cause a large, local structural changes in a peptide loop, here denoted the steroid binding loop. All these contribute to the adaptation of the substrate channel in SS-ADH to accommodate steroid substrates. The major structural difference between EE- and SS-ADH

concerns the substrate binding channel. A certain flexibility of hydrophobic side chains allows for a limited variation in the shape of the channel within the central compartment. Phe110 and Leu116 in EE-ADH limit the compartment size to better fit smaller substrates rather than steroids (see Figure 2, panel A). AspEE115 has the largest impact on the size of the channel (Figure 2, panel B). The position of AspEE115 is inside the loop to which the carboxyl group forms hydrogen bonds to main chain amide nitrogens. LeuEE116, which occupies a volume in the vicinity of the active site metal, becomes buried inside the truncated loop. The cholic acid in the EE-enzyme occupies the full length of the substrate channel. Its carboxyl group ligates to the active site zinc ion. The results obtained provide a solid structural basis for analysis of differences in substrate specificity and stereoselectivity of the ADH isozymes.



**Figure 2.** The entrance to the inner compartment of the substrate channel in EE- and SS-ADH. (A) The EE-ADH-NADH-DMSO complex at 1.8- Å resolution. The dihydropyridine ring is in magenta and DMSO in green. LeuEE116 is closest neighbour to DMSO and PheEE110 blocks the entrance into the interior of the substrate channel. (B) The same region of the SS-isozyme. LeuSS115 takes the position of the deleted AspEE115 after a conformational change in the loop. The PheEE-110LeuSS mutation further enlarges the active site allowing the bulky cholic acid to dock.

## On the catalytic histidine in the NAD-specific D-lactate dehydrogenase

With S. Kochhar, M. Delley, J-E. Germond and H. Hottinger (Nestle Research Centre, Lausanne)

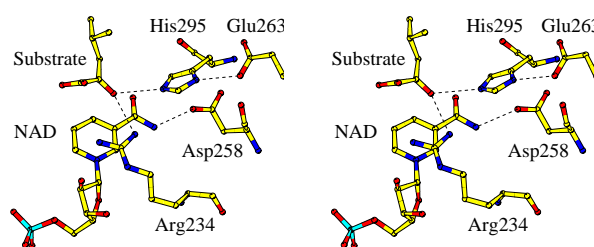
*Lactobacillus delbrueckii ssp. bulgaricus* (*L. bulgaricus*) is a Gram-positive, homofermentive, facultative anaerobic microorganism commonly used in the dairy industry for production of yoghurt. The end product of the glycolytic pathway is the D-isomer of lactic acid. The reduction of pyruvate to D-lactic acid, the last step in the pathway, is catalysed by NAD<sup>+</sup>-specific D-lactate dehydrogenase. A histidine residue, His296, is highly conserved in the sequence alignment and in homologous structures is located in the middle of the hydrophobic cluster near the binding site of the nicotinamide moiety and has been proposed to function as an acid/base catalyst by donating (or abstracting) a proton to the pyruvate carbonyl (or the lactate hydroxyl). Therefore, for the reversibility of the reaction, it is important that this histidine is able to switch between protonated and deprotonated forms.

We replaced His296 with lysine residues in D-lactate dehydrogenase with site-directed mutagenesis. The authenticity of each mutation was confirmed by sequencing the cloned PCR amplified fragment. The mutant, and also the wild-type recombinant enzyme, were produced successfully by the transformed *E. coli* cells. It exhibited a single band on SDS/PAGE electrophoresis, a single protein band on isoelectric focusing and identical cross-reactivity with antibodies raised against the wild-type enzyme. The N-terminal sequences of the mutant was identical to wild-type.

The His296 mutant was enzymatically active showing the specific activity of 50% of that for wild-type although mutation resulted in an acidic shift in optimum pH by about 1 unit. The pH profiles of  $K_m$  for substrate pyruvate showed behaviour qualitatively similar to that of the wild-type enzyme.  $K_m$  for NADH was reduced somewhat. There were no appreciable changes in  $k_{cat}$  values within the pH range

4.0–6.5. At very acidic pH the mutant is more active compared to wild-type. In addition, the presence of acetate ions increased the reaction rate for the mutant by up to a factor of 2.

In summary, the replacement of His296 with lysine does not affect the catalysis considerably and results only in poorer substrate binding. We propose that the acid catalyst, as well as the substrate-binding function, is accomplished by lysine with pKa lowered to < 6. Such a pKa value has been reported for a lysine in, e.g. acetoacetate decarboxylase and was attributed to its spatial proximity to another lysine residue (Highbarger *et al.*, 1996).



**Figure 3.** Stereo view of the active site of a representative of the D-specific dehydrogenase family, D-2-hydroxyisocaproate dehydrogenase. PDB id code 1DXY.

There are two structure-based factors that could contribute to the lowering of the lysine pKa in D-lactate dehydrogenase. First, there is the positively charged Arg235 side chain in the close vicinity which should destabilise the charge at the lysine. Second, the lysine side chain is longer than that of histidine. This prevents it forming a hydrogen bond with the negatively charged Glu264, which in the wild-type stabilises the positive charge on the His296, Figure 3. Poorer affinity to pyruvate is probably caused by a weaker hydrogen bond interaction of lysine and substrate carbonyl. Details of its interaction with the substrate and the active site environment remain an open question and await accurate three-dimensional structural information.

## Automated crystallographic data analysis

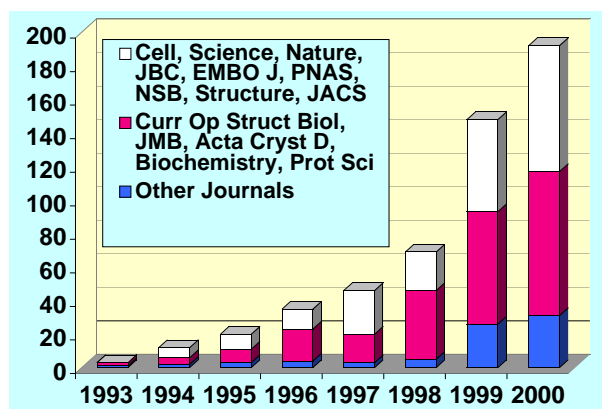
*With A. Perrakis (EMBL Grenoble and NKI Amsterdam)*

As soon as a well diffracting crystal of the macromolecule is available, several consecutive steps have to be taken before a reliable model can be deposited in a database. Good *diffraction data* must be collected initially and subsequent data processing and reduction derives the *intensities* of the diffraction spots. In order to construct an *electron density map*, both the *amplitudes* and the *phases* of the diffracted X-rays need to be known. While the amplitudes can be straightforwardly derived from the intensities, the phases have to be obtained indirectly. Recent advances in molecular biology (expression of selenomethionine substituted proteins) coupled with the availability of tuneable synchrotron radiation and development of excellent software, were the main driving forces for the success of the multi-wavelength anomalous dispersion (MAD) technique (Ogata, 1998). Metalloproteins, or the soaking of protein crystals in solutions containing heavy atoms and, more recently, halides (Dauter *et al.*, 2000) have popularised the MAD and single-wavelength (SAD) measurements even further. This technology has essentially re-defined the crystallographic phase problem and the emphasis is now shifting to obtaining the highest quality experimental data. Considering the rapidly growing database of known macromolecular structures, developments are foreseen towards the *generalised molecular replacement*, where various topological motifs and templates will be screened for potential structural homology. When phases are available, the *electron density map* - the net result of a crystallographic experiment - can be computed and has to be interpreted in terms of a *macromolecular model*. The macromolecular model then undergoes a *refinement* procedure in which the parameters are optimised to best fit the experimental data and stereochemical expectations. During the refinement the

electron density map improves and may require considerable adjustment and re-building of the model. The structure determination finishes with the model *validation* and deposition.

In an automated process all these modules in structure determination should not only be tied together with a user-friendly interface. Rather, information gathered from any step within the process should be fed back to the optimisation of the preceding steps. This is often utilised by expert crystallographers, but remains the major challenge for automation. High throughput plans can be hampered by the time which is required to complete a crystallographic structure determination and may vary from hours to years.

We have developed the ARP/wARP software suite (Perrakis *et al.*, 1999) that offers complete automation in the iteration between model building and model refinement, Figure 4. The basic concept underlying the ARP/wARP approach is the unified view of the refinement and the iterative building of model fragments in the interpretable regions of the density map. The hybrid macromolecular model (a mixture of protein fragments and free atoms) has been introduced and not only the parameters of the model but the model itself are allowed to change *on the fly* during the refinement. Accounting for the limitations in resolution of the X-ray data required (2.3 or higher), which spans about two thirds of the Protein Data Bank content, and allowing a reasonable failure rate, it is expected that when initial phases are available, about 50% of structure solutions can proceed with automatic model building using ARP/wARP. Although this is a step towards full automation, further developments are needed to address the remaining 50% (often the most difficult ones) to deliver a final and well-validated model.



**Figure 4.** Citations on the use of ARP/wARP. In December 1999 the Version 5.1 for automatic tracing of protein molecules has been released. The software suite has been obtained by about 400 laboratories world-wide.

### Publications during the year

Jelsch C., Teeter M.M. Lamzin V., Pichon-Pesme V., Blessing R.H. & Lecomte C. (2000) Accurate protein crystallography at ultra-high resolution: valence electron distribution in crambin. *Proc. Natl. Acad. Sci. USA*. **97**, 3171-3176.

Kochhar, S., Lamzin, V.S., Razeto, A., Delley, M., Hottinger, H. & Germond, J.-E. (2000) Roles of His205, His296, His303 and Asp259 in catalysis by NAD<sup>+</sup>-specific D-lactate dehydrogenase. *Eur. J. Biochem.* **267**, 1633-1639.

Samygina, V.R., Antonyuk, S.V., Lamzin, V.S. & Popov, A.N. (2000) Improving the X-ray resolution by reversible flash-cooling combined with concentration screening, as exemplified with PPase. *Acta Crystallogr.* **D56**, 595-603.

Adolph, H.-W., Zwart, P., Meijers, R., Hubatsch, I., Kiefer, M., Lamzin, V. & Cedergren-Zeppezauer, E. (2000) Structural basis for substrate specificity differences of horse liver alcohol dehydrogenase isozymes. *Biochemistry* **39**; 12885-12897.

Lamzin, V.S., Perrakis, A., Bricogne, G., Jiang, J., Swaminathan, S. & Sussman, J.L. (2000)

Apotheosis, not apocalypse: methods in protein crystallography. *Acta Crystallogr.* **D56**, 1510-1511.

Lamzin, V.S. & Perrakis, A. (2000) Current state of automated crystallographic data analysis. *Nature Struct. Biol.* **7**, Structural Genomics Supplement, 978-981.

### Other references

Blessing, R. H. (1997) *J. Appl. Crystallogr.* **30**, 421—426.

Dauter, Z., Dauter, M. & Rajashankar, K.R. (2000) Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr.* **D56**, 232-237.

Hansen, N. K. & Coppens, P. (1978) *Acta Crystallogr.* **A34**, 909—921.

Highbarger, L.A., Gerlt, J.A. & Kenyon, G.L. (1996) Mechanism of the reaction catalyzed by acetoacetate decarboxylase. Importance of lysine 116 in determining the pK<sub>a</sub> of active-site lysine 115. *Biochemistry* **35**, 41-46.

Ogata, C.M. MAD phasing grows up. (1998) *Nature Struct. Biol.* **5**, 638-640.

Otwinowski, Z. & Minor, W. (1997) *Methods Enzymol.* **276**, 307—326.

Perrakis, A., Morris, R. & Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463.

Pietruszko, R., Clark, A., Graves, J. M. H., & Ringold, H. J. (1966) *Biochem. Biophys. Res. Comm.* **23**, 526-534.