

Lamzin Group

Protein crystallography

Group leader: Victor S. Lamzin

Postdoctoral fellows: Adelia Razeto^{*},
Gerlind Sulzenbacher^{**}

PhD students: Rita Berisio^{*}, Rob Meijers,
Richard J. Morris^{*}, Adelia Razeto^{**},
Gerlind Sulzenbacher^{**}, Petrus H. Zwart^{*}

Collaborators and visitors: H.W. Adolph,
P. Alzari, J. Benach, E. Cedergen-Zeppezauer,
Z. Dauter, B. Galunsky, S. Henk, V. Kasche,
S. Kochhar, A. Kornelyuk, C. Kratky,
I. Kuranova, R. Ladenstein, L. Mazzarella,
A. Perrakis, S. Santesson, J. Sevcik,
G.M. Sheldrick, F. Sica, O. Sinitzina,
O. Tapia, M.M. Teeter, M. Vudmaska,
K.S. Wilson, A. Zagari

The structural investigation of macromolecules provides a deeper understanding of the properties and function of individual proteins, of their complexes and even of more complicated biological systems. Protein function and stability depend critically on a delicate balance of numerous interactions of similar strength, which are markedly influenced by their environment. The presence of a liquid solvent in protein crystals allows the manipulation of the physico-chemical properties of the protein. It is possible to study processes where proton transfer plays a crucial role, or where the ligand-binding affinity is strongly regulated by pH or ionic strength. X-ray crystallography can be used to single out subtle alterations caused by a systematic stepwise modification of a physico-chemical parameter, which may be relevant to the biological function of the protein, while keeping all the other parameters which may affect the overall structure essentially constant. Conformational changes that range from movement of large domains to a "motion" of a side chain of a catalytic residue or bound ligand, have direct biological implications. In particular, crystal structures determined at atomic resolution (1.2 Å or higher) provide invaluable information. This resolution is essential to provide the electronic details necessary to obtain the information for a deeper understanding 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 vast majority of macromolecular structures is determined by X-ray crystallography. The traditional approach involves a stage of manual building of a molecular model that is often tedious, time demanding, subjective and relies heavily on user experience. There is an anticipation of considerable human effort involved within high throughput crystal structure determination. It is the unified process of building and 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 structural principles of macromolecular function and enzymatic catalysis on representative examples of proteases, dehydrogenases and ribonucleases are presented. The group is also active in the development of automated objective procedures for crystallographic structure solution, refinement and modelling (ARP/wARP) as outlined below.

Protein titration in the crystal state

With F. Sica, A. Zagari and L. Mazzarella
(University of Naples "Federico II")

For pH-controlled processes, the location of hydrogen atoms can provide a direct determination of the protonation state of residues essential for macromolecule stability or function. We have used X-ray diffraction to uncover how pH affects a protein structure, with the purpose of highlighting the role of the protonation equilibrium in a biological process.

We present an analysis of the effect of pH on a protein structure in the crystalline state using RNase A as a model system. This enzyme shows a pH dependent reaction mechanism involving a "2-proton shift". Although RNase A is a very well studied enzyme, its two step reaction mechanism has recently been the subject of some reinterpretation. The first step of the reaction, transphosphorylation, was proposed to proceed through the formation of an intermediate phosphorane monoanion (Breslow & Chapman, 1996). Here we focus on the structural aspects related to the pH-

titration, which has been achieved for the first time for a protein in a crystal state.

RNase A crystals were transferred to the buffered solutions, using a stepwise procedure in which the pH* was adjusted in steps of 0.2 units, and were maintained under the final conditions for a week with daily changes of the mother liquor. Diffraction experiments were carried out at room temperature on the EMBLX11 beam line at the DORIS storage ring. Atomic resolution (around 1.1 Å) X-ray diffraction data were collected for crystals at six pH*s ranging from 5.2 to 8.8. The high stability of the crystals in the X-ray beam allowed the collection of complete data using only a single crystal for each pH*. Six independently refined structures reveal subtle, albeit well-defined variations directly related to the pH titration of the protein.

The main structural alterations induced by the changes in pH occur in the active site. Previous crystallographic studies indicated that His-12 is rigidly anchored to the protein matrix. The present data also show that the overall conformation of the His-12 side chain is not influenced by the change of pH. In addition, His-12 shows very low mobility in all six models, as judged from the refined values of atomic displacement parameters. Consequently, the noise in the electron density around this residue is considerably lower than the

average. A comparison of the six electron density maps at different pH* allows a direct observation of the His-12 deprotonation process. The figure shows the presence of hydrogen linked to the proximal nitrogen ND1 over the whole pH range. In contrast, hydrogen linked to NE2 is present at acidic pH*, but disappears at higher pH*. This provides a further experimental support for the proposed mechanism, which suggests that the NE2 atom of His-12 exchanges a proton with substrate (Cuchillo *et al.*, 1997). To our knowledge, this is the first direct structural evidence for the pH-dependent deprotonation of a residue in a crystalline environment (see Figure 1).

The values bond angles of the imidazole ring of His-12, which were not restrained during the structure refinement, correlate with the change of the protonation state. The two endocyclic bond angles show the highest deviation as a function of pH. The overall trend generally agrees with the reference values typically used as targets in restrained refinement of protein crystal structures, even though the value of the bond angle at CD2 is systematically higher. Consistent with the trend there is a considerable lengthening of the hydrogen bond between the donor nitrogen ND1 of His-12 ND1 and the carbonyl O of Thr-45 (data not shown). The dependence is of a sigmoidal shape with an apparent pK_a of 7.0±0.1.

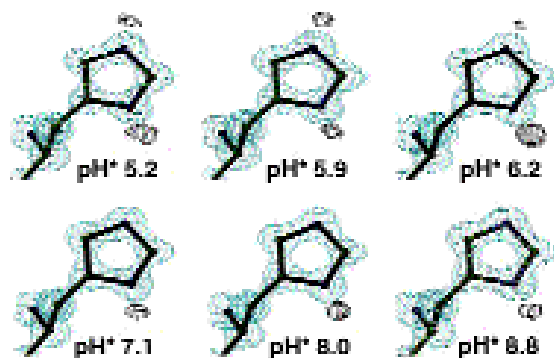


Figure 1

Sequence of snapshots showing the deprotonation of the imidazole ring of the His-12 NE2 atom: the difference ($F_o^{1/2}E_o^{1/2} - F_c^{1/2}E_c^{1/2}$) maps are contoured at 1.6 σ .

It is generally believed that information coming from an X-ray diffraction experiment is static. One can however use a series of static images to obtain a dynamic picture. The availability of a sequence of structures in different conditions provided information about a dynamic process. The structural changes are small, but significant, given the high accuracy (0.02 Å coordinate error) of the refined models. This study gives further evidence for the advantage of atomic resolution X-ray crystallographic analyses for revealing small but significant structural changes which provide clues to the function of a biological macromolecule (see Figure 2)

Stereoselectivity and catalysis by D-lactate dehydrogenase

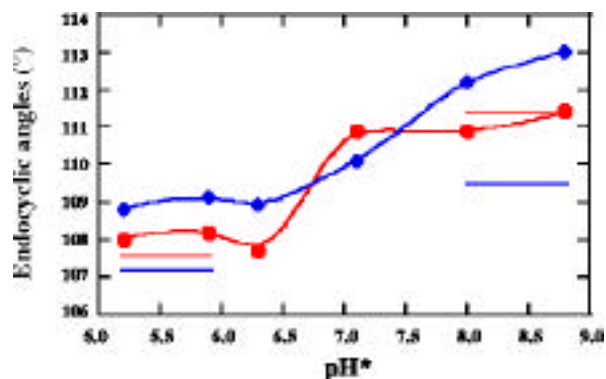
With S. Kochhar (Nestle Research Centre Lausanne) and K.S. Wilson (York University)

NAD-dependent D- and L-lactate dehydrogenases (LDH) catalyse the reversible interconversion of pyruvate to D- and L-lactate respectively. Despite the fact that the enzymes catalyse the reaction that differs only in the chirality of the product/substrate, they belong to the two distinct families, the D- and L-2-hydroxyacid dehydrogenases, with very low sequence and structural similarity between them. Their stereoselectivity has a great potential in biotechnology for the production of chiral synthons from prochiral precursors. However, the use of D- and L-LDH is still limited since the enzymes have a narrow specificity. Protein engineering and conventional screening are currently considered to widen their application in synthesis of chiral compounds.

Structural studies on D-LDH and related enzymes show that each subunit of the dimer consists of two globular domains, connected by a flexible hinge. The active site, located in the interdomain cleft, is shield-

Figure 2

Variation of the endocyclic bond angles at CD2 (dark line) and CE1 (grey line) of the catalytic residue His-12 versus pH*. The estimated uncertainties are of the order of 0.8°. The straight lines indicate the corresponding target values (Engh & Huber, 1991) for the protonated and deprotonated states of a histidine.



ed from the solvent after the catalytic domain moves as a rigid-body towards the coenzyme binding domain on NADH and substrate binding. Mutagenesis studies indicate three residues essential for catalysis (Kochhar *et al.*, 1992). His-297 is the acid/base catalyst. Glu-265 forms a strong H-bond with His-297 and slightly increases its pKa to 7.5 from the typical value of about 7. The third residue is Arg-236 whose putative role in binding the carboxylate and polarising the carbonyl bond is still under debate. The mechanism of domain closure, substrate binding and catalysis is still unclear. Furthermore, the substitution of the acid-base catalyst His-297 to lysine does not result in a significant change in k_{cat} and K_m in pyruvate reduction, but only in a shift in the pH optimum from 7-7.5 to 6.

To gain a better understanding of the catalytic mechanism of D-LDH, crystallographic studies have been carried out on *L. Bulgaricus* D-LDH complexed with NADH and on the D-LDH H297K mutant (D-LDH297).

Crystals of D-LDH were grown with Me PEG 2K and ammonium sulphate in presence of NADH in the space group P43212, with unit cell parameters $a=79.4$, $b=79.4$, $c=228.5$ Å and one dimeric molecule in the asymmetric unit. X-ray diffraction data have been collected to 2.2 Å resolution. The structure was solved by a "domain" molecular replacement, using *L. helveticus* D-LDH as a model: independent rotation and translation solutions had to be found for individual domains. The structure was refined to an R-factor of 20.9 %. The two subunits display strong asymmetry: the opening angle of the two catalytic domains with respect to the core coenzyme binding domains differ by 16°. The conformation of subunit A is "open", typical of an apo-enzyme, while the conformation of subunit B is "closed". NADH in subunit A is only 30 % occupied, while in subunit B it is fully present and there is a sulphate in the substrate binding pocket. The conformational asymmetry is attributed to different crystal environments around the two subunits (see Figure 3)

The structure allowed modelling the substrate binding in the closed (active) conformation of the enzyme. Stereospecificity of the catalysis is reflected by the fact that pyruvate in one orientation only could have been fitted into the active site. The modelling is in agreement with kinetic data and with one of the models proposed earlier. In addition, this structure provided a deeper insight into substrate specificity: on domain closure a cluster of hydrophobic residues packs tightly around the methyl of the modeled pyruvate. This is the reason why D-LDH displays the highest specificity for pyruvate, while it has a much lower affinity for bulkier substrates. Substrate binding plays an important role in the stabilisation of domain closure and activation of the enzyme. The substrate anchors to the coenzyme binding domain, through interactions with the positively charged His-297, Arg-236, and to the catalytic domain, the main chain nitrogens of Val-79, Gly-80 and the side chain of Asn-78. The carboxylate of pyruvate approaches the dihydronicotinamide moiety and activates it for the hydride transfer (see Figure 4).

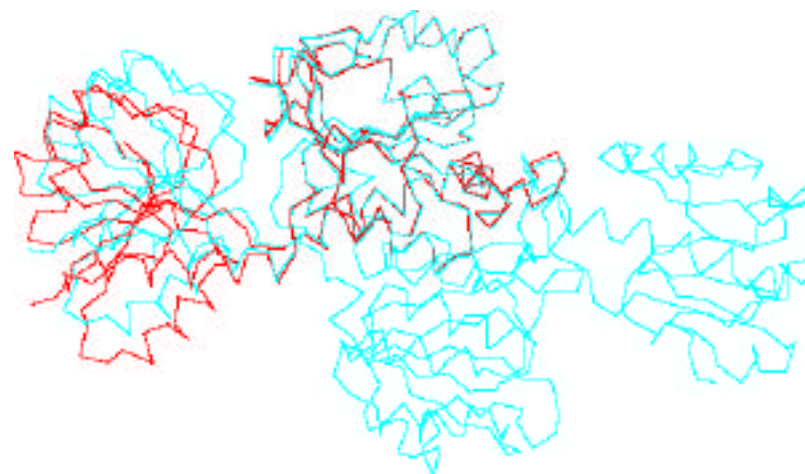


Figure 3

The D-LDH dimer with subunits superimposed onto each other.

Crystals of the mutant D-LDH297 enzyme belong to a different space group C2221 with unit cell $a=94.3$, $b=188.0$, $c=193.4$ Å and there are two dimers in the asymmetric unit. Data to 1.9 Å resolution have been collected. The structure was solved by molecular replacement, using the wild type model described above, and was refined to an R-factor of 20.6 %. Electron density maps confirmed the replacement of the histidine to lysine in the enzyme active site. Lys-297 assumes different conformations in different subunits of D-LDH297. Only in subunit A Lys-297 points towards the catalytic Arg-236 as a result of the sulphate ion neutralising the positively charged active site. The side-chain of Glu-265 is poorly defined in the density, in contrast to the wild-type enzyme. This is due to the lack of a strong interaction with a positively

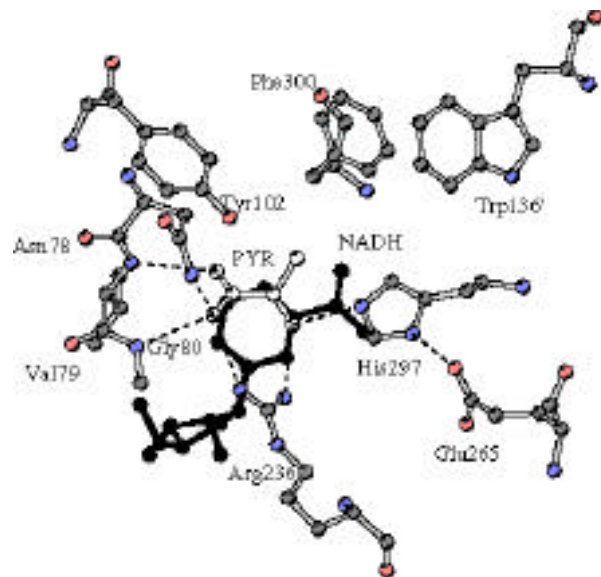


Figure 4

Substrate modeled in the active site of D-LDH.

charged histidine, forming the charge-relay system in the wild-type. Lys-297 positive charge is thereby not neutralised by Glu-265 and is destabilised by the vicinity of Arg-236. This arginine is forced to be close to Lys-297 on substrate binding. These can be the reasons that pKa for the lysine is lowered to make it catalytically competent.

Crystal structure of a glutamic acid specific serine protease

With O. Sinitzina and I.P. Kuranova
(Institute of Crystallography, Moscow)

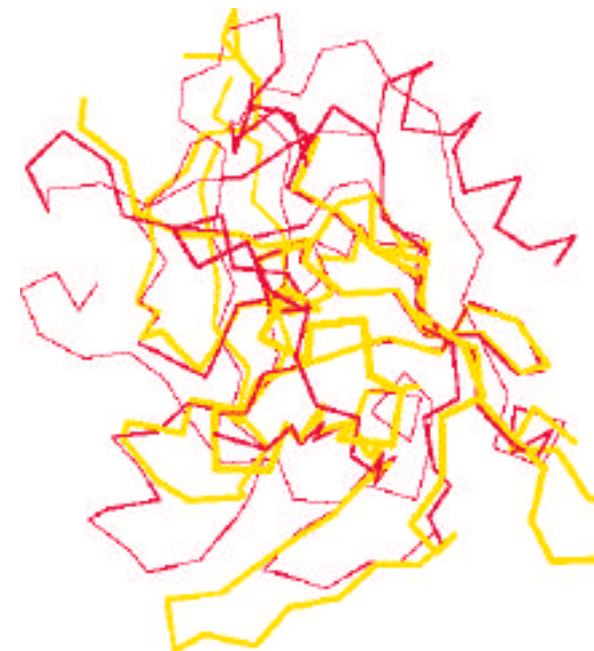
Glutamic acid specific serine protease (GASSP) from *Bacillus intermedius* belongs to the subgroup of serine proteases which cleave the peptide bonds on the carboxylic side of either a glutamate or aspartate residue. These enzymes play an important biological role, taking part in viral processing and some disease states. GASSP is used extensively for the fragmentation of proteins and the enzymatic syntheses of new peptide bonds. We have undertaken a project to solve the X-ray crystal structure of this commonly used protease to learn its structural relation to other serine proteases. Knowledge of the 3-D structure would offer a rational basis for optimisation of the enzyme performance and would help in understanding the mechanism that rules the specific substrate recognition.

Crystals of GASSP were grown in 0.1 M Tris-HCl buffer, pH 7.0, with 1.3 M phosphate and 3 % MPD at a protein concentration of 15 mg/ml. The use of macroseeding has optimised the growth of single crystals. The dimensions of the crystals were approximately 0.4x0.2x0.1 mm. X-ray data were collected at the X11 beamline of EMBL at DESY to a resolution of 1.5 Å at room temperature.

An attempt was made to solve the GASSP structure using the molecular replacement technique. Based on

Figure 5

Superposition of the CAatom representation of the search model derived from ETA (thick line) and the final model of GASSP (thin line).



sequence analysis, the most closely related protease for which the X-ray structure is known is epidermic toxin A (ETA, Cavarelli et al, 1997). The sequence identity between the two proteins is 32 %. The only available structure of glutamic acid specific serine protease from *S. griseus* (HPG) is even more distant and has only 25 % sequence identity. The molecular replacement only succeeded after two sections inside the search model of ETA were clipped off. In effect, 30 residues out of 170 were omitted. The solution still had to go

through a remarkable metamorphosis before the final model for GASSP was obtained (see Figure 5).

Attempts to refine the model using conventional techniques have not been successful. The successful protocol employed parallel refinement of 4 free atom models within the ARP/wARP suite (Perrakis *et al.*, 1999). A map based on weighed observed structure factors was further used for placement of free atoms, refinement, updating and automatic model building using the warpNtrace protocol. Fully automated structure construction resulted in 210 of the 215 peptide units being traced correctly and 70 % of the side chains

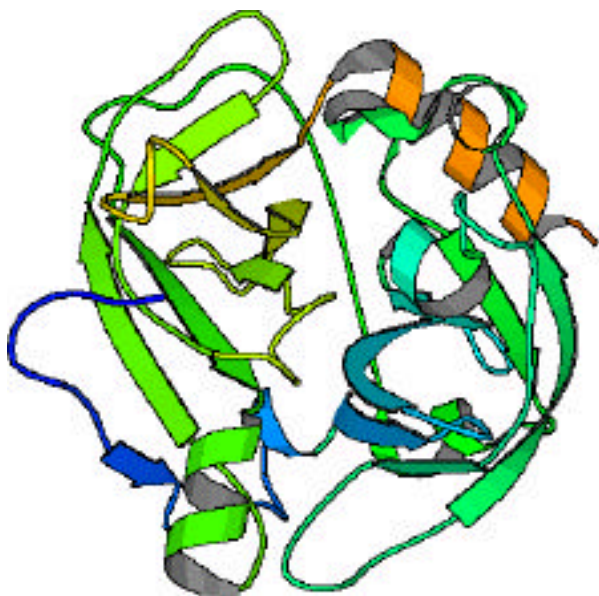


Figure 6

Ribbon representation of the GASSP molecule. The catalytic domain is on the right and the substrate recognition domain on the left.

built. Post mortem analysis indicated that the critical entity for the molecular replacement solution was a single β -barrel conserved in all serine proteases. The protocol that was used to obtain the structure from a marginal solution may have a wider application and further methodological studies are underway.

The structure of GASSP consists of two subdomains and seems to reveal a new variation on old themes. The structure contains one sulphur bridge in the same part of the fold as the single sulphur bridge that is found in HPG. The core of the catalytic domain is based on the β -barrel typical for serine proteases. The substrate recognition domain is unique in architecture containing four parallel beta sheets (See Figure 6).

Atopology search on the TOPS server of EBI-Hinxton revealed that the arrangement of secondary structure elements of the GASSP substrate binding domain has some resemblance to the defective binding domain of bovine procarboxypeptidase A-S6, a zymogen E (Pignol *et al.*, 1994). The latter is a pro-enzyme that is protected from proteases and extreme environments. Upon activation of zymogen E, conformational changes occur. Since GASSP is fully active, the conformational changes observed when the two domains are superimposed, give insight into the activation of a zymogen E related protease.

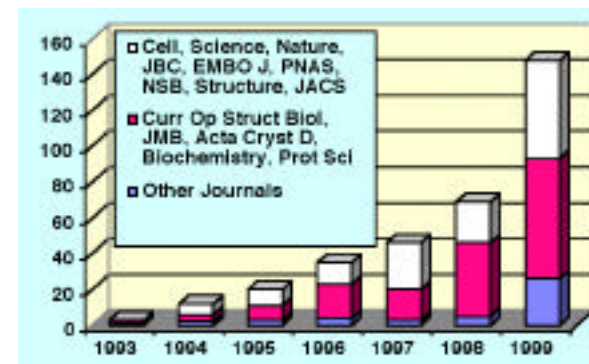
The ARP/wARP suite for automated building and refinement of macromolecular crystal structures

With A. Perrakis (EMBL Grenoble)

In view of the rapidly developing challenging areas of structural molecular biology, there is now a pronouncing demand for automated methods for solving macromolecular crystal structures. Piloting projects are now starting to emerge to develop procedures for

Figure 7

Citations on the use of ARP/wARP. In September 1998 the Version 5.0, including automatic tracing of protein molecules, has been released. The software suite has been obtained by about 230 laboratories world-wide.



high-throughput determination of macromolecular structures by X-ray crystallography. Availability of synchrotron radiation facilities, fast and accurate 2-D detectors and techniques for cryogenic freezing of protein crystals have made it possible to collect X-ray diffraction data from a large number of protein crystals in record time. The procedure of structure determination may be eased by merely tying together existing methods and programs with a user-friendly interface. However the task of automation should now be really viewed as a need for re-examination of the whole process of structure determination.

The conventional view of the refinement of macromolecular crystal structures is the optimisation of the model parameters to fit both the experimental data and a set of *a priori* stereochemical observations. While the values of the parameters are optimised during the minimisation cycles, the parameterisation itself remains fixed. The ARP/wARP suite challenges this classical view by allowing the actual model to vary via



real space manipulation. The refinement of a hybrid model, a combination of automatically recognised polypeptide fragments with free atoms, constitutes the basic ARP/wARP concept. The polypeptides provide additional information in a form of restraints while the free atoms describe remaining prominent features in the electron density. As refinement proceeds, the hybrid model converges towards a conventional protein structure: the polypeptide chains transform to a (essentially) complete protein chain while free atoms convert to solvent structure.

The method in its present implementation requires good quality X-ray data to a sufficiently high resolution (Perrakis *et al.*, 1999). The main applications are: (1) Delivery of a model starting from MR or MIR(AS) phases. The X-ray data should extend to resolution of 2.3 Å or higher. The phases do not have to necessarily extend to that resolution. (2) *Ab initio* solution of metalloproteins or heavy atom derivatives. A few single site heavy atom derivatives of proteins with a molecular mass of about 20 kDa have been solved automatically with SAS data extending to 1.7 - 2.0 Å. (3) Building solvent structure. The X-ray data should be to at least 2.5 Å and the model available. (4) Partial model re-building into the density map ("mutations" of side chains) - some test examples have shown the feasibility of this with data extending to 3.0 Å resolution only (see Figure 7).

The ARP/wARP suite for refinement and building of protein structures in its present state is already in a position to promote progress in automating the steps of deriving a complete structural model from the X-ray data. The time required for building a protein structure - still one of the main bottlenecks of the structure determination process - can be shortened from several man-days or even man-months to a few CPU hours on inexpensive workstations.

Publications during the year

Berisio, R., Lamzin, V.S., Sica, F., Wilson, K.S., Zagari, A. & Mazzarella, L. (1999). Protein titration in the crystal state. *J. Mol. Biol.*, 292, 845-854

Declercq, J.P., Evrard, C., Lamzin, V. & Parello, J. (1999). Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 Å) and at low temperature (100 K). Evidence for conformational multistates within the hydrophobic core. *Protein Sci.*, 8, 2194-2204

Lamzin, V.S., Morris, R.J., Dauter, Z., Wilson, K.S. & Teeter, M.M. (1999). Experimental observation of bonding electrons in proteins. *J. Biol. Chem.*, 274, 20753-20755

Parisini, E., Capozzi, F., Lubini, P., Lamzin, V., Luchinat, C. & Sheldrick, G.M. (1999). *Ab initio* solution and refinement of two high-potential iron protein structures at atomic resolution. *Acta Crystallogr. D Biol. Crystallogr.*, 55, 1773-1784

Perrakis, A., Antoniadou-Vyza, E., Tsitsa, P., Lamzin, V.S., Wilson, K.S. & Hamodrakas, S.J. (1999). Molecular, crystal and solution structure of a beta-cyclodextrin complex with the bromide salt of 2-(3-dimethylaminopropyl)tricyclo[3.3.1.1(3,7)]decan-2-ol, a potent antimicrobial drug. *Carbohydr. Res.*, 317, 19-28

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

Sabini, E., Sulzenbacher, G., Dauter, M., Dauter, Z., Jorgensen, P.L., Schulein, M., Dupont, C., Davies, G.J. & Wilson, K.S. (1999). Catalysis and specificity in enzymatic glycoside hydrolysis: a

2,5B conformation for the glycosyl-enzyme intermediate revealed by the structure of the *Bacillus agaradhaerens* family 11 xylanase. *Chem. Biol.*, 6, 483-492

Sulzenbacher, G., Mackenzie, L.F., Wilson, K.S., Withers, S.G., Dupont, C. & Davies, G.J. (1999). The crystal structure of a 2-fluorocellotriosyl complex of the *Streptomyces lividans* endoglucanase CelB2 at 1.2 Å resolution. *Biochemistry*, 38, 4826-4833

Other references

Breslow, R. & Chapman WH, Jr. (1996). On the mechanism of action of ribonuclease A: relevance of enzymatic studies with a p-nitrophenylphosphate ester and a thiophosphate ester. *Proc Natl Acad Sci U S A*, 93, 10018-10021

C. Engh, R.A. & Huber R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, 47, 392-400

Cavarelli, J., Prevost, G., Bourguet, W., Moulinier, L., Chevrier, B., Delagoutte, B., Bilwes, A., Mourey, L., Rifai, S., Piemont, Y. & Moras, D. (1997). The structure of *Staphylococcus aureus* epidermolytic toxin A, an atypical serine protease, at 1.7 Å resolution. *Structure*, 5, 813-824

Cuchillo, C.M., Vilanova, M. & Noguès, M.V. (1997). Pancreatic Ribonucleases. In "Ribonucleases: Structures and Functions", G. D'Alessio & J.F. Riordan (eds.) Academic Press, pp. 271-304

Kochhar, S., Hunziker, P.E., Leong-Morgenthaler, P. & Hottinger, H. (1992). Primary structure, physicochemical properties, and chemical modification of NAD(+)-dependent D-lactate dehydrogenase.

Evidence for the presence of Arg-235, His-303, Tyr-101, and Trp-19 at or near the active site. J. Biol. Chem., 267 8499-8513

Pignol, D., Gaboriaud, C., Michon, T., Kerfelec, B., Chapus, C. & Fontecilla-Camps, J.C. (1994). Crystal structure of bovine procarboxypeptidase A-S6 subunit III, a highly structured truncated zymogen E. EMBO J., 13, 1763-1771