Cell. Mol. Life Sci. 64 (2007) 1959 – 1969
1420-682X/07/151959-11
DOI 10.1007/s00018-007-7195-7

Cellular and Molecular Life Sciences

Review

From atoms to proteins

A. Schmidt* **and V. S. Lamzin**

European Molecular Biology Laboratory c/o DESY, Notkestrasse 85, 22603 Hamburg (Germany),
Fax: +49 40 8 99 02 149, e-mail: andrea@embl-hamburg.de

Received 20 April 2007; accepted 20 April 2007
Online First 11 May 2007

Abstract. The investigation of biological macromolecules and the characteristics that determine their function has been of particular interest over the last decades. Here we overview some modern approaches for making the most of the 3-D protein structural information, with a distinctive emphasis on macromolecular crystallography and complementary techniques used to establish the structure-function relationship. A tight link between the biology of the cellular processes and the underlying chemistry of protein function governs the flow of the presented material. The reader will be lead through the basic principles of protein structure analysis and the means to capture the characteristics that portray the function. The techniques exploiting high-resolution data and allowing quantification of molecular motion and structure-activity relationship are given particular attention.

Keywords. Proteins, macromolecular crystallography, structure determination, atomic resolution, catalytic function, structure-function relationship.

Introduction

Proteins are large biological molecules that are capable of performing a wide range of functions. It is not surprising that the investigation of these macromolecules and the characteristics that determine their function has been of particular interest over the last decades. The formation, development and the sustenance of a living organism are governed by the correct function of proteins. Metabolic pathways such as the fatty acid synthesis ([1] and references therein) or the citrate cycle [2], both essential for almost any organism, are probably the best-known examples. There are numerous other processes, such as signal transduction or expression regulation, where the high degrees of specificity and efficiency of proteins and their interactions are indispensable for the correct biological function [3].

Proteins may act as a specific recognition site (good examples would be antibodies) or play a framework role, acting as "glue" or forming tissue such as skin or muscles. Proteins may tunnel substances through otherwise hostile environment, as can be seen from ion transport through cell membranes. Finally, proteins can be highly efficient, specific and versatile chemical catalysts (enzymes), on which we will place particular emphasis throughout this text.

In enzymatic catalysis two compounds (small organic or inorganic molecules) may be combined into a larger molecule (by synthetic enzymes), or large compounds dismantled into smaller ones (by degrading enzymes). Compounds may also be stereochemically rearranged (by racemases). Enzymes are known to catalyse about 4000 reactions [4] and catalysis generally involves the steps of substrate binding, chemical modification (intermediate state) and product release. For each of

* Corresponding author.

these steps a number of structural requirements need to be fulfilled. Binding small molecules to the protein in its active site lowers the energy barrier of the chemical reaction, which at the same environmental conditions but without the enzyme would not normally occur. Enzymes are not only capable of binding substrates specifically, but also of providing the environment that stabilises high-energy transition states, which favours the chemical conversion. Binding sites and contact properties, as well as chemical properties such as reactivity and specificity are determined by surface characteristics, which in turn are set by the arrangement of the atoms within that area. It is the protein as a whole that determines the specific properties of the atoms constituting its active site.

What makes a protein?

Proteins are composed of non-branched chains of L-α-amino acids that are connected to each other through the so-called peptide bonds. In a cell, the polypeptide chains are assembled in the ribosomes, a process that in itself is dependent on the correct functioning of a series of enzymes [5]. Ribosomes produce a protein from an RNA sequence template and are large, complex arrangements of both RNA (ribonucleic acid) and protein molecules that play various roles. Deciphering the sequence of the mRNA threaded through the ribosome in order to deliver the correct amino acid residue is achieved at a specific codon-reading site. The amino acid residues are linked to each other at a synthesis site and the new protein chain is threaded out of the ribosome complex. The subsequent assembly of the polypeptide chain into the correct 3-D structure involves the formation of secondary structural elements such as helices and strands, and the arrangement of those elements into the tertiary structure, which gives the molecule its shape. Conformational freedom along the protein backbone exists only as rotations around the N-Cα and the Cα-C bonds where, due to stereochemical constraints, only a limited number of combinations are possible. For decades this has been exploited in the form of the Ramachandran plot, widely used in structure determination, validation and analysis [6].

Protein folding is a complex process for which the driving forces are still only understood to a small extent. Extensive efforts continue to be put into the elucidation of this process from the bioinformatics side using molecular dynamics simulations [7] as well as a variety of structural studies [8].

Proteins through the looking glass

A large number of methods have been established with an aim of unraveling the many questions imposed by the mystery of protein function. Some of them focus on the biophysical properties, some on the chemical composition, while others are geared towards the study of 3-D structural properties and the relation of that knowledge to the catalytic function. There is a variety of means by which the structural information of a macromolecule can be obtained. Biochemical or biophysical methods, *e.g.* gel filtration, electrophoresis or dynamic light scattering are used routinely for the validation of the protein sample during production and reveal basic information about the molecular weight and the oligomerisation state of a macromolecule. X-ray small angle scattering (SAXS) on protein solutions can be employed to compute the molecular shape and known 3-D structures can be exploited for the determination of their relative arrangement in multi-domain proteins or large assemblies. For studies of large macromolecules and biomolecular complexes the use of electron microscopy (EM) is emerging. Similar to SAXS, it provides information on the size and shape of the macromolecule up to a resolution of about 10 Å and requires only very small amounts of the sample. Nuclear magnetic resonance (NMR) is well suited for solution studies of biological macromolecules, targeting the investigation of molecular interaction and molecular flexibility.

Macromolecular X-ray crystallography (MX) provides perhaps the most detailed picture of the biological molecules. MX has become the predominant technique for 3-D structure determination and is the main contributor to the rapid increase in the number of structures available from the Protein Data Bank (PDB) [9] (Fig. 1). The resolution that can be obtained in an MX experiment is typically high (Fig. 2), and allows visualisation of the macromolecular 3-D structure down to very small building blocks – the individual atoms and, sometimes, even electrons. At the same time, it provides long-range structural information. All this has opened up new opportunities for establishing structure-function relationships that convert the mere assembly of atoms in space into an appropriately folded biological molecule with a dedicated purpose and function.

The last decade has witnessed a rapidly emerging combined use of crystallography with complementary methods with an aim to provide structural information at different levels. Examples include the placement of crystal structures into SAXS envelopes of large multi-protein complexes [10], the use of EM envelopes for crystallographic structure solution by molecular replacement [11] and the combination of X-ray structure analysis with molecular dynamics simulations [12] or quantum chemical calculations [13] for highlighting the link between a structure and the underlying chemistry of protein function.

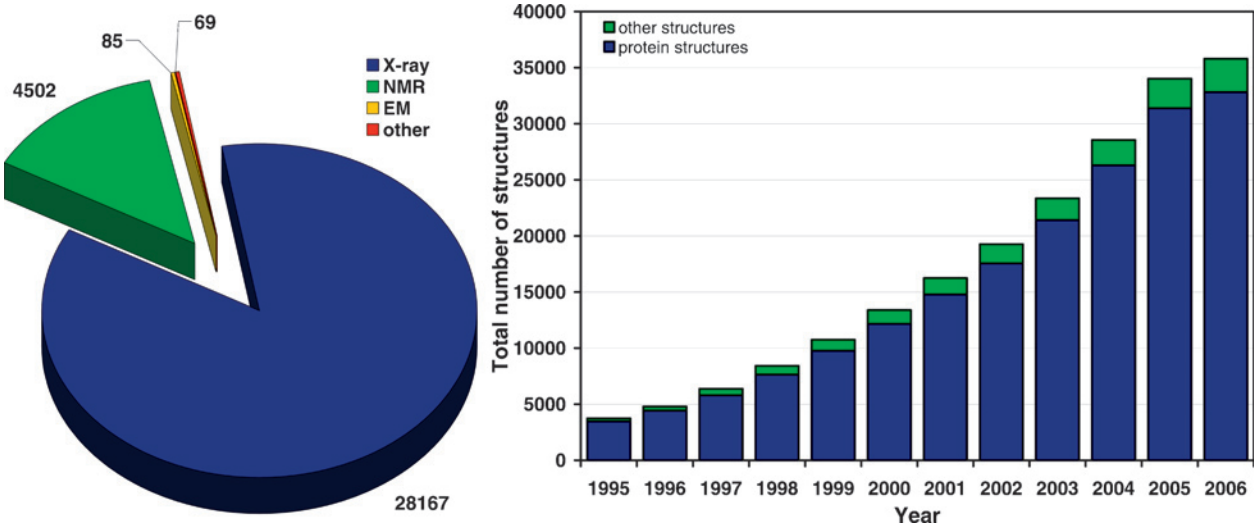

Figure 1. Statistics from the Protein Data Bank (PDB) as of 6 April 2006. (*a*) The methods used for structure determination. (*b*) The growing number of structures accumulating in the PDB.

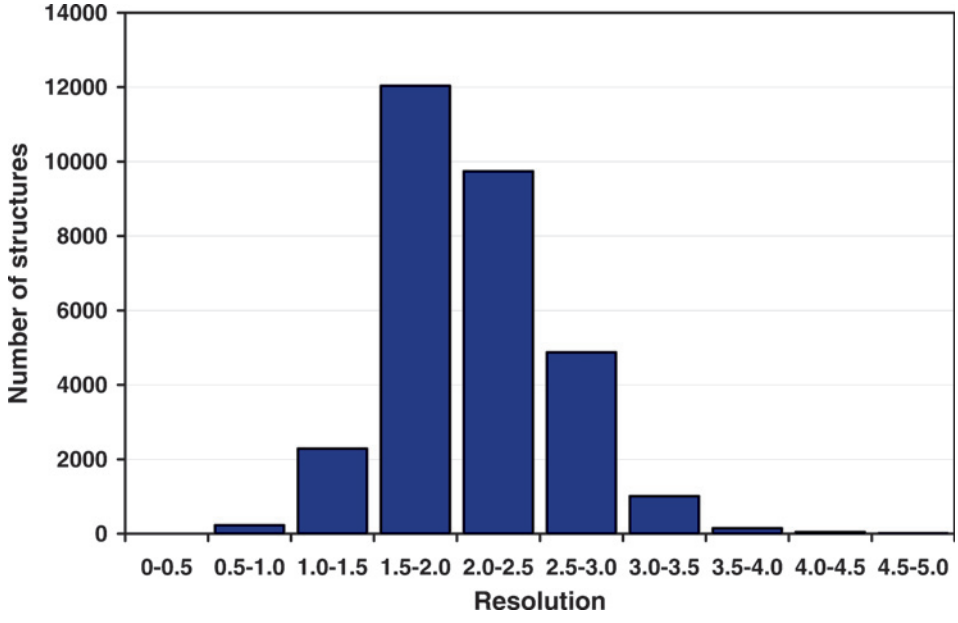

Figure 2. Histogram on crystallographic resolution to which the structures in the PDB have been determined. The majority of structures are in the 2.0 Å range.

From proteins to crystals and to crystal structures

Since MX relies on the analysis of diffraction patterns from a crystalline sample, the proteins need to be crystallised. Crystallisation can be viewed as an adjustment of the external conditions so that regular packing of the molecules to each other is favoured. Bringing flexible macromolecular entities into regular arrangement is a major challenge and still remains rather art than science. It is therefore not surprising that crystallisation of biological macromolecules has become a massive field of research. Most commonly used are vapour diffusion methods (see for example the methods outlined in [14, 15]) where a drop containing a protein solution and a precipitant is equilibrated *via* the gas phase against a higher concentrated precipitant solution. The advantages are in the large variability of the amount of sample needed (*i.e.* the drop sizes) and the ease of retrieval of the crystals for subsequent measurements. An example of a newer technique is counter diffusion [16], where a capillary with the protein solution is inserted into a gel layer, through which a gentle interaction with the precipitant solution takes place. This method gained new attention through microgravity experi-

ments. However, the crystal retrieval, *e.g.* for cryogenic cooling and data collection, may be problematic. The list of possibilities for the crystallisation set-up and the trial conditions is almost endless. Therefore, the number of available crystallisation screens and, at the same time, the degree of automation in setting up large numbers of crystallisation trials is rapidly growing [17].

Protein molecules pack loosely into the crystalline lattice and are surrounded by layers of solvent molecules from the crystallisation cocktail. The high solvent content influences the diffraction properties but also provides accessibility of the protein molecules to small molecule compounds. The presence of the solvent and high ionic strength of the crystallisation mother liquor also allows the relay of the information gathered from the crystal structures back to the *in vivo* environment [18]. The structures observed in the crystal are normally the same as the ones in solution, the exceptions being loops or amino acid side chains involved in crystal lattice contacts. The strain imposed on the molecules by the crystal lattice as a whole is almost insignificant as was demonstrated by a study on flexible sarcin-ricin domain motifs from 23S (*E. coli*) and 28S (rat) rRNAs, which showed that their crystal structures corresponded very well to the molecular dynamics-simulated structures representing energy-minimised states [19]. Also, the differences between NMR structures determined in solution (*i.e.* in absence of a crystal lattice) and X-ray crystal structures were generally within the margins of error and may be attributed to the different (mathematical) treatment of the models, *e.g.* different restraints, simplified treatment of non-bonded interactions and exclusion of explicit solvent molecules in the NMR refinement process ([20] and references therein). One example is the X-ray crystal and NMR of myoglobin structures, where the only relevant observed changes were in the conformation of two side chains in the haem pocket [21]. Efforts to make structure comparison easier are manifested in the creation of databases for NMR structures [22] The overall molecular shapes of macromolecules also remain unchanged in the crystal, as was illustrated by their comparison to low-resolution envelopes obtained by SAXS experiments in solution [23]. However, there could be exceptions, *e.g.* multi-domain proteins or proteins occurring as multimers, where the relative arrangement of the subunits is energetically poorly defined and thus may be affected by the weak crystal lattice forces [24].

Enormous efforts have been put into the development of crystallographic methodology and technology, which were to a large extent driven by biotechnological, pharmaceutical or academic interest, as well as large scale initiatives such as structural genomics [25] and industrial scale structure-based drug design [26]. We now witness an upcoming era of high-throughput crystallisation facilities and powerful synchrotron X-ray sources with automated beamlines and sample handling environment that dramatically enhance the process of 3-D structure determination [27, 28]. Complemented by the development of efficient software that minimises the user effort and spans the whole range of steps from processing of diffraction data to the validation of the macromolecular model at the end of the refinement, macromolecular crystallography is now evolving into a widely used and easy-to-apply method. The reader is referred to a comprehensive review by Dauter [29] on the past and recent developments in macromolecular crystallography.

Seeing is believing

The direct result of the crystallographic structure determination is an electron-density distribution in 3-D space from which the atomic positions, expressed as a set of 3-D coordinates, can be derived and the molecular structure modelled. Crystallographic model building hence forms a major step in the structure determination process. Nowadays automatic routines can aid building protein models into electron density [30] in largely unbiased manner.

The accuracy of the atomic positions is defined by the strength of crystal diffraction and thus by the resolution to which the X-ray data have been measured. The crystallographic resolution is approximately equal to the minimum distance between two points that can still be resolved in the electron density map (see Fig. 3 for a gallery of maps at different resolution). This in turn defines the amount of detail that can be obtained from the resulting macromolecular model. During the crystal structure refinement the protein stereochemistry is usually restrained to the target values that are ordinarily taken from the library of small molecule compounds. At a resolution of 3.0 Å or lower the amount of the observed data is very limited. The individual atoms and often even the side chains cannot be resolved in the electron density map. Consequently, the derived models are heavily influenced by predefined stereochemical restraints and the coordinate error in the atomic positions can be in the range of 0.5 Å. The resolution of 2.2 Å corresponds to about 'typical' structures available from PDB (see also Fig. 2), where chemical groups (phenyl rings, carboxylates, etc) and isolated atoms (e.g. solvent or metals) are already resolved. The coordinate error is of an order of 0.2 Å. At atomic resolution (1.2 Å or higher [31]) not only the bonded atoms are resolved and their types become distinguishable but also the atomic positions can be accurately determined to 1/100th of an Ångstrom. At ultra-high resolution (0.8 Å

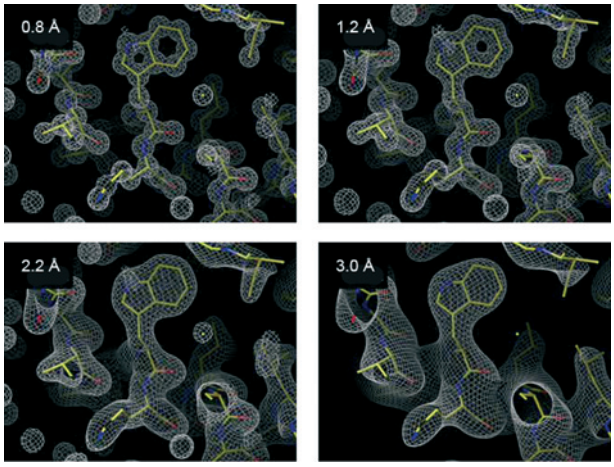

Figure 3. Examples for electron density maps at different resolution of the X-ray data with error-free crystallographic phases. Maps are contoured at the same level, 1.0 σ above the mean. The amount of detail increases impressively with the resolution. Pictures were kindly provided by James Holton (Lawrence Berkeley National Laboratory, One Cyclotron Road, MS 6R2100, Berkeley, CA 94720, USA).

or higher), the bonding electrons and hydrogen atoms can be detected and thus the chemical properties can be directly inferred from the crystal structure. The benefits of atomic and ultra-high resolution have been reviewed elsewhere [32–34] and the idiom "seeing is believing" is perfectly applicable for these structures.

Protein structure interpretation

In high-resolution protein structures, where the X-ray term dominates over stereochemical restraints, the deviations from the library values can be considerable. An analysis of squid ganglion DFPase determined to 0.85 Å shows that individual bond lengths and angles may differ significantly from the usually employed targets and even suggests the revision of those libraries for macromolecules based on data from ultra-high resolution structures [35]. The high accuracy of the model becomes most relevant for the interpretation of active sites or ligand binding studies where, at ultra-high resolution, the analysis of interatomic distances may be followed by the interpretation of chemical bonding or the charge distribution. Charge density analysis has so far almost exclusively been done on small molecule compounds, but recent advances demonstrated its applicability also to macromolecules [36]. An example is the exact characterisation of ligand-protein contacts in complexes of aldose reductase at a resolution of 0.66 Å [37], where the complementarity of the surface charges on the protein and the ligands could be directly observed. Such complementarity is necessary for strong binding interaction and the exact placement of the ligand as well as the correct spatial arrangement of the chemically active groups involved in the catalytic reaction.

From atomic structure to protein chemistry

The fine details attainable in high-resolution electron density maps are now increasingly used for extensive probing of active sites in terms of their chemical properties and charge distribution. The level of accuracy can be so high that electronic effects caused by the protein environment can be monitored. These effects can be quite far-reaching and may originate from the overall molecular architecture. Lario and Vrielink [38] observed that within α-helices the electron density of the main chain carbonyl groups is polarised toward the oxygen atoms, while in β-sheet structures the charge density is higher in between the carbon and oxygen atoms. Following this, one may be able to quantify the complex polarisation effects that influence the chemical properties of the active site residues by, for example, altering their pKa values [39–41]. Considerable effort has been put into the development of refinement software to take into account the atomic properties that go beyond the regular spherical model. The first attempts to model delicate features in the electron density for small molecules were made by Hansen and Coppens [42] and have then over the years been tuned towards application to macromolecules [36, 43]. Differences in bond order or electron lone pairs can now be detected in ultra-high resolution difference electron density maps and modelled accordingly [44]. The main requirement for the application of a non-spherical model is the sufficient "sharpness" of the atomic peaks in the electron density, which sets a current resolution limit of about 0.9 Å.

Small size with big impact – hydrogen atoms in protein structures

In structures determined to atomic resolution or higher, most hydrogen atoms can be seen directly, and the detection of the bonding electrons around the parent atom allows determination of its protonation state, the pKa value, or H-bonding interactions in general. Since most of the catalytic reactions involve hydrogen transfer of some sort, such information can be of high importance for establishing the catalytic mechanism. The large number of examples where the protonation state of residues or the involvement of exceptional H-bonding contacts in catalysis could be observed from MX structure analysis makes extensive discussion on individual proteins close to impossible. The reader is therefore referred to the specialised literature (e.g. [37, 45–49] and the references therein). We illustrate the importance of direct observation of hydrogen atoms through the studies of aldose reduc-

tase, a highly relevant drug target, which showed the full potential of the characterisation of the ligand and cofactor binding sites in terms of electron and charge distribution. In aldose reductase unusual hydrogen donors and acceptors have been identified unambiguously from the appearance of hydrogen peaks in the 0.66 Å electron density maps [50], which revealed the chemical properties of the protein as well as the bound inhibitors [37, 50] (Fig. 4). The charge distribution directly determined from the diffraction data using multipole refinement [37, 51] (Fig. 5), showed remarkable electrostatic apoenzyme–cofactor complementarity within the active site. This work pioneered the experimental analysis of the protein host–guest interaction at the electrostatic level, beyond geometrical or putative hydrogen bonding considerations.

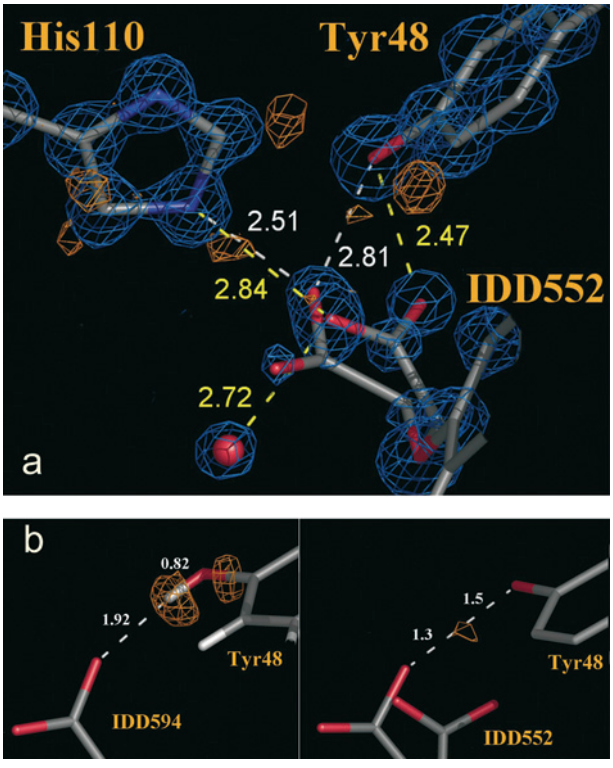

Figure 4. The active site of human aldose reductase, showing the double conformation of the inhibitor IDD552, the contact distances and the protonation state of the residues involved in binding. (*a*) In blue is the $2F_o$-F_c map, in red is the F_o-F_c map showing the hydrogen atom positions and the trace of a lower occupancy alternate conformer for Tyr 48. Both maps are contoured at the level of 2 σ above their mean. The yellow numbers and lines are distances for contacts made by the conformer of the inhibitor IDD552 with higher occupancy; the white numbers correspond to its lower-occupancy conformer. (*b*) The difference in H-bonding in the two aldose reductase inhibitor structures. In one case (left, inhibitor IDD594) the proton of Tyr 48 is clearly attached to the tyrosine OH atom. In the other case (right) it is involved in an almost symmetrical H-bond to one of the conformers of inhibitor IDD552; the H atom is delocalised along the bond and shared equally between the binding partners. The numbers show the hydrogen bonding distances [50].

Determination of atomic resolution structures of complexes that represent various stages of the catalytic reaction can help in mapping the complete reaction path and understanding the catalytic mechanism. The high accuracy of the atomic positions allows reliable interpretation of ligand-induced structural changes, even if they seem small. This was beautifully demonstrated in a study of horseradish peroxidase, where the X-ray radiation during the diffraction data collection was used to create various binding states of oxygen, hydrogen peroxide and water to the haem group, thus mapping the catalytic pathway [52]. This study also served as a neat example of the complementary application of UV/VIS spectroscopy during the crystallographic measurements.

Quantitative structure-ativity relationship

"Quantitative structure-activity relationship" (QSAR) is a technique that has emerged recently in the context of drug design ([53] and references therein) and aims at deriving the energetics of substrate binding, recognition or catalysis from biophysical data. This is achieved by correlating the chemical, geometrical and electronic properties of the ligands with kinetic data. The effects of ligand binding were studied in horseradish peroxidase, where the correlations between molecular structure, molecular orbital energies (derived by quantum chemical calculations) and the experimental reaction rates were used to simulate the optimal enzyme-substrate conformations [54]. It was shown that proteomic modifications of the enzyme, designed to decrease the H-bonding distance between the substrates of concern and the catalytic histidine, may be useful in enhancing the reactivity towards recalcitrant phenolic contaminants and improving degradation efficiencies in engineering applications.

In another case the crystal structures were used to verify the simulated models of active conformations of HIV-1 protease [55]. This 4-D QSAR approach yielded not only the most favoured conformational states, but also showed the importance of the H-bonding patterns between enzyme and ligand and allowed characterisation of the active conformation of the inhibitors. Following this, a use of highly accurate crystal structures would make it possible to relate the extent of weakening of hydrogen bonds to the electronic and steric properties of ligands. If those ligands are lead compounds, these results can be correlated to the known pharmacological data.

Structure-activity relationship is currently based on tabulated atomic properties that are used in docking or other computational approaches for surveying the structure-target-ligand interactions. These atomic properties can change substantially as a function of the specific environment. Therefore, the availability

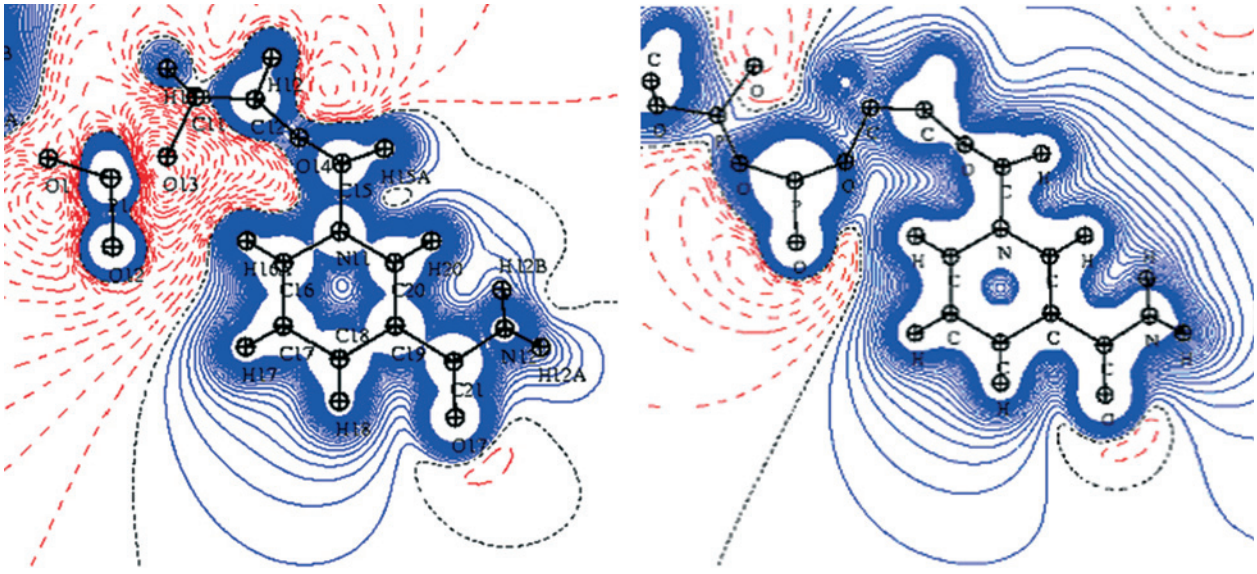

Figure 5. Experimentally determined (left) and theoretical electrostatic potential (right) on the $NADP^+$ cofactor in the active site of human aldose reductase. Positive electrostatic charge is contoured in blue, negative charge in red. Hydrogen atoms and lone electron pairs as well as the surface charge distribution on the nicotinamide group of $NADP^+$ are shown [51].

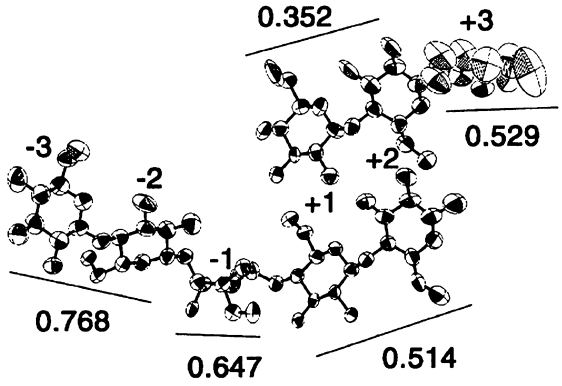

Figure 6. TELP (thermal ellipsoids) representation of the oligosaccharide ligand in CelA, showing the variations in mobility of the sugar rings dependent on their respective binding sites. Larger ellipsoids indicate higher mobility (larger displacement), which is most pronounced for subsite +3. The sugar subsites from –3 to +3 were grouped as indicated by thin lines and the numbers display refined group occupancies [61].

of highly accurate crystal structures, where the atomic (electronic) properties are directly available, would permit rational design of interactions with inhibitors. In addition to the atomic coordinates and the 3-D structural properties, the energetics of ligand binding can then be taken into account for structure interpretation. This method was tried on aldose reductase [37] (and literature cited above), for which complexes with inhibitors and cofactor were available to ultra-high resolution.

Static and dynamic

A number of approaches have been employed to go beyond a static, time and space averaged image of the structures in the crystallographic analysis. Some involved time-resolved measurements on a series of crystals, or applying multi-wavelength techniques such as Laue diffraction for fast data collection, with an aim to catch glimpses of transient chemical states in the crystal. However, the generally fast rate of chemical turnover may pose a problem. For an extensive overview on the past and recent developments in this area see [56] and the references therein. Other methods focus on the inherent molecular motion and aim at the investigation of what can be called "dynamic properties". Even though a protein molecule is embedded in a crystal lattice, it is still surrounded by a great deal of solvent and the contacts between the molecules are rather loose. Therefore one can expect that traces of the intrinsic mobility are retained even though large-scale motion may be impeded by the crystal lattice [18].

Some time ago Dunitz et al. [57] made a comment referring to the crystallographic analysis of small molecule compounds: "crystals should not be seen as a chemical cemetery". Recent developments have demonstrated that also for macromolecules it is feasible to extract information that goes beyond the mere atomic positions. Displacement of atoms around their equilibrium position has been routinely modelled by an atomic displacement parameter, ADP, (the so-called temperature factor) that describes a probability density function for the location of a given atom

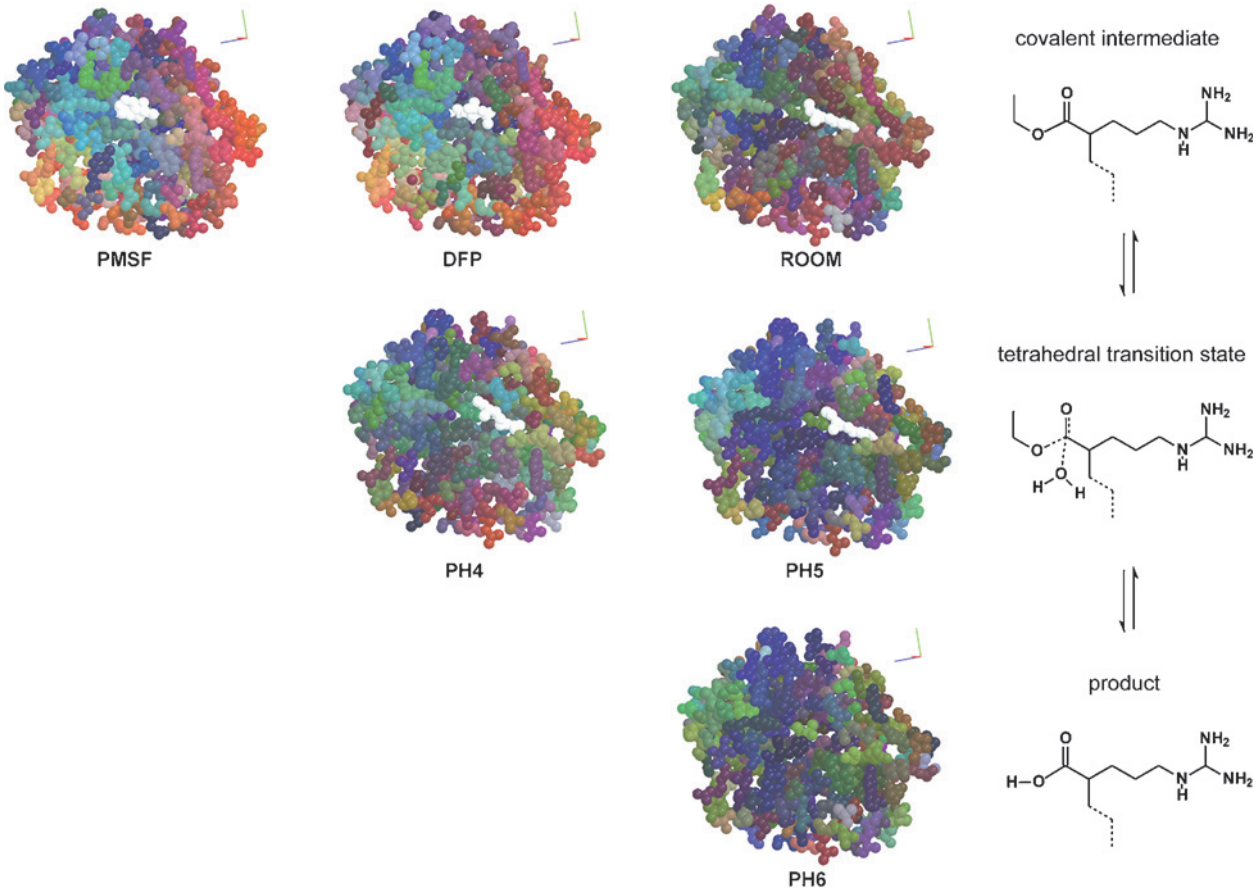

Figure 7. CPK (1 Å spheres) representation of a series of trypsin structures in various ligand-binding states that map the reaction pathway from covalent intermediate to product. The structures are named after their crystallisation conditions (PH4, PH5 and PH6) or the added covalently binding inhibitors phenylmethylsulphonyl fluoride (PMSF) or di-isopropyl fluorophosphate (DFP). The ROOM structure has been determined at room temperature, while all others at 100 K. The colours represent the vectors of principal direction of motion determined from the anisotropic atomic displacement parameters where the vector components X,Y,Z have been translated into RGB colour code. The change in colours and hence the pattern of directional motion in the molecule is dependent of the ligand's nature and its binding state [62].

[58]. At low resolution, modelling of this displacement is restricted to the spherical shape (isotropic). At atomic resolution (1.2 Å or higher), the high data to parameter ratio allows extension of this model to a 3-D ellipsoid (anisotropic), which is defined by the lengths, and the directions of its three principal axes. Such anisotropic ADPs carry valuable information about directional motion in a macromolecule. Indeed, any enzymatic reaction requires motion of some sort. Substrates need to be admitted and bound, cofactors need to be brought to contact and often the active site needs to be shielded against the adverse effects of the surrounding (aqueous) solvent. Furthermore, the protein needs to adjust itself to accommodate the changing stereochemistry along the reaction path, which is also a prerequisite for the achievement of the high turnover rate [59, 60].

The inclusion of anisotropic ADPs into the high-resolution macromolecular model [58] can elucidate molecular properties that hint at the crystal structures of biological molecules not being as static and rigid as they had been thought to be. One example is cellulase A from *Clostridium thermocellum* where the flexibility of the bound ligand could be visualised in a TELP representation [61] (Fig. 6). This, in addition to the analysis of the bond distances, allowed better characterisation of the binding sites and visualisation of considerable torsional strain in the sugar moieties of the cellopentaose ligand, in particular in the region close to the cleavage site. The distortion may help bring the sugar subunits in binding sites +1 and -1 into ideal contacts with the catalytically active residues for cleavage of the glycosidic bond, while the mobility of the sugar moieties, as seen from the TELP picture, may be used for estimating the strength of their binding contacts.

At atomic resolution, the direct availability of individual anisotropic ADPs permits straightforward

extraction of directional motion from the principal components of the thermal ellipsoid. This may allow monitoring both subtle changes as well as larger scale motion. On a series of trypsin structures that represented various substrate and inhibitor binding states, a noticeable change of direction of motion was detected that was dependent on the binding state [62]. This change of directionality concurred with the observed structural alterations such as the change of width and length of the specificity pocket, as well as the contact distances between active site residues and ligands (Fig. 7).

At 'typical' resolution, around 2 Å, the individual anisotropic ADPs cannot yet be derived and motion within a macromolecule can to an extent be modelled by the so-called translation-libration-screw (TLS) refinement [63], describing the rigid-body motion of either regions in the molecule or the molecule as a whole. Using a TLS approach Harata and coworkers characterised the molecular motion of lactalbumin [64], lysozyme [65] and recently xylanase II [66]. They attempted to decouple the TLS and the atomic contribution to the overall atomic motion in the molecule and to characterise the rigid body motions of protein fragments in the structure. Also in their most recent study they monitored the influence of lattice contacts. Another tool, a vibrational (normal mode) analysis [67] has been successfully tried on a number of structures to extract characteristic domain motion. This method can be used to characterise (theoretically) structural subunits in the protein and may also indicate internal enzymatic cooperativity. For example, in the case of *S*-adenosyl-L-homocysteine hydrolase normal mode calculations show strong coupling of the hinge-bending motions of the individual subunits to each other and to other low-frequency vibrations [68]. The use of the low-frequency vibrational modes can also be promising for facilitating the structure solution by molecular replacement [69] for cases where variations in domain arrangement can be expected and ensembles of the search model are needed.

Outlook

With low coordinate error and availability of data on thermal motion from atomic resolution structures, it now becomes possible to catch the snapshots of enzymatic reactions. This bridges the gap between the 3-D atomic structure and the underlying chemistry that governs the biological function. With the upcoming third generation synchrotron sources, novel data collection and structure interpretation methods we can foresee that high-accuracy atomic models will become of increasing importance in the near future and that the already ongoing developments of software tools will make the wealth of detail present in such structures easily accessible.

1 White, S. W., Zheng, J., Zhang, Y. M. and Rock (2005) The structural biology of type II fatty acid biosynthesis. Annu. Rev. Biochem. 74, 791–831.
2 Owen, O. E., Kalhan, S. C. and Hanson, R. W. (2002) The key role of anaplerosis and cataplerosis for citric acid cycle function. J. Biol. Chem. 277, 30409–30412.
3 Voet, D. and Voet, J. D. (2005) Biochemistry, 3rd edn, Wiley, New York.
4 Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G. and Schomburg, D. (2004) BRENDA, the enzyme database: updates and major new developments. Nucleic Acids Res. 32, D431–433.
5 Moore, P. B. and Steitz, T. A. (2005) The ribosome revealed. Trends Biochem. Sci. 30, 281–283.
6 Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, 95–99.
7 Abseher, R. and Nilges, M. (2002) Protein folding in mode space: a collective coordinate approach to structure prediction. Proteins 49, 365–377.
8 Fersht, A. R. and Daggett, V. (2002) Protein folding and unfolding at atomic resolution. Cell 108, 573–582.
9 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.
10 Petoukhov, M. V. and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J. 89, 1237–1250.
11 Tama, F., Miyashita, O. and Brooks, C. L., 3rd (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J. Struct. Biol. 147, 315–326.
12 Castillo, R., Silla, E. and Tunon, I. (2002) Role of protein flexibility in enzymatic catalysis: quantum mechanical- molecular mechanical study of the deacylation reaction in class A β-lactamases. J. Am. Chem. Soc. 124, 1809–1816.
13 Heikinheimo, P., Tuominen, V., Ahonen, A.-K., Teplyakov, A., Cooperman, B. S., Baykov, A. A., Lahti, R. and Goldman, A. (2001) Toward a quantum-mechanical description of metal-assisted phosphoryl transfer in pyrophosphatase. Proc. Natl. Acad. Sci. USA 98, 3121–3126.
14 Ducruix, A. (1999) Crystallization of Nucleic Acids and Proteins: A Practical Approach (Practical Approach Series). Ducruix, A. and Giege, R. (eds.) Oxford University Press, Oxford.
15 Drenth, J. (1999) Principles of Protein X-ray Crystallography. Springer-Verlag, Heidelberg.
16 Tanaka, H., Inaka, K., Sugiyama, S., Takahashi, S., Sano, S., Sato, M. and Yoshitomi, S. (2004) A simplified counter diffusion method combined with a 1D simulation program for optimizing crystallization conditions. J. Synchrotron Radiat. 11, 45–48.
17 Hiraki, M., Kato, R., Nagai, M., Satoh, T., Hirano, S., Ihara, K., Kudo, N., Nagae, M., Kobayashi, M., Inoue, M., Uejima, T., Oda, S., Chavas, L. M., Akutsu, M., Yamada, Y., Kawasaki, M., Matsugaki, N., Igarashi, N., Suzuki, M. and Wakatsuki, S. (2006) Development of an automated large-scale protein-crystallization and monitoring system for high-throughput protein-structure analyses. Acta Crystallogr. D Biol. Crystallogr. 62, 1058–1065.
18 Wilson, M. A. and Brunger, A. T. (2000) The 1.0 Å crystal structure of Ca(2+)-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity. J. Mol. Biol. 301, 1237–1256.

19 Spackova, N. and Sponer, J. (2006) Molecular dynamics simulations of sarcin-ricin rRNA motif. Nucleic Acids Res. 34, 697–708.
20 Garbuzynskiy, S. O., Melnik, B. S., Lobanov, M. Y., Finkelstein, A. V. and Galzitskaya, O. V. (2005) Comparison of X-ray and NMR structures: is there a systematic difference in residue contacts between X-ray- and NMR-resolved protein structures? Proteins 60, 139–147.
21 Nguyen, B. D., Zhao, X., Vyas, K., La Mar, G. N., Lile, R. A., Brucker, E. A., Phillips, G. N., Jr., Olson, J. S. and Wittenberg, J. B. (1998) Solution and crystal structures of a sperm whale myoglobin triple mutant that mimics the sulfide-binding hemoglobin from *Lucina pectinata*. J. Biol. Chem. 273, 9517–9526.
22 Nabuurs, S. B., Nederveen, A. J., Vranken, W., Doreleijers, J. F., Bonvin, A. M., Vuister, G. W., Vriend, G. and Spronk, C. A. (2004) DRESS: a Database of refined solution NMR structures. Proteins 55, 483–486.
23 Kozak, M. (2005) Direct comparison of the crystal and solution structure of glucose/xylose isomerase from *Streptomyces rubiginosus*. Protein Pept. Lett. 12, 547–550.
24 Kozak, M. and Jurga, S. (2002) A comparison between the crystal and solution structures of *Escherichia coli* asparaginase II. Acta Biochim. Pol. 49, 509–513.
25 Zhang, C. and Kim, S. H. (2003) Overview of structural genomics: from structure to function. Curr. Opin. Chem. Biol. 7, 28–32.
26 Kuhn, P., Wilson, K., Patch, M. G. and Stevens, R. C. (2002) The genesis of high-throughput structure-based drug discovery using protein crystallography. Curr. Opin. Chem. Biol. 6, 704–710.
27 Arzt, S., Beteva, A., Cipriani, F., Delageniere, S., Felisaz, F., Forstner, G., Gordon, E., Launer, L., Lavault, B., Leonard, G., Mairs, T., McCarthy, A., McCarthy, J., McSweeney, S., Meyer, J., Mitchell, E., Monaco, S., Nurizzo, D., Ravelli, R., Rey, V., Shepard, W., Spruce, D., Svensson, O. and Theveneau, P. (2005) Automation of macromolecular crystallography beamlines. Prog. Biophys. Mol. Biol. 89, 124–152.
28 Jiang, J. and Sweet, R. M. (2004) Protein Data Bank depositions from synchrotron sources. J. Synchrotron Radiat. 11, 319–327.
29 Dauter, Z. (2006) Current state and prospects of macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 62, 1–11.
30 Badger, J. (2003) An evaluation of automated model-building procedures for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. 59, 823–827.
31 Morris, R. J. and Bricogne, G. (2003) Sheldrick's 1.2 Å rule and beyond. Acta Cryst. 59, 615–617.
32 Dauter, Z., Lamzin, V. S. and Wilson, K. S. (1997) The benefits of atomic resolution. Curr. Opin. Struct. Biol. 7, 681–688.
33 Longhi, S., Czjzek, M. and Cambillau, C. (1998) Messages from ultrahigh resolution crystal structures. Curr. Opin. Struct. Biol. 8, 730–737.
34 Schmidt, A. and Lamzin, V. S. (2002) Veni, vidi, vici: atomic resolution unravelling the mysteries of protein function. Curr. Opin. Struct. Biol. 12, 698–703.
35 Koepke, J., Scharff, E. I., Lucke, C., Ruterjans, H. and Fritzsch, G. (2003) Statistical analysis of crystallographic data obtained from squid ganglion DFPase at 0. 85 Å resolution. Acta Crystallogr. D Biol. Crystallogr. 59, 1744–1754.
36 Guillot, B., Viry, L., Guillot, R., Lecomte, C. and Jelsch, C. (2001) Refinement of proteins at subatomic resolution with MOPRO. J. Appl. Cryst. 34, 214–223.
37 Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. and Podjarny, A. (2004) Ultrahigh resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 Å. Proteins 55, 792–804.
38 Lario, P. I. and Vrielink, A. (2003) Atomic resolution density maps reveal secondary structure dependent differences in electronic distribution. J. Am. Chem. Soc. 125, 12787–12794.
39 Merkel, J. S., Sturtevant, J. M. and Regan, L. (1999) Sidechain interactions in parallel beta-sheets: the energetics of cross-strand pairings. Structure 7, 1333–1343.
40 Yang, A. S., Gunner, M. R., Sampogna, R., Sharp, K. and Honig, B. (1993) On the calculation of pKas in Proteins. Proteins 15, 252–265.
41 Davoodi, J., Wakarchuk, W. W., Campbell, R. L., Carey, P. R. and Surewicz, W. K. (1995) Abnormally high pKa of an active-site glutamic acid residue in *Bacillus circulans* xylanase: The role of electrostatic interactions. Eur. J. Biochem. 232, 839–843.
42 Hansen, N. K. and Coppens, P. (1978) Testing aspherical atom refinements on small molecule data sets. Acta Cryst. A34, 909–921.
43 Afonine, P., Pichon-Pesme, V., Muzet, N., Jelsch, C., Lecomte, C. and Urzhumtsev, A. (2002) Modelling of bond electron density by Gaussian scatterers at subatomic resolution. CCP4 Newsletter 41.
44 Afonine, P. V., Lunin, V. Y., Muzet, N. and Urzhumtsev, A. (2004) On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps. Acta Crystallogr. D Biol. Crystallogr. 60, 260–274.
45 Varrot, A. and Davies, G. J. (2003) Direct experimental observation of the hydrogen-bonding network of a glycosidase along its reaction coordinate revealed by atomic resolution analyses of endoglucanase Cel5A. Acta Crystallogr. D Biol. Crystallogr. 59, 447–52.
46 Stranzl, G. R., Gruber, K., Steinkellner, G., Zangger, K., Schwab, H. and Kratky, C. (2004) Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large pKa shift of the catalytic base induced by the reaction intermediate. J. Biol. Chem. 279, 3699–3707.
47 Jogl, G., Rozovsky, S., McDermott, A. E. and Tong, L. (2003) Optimal alignment for enzymatic proton transfer: structure of the Michaelis complex of triosephosphate isomerase at 1.2-Å resolution. Proc. Natl. Acad. Sci. USA 100, 50–55.
48 Minasov, G., Wang, X. and Shoichet, B. K. (2002) An ultrahigh resolution structure of TEM-1 beta-lactamase suggests a role for Glu166 as the general base in acylation. J. Am. Chem. Soc. 124, 5333–5340.
49 Stirnimann, C. U., Rozhkova, A., Grauschopf, U., Bockmann, R. A., Glockshuber, R., Capitani, G. and Grutter, M. G. (2006) High-resolution structures of *Escherichia coli* cDsbD in different redox states: A combined crystallographic, biochemical and computational study. J. Mol. Biol. 358, 829–845.
50 Ruiz, F., Hazemann, I., Mitschler, A., Joachimiak, A., Schneider, T., Karplus, M. and Podjarny, A. (2004) The crystallographic structure of the aldose reductase-IDD552 complex shows direct proton donation from tyrosine 48. Acta Crystallogr. D Biol. Crystallogr. 60, 1347–1354.
51 Muzet, N., Guillot, B., Jelsch, C., Howard, E. and Lecomte, C. (2003) Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations. Proc. Natl. Acad. Sci. USA 100, 8742–8747.
52 Berglund, G. I., Carlsson, G. H., Smith, A. T., Szoke, H., Henriksen, A. and Hajdu, J. (2002) The catalytic pathway of horseradish peroxidase at high resolution. Nature 417, 463–468.
53 Cachau, R. E. and Podjarny, A. D. (2005) High-resolution crystallography and drug design. J. Mol. Recognit. 18, 196–202.
54 Colosi, L. M., Huang, Q. and Weber, W. J., Jr. (2006) Quantitative structure-activity relationship based quantification of the impacts of enzyme-substrate binding on rates of peroxidase-mediated reactions of estrogenic phenolic chemicals. J. Am. Chem. Soc. 128, 4041–4047.
55 Santos-Filho, O. A. and Hopfinger, A. J. (2006) Structure-based QSAR analysis of a set of 4-hydroxy-5,6-dihydropyrones as inhibitors of HIV-1 protease: an application of the receptor-

dependent (RD) 4D-QSAR formalism. J. Chem. Inf. Model. 46, 345–354.
56 Schmidt, M., Ihee, H., Pahl, R. and Srajer, V. (2005) Protein-ligand interaction probed by time-resolved crystallography. Methods Mol. Biol. 305, 115–154.
57 Dunitz, J. D., Shomaker, V. and Trueblood, K. N. (1988) Interpretation of atomic displacement parameters from diffraction studies of crystals. J. Phys. Chem. 92, 856–867.
58 Merritt, E. A. (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement. Acta Crystallogr. D Biol. Crystallogr. 55, 1109–1117.
59 Benkovic, S. J. and Hammes-Schiffer, S. (2003) A perspective on enzyme catalysis. Science 301, 1196–1202.
60 Agarwal, P. K., Billeter, S. R., Rajagopalan, P. T. R., Benkovic, S. J. and Hammes-Schiffer, S. (2002) Network of coupled promoting motions in enzyme catalysis. Proc. Natl. Acad. Sci. USA 99, 2794–2799.
61 Guerin, D. M., Lascombe, M.-B., Costabel, M., Souchon, H., Lamzin, V. S., Beguin, P. and Alzari, P. M. (2002) Atomic (0.94 Å) resolution structure of an inverting glycosidase in complex with substrate. J. Mol. Biol. 316, 1061–1069.
62 Schmidt, A. and Lamzin, V. S. (2005) Extraction of functional motion in trypsin crystal structures. Acta Crystallogr. D Biol. Crystallogr. 61, 1132–1139.
63 Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. 57, 122–133.
64 Harata, K., Abe, Y. and Muraki, M. (1999) Crystallographic evaluation of internal motion of human alpha-lactalbumin refined by full-matrix least-squares method. J. Mol. Biol. 287, 347–358.
65 Harata, K. and Kanai, R. (2002) Crystallographic dissection of the thermal motion of protein-sugar complex. Proteins 48, 53–62.
66 Watanabe, N., Akiba, T., Kanai, R. and Harata, K. (2006) Structure of an orthorhombic form of xylanase II from *Trichoderma reesei* and analysis of thermal displacement. Acta Crystallogr. D Biol. Crystallogr. 62, 784–792.
67 Kidera, A. and Go, N. (1992) Normal mode refinement: crystallographic refinement of protein dynamic structure. I. Theory and test by simulated diffraction data. J. Mol. Biol. 225, 457–475.
68 Wang, M., Borchardt, R. T., Schowen, R. L. and Kuczera, K. (2005) Domain motions and the open-to-closed conformational transition of an enzyme: a normal mode analysis of *S*-adenosyl-L-homocysteine hydrolase. Biochemistry 44, 7228–7239.
69 Suhre, K. and Sanejouand, Y. H. (2004) On the potential of normal-mode analysis for solving difficult molecular-replacement problems. Acta Crystallogr. D Biol. Crystallogr. 60, 796–799.