

Crystal structures of Aldose Reductase, C2A domain of Rabphilin3A and tests of new restraints

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaften Fakultäten
der Georg-August Universität zu Göttingen

vorgelegt von
Marianna Biadene
aus Perugia, Italien

Göttingen, 2006

D7

Referent: Prof. G. M. Sheldrick

Korreferent: Jr. Prof. O. Einsle

Tag der mündlichen Prüfung: 04.07.06

CONTENTS

1. Summary	1
2. Theoretical background	3
2.1. X-ray crystallography	3
2.2. The phase problem	5
2.2.1. Isomorphous replacement	5
2.2.2. Anomalous dispersion	7
2.2.3. Molecular replacement	9
2.2.4. Direct methods	10
2.3. Density modification	12
2.4. Refinement	14
3. Aldose Reductase	16
3.1. Introduction	16
3.1.1. Aldose Reductase	16
3.1.2. High resolution refinement	20
3.2. Materials and methods	22
3.2.1. Crystallization and data collection	22
3.2.2. Structure determination	24
3.2.3. Refinement	24
3.2.3.1. Refinement of disordered side chains	26
3.2.3.2. Refinement of waters	29
3.2.3.3. Hydrogens	30
3.2.3.4. Disagreeable restraints	31
3.3. Results and discussion	32
3.3.1. Aspects of high resolution refinement	32
3.3.1.1. Disordered regions	32

3.3.1.2.	Waters	37
3.3.1.3.	Hydrogens	38
3.3.1.4.	Analysis of disagreeable restraints	40
3.3.1.5.	Quality of the model	42
3.3.2.	Structure	46
3.3.2.1.	Overall structure	46
3.3.2.2.	Active site	47
3.3.2.3.	Safety-belt loop	49
3.3.2.4.	The tongs movement	54
3.4.	Conclusions and perspectives	57
4.	C2A domain of Rabphilin3A	59
4.1.	Introduction	59
4.1.1.	The synaptic vesicle cycle	59
4.1.2.	Rabphilin3A	61
4.1.3.	C2 domains	63
4.2.	Materials and methods	65
4.2.1.	Crystallization and data collection	65
4.2.2.	Structure determination and refinement	66
4.3.	Results and discussion	69
4.3.1.	Quality of the model	69
4.3.2.	Structure description	71
4.3.3.	Comparison with other C2 domains	75
4.3.4.	DED motif	78
4.3.5.	Comparison of electrostatic surfaces	80
4.4.	Conclusions and perspectives	83
5.	Restraints	84
5.1.	Introduction	84
5.1.1.	Atomic displacement parameters	85

5.1.2. TLS	86
5.1.3. New restraints in <i>SHELXL</i>	88
5.2. Materials and methods	90
5.3. Results and discussion	92
5.3.1. Comparison between TLS in <i>REFMAC</i> and new <i>SHELXL</i>	92
5.3.2. Comparison between old and new <i>SHELXL</i>	93
5.4. Conclusions and perspectives	97
 Appendix A	 98
Appendix B	104
Appendix C	107
 References	 110

LIST OF ABBREVIATIONS

ADP	Atomic Displacement Parameter
ADPRP	2'-monophosphoadenosine-5'-diphosphoribose
APS	Advanced Photon Source
AR	Aldose Reductase
CBL	Calcium Binding Loop
CBR	Calcium Binding Region
CC	Correlation Coefficient
CCD	Charge Coupled Device
CGLS	Conjugate Gradient Least Squares
CSD	Cambridge Structures Database
DDT	Dichlorodiphenyltrichloroethane
DOC2	Double C2-like domain-containing proteins
ESD	Estimated Standard Deviation
GTP	Guanosine Triphosphate
HEPES	N-2-hydroxyethylpiperazine-N'-2'-ethanesulfonic acid
kDa	kiloDalton
LS	Least Squares
MAD	Multiple Wavelength Anomalous Diffraction
MIR	Multiple Isomorphous Replacement
MR	Molecular Replacement
NADPH	Dihydro-nicotinamide-adeninde-dinucleotide-phosphate
NCS	Non-Crystallographic Symmetry
NMR	Nuclear Magnetic Resonance
NPD	Non Positive Definite
PEG	Polyethylenglycole
PIP2	Phosphatidylinositol-4,5-bisphosphate

PKC	Protein Kinase C
R.M.S.D.	Root Mean Standard Deviation
SAD	Single Wavelength Anomalous Diffraction
SIR	Single Isomorphous Replacement
SM	Sec1/Munc18-like protein
SNAP	Synaptosome-Associated Protein of 25 kDa
SNARE	Soluble <i>N</i> -ethylmaleimide-Sensitive Factor
	Attachment Protein Receptor
SV2	Synaptic vesicle proteins 2a
TLS	Translation, Libration, Screw motion
2D	Two Dimensional

THREE LETTERS CODE FOR AMINO ACIDS

Ala	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
Cys	Cysteine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
Ile	Isoleucine
Leu	Leucine
Lys	Lysine
Met	Methionine
Phe	Phenylalanine
Pro	Proline
Ser	Serine
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

MOLECULAR GRAPHICS

Pictures containing molecular graphics were created with the programs *BOBSCRIPT* (Esnouf 1997), *PYMOL* (DeLano 2002) and *COOT* (Emsley *et al.* 2004). Some of the pictures were rendered with the program *RASTER3D* (Merritt *et al.* 1997).

Fig. 3.20 and 4.8 were made using the *PARVATI* webserver (<http://www.bmsc.washington.edu/parvati>; (Merritt 1999)).

Fig. 4.19B and 4.20 were created with the program *DELPHI* (Honig *et al.* 1995) and *PYMOL*.

Fig. App_A1 was made with the Bruker program *XP*.

1 Summary

The high resolution (0.82 Å) structure of Aldose Reductase (an enzyme implicated in the polyol pathway) is presented. The aim of the work was a careful inspection of the model with particular attention to features visible at high resolution. Atomic resolution data allowed to model more multiple conformations than in the previously determined structure and to improve the fit between model and data. Particular attention was paid to the examination of the list of disagreeable restraints: some of them are probably due to wrong restraints while others are real features of the structure observable only at high resolution. These observations should be used both to build a new library of restraints and to explain in more details the molecule structure and function. The most interesting feature of this model is the safety-belt loop modeled in two discrete conformations: the common closed and a partially open conformation (possible because the holo-enzyme is in a post-reaction state). The removal of one interaction is enough to trigger the release of the cofactor and to let the safety-belt loop open. Accurate inspection of various Aldose Reductase high resolution structures reveals a domain movement that has been called "tongs movement": loops at the top and at the bottom are closer to or farther away from each other while the β -barrel is rigid. This movement could be related to the post reaction state of the molecule and to the release of the cofactor.

The crystal structure of the Ca^{2+} -free form of the C2A domain of Rabphilin3A (a neuronal protein involved in the synaptic vesicle cycle for which the exact biological role still remains to be elucidated) to a resolution of 1.92 Å is presented. It has a type I topology and the structure is characterized by an eight-stranded antiparallel β -sandwich. Its electrostatic surface is highly basic and similar to the one of the C2B domain of Rabphilin3A. In the C2A domain of Rabphilin3A two specific structural features related to deviations from the *consensus* sequence were found: a different position of the calcium binding loop 1 (CBL1) that causes a significant shift of the conserved Ca^{2+} binding residue Asp413, and the presence of the residue Glu475 (DED motif) that creates a

localized negative patch close to the calcium binding region. The structure presents new features with respect to other C2 domains suggesting a new function for this domain.

In the last chapter tests of a new version of *SHELXL* (Sheldrick *et al.* 1997), in which a restraint applied to anisotropic displacement parameters was implemented (DELU restraint) are presented. The new restraint mimics a TLS refinement and it should be useful at medium resolution (1.5-2.0 Å). Comparisons between structures refined at various resolutions both with TLS refinement in *REFMAC* and with two versions of *SHELXL* were carried out. The results show that at medium resolution the two programs are comparable. The new restraint is, however a large improvement compared to the old version of *SHELXL*.

2 Theoretical background

2.1 X-ray crystallography

The aim of X-ray crystallography is to resolve the three dimensional structure of molecules. Because atoms are separated by distances in the range of Ångstroms it is mandatory to use X-rays which have wavelengths of 0.1-100 Å.

X-rays can be described as an electromagnetic wave that interacts with the electron cloud that surrounds the atoms. When electrons get hit by X-rays they emit radiation that has the same frequency of the incident beam (elastic scattering) or a different one (inelastic scattering).

The atomic scattering factor of a single atom can be described by the equation:

$$\bar{f}_a = \int_r \rho_a(\bar{r}) \exp[2\pi i \bar{r} \cdot \bar{S}] d\bar{r} \quad 2.1$$

where $\rho_a(\bar{r})$ is the electron density (approximated to be spherical) of an atom at position \bar{r} and \bar{S} is the scattering vector. The atomic scattering factor is function of λ (wavelength of the incident beam) and θ (half of the diffraction angle) because it depends on the length of the scattering vector \bar{S} that is equal to:

$$|\bar{S}| = \frac{2 \sin \theta}{\lambda} \quad 2.2$$

Because the result of an X-ray experiment is an average in space and time the atoms can be described by a mean position and the displacement around it depending on the temperature:

$$T = \exp \left[-B \frac{\sin^2 \theta}{\lambda^2} \right] \quad 2.3$$

where $B=8\pi^2\langle u^2 \rangle$ is the atomic displacement parameter (isotropic or anisotropic) and u is the square mean displacement of the atom. Consequently the atomic scattering factor can be rewritten as:

$$\bar{f}_a' = \bar{f}_a \exp \left[-B \frac{\sin^2 \theta}{\lambda^2} \right] \quad 2.4$$

The radiation emitted by the whole crystal is the summation of waves each scattered by one electron. Positive interference takes place when the phase shift of the waves is an integer multiple of 2π and this phenomenon is described by Bragg's law (Fig. 2.1):

$$2d \sin \theta = n\lambda \quad 2.5$$

The structure factor for the crystal is then the summation of the atomic scattering factors over all the j atoms of the unit cell:

$$\bar{F}(\bar{S}) = \sum_j \bar{f}_j \exp[2\pi i \bar{r}_j \cdot \bar{S}] \quad 2.6$$

or considering the electron cloud around the atoms:

$$\bar{F}(\bar{S}) = \int_{\text{cell}} \rho(\bar{r}) \exp[2\pi i \bar{r} \cdot \bar{S}] dV \quad 2.7$$

The structure factor \bar{F} can also be expressed in terms of fractional coordinates (x, y, z) and Bragg indices (h, k, l), which describe the lattice planes that cause the diffraction:

$$\bar{F}(hkl) = V \int_{x=0}^1 \int_{y=0}^1 \int_{z=0}^1 \rho(xyz) \exp[2\pi i(hx + ky + lz)] dx dy dz \quad 2.8$$

The structure factor is the Fourier transform of the electron density and *viceversa*. Therefore the electron density can be written as:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l \bar{F}(hkl) \exp[-2\pi i(hx + ky + lz)] \quad 2.9$$

The structure factor \bar{F} is a vector that can be expressed also as:

$$\bar{F}(hkl) = |\bar{F}(hkl)| \exp[i\alpha(hkl)] \quad 2.10$$

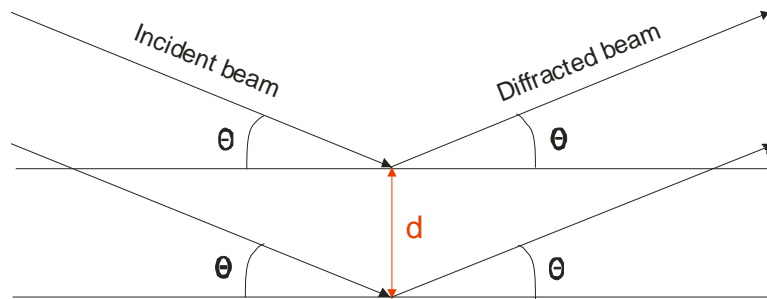


Figure 2.1 The two lattice planes are separated by the distance d . The X-rays are reflected by the planes with an angle Θ .

and the electron density as:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |\bar{F}(hkl)| \exp[-2\pi i(hx + ky + lz) + i\alpha(hkl)] \quad 2.11$$

2.2 The phase problem

In an X-ray experiment it is only possible to measure spot intensities that are proportional to the square of the amplitudes of the structure factors:

$$I(hkl) \propto |\bar{F}(hkl)|^2 \quad 2.12$$

while the phases that contain the most important information can be obtained only indirectly: this is the so called *phase problem*.

In macromolecular crystallography phases can be estimated by different approaches like:

- a) Isomorphous replacement (MIR)
- b) Anomalous diffraction (MAD, SAD)
- c) Molecular replacement (MR)
- d) Direct methods

2.2.1 Isomorphous replacement

With the isomorphous replacement method the presence of heavy atoms in crystals is used to get initial phases. Heavy atoms are introduced in the crystals by soaking them in solution containing the heavy atoms or by cocrystallization. The only difference between native and derivative crystals should then be the presence of the heavy atoms (hence the name isomorphous replacement). Therefore the structure factor of the derivatives can be written as:

$$\bar{F}_{PH} = \bar{F}_H + \bar{F}_P \quad 2.13$$

where \bar{F}_P is the structure factor of the native protein and \bar{F}_H the contribution of the heavy atom. Practically intensities of native crystals and of one or more derivatives are collected. The differences between derivative and native structure factors are then used for phasing:

$$|\bar{F}_H|^2 = \left(|\bar{F}_{PH}| - |\bar{F}_P| \right)^2 \quad 2.14$$

The positions of the heavy atoms are found by examination of Patterson maps. Graphically a multiple isomorphous replacement (MIR) can be described as in Fig. 2.2 where it can be seen that the phase can be unambiguously determined if two or more derivatives are available. On the contrary in case of single isomorphous replacement (SIR) there is a two fold phase ambiguity. The major problem of this technique is that often native and derivative crystals are not perfectly isomorphous.

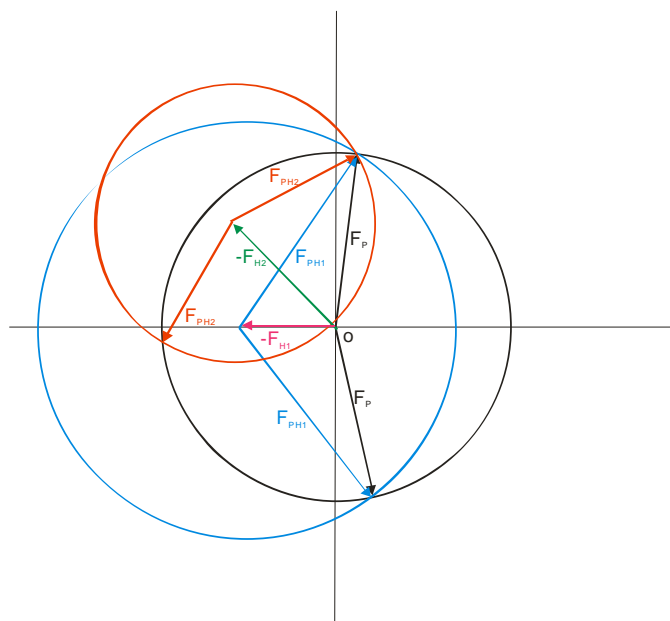


Figure 2.2 Harker construction of the MIR method. The black circle with radius $|\bar{F}_P|$ represents the native crystal with unknown phases. The light blue circle with radius $|\bar{F}_{PH1}|$ represents the first derivative with unknown phases. The vector \bar{F}_{H1} (depicted in magenta) is calculated from the heavy atom position obtained by inspection of the Patterson map. The intersections of the two circles define the two possible values for the phase of the native crystal (SIR). If a second derivative is available (red circle) the three circles intersect in a single point that defines the only possible value of the native crystal phase.

2.2.2 Anomalous dispersion

In SAD (Single Wavelength Anomalous Diffraction) and MAD (Multiple Wavelength Anomalous Diffraction) experiments the anomalous scattering of particular atoms present in the crystal is used to derive the phases of the protein. When the frequency of the X-rays is near the absorption edge of an element, an electron is ejected from the core of the atom and a photon can be absorbed: electrons from higher energy shells decay to lower energy ones emitting a photon with lower energy (fluorescence). The atomic scattering factor of an anomalous scatterer can be then described as:

$$f_{\text{ano}} = f_0 + f'(\lambda) + if''(\lambda) \quad 2.15$$

where f_0 is the normal scattering factor contribution, f' is the dispersive signal and f'' is the anomalous signal. f'' can be measured by fluorescence spectroscopy while f' can be calculated from f'' by the Kramers-Kronig equation (Fig. 2.3):

$$f'(E) = \frac{2}{\pi} \int_0^\infty \frac{E' f''(E')}{(E^2 - E'^2)} dE' \quad 2.16$$

The result of anomalous dispersion is that Friedel's law breaks down:

$$|\bar{F}_{hkl}| \neq |\bar{F}_{-h-k-l}| \quad 2.17$$

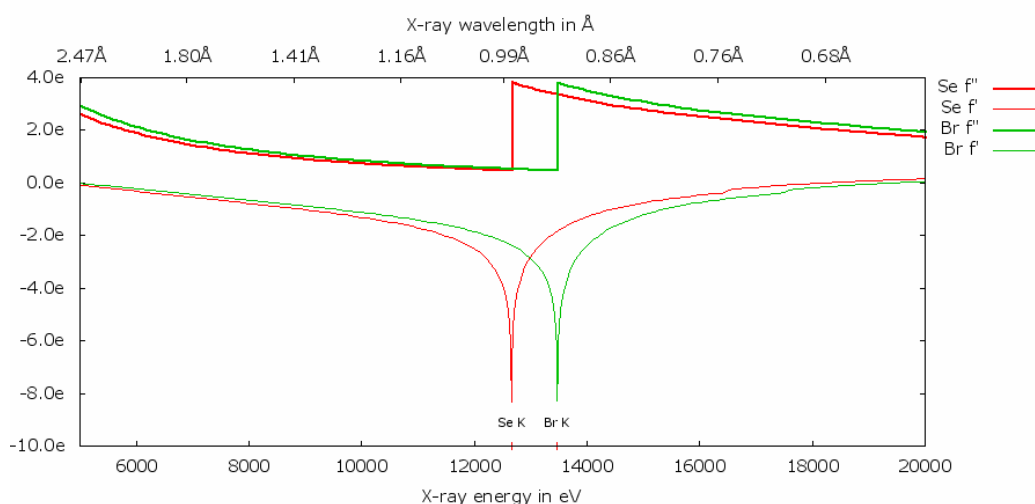


Figure 2.3 Theoretical absorption edges and values of f' and f'' for Selenium and Bromine (<http://www.bmsc.washington.edu/scatterer>).

The amplitudes of structure factors containing an anomalous scatterer are (Hendrickson *et al.* 1985):

$$|\bar{F}_{\text{obs}}|^2 = |\bar{F}_T|^2 + a(\lambda)|\bar{F}_A|^2 + b(\lambda)|\bar{F}_T||\bar{F}_A|\cos\alpha \pm c(\lambda)|\bar{F}_T||\bar{F}_A|\sin\alpha \quad 2.18$$

where \bar{F}_T is the total structure factor, \bar{F}_A the contribution of the anomalous scatterer, $a = (f''^2 + f'^2)/f_0^2$, $b = 2f'/f_0$, $c = 2f''/f_0$ and $\alpha = \varphi_T - \varphi_A$.

Experimentally the Bijvoet differences are measured (Fig. 2.4):

$$\Delta|\bar{F}_{\text{ano}}| \approx ||\bar{F}_{\text{hkl}}| - |\bar{F}_{-h-k-l}|| \quad 2.19$$

and because in a MAD experiment $\Delta|\bar{F}_{\text{ano}}|$ s are measured at least at two different wavelengths, $|\bar{F}_T|$, $|\bar{F}_A|$ and α can be extracted.

On the contrary in a SAD experiment some approximations have to be made:

$$|\bar{F}_T| \approx \frac{1}{2}(|\bar{F}^+| + |\bar{F}^-|) \quad 2.20$$

$$|\bar{F}_A|\sin\alpha \approx \frac{1}{2}(|\bar{F}^+| - |\bar{F}^-|) \quad 2.21$$

If $|\bar{F}^+| \gg |\bar{F}^-|$ than addition of $\alpha = 90^\circ$ to the phase of $|\bar{F}_A|$ is a good approximation of $|\bar{F}_T|$, if $|\bar{F}^+| \ll |\bar{F}^-|$ than addition of $\alpha = 270^\circ$ to the phase of $|\bar{F}_A|$ is a good approximation of $|\bar{F}_T|$ and if $|\bar{F}^+| \approx |\bar{F}^-|$ there is a two fold ambiguity for the value of the phase.

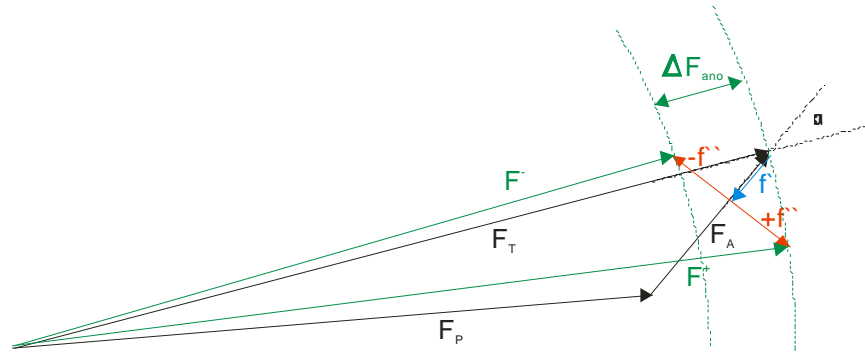


Figure 2.4 Argand diagram of a structure with significant anomalous signal. The diagram illustrates the origin of the inequality of phase and amplitude for the resulting structure factors \bar{F}^+ and \bar{F}^- and consequently the origin of the Bijvoet differences.

2.2.3 Molecular replacement

The structure of a related protein (determined either by X-ray crystallography or by NMR) with a sequence homology of at least 40%, can be used in molecular replacement as search model for providing initial estimated phases of the unknown structure. The information contained in the known structure has to be transferred to the crystal of the unknown protein: a proper orientation (described by three angles) and a proper translation (described by three translation parameters) have to be found. For doing that the Patterson function, which is the Fourier transform of the intensities and which can have phases only equal to 0° or 180° is calculated:

$$P(uvw) = \frac{1}{V} \sum_h \sum_k \sum_l |\bar{F}(hkl)|^2 \cos 2\pi(hu + kv + lw) \quad 2.22$$

where u , v and w are relative coordinates in the unit cell. The rotation function R , used to find the right orientation, is the result of the maximum overlap between two self-Patterson functions:

$$R(\alpha\beta\gamma) = \int_U P_{\text{obs}}(\bar{u}) \times P_{\text{calc}}(\bar{u}) d\bar{u} \quad 2.23$$

where U is the volume of the Patterson map, P is the Patterson map value at point \bar{u} and α , β and γ are the three rotation Euler angles.

Once the orientation has been fixed, the right position in the unit cell is described by the translation function T :

$$T(t_x t_y t_z) = \int_U P_{\text{calc}}(\bar{u}, t) \times P_{\text{obs}}(\bar{u}) d\bar{u} \quad 2.24$$

This function is optimized via a trial-error procedure: after each trial the calculated structure factors (\bar{F}_{calc}) are compared with the observed ones (\bar{F}_{obs}) and an R value or a correlation coefficient CC is calculated to rank the solutions:

$$R = \frac{\sum_{hkl} \left| \bar{F}_{hkl}^{\text{obs}} - k \bar{F}_{hkl}^{\text{calc}} \right|}{\sum_{hkl} \left| \bar{F}_{hkl}^{\text{obs}} \right|} \times 100 \quad 2.25$$

$$CC = \frac{\sum_{hkl} \left(|\bar{F}_{hkl}^{obs}|^2 - \left| \langle \bar{F}_{hkl}^{obs} \rangle \right|^2 \right) \left(|\bar{F}_{hkl}^{calc}|^2 - \left| \langle \bar{F}_{hkl}^{calc} \rangle \right|^2 \right)}{\left[\sum_{hkl} \left(|\bar{F}_{hkl}^{obs}|^2 - \left| \langle \bar{F}_{hkl}^{obs} \rangle \right|^2 \right)^2 \sum_{hkl} \left(|\bar{F}_{hkl}^{calc}|^2 - \left| \langle \bar{F}_{hkl}^{calc} \rangle \right|^2 \right)^2 \right]^{1/2}} \quad 2.26$$

where $\langle \bar{F}_{hkl} \rangle$ is the average structure factor amplitude over the reflections.

Most of the molecular replacement programs use a two steps procedure in which first a rotation function search is performed, followed by a translation function search. In the last years a six-dimensional search where all the orientations and translations are checked simultaneously, due to progress in the computer field, has become possible.

2.2.4 Direct methods

Direct methods are statistical methods that use probabilistic relationships between the phases. With the assumption that the electron density in a structure is concentrated in well defined maxima, Sayre (Sayre 1952) formulated the following relationship:

$$\bar{F}_h = q \sum_{h'} (\bar{F}_h \bar{F}_{h-h'}) \quad 2.27$$

where q is a constant dependent on $\sin\theta/\lambda$ for the reflection $h(hkl)$. Because the atomic scattering factors are dependent on the diffraction angles θ , the amplitudes $|\bar{F}|$ of reflections measured at different angles cannot be compared with each other directly and the use of normalized structure factors \bar{E} instead of \bar{F} is recommended.

From the Sayre's equation the so-called triple phase relation can be extracted:

$$\phi_h \approx \phi_{h'} + \phi_{h-h'} \quad 2.28$$

This relationship is true in a probabilistic sense and its probability function can be written as (Cochran 1955):

$$P(\Phi) = g \exp \left(\frac{2 |\bar{E}_h \bar{E}_{h'} \bar{E}_{h-h'}|}{N^{1/2}} \right) \quad 2.29$$

where g is a normalizing factor, $\Phi = \phi_h - \phi_{h'} - \phi_{h-h'}$ and N the number of atoms present in the structure. It is easy to see that direct methods are more successful for structures with a limited number of atoms.

Phases of non-centrosymmetric structures can be derived using the tangent-formula (Karle *et al.* 1956):

$$\tan(\phi_h) = \frac{\sum_{h'} |\bar{E}_h \bar{E}_{h'} \bar{E}_{h-h'}| \sin(\phi_{h'} + \phi_{h-h'})}{\sum_{h'} |\bar{E}_h \bar{E}_{h'} \bar{E}_{h-h'}| \cos(\phi_{h'} + \phi_{h-h'})} \quad 2.30$$

in which the summation is carried out over all relevant triplets.

For the solution of macromolecular structures dual-space recycling is used (Miller *et al.* 1993): a starting set of random atoms is generated, the corresponding structure factors are calculated, phases are refined and an electron density map is generated. Peaks are searched in the map and a list of new atoms is made. The whole process is repeated over several cycles (Fig. 2.5).

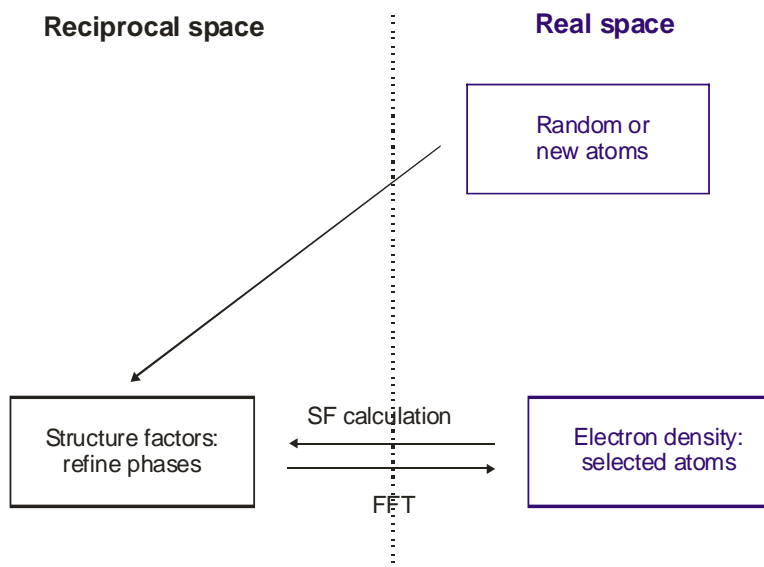


Figure 2.5 Dual-space recycling (Shake-and-Bake procedure) as implemented in *SHELXD*.

2.3 Density modification

The phases obtained experimentally or by molecular replacement are generally not very accurate. They can be improved by iterative density modification cycles. Density modification techniques use some conserved features of correct electron density maps to improve the observed map. Density modification is normally carried out in a two steps procedure: first the electron density map obtained experimentally is modified in real space in such a way that it is consistent with the expectations and then phases are calculated from this modified map and combined with the experimental ones to form a new set of phases. The major problem of this approach is how to weight the observed phases with the ones obtained from the modified map. The most common approaches used for density modification are *solvent flattening*, *histogram matching*, *NCS averaging*, *maximum-likelihood density modification* and the *sphere of influence method*.

Solvent flattening: Protein crystals contain often a large amount of solvent (30 to 80%). Only a small amount of solvent molecules is ordered in the crystal while the majority of the molecules form a disordered *continuum*. As result the solvent electron density map is evenly distributed and consequently can be flatted down so to emphasize the protein electron density that has higher average value and higher variance. In the Wang method (Wang 1995) a three dimensional grid is placed in the asymmetric unit of the cell. Around each grid point the standard deviation $\sigma(\rho)$ of the electron density is calculated for spheres with radius of approximately 3 Å:

$$\sigma(\rho) = \sqrt{\frac{\sum_{i=1}^N (\rho_i - \bar{\rho})^2}{N}} \quad 2.31$$

where ρ_i is the individual electron density at the grid point i , $\bar{\rho}$ is the average electron density in the sphere and N is the total number of grid points. The calculated σ values are ranked by height and divided into two groups corresponding to the percentage of solvent and protein in the cell. The points of the group with lower values of σ are considered as solvent and the density of them is set to a constant low value so that the noisy peaks are removed and the density distribution becomes flat.

Histogram matching: It takes into account features of the protein and can therefore be considered complementary to *solvent flattening*. The ideal electron density histogram of a protein does not depend on the structural conformation but it is influenced by the overall temperature factor, the resolution and the phase error (Zang *et al.* 2001). If resolution and overall temperature factor are known one can improve the phases by histogram matching. This technique can also be used for phase extension.

NCS averaging (Vellieux *et al.* 1997): This method is used when there are molecules in the unit cell related by non-crystallographic symmetry (NCS). In that case the electron density of NCS-related molecules is assumed to be similar and can thus be averaged. In this way density errors are minimized and phases can be improved after re-inversion of the map. Both NCS rotation matrix and protein density mask, which describes the shape of the electron density of a single molecule, have to be determined.

Maximum-likelihood density modification (Terwilliger 2000): The basic idea of this method is that there are two kinds of information needed for density modification. Information coming from experimental phases and amplitudes are used to calculate the probability distribution of the structure factors while the likelihood of a map resulting from the structure factors is used to determine the probability distribution of the map. Taking into account also the environment of different points in the map (histogram-based approach) and combining the two probabilities just described, a total likelihood is maximized. This idea is implemented in the program *RESOLVE* (Terwilliger 2000).

Sphere of influence method (Sheldrick 2002): In the sphere of influence method the variance of the density over the surface at a sphere of radius 2.42 Å (corresponding to 1,3-interatomic distance in macromolecules) is calculated. Pixels with the smallest variance are assigned to the solvent, pixels with the highest one to the boundary region and the rest to the protein region. The use of a crossover region allows to avoid that density modification will be locked into an initial poor solvent boundary.

2.4 Refinement

In most cases the model built manually or automatically in the electron density map obtained after density modification is incomplete and partially wrong. Iterative model building via inspection of electron density map and cycles of refinement should increase the agreement between observed and calculated data. The agreement can be expressed by the index R defined as:

$$R = \frac{\sum_{hkl} \left| |\bar{F}_{obs}| - |\bar{F}_{calc}| \right|}{\sum_{hkl} |\bar{F}_{obs}|} \quad 2.32$$

In macromolecular crystallography an additional index R_{free} (Brunger 1992) is used for cross-validation and to avoid overfitting: a small percentage of the data (normally around 5%) are not used in refinement and their agreement with the refined data is monitored.

A model is described by atomic coordinates and by atomic displacement parameters that, depending on the resolution and consequently on the data-to-parameter ratio, can be modeled as isotropic, anisotropic or via TLS. The data-to-parameter ratio can be artificially increased using constraints or restraints. The first are exact mathematical conditions that decrease the number of parameters while the second are additional observational equations that increase the number of data. More details about high resolution refinement and restraints are discussed in Chapters 3 and 5 respectively.

Basically there are two different mathematical approaches that are used in the refinement of macromolecules: *least-squares* and *maximum likelihood*.

In the *least-squares* method the observations (experimental intensities) have fixed values and the parameters are varied in such a way that the calculated values approach the observed ones as closely as possible. In the program *SHELXL* (Sheldrick *et al.* 1997) the *least-squares* method has been implemented using the function Q :

$$Q = \sum_{hkl} w_{hkl} \left(|\bar{F}_{obs}|^2 - |\bar{F}_{calc}|^2 \right)^2 \quad 2.33$$

where w_{hkl} is a weighting factor by which less accurate structure factors are down weighted. This method is usually too expensive for macromolecules for which the *conjugate gradient* solution of the sparse *least squares* normal equations (Tronrud 1992)

is often used. This kind of refinement works well only for models with very small errors and for which high resolution data are available.

In the *maximum-likelihood* method (implemented in the program *REFMAC* (Murshudov *et al.* 1997)) the theoretical approach is different. Given the model the probability function that a set of data would be observed is calculated. This can be described as:

$$L = \prod_i P(\bar{F}_{i,obs}; \bar{F}_{i,calc}) \quad 2.34$$

where $P(\bar{F}_{i,obs}; \bar{F}_{i,calc})$ is the probability distribution of the structure factor $\bar{F}_{i,obs}$ given the model structure factor $\bar{F}_{i,calc}$ (Pannu *et al.* 1996). For computational reasons the logarithm of the likelihood function is calculated and minimized.

3 Aldose Reductase

3.1 Introduction

In this chapter the crystal structure of the holo-enzyme human Aldose Reductase will be presented. This enzyme, implicated in the polyol pathway, has been the object of different studies of drug design and structures with different inhibitors have already been solved. The aim of the work was a careful investigation of the model with particular attention to the features characteristic of high resolution refinement like disordered residues, clusters of disordered parts of the molecule, hydrogen bond networks, anisotropic temperature factors, calculation of estimated standard deviations and accurate examination of disagreeable restraints.

3.1.1 Aldose Reductase

Human Aldose Reductase (AR; EC 1.1.1.21) is a member of the aldoketoreductase superfamily and catalyzes the NADPH-dependent reduction of a wide variety of aldehydes to their corresponding alcohols. It is implicated in the polyol pathway converting glucose to sorbitol, which is then transformed into fructose by sorbitol dehydrogenase (Fig. 3.1). The increasing level of glucose through the polyol pathway induces biochemical imbalances in some tissues and leads to diabetic complications such as the development of glaucoma, retinopathies and cataract (Kinoshita *et al.* 1988; Yabe-Nishimura 1998; Costantino *et al.* 1999; Dunlop 2000). *In vitro* and *in vivo* studies suggest that a high level of sorbitol can be avoided by inhibition of Aldose Reductase. Inhibitors of AR thus represent promising targets for the development of new drugs (Costantino *et al.* 2000; El-Kabbani *et al.* 2004b; Klebe *et al.* 2004; Petrash 2004; Podjarny *et al.* 2004).



Figure 3.1 Polyol pathway showing the chemical reactions and the enzymes involved in the transformation of glucose into fructose via sorbitol.

Human Aldose Reductase consists of 316 residues, weights 36 kDa and the crystals have a 43% solvent content. It folds as an eight-stranded β/α barrel with a small β -sheet on the N-terminus (residues 2-14) and a C-terminal extension (residues 275-315). The active site is located on the C-terminus of the barrel and the NADP^+ -binding site is near the hydrophobic active site pocket (Urzhumtsev *et al.* 1997; El-Kabbani *et al.* 1998). The catalytic residues of the active site are Lys77, His110 and Tyr48 (which most likely acts as proton donor, (Bohren *et al.* 1994)) together with the nicotinamide moiety of the cofactor NADPH. The active site can be divided into a rigid invariant region called anion binding pocket (residues Trp20, Val47, Tyr48, Trp79, His110) where the residues are interacting with the negatively charged part of the inhibitor (Harrison *et al.* 1994; Klebe *et al.* 2004; Sotriffer *et al.* 2004); in a quite flexible part (induced-fit phenomena) called specificity pocket (residues Thr113, Phe115, Phe122, Cys303, Tyr309) where the residues are interacting with only some of the inhibitors (El-Kabbani *et al.* 2004a; Howard *et al.* 2004; Ruiz *et al.* 2004; Steuber *et al.* 2006), and in a very flexible part formed by the strand between Val297 and Leu300.

A variety of compounds that inhibit Aldose Reductase have been investigated for many years. Inhibitors of AR can be divided into four major classes (Miyamoto 2002):

- hydantoin type inhibitors
- new sulfonyl-pyridazinone type inhibitors
- carboxylic acid type inhibitors
- arylsulphonylnitromethanes inhibitors.

Hydantoin type inhibitors (Fig. 3.2A), like Sorbinil, Fidarestat and Minalrestat, have a pK_a value of about 8 and they can easily cross the membrane. Unfortunately Sorbinil induces hypersensitivity reactions, through toxic intermediate (Spielberg *et al.* 1991).

On the contrary, due to their low pK_a , carboxylic acid type inhibitors (Fig. 3.2B) are ionized under physiological conditions and they can hardly cross the membrane (Krentz *et al.* 1992). The structure of the inhibitor IDD594 is presented in Appendix A.

Recent studies on phenylsulphonylnitromethanes (Fig. 3.2C) show that these compounds too are AR inhibitors.

In 2003 a new class of inhibitors with excellent properties was proposed (Fig. 3.2D,

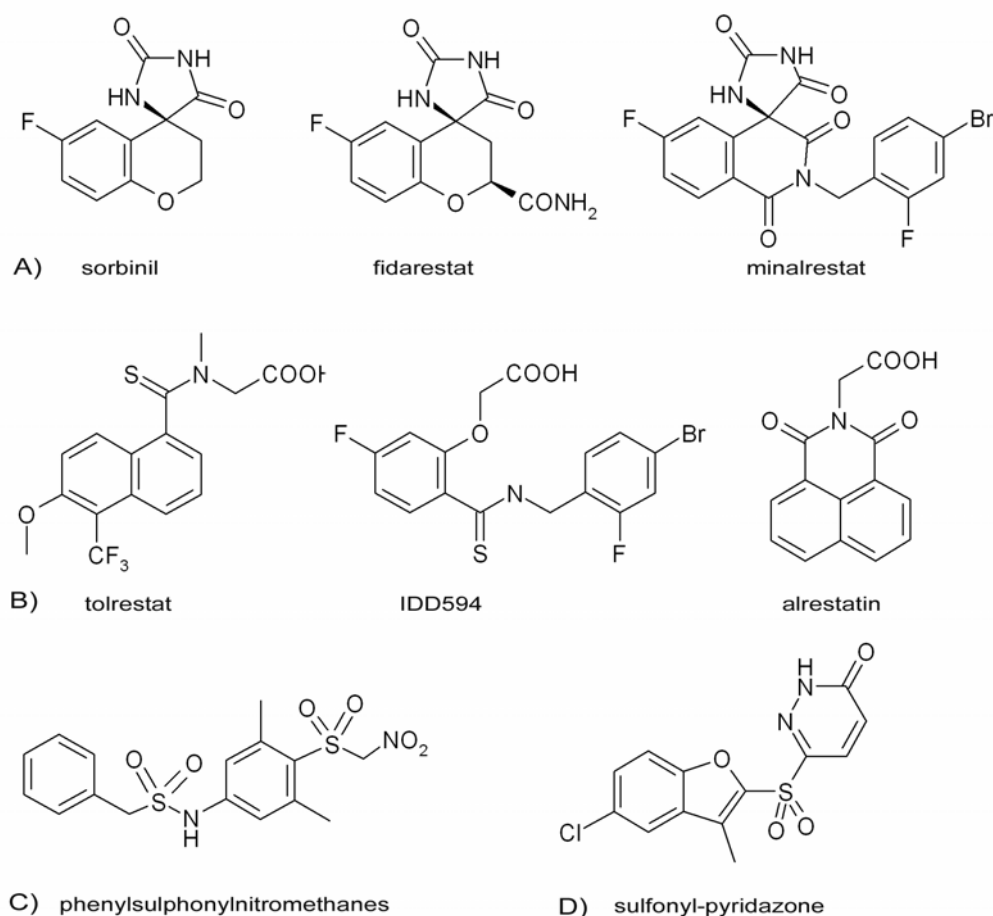


Figure 3.2 Some structures of different AR inhibitors. A) Hydantoin type inhibitors. B) Carboxylic acid type inhibitors. C) Arylsulphonylnitromethanes inhibitor. D) Novel sulfonyl-pyridazone inhibitor.

(Mylari *et al.* 2003)). It is a non-carboxylic, non-hydantoin inhibitor with good *in vivo* efficiency and almost perfect bioavailability.

The mechanism of the forward-reaction catalyzed by Aldose Reductase can be described as a sequential ordered mechanism in which NADPH binds first, followed by the binding of the substrate (Kubiseski *et al.* 1992). Once the aldehyde has been converted into the corresponding alcohol, the product is released and only then is the oxidized cofactor disengaged. Before crystal structures were determined, kinetic studies already showed that the cofactor up-take and release are accompanied by a conformational change of the enzyme*cofactor complex (Kubiseski *et al.* 1992). The conformational change of the protein upon NADPH binding was confirmed monitoring via spectroscopic fluorescence the quenching of the protein (that contains 6 Trp residues) and a reaction scheme with a fast bimolecular ligand-protein association followed by a unimolecular protein isomerization was proposed. It had been previously observed in other kinetic experiments that the conformational change preceding the cofactor dissociation is the overall rate-limiting step in the enzymatic cycle (Grimshaw *et al.* 1990; Grimshaw *et al.* 1995). Different catalytic mechanisms have been proposed during the last decades but the most plausible one is based on quantum chemical calculations and high resolution structures (Cachau *et al.* 2002). The protonated Lys77 makes a salt-bridge with the negatively charged Asp43 and polarizes Tyr48. The singly protonated and flexible His110 takes one hydrogen from Tyr48 and then moves toward the substrate and donates it one proton. His110 acts as a proton shuttle (Fig. 3.3).

The first crystal structure of Human Aldose Reductase (Wilson *et al.* 1992) showed the cofactor bound to a cleft running across the C-terminal side of the β/α barrel. The cofactor is kept in position by residues which are mostly located in the loop that extends from residue Gly213 to residue Leu227 and has been called safety-belt. Inspection of the crystal structure of a mutated version of Human Aldose Reductase (Cys298Ser) in complex with NADP(H) (Bohrani *et al.* 1992) and of Porcine Aldose Reductase in complex with ADPRP (Rondeau *et al.* 1992) revealed that the conformation of the safety-belt strongly depends on the cofactor bound and suggested that a motion of this loop between an open and a closed conformation is related to the cofactor up-take and release.

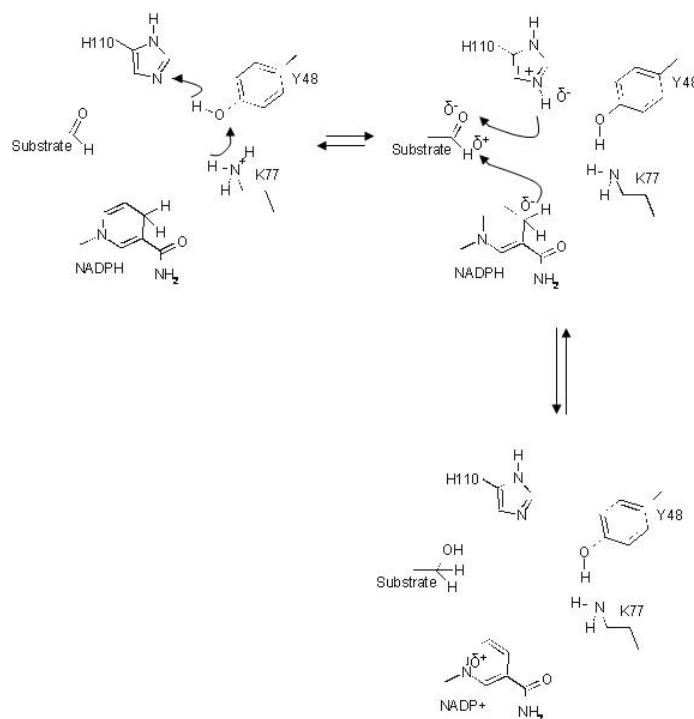


Figure 3.3 Mechanism of aldehyde reduction by AR (Cachau *et al.* 2002): Tyr48 acts as first proton donor, His110 as proton shuttle, the substrate is reduced to alcohol and the cofactor NADPH is oxidized to NADP⁺.

A mutated version of Human Aldose Reductase (Arg268Ala) could be crystallized without the cofactor (Bohren *et al.* 2005). Indeed the mutation removes an important interaction between the cofactor and the enzyme, thereby reducing the affinity of NADPH by several orders of magnitude. In this structure of the apo-enzyme the safety-belt is in a completely open and ordered conformation which is stabilized by a π -stacking interaction between Trp219 and Arg293.

3.1.2 High resolution refinement

Over the last few years developments in X-ray sources, in efficiency of 2D detectors and in cryo-freezing techniques have made it possible to obtain macromolecular crystal structures at high resolution. The advantages of high resolution and, consequently, of a large amount of data are the possibility to solve the structure with direct methods and

to appreciate fine details of the structure that can provide additional information useful to understand the mechanism of a reaction and the biological role of the molecule (Dauter 2003). Moreover, it is possible to study the enzyme interactions with ligands like inhibitors or substrates, thereby enhancing the development of the drug-design sector.

High resolution means to have data to a resolution better than 1.2 Å and with 50% of the theoretically measurable reflections in the outer shell with an intensity higher than 2 sigma(I) (Sheldrick 1990). With this resolution it is possible to invert the *least-squares* matrix and to calculate the estimated standard deviations (ESDs) of each individual atomic parameter. Normally, ESDs for high resolution structures are much smaller than the ones for medium or low resolution structures. This also allows to model the atomic displacement parameters (ADPs) anisotropically with the consequent decrease in the noise of the maps and the possibility to model flexible parts of the molecule not fully occupied. At very high resolution, hydrogens start to be visible in difference maps ($F_o - F_c$) for the well ordered part of the molecule providing the opportunity to observe protonation states, to elucidate the catalytic mechanism and to study the network of hydrogen bonds. Structures refined at high resolution quite often have lots of disagreeable restraints. The restraints used in *SHELXL* (Sheldrick *et al.* 1997) are based on the ones of Engh and Huber (Engh *et al.* 1991). They were compiled looking at the structures of amino acids and peptides present in the Cambridge Structural Database (Allen 2002). Amino acids in proteins are not isolated molecules and can be distorted by the surrounding residues or, eventually, by an inhibitor or other molecules present in the crystal. For these reasons, structures refined at high resolution show deviations from standard geometry such as non planar peptide bond, non planar aromatic ring for aromatic side chains, unusual values of chiral volumes and so on. These deviations can indicate special features of the molecule that are observable only at high resolution. They may otherwise indicate that the restraints used are wrong and the new values found should be used to create a new library of restraints parameters.

3.2 Material and methods

3.2.1 Crystallization and data collection

Human Aldose Reductase was crystallized by the vapour diffusion method. The protein concentrated to 16mg/ml in 50 mM Ammonium Citrate buffer (pH 5), was mixed with the cofactor NADP⁺ and the stereoisomer 2R4R of the inhibitor Fidarestat in a 1:2:2.5 ratio. The solution containing the protein, the cofactor and the inhibitor was mixed with an equal volume of 15% (w/v) PEG6000 in 50 mM Ammonium Citrate buffer at pH 5. The reservoir solution (1 mL) contained 120 mM Ammonium Citrate and 20% (w/v) PEG6000. The hanging drops had a volume of 10 μ L. Crystals grew at 4° C

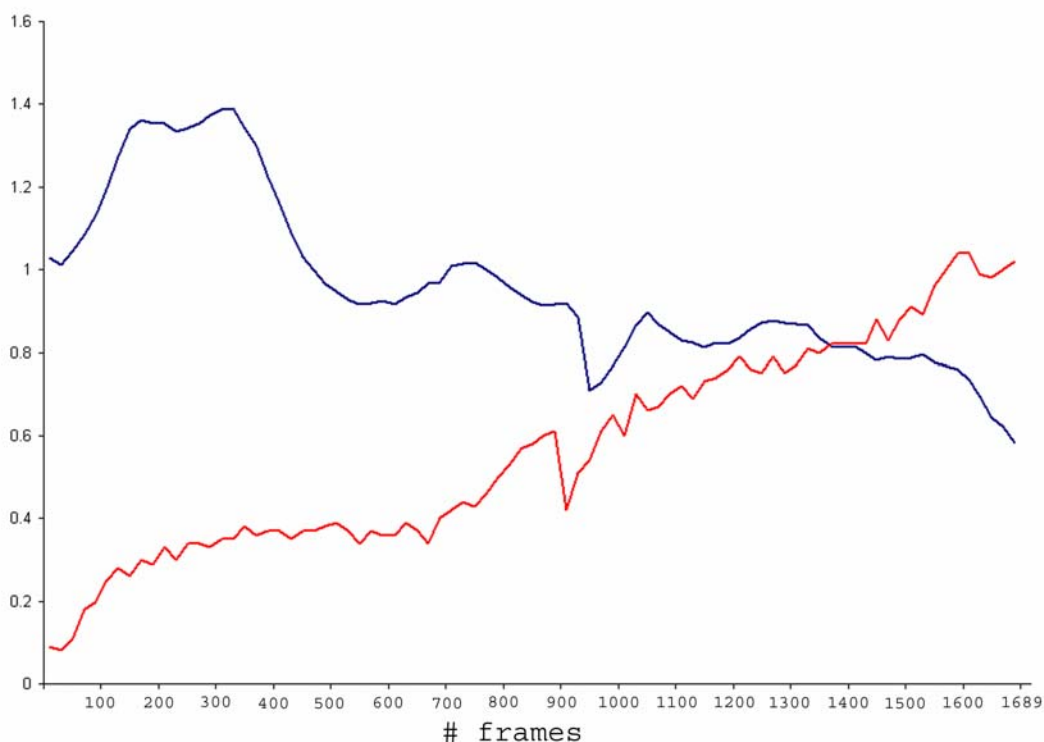


Figure 3.4 Graph showing the values of the B factors (in red) and of the scale factors (in blue) for each 20th frame of the high resolution pass. The decrease of the scale factors and the increase of the B factors could be a sign of radiation damage.

and, prior to data collection were transferred first into a stabilization solution (25% (w/v) PEG 6000) and then into a cryoprotecting solution (40% (w/v) PEG6000). Data to 0.82 Å were collected from a single crystal at the beamline ID19 at the APS, Structural Biology Centre at 10 K, $\lambda = 0.7999$ Å using a marCCD detector (Mar Research, Norderstedt, Germany). Two data sets were collected: a low-resolution one extending to 1.44 Å (676 frames with an oscillation angle of 0.5°) and a high-resolution one extending to 0.82 Å (1690 frames with an oscillation angle of 0.2°). In the high-resolution data set, refined scale factors decreased from 0.8844 to 0.5458 and B values increased from 0.42 to 0.98 between frames 900 and 1690. Both these observations might indicate some radiation damage (Fig. 3.4). Anyway other sources like incorrect centering of the crystal, inaccurate spindle movement or imperfect stability of the beam cannot be excluded. Thus, only intensities from the low-resolution data set (cut at 1.8 Å) and the first 900 frames of the high-resolution data set (cut at 8 Å, corresponding to 180° rotation of the crystal) were used to construct the merged data set for refinement. The data were integrated and scaled using *HKL2000* and *SCALEPACK* (Otwinowski *et al.* 1997) while further data analysis and merging were performed using the Bruker program *XPREF*. 5% of the reflections were randomly flagged by *XPREF* for use in cross-validation. The crystal belongs to the monoclinic space group P2₁, with the unit cell parameters $a = 47.44$, $b = 66.73$, $c = 49.29$ Å, $\alpha = 90$, $\beta = 91.55$, $\gamma = 90^\circ$. Data statistics are shown in Table 3.1.

Resolution range (Å)	34.6 - 0.82
No. of observations	394966
No. of unique reflections	131238
Redundancy	4.26 (3.34)
$\langle I/\sigma(I) \rangle$	26.11 (12.24)
Completeness (%)	94.3 (90.3)
$R_{\text{int}}^{\S}(\%)$	4.21 (8.43)

Table 3.1 Data statistics of native Human Aldose Reductase in complex with NADP⁺. Values between parenthesis are for the highest resolution shell (0.92-0.82 Å, corresponding to 96796 reflections; § $R_{\text{int}} = \sum (|I - \langle I \rangle|) / \sum I$).

3.2.2 Structure determination

The structure was solved by molecular replacement. Data in the resolution range 15-4 Å were used. The atoms belonging to the protein and the cofactor NADP⁺ of the crystal structure of the complex Aldose Reductase-IDD384 (Calderone *et al.* 2000) were chosen as search model. The molecular replacement was carried out with the program *EMPR* (*Evolutionary Programming for Molecular Replacement*, (Kissinger *et al.* 1999)). In this program an efficient six-dimensional search is implemented using an evolutionary optimization algorithm: a starting set of random molecular replacement solutions is iteratively optimized against the correlation coefficient between observed and calculated structure factors. In the case of human Aldose Reductase the best solution had a correlation coefficient of 68.9% and an R factor of 34.3%. The rotation function was characterized by Euler angles of $\theta = 261.51^\circ$, $\phi = 91.77^\circ$ and $\omega = 269.68^\circ$ and the translation function by three vectors $t_x = 23.59$ Å, $t_y = -0.10$ Å and $t_z = 9.42$ Å.

3.2.3 Refinement

The refinement was carried out with the program *SHELXL* (Sheldrick *et al.* 1997) using the normal procedure for high resolution structure (Table 3.2). The model coordinates found by molecular replacement were first refined with twenty cycles of *least-squares* (LS command) refinement. For this purpose the molecule was considered as a single rigid-body and only data to 2.5 Å were used. *Least-squares* refinement was possible at this stage because the amount of parameters was still small. For a better convergence, the resolution was increased from 4 Å to 2.5 Å in step of 0.1 Å (STIR command). The B factors were initially refined as isotropic using data to 1.5 Å and *conjugate gradient least-squares* refinement (CGLS command). At this point water molecules were added automatically with the program *SHELXWAT*. The initial isotropic value U for waters was set to 0.1 Å², in each cycle up to 50 solvent molecules were added and waters with isotropic U value higher than 0.2 Å² were rejected. In each job waters were added in ten cycles and after each cycle a normal *SHELXL* refinement was performed. In total four jobs of *SHELXWAT* were carried out. The waters were added

Step(s)	Remark	Resol. (Å)	# data	# param.	R _{work} (%)	R _{free} (%)
1	Rigid body refinement	10-2.5	10008	2582	40.0	39.7
2	Isotropic B refinement	10-1.5	46050	10295	28.0	31.2
3-6	Waters added	10-1.2	89112	11243	19.5	21.9
7	All data	10-0.82	268333	11243	18.2	19.8
8	Anisotropic refinement	10-0.82	268333	25293	14.5	15.9
9-46	Rebuilding	10-0.82	268333	32565	10.7	12.4
47	Add hydrogens	Inf-0.82	268491	32565	9.8	11.1
48	All data (no R _{free})	Inf-0.82	282631	32565	9.8	

Table 3.2 Major refinement steps with the corresponding R_{work} and R_{free}. The column data list the unique data including the 5% of them used for the cross-validation (R_{free}).

automatically looking at the peaks in the difference electron density map and taking into account also chemical information such as the possibility to make hydrogen bonds. The model was then refined against all data and the B factors were refined anisotropically (ANIS command): R_{work} and R_{free} drop to 14.5% and to 15.9% (with a difference in R_{free} equal to 3.9%) respectively. It was only at this stage, that a systematic checking of the model with particular attention to disordered side chains and to high peaks visible in difference electron density maps was made by visual inspection with the program *XFIT* (McRee 1999). More details are given in the next subchapters. Cycles of iterative rebuilding and refinement were carried out until the end of the refinement was reached. Finally twenty cycles of *conjugate gradient least-squares* refinement were carried out using all data. The trend of the refinement is shown in Fig. 3.5. Error estimates of all parameters (bond lengths and angles, planarity etc) were obtained through inversion of the *least-squares* matrix. Because the amount of parameters was extremely large, the block matrix *least-squares* approximation was used. Simultaneously all the restraints were removed and the damp for all parameters was set to zero (BLOC and DAMP 0 0 commands). The refinement was considered to be completed when no more entries

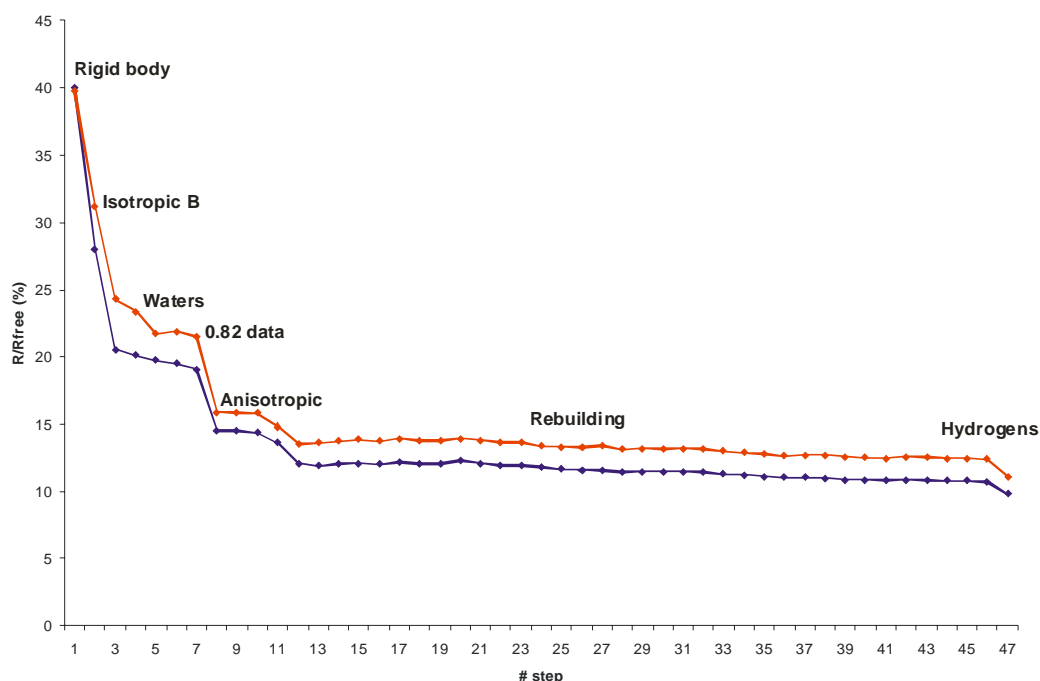


Figure 3.5 Trend of R_{work} (blue squares) and R_{free} (red squares) during the refinement.

unjustifiable in the list of disagreeable restraints, no peaks higher than 5 sigma in the difference map and no decrease of R_{work} and R_{free} were seen.

3.2.3.1 Refinement of disordered side chains

At high resolution it becomes possible to model disordered residues in two conformations. The disorder can involve only the side chain atoms or also backbone ones but the way to proceed to model them is the same in both cases. There are different things to look at to detect possible disordered residues:

- atoms with extremely high B values
- list of disagreeable DELU/SIMU restraints (longer than 3 sigma) in the *.lst* file
- atoms flagged as "may be split" or "non positive definite"
- peaks higher than 5 sigma in the difference electron density map ($F_o - F_c$) near an atom.

Once the candidate disordered residues had been detected, the electron density was checked. Normally there was negative density in correspondence of the conformation

already modeled and some positive density in correspondence of the probable second conformation (Fig. 3.7A). The occupancy of the main conformation was fixed to 65% and some cycles of refinement were carried out. Then the electron density maps were again carefully inspected and the second conformation was modeled in the positive electron density (Fig. 3.7B). The occupancy was refined using free variables and restraining the sum to be equal to one. After refinement of the second conformation neither positive nor negative density was visible and the disordered residue was fitting nicely in the σ_A weighted ($2mF_o - DF_c$) electron density map (Fig. 3.7C). An example of part of the *SHELXL* instruction file is shown in Fig. 3.6.

```

WGHT      0.100000
SWAT      0.288673  391.7671
FVAR              0.08076  0.5

RESI 133  SER
N      3      0.942239      0.063359      0.347333      11.00000      0.06452      0.09018 =
      0.09876      0.02123      0.00358      -0.00286
CA     1      0.946849      0.056943      0.375316      11.00000      0.10634      0.09692 =
      0.09749      0.02658      0.00714      0.02208
C      1      0.961583      0.036783      0.376590      11.00000      0.06130      0.08738 =
      0.10873      0.01548      0.01046      0.00028
O      4      0.957089      0.024340      0.359141      11.00000      0.12703      0.11088 =
      0.12922      0.00168      -0.00659      0.01988

PART 1 21.00
CB      1      0.919840      0.056411      0.390916      10.65000      0.14134
OG      4      0.905980      0.043784      0.373185      10.65000      0.21531
PART 2 -21.00
CB      1      0.918758      0.055662      0.389901      10.35000      0.14134
OG      4      0.923591      0.038107      0.403636      10.35000      0.21531
PART 0

```

Figure 3.6 Example of part of the instruction file of *SHELXL* (.ins). FVAR are the free variables.

In this case the second one is associated with the occupancy of residue 133. The atoms of the main conformation (PART 1 21.00) will have an occupancy equal to the refined value of FVAR while the ones of the second conformation (PART 2 -21.00) will have one equal to (1-FVAR).

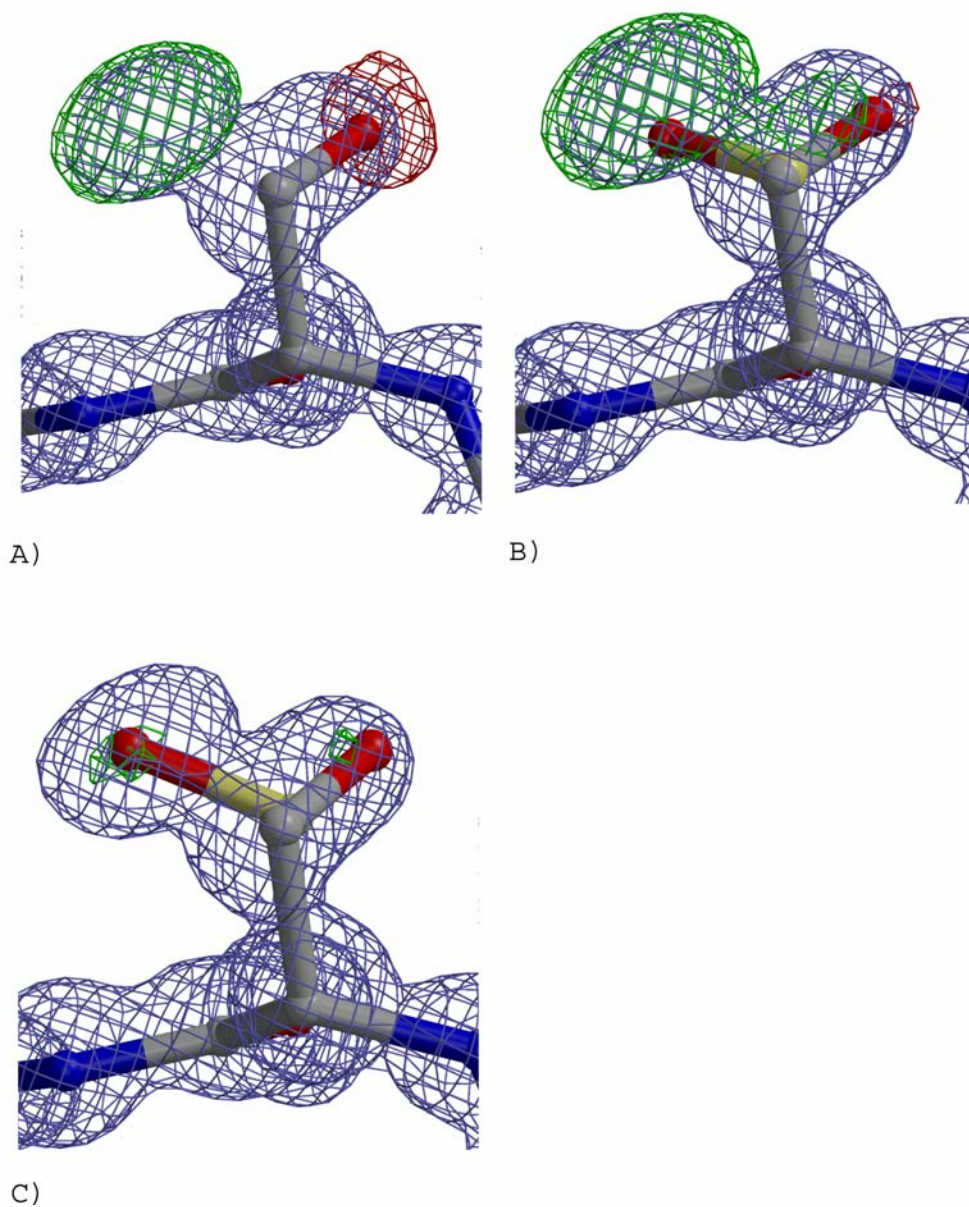


Figure 3.7 Example of disordered residue modeling (Ser133). σ_A weighted $(2mF_o - DF_c)$ map contoured at 1σ is colored in blue while positive and negative difference maps ($F_o - F_c$) contoured at 2.5σ are colored in green and red respectively. A) The residue is fully occupied in a single conformation. B) The occupancy of the atoms CB and SG has been decreased to 65% and the second conformation has been modeled. C) The final model with occupancy refined to 59.3% for the main conformation (in grey) and to 41.7% for the second conformation (in yellow).

Part of the disordered residues was already visible in the first stages of the refinement, while other regions became clearer only when the phases got better. Some residues showed a disorder that could have been modeled in more than just two conformations. Attempts to model them in three conformations did not improve the model and it was therefore decided to refine them in two conformations only. The main conformation of other residues was not fully occupied but it was not possible to model a second one. The occupancy of connected disordered residues was refined to the same value using the same free variable.

Particular attention was given to the active site and the safety-belt loop. A molecule of the stereoisomer 2R4R of the inhibitor Fidarestat should have been present in the active site. When the inhibitor was placed in the active site and refined, even if only partially occupied, negative difference electron density ($F_o - F_c$) appeared. It was then decided to put a molecule of citrate instead, but after some cycles of refinement it became clear that this molecule too was not fully occupied. The occupancy of the citrate in the major conformation was then fixed to 64% and after some cycles of refinement a second conformation was modeled in the appearing positive difference electron density map ($F_o - F_c$). A similar approach was used to model the second conformation of the safety-belt loop: the occupancy of the main conformation was fixed to 64% together with the B values and the minor conformation was partially built in the positive difference electron density map ($F_o - F_c$). The occupancy of the citrate in the active site, of the residues of the safety-belt loop, of other residues connected to them and of some waters was refined using only one free variable (see Appendix B).

3.2.3.2 Refinement of waters

The first 237 water molecules were added automatically by the program *SHELXWAT*. More waters were then added manually looking at peaks higher than 5 sigma in the difference electron density map ($F_o - F_c$). If a peak could make hydrogen bonds with the surrounding residues or waters and if it had tetrahedral coordination geometry, then it was modeled as the oxygen atom of a water molecule (Fig. 3.8A/B).

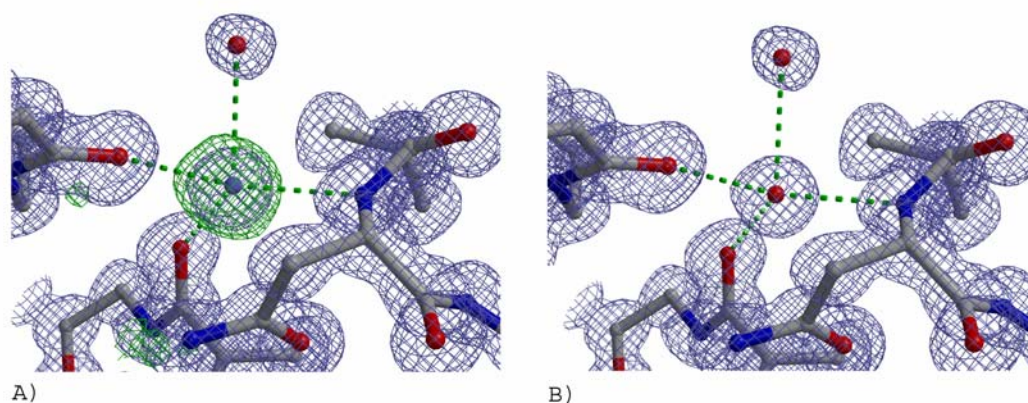


Figure 3.8 Example of water modeling. A) Peak higher than 5σ in the difference ($F_o - F_c$) electron density map (shown in green and contoured at 2.5σ) has a round shape σ_A weighted ($2mF_o - DF_c$) density map (shown in blue and contoured at 1σ) and makes hydrogen bonds with the surrounding residues showing a tetrahedral coordination. B) The peak was refined as the oxygen atom of a water molecule with occupancy set to 1.

Waters were refined into three major groups:

- fully occupied waters
- partially occupied waters connected to disordered residues or to at least two water molecules
- couples of half occupied waters.

3.2.3.3 Hydrogens

Hydrogen atoms have only a single electron and consequently their scattering power is weak. For this reason, even at high resolution, their signal disappears in the noise and hydrogens are visible in positive difference electron density map ($F_o - F_c$) only for the well ordered parts of the molecule. Hydrogens are normally added in calculated positions. They are not refined as independent atoms but as "riding" on the atom they are attached to (HFIX command). Hydrogens attached to nitrogen atoms of the imidazole ring of His, to the oxygen atom of the hydroxyl group of Tyr, Thr and Ser were not

added: they can occupy several possible positions and if a wrong one is automatically chosen, the antibumping restraints could cause problems.

3.2.3.4 Disagreeable restraints

A careful inspection of the list of the disagreeable restraints was made. The major restraints used in *SHELXL* and interesting for such a study are: antibumping restraint, chiral volume restraint and flatness restraint. The antibumping restraints (BUMP) are generated automatically after each cycle of refinement and they prevent the distance between two non-bonded atoms from being shorter than a certain value. They should avoid unfavorable side chain conformations. The chiral volume restraints (CHIV) are used to prevent wrong chirality or disallowed conformations. The flatness restraints (FLAT) ensure peptide bonds and some groups (aromatic ring, guanidinium group) to be flat. The list of disagreeable restraints written out after the last refinement was checked together with the model and with the electron density map after refinement with and without restraints in order to see if the restraints were wrong or if the structure was presenting special features.

3.3 Results and discussion

3.3.1 Aspects of high resolution refinement

Some aspects characteristic of high resolution refinement are analyzed in the following subchapters. First a careful inspection of disordered residues will be shown with particular attention to the presence of clusters; then analysis of waters and hydrogens will be carried out followed by an examination of disagreeable restraints. At the end the calculation of the estimated standard deviations (ESDs) of the parameters used and an overall picture of the quality of the model will be presented.

3.3.1.1 Disordered regions

The modeling of disordered residues in two conformations generally decreases the B values of the atoms involved, the amount of peaks visible in positive difference electron density map and both R_{work} and R_{free} . Due to the absence of the inhibitor in the active site, the whole molecule of Aldose Reductase studied here presents a high level of disorder with 89 out of 315 (27.8 %) residues modeled in more than one conformation. Most of the disordered residues are concentrated in the loops (66 out of 89 total disordered residues corresponding to 75%). 52 residues are disordered and isolated (Fig. 3.9A/B/C). The majority of the disordered residues (40) also have their backbone refined in more than one conformation (Fig. 3.9D) while in 29 residues (9 Lys) it is only the side chain that is disordered. For 21 residues it was possible to model only one conformation but not fully occupied. Part of the second conformation of the safety-belt loop (Ala220 to Asp224) and the last three residues at the C-terminal side (Glu313, Glu314 and Phe315) were not visible in the electron density map. The residues with disordered or not fully occupied side chains are in most of the cases located on the surface of the protein: probably the side chains are just floating around in the solvent. The

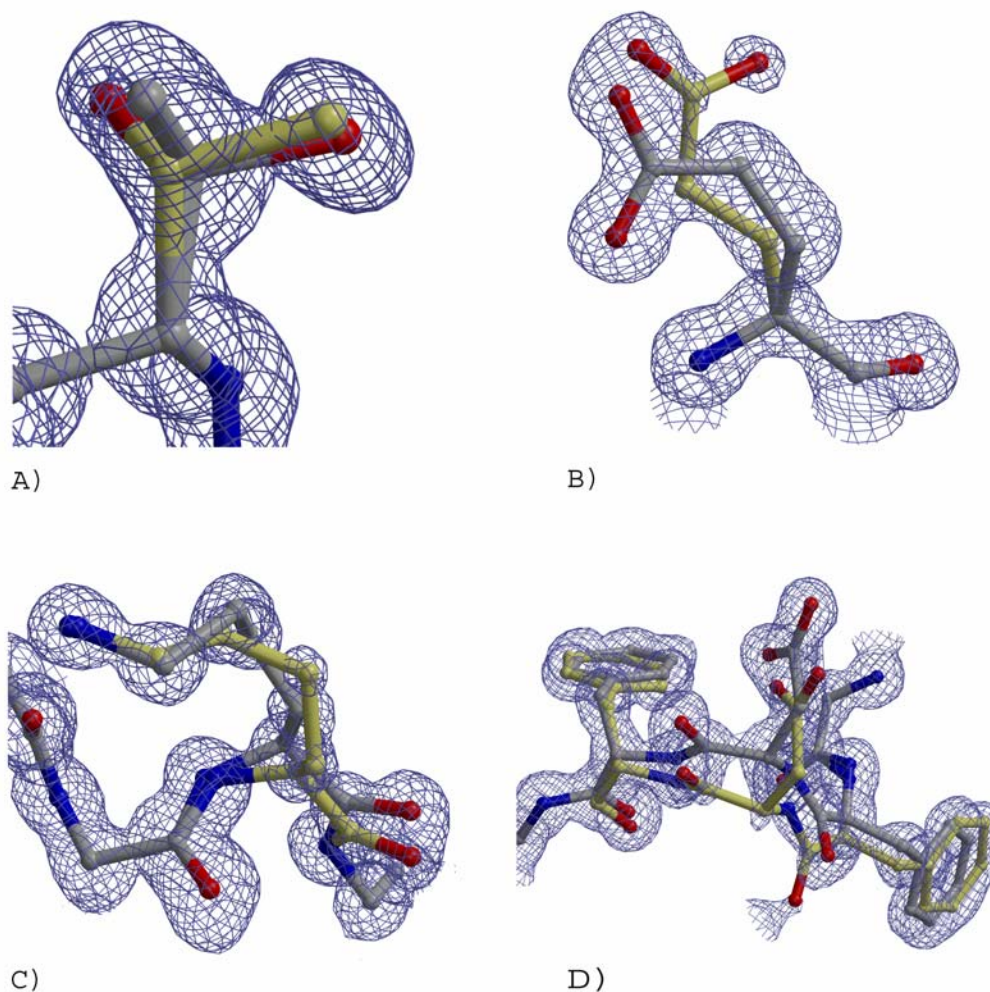


Figure 3.9 Example of disordered residues. The main conformation is always drawn in grey while the minor one in yellow. σ_A weighted $(2mF_o - DF_c)$ density map (contoured at 1σ) is shown in blue. A) Residue Thr144 with an occupancy for the main conformation equal to 80.5% and of the second one to 19.5%. B) Residue Glu193 with the main conformation refined with an occupancy of 72.9% and the second one of 27.1%. C) Residue Lys100 for which also the backbone has been modeled in two conformations (occupancy of the main one equal to 60.9% and of the second one equal to 39.1%). D) Example of backbone disorder: residues Phe276, Asp277, Phe278 have two conformations characterized by a shift of the main chain.

side chains of some Glu residues show a clear negative difference electron density ($F_o - F_c$): due to radiation damage (Weik *et al.* 2000) these residues may have been decarboxylated during the data collection. The disorder of some residues had to be coupled because they were directly interacting through hydrogen bonds or through waters or they were clashing with each other (Fig. 3.10). Four couples of disordered residues were indeed modeled using the same free variables. A significant part of the residues modeled in multiple conformations were refined in clusters (Table 3.3). One of the clusters includes residues of the strand from His83 to Leu87 and of the loop from Ser133 to Ile137 plus seven water molecules.

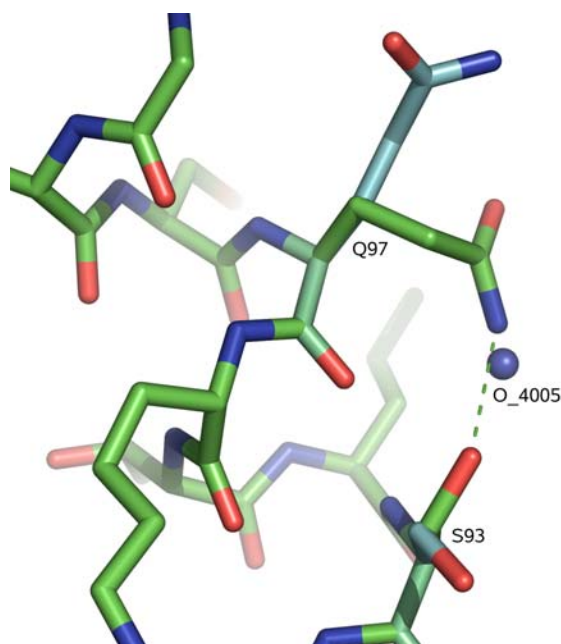


Figure 3.10 Example of residues refined using the same FVAR. In green the main conformation where Ser93 and Gln97 are making a strong hydrogen bond (green dashed line) is shown while the second conformation, where a water molecule was also refined, is shown in cyan.

Couple	Residues	Waters	Occupancy main conformation (%)
I	N40-V67		51.1
II	Q93-S97	4005, 4029	56.5
III	E229-K234	4004, 4028	57.2
IV	I4-K194	4009	62.0
Cluster	Residues	Waters	Occupancy main conformation (%)
I	H83-L87; S133-L137	A: 4034, 4036, 8001, 8002, 8003, 8004, 8005 B: 4035, 8006, 8007, 8008	58.7
II	K21; D216-D224; N256; C298; CIT320	A: 4013, 4018, 4020 B: 4001, 4006, 4016, 4017, 4019, 4021	63.8
III	L124-D125; S127-V131	4014, 4036	69.1

Table 3.3 Residues and waters refined in couples and clusters. In the case of waters A indicates the most occupied site and B the second one.

The main conformation has an occupancy of 58.3% (Fig. 3.11). Another cluster includes the safety-belt loop residues (Gly213 to Ser227), the citrate in the active site, the residues Lys21, Cys296, Asp256 and nine waters. The main conformation has an occupancy of 66%. A detailed discussion of the safety-belt loop is presented in section 3.3.2.3. Lys21 in the main conformation makes the characteristic salt-bridge with Asp216 while this does not occur in the second one (Fig. 3.12). The occupancy of the eight residues of the loop from Leu124 to Val131 together with two waters was also refined using a single free variable (see Appendix B).

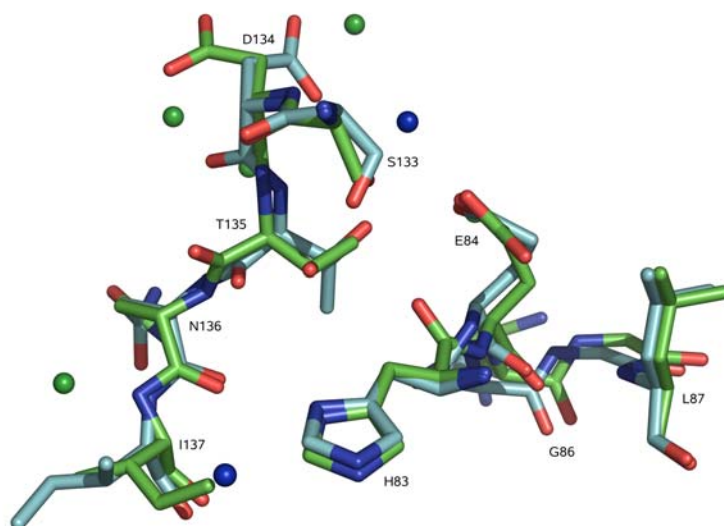


Figure 3.11 Residues refined in cluster I. The main conformation (occupancy = 58.7%) is shown in green and the second one (occupancy = 41.3%) in cyan. Some of the waters refined in this cluster are also shown here.

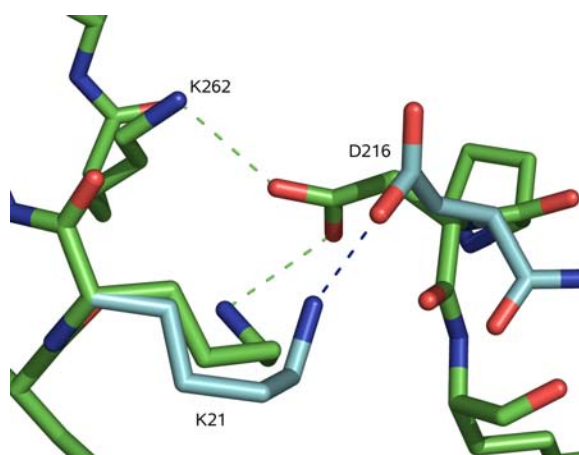


Figure 3.12 Salt-bridge (drawn as green dashed line) between Lys21, Lys262 and Asp216 in the closed conformation of the safety-belt loop (in green). In the open conformation (shown in cyan) only Lys21 is interacting with Asp216 (blue dashed line).

3.3.1.2 Waters

The solvent content of human Aldose Reductase crystals is 43%; that means that about 2000 molecules of water should be present in the model. The problem of the bulk solvent persists also at high resolution and most waters are still disordered and form a *continuum* of electron density. At this resolution is usually possible to see ordered waters of the first and second solvation shell.

The final model contains a total of 477 water sites (Fig. 3.13), which can be divided into three groups:

- 373 fully occupied sites
- 38 pairs of partially occupied sites (Fig. 3.14)
- 66 partially occupied sites connected to disordered protein atoms or to at least two other water molecules.

Almost all the fully occupied sites were found automatically in the first stages of the refinement. Fully occupied water molecules form networks of hydrogen bonds either with atoms of the protein side chains or with other waters; consequently they have quite low B values (20 \AA^2). During the refinement, some of the waters showed an ellipsoid shape of the electron density and high B values: they were thus refined in two positions with an occupancy of 50% each. The occupancy of waters that were clashing with disordered residues or that were interacting with them through hydrogen bonds was refined to the

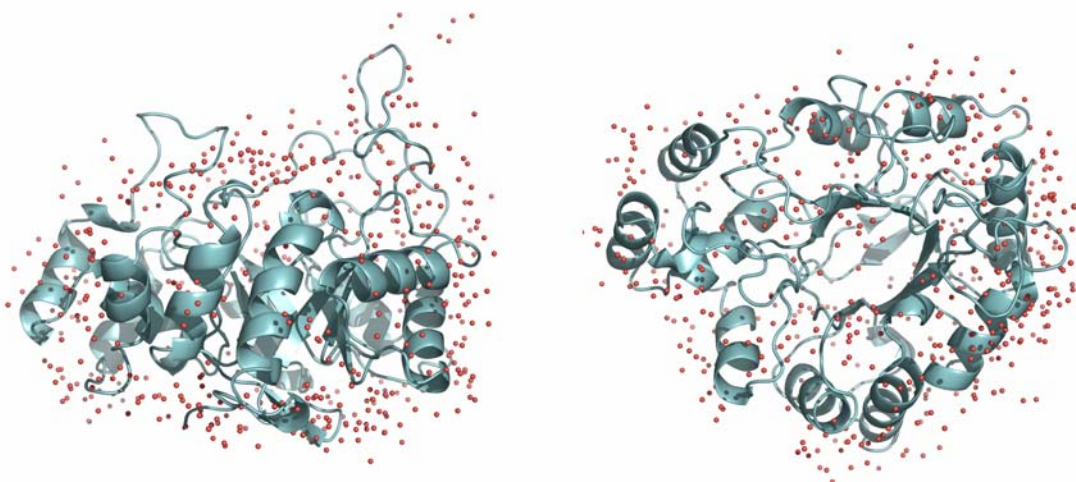


Figure 3.13 Water molecules refined in the Aldose reductase structure.

same value using a single free variable.

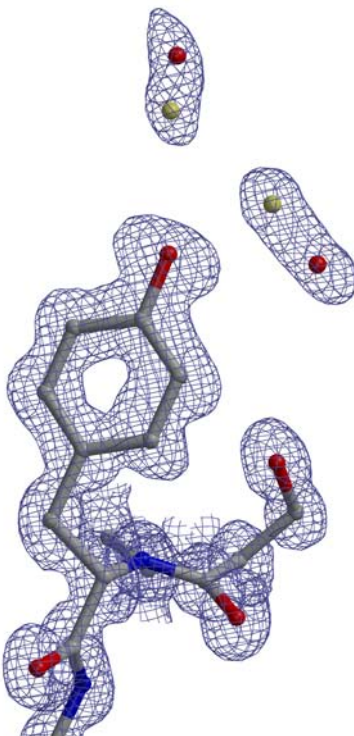


Figure 3.14 Example of partially occupied sites for two water molecules. The σ_A weighted ($2mF_o - DF_c$) density map (contoured at 1σ) has an ellipsoid shape. The two sites, each 50% occupied, are shown as red and yellow spheres.

3.3.1.3 Hydrogens

At high resolution the data-to-parameter ratio allows the addition of hydrogens to the model. In the very well ordered regions of the model with low B values hydrogens become visible in the positive difference electron density map ($F_o - F_c$) contoured at 2.5 sigma (Fig. 3.15A). In other parts of the molecule hydrogens, which have very low scattering power, were not detectable. In both cases they have been included in the calculated positions using the "riding" model. After putting the hydrogens R_{work} and R_{free} dropped down by 1.2% and 1.3%, respectively. In the difference electron

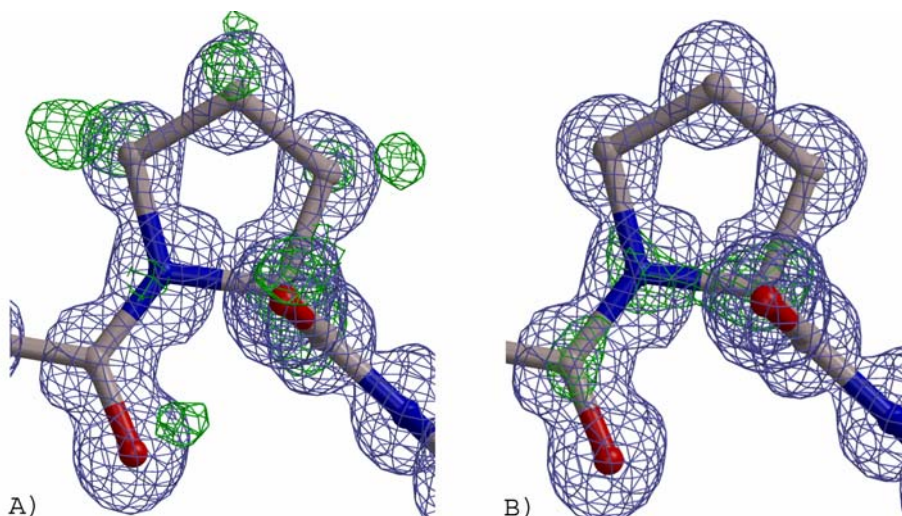


Figure 3.15 A) σ_A weighted ($2mF_o - DF_c$) density map (contoured at 1σ) in blue and difference ($F_o - F_c$) electron density map contoured at $+2.5\sigma$ (in green): most hydrogens are visible. B) The two maps after the refinement with hydrogens placed in "riding" position.

density map ($F_o - F_c$) of the well ordered regions of the protein peaks were no longer visible and the noise of the maps generally decreased (Fig. 3.15B).

Hydrogens of the hydroxyl group of Thy, Ser, Thr, hydrogens attached to the imidazole ring of His and hydrogens of the cofactor were not modeled. In the positive difference electron density map ($F_o - F_c$) it could clearly be seen that the catalytic residue His110 is singly protonated on NE1 (Fig. 3.16A). The analysis of the difference electron density maps also allowed the unequivocal identification of the redox state of the cofactor NADP⁺ (Fig. 3.16B).

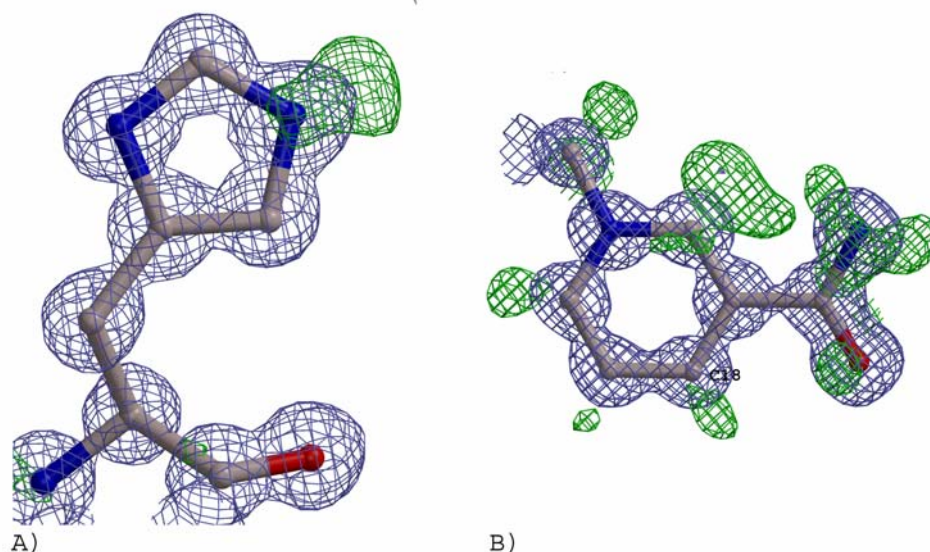


Figure 3.16 In blue is shown the σ_A weighted $(2mF_o - DF_c)$ density map contoured at 1σ and in green is shown the positive difference $(F_o - F_c)$ electron density map contoured at 2.5σ . A) His110 is singly protonated on NE1. B) The hydrogen on C18 shows that the cofactor is oxidized (NADP⁺) and in a post-reaction state.

3.3.1.4 Analysis of disagreeable restraints

Structures refined at high resolution show deviations from standard geometry such as non planar peptide bond, non planar aromatic ring for aromatic side chains, unusual values of chiral volumes and so on. In the case of Aldose Reductase a total of 268 disagreeable restraints are listed at the end of the last *conjugate gradient least-squares* refinement (see Appendix C).

41 of them are antibumping (BUMP) restraints. 10 of them involve hydrogens, 3 atoms belonging to the cofactor NADP⁺ and 12 molecules of water. The largest violations involve atoms of disordered side chains and carbonyl groups of the backbone. Tyr209 shows two violations: an excessively short distance between the C carbon atom of the carbonyl group and the CD1 of the side chain (3.18 \AA instead of 3.3 \AA) and an excessively short distance between CD1 and C17 of the cofactor (3.19 \AA instead of 3.3 \AA). These interactions are due to the bend of the aromatic ring of Tyr209 that is

interacting with the cofactor nicotinamide ring via a π -stacking interaction. The bend of the aromatic ring was visible in the electron difference map too (Fig. 3.17A/B). Also the oxygen atom of Ala299 is too near the carbon atom C of Leu300 (2.69 Å instead of 2.8 Å). In this region the molecule is distorted because of a strong hydrogen bond between the oxygen of Ala299 and the hydroxyl group of Tyr309 (O...H 2.60 Å).

There are 21 disagreeable restraints of the distance between two atoms (DFIX) and 85 of the angle distance (1-3 distance, DANG). Of these 12 involve the cofactor or the citrate molecules, 10 safety-belt loop residues and all of them involve disordered residues.

Almost all the 18 violations on the restraints of chiral volume (CHIV) involve the didahedral angle formed by N, C, CA and CD of Pro. The residues are anyway well ordered with low B values and they fit nicely in the electron density map.

The list of the 24 disagreeable flat (FLAT) restraints is more interesting. 13 of them involve peptide bonds. In recent years, with an increasing number of high resolution structures coming out, it has been noticed that generally peptide bonds are not perfectly planar (Wilson *et al.* 1998). In the structure of Aldose Reductase ω angles show a broad distribution centered around 179.23° (Fig. 3.18). The ω angle between residues 132 and 133a shows an extreme value of 159.95°. This region is disordered and in both conformations Ser133 is making strong hydrogen bonds with surrounding residues like

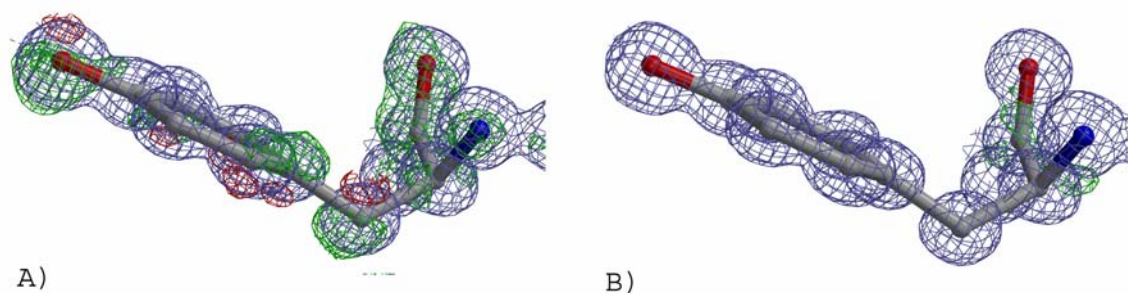


Figure 3.17 Residue Tyr209 is drawn as ball-and-stick, σ_A weighted ($2mF_o-DF_c$, contoured at 1σ) electron density map is shown in blue and difference (F_o-F_c , contoured at 2.5σ) electron density maps are shown in green (positive one) and red (negative one). A) After refinement using the FLAT restraint. B) After refinement with the FLAT restraint taken out: the ring is clearly not planar but bent.

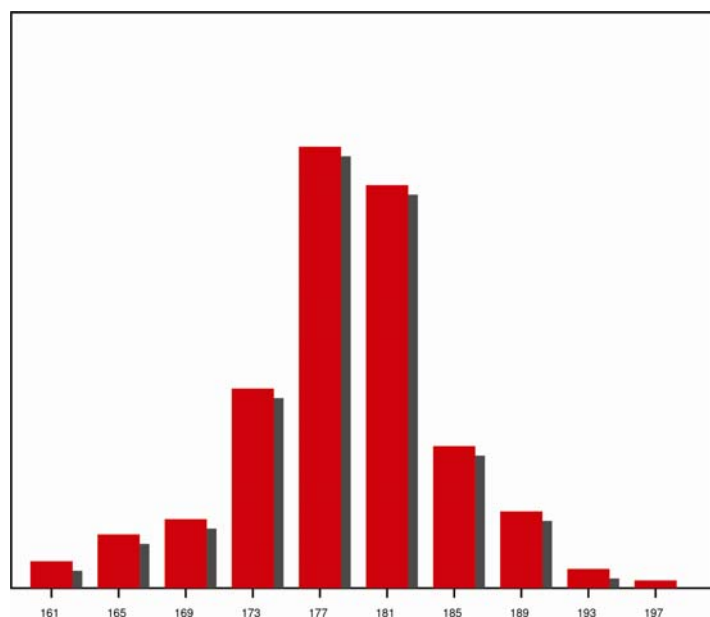


Figure 3.18 Histogram showing the distribution of the ω angle around the mean value of 179.23° .

Glu84 (2.71 \AA) and Thr135 (2.94 \AA). Even all the other disagreeable flat restraints involving peptide bonds are due to strong hydrogen bonds (the shortest being the one between the oxygen atom of Val297 and NE1 of Trp219 equal to 2.62 \AA , see also subchapters 3.3.2.2. and 3.3.2.3). Flatness disagreeable restraints are also found for the plane of aromatic rings of some side chains (one Phe, three Trp and five Tyr residues). The violation of the flat restraint for the Tyr209 aromatic ring, which interacts with the cofactor nicotinamide ring via a π -stacking interaction is particularly big (1.73 equal to 17.2 sigma).

The restraints of anisotropic parameters (DELU and SIMU) are quite well fulfilled and the violations generally involve disordered or not fully occupied sites.

3.3.1.5 Quality of the model

The agreement of the model with the data is excellent as shown by the low crystallography R values of $R_{\text{work}} = 9.8\%$ and $R_{\text{free}} = 11.2\%$ (for $F_o > 4\sigma(F_o)$). Out of the 316 residues present in the construct used for crystallization, the model of Aldose

Reductase contains 315 residues, the cofactor NADP^+ , two citrate molecules, and 555 water sites. The protein is described by 1968 fully occupied and 1033 partially occupied sites, while in the solvent region 383 fully and 142 partially occupied sites were modeled. The cofactor is fully ordered. One of the citrate molecules is fully occupied and located on the surface of the protein close to Gln49, Glu53, Asp98, Lys94; the other one is present in two conformations in the active site.

All dihedral angles of the protein backbone fall in the allowed (88.9%) and additionally allowed (10.4%) regions of the Ramachandran plot (Fig. 3.19) (Ramachandran *et al.* 1963) as defined in the program *PROCHECK* (Laskowski *et al.* 1993).

The mean B values for the fully and partially occupied sites of the protein are $7.5 \pm 3.7 \text{ \AA}^2$ and $12.1 \pm 3.7 \text{ \AA}^2$, respectively. The fully occupied waters have a mean B value of $20.3 \pm 3.7 \text{ \AA}^2$ and the partially occupied ones of $15.7 \pm 3.7 \text{ \AA}^2$. The average anisotropy (the ratio of the minimum to the maximum eigenvalue of the U_{ij} tensor, E_{11}/E_{33} (Merritt

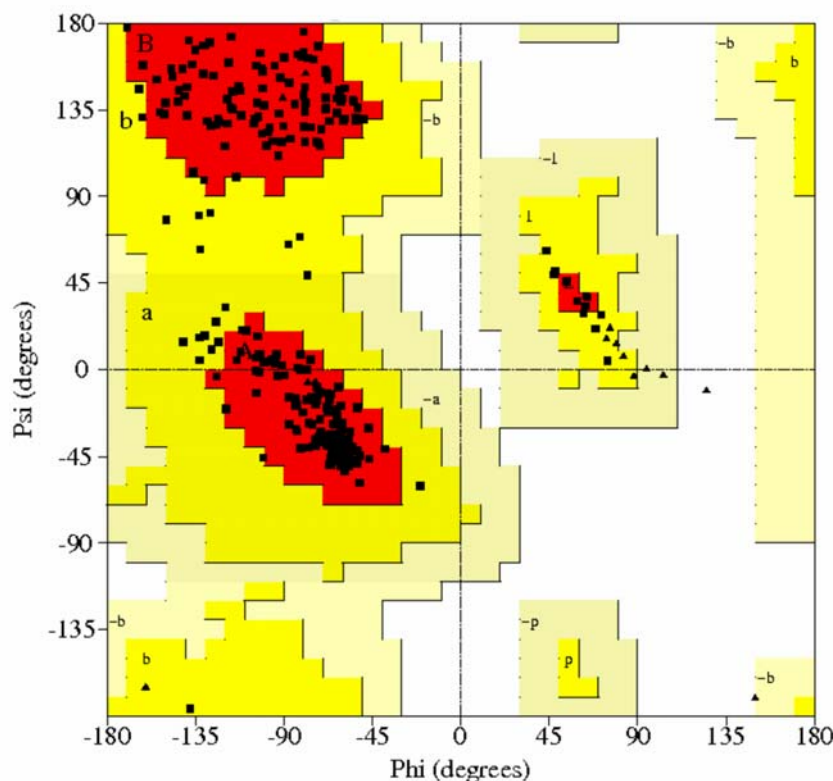


Figure 3.19 Ramachandran plot for human Aldose Reductase. Almost all residues are nicely packed in the most favored regions.

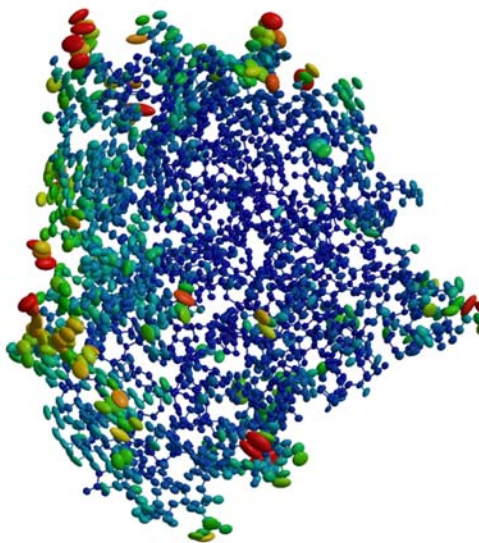


Figure 3.20 Structure of anisotropically refined human Aldose Reductase. Atoms with low B values are coloured in blue while the ones with high B values are shown in red.

1999)) is about 0.45, a normal value for high resolution structures (Fig. 3.20).

The coordinate uncertainties for fully and partially occupied sites of the protein are 0.011 Å and 0.059 Å respectively. The cofactor, which is fully occupied and in a rigid conformation, has very low ESD (0.007 Å) while the disordered citrate molecules have higher ESD (0.046 Å). The fully occupied water molecules have estimated coordinate uncertainties of 0.031 Å. Refinement statistics are shown in Table 3.4.

Resolution range (Å)	Inf - 0.82
No. of reflections	282631
$R_{\text{free}} / R_{\text{work}} [F_o > 4\sigma (F_o)]$ (%)	11.11 / 9.82
$R_{\text{free}} / R_{\text{work}}$ (all data) (%)	11.34 / 10.04
No. of reflections $[F_o > 4\sigma (F_o)]$	265969
$R1$ for $F_o > 4\sigma (F_o)$ (%)	9.84
$R1$ for all data (%)	10.06
No. of parameters	32562
No. of refined non-H sites	
Total (full / partial)	2087 / 1222
Protein (full / partial)	1968 / 1033
Cofactor (full / partial)	48 / 0
Water (full / partial)	383 / 142
Citrate (full / partial)	0 / 39
R.m.s. deviations from ideal values	
Bond distances (Å)	0.017
Angle distances (Å)	0.045
Planar groups (Å)	0.417
DELU (Å ²)	0.006
SIMU (Å ²)	0.034
ISOR (Å ²)	0.082
Average isotropic B factors	
Protein (full / partial) (Å ²)	7.5 / 12.1
Cofactor (full / partial) (Å ²)	5.6 / -
Water (full / partial) (Å ²)	20.3 / 15.7
Citrate (full / partial) (Å ²)	- / 12.4
Average coordinate uncertainties	
Protein (full / partial) (Å)	0.011 / 0.059
Cofactor (full / partial) (Å)	0.007 / -
Water (full / partial) (Å)	0.031 / 0.041
Citrate (full / partial) (Å)	- / 0.046

Table 3.4 Statistics for the refinement of human Aldose Reductase ($R1 = \sum \|F_o\| - \|F_c\| / \sum \|F_o\|$).

3.3.2 Structure

3.3.2.1 Overall structure

The overall structure of human Aldose Reductase is the same as in the previously determined structures of the molecule (Fig. 3.21). It folds as an eight-stranded β/α barrel with a small β -sheet on the N-terminal (residues 2-14) and a C-terminal extension (residues 275-315). The active site is located at the end of the barrel and the NADP⁺-binding site is near the hydrophobic active site pocket. The cofactor is bound to the protein and the active site contains one citrate molecule and several molecules of water. The most interesting features of the structure are the safety-belt loop modeled in a close conformation and a partially open one, and the presence of a disordered citrate molecule in the active site.

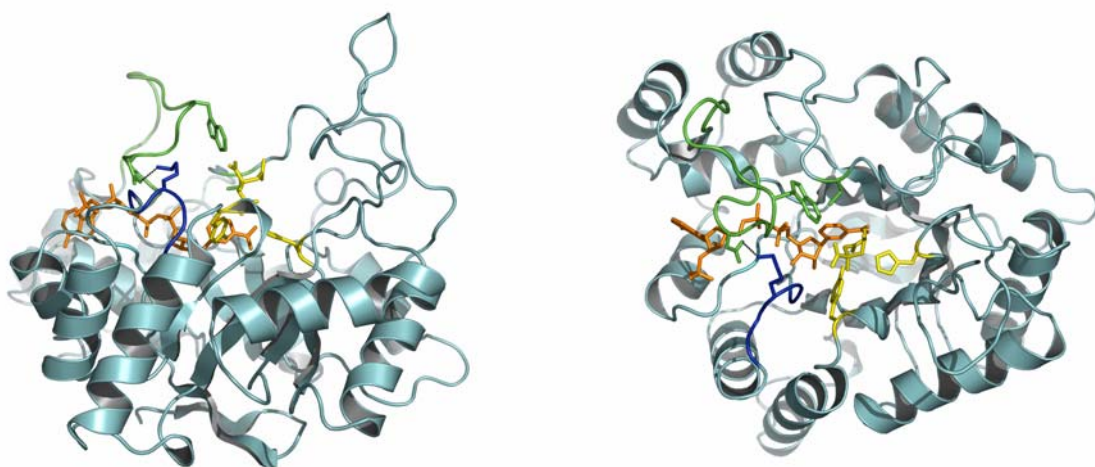


Figure 3.21 Overall structure of human Aldose Reductase. The safety-belt loop (residues 213-226) and Cys298 are shown in green; for Asp216, Trp219 and Cys298 the side chains are depicted. The loop opposing the safety-belt loop (containing residue Lys21 that makes a crucial salt-bridge to Asp216 in the closed conformation of the safety-belt loop) is shown in blue. The cofactor and the ligand are shown in orange and yellow, respectively. Side chains of Tyr48 and His110 are shown as yellow sticks.

3.3.2.2 Active site

Although the crystals used in this study were obtained by cocrystallization of the protein with the cofactor and the stereoisomer 2R4R of the inhibitor Fidarestat (El-Kabbani *et al.* 2004a), it was not possible to model the difference electron density present in the active site by placing fully or partially occupied copies of the inhibitor molecule. After careful inspection, it was decided to model the density as two overlapping partial conformations of citrate (which was present in the crystallization buffer) leading to a satisfactory explanation of all peaks in the difference electron density map (Fig. 3.22A/B).

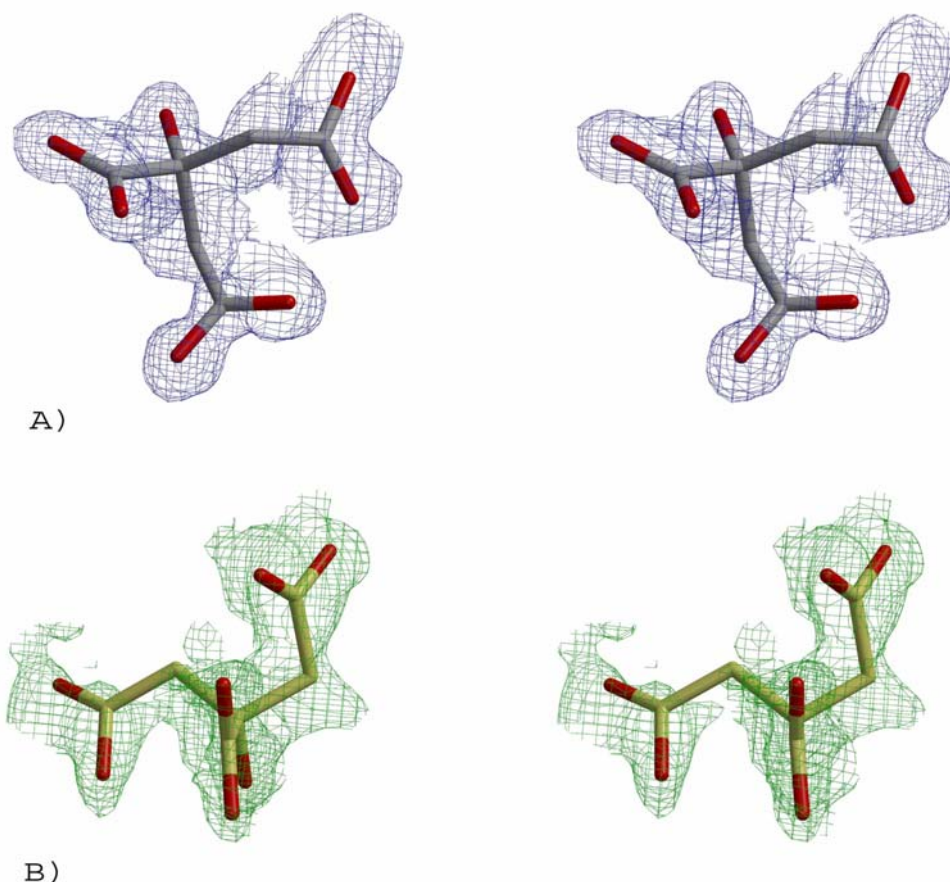


Figure 3.22 Crossed-eyes stereo views of electron density for citrate bound in the active site. A) σ_A weighted ($2mF_o - DF_c$) electron density map (contoured at 1σ) and the main conformation of the citrate (occupancy = 64%). B) Difference ($F_o - F_c$) electron density map (contoured at 2.5σ) as calculated in the presence of the citrate main conformation but before adding the minor conformer (occupancy = 36%). The refined position of the second conformation is shown.

As in many other determined structures (Bohrani *et al.* 1992; Wilson *et al.* 1992; Wilson *et al.* 1993; Bohren *et al.* 1994; Harrison *et al.* 1994; Harrison *et al.* 1997; Calderone *et al.* 2000; Kinoshita *et al.* 2002; El-Kabbani *et al.* 2003; El-Kabbani *et al.* 2004a; El-Kabbani *et al.* 2004c; Howard *et al.* 2004; Ruiz *et al.* 2004), the catalytic residues Tyr48 and His110 interact with a carboxylate group of the ligand, in this case a molecule of citrate. In both conformations the citrate molecule is indeed anchored to the active site via hydrogen bonds to Tyr48 and His110 (Table 3.5 and Fig. 3.23A/B). Except for this strong interaction the second conformation is interacting only weakly with Cys298 via a normal hydrogen bond and with Trp20 via a π -hydrogen bond.

Instead the main conformation interacts with residues situated in different parts of the active site, connecting them to each other. The same carboxylic group that interacts with Tyr48 and His110 is making a π -hydrogen bond with the cofactor nicotinamide ring and a normal hydrogen bond with the nitrogen atom NE1 of Trp111. The carboxylic group on the other side of the citrate molecule is acting as a bridge between Trp20 and Trp219 of the closed conformation of the safety-belt loop. The O6 atom is located in the plane of the Trp219 imidazole at a distance of 4.41 Å from the center of mass of the six-membered ring of imidazole, indicating an interaction between O6 and the edge of the aromatic system (Meyer *et al.* 2003). The geometry of this interaction is in good agreement both with the results of quantum mechanical calculations for the interaction between oxygen and the edge of benzene (Thomas, 1982) and with statistical analyses of

Citrate conformation A			Citrate conformation B		
atom citrate	atom	distance (Å)	atom citrate	atom	distance (Å)
O3	C18_318	3.29	O1	O_5039b	3.22
	OH_48	3.14	O2	O_4033b	2.89
	NE2_110	2.81	O3	C18_318	3.32
O4	C18_318	4.14	O4	C18_318	3.13
	OH_48	2.71		NE2_110	2.72
	NE2_110	3.34	O5	O_3143	2.30
	NE1_111	2.99		O_4017b	2.74
O5	NE1_20	2.95		O_4021b	3.33
O6	CH2_219a	3.21		O_4033b	3.22
			O6	SG_298	3.22

Table 3.5 Interactions of the two conformations of the citrate molecule in the active site with surrounding residues and water molecules.

protein structures (Hakansson 1996).

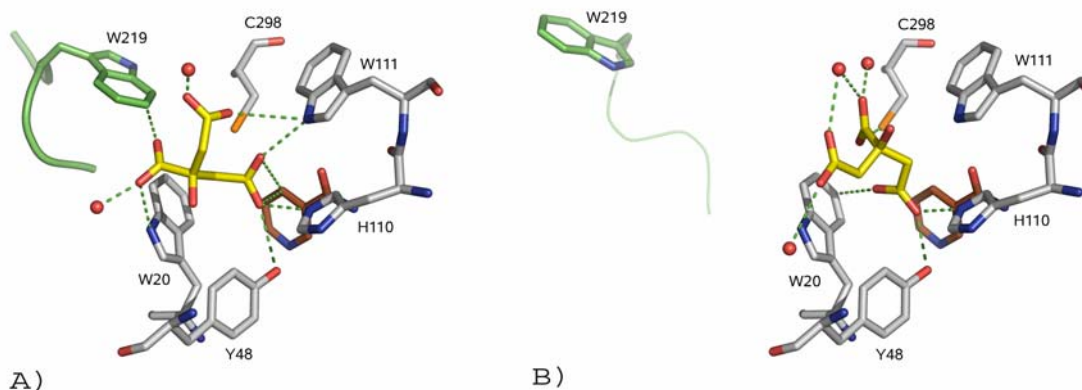


Figure 3.23 Interactions of the citrate (shown in yellow) in the active site. Residues belonging to the protein are drawn in grey (safety-belt loop in green), the nicotinamide ring of the cofactor in orange. A) Main conformation. B) Second conformation.

3.3.2.3 The safety-belt loop

Correlated with the two conformations of citrate in the active site, the safety-belt loop is also present in two conformations. For the major conformation, the entire polypeptide chain between Gly213 and Ser226 could be placed into electron density, although for residues Asp216 to Asp224, the atomic sites are only partially occupied. The structure of the major conformation (occupancy = 64%) of the safety-belt corresponds to the closed conformation present in most of the structures of the holo-enzyme and it is characterized by hydrogen bonds between the carboxylate of Asp216 and the amino-groups of Lys21 and Lys262 (Wilson *et al.* 1992) with lengths of 2.95Å and 2.82Å, respectively. For the minor conformation (occupancy = 36%), only residues Asp216 to Trp219 are visible in the electron density (Fig. 3.24). This intermediate open conformation is stabilized by interactions between Pro218 and Trp219 with Asn256 of a symmetry-related molecule: Pro218 makes a hydrogen bond via its carbonyl oxygen to

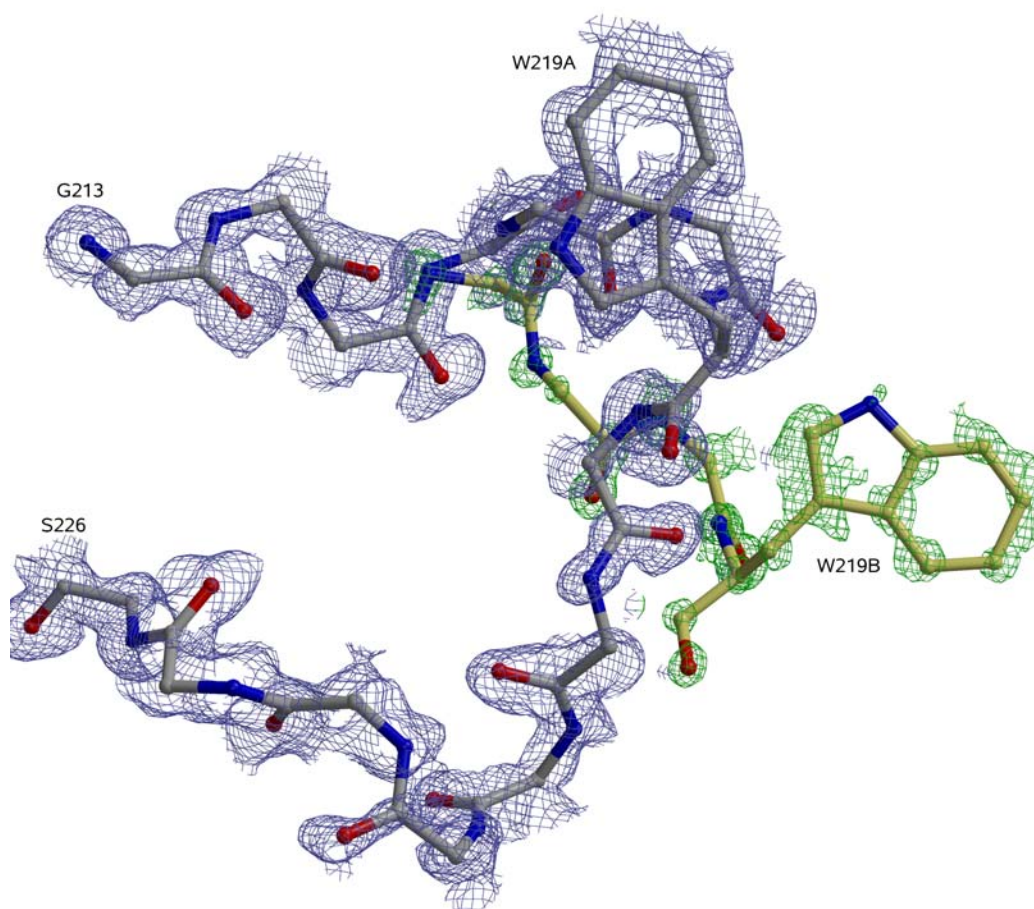


Figure 3.24 The σ_A weighted ($2mF_o - DF_c$, contoured at 1σ) map and the difference map ($F_o - F_c$, contoured at 2σ) as calculated in the presence of the safety-belt loop main conformation (drawn in grey) but before adding the minor conformer, are drawn in blue and green respectively. The refined position of the second conformation of the safety-belt loop is shown in yellow.

the amide nitrogen while the imidazole of Trp219 makes a π -stacking interaction with the guanidinium group of Asn256 (Fig. 3.25).

In terms of opening of the safety-belt the minor conformation seen here corresponds to an intermediate conformation between the closed and the open conformation as observed in the mutant Arg268Ala of Aldose Reductase by Bohren *et al.* (2005) as residues Gly213 to Pro215 remain in the closed conformation and the region between Ala220 and Ser226 is disordered (Fig. 3.26)(Bohren *et al.* 2005). The necessary but not sufficient condition

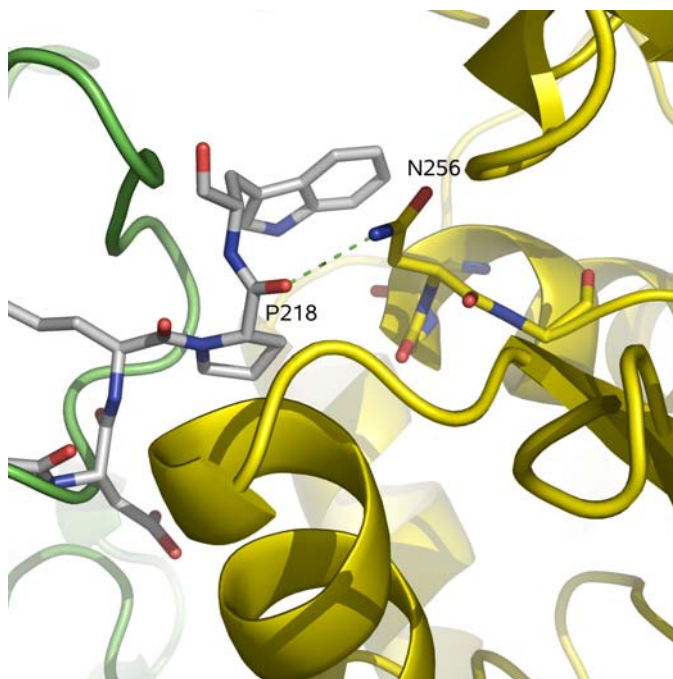


Figure 3.25 Interaction between the open conformation of the safety-belt loop in grey (Pro218) and a symmetry related molecule in yellow (Ans256).

for the opening of the safety-belt loop is the possibility for the sulphur atom of Cys298 to interact with the positive charged nicotinamide ring of NADP^+ . It is known (Duan *et al.* 2001) that Cys can interact with aromatic rings or through π -hydrogen bonds or through electrostatic potential attraction between the negative charged lone pairs of the sulphur atom and the positive electrostatic potential of the aromatic ring. In all but one of the structures solved until now the safety-belt loop is in the closed conformation. In some structures the inhibitor binds in the specificity pocket (residues Thr113, Phe115, Phe122, Cys303, Tyr309) making the whole molecule and consequently the loop 213-226 more rigid (PDB-ID: 1pwl (El-Kabbani *et al.* 2004b), 1us0 (Howard *et al.* 2004), 1t40, 1t41 (Ruiz *et al.* 2004)). When the inhibitor interacts strongly with Leu300 (PDB-ID: 1iei (Kinoshita *et al.* 2002) and 1pwm (El-Kabbani *et al.* 2004b)) the whole molecule becomes more rigid. In other two structures Cys298 (PDB-ID: 1iei and 1pwl) strongly interacts with the inhibitor thereby excluding a possible interaction with the nicotinamide

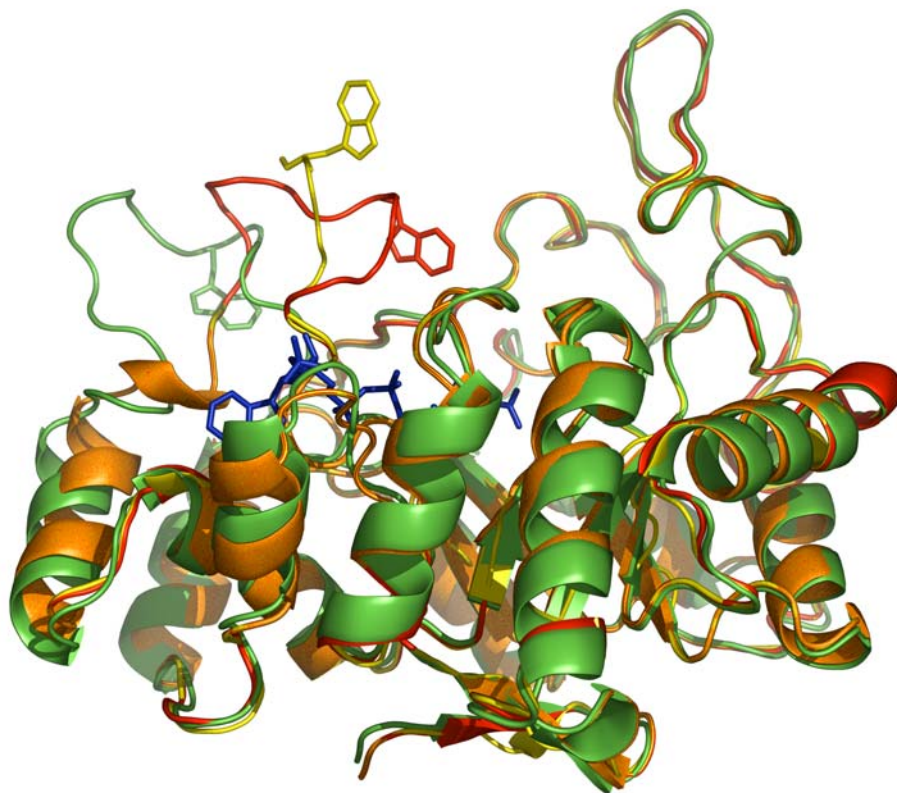


Figure 3.26 Safety-belt loop in different conformations. The closed and open conformation of the present model are shown in red and yellow, while the open conformation found in the structure of the R268A mutant apo-AR is shown in green. The side-chain of Trp219 is drawn as sticks and the cofactor is depicted in blue.

ring of the cofactor NADP^+ . If Trp219 (PDB-ID: 2acq (Harrison *et al.* 1994), 2acu (Bohren *et al.* 1994) and 1el3 (Calderone *et al.* 2000)) or Pro218 (1el3) are making strong interactions with the inhibitor, the safety-belt loop cannot be in the open conformation. For other structures the open conformation could be possible: for some of them the electron density map for that region (checked using the Uppsala Electron Density Server, <http://fsrv1.bmc.uu.se/eds>, (Kleywegt *et al.* 2004)) is not clear (PDB-ID: 1az1 (Harrison *et al.* 1997)). For others, for which no structure factors are deposited, the B values of the residues 213-226 are extremely high (PDB-ID: 1ads, 2acs and 2acr (Harrison *et al.* 1994)). In the structure of the apo Arg268Ala Human Aldose Reductase (PDB-ID: 1xgd,

(Bohren *et al.* 2005)) the safety-belt loop is in the open conformation. In this case the affinity of the enzyme for the cofactor is decreased by a point mutation of residue Arg268 to Ala. The cofactor is not present in the crystal structure and Cys298 can be involved in an aromatic sulphur interaction with Trp22. The open conformation, where the safety-belt loop is displaced by 17 Å at most, is stabilized through an aromatic π -stacking interaction between the aromatic ring of Trp219 and the guanidinium group of Arg293. The movement of the safety-belt seems not to be a rigid one but probably the exposed loop is moving freely in the solvent (Table 3.6).

pdb code	distance (Å) C ^{α} 213-C ^{α} 216	distance (Å) C ^{α} 216-C ^{α} 219	distance (Å) C ^{α} 219-C ^{α} 222	distance (Å) C ^{α} 222-C ^{α} 225
2acq	8.74	8.36	10.43	8.29
2acr	8.66	8.50	9.98	8.25
2acs	8.73	8.51	7.12	9.32
2acu	8.67	8.42	10.11	7.85
1ads	8.83	8.58	8.51	6.93
1az1	8.67	8.31	9.05	9.04
1az2	8.41	7.90	9.79	9.75
1ef3 (chainA)	8.80	8.60	10.00	8.02
1el3	8.73	8.48	9.34	9.42
1iei	9.08	8.18	8.56	7.02
1mar	8.75	8.53	9.83	8.49
1pwl (chainA)	8.67	8.74	9.82	8.34
1pwm (chainA)	8.77	8.40	10.10	8.47
1t40	8.70	8.50	9.63	9.17
1t41	8.70	8.57	9.78	8.43
1us0	8.67	8.35	9.92	8.58
1x96	8.77	8.48	10.03	8.19
1x97	8.78	8.51	9.95	8.30
1x98	8.76	8.52	10.11	8.48
1xdg	8.88	9.00	9.89	8.17
cit. (conf.A)	8.77	8.54	10.16	8.44

Table 3.6 Distances between C ^{α} atoms of residues of the safety-belt loop. The distances are different in the different structures showing that the movement of the loop is probably not rigid.

3.3.2.4 The tongs movement

The structure of Aldose Reductase presented here was compared with other four high resolution structures of Aldose Reductase (PDB-ID: 1pwm, 1pwl, 1us0, 1t41). A qualitative comparison was carried out by visual inspection of the models in *PYMO*L (DeLano 2002). Since studies on the different binding modes of inhibitors and on the so-called induced-fit phenomena have already been made (Klebe *et al.* 2004; Podjarny *et al.* 2004; Sotriffer *et al.* 2004), the aim of this comparison was to detect the possible presence of domains movements.

The overall structures of 1pwl, 1pwm and 1t41 are very similar to each other while they differ from the ones of 1us0 and the one described in this paper (abbreviated as cit). The rigid β -sandwich (domain S3, S4, S5, S6, S7, S8, S9 according to the topology diagram of AR (Harrison *et al.* 1994), Fig. 3.27) is the same in all five structures as well as the strand S2, part of the helices H3 and H4, the loop between H4 and S7, the helices H7, H8 and the loop connecting the two.

In the structure 1pwl the loop 112-137 (between S6 and H4) and the loop 214-231 (between S9 and H7) are much closer to each other than in 1us0 and cit (Table 3.7).

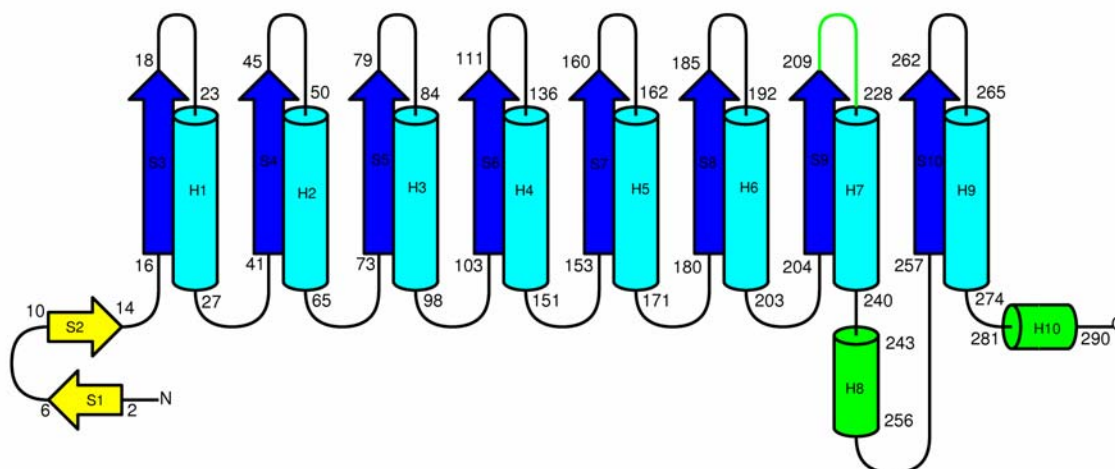


Figure 3.27 Topology diagram of Aldose Reductase. The β/α barrel is shown in dark and light blue, the safety-belt loop between S9 and H7 and two extra helices are shown in green, and the two short β strands at the N-terminal side in yellow.

res. A	res. B	distance (A-B) Å			Δ distance(A-B) Å	
		1pwl (closed)	1us0 (medium)	Cit (open)	1pwl-1us0	cit -1us0
D224	N129	24.19	25.27	26.03	-1.08	0.76
E223	G128	24.56	26.11	26.81	-1.55	0.8
P122	S127	26.42	28.17	29.61	-1.75	1.44
K221	E126	25.12	26.8	28.12	-1.68	1.32
W219	G128	15.64	17.16	18.13	-1.52	0.97
average Δ dist(A-B)					-1.52	1.06
		1pwl (open)	1us0 (closed)	cit (med.)	1pwl-1us0	cit-1us0
K172	S282	39.01	37.34	37.45	1.67	0.11
	Q283	41.06	39.6	39.7	1.46	0.1
P173	S282	41.38	40.53	40.69	0.85	0.16
	Q283	43.62	42.9	43.04	0.72	0.14
G174	S282	42.52	41.12	41.3	1.4	0.18
	N283	44.87	43.72	43.88	1.15	0.16
L175	S282	39.13	37.71	37.89	1.42	0.18
	N283	41.48	40.34	40.1	1.14	0.24
average Δ dist(A-B)					1.23	0.16
		1pwl	1us0	cit	1pwl-1us0	cit-1us0
G18	D43	5.01	5.03	5.02	-0.02	-0.01
L17	D43	5.89	5.93	5.94	-0.04	0.01
G16	I42	5.27	5.21	5.25	0.06	0.04
L15	H41	6.76	6.79	6.79	-0.03	0
D43	V75	4.98	4.96	4.96	0.02	0
I42	I74	5.21	5.22	5.21	-0.01	-0.01
H41	F73	4.52	4.54	4.54	-0.02	0
average Δ dist(A-B)					0.02	0.01

Table 3.7 Distances between residues of the regions of the molecule involved in the "tongs" movement.

On the opposite side of the molecule, the helix 280-289 (H10) and the region 164-175 (loop after S7 and part of H5) are farther away from each other in 1pwl than in 1us0. In the present model this part does not move so much (see Fig. 3.28). It can be asserted that a domain movement has been detected, where the upper part is closing, the central one is motionless and the lower part is opening. This domain movement as been called "tongs movement". The loops involved in this movement are quite flexible, far from the active site and they show high B values. Therefore this movement is probably not related with the size or the chemical structure of the inhibitor. The inhibitors Minarelstat in 1pwl and IDD594 in 1us0 are indeed similar: both interact with the specificity pocket via a

Bromine and a Fluorine atom on the benzyl ring and both make interactions with the anion binding site, the first one through the cyclized imide ring and the second through the carboxylic head. It could also be that the loops show different positions because of their intrinsic flexible nature or the movement could be just a crystal packing artefact. Inspection of the models did not show any strong interactions between the loops and symmetry related molecules disproving this last hypothesis.

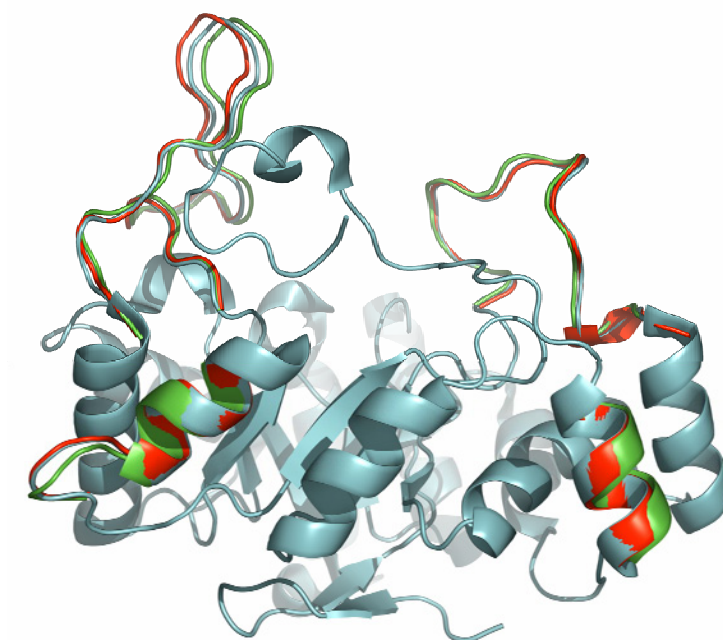


Figure 3.28 Superposition of 1us0 (in cyan), 1pwl (in red) and cit (in green) that shows the loops involved in the "tongs" movement.

3.4 Conclusions and perspectives

In this chapter the structure of human Aldose Reductase in complex with citrate to a resolution of 0.82 Å has been described. The atomic resolution data allowed to model more multiple conformations than in the previously determined structure (88 vs. 47) and to improve the fit between model and the data (R_{free} 11.2% vs. 20.1%).

Particular attention was paid to the inspection of the list of disagreeable restraints. Some of them (chiral volume of nitrogen atom of Pro) are probably due to wrong restraints while others (flatness of the peptide bond or of the ring of some aromatic residues) are real features of the structures observable only at high resolution and probably caused by strong interactions. These observations should be used to build a new library of restraints, to explain in more details the structure and to better interpret the role of the molecule.

The most interesting feature of this model is the safety-belt loop modeled in two discrete conformations. The common closed conformation with an occupancy of 64% is stabilized by the main conformation of the citrate molecule in the active site that connects Trp219 to residues of the active site, some of which (Tyr48 and His110) are known to be very rigidly placed. The partially open conformation was possible because the holo-enzyme is in a post-reaction state. Indeed, from the hydrogen atoms present on the nicotinamide moiety of the cofactor the reduced state of it was confirmed. In the minor conformation, the citrate molecule does not make an interaction with Trp219. This conformation of the citrate is correlated with a 'half-open' conformation of the safety-belt loop, indicating that the removal of one interaction is enough to trigger the release of the cofactor. Such a transition, from a situation where the safety-belt is kept closed due to interactions with substrate or product of the catalytic reaction to a situation where the safety-belt can open because important interactions can not be kept after release of the product, could well be the structural basis for the regulation of cofactor release.

Careful inspection of different high resolution structures reveals a domain movement that has been called "tongs movement": loops at the top and at the bottom are closer to or farther away from each other, while the β -barrel stays rigid. This movement could be related to the post reaction state of the molecules and to the release of the cofactor.

In the context of drug-design, the present study shows that Trp219 in the safety-belt in its closed conformation provides an interaction point for ligands. This interaction could be exploited in the construction of AR inhibitors to achieve an increase of both affinity and specificity.

4 C2A domain of Rabphilin3A

4.1 Introduction

In this chapter the crystal structure of the C2A domain of Rabphilin3A to a resolution of 1.92 Å is presented (PDB-ID: 2ch0, (Biadene *et al.* 2006)). Rabphilin3A is a neuronal protein implicated in the synaptic vesicle cycle and its exact biological role still remains to be elucidated. The structure presents new features with respect to other C2 domains suggesting a new function for this domain.

4.1.1 The synaptic vesicle cycle

Signal transduction processes can be carried out under hormonal control or via synaptic transmission of nervous impulses. Inside a single neuron the impulse is transmitted by an action potential generated through the polarization and depolarization

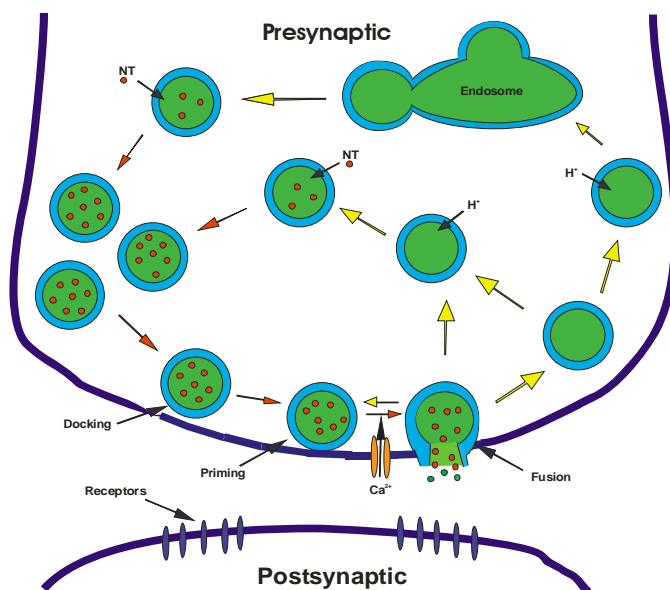


Figure 4.1 Scheme of the synaptic vesicle (green circles) trafficking. Red arrows show exocytosis steps while yellow arrows show the different possibilities for the endocytosis process. The small red circles represent neurotransmitters molecules (NT).

of the membrane. When the impulse has to go through different neurons, it has to cross the synaptic junction. Neurotransmitter release is then mediated by exocytosis and endocytosis of synaptic vesicles at the presynaptic active zone of the nerve terminal. Exocytosis is triggered by Ca^{2+} : Ca^{2+} ions enter through Ca^{2+} channels and bind to Ca^{2+} sensors inducing fusion-pore opening. The endocytosis follows three different possible routes: the vesicles can be refilled without undocking, they can undock and then be recycled locally or they can be refilled through a clathrin intermediate (Fig. 4.1, for review (Sudhof 2004)). Synaptic vesicles contain two major classes of proteins: the ones involved in neurotransmitter up-take (called transport proteins) and the ones that participate in exo-, endocytosis and recycling (named trafficking proteins) (Fig. 4.2). Transport proteins form a large multiprotein complex that works as a proton pump that, generating an electrochemical gradient, up-takes neurotransmitter molecules (for review (Fykse *et al.* 1996)). Trafficking proteins can be intrinsic membrane proteins, proteins associated via posttranslational lipid modification and peripherally bound proteins. The four different classes of soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) proteins characterized by the so called SNARE motif (70 residues) take part in intracellular membrane fusion (for review (Chen *et al.* 2001; Jahn *et al.* 2003)). The function of SNARE proteins is regulated by a class of essential fusion

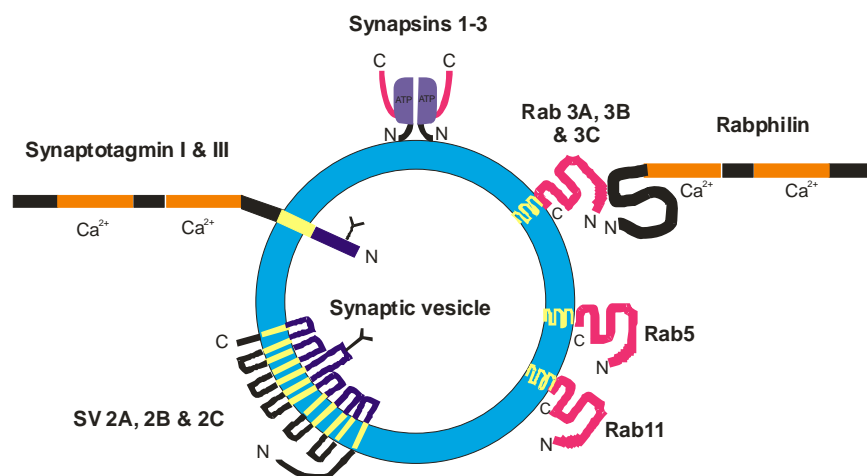


Figure 4.2 Schematic representation of a synaptic vesicle that shows the most important trafficking proteins. The C2 domains are shown in orange and the intermembrane regions in yellow. For simplicity Rabphilin has been associated only to the protein Rab3A, B and C.

proteins called Sec1/Munc18-like (SM) proteins (Jahn *et al.* 2003) and by Synaptophysins (Johnston *et al.* 1990; Calakos *et al.* 1994; Edelmann *et al.* 1995; Washbourne *et al.* 1995). Synaptotagmin I and II are involved in endocytosis and synaptic fusion: they can bind up to five Ca^{2+} ions through two C2 domains (Ubach *et al.* 1998; Fernandez *et al.* 2001), interact with SNARE complexes (Bennett *et al.* 1992; Chapman *et al.* 1995; Li *et al.* 1995a; Shin *et al.* 2003) and form calcium/phospholipid complexes. Rab3 (Schluter *et al.* 2002), Rab5 (von Mollard *et al.* 1994) and Rab11 (Khvotchev *et al.* 2003) are regulators of synaptic vesicles fusion and they can bind to the effectors Rabphilin (Shirataki *et al.* 1993; Li *et al.* 1994) and RIM1 α /2 α (Wang *et al.* 1997; Wang *et al.* 2000; Wang *et al.* 2003) that are located in the active zone and that in turn interact directly or indirectly with many other synaptic proteins. The function of Synaptic vesicle proteins 2a (SV2s) (Buckley *et al.* 1985) is not yet fully understood but they may regulate Ca^{2+} levels (Janz *et al.* 1999). Finally Synapsins (see Fig. 4.2) are probably required to maintain the right number of synaptic vesicles and to regulate short-term synaptic plasticity (Rosahl *et al.* 1993; Li *et al.* 1995b; Rosahl *et al.* 1995).

4.1.2 Rabphilin3A

Rabphilin3A is a peripheral membrane protein found on synaptic vesicles. It is expressed in neurons and neuroendocrine cells. It has a relative molecular mass of 78 kDa and comprises 710 residues (Shirataki *et al.* 1993). The N-terminal region (residues 40-170) contains two α -helical regions and a cysteine rich zinc-finger domain that can interact with Rab3A-GTPase (Stahl *et al.* 1994; McKiernan *et al.* 1996; Chung *et al.* 1997), Rab27 (Fukuda 2003), Rabaptin5 (Ohya *et al.* 1998) and α -actinin (Kato *et al.*

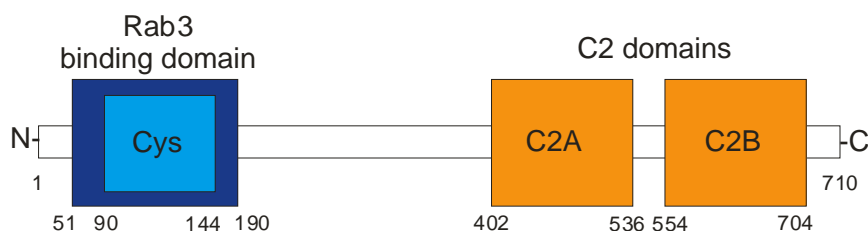


Figure 4.3 Schematic representation of the domain organization of Rabphilin3A.

1996; Baldini *et al.* 2005). Next a long linker region of about 120 residues, the C-terminal part (residues 396-704) contains two C2 domains (homologous to the ones of Synaptotagmin) which can bind phospholipids in a Ca^{2+} -dependent manner (Yamaguchi *et al.* 1993; McKiernan *et al.* 1996; Chung *et al.* 1998; Ubach *et al.* 1999) and β -adducin (Miyazaki *et al.* 1994) (Fig. 4.3). The function of Rabphilin3A is not yet fully understood. Studies carried out in recent years showed that Rabphilin3A was involved in neurotransmitter release with a subsection of reactions and that it could eventually be considered a positive regulator of exocytosis (Chung *et al.* 1995; Burns *et al.* 1998). An additional role in endocytosis through interaction of the N-terminal region with Rabaptin5 was proposed by Ohya (Burns *et al.* 1998; Ohya *et al.* 1998). On the opposite, later studies based on genetic mutations considered Rabphilin3A as a non-specific Rab3 effector (Schluter *et al.* 1999; Staunton *et al.* 2001). Rabphilin3A needs Rab3 for stability and synaptic location but the *viceversa* does not apply. At the same time Rab3A can also work perfectly in the absence of Rabphilin3A. Moreover the role of Rabphilin3A as a major synaptic Ca^{2+} sensor in exocytosis was excluded. It was instead proposed that the function of the C2 domains is probably independent from the binding to Rab3A and that Rabphilin3A has not a fundamental but a specialized role under extreme physiological conditions (Schluter *et al.* 1999). Fukuda (Fukuda 2003) proposed Rabphilin3A as an effector of Rab8A, Rab27A and Rab3A. Other studies claimed that Rabphilin3A could be involved in the docking step of regulated exocytosis through a direct interaction between the C2B domain and the plasma membrane protein Synaptosome-associated protein of 25 kDa (SNAP25), a target SNARE (Tsuboi *et al.* 2005). A function of Rabphilin3A as a linker for Rab3A and different cytoskeletal proteins has been proposed. Rabphilin3A interacts through the C2 domain with β -adducin (Miyazaki *et al.* 1994). The protein α -actinin instead interacts with the N-terminal region and this interaction stimulates the activity of α -actinin to cross-link actin filaments into a bundle (Kato *et al.* 1996; Baldini *et al.* 2005). In the light of all these observations the exact biological role of Rabphilin3A still remains to be elucidated.

4.1.3 C2 domains

C2 domains are intracellular protein modules about 130 residues long with a structure characterized by a pair of four-stranded β -sheets that form an eight-stranded antiparallel β -sandwich. They adopt alternative type I (if located more towards the C-terminus) and type II (if located more towards the N-terminal region) connectivities that differ by circular permutation of their topology (Fig. 4.4, for review (Nalefski *et al.* 1996; Rizo *et al.* 1998)). The loops between strand $\beta 2$ and $\beta 3$ (called CBL1) and the one between $\beta 6$ and $\beta 7$ (called CBL3) form a bidentate Ca^{2+} binding site (Sutton *et al.* 1995; Essen *et al.* 1996). Two to three Ca^{2+} ions can be coordinated with low intrinsic affinity (Shao *et al.* 1996; Ubach *et al.* 1998; Fernandez *et al.* 2001) but the binding does not induce large conformational changes (Shao *et al.* 1998). C2 domains are widely distributed in eukaryotes and the proteins in which they are found can be divided into four classes: Ca^{2+} -sensory vesicular transport proteins (Synaptotagmin, Rabphilin, Double C2-like domain-containing proteins), lipid-modifying enzymes (Cytoplasmic phospholipase A₂, Phosphoinositide-specific phospholipase C, Plant phospholipase D), protein kinases regulated by Ca^{2+} or phospholipids (Protein kinase C, Protein kinase C-

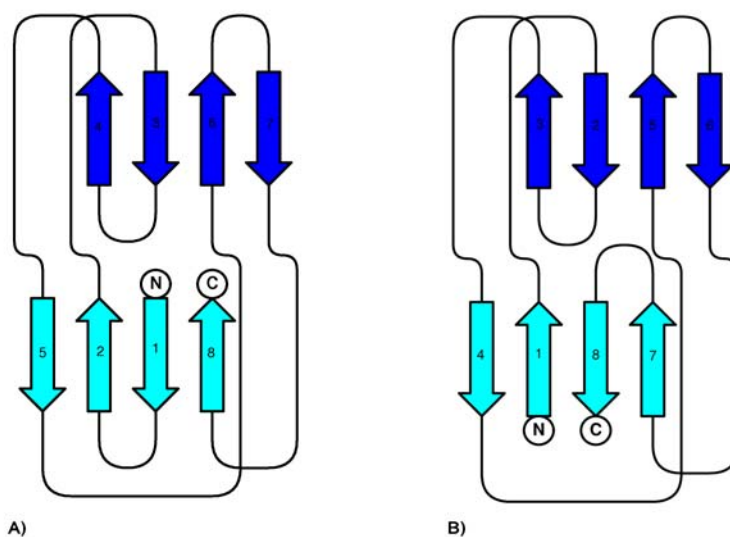


Figure 4.4 The two different topologies found in C2 domains. A) Topology I (found also in Rabphilin3A). B) Topology II.

like kinase, Serine/Threonine-protein kinase SCH9) and GTPase activating proteins (Ras-GTPase-activating protein, Ras-GTPase-activating protein 1). C2 domains seem to have diverse functions. They can bind Ca^{2+} and can interact with cellular membranes and mediate both protein-protein interactions and small molecule binding. In the structure of the complex of Protein Kinase C α (PKC α) with phosphatidyl serine the Ca^{2+} ions act as a bridge between the ligand and the membrane (Verdaguer *et al.* 1999). A two steps mechanism has been proposed in which the first Ca^{2+} ion triggers the interaction with the negatively charged phosphatidyl serine while the other calcium binding region (CBL3) penetrates into the phospholipid bilayer.

C2 domains can act as a Ca^{2+} -dependent electrostatic switch (Shao *et al.* 1997). The interaction between Synaptotagmin I and Syntaxin via CBL3 is mediated by Ca^{2+} binding. In the absence of Ca^{2+} the calcium binding site formed by five negatively charged Asp residues is surrounded by four basic residues and the electrostatic potential of the region is zwitterionic. The electrostatic surface of this region becomes positive upon Ca^{2+} binding and consequently the interactions with the Asp rich syntaxin become stronger.

Rabphilin3A has two C2 domains in its C-terminal region: C2A that includes residue 396 to residue 530 and C2B that includes residue 548 to residue 669. The structure of the C2B domain has been elucidated both by NMR (Ubach *et al.* 1999) and X-ray crystallography (Schlicker *et al.*, personal communication). The C2 domain tandem, characteristic also for the synaptotagmin protein family, shows a functional cooperativity and it is essential for the function of these proteins (Bai *et al.* 2004). Elucidating the structural features of the C2A domain will then help to highlight functional properties of Rabphilin3A.

4.2 Materials and methods

4.2.1 Crystallization and data collection

For crystallization the protein was concentrated in 10 mM HEPES, pH 7.0, 150 mM NaCl, 1 mM DTT to a final concentration of 10 mg/mL. The reservoir solution contained 25% PEG 4000 (w/v), 0.1 M Ammonium Sulphate and 0.1 M Sodium Acetate, pH 4.6. Drops were set up in 96 well Greiner low profile plates using a Mosquito robot (TTP Labtech). 100 nL of the protein solution were mixed with 100 nL of the reservoir solution. Crystals (Fig. 4.5) appeared overnight and grew for about two weeks. Trials to scale up these conditions in 24-well plates with the hanging-drop method were not successful.



Figure 4.5 Crystal of C2A domain of Rabphilin3A of dimension 0.25 x 0.03 x 0.05 mm³.

Crystals were soaked in a cryosolution containing 30% (v/v) of glycerol, mounted in Hampton loops and flash-cooled in a stream of liquid nitrogen. Data were collected in-house in fine-sliced mode (0.2°) using a Bruker rotating-anode generator, CuK α radiation, Osmic focusing mirrors and a Bruker *SMART6000* 4K CCD detector placed at a distance of 7 cm. The exposure time was 70 seconds for the low-resolution pass and 120 seconds for the high-resolution data collection. The best crystal was diffracting to 1.82 Å. The data were indexed with the Bruker program *PROTEUM*, integrated with the Bruker program *SAINT* and scaled with the Bruker program *SADABS*. After scaling, the statistics

Resolution range (Å)	44.41-1.92
No. of observations	65418
No. of unique reflections	10435
Redundancy	6.13 (2.74)
$\langle I \rangle / \sigma(I)$	19.11 (3.09)
Completeness (%)	97.8 (92.6)
R_{int}^{\S} (%)	5.15 (28.37)

Table 4.1 Data statistics for the C2A domain of Rabphilin3A. Values between parentheses are for the highest resolution shell (1.95–1.92 Å, corresponding to 439 observations; § $R_{\text{int}} = \sum (|I - \langle I \rangle|) / \sum I$).

and the space group determination were performed with the Bruker program *XPREF*. In the light of the statistics the data were cut at 1.92 Å. The crystals belong to the orthorhombic space group $P2_12_12_1$, with the unit cell parameters $a = 37.84$, $b = 39.16$, $c = 88.95$ Å, $\alpha, \beta, \gamma = 90^\circ$. There is one molecule per asymmetric unit and the solvent content of the crystals is 39.8%, with a Matthews coefficient of $2.1 \text{ Å}^3 \text{Da}^{-1}$. The data statistics are shown in Table 4.1.

4.2.2 Structure determination and refinement

The structure has been determined by molecular replacement using data in the range 35.8–2.5 Å. The search model was based on the structure of the C2A domain of Synaptotagmin I (PDB-ID: 1rsy, residues 140–265, (Sutton *et al.* 1995)) which has a sequence identity of 40% (Fig.4.6). The final search model was generated by replacing the residues that were different with Ala. Using this mixed model a solution has been found with the program *PHASER* (McCoy *et al.* 2005) of the CCP4 package (Collaborative Computational Project 1994). In *PHASER* a two steps molecular replacement search is made using likelihood method: first a rotation function search is performed, followed by a translation function search. The best solution had a log-likelihood gain of 10.52 and a Z score of 10.20. The rotation function was characterized

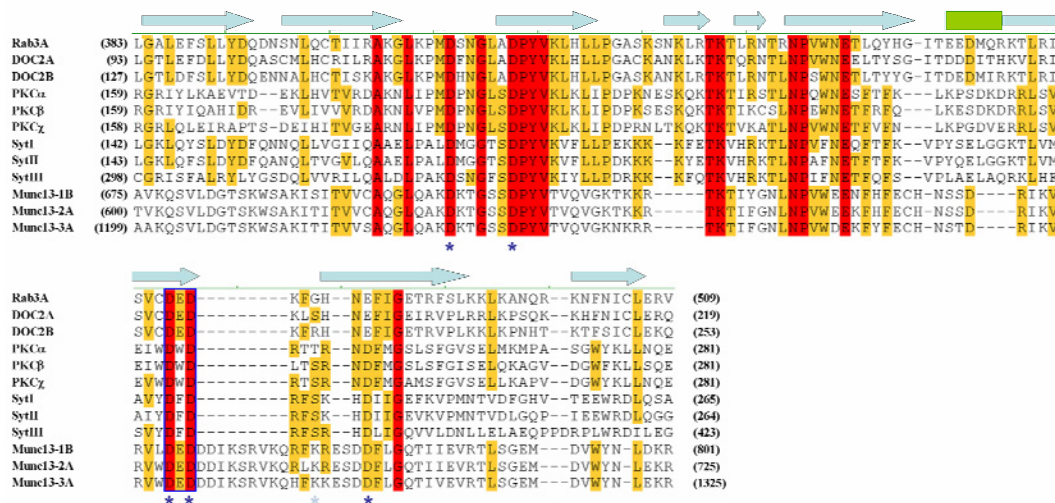


Figure 4.6 Sequence alignment for several C2 domains (performed with the program *Vector NTI*). Red boxes highlight identical residues, orange boxes conserved ones. The blue box shows the DED motif present in Rabphilin, DOC2A/B and Munc13-1/2/3 proteins. The secondary structure elements were drawn according to the structure of the C2A domain of Rabphilin3A. The blue arrows represent the β -sheet and the green box the α -helix. The Ca^{2+} binding residues conserved in the structure of Synaptotagmin I and C2A domain of Rabphilin3A are represented by the dark blue asterisks while the only non-conserved one by the light blue asterisk.

by three Euler angles $\theta = 24.3^\circ$, $\phi = 127.7^\circ$ and $\omega = 138.4^\circ$ and the translational function by three vectors $t_x = -0.419$, $t_y = -0.25$ and $t_z = -0.073$. After twenty cycles of rigid-body refinement in *REFMAC* ($R_{\text{work}} = 53\%$, $R_{\text{free}} = 54\%$) (Murshudov *et al.* 1997) the program *RESOLVE* was employed for density modification and for model building (Terwilliger 2002). Density modification was performed with the prime-and-switch procedure of *RESOLVE* to reduce model bias. *RESOLVE* uses maximum-likelihood density modification based on three quantities: the prior knowledge of the structure factors, the likelihood of the observed set of structure factors and the likelihood of the map resulting from them (Terwilliger 2000). In the prime-and-switch procedure phases of an atomic model are used to calculate an initial electron density map. The probability function for the map is computed and phases are adjusted to maximize this probability function (Terwilliger 2004). The partial model obtained from *RESOLVE* contained 111 correctly

placed residues (104 of them with side chains built) out of 127 ($R_{\text{work}} = 36\%$, $R_{\text{free}} = 38\%$). Further tracing and rebuilding was carried out using the graphics program *COOT* (Emsley *et al.* 2004) alternating with refinement in *REFMAC*.

After a few refinement cycles 84 waters have been automatically added with *ARP/wARP* (Lamzin *et al.* 2001) in *REFMAC* with a drop in R_{work} and R_{free} of 3.2% and 1.9%, respectively. *ARP/wARP* adds waters in place of peaks higher than 0.5 sigma. If two peaks are closer than 2.2 Å they are merged together. New waters are added if closer than 3.3 Å and farther than 2.2 Å to each other. The B values were initially refined isotropically while at the end TLS refinement (Winn *et al.* 2001) was carried out considering the whole protein as a single domain. The final model from the *REFMAC* refinement has an $R_{\text{work}} = 19.6\%$ and $R_{\text{free}} = 26.0\%$, contains 127 residues, 66 water molecules and one glycerol molecule. The model of the last *REFMAC* refinement was further refined using *SHELXL* (Sheldrick *et al.* 1997). The version of *SHELXL* used was a test version with restraints that mimic a TLS refinement (see Chapter 5 for more details). An $R_{\text{work}} = 19.31\%$ and $R_{\text{free}} = 23.97\%$ were obtained (Fig.4.7).

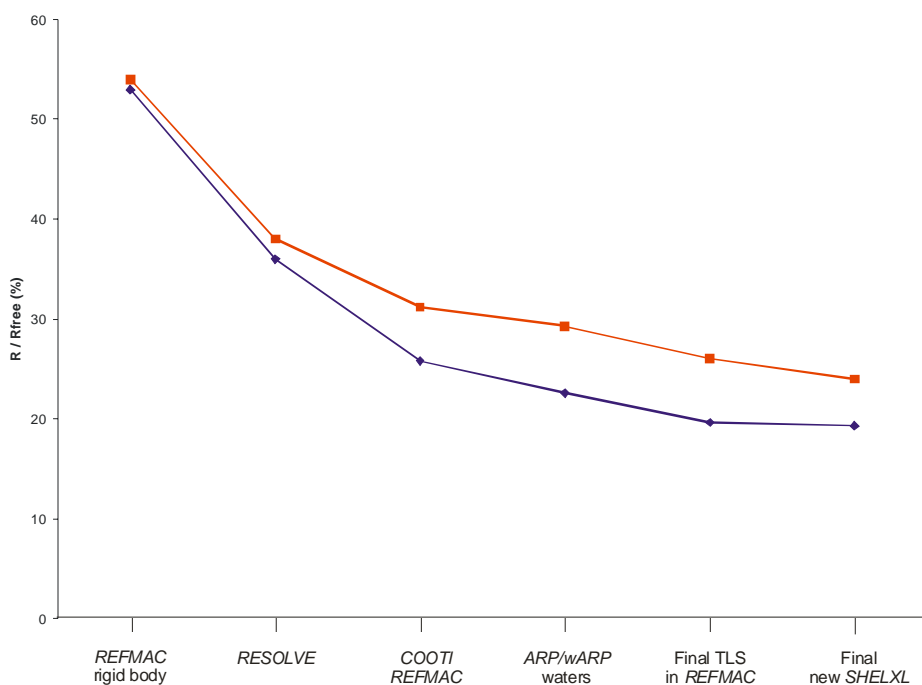


Figure 4.7 Graphic that shows the trend of R_{work} (in blue) and R_{free} (in red) during the major steps of the refinement.

4.3 Results and discussion

4.3.1 Quality of the model

As described in the previous section the model of the C2A domain of Rabphilin3A was refined in *REFMAC* using TLS minimization considering the whole molecule as a single domain. The results of the TLS refinement have been analyzed with the program *TLSANL* (Howlin *et al.* 1993). The analysis of the TLS refinement shows that the main libration is along the z axis (Fig. 4.8). No atoms are non positive definite. After refinement with *REFMAC*, the gap between R_{work} (19.6%) and R_{free} (26.0%) was relatively large (6.4%). On the contrary the refinement with *SHELXL* resulted in a considerably smaller gap between the two values (4.7%) and in a lower R_{free} (24.0%) suggesting that the restraints applied are more suitable at this resolution (for more detailed discussion see Chapter 5).

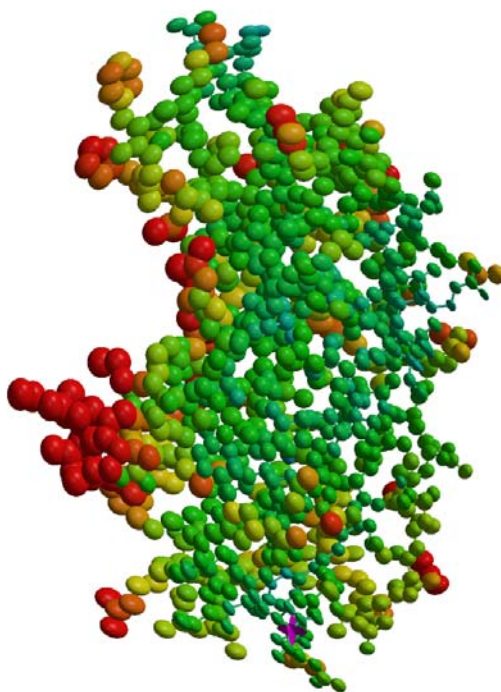


Figure 4.8 Model of the C2A domain after TLS refinement in *REFMAC* and analysis with *TLSANL*. Atoms with high anisotropy are drawn in red.

Analysis with *PROCHECK* (Laskowski *et al.* 1993) shows that the overall geometry of the structure is quite good: 83.5% of the residues are in the most favorite regions, 14.8% in the additional allowed ones and 1.7% in the disallowed region of the Ramachandran plot (Fig. 4.9).

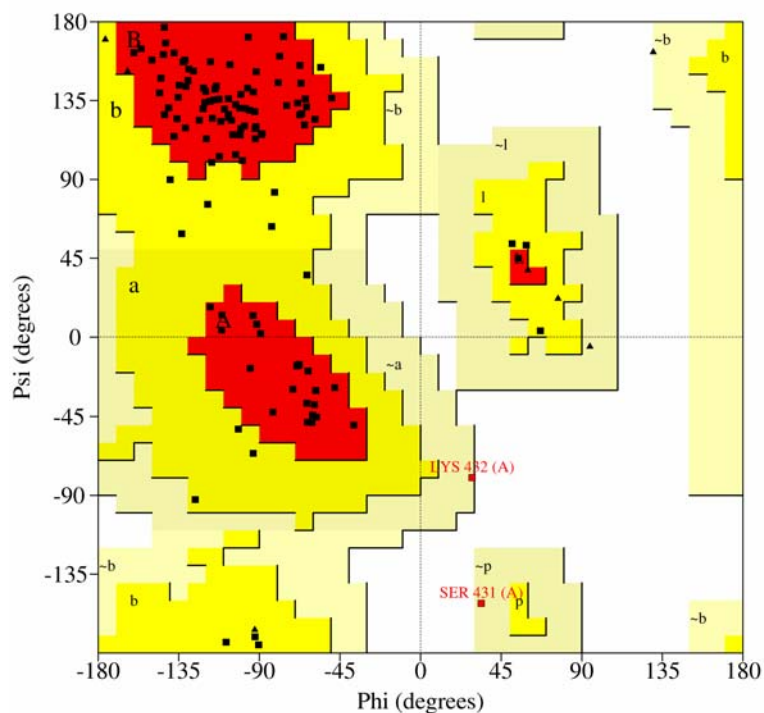


Figure 4.9 Ramachandran plot for the final model of the C2A domain of Rabphilin3A. Ser431 and Lys432, for which the electron density is not clear are in disordered region.

The only two residues in the disallowed region are Ser431 and Lys432 for which the electron density is not well defined, in line with a relatively high B value (45 \AA^2). The statistics for the *SHELXL* refinement are shown in Table 4.2.

R_{work} (%)	19.3
R_{free} (%)	24.0
No. of protein atoms per asymmetric unit	1056
No. of non-protein atoms	72
r.m.s.d bond lengths (\AA)	0.005
r.m.s.d. angle distance (\AA)	0.023
Average B (\AA^2)	27.7
Ramachandran plot statistic	
Residues in most favored regions (%)	83.5
Residues in additional allowed regions (%)	14.8
Residues in disallowed regions (%)	1.7

Table 4.2 Data statistics for the final model of the C2A domain of Rabphilin3A refined with *SHELXL*.

The average B value of the 1056 protein atoms in the asymmetric unit calculated with *Baverage* in the CCP4 package is 27.7 \AA^2 . The root mean standard deviation (r.m.s.d.) for the bond lengths is 0.005 \AA while the r.m.s.d. for the angle distances (distance 1-3) is 0.023 \AA .

4.3.2 Structure description

The refined crystal structure of the C2A domain of Rabphilin3A includes 127 residues (residue Tyr382 to residue Val509) (Fig. 4.10), 66 water molecules and one glycerol molecule. The structure consists of a compact β -sandwich formed by two four-stranded β -sheets, one small helix (residues 460-465) and three loops (loop1: residues 407-419, loop 2: residues 441-447 and loop3: residues 476-479) at the top and the bottom of the sandwich, respectively. The C2A domain has the topology I found in all other C2 domains of synaptotagmin proteins (Fig. 4.11).

It was also attempted to obtain the structure of Rabphilin3A with Ca^{2+} ions bound but, after soaking the crystals in solution containing different concentrations of Ca^{2+} their diffraction quality dramatically decreased. Also trials to cocrystallize Rabphilin3A with Ca^{2+} did not succeed.

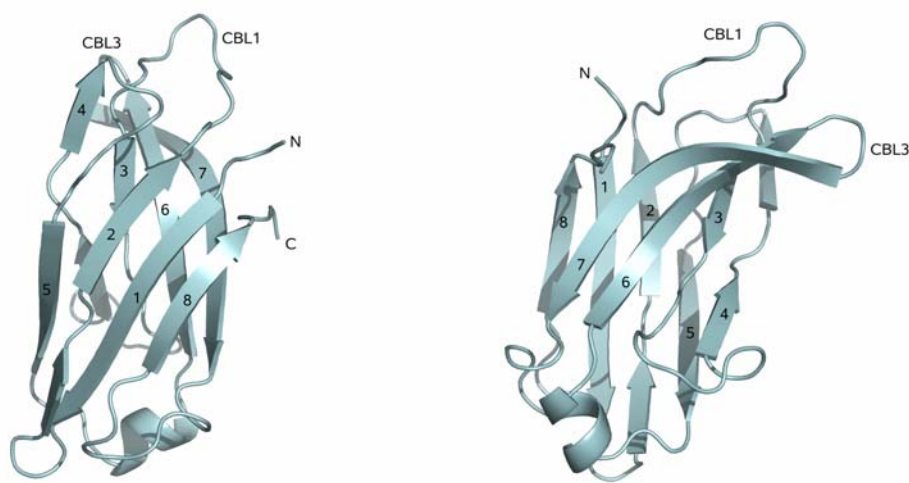


Figure 4.10 Overall structure of the C2A domain of Rabphilin3A. The two calcium binding loops are denoted as CBL1 and CBL3.

Although Ca^{2+} ions have not been found in the structure, the loops located in the Ca^{2+} binding region (residues 411-421, called CBL1 and 474-484, called CBL3) are well ordered: for CBL1 the C^α atoms have an average B factor of 19 \AA^2 , while for CBL3 the average B factor is 23 \AA^2 (Fig. 4.12).

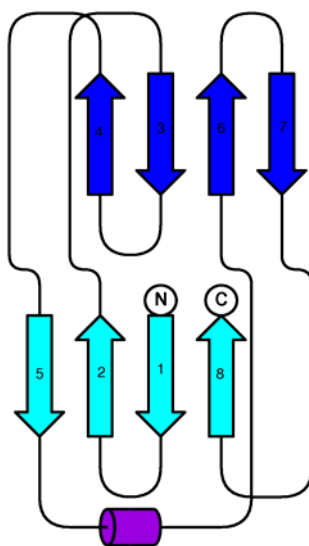


Figure 4.11 Topology diagram of type I of C2A domain of Rabphilin3A.

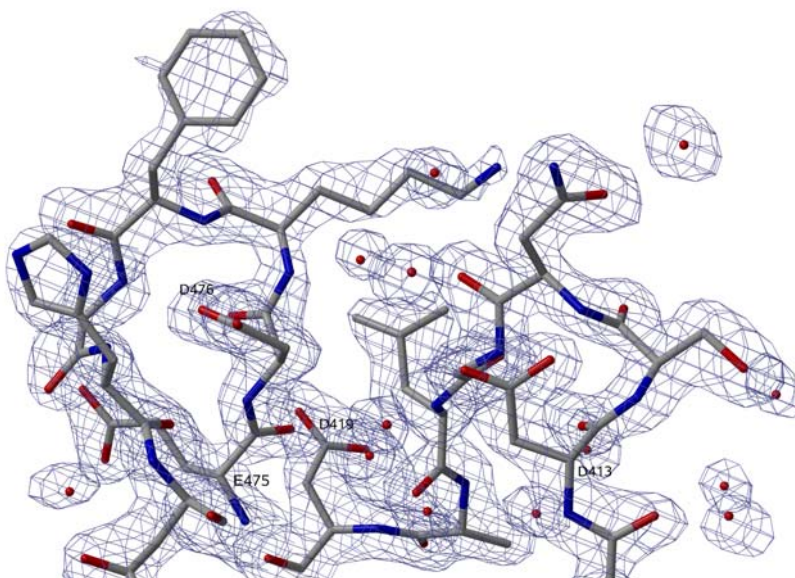


Figure 4.12 σ_A weighted ($2mF_o - DF_c$) electron density map (contoured at 1σ) of the regions CBL1 and CBL3 and some of the residues normally involved in Ca^{2+} binding (Asp413, Asp419, Glu475, Asn481) are shown. Even if no Ca^{2+} ions are present the loops are well ordered. Water molecules are shown as little red spheres.

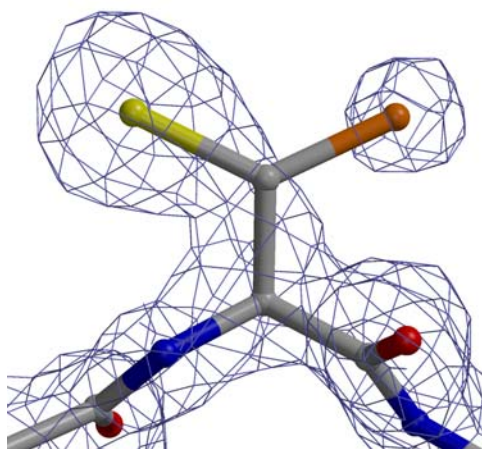


Figure 4.13 σ_A weighted ($2mF_c - DF_o$) electron density map (contoured at 1σ) of residue Cys401. The two conformations of the sulfur atom (each with occupancy = 50%) are shown in yellow and orange.

The residues conserved among different C2 domains and involved in Ca^{2+} binding (Asp 413, Asp419, Asp474, Asp476 and Glu482) are in the Ca^{2+} -free conformation with the side chains in an orientation similar to the one described in the Synaptotagmin I C2A domain crystal structure (Sutton *et al.* 1995).

The electron density was generally of good quality except for the N-terminal part of the fourth β -strand where it was not possible to build the side chains of some residues while the main chain could be unambiguously traced. Double conformations of the side chain were modeled for 6 out of the 127 residues (Cys401, His425, Asp434, Lys439, Asn450, Arg469; Fig. 4.13). For four residues (Gln460, Glu461, Lys492, Lys495) located on the surface of the protein only part of the side chain has been modeled. A cis-peptide bond was detected between Leu427 and Pro428 in the same position as in other C2 domains structures (Pro202 in the PKC α (Verdaguer *et al.* 1999) and PKC β (Sutton *et al.* 1998), Pro411 in the C2A domain of Synaptotagmin III (Sutton *et al.* 1999) and Pro187 in the C2A domain of Synaptotagmin I (Sutton *et al.* 1995)).

One glycerol molecule, presumably introduced by soaking with glycerol for cryoprotection, has been modeled near Asn481, Lys423 and Lys410 of a symmetry-

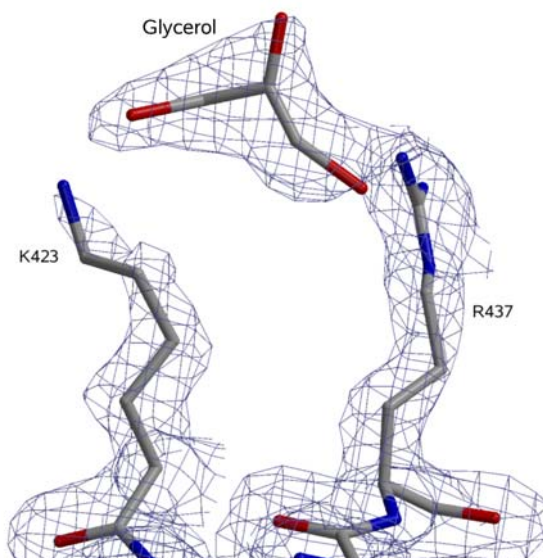


Figure 4.14 The glycerol molecule found in the structure and the residues that interact with it (Lys423 and Arg437) are shown in the σ_A weighted ($2mF_o - DF_c$) electron density map (contoured at 1σ).

related C2A molecule (Fig. 4.14). The glycerol molecule is in the same position as a phosphate found in the structure of Synaptotagmin III (Sutton *et al.* 1999) and it might mimic the negative head of a phospholipid molecule. As Sodium Sulphate was a component of the crystallization condition, it was also tried to place a sulphate ion instead of glycerol into this electron density, but this resulted in very high B values, suggesting that glycerol was the better choice.

4.3.3 Comparison with other C2 domains

The crystal structure of the C2A domain of Rabphilin3A has been compared with the structures of the C2A domains of Synaptotagmin I (PDB-ID: 1rsy, (Sutton *et al.* 1995)), Synaptotagmin III (PDB-ID: 1dqy, (Sutton *et al.* 1999)) and the C2 domains of PKC α (PDB-ID: 1dsy, (Verdaguer *et al.* 1999)) and of PKC β (PDB-ID: 1a25, (Sutton *et al.* 1998)). A superposition of the structures was carried out using the program *SSM* (Krissinel *et al.* 2004) within *COOT* (Emsley *et al.* 2004). All five structures are similar to each other with differences concentrated in the loops. The root mean square deviations between the structure described here and the C2A domains of Synaptotagmin I and Synaptotagmin III are 1.21 Å and 1.81 Å, respectively and the ones with the C2 domains of PKC α and PKC β are 1.44 Å and 1.45 Å, respectively. In all the cases the differences are quite small.

The major differences between the structure of the C2A domain of Rabphilin3A and of the C2A domain of Synaptotagmin I, which was used as a search model in the molecular replacement, are visible in calcium binding loops 1 (CBL1) and 3 (CBL3) that connect strands β 2 and β 3 and strands β 6 and β 7, respectively (Fig. 4.15). CBL1 of Rabphilin3A is flipped by 20° and rotated by 90° against CBL1 of the homologous C2 domains of Synaptotagmin I, III and PKC α and PKC β , but in all the structures it is well ordered (B values of 15-20 Å²). A consequence of these structural changes is a significant shift of the position of residue Asp413 compared to the equivalent residue Asp172 of the C2A domain of Synaptotagmin I. The longer side chain of Glu482, as compared to the oxylic

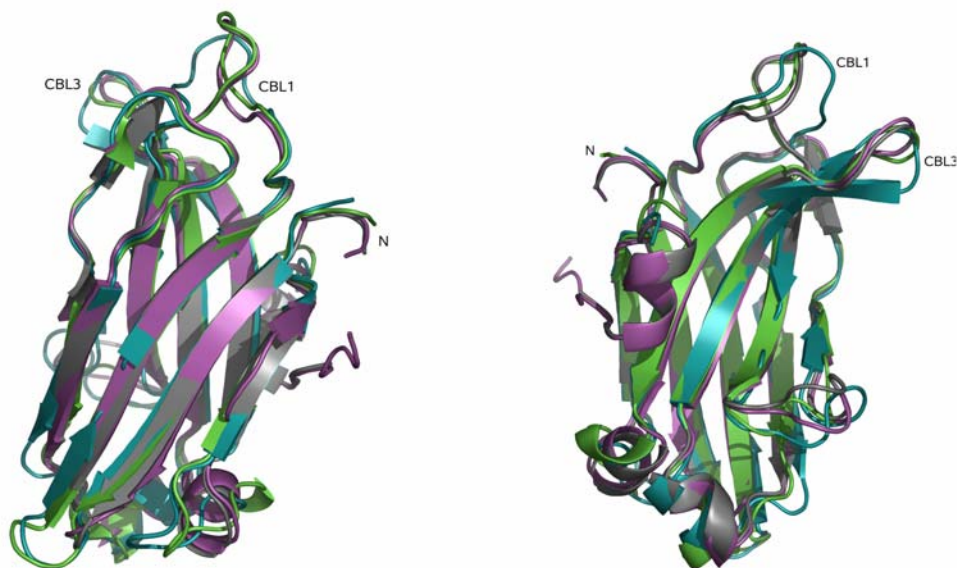


Figure 4.15 Superposition of C2 domains. The C2A domain of Rabphilin3A is shown in cyan, the one of Synaptotagmin I in green, the C2 domain of PKC α is drawn in magenta and the one of PKC β in grey.

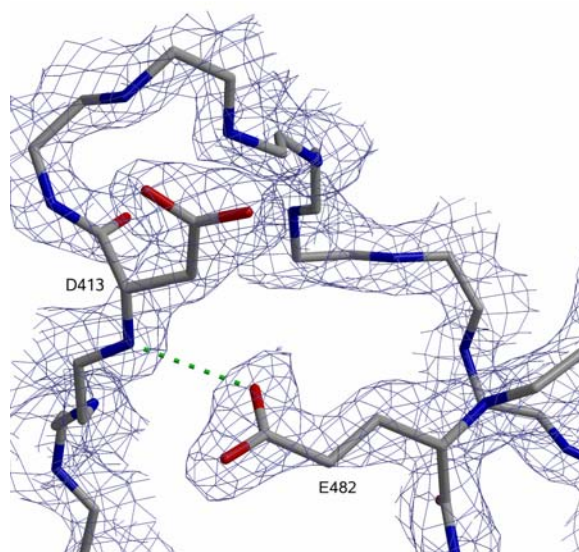


Figure 4.16 : σ_A weighted ($2mF_o - DF_c$) electron density map (contoured at 1σ) of the loop between residues 411-421 and the strand from residue 482 to 484. The density for some side chains has been omitted for clarity. The hydrogen bond between Asp413 and Glu482 is indicated by a dashed line.

moiety and the backbone NH of conserved Asp found in other C2 domains (Asp238 in the case of Synaptotagmin I), facilitates a hydrogen bond between its carboxylic moiety and the backbone NH of Asp413 located in CBL1 (Fig.4.16). Part of CBL3 (residues 476-481) of the C2A domain of Rabphilin3A is flipped by 56° and rotated by 95° against CBL3 of the homologous C2 domains of Synaptotagmin I, Synaptotagmin III, PKC α and PKC β . The loop between the strands $\beta 7$ and $\beta 8$ of Rabphilin3A (residues 494-501) is similar to the ones of Synaptotagmin I, PKC α and PKC β . But in the structure of the C2A domain of Synaptotagmin III this loop partially folds as a short helix (residues 406-409 of Synaptotagmin III) and the backbone of the following six residues (in which are included also two inserted residues Pro412 and Asp413, see Fig. 4.6) is tilted in completely different direction in comparison to Rabphilin3A, Synaptotagmin I, PKC α and PKC β (Fig. 4.17).

Among the Ca^{2+} binding residues commonly found in type I C2 domains, the only residue which is not conserved in the C2A domain of Rabphilin3A is Gly479, structurally homologous to Ser235 located in CBL3 of the C2A domain of Synaptotagmin I. In the structure of Rabphilin3A C2A domain the residues normally involved in the Ca^{2+} binding are in the non-bound conformation. Their side chains are orientated in a different way in

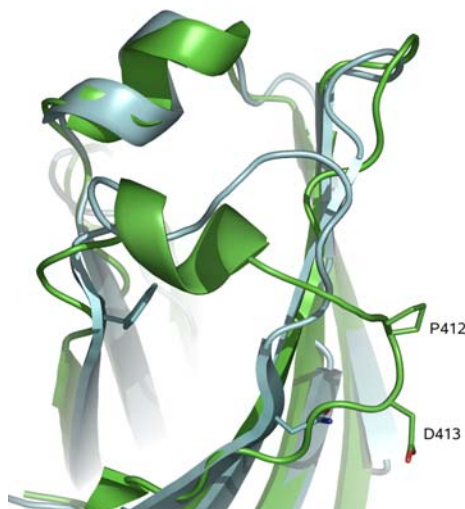


Figure 4.17 Superposition of the C2A domain of Rabphilin3A (in cyan) and of Synaptotagmin III (in green) that shows the different path of the loop between $\beta 7$ and $\beta 8$. The two residues present in Synaptotagmin III and absent in Rabphilin3A are shown as sticks.

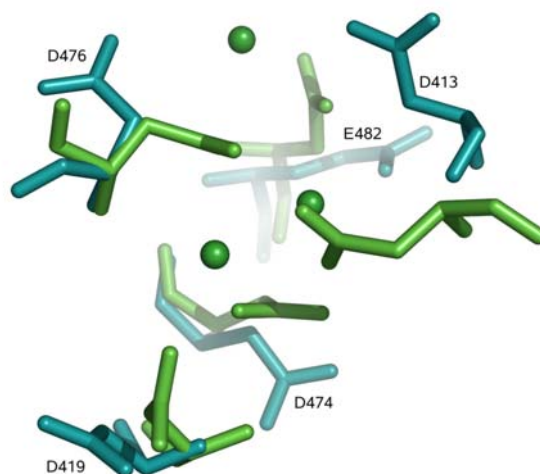


Figure 4.18 Superposition of residues normally involved in Ca^{2+} binding. Residues of the C2A domain of Rabphilin3A are painted in cyan while the ones of the C2A domain of Synaptotagmin I and the Ca^{2+} ions present in that structure are shown in green.

comparison to the ones observed in the structure of Synaptotagmin I with Ca^{2+} bound (PDB-ID: 1byn, Fig. 4.18).

The quality of the electron density of the loop between $\beta 3$ and $\beta 4$ is limited in all structures discussed so far: in the structure of Synaptotagmin III it is disordered with very high B values ($90\text{-}100 \text{ \AA}^2$), in the one of Synaptotagmin I four of the residues have zero occupancy (Lys189, Lys190, Lys191, Lys192), in the structure of PKC α residues 205 to 209 have relatively high B values (30 \AA^2) and in the structure of the C2A domain of Rabphilin3A the loop is disordered.

4.3.4 DED motif

In a recent study the Ca^{2+} sensitivity of Double C2-like domain-containing proteins (DOC2A and DOC2B), two soluble proteins containing C2 domains, has been analyzed. It was found that these two proteins, highly related to the synaptotagmin

protein family and to Rabphilin, have a high Ca^{2+} -dependent phospholipid binding affinity. Therefore it has been hypothesized that this feature could be linked to a DED motif found in CBL3 (the two Asp residues are part of the five conserved residues involved in Ca^{2+} binding (Groffen *et al.* 2006)). In the case of topology type I C2 domains this motif has been found in C2 domains of DOC2A and DOC2B, MUNC13 and Rabphilin3A. In other C2 domains the Glu positioned between the two highly conserved Asp residues that are normally involved in the Ca^{2+} binding, is usually a Tyr or a Phe. The three negatively charged residues (DED) probably contribute to the high Ca^{2+} -dependent phospholipid binding affinity of DOC2A and DOC2B while some minor differences in the surrounding residues could simply stabilize the extra negative charge or interact with the negative head of phospholipids at the plasma membrane. The structure presented here is the first C2 domain structure containing this DED motif. The side chain of Glu275 points towards the concave side of the structure, opposite to the Ca^{2+} binding region. It interacts with the conserved residue Glu419 (the distance between the nitrogen atom of Glu475 and the carboxyl oxygen atom of Asp419 is 2.98 Å) while the side chain is close to the hydroxyl group of Thr440 (distance between OE1 of Glu475 and OG1 of Thr440 is 3.31 Å) and the guanidinium group of Arg437 (distance between OE1 of Glu475 and NH2 of Arg437 is 3.29 Å). The electrostatic potential of the C2A domain of Rabphilin3A, calculated with the Boltzmann normal equation using the program *DelPhi* (Honig *et al.* 1995) and mapped onto the molecular surface, shows that the three acidic residues of the DED motif form a negatively charged patch together with Asp419 and Thr440 (Fig. 4.19). The potential role of this atypical acidic patch might be to be involved in a new Ca^{2+} binding site or to modify the ability of CBL3 to bind Ca^{2+} . Further investigations will be necessary to describe the exact role of this specific motif.

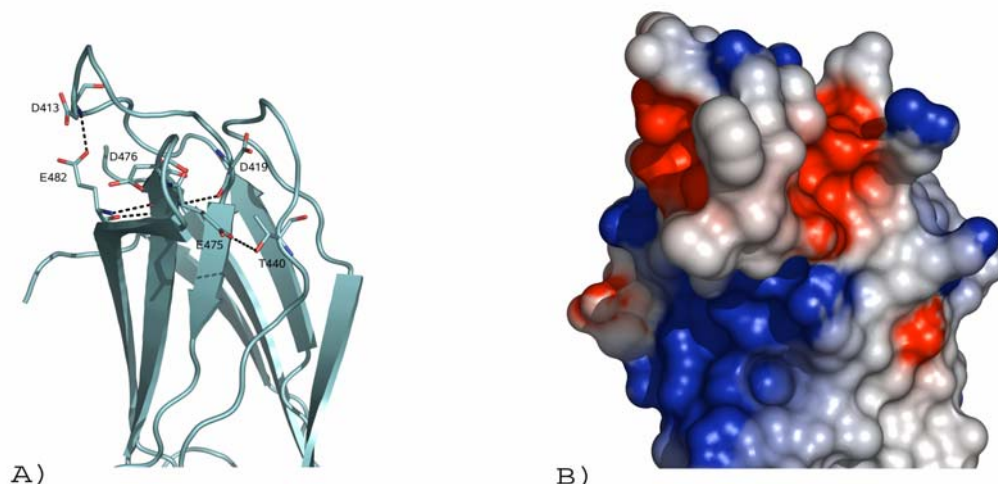


Figure 4.19 Influence of the DED motif on the electrostatic potential surface close to the calcium binding region. A) Ribbon representation of the calcium binding region of the C2A domain of Rabphilin3A. The residues Glu482, Asp413, Asp476 and Asp419 are the conserved residues involved in the Ca^{2+} binding (in this view Asp474 lies behind the β -strand 6). The residue Glu475 from the DED motif forms hydrogen bonds with the Thr440 side chain and the Asp419 backbone oxygen. B) The Poisson-Boltzmann electrostatic potential was calculated and mapped onto the same view of the C2A domain as shown in A. The close proximity of the residues Glu475, Thr440 and Glu419, creates a new negative patch next to the calcium binding region.

4.3.5 Comparison of electrostatic surfaces

The electrostatic potential of the C2A domain of Rabphilin3A shows a positively charged belt surrounded by two localized negative patches (Fig. 4.20). The first negative patch is located in the calcium binding region. It contains all the acidic residues needed for chelating Ca^{2+} ions described in Synaptotagmin I C2A and PKC α C2 domains. The acidic character is then reversed upon binding of Ca^{2+} ions. This electrostatic switch is strictly required to approach the anionic phospholipid subdomains (Murray *et al.* 2002). Previous studies showed that the C2A domain of Rabphilin3A is able to interact with phospholipids in a Ca^{2+} -dependent manner suggesting that these conserved residues are indeed able to bind Ca^{2+} (Chung *et al.* 1998).

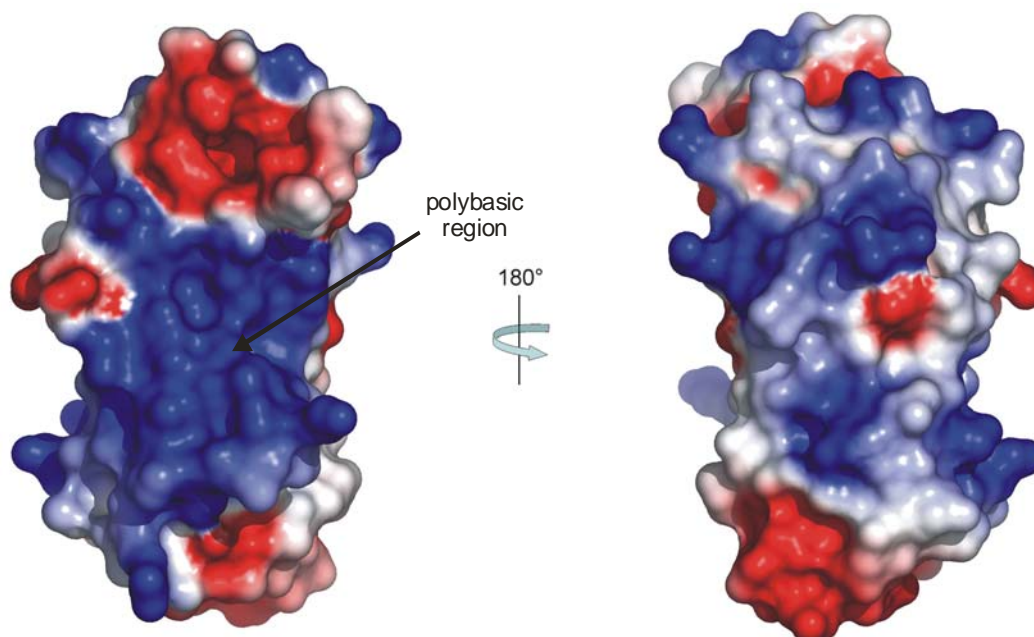


Figure 4.20 Electrostatic potential of the surface of C2A domain of Rabphilin3A calculated as in Fig. 4.19. The left view shows the concave site of the C2A domain, the Ca^{2+} binding region is on the top; the right view shows the other side of the β -sandwich. The arrow points to the PIP2 binding polybasic region.

The second negative patch sits at the tip located at the opposite end of the domain relative to the Ca^{2+} binding region (CBR) and it involves the acidic residues located in the short helix connecting the strands $\beta 5$ and $\beta 6$ (residues 460-465), and in the turn between the strands $\beta 1$ and $\beta 2$ (residues 394-396). This region might be involved in the interaction of the C2A domain with a yet unidentified partner.

The C2A domain of Rabphilin3A forms part of a C2 domain tandem. The electrostatic surfaces of both the C2A and the C2B domains show a strongly positive character contrary to the Synaptotagmin I C2 domain tandem where the C2A domain is negatively charged and the C2B domain is highly basic. This observation suggests different roles for both C2A domains and thus for the C2 domain tandems of Rabphilin3A and Synaptotagmin I.

The residues of the polybasic region of the β -strand 4 of the C2 domain of PKC α that have been reported to be responsible for the binding of phosphatidylinositol-4,5-bisphosphate (PIP2) (Lys197, Lys209, Lys211, (Corbalan-Garcia *et al.* 2003)), are conserved in the C2A domain of Rabphilin3A (Lys423, Lys435, Arg437, see Fig. 4.6). It has been shown that the C2A domain of the structure presented here is able to bind specifically PIP2 in presence of Ca²⁺ (Chung *et al.* 1998). In the structure of the C2A domain of Rabphilin3A these conserved amino acids form part of the large basic belt (see Fig. 4.20), while in PKC α they were involved in a restricted positive patch. In addition to the PIP2 binding, it has been shown that this polybasic region of PKC α is involved in the proper subcellular localization of the whole protein upon Ca²⁺ signaling. It will be interesting to study whether this feature might be essential for the role of the Rabphilin3A C2A domain in Rabphilin3A function.

4.4 Conclusion and perspectives

The crystal structure of the Ca^{2+} -free form of the C2A domain of Rabphilin3A has been determined at 1.92 Å resolution. It has a type I topology and the structure is characterized by an eight-stranded antiparallel β -sandwich also found in the C2 domains of all synaptotagmins and several protein kinases. The C2A domain of Rabphilin3A has Ca^{2+} -dependent membrane binding properties and consequently contains all the acidic residues involved in Ca^{2+} binding. Its electrostatic surface is highly basic and similar to the one of the C2B domain of Rabphilin3A. The Ca^{2+} binding to C2 domains does not induce large conformational changes and consequently the Ca^{2+} binding properties of a specific C2 domain cannot readily be inferred from its primary sequence (Dai *et al.* 2004). Thus, small structural modifications may be responsible for differences in functional specificities from one C2 domain to another. In the C2A domain of Rabphilin3A two specific structural features related to deviations from the *consensus* sequence have been found: a different position of the calcium binding loop 1 (CBL1) that causes a significant shift of the conserved Ca^{2+} binding residue Asp413, and the presence of the residue Glu475 that creates a localized negative patch close to the calcium binding region. Further investigations are needed to determine the relationship between these structural particularities of the C2A domain of Rabphilin3A and its specific biophysical properties.

5 Restraints

5.1 Introduction

In this chapter tests of a new version of *SHELXL* (Sheldrick *et al.* 1997) are presented. In macromolecular crystallography one of the biggest problems to overcome is the relatively small amount of data in comparison with the parameters refined (low data-to-parameter ratio, Table 5.1). Two major strategies have been used to get over this problem: decrease the amount of parameters introducing constraints into the model or artificially increase the amount of data using additional restraints. Known parameters are fixed to target values and constraints can be seen as exact mathematical conditions applied for example to rigid groups of a molecule or to riding hydrogen atoms. Restraints are instead used when the parameters are allowed to deviate from the target values. Restraints can then be seen as additional observational equations involving the target values T and their standard deviations σ and this term is added to the quantity that has to be minimized during the refinement:

$$M = \sum w_x (\bar{F}_{\text{obs}}^2 - \bar{F}_{\text{calc}}^2)^2 + \sum w_r (T_{\text{target}} - T_{\text{calc}})^2 \quad 5.1$$

where w_x are the X-ray weights that should increase as the agreement between observed and calculated structure factors improves and w_r is equal to $1/\sigma^2$.

Both restraints and constraints are based on previous chemical knowledge from small molecule structures (Cambridge Structures Database, (Allen 2002)) or high resolution

resolution (Å)	# reflections per atom	parameters refined	# data per parameter
3.0	2	x y z	0.6
2.5	4	x y z (B)	1
2.0	8	x y z B	2
1.5	20	x y z B	5
1.1	50	x y z U ₁₁ U ₂₂ U ₃₃ U ₁₂ U ₁₃ U ₂₃	5
0.9	90	x y z U ₁₁ U ₂₂ U ₃₃ U ₁₂ U ₁₃ U ₂₃	10

Table 5.1 Data-to-parameter ratio at different resolutions.

protein structures and normally involve bond lengths, bond angles, planarity or dihedral angles. Restraints are also applied to the atomic displacement parameters (ADPs) so that the motion described has physical sense. In the next subchapters an introduction of the different ways to describe the thermal motion will be presented together with the results of a new version of *SHELXL*.

5.1.1 Atomic displacement parameters

Atoms of a molecule are fully described by their mean average position and by the atomic displacement parameters. ADPs express the mean-square deviation from the average position within the normal Gaussian approximation (Willis *et al.* 1975; Trueblood *et al.* 1996):

$$B = 8\pi^2 \langle \mu^2 \rangle \quad 5.2$$

where $\langle \mu^2 \rangle$ is the mean square displacement.

The thermal motion of the atoms has different origins. Internal statistic disorder is caused by differences in equivalent atomic coordinates between different unit cells while internal dynamic disorder is due to the fact that the motion of the atoms is much faster than the time of an X-ray experiment. Moreover lattice defects of the crystals and errors in the model have to be taken into account.

Depending on the resolution and consequently on the amount of data, this thermal motion can be described in different ways. At very low resolution the B factors are fixed to an arbitrary value and not refined. At low resolution an isotropic B factor is associated with each atom: the motion of the atom is described by a sphere centered at the average position. One parameter (the radius of the sphere) in addition to the position coordinates is necessary to fully describe each atom. On the opposite side when high resolution data are available and the data-to-parameter ratio is high enough, the thermal motion of each atom can be described by an ellipsoid centered at the mean position. Six parameters describe the orientation and the elongation of the ellipsoid. An atom is then fully

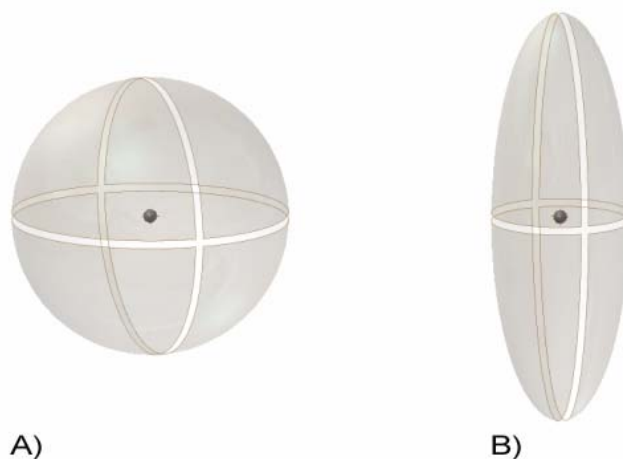


Figure 5.1 Description of the thermal motion of an atom. A) Isotropic description. B) Anisotropic description.

described by nine parameters (Fig. 5.1). At medium resolution, when the amount of data is enough to describe the model better than as simply isotropic but not enough for an anisotropic description of all the individual atoms, a compromise is to use TLS (translation, libration and screw motion) refinement.

5.1.2 TLS

The total atomic displacement parameter of a crystal can be considered as the sum of four separate contributions (Winn *et al.* 2001):

$$U = U_{\text{crys}} + U_{\text{TLS}} + U_{\text{internal}} + U_{\text{atom}} \quad 5.3$$

where U_{crys} is the anisotropy of the crystal and it has to obey the symmetry of the crystal space group, U_{TLS} is the TLS motion of pseudo-rigid bodies in the asymmetric unit, U_{internal} is the collective torsional libration or internal normal modes and U_{atom} is the displacement of individual atoms. The TLS description is based on the assumption of the presence of pseudo-rigid bodies in the molecule where a rigid-body is defined as a part of the molecule in which all intramolecular links (non-bonded intramolecular distances) are

rigid and in which all the atoms move in phase (Didisheim *et al.* 1987). The theory of TLS (Schomaker *et al.* 1968; Howlin *et al.* 1989; Schomaker *et al.* 1998) assumes that the instantaneous displacement \bar{u} at a point \bar{r} in a rigid-body can be specified as a screw motion, that is a rotation of magnitude $|\bar{\lambda}|$ (in radians) about a suitably positioned and oriented axis $\bar{\lambda}$, correlated with a translation \bar{t} parallel to that axis, all vectors being specified relative to some arbitrarily positioned and orientated fixed Cartesian axes (Fig. 5.2):

$$\bar{u} \cong \bar{t} + \lambda \bar{r} \times \bar{r} \quad 5.4$$

where "x" denotes the vector product. The equation is more conveniently written in matrix notation:

$$\bar{u} = \bar{t} + \bar{D} \cdot \bar{r} \quad 5.5$$

where \bar{D} is the rotation matrix and \bar{t} the translation vector.

The dyad product is then:

$$\overline{uu}^T = \overline{tt}^T + \bar{t} \bar{\lambda}^T \bar{x} \bar{r}^T - \bar{r} \bar{x} \bar{t}^T - \bar{r} \bar{x} \bar{\lambda} \bar{\lambda}^T \bar{x} \bar{r}^T \quad 5.6$$

and averaging in time and space over all displacements gives:

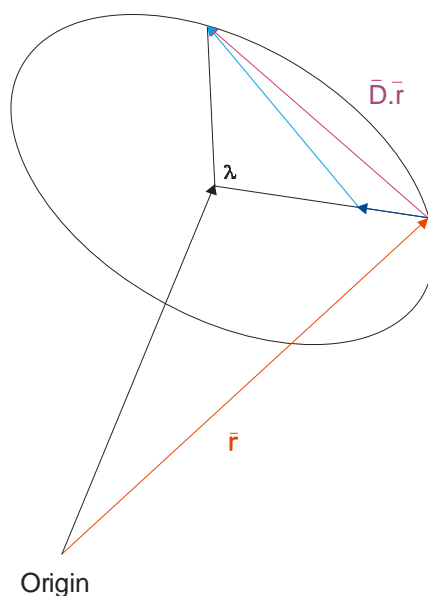


Figure 5.2 Original description of TLS (Schomaker *et al.* 1968). The instantaneous rotation displacement of an atom originally at \bar{r} relative to an arbitrary molecule-fixed origin is given by:

$$\bar{D} \cdot \bar{r} = \bar{\lambda} \times \bar{r} (\sin \lambda) / \lambda + \left[(\bar{r} \cdot \bar{\lambda}) \bar{\lambda} - \lambda^2 \bar{r} \right] (1 - \cos \lambda) / \lambda^2 .$$

$$U \equiv \langle \overline{uu}^T \rangle = T + S^T x \bar{r}^T - \bar{r} x S - \bar{r} x L x \bar{r}^T \quad 5.7$$

where T , L and S are three tensors. The tensor T is symmetric, dependent on the origin chosen and it describes the translation. The tensor L is also symmetric but independent from the origin chosen and it describes the libration (rotational oscillation). The tensor S is asymmetric, dependent on the origin chosen and it describes the average quadratic correlation of translation and libration. The trace of S cannot be determined from diffraction data alone and it is normally constrained to the arbitrary value of zero. Normally the origin is chosen so that T is minimized and S becomes symmetric. There are then 20 parameters to be determined for each rigid-body group: 3 rotation amplitudes, 3 translation amplitudes, 3 screw correlations (out of which only 2 are determined), 9 angles that describe the orientation of the principal-axes system and 3 coordinates of the origin. In the program *REFMAC* (Murshudov et al. 1997) TLS constraints are combined with isotropic refinement of the B factors while in a test version of *SHELXL* restraints to be applied to the ADPs have been modified as to mimic TLS refinement.

5.1.3 New restraints in *SHELXL*

In *SHELXL* normally three restraints are applied to the atomic displacement parameters: SIMU, ISOR and DELU (Fig. 5.3).

The *similar ADP* restraint (SIMU) takes into account the fact that two spatially closed atoms tend to move in the same way (in phase) and consequently the ellipsoids describing their motion have to be similar to each other. Six restraints are applied to each pair of atoms.

The *isotropic* restraint (ISOR) is normally applied to solvent molecules (which are not bonded to any other atoms). For these molecules the atomic displacement parameters should be approximately isotropic. This restraint is also useful to prevent getting ellipsoids with no physical sense (imaginary length of one or more of the axes).

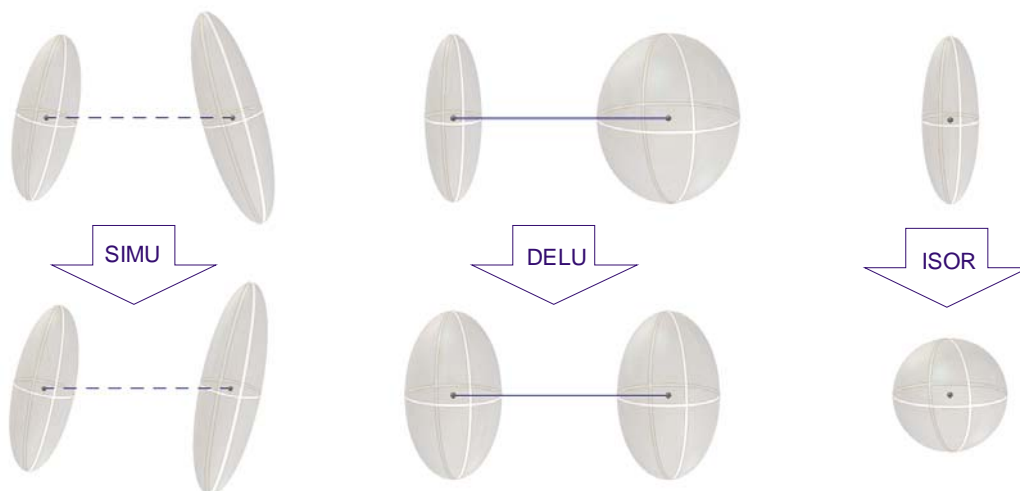


Figure 5.3. The three restraints applied to the ADPs used in *SHELXL* refinement are shown.

With the *rigid-body* restraint (DELU) the components of the anisotropic motion of two atoms along the line joining them are restrained to be equal (Rollett 1970; Hirshfeld 1976; Trueblood *et al.* 1983). One strict rigid-bond restraint is applied to each couple of atoms and low estimated standard deviations (ESDs) should be used (Didisheim *et al.* 1987).

In a test version of *SHELXL* a new DELU restraint is implemented. First of all the restraint is not applied only to bonded atoms but to atom pairs within a distance of about 8 Å. In this way not only the component along a bond but the components in any directions are restrained. Secondly the differences between the ADP components and U_{eq} of the atoms along the ideal line joining them are restrained to be equal. The ADP component along this line and U_{eq} are both linear combinations of the U_{ij} and consequently also the new DELU restraint is a linear restraint. If very small estimated standard deviations are used for atom pairs within one domain the DELU restraint can be seen as the TLS constraint implemented in *REFMAC*. Because the ESDs are multiplied by the distance between the atoms so that the closer interactions get higher weights, it is not necessary to define the domains.

5.2 Materials and methods

The new version of *SHELXL* was tested using different data sets from various macromolecules with a typical resolution between 1.2 and 2.3 Å. At this resolution a simple isotropic refinement is not satisfactory while a fully anisotropic refinement, due to the still low data-to-parameter ratio is not possible. The best way to describe the thermal motion of the atoms is then TLS refinement or something similar. For some data sets a comparison was made between TLS refinement in *REFMAC* and the new version of *SHELXL* while for other data sets the comparison was carried out between two versions of *SHELXL* (Table 5.2).

In the *SHELXL* instruction file (Fig. 5.4) the new DELU command looks like:

DELU [k] [d] atomnames

where:

- k is the effective standard deviation for which the default value is set to 0.01 Å
- d is the maximum distance between the atoms to which the restraint is applied. The default value is equal to 8 Å.

macromol. name	data set name	#residues	resolution (Å)
Wind	W-His	450	2.35
Wind Y53F mutant	Y53F	449	2.28
Wind Y55K mutant	Y55K	347	2.25
Aldose Reductase	AR	315	2.00
Concavalin	Con20	237	2.00
C2A domain Rabphilin3A	C2A	127	1.92
Concavalin	Con17	237	1.75
Viscotoxin-a3	Visc-a3	92	1.75
Wind Y53S mutant	Y53S	227	1.75
Thermolysin	Therm	316	1.65
Glucosoisomerase with Calcium	Gi_ca	385	1.60
Thaumatococcus	Thau	206	1.55
C2B domain of Rabphilin3A	C2B	153	1.28
Phosphatase	tPphA	240	1.19

Table 5.2 Data sets used for testing the new version of *SHELXL*. The common names, their abbreviations used in this chapter, the number of residues, and the resolution are listed.

```

TITL c2a with glycerol
CELL 1.54178 37.870 39.160 88.850 90.00 90.00 90.00
ZERR 2 0.038 0.039 0.089 0.00 0.00 0.00

REM Space group P 21 21 21

LATT -1
SYMM 1/2-X, -Y, 1/2+Z
SYMM -X, 1/2+Y, 1/2-Z
SYMM 1/2+X, 1/2-Y, -Z
SFAC C H N O S
UNIT 2576 5224 736 1040 20

DEFS 0.02 0.1 0.01 0.04
CGLS 20 -1
SHEL 10 0.1
FMAP 2
PLAN 200 2.3
LIST 6
WPDB 2

ANIS

DELU 0.01 8 N_1382 > O1_3647
SINU 0.1 0.1 2.1 N_1382 > O1_3647
ISOR 0.05 O_4001 > LAST
CONN 0 O_4001 > LAST
BUMP
SWAT
REM HOPE

```

Figure 5.4 Example of the instruction file of *SHELXL* with the new DELU restraints.

atomnames are the atoms to which the restraint is applied, normally all the atoms of the protein.

For each structure a final refined isotropic model was taken together with a data file (mtz or hkl format but with the same reflections flagged for cross-validation in R_{free}). The structures were then refined with *REFMAC* with the TLS and restrained refinement option on and with the new version of *SHELXL* using the default setting for the restraints. The results of TLS refinement in *REFMAC* and new *SHELXL* were compared looking at the values of R_{work} , R_{free} and the gap between them. Some of the structures were refined using the two versions of *SHELXL*. In that case the results were compared looking also at the number of atoms that got flagged as "may be split" or "non positive definite".

5.3 Results and discussion

5.3.1 Comparison between TLS in *REFMAC* and new *SHELXL*

Eight data sets were used to make a comparison between TLS refinement as implemented in *REFMAC* and the new DELU restraint in *SHELXL*. At low resolution (worse than 2 Å; data sets W-His, Y53K, Y55K) *REFMAC* seems to work better. Both R_{work} and R_{free} are generally lower by about 2% and also the gap between R_{free} and R_{work} is smaller by about 2%. At medium resolution (between 1.65 and 2.0 Å; data sets C2A, Y53S and Thermo) the results obtained with the two programs are more or less the same. At higher resolution (better than 1.6 Å; data sets Gi_ca and Thau) *SHELXL* seems to be a bit better relative to *REFMAC* with a difference in R_{work} and R_{free} of about 2% (Table 5.3,

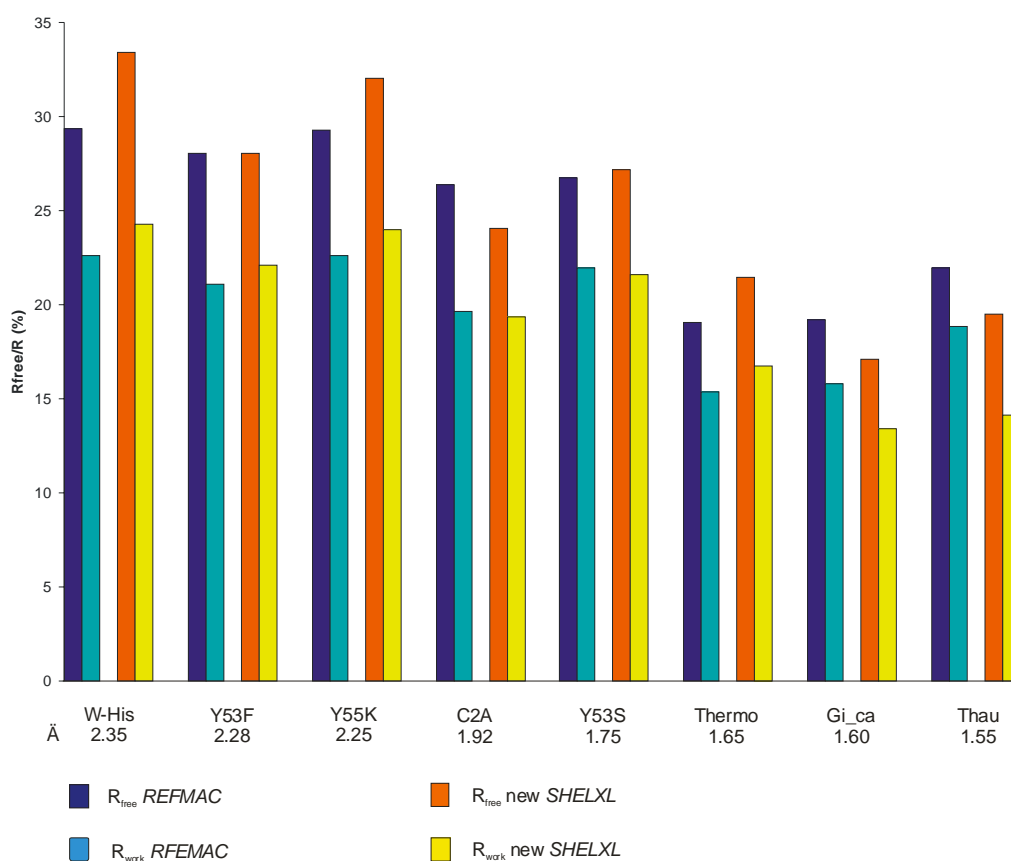


Figure 5.5 Graph that shows the results (R_{work} and R_{free}) obtained using TLS refinement in *REFMAC* and the new version of *SHELXL*.

Data set	job	$R_{\text{work}}(\%)$	$R_{\text{free}}(\%)$	$R_{\text{free}}-R_{\text{work}}(\%)$
W-His	Refmac	22.60	29.34	7.94
	Shelxl	24.28	33.38	9.10
Y53F	Refmac	21.19	28.06	6.87
	Shelxl	22.09	28.06	5.97
Y55K	Refmac	22.63	29.26	6.63
	Shelxl	24.01	32.06	8.03
C2A	Refmac	19.64	26.39	6.75
	Shelxl	19.36	24.10	4.74
Y53S	Refmac	21.93	26.75	4.12
	Shelxl	21.57	27.17	5.60
Thermo	Refmac	15.37	19.09	3.72
	Shelxl	16.70	21.47	4.77
Gi_ca	Refmac	15.79	19.60	3.81
	Shelxl	13.42	17.12	3.70
Thau	Refmac	18.84	21.98	3.14
	Shelxl	14.14	19.49	5.35

Table 5.3 The different data sets used and jobs carried out with the corresponding R_{free} and R_{work} are listed.

Fig. 5.5).

From these results it is evident that *REFMAC* is working better at low resolution while the new version of *SHELXL* is a good choice at medium resolution. *REFMAC* is probably using a better solvent model and non-crystallographic symmetry is handled better when present. The big advantage of *SHELXL* versus *REFMAC* is certainly the fact that no domains have to be defined.

5.3.2 Comparison between old and new *SHELXL*

Fourteen data sets were used to compare the old DELU restraint in *SHELXL* with the new one. Except for two data sets at high resolution (better than 1.3 Å; data sets C2B and tPphA), for which a complete anisotropic refinement is justified, in all the other cases but one the results of the new version of *SHELXL* are better than the ones obtained with the old version (Table 5.4, Fig. 5.6). The R_{free} are generally lower but more importantly the differences between R_{free} and R_{work} became smaller.

		$R_{\text{work}}(\%)$	$R_{\text{free}}(\%)$	$R_{\text{free}}-R_{\text{work}}(\%)$	# may be split atoms	# NPD atoms
W-His	old	20.93	32.64	11.71	232	245
	new	24.28	33.38	9.1	0	0
Y53F	old	18.28	30.11	11.83	65	3
	new	22.09	28.06	5.97	0	0
Y55K	old	23.32	32.26	8.94	72	1
	new	24.01	32.06	8.03	0	0
AR	old	14.98	23.73	8.75	580	25
	new	19.36	24.10	4.74	95	1
Con20	old	14.58	26.04	11.46	126	21
	new	16.94	24.24	7.30	5	0
C2A	old	14.98	23.73	8.75	121	20
	new	19.36	24.10	4.74	8	0
Con17	old	19.61	28.24	8.63	299	58
	new	21.96	26.49	4.53	1	0
Visc-a3	old	15.86	26.67	10.81	44	3
	new	18.17	24.00	5.83	1	0
Y53S	old	18.46	27.44	8.98	37	0
	new	21.57	27.17	5.60	0	0
Thermo	old	15.25	22.19	6.94	29	11
	new	16.70	21.47	4.77	0	0
Gi_ca	old	12.1	17.6	5.5	190	36
	new	13.4	17.1	3.7	14	0
Thau	old	14.08	19.99	5.91	109	16
	new	14.14	19.49	5.35	17	0
C2B	old	14.98	19.53	4.55	114	0
	new	16.80	19.93	3.13	3	0
tPphA	old	16.44	20.54	4.10	102	0
	new	17.72	21.24	3.52	3	0

Table 5.4 The different data sets used and jobs carried out with the corresponding R_{free} , R_{work} are listed. In the last two columns the number of atoms that were flagged as "may be split" or "NPD" are listed.

At low and medium resolution (up to 1.75 Å; data sets Y55K, AR, Con20, C2A, Con17, Visc-a3, Y53S) this gap is reduced on average by about 4%. At higher resolution (data sets Thermo, Gi_ca, Thau, C2B, tPphA) an anisotropic refinement for individual atoms start to be plausible and the difference in the gap between R_{free} and R_{work} is not so evident (between 0.5 and 1.8%). *SHELXL* produces also a list file (*.lst*) where the trend of

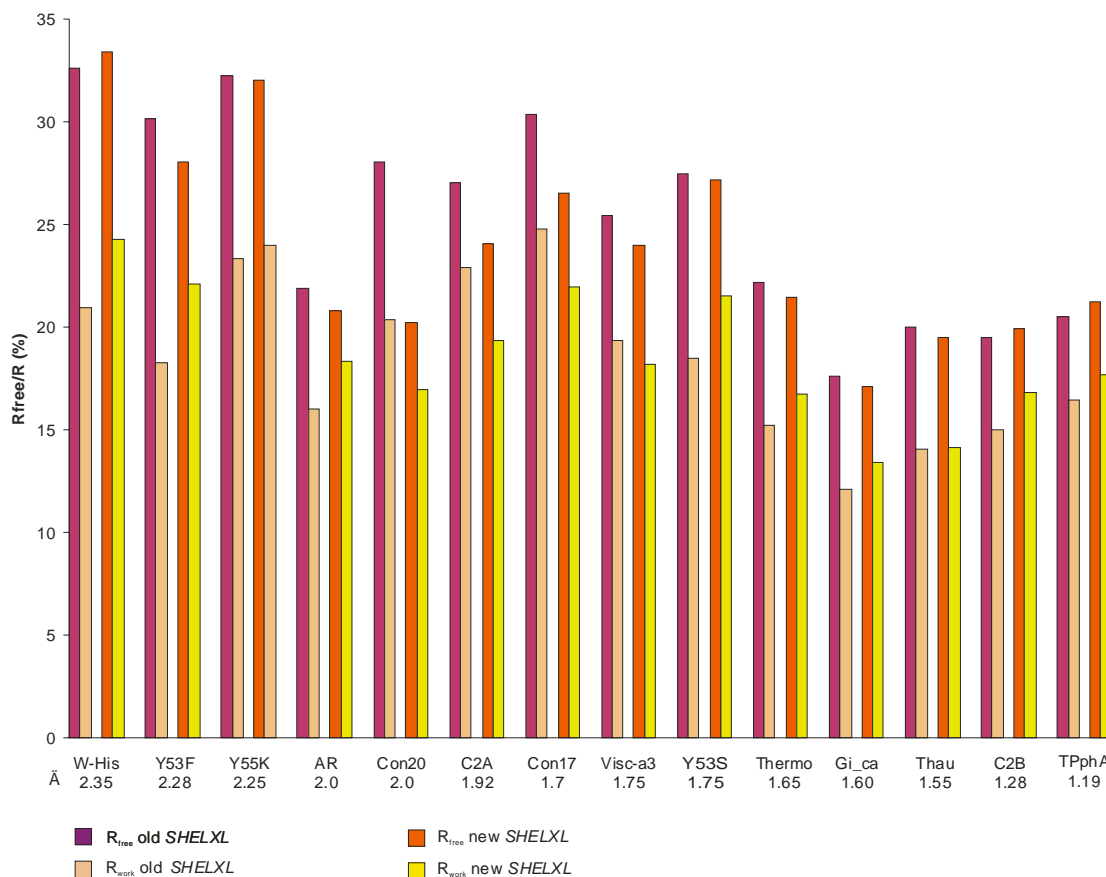


Figure 5.6 Graph that shows the results (R_{work} and R_{free}) obtained using the old and the new version of *SHELXL*.

refinement is stored together with other useful information. In this file are listed all the atoms that were flagged as "non positive definite" (NPD), which means that the atoms are described by an ellipsoid that has no physical sense. Using the old version all data sets except C2B, tPphA and Y53S got at least one atom NPD with the worst being W-His with 245 NPD atoms. On the contrary using the new version of the DELU restraint all atoms for all data sets behave well with just one NPD atom for the data set C2A. In this file are written out also atoms for which the anisotropic displacement parameters are quite big meaning that these atoms could be refined at more than one site. These atoms got flagged as "may be split". Using the old version of *SHELXL* at least 29 atoms per data set (Thermo) were flagged as "may be split". In the structure of Concavalin at 1.7 Å

resolution (Con17) even 299 atoms out of 1874 were suggested to be refined at two sites. It seems that the new version of the DELU restraint improves also this parameter. Indeed an average of only 4.3 atoms per structure were flagged as "may be split". Thaumatin was the worst data set with 17 atoms that could be refined in double position. Con17 that was the worst with the old version instead got only one "may be split" warning.

All these observations clearly show that the newly implemented DELU restraint works much better than the old one especially at medium resolution (1.5-2.0 Å).

5.4 Conclusion and perspectives

A new version of *SHELXL* in which a new DELU restraint was implemented was tested. This new restraint mimics a TLS refinement and it should consequently be useful at medium resolution when a simply isotropic refinement is not satisfactory and a fully anisotropic refinement is not justifiable. Comparisons were made between TLS refinement as implemented in *REFMAC* and the new version of *SHELXL* and between two versions of *SHELXL*. The results show that at low resolution (worse than 2.0 Å) *REFMAC* works better than *SHELXL* while at medium resolution (1.6-2.0 Å) the results are comparable. The new implementation of DELU restraint, however, is a big improvement compare to the previous one.

More tests have to be made to find the best values for the default parameters both for the distance between the atoms considered and for the estimated standard deviation that has to be used.

APPENDIX A

Structure of the Aldose Reductase inhibitor IDD594

Introduction

IDD594 is an Aldose Reductase inhibitor that belongs to the carboxylic acid type classes. It has a good affinity *in vitro* but a low efficiency *in vivo*. Due to its low pK_a value, at physiological pH it is ionized and does not easily cross the membrane (Costantino *et al.* 1999; Miyamoto 2002). The carboxylate group is necessary for efficient inhibition while the aromatic side chain is critical for both potency and selectivity. The structure to 0.66 Å of the complex of IDD594 with AR has been solved (PDB-ID: 1us0, (Howard *et al.* 2004)). The structure of the inhibitor has been determined to 0.83 Å with the aim to use the data in further charge density studies to describe its electrostatic properties.

Materials and methods

The inhibitor IDD594 was dissolved in acetone and crystallized by slow solvent evaporation. Single yellowish crystals were obtained after one month. Data to 0.838 Å from a crystal of dimension 0.2x0.2x0.15 mm³ coated with paraffin oil were collected on a Stoe Image Plate IPDS-II system equipped with a Bruker AXS CCD detector using MoK α radiation ($\lambda = 0.71073$ Å) at 133 K. Data were indexed, integrated and scaled with the Stoe program *X-AREA*. Data analysis and merging was performed with the Bruker program *XPREF*. The structure was solved by direct methods with the program *SHELXD* (Sheldrick 2004) and further refined against intensities with *SHELXL* (Sheldrick *et al.* 1997). Data and refinement statistics are shown in Table A.

Chemical formula	C ₁₆ H ₁₂ BrF ₂ NO ₃ S
Space group	P-1
Unit cell parameters (Å, °)	a = 10.740(2), b = 12.006(2), c = 13.180(3); $\alpha = 77.75(3)$, $\beta = 69.08(3)$, $\gamma = 82.81(3)$
Volume (Å ³)	1549.04
Z	4
Crystal size (mm)	0.20 x 0.20 x 0.15
D_x (mg/m ³)	1.785
Absorption coefficient (mm ⁻¹)	2.826
$F(000)$	832
Reflections collected	23262
Independent reflections ($I > 2\sigma$)	5335
R_{int} § (%)	7.79
Mean (I / σ)	16.2
θ range (°)	1.68 - 24.86
Data/restraints/parameters	5335/0/451
Goodness of fit on F^2	1.050
R1 (%) ($I > 2\sigma(I)$)	4.31
Largest diff. peak/hole	0.463/-0.628

Table A Data and refinement statistics for inhibitor IDD594 ($\% R_{\text{int}} = \frac{\sum (|I - \langle I \rangle|)}{\sum I}$;
 $R1 = \frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$).

Structure

The structure of the inhibitor IDD594 was determined to a resolution of 0.83 Å. It is characterized by a disordered Fluorine atom on the benzyl ring that has attached also a Bromine atom. When the Fluorine is bound to C27 has an occupancy of 80.2 % and when it is bound to C26 is occupancy is 19.8% (Fig. AppA_1A). Two molecules of the inhibitor are coupled by hydrogen bonds between the two carboxylic heads (Fig. AppA_1B). Bond lengths and bond angles are listed in Tables B/C respectively.

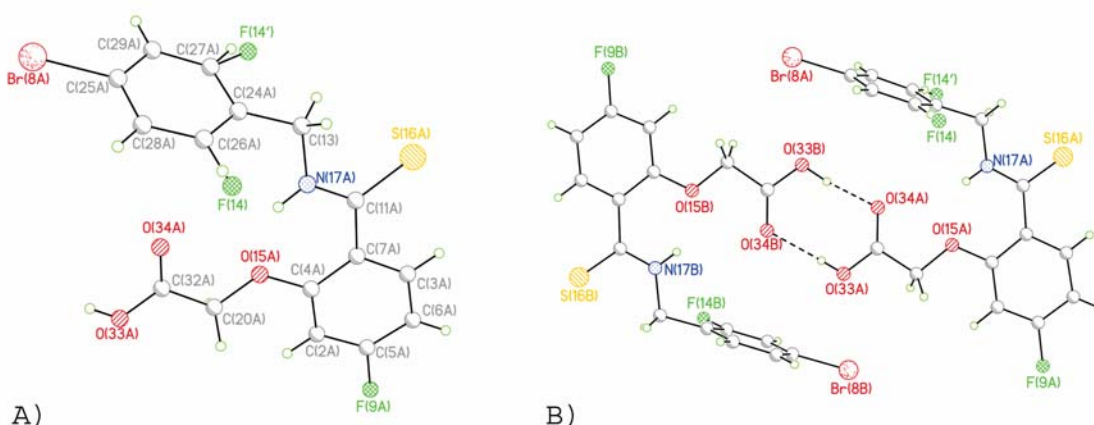


Figure AppA_1. A) Monomer of the inhibitor IDD594. B) Dimer stabilized by hydrogen bonds between the carboxylic heads.

The structure of the inhibitor was compared to the one of IDD594 in the crystal structure of the complex with AR refined to a resolution of 0.66 Å (Fig. AppA_2). The largest difference between the two structures is the protonation state of the carboxylic head: the isolated inhibitor is protonated as it can be seen both from the hydrogen bond and from the oxygen carbon distances ($O33A...C32A = 1.233$ Å, $O34A...C32A = 1.291$ Å) while in the complex with AR it is negatively charged ($O33...C32 = 1.23(9)$ Å and $O34...C32 = 1.26(2)$ Å). The orientation of the carboxylic head is different in the two structures due to diverse modes of interaction with surrounding atoms. Also the mutual orientation of the aromatic rings is different in the two structures. The distance between Br8 and C25 is instead more or less the same (1.89 Å and 1.88 Å in the isolated and complex structure respectively).

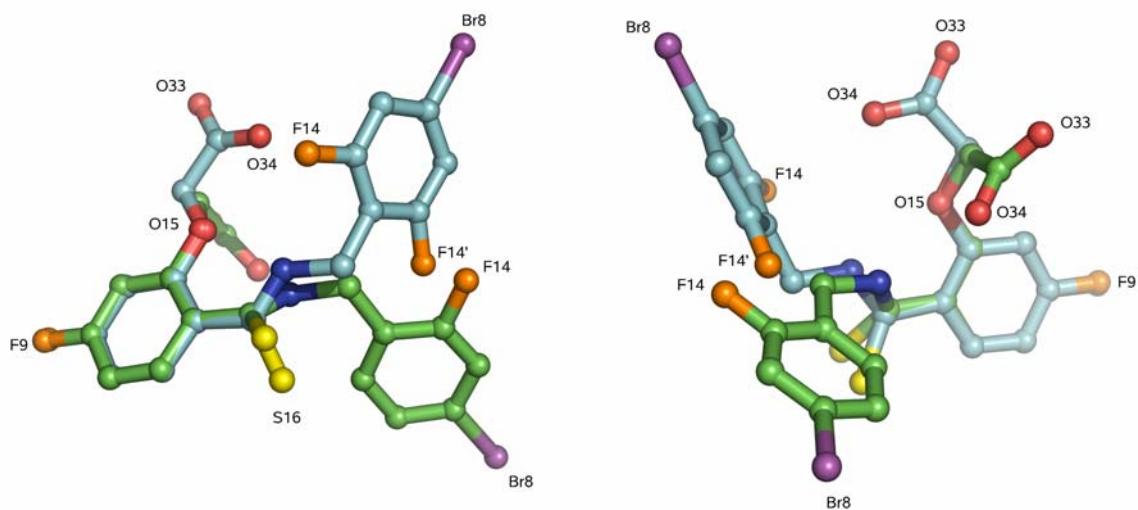


Figure AppA_2 Superposition of inhibitor IDD594 in the isolated structure (Carbon atoms depicted in cyan) and in the complex with Aldose Reductase (Carbon atoms depicted in green). Bromines are drawn in magenta, Fluorines in orange, Sulfurs in yellow, Nitrogens in blue and Oxygens in red.

Molecule A		Molecule B	
Br(8A)-C(25A)	1.890(4)	Br(8B)-C(25B)	1.887(4)
S(16A)-C(11A)	1.684(4)	S(16B)-C(11B)	1.667(4)
F(9A)-C(5A)	1.358(4)	F(9B)-C(5B)	1.357(5)
O(34A)-C(32A)	1.233(5)	O(34B)-C(32B)	1.221(5)
O(15A)-C(4A)	1.376(5)	O(15B)-C(4B)	1.367(5)
O(15A)-C(20A)	1.427(5)	O(15B)-C(20B)	1.423(5)
O(33A)-C(32A)	1.291(5)	O(33B)-C(32B)	1.300(5)
C(5A)-C(6A)	1.371(6)	C(6B)-C(5B)	1.368(6)
C(5A)-C(2A)	1.374(6)	C(5B)-C(2B)	1.377(6)
F(14)-C(26A)	1.331(5)		
F(14')-C(27A)	1.276(13)	F(14B)-C(27B)	1.368(5)
C(4A)-C(2A)	1.398(5)	C(2B)-C(4B)	1.392(5)
C(4A)-C(7A)	1.405(6)	C(4B)-C(7B)	1.414(6)
N(17A)-C(11A)	1.329(5)	N(17B)-C(11B)	1.324(5)
N(17A)-C(13)	1.472(5)	N(17B)-C(13B)	1.465(5)
C(11A)-C(7A)	1.503(5)	C(11B)-C(7B)	1.515(5)
C(25A)-C(29A)	1.380(6)	C(25B)-C(29B)	1.377(6)
C(25A)-C(28A)	1.389(6)	C(25B)-C(28B)	1.391(6)
C(28A)-C(26A)	1.380(6)	C(28B)-C(27B)	1.384(6)
C(7A)-C(3A)	1.397(6)	C(3B)-C(7B)	1.394(6)
C(6A)-C(3A)	1.385(6)	C(6B)-C(3B)	1.391(6)
C(32A)-C(20A)	1.498(5)	C(32B)-C(20B)	1.509(5)
C(24A)-C(26A)	1.382(7)	C(24B)-C(26B)	1.399(6)
C(24A)-C(27A)	1.388(6)	C(24B)-C(27B)	1.384(6)
C(24A)-C(13)	1.501(6)	C(24B)-C(13B)	1.502(6)
C(27A)-C(29A)	1.388(6)	C(29B)-C(26B)	1.391(6)

Table B Bond distances (Å) of isolated IDD594.

Molecule A		Molecule B	
C(4A)-O(15A)-C(20A)	117.9(3)	C(4B)-O(15B)-C(20B)	117.6(3)
F(9A)-C(5A)-C(6A)	119.3(4)	F(9B)-C(5B)-C(6B)	119.1(4)
F(9A)-C(5A)-C(2A)	117.2(4)	F(9B)-C(5B)-C(2B)	117.6(4)
C(6A)-C(5A)-C(2A)	123.5(4)	C(6B)-C(5B)-C(2B)	123.3(4)
O(15A)-C(4A)-C(2A)	120.8(4)	O(15B)-C(4B)-C(2B)	121.5(4)
O(15A)-C(4A)-C(7A)	118.1(3)	O(15B)-C(4B)-C(7B)	117.8(3)
C(2A)-C(4A)-C(7A)	121.1(3)	C(2B)-C(4B)-C(7B)	120.7(4)
C(11A)-N(17A)-C(13)	121.6(4)	C(11B)-N(17B)-C(13B)	123.4(3)
N(17A)-C(11A)-C(7A)	119.1(4)	N(17B)-C(11B)-C(7B)	118.4(4)
N(17A)-C(11A)-S(16A)	119.9(3)	N(17B)-C(11B)-S(16B)	120.8(3)
C(7A)-C(11A)-S(16A)	121.0(3)	C(7B)-C(11B)-S(16B)	120.8(3)
C(29A)-C(25A)-C(28A)	121.6(4)	C(29B)-C(25B)-C(28B)	120.9(4)
C(29A)-C(25A)-Br(8A)	118.7(3)	C(29B)-C(25B)-Br(8B)	120.7(3)
C(28A)-C(25A)-Br(8A)	119.8(3)	C(28B)-C(25B)-Br(8B)	118.5(3)
C(26A)-C(28A)-C(25A)	116.9(4)	C(27B)-C(28B)-C(25B)	117.1(4)
C(3A)-C(7A)-C(4A)	116.8(3)	C(3B)-C(7B)-C(4B)	117.2(4)
C(3A)-C(7A)-C(11A)	117.7(4)	C(3B)-C(7B)-C(11B)	117.7(4)
C(4A)-C(7A)-C(11A)	125.5(3)	C(4B)-C(7B)-C(11B)	125.0(4)
C(5A)-C(6A)-C(3A)	116.9(4)	C(5B)-C(6B)-C(3B)	117.3(4)
O(34A)-C(32A)-O(33A)	126.2(4)	O(34B)-C(32B)-O(33B)	125.6(4)
O(34A)-C(32A)-C(20A)	122.4(3)	O(34B)-C(32B)-C(20B)	123.7(4)
O(33A)-C(32A)-C(20A)	111.4(3)	O(33B)-C(32B)-C(20B)	110.6(3)
C(26A)-C(24A)-C(27A)	116.1(4)	C(27B)-C(24B)-C(26B)	116.3(4)
C(26A)-C(24A)-C(13)	123.2(4)	C(26B)-C(24B)-C(13B)	121.9(4)
C(27A)-C(24A)-C(13)	120.6(4)	C(27B)-C(24B)-C(13B)	121.7(4)
F(14)-C(26A)-C(28A)	120.0(4)	F(14B)-C(27B)-C(28B)	118.2(4)
F(14)-C(26A)-C(24A)	115.6(4)	F(14B)-C(27B)-C(24B)	117.3(4)
C(28A)-C(26A)-C(24A)	124.4(4)	C(24B)-C(27B)-C(28B)	124.5(4)
C(5A)-C(2A)-C(4A)	118.3(4)	C(5B)-C(2B)-C(4B)	118.6(4)
C(6A)-C(3A)-C(7A)	123.4(4)	C(6B)-C(3B)-C(7B)	122.8(4)
N(17A)-C(13)-C(24A)	112.4(3)	N(17B)-C(13B)-C(24B)	111.9(3)
F(14')-C(27A)-C(29A)	132.9(8)		
F(14')-C(27A)-C(24A)	104.8(7)		
C(29A)-C(27A)-C(24A)	122.2(4)	C(25B)-C(29B)-C(26B)	120.2(4)
O(15A)-C(20A)-C(32A)	108.8(3)	O(15B)-C(20B)-C(32B)	108.5(3)
C(25A)-C(29A)-C(27A)	118.7(4)	C(29B)-C(26B)-C(24B)	121.0(4)

Table C Bond angles (°) of isolated IDD594

APPENDIX B

Free variables used in the refinement of Aldose Reductase

FVAR	occupancy main conformation (%)	Residue	Atoms
2	58.3	E126	side chain
3	51.1	R40	all
		V67	side chain
4	60.9	K100	all
		HOH4030	
		HOH4031	
5	89.8	L104	side chain
6	72.7	P117	side chain
7	50.5	P132	all
8	68.3	H83	all
		E84	all
		K85	all
		G86	all
		L87	all
		S133	all
		D134	all
		T135	all
		N136	all
		I137	all
		HOH4034	
		HOH4035	
		HOH8001	
		HOH8002	
		HOH8003	
		HOH8004	
		HOH8005	
		HOH8006	
		HOH8007	
		HOH8008	
9	51.5	H163	all
10	71.3	M168	side chain
11	53.1	I169	side chain
12	94.4	L175	all
13	64.3	F276	all
		D277	all
		F278	all
14	64.5	K274	side chain
		HOH4022	
		HOH4023	
		HOH4024	

FVAR	occupancy main conformation (%)	Residue	Atoms
15	33.0	E279	side chain
16	64.4	S282	side chain
17	80.5	T244	side chain
18	96.3	T243	side chain
19	54.8	T304	all
		S305	all
		HOH4025	
		HOH4026	
		HOH4027	
20	54.5	Q93	side chain
		S97	side chain
		HOH4005	
		HOH4029	
21	55.3	M2001	all
22	59.6	D102	all
		HOH4002	
		HOH4003	
		HOH4010	
		HOH4011	
23	55.8	K116	side chain
		HOH4008	
24	73.8	E64	all
25	58.3	I152	side chain
26	68.2	K154	side chain
27	53.2	P179	CG
28	72.1	E193	side chain
		HOH4015	
29	81.1	K202	side chain
30	57.2	E229	side chain
		K234	side chain
		HOH4004	
		HOH4028	
31	63.6	K239	side chain
33	63.7	K21	side chain
		D216	all
		R217	all
		P218	all
		W219	all
		A220	all
		K221	all
		P222	all
		E223	all
		D224	all
		N256	all
		C298	all
		CIT320	all

FVAR	occupancy main conformation (%)	Residue	Atoms
33	63.7	HOH4110 HOH4006 HOH4013 HOH4016 HOH4017 HOH4018 HOH4019 HOH4020 HOH4021	
35	62.0	I4 K194 HOH4009	all side chain
37	59.9	E60 HOH4032	side chain
38	53.4	Q197 HOH4012	side chain
39	66.2	N241	side chain
40	58.3	I15	all
41	68.5	V37 G38	all all
42	56.5	K119	side chain
43	69.1	L124 D125 S127 G128 N129 V130 V131 HOH4014 HOH4036	all all all all all all all
44	79.6	E271 HOH4007	side chain
45	69.4	R293	all

APPENDIX C

List of disagreeable restraints

List of antibump (BUMP) disagreeable restraints longer than 5 sigma

Observed	Target	Δ (Deviation)	Δ / σ	Atoms
1.9598	2.1000	-0.1402	7.01	HD2B_136a H0_139
1.9691	2.1000	-0.1309	6.55	H0A_84b HG3_84b
3.1699	3.3000	-0.1301	6.50	C_197 CD_197b
3.1762	3.3000	-0.1238	6.16	C_76 CG_77
3.1850	3.3000	-0.1150	5.75	C_209 CD1_209
2.6861	2.8000	-0.1139	5.69	O_299 C_300
1.9865	2.1000	-0.1135	5.67	HG1C_66 HG2_132a
3.1927	3.3000	-0.1073	5.32	CD1_209 C17_318

List of angle distance (DANG) disagreeable restraints longer than 5 sigma

Observed	Target	Δ (Deviation)	Δ / σ	Atoms
2.0940	2.4350	-0.3410	7.12	C_136a CA_137a
2.0370	2.3690	-0.3320	7.14	C3_321a O5_321a
2.7994	2.5100	0.2894	7.23	CG_194a CE_194a
2.2383	2.5260	-0.2877	7.14	CB_84b CD_84b
2.1516	2.4350	-0.2834	7.08	C_215 CA_216b
2.8430	2.5590	0.2840	7.10	CA_126 CG_126a
2.7913	2.5100	0.2813	7.04	CB_194a CD_194a
2.7538	2.4790	0.2748	6.85	N_137b CB_137b
2.2060	2.4790	-0.2730	6.82	N_137a CB_137a
2.1887	2.4550	-0.2663	6.66	CB_129b N_129b
2.6350	2.3730	0.2620	6.55	OG1_135a CG2_135a
2.1376	2.3950	-0.2574	6.39	C6_321a O7_321a
2.2508	2.5040	-0.2532	6.34	CG1_67b CG2_67b
2.6616	2.4350	0.2266	5.66	C_222a CA_223a
2.6419	2.4250	0.2169	5.40	CA_305b N_306
2.7740	2.5590	0.2150	5.37	CA_223a CG_223a
2.3665	2.5760	-0.2095	5.23	CA_169 CG1_169a
2.7018	2.4930	0.2088	5.22	CD_221a NZ_221a
2.5869	2.3790	0.2079	5.14	CG_70a OE1_70a
2.3330	2.5340	-0.2010	5.02	CA_136b CG_136b

List of flatness (FLAT) disagreeable restraints bigger than 5 sigma

Δ (Deviation)	Δ / σ	Atoms
1.7270	17.23	CB_209 CD1_209 CE2_209 OH_209 etc
0.9871	9.87	CB_39 CD1_39 CE2_39 OH_39 etc.
0.9540	9.54	CB_79 CE2_79 CZ3_79 CG_79
0.8539	8.54	CB_107 CD1_107 CE2_107 OH_107 etc.
0.7579	7.58	CB_48 CD1_48 CE2_48 OH_48 etc.
0.5945	5.94	CD_255 NH1_255 NH2_255 NE_255
0.5709	5.71	CD_69 NH1_69 NH2_69 NE_69

List of flatness (FLAT) disagreeable restraints for peptide bond bigger than 3 sigma

Δ (Deviation)	Δ / σ	Atoms
1.9509	3.90	O_132 CA_132 N_133a CA_133a
1.9175	3.83	O_181 CA_181 N_182 CA_182
1.7576	3.51	O_108 CA_108 N_109 CA_109
1.7205	3.44	O_132 CA_132 N_133b CA_133b
1.6890	3.38	O_276b CA_276b N_277b CA_277b
1.6740	3.35	O_297 CA_297 N_298b CA_298b
1.6147	3.23	O_310 CA_310 N_311 CA_311
1.5996	3.20	O_257 CA_257 N_258 CA_258
1.5591	3.12	O_296 CA_296 N_297 CA_297
1.5317	3.06	O_311 CA_311 N_312 CA_312
1.5227	3.05	O_102a CA_102a N_103 CA_103
1.5128	3.03	O_124a CA_124a N_125a CA_125a

List of chiral volume (CHIV) disagreeable restraints bigger than 5 sigma

Observed	Target	Δ (Deviation)	Δ / σ	Atoms
-0.7857	0.0000	-0.7857	7.86	N_310 C_309 CA_310 CD_310
-0.5948	0.0000	-0.5948	5.85	N_117 C_116 CA_117 CD_117b
-0.5252	0.0000	-0.5252	5.25	N_188 C_187 CA_188 CD_188
1.9827	2.5030	-0.5203	5.20	CA_21 C_21 CB_21b N_21

List of distance (DFIX) disagreeable restraints longer or shorter than 3 sigma

Observed	Target	Δ (Deviation)	Δ / σ	Atoms
1.1512	1.2490	-0.0978	-4.89	CD_150a OE2_150a
1.2384	1.3310	-0.0926	-4.63	C6_320a O5_320a
1.1578	1.2490	-0.0912	-4.56	C_315a OT2_315a
1.4304	1.5210	-0.0906	-4.53	CG_152b CD1_152b
1.4178	1.3290	0.0888	4.44	C_305a N_306
1.4446	1.5300	-0.0854	-4.27	CA_216b CB_216b
1.4071	1.4920	0.0849	-4.25	CG_231 CB_231
1.7079	1.7910	-0.0831	-4.16	SD_2001a CE_2001a
1.2455	1.3240	-0.0785	-3.93	C5_321a O3_321a
1.5808	1.5030	0.0778	3.89	CD_132 CG_132b
1.3426	1.4170	-0.0744	3.72	CB_305a OG_305a
1.2562	1.3290	-0.0728	-3.64	C_222a N_223a
1.4451	1.5160	-0.0709	-3.55	CB_216b CG_216b
1.3997	1.3290	0.0707	3.54	C_100b N_101
1.4506	1.5200	-0.0694	-3.47	CG_11 CD_11
1.4855	1.4170	0.0685	3.43	CB_133a OG_133a
1.5981	1.5300	0.0681	3.41	CA_305b CB_305b
1.5880	1.5210	0.0670	3.35	CG_152b CD2_152b
1.5551	1.4890	0.0661	3.31	CE_242 NZ_242
1.3146	1.2490	0.0656	3.28	CD_70a OE2_70a
1.5364	1.4730	0.0634	3.17	N_218a CD_218a
1.4933	1.4300	0.0633	3.17	C3_321a O7_321a
1.4731	1.5350	-0.0619	-3.10	C2_321a C3_321a
1.4546	1.5160	-0.0614	-3.07	CG_65 CD_65
1.3990	1.4600	-0.0610	-3.05	CD_40b NE_40b
1.3097	1.2490	0.0607	3.04	CD_70a OE1_70a

REFERENCES

- Allen, F. H., *Acta Cryst. B*, (2002), **B58**, 380-388.
- Bai, J. and Chapman, E. R., *Trends Biochem. Sci*, (2004), **29**, 143-151.
- Baldini, G., Martelli, A. M., Tabellini, G., Horn, C., Machaca, K., Narducci, P. and Baldini, G., *J. Biol. Chem.*, (2005), **280**, 34974-34984.
- Bennett, M. K., Calakos, N. and Scheller, R. H., *Science*, (1992), **257**, 255-259.
- Biadene, M., Montaville, P., Sheldrick, G. M. and Becker, S., *Acta Cryst. D*, (2006), **62**, 793-799.
- Bohrani, D. W., Harter, T. M. and Petrash, J. M., *J. Biol. Chem.*, (1992), **267**, 24841-24847.
- Bohren, K. M., Brownlee, J. M., Milne, A. C., Gabbay, K. H. and Harrison, D. H., *Biochim. Biophys. Acta*, (2005), **1748**, 201-212.
- Bohren, K. M., Grimshaw, C. E., Lai, C. J., Harrison, D. H., Ringe, D., Petsko, G. A. and Gabbay, K. H., *Biochemistry*, (1994), **33**, 2021-2032.
- Brunger, A. T., *Nature*, (1992), **355**, 472-475.
- Buckley, K. and Kelly, R. B., *J. Cell Biol.*, (1985), **100**, 1284-1294.
- Burns, M. E., Sasaki, T., Takai, Y. and Augustine, G. J., *J. Gen. Physiol.*, (1998), **111**, 243-255.
- Cachau, R. E., Howard, E. I., Barth, P., Mitschler, A., Chevrier, B., Lamour, V., Joachimiak, A., Saninhvili, A., Van Zandt, M., Sibley, E., Moras, D. and Podjarny, A., *J. Phys. IV*, (2002), **10**, 3-13.
- Calakos, N. and Scheller, R. H., *J. Biol. Chem.*, (1994), **269**, 24534-24537.
- Calderone, V., Chevrier, B., Van Zandt, M., Lamour, V., Howard, E., Poterszman, A., Barth, P., Mitschler, A., Lu, J., Dvornik, D. M., Klebe, G., Kraemer, O., Moorman, A. R., Moras, D. and Podjarny, A., *Acta Cryst. D*, (2000), **D56**, 536-540.
- Chapman, E. R., Hanson, P. I., An, S. and Jahn, R., *J. Biol. Chem.*, (1995), **270**, 23667-23671.
- Chen, Y. A. and Scheller, R. H., *Nat. Rev. Mol. Cell Biol.*, (2001), **2**, 98-106.

- Chung, S.-H., Song, W.-J., Kim, K., Bednarski, J. J., Chen, J., Prestwich, G. D. and Holz, R. W., *J. Biol. Chem.*, (1998), **273**, 10240-10248.
- Chung, S. H., Stabila, P., Macara, I. G. and Holz, R. W., *J. Neurochem.*, (1997), **69**, 164-173.
- Chung, S. H., Takai, Y. and Holz, R. W., *J. Biol. Chem.*, (1995), **270**, 16714-16718.
- Cochran, W., *Acta Cryst.*, (1955), **8**, 473-478.
- Collaborative Computational Project, N., *Acta Cryst. D*, (1994), **D50**, 760-763.
- Corbalan-Garcia, S., Garcia-Garcia, J., Rodriguez-Alfaro, J. A. and Gomez-Fernandez, J. C., *J. Biol. Chem.*, (2003), **278**, 4972-4980.
- Costantino, L., Rastelli, G., Gamberini, M. C. and Barlocco, D., *Exp. Opinion on Therapeutic Patents*, (2000), **10**, 1245-1262.
- Costantino, L., Rastelli, G., Vianello, P., Cignarella, G. and Barlocco, D., *Med. Res. Rev.*, (1999), **19**, 3-23.
- Dai, H., Shin, O. H., Machius, M., Tomchick, D. R., Sudhof, T. C. and Rizo, J., *Nat Struct Mol Biol*, (2004), **11**, 844-9.
- Dauter, Z., *Methods Enzymol.*, (2003), **368**, 288-337.
- DeLano, W. L., *www.pymol.org*, (2002).
- Didisheim, J.-J. and Schwarzenbach, D., *Acta Cryst. A*, (1987), **A43**, 226-232.
- Duan, G., Smith Jr, V. H. and Weaver, D. F., *Mol. Phys.*, (2001), **99**, 1689-1699.
- Dunlop, M., *Kidney. Int. Suppl.*, (2000), **77**, S3-S12.
- Edelmann, L., Hanson, P. I., Chapman, E. R. and Jahn, R., *EMBO J.*, (1995), **14**, 224-231.
- El-Kabbani, O., Darmanin, C., Oka, M., Schulze-Briese, C., Tomizaki, T., Hazemann, I., Mitschler, A. and Podjarny, A., *J. Med. Chem.*, (2004a), **47**, 4530-4537.
- El-Kabbani, O., Darmanin, C., Schneider, T. R., Hazemann, I., Ruiz, F., Oka, M., Joachimiak, A., Schulze-Briese, C., Tomizaki, T., Mitschler, A. and Podjarny, A., *Proteins*, (2004b), **55**, 805-813.
- El-Kabbani, O., Ramsland, P., Darmanin, C., Chung, R. P.-T. and Podjarny, A., *Proteins*, (2003), **50**, 230-238.

- El-Kabbani, O., Ruiz, F., Darmanin, C. and Chung, R. P.-T., *Cell. Mol. Life Sci.*, (2004c), **61**, 750-762.
- El-Kabbani, O., Wilson, D. K., Petrash, M. and Quioco, F. A., *Mol. Vis.*, (1998), **4**, 19.
- Emsley, P. and Cowtan, K., *Acta Cryst. D*, (2004), **D60**, 2126-2132.
- Engh, R. and Huber, R., *Acta Cryst. A*, (1991), **A47**, 392-400.
- Esnouf, R. M., *J. Mol. Graphics Modell.*, (1997), **15**, 132-134.
- Essen, L.-O., Perisic, O., Cheung, R., Katan, M. and Williams, R. L., *Nature*, (1996), **380**, 595-602.
- Fernandez, I., Arac, D., Ubach, J., Gerber, S. H., Shin, O., Gao, Y., Anderson, R. G. W., Sudhof, T. C. and Rizo, J., *Neuron*, (2001), **32**, 1057-1069.
- Fukuda, M., *J. Biol. Chem.*, (2003), **278**, 15373-15380.
- Fykse, E. M. and Fonnum, F., *Neurochem. Res.*, (1996), **21**, 1053-1060.
- Grimshaw, C. E., Bohren, K. M., Lai, C. J. and Gabbay, K. H., *Biochemistry*, (1995), **34**, 14356-14365.
- Grimshaw, C. E., Shahbaz, M. and Putney, C. G., *Biochemistry*, (1990), **29**, 9936-46.
- Groffen, A. J. A., Friedrich, R., Brian, E. C., Ashery, U. and Verhage, M., *J. Neurochem.*, (2006), **97**, 818-833.
- Hakansson, K., *Int. J. Biol. Macromol.*, (1996), **18**, 189-194.
- Harrison, D. H., Bohren, K. M., Petsko, G. A., Ringe, D. and Gabbay, K. H., *Biochemistry*, (1997), **36**, 16134-16140.
- Harrison, D. H., Bohren, K. M., Ringe, D., Petsko, G. A. and Gabbay, K. H., *Biochemistry*, (1994), **33**, 2011-2020.
- Hendrickson, W. A., Smith, J. L. and Sheriff, S., *Methods Enzymol.*, (1985), **115**, 41-55.
- Hirshfeld, F. L., *Acta Cryst. A*, (1976), **A32**, 239-244.
- Honig, B. and Nicholls, A., *Science*, (1995), **268**, 1144-1149.

- 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., *Proteins*, (2004), **55**, 792-804.
- Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. and Driessen, H. P. C., *J. Appl. Cryst.*, (1993), **26**, 622-624.
- Howlin, B., Moss, D. S. and Harris, G. W., *Acta Cryst. A*, (1989), **A45**, 851-861.
- Jahn, R., Lang, T. and Sudhof, T. C., *Cell*, (2003), **112**, 519-533.
- Janz, R., Goda, Y., Geppert, M., Missler, M. and Sudhof, T. C., *Neuron*, (1999), **24**, 1003-1016.
- Johnston, P. A. and Sudhof, T. C., *J. Biol. Chem.*, (1990), **265**, 10524-10528.
- Karle, J. and Hauptman, H., *Acta Cryst.*, (1956), **9**, 635-651.
- Kato, M., Sasaki, T., Ohya, T., Nakanishi, H., Nishioka, H., Imamura, M. and Yoshimi, T., *J. Biol. Chem.*, (1996), **271**, 31775-31778.
- Khvotchev, M. V., Ren, M., Takamori, S., Jahn, R. and Sudhof, T. C., *J. Neurosci.*, (2003), **23**, 10531-10539.
- Kinoshita, J. H. and Nishimura, C., *Diabetes Metab. Rev.*, (1988), **4**, 323-337.
- Kinoshita, T., Miyake, H., Fujii, T., Takakura, S. and Goto, T., *Acta Cryst. D*, (2002), **D58**, 622-626.
- Kissinger, C. R., Gehlhaar, D. K. and Fogel, D. B., *Acta Cryst. D*, (1999), **D55**, 484-491.
- Klebe, G., Kramer, O. and Sotriffer, C., *Cell. Mol. Life Sci.*, (2004), **61**, 783-793.
- Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wahlby, A. and Jones, T. A., *Acta Cryst. D*, (2004), **60**, 2240-2249.
- Krentz, A. J., Honigsberger, L., Ellis, S. H., Hardman, M. and Nattrass, M., *Diabet. Med.*, (1992), **9**, 463-468.
- Krissinel, E. and Henrick, K., *Acta Cryst. D*, (2004), **D60**, 2256-2268.
- Kubiseski, T. J., Hyndman, D. J., Morjana, N. A. and Flynn, T. G., *J. Biol. Chem.*, (1992), **267**, 6510-6517.
- Lamzin, V. S., Perrakis, A. and Wilson, K. S., *International Table for Crystallography Volume F: Crystallography for biological macromolecules*

- (2001), M.G. Rossmann and E. Arnold, Dordrecht, The Netherlands, Kluwer Academic Publisher, 720-722.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M., *J. Appl. Cryst.*, (1993), **26**, 283-291.
- Li, C., Takei, K., Geppert, M., Daniell, L., Stenius, K., Chapman, E. R., Jahn, R., De Camilli, P. and Sudhof, T. C., *Neuron*, (1994), **13**, 885-898.
- Li, C., Ullrich, B., Zhang, J. Z., Anderson, R. G. W., Brose, N. and Sudhof, T. C., *Nature*, (1995a), **375**, 594-599.
- Li, L., Chin, L. S., Shupliakov, O., Brodin, L., Sihra, T. S., Hvalby, O., Jensen, V., Zheng, D., McNamara, J. O., Greengard, P. and Andersen, P., *Proc. Natl. Acad. Sci. USA*, (1995b), **92**, 9235-9239.
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. and Read, R. J., *Acta Cryst. D*, (2005), **D61**, 458-464.
- McKiernan, C. J., Stabila, P. and Macara, I. G., *Mol. Cell Biol.*, (1996), **16**, 4985-4995.
- McRee, D. E., *J. Struct. Biol.*, (1999), **125**, 156-165.
- Merritt, E. A., *Acta Cryst. D*, (1999), **D55**, 1109 - 1117.
- Merritt, E. A. and Bacon, D. J., *Methods Enzymol.*, (1997), **277**, 505-524.
- Meyer, E. A., Castellano, R. K. and Diederich, F., *Angew. Chem. Int. Ed. Engl.*, (2003), **42**, 1210-1250.
- Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. and Hauptman, H. A., *Science*, (1993), **259**, 1430-1433.
- Miyamoto, S., *Chem-Bio. Informatics J.*, (2002), **2**, 74-85.
- Miyazaki, M., Shirataki, H., Kohno, H., Kaibuchi, K., Tsugita, A. and Takai, Y., *Biochem. Biophys. Res. Commun.*, (1994), **205**, 460-466.
- Murray, D. and Honig, B., *Mol. Cell* (2002), **9**, 145-154.
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J., *Acta Cryst. D*, (1997), **D53**, 240-255.
- Mylari, B. L., Armento, S. J., Beebe, D. A., Conn, E. L., Coutcher, J. B., Dina, M. S., O' Gorman, M. T., Linhares, M. C., Martin, W. H., Oates, P. J., Tess, D. A., Withbroe, G. J. and Zembrowski, W. J., *J. Med. Chem.*, (2003), **46**, 2283-2286.

- Nalefski, E. A. and Falke, J. J., *Protein Sci.*, (1996), **5**, 2375-2390.
- Ohya, T., Sasaki, T., Kato, M. and Takai, Y., *J. Biol. Chem.*, (1998), **273**, 613-617.
- Otwinowski, Z. and Minor, W., *Methods Enzymol.*, (1997), **276**, 307-326.
- Pannu, N. S. and Read, R. J., *Acta Cryst. A*, (1996), **A52**, 659-668.
- Petrash, J. M., *Cell. Mol. Life Sci.*, (2004), **61**, 737-749.
- Podjarny, A., Cachau, R. E., Schneider, T., Van Zandt, M. and Joachimiak, A., *Cell. Mol. Life Sci.*, (2004), **61**, 763-773.
- Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V., *J. Mol. Biol.*, (1963), **7**, 95-99.
- Rizo, J. and Sudhof, T. C., *J. Biol. Chem.*, (1998), **273**, 15879-15882.
- Rollett, J. S. (1970). Crystallographic Computing. Copenhagen, Munksgaard.
- Rondeau, J.-M., Tete-Favier, F., Podjarny, A., Reymann, J.-M., Barth, P., Biellmann, J.-F. and Moras, D., *Nature*, (1992), **355**, 469-472.
- Rosahl, T. W., Geppert, M., Spillane, D., Herz, J., Hammer, R. E., Malenka, R. C. and Sudhof, T. C., *Cell*, (1993), **75**, 661-670.
- Rosahl, T. W., Spillane, D., Missler, M., Herz, J., Selig, D. K., Wolff, J. R., Hammer, R. E., Malenka, R. C. and Sudhof, T. C., *Nature*, (1995), **365**, 488-493.
- Ruiz, F., Hazemann, I., Mitschler, A., Joachimiak, A., Schneider, T., Karplus, M. and Podjarny, A., *Acta Cryst. D*, (2004), **D60**, 1347-1354.
- Sayre, D., *Acta Cryst.*, (1952), **5**, 60-65.
- Schluter, O. M., Khvotchev, M., Jahn, R. and Sudhof, T. C., *J. Biol. Chem.*, (2002), **277**, 40919-40929.
- Schluter, O. M., Schnell, E., Verhage, M., Tzonopoulos, T., Nicoll, R. A., Janz, R., Malenka, R. C., Geppert, M. and Sudhof, T. C., *J. Neurosci.*, (1999), **19**, 5834-5846.
- Schomaker, V. and Trueblood, K. N., *Acta Cryst. B*, (1968), **B24**, 63-76.
- Schomaker, V. and Trueblood, K. N., *Acta Cryst. B*, (1998), **B54**, 507-514.
- Shao, X., Davletov, B. A., Sutton, R. B., Sudhof, T. C. and Rizo, J., *Science*, (1996), **273**, 248-251.

- Shao, X., Fernandez, I., Sudhof, T. C. and Rizo, J., *Biochemistry*, (1998), **37**, 16106-16115.
- Shao, X., Li, C., Fernandez, I., Zhang, X., Sudhof, T. C. and Rizo, J., *Neuron*, (1997), **18**, 133-142.
- Sheldrick, G. M., *Acta Cryst. A*, (1990), **A46**, 467-473.
- Sheldrick, G. M., *Z. Kristallogr.*, (2002), **217**, 644-650.
- Sheldrick, G. M., *University of Goettingen*, (2004).
- Sheldrick, G. M. and Schneider, T. R., *Methods Enzymol.*, (1997), **277**, 319-343.
- Shin, O. H., Rhee, J. S., Tang, J., Sugita, S., Rosenmund, C. and Sudhof, T. C., *Neuron*, (2003), **37**, 99-108.
- Shirataki, H., Kaibuchi, K., Sakoda, T., Kishida, S., Yamaguchi, T., Wada, K., Miyazaki, M. and Takai, Y., *Mol. Cell Biol.*, (1993), **13**, 2061-2068.
- Sottriffer, C. A., Kramer, O. and Klebe, G., *Proteins*, (2004), **56**, 52-66.
- Spielberg, S. P., Shear, N. H., Cannon, M., Hutson, N. J. and Gunderson, K., *Ann. Intern. Med.*, (1991), **114**, 720-724.
- Stahl, B., von Mollard, G. F., Walch-Solimena, C. and Jahn, R., *J. Biol. Chem.*, (1994), **269**, 24770-24776.
- Staunton, J., Ganetzky, B. and Nonet, M. L., *J. Neurosci.*, (2001), **21**, 9255-9264.
- Steuber, H., Zentgraf, M., Podjarny, A., Heine, A. and Klebe, G., *J. Mol. Biol.*, (2006), **356**, 45-56.
- Sudhof, T. C., *Annu. Rev. Neurosci.*, (2004), **27**, 509-547.
- Sutton, R. B., Davletov, B. A., Berghuis, A. M., Sudhof, T. C. and Sprang, S. R., *Cell*, (1995), **80**, 929-938.
- Sutton, R. B., Ernst, J. A. and Brunger, A. T., *J. Cell Biol.*, (1999), **147**, 589-598.
- Sutton, R. B. and Sprang, S. R., *Structure*, (1998), **6**, 1395-1405.
- Terwilliger, T., *Acta Cryst. D*, (2000), **D56**, 965-972.
- Terwilliger, T., *Acta Cryst. D*, (2002), **D58**, 1937-1940.

- Terwilliger, T., *Acta Cryst. D*, (2004), **D60**, 2144-2149.
- Tronrud, D., *Acta Cryst. A*, (1992), **A48**, 912-916.
- Trueblood, K. N., Burgi, H.-B., Burzlaff, H., Dunitz, J. D., Gramaccioni, C. M., Schulz, H. H., Shmueli, U. and Abrahams, S. C., *Acta Cryst. A*, (1996), **A52**, 770-781.
- Trueblood, K. N. and Dunitz, J. D., *Acta Cryst. B*, (1983), **B39**, 120-133.
- Tsuboi, T. and Fukuda, M., *J. Biol. Chem.*, (2005), **280**, 39253-39259.
- Ubach, J., Garcia, J., Nittler, M. P., Sudhof, T. C. and Rizo, J., *Nat. Cell Biol.*, (1999), **1**, 106-112.
- Ubach, J., Zhang, X., Shao, X., Sudhof, T. C. and Rizo, J., *EMBO J.*, (1998), **17**, 3921-3930.
- Urzhumtsev, A., Tete-Favier, F., Mitschler, A., Barbanton, J., Barth, P., Urzhumtseva, L., Biellmann, J.-F., Podjarny, A. and Moras, D., *Structure*, (1997), **5**, 601-612.
- Vellieux, F. M. D. and Read, R. J., *Methods Enzymol.*, (1997), **277**, 18-53.
- Verdaguer, N., Corbalan-Garcia, S., Ochoa, W. F., Fita, I. and Gomez-Fernandez, J. C., *EMBO J.*, (1999), **18**, 6329-6338.
- von Mollard, G. F., Stahl, B., Walch-Solimena, C., Takei, K., Daniels, L., Khoklatchev, A., De Camilli, P., Sudhof, T. C. and Jahn, R., *Eur. J. Cell Biol.*, (1994), **65**, 319-326.
- Wang, B. C., *Methods Enzymol.*, (1995), **115**, 90-112.
- Wang, Y., Okamoto, M., Schmitz, F., Hofman, K. and Sudhof, T. C., *Nature*, (1997), **388**, 593-598.
- Wang, Y. and Sudhof, T. C., *Genomics*, (2003), **81**, 126-137.
- Wang, Y., Sugita, S. and Sudhof, T. C., *J. Biol. Chem.*, (2000), **275**, 20033-20044.
- Washbourne, P., Schiavo, G. and Montecucco, C., *Biochem. J.*, (1995), **305**, 721-724.
- Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J. L., *Proc. Natl. Acad. Sci. USA*, (2000), **97**, 623-628.
- Willis, B. T. M. and Pryor, A. W., *Thermal vibrations in crysalligraphy*, (1975), Cambridge University Press.

-
- Wilson, D. K., Bohren, K. M., Gabbay, K. H. and Quioco, F. A., *Science*, (1992), **257**, 81-84.
- Wilson, D. K., Tarle, I., Petrash, J. M. and Quioco, F. A., *Proc. Natl. Acad. Sci. USA*, (1993), **90**, 9847-9851.
- Wilson, K. S., Butterworth, S., Dauter, Z., Lamzin, V. S., Walsh, M., Wodak, S., Pontius, S., Richelle, J., Vaguine, A., Sander, C., Hooft, R. W. W., Vriend, G., Thornton, J. M., Laskowski, R. A., MacArthur, M. W., Dodson, E. J., Murshudov, G. N., Oldfield, T. J., Kaptein, T. and Rullman, J. A. C., *J. Mol. Biol.*, (1998), **276**, 417-436.
- Winn, M. D., Isupov, M. N. and Murshudov, G. N., *Acta Cryst. D*, (2001), **D57**, 122-133.
- Yabe-Nishimura, C., *Pharmacol. Rev.*, (1998), **50**, 21-33.
- Yamaguchi, T., Shirataki, H., Kishida, S., Miyazaki, M., Nishikawa, J., Wada, K., Numata, S., Kaibuchi, K. and Takai, Y., *J. Biol. Chem.*, (1993), **268**, 27164-27170.
- Zang, K. Y. J., Cotwan, K. and Main, P., *International Table for Crystallography Volume F: Crystallography for biological macromolecules*, (2001), M. G. Rossmann and E. Arnold, Dordrecht, The Netherlands, Kluwer Academic Publisher, 311-324.

PUBLICATIONS

- [1] **“Crystallization and preliminary X-ray data of amine oxidase from bovine serum”**, V. Calderone, M.L. Di Paolo, M. Trabucco, M. Biadene, R. Battistutta, A. Rigo and G. Zanotti, *Acta D* (2003), **D59**, 727 - 729
- [2] **“Tetranuclear homo- and heteroalum-oxanes containing reactive functional groups: syntheses and X-ray crystal structures of $[[\text{Lal}(\text{Me})](\mu\text{-O})(\text{MH}_2)_2]$ ”**, S. Singh, S.S. Kumar, V. Chandrasekhar, H.-J. Ahn, M. Biadene, H.W. Roesky, N.S. Hosmane, M. Noltemeyer, H.G. Schmidt, *Angew. Chem. Int. Ed.* (2004), **43**, 4940-4943
- [3] **“Crystal structure of amine oxidase from bovine serum”**, M. Lunelli, M.L. Di Paolo, M. Biadene, V. Calderone, R. Battistutta, M. Scarpa, A. Rigo, G. Zanotti, *Journal Molecular Biology*, (2005), **346**, 991-1004
- [4] **"Syntheses, characterization, and X-Ray Crystal structure of β -Diketimate Group 13 Hydrides, Chlorides, and Fluorides"**, S. Singh, H.-J. Ahn, A. Stasch, V. Jancik, H. W. Roesky, A. Pal, M. Biadene, R. Herbst-Irmer, M. Noltemeyer, H.-G. Schmidt, *Inorganic Chemistry*, (2006), **4**, 1853-1860
- [5] **“Structure elucidation of the DPI-related chaperone Wind with help of mutants”**, M. Sevvana, M. Biadene, Q. Ma, C. Guo, H.-D. Söling, G.M. Sheldrick, D.M. Ferrari, *Acta D* (2006), **D62**, 589-594
- [6] **"Crystal structure of the C2A domain of Rabphilin3A"**, M. Biadene, P. Montaville, G. M. Sheldrick, S. Becker, *Acta D* (2006), **D62**, 793-799
- [7] **"Atomic resolution structure of human Aldose Reductase reveals that breakage of a single hydrogen bond is sufficient to allow the opening of the safety-belt"**, M. Biadene, A. Podjarny, T. R. Schneider *et al.*, in preparation

LEBENS LAUF

Persönliche Daten

Name	Marianna Biadene
Geburtsdatum	26. März 1977
Geburtsort	Perugia, Italien
Staatsangehörigkeit	niederländisch und italienisch
Familienstand	ledig

Studium und schulische Ausbildung

Juli 1996	Italienisches Abitur, Liceo Scientifico "G. Marinelli", Udine, Italien
Sept. 1997	Diplomstudiengang Chemie, Universität von Padova, Italien
Nov. 2002	Diplomarbeit in der Abteilung für Strukturchemie der Universität Padova, Italien, im Arbeitskreis von Prof. Giuseppe Zanotti: "Ammino ossidasi da siero bovino: studi strutturali mediante diffrazione ai Raggi X". (Rinderserum Aminooxidasen: Strukturstudien mit Röntgenstrahlen)

Promotion

Okt. 2003- Mai 2006	Dissertation am Lehrstuhl für Strukturchemie der Georg-August-Universität Göttingen im Arbeitskreis von Prof. G.M. Sheldrick: "Crystal structures of Aldose Reductase, C2A domain of Rabphilin3A, and tests of new restraints"
---------------------	---

ACKNOWLEDGEMENTS

First of all I would like to thank Prof. G. M. Sheldrick for giving me the opportunity to work in a very stimulant but at the same time relaxing environment. I am also grateful to him to have sent me to various interesting workshops and meetings all around Europe. And last but not least I cannot forget the hundred kilos of cookies and cakes I have eaten during the coffee breaks.

Thanks to Dr. T. Schneider who has driven me through the tortuous street of high resolution refinement, for the always helpful discussions. Mille grazie also to the whole group at the IFOM (Barbara, Fabio and Roberto) for the nice weeks in Milano.

I am in debt with Dr. S. Becker for introducing me to the molecular biology world, for the more than satisfactory collaboration and for the numerous and nice conversations not only about crystallography.

For the corrections of the manuscript I am grateful to Ina Dix, Julia Wittmann, Paola Gringiani and especially Stefan Becker and Tim Grüne.

Thanks to all former and present members of the Sheldrick Abteilung for the continued help and for the friendly atmosphere and a special thank to my office colleagues Jose', Burkhard, Roland and Madhoo.

My gratitude and my huge thank to people who have made my life in Göttingen more than just nice and pleasant: Tine, Jose', Heike and Eugenio.

There will be never enough words to thank my Italian friends who were and are always at my side: Claudia, Marcello and Maria. And grazie to the patavin group for the funny and relaxing days we had every time I was back: Teo, Marianna, Beppe, Carlo e le Elene (Brasola, Sperotto, Papinutto).

I am especially grateful to my parents because without their continued support I would have never made it!!!!