

**The crystal structure of a bacterial lysozyme at
atomic resolution**

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Chemisch-Geowissenschaftlichen Fakultät
der Friedrich-Schiller-Universität Jena

von Diplom-Chemikerin Astrid Rau
geboren am 01.06.1974 in Saalfeld

Gutachter:

1. Prof. Dr. R. Hilgenfeld

2. Prof. Dr. D. Klemm

Tag der öffentlichen Verteidigung: 01.06.2005

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1 Milestones in lysozyme research	1
1.2 Definition and classification of lysozymes	5
1.3 Catalytic mechanisms of lysozymes	7
1.4 <i>Chalaropsis</i> -type lysozymes	9
1.5 Cellosyl – a Ch-type lysozyme from <i>Streptomyces coelicolor</i>	12
1.7 Aim of the project	13
2. MATERIALS AND METHODS	14
2.1 Materials	14
2.1.1 Proteins	14
2.1.2 Carbohydrates	14
2.1.3 Chemicals	15
2.1.4 Crystallisation screens	15
2.1.5 Dialysing tools, assays, crystallisation materials and cryo-tools	15
2.1.6 Laboratory equipment and synchrotron facilities	16
2.2 Methods	17
2.2.1 Determination of protein purity	17
2.2.2 Determination of protein concentration	18
2.2.3 Dialysis	18
2.2.4 Sample concentration	18
2.2.5 Crystallisation	18
2.2.6 Heavy atom and polysaccharide soaks	19
2.2.7 Cryocooling	20
2.2.8 Data acquisition and processing	20
2.2.8.1 <i>Native data collection on the monoclinic crystal form</i>	21
2.2.8.2 <i>Native data collection on the hexagonal crystal form</i>	22
2.2.8.3 <i>MAD data collection</i>	22
2.2.8.4 <i>Data collection on heavy-atom derivatised crystals</i>	24
2.2.8.5 <i>Collection and processing of atomic-resolution data</i>	25

2.2.9	Phase determination	27
2.2.9.1	<i>Molecular replacement</i>	28
2.2.9.2	<i>Multiple wavelength anomalous dispersion</i>	29
2.2.9.3	<i>Multiple isomorphous replacement with anomalous scattering</i>	29
2.2.10	Model building and electron-density maps	31
2.2.11	Structure refinement	33
2.2.12	Validation of model quality	35
3.	RESULTS	36
3.1	Preparation of Cellosyl	36
3.2	Crystallisation	36
3.2.1	Hexagonal crystals	36
3.2.2	Monoclinic crystals	37
3.3	Structure elucidation of the monoclinic crystal form	38
3.3.1	Native data collection	38
3.3.2	Initial attempts to solve the structure by molecular replacement	40
3.3.3	Preparation of heavy-atom derivatives	43
3.3.4	Initial attempts to solve the structure by MAD	44
3.3.4.1	<i>Data collection</i>	44
3.3.4.2	<i>Location of the heavy-atom sites and MAD phasing</i>	45
3.3.5	Structure elucidation by MIRAS	47
3.3.5.1	<i>Data collection</i>	47
3.3.5.2	<i>Location of the heavy atoms in the unit cell</i>	48
3.3.5.3	<i>Phase determination</i>	49
3.3.6	Model building and refinement	51
3.3.7	Assessment of model quality	52
3.4	Atomic-resolution structure	55
3.4.1	Data acquisition and processing	55
3.4.2	Refinement of the high-resolution structure	57
3.4.3	Validation of the model quality	63

3.5	Structure elucidation of the hexagonal crystal form	67
3.5.1	Data collection and space group determination	67
3.5.2	Structure solution by molecular replacement	69
3.5.3	Refinement	70
3.5.4	Validation of the model quality	72
4.	DISCUSSION	74
4.1	Overall structure of the monoclinic crystal form at 1.65 Å	74
4.2	Overall structure of the hexagonal crystal form at 2.32 Å	76
4.3	Atomic-resolution structure of Cellosyl at 0.83 Å resolution	77
4.3.1	Advantages and biological relevance of macromolecular crystal structures at atomic resolution	77
4.3.2	The overall structure	78
4.3.3	Alternate conformations and disordered residues	79
4.3.4	Solvent region	81
4.3.5	Anisotropic displacement parameters	83
4.4	Active site and mechanism	85
4.5	Relationship to other lysozyme structures	90
4.6	Relationship to other β/α barrels enzymes	92
4.7	Evolution of $(\beta/\alpha)_8$ barrel structures	94
5.	SUMMARY	97
6.	ZUSAMMENFASSUNG	99
7.	REFERENCES	102
8.	APPENDIX	113

FIGURE INDEX

Figure		Page
1.1-1	Sir Alexander Flemming	2
1.1-2	Solid model of the lysozyme electron density - from the original publication by Blake <i>et al.</i>	3
1.1-3	Ribbon representation of the crystal structure of HEWL	4
1.2-1	(a) Bacterial peptidoglycan subunit of Gram-positive bacteria, (b) Chemical formulas of NAM and NAG	5
1.3-1	Reaction sequence of retaining glycosyl hydrolases	8
1.3-2	Reaction sequence of inverting glycosyl hydrolases	9
1.4-1	Substrate of the Ch-type lysozymes	10
2.2.8.3-1	Theoretical plots of f' and f'' for osmium	23
2.2.8.4-1	Theoretical plots of f' and f'' for osmium and mercury	25
3.1-1	SDS-PAGE analysis of the original Cellosyl samples	36
3.2-1	The two crystal forms of Cellosyl: hexagonal and monoclinic	38
3.3.1-1	Diffraction image of a monoclinic Cellosyl crystal	39
3.3.2-1	Sequence comparison between Cellosyl and <i>S. erythraeus</i> lysozyme	40
3.3.2-2	Ramachandran plot of <i>Streptomyces erythraeus</i> lysozyme	43
3.3.4.1-1	X-ray fluorescence scan of the Os-derivatised crystal	44
3.3.4.2-1	Anomalous and dispersive difference Patterson map	46
3.3.5.2-1	Isomorphous and anomalous difference Patterson map for the osmium mercury double derivative	48
3.3.6-1	Final electron density map with anomalous difference density	51
3.3.7-1	Ramachandran plot of the final structural model of Cellosyl	53
3.3.7-2	Stereo representation of the bend of the loop between α_3 and β_4	54
3.4.1-1	Diffraction image of a Cellosyl crystal.	55
3.4.1-2	Completeness and I/σ in dependence of the resolution	56
3.4.2-1	Met107 at different stages of refinement	59
3.4.2-2	Ile99 during refinement	60
3.4.2-3	Example of a chloride ion before and after anisotropic refinement	61
3.4.2-4	The course of refinement of the high-resolution structure of Cellosyl	62
3.4.3-1	Final electron density map for residue Thr132	64
3.4.3-2	Ramachandran plot of the Cellosyl structure at 0.83 Å resolution	65

3.4.3-3	Luzzati plot of the final model of Cellosyl at 0.83 Å resolution	66
3.5.1-1	The <i>hkl</i> planes of the hexagonal crystal at $l = 0, 1, 2$ and 3	68
3.5.3-1	Section of the final electron density map covering two sulphate ions	71
3.5.4-1	Ramachandran plot of the model derived from the hexagonal crystals	73
4.1-1	Stereo $C\alpha$ -trace of the Cellosyl molecule	75
4.1-2	Stereographic ribbon representation of the overall fold of Cellosyl	75
4.2-1	Backbone of Cellosyl structure derived from hexagonal crystals	77
4.3.3-1	Overall fold of the high-resolution model of Cellosyl with the alternative side-chain conformations depicted in red as ball-and-stick.	79
4.3.3-2	Depiction of residue Ser24	80
4.3.4-1	Stereographic depiction of the completely buried water 4	82
4.3.5-1	Illustration of the anisotropic displacement parameters for a selection of residues containing atoms with an anisotropy < 0.2	83
4.3.5-2	(a) Mean equivalent B of main-chain atoms. (b) Mean anisotropy of main-chain atoms. (c) Mean equivalent B of side-chain atoms. (d) Mean anisotropy of side-chain atoms	84
4.4-1	Charge distribution on the surface of Cellosyl	85
4.4-2	Comparison of the amino acid sequence of the Ch-type lysozymes	87
4.4-3	Depiction of the two pairs of carboxylic residues lining opposite faces of the active site cleft.	88
4.5-1	Comparison of the three-dimensional structures of HEWL, GEWL, T4L, and Cellosyl.	91
4.6-1	Superposition of Cellosyl and CiX1	92
4.6-2	Topology diagrams of a conventional TIM barrel, Cellosyl and enolase	93

TABLE INDEX

Table		Page
2.1.1-1	Proteins and their manufactures	14
2.1.2-1	Carbohydrates and their manufactures	14
2.1.3-1	Chemicals and their manufactures	15
2.1.4-1	Crystallisation screens	15
2.1.5-1	Equipment and manufactures	15
2.1.6-1	Equipment and manufacturers	16
2.2.8.5-1	Data collection parameters for the three passes on a single Cellosyl crystal	26
3.3.1-1	Scaling statistics of the monoclinic data set	39
3.3.2-1	Results of the rotation search	41
3.3.2-2	Rotation and translation function	41
3.3.2-3	Solution obtained with the program <i>EPMR</i>	42
3.3.4.1-1	Selected wavelengths for the MAD experiment	44
3.3.4.1-2	Scaling statistics of the three passes on an Os-derivatised crystal	45
3.3.5.1-1	Scaling and merging statistics of the heavy atom derivatives	47
3.3.5.2-1	Initial heavy atom sites	49
3.3.5.3-1	Phasing statistics for the three heavy atom derivatives	50
3.3.6-1	Refinement statistics of the Cellosyl structure at 1.65 Å resolution	52
3.4.1-3	Scaling and merging statistics	56
3.4.2-1	Refinement statistics of the high-resolution structure	63
3.5.1-1	Scaling statistics of space groups $P6_1 / P6_5$ and $P6_122 / P6_522$	67
3.5.2-1	Matthews coefficient and solvent content of the hexagonal crystal	69
3.5.2-2	Correlation coefficients and <i>R</i> -factors	70
3.5.3-1	Refinement statistics of the Cellosyl structure in the hexagonal crystal form	72
8.-1	Hydrogen-bonding interactions of the chloride ions	113
8.-2	Hydrogen-bonding interactions of the sulphate ions	114

ABBREVIATIONS

ADP	anisotropic displacement parameter
<i>B</i> -value	isotropic temperature factor
χ^2	goodness of fit
cc	correlation coefficient
CelC	<i>Clostridium thermocellum</i> endoglucanase
Cellosyl	<i>Streptomyces coelicolor</i> lysozyme
Ch	<i>Chalaropsis</i>
CiX1	<i>Coccidioides immitis</i> chitinase-1
Cpl-1	pneumococcal phage lysin
f'	anomalous scattering contribution (real part)
f''	anomalous scattering contribution (imaginary part)
F	structure factor amplitude
F_c	calculated structure factor amplitude
F_o	observed structure factor amplitude
FOM	figure of merit
GEWL	goose egg-white lysozyme
GH	glycosyl hydrolase
HEWL	hen egg-white lysozyme
HEPES	N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid
I	reflection intensity
LBHB	low-barrier hydrogen bond
LOI	lack of isomorphism
MAD	multiple wavelength anomalous dispersion
MIRAS	multiple isomorphous replacement with anomalous scattering
MLE	muconate lactonizing enzyme

mM	millimolar
MR	mandelate racemase
NAG	N-acetyl glycosamine
NAM	N-acetylmuramic acid
NCS	noncrystallographic symmetry
PDB	Protein Data Bank
PEG	Polyethylene glycol
pv	packing value
r.m.s.d.	root-mean-square deviation
$R_{p.i.m.}$	precision indicating <i>R</i> -factor
$R_{r.i.m.}$	redundancy independent <i>R</i> -factor
σ	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide-gel electrophoresis
SIR	Single isomorphous replacement
T4L	bacteriophage T4 lysozyme
TIM	triosephosphate isomerase
TLS	tensors describing translation, libration and cross-correlation in the refinement of a rigid body
Tris	Tris(hydroxymethyl)-aminomethane

1. INTRODUCTION

1.1 Milestones in lysozyme research

Bacteriology textbooks of the late 19th century describe the cultivation of bacteria in the interior of hen eggs. Laschtschenko, a professor of hygiene at the university of Tomsk, questioned the feasibility of eggs for culturing microorganisms and discovered the bactericidal power of hen egg white as early as 1909 (Laschtschenko, 1909). He noticed that the content of a hen egg emptied on an uncovered glass plate at room temperature dries up without putrefaction and development of malodorous gases and concluded that germs contained in air dust find no conditions to propagate in hen egg white. In further experiments, he demonstrated the inhibitory action of hen egg white on *Bacillus subtilis*. Even high concentrations of *Bacillus subtilis* were killed by small amounts of hen egg white. The bactericidal power of hen egg white was also demonstrated for *Bacillus anthrax*, *Proteus zopfii*, *Proteus zencker* and *Bacillus megaterium*. Laschtschenko experimentally excluded the death of bacteria as result of a plasmolytic process or insufficient nutritional conditions and concluded that hen egg white contains enzymes of proteolytic character (Laschtschenko, 1909). Following him, Rettger and Sperry noticed the antibacterial action of hen egg white, but also failed to identify the source of its bactericidal power (Rettger and Sperry, 1912).

Since 1900, several scientists have described the antibacterial action of saliva and body secretions, among them Bloomfield, who investigated the fate of bacteria introduced into the upper air passages of man (Bloomfield, 1919). The aim of his work was to obtain more information about contact infections and details of the spread of bacteria in the upper air passages. Bloomfield stated: "It seemed advisable to make a preliminary study of the fate of non-pathogenic organisms before employing possible disease producers." (Bloomfield, 1919). For this reason *Sarcina lutea* was chosen, since besides its non-pathogenicity, it is normally not present in the upper air passages of men and could be readily grown on simple media. Individuals with no disease or abnormalities in the upper air passages were selected and solid masses of a 24-hour growth of *Sarcina lutea* were swabbed on their tongue and nasal mucosa. Cultures were made at certain time intervals. *Sarcina lutea* wiped on the tongue disappeared rapidly and was extinct after an hour while *Sarcina lutea* swabbed on the nasal mucosa was nearly extinct after 24 hours. As a mechanism for the disappearance of *Sarcina lutea*,

mechanical cleansing and mouth bacteria were considered but the main effect was attributed to the bactericidal effect of saliva and mouth secretions.

Alexander Fleming made a similar discovery in 1921. Scientific lore tells that while he was suffering from a cold, a drop from his nose fell onto an agar plate where cultures of bacteria had grown. Around the drop, the bacterial cultures began to dissolve and Fleming was the first one to conclude that nasal secretions contain a lytic substance. Fleming called



Figure 1.1-1 Sir Alexander Fleming ¹

this substance “Lysozyme”. He confirmed his observations with several simple but very resourceful experiments: A drop of nasal mucus diluted with salt solution was placed on an agar plate thickly planted with *Micrococcus lysodeikticus*. After incubation for 24 hours, the cocci had copiously grown except in the regions covered with nasal mucus, where complete inhibition of growth was observed. In a further experiment, a drop of diluted nasal secretion was added to 1 cm³ of a thick suspension of *M. lysodeikticus* and within a few minutes, the cocci completely disappeared (Fleming, 1922). In his first paper on lysozyme, Fleming describes that the enzyme is present in human body

secretions such as tears, nasal mucus, saliva and sputum, and in body tissues, especially in cartilage. Lysozyme was furthermore found in animal and vegetable tissues and in a very large concentration in egg white (Fleming, 1922). Lysozyme of different tissues and secretions showed varied bactericidal action towards different microorganisms, therefore the existence of different lysozymes was concluded (Fleming, 1932). Fleming demonstrated that lysozyme is active towards a great variety of bacteria non-pathogenic to man and towards some pathogenic bacteria. However, he had to realise that lysozyme is inactive against numerous very pathogenic bacteria such as *Micrococcus melitensis* and *Bacillus pestis* (nowadays named *Brucella melitensis* and *Yersinia pestis*, respectively) (Fleming, 1929).

Besides his outstanding work in lysozyme research, Sir Alexander Fleming's major achievement was the discovery of penicillin in 1928, for which he was awarded the Nobel Prize for Physiology or Medicine together with Ernst Boris Chain and Howard Walter Florey in 1945.

¹ Source of picture: <http://www.estadao.com.br/divirtaseonline/fotos/retrospectiva/1928.htm>

Since its discovery, lysozyme served as model system for many studies, with hen egg-white lysozyme (HEWL) attracting a great deal of attention. In 1963, the primary structure of HEWL was published by two different research groups (Canfield, 1963; Jollès *et al.*, 1963) and it was the first enzyme reported that contained all usual 20 amino acids.

Only 2 years later, HEWL was the first enzyme to have its three-dimensional structure determined (Blake *et al.*, 1965). The structure was solved by the method of multiple isomorphous replacement exploiting the anomalous scattering effect of mercury and uranium derivatives using copper K_{α} -radiation. The authors reported that approximately 16 crystals had been used to obtain a complete data set for each derivative. Phases were determined according to the phase probability method (Blow and Crick, 1959) and a model of the electron-density distribution at 6 Å resolution is depicted in Fig. 1.1-2 (Blake *et al.*, 1965).

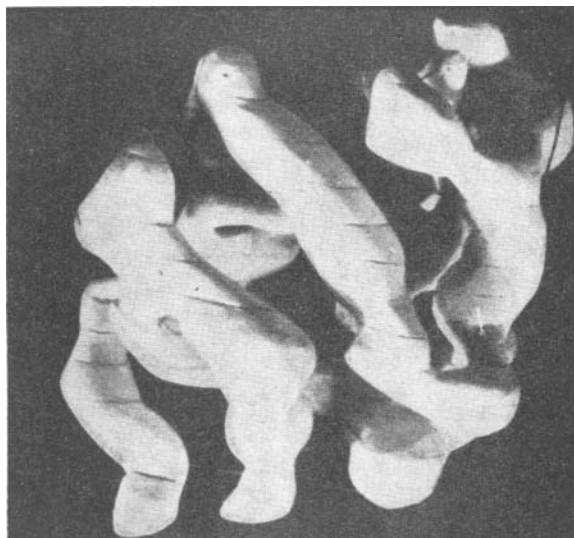


Figure 1.1-2 Solid model of the lysozyme electron density greater than about 0.5 electrons/Å³ at 6 Å resolution - from the original publication by Blake *et al.* (1965).

The model allowed the authors to follow the course of the folded polypeptide chain roughly, but for a detailed structure determination, a Fourier map of electron-density distribution at 2 Å was calculated and plotted in 60 sections. All 129 amino acid residues could be located and several side chains, including the 4 disulphide bridges of HEWL, could be unambiguously determined (Blake *et al.*, 1965).

The molecule is comprised of two domains, with one domain dominantly α -helical and the second showing a particularly complex folding pattern, in which the protein chain was described to be arranged in three sections orientated roughly antiparallel. This was the first description of a β -sheet in a globular protein.

Fig. 1.1-3 shows a ribbon representation of the crystal structure of HEWL (PDB code: 3LZT, Walsh *et al.*, 1998), clearly showing the two domains already described by Blake *et al.* (1965).

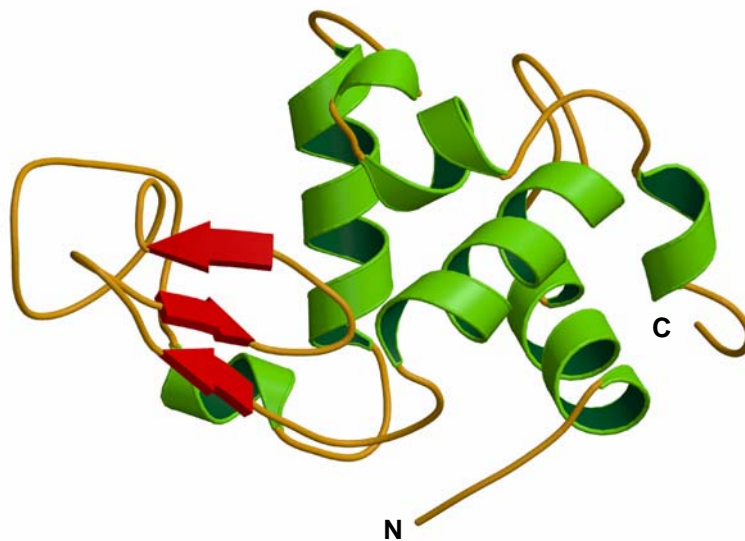


Figure 1.1-3 Ribbon representation of the crystal structure of HEWL (PDB code: 3LZT, Walsh *et al.*, 1998), generated with the programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt and Bacon, 1997).

Furthermore, HEWL was the first enzyme for which a detailed mechanism of action was proposed based on model building studies (Phillips, 1966). The Phillips mechanism was widely held as the paradigm for the cleavage of lysozyme substrates under retention of configuration. It was only corrected in 2001 after extensive studies of a covalent substrate-enzyme intermediate applying electrospray ionisation mass spectrometry and high-resolution X-ray crystallography (Vocadlo *et al.*, 2001).

Nearly 100 years of research have made lysozymes one of the best-studied classes of enzymes. As evident from the vast body of literature, they remain the subject of extensive ongoing studies, which is in accordance with Fleming's prophecy: "We shall hear more about lysozyme" (Jollès, 1996).

1.2 Definition and classification of lysozymes

Bacteria are enclosed by rigid cell walls, which offer mechanical protection and prevent them from bursting due to their high inner osmotic pressure. The bactericidal action of lysozyme is attributed to its ability to cleave a particular polysaccharide bond (Fig. 1.2-1 and Fig. 1.4-1) in the cell wall of Gram-positive bacteria (Salton, 1952). Gram-negative bacteria are not susceptible to enzymatic breakdown by lysozyme since their peptidoglycan layer is protected by an outer membrane (Ghuysen, 1968). A schematic representation of a peptidoglycan subunit constituting the cell wall of Gram-positive bacteria is depicted in Fig. 1.2-1a. The peptidoglycan forms a bag-shaped sacculus, which is composed of three components: polysaccharide chains, oligopeptide subunits and peptide cross-linking bridges.

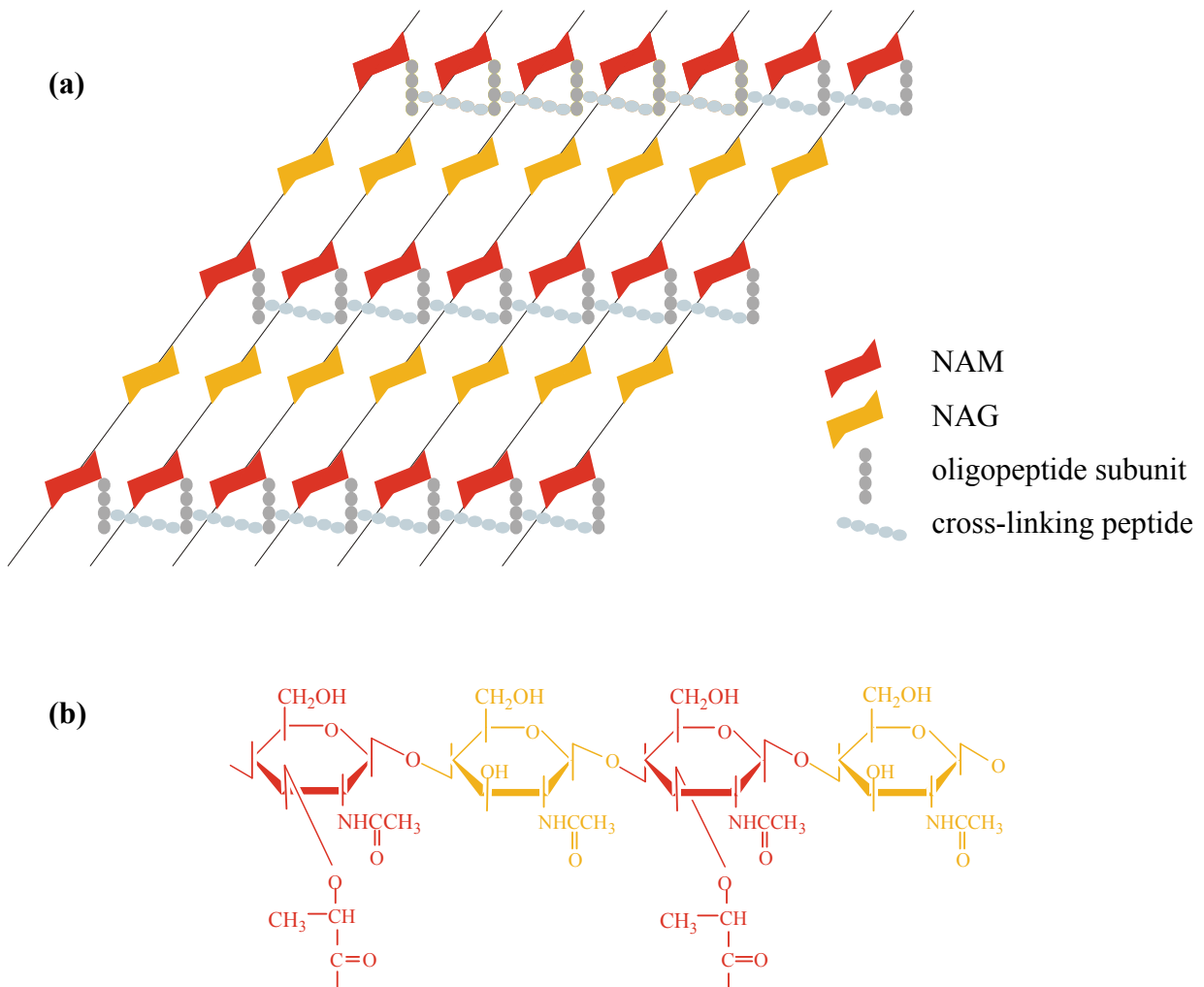


Figure 1.2-1 (a) Schematic representation of the bacterial peptidoglycan subunit of Gram-positive bacteria, (b) Chemical formulas of NAM (red) and NAG (yellow) connected via a β -1,4-glycosidic bond.

The linear polysaccharide chains consist of two alternating sugar moieties, N-acetylmuramic acid (NAM) and N-acetylglucosamine (NAG), which are connected via a β -1,4-glycosidic bond between the C-1 atom of NAM and the C-4 atom of NAG (Fig 1.2-1b). All *O*-glycosidic bonds in the bacterial cell wall adopt the β -configuration, implying that the glycosidic oxygen is located above the plane of the sugar rings. The lactic acid component of NAM forms an amide bond to the oligopeptide subunit, usually a tetrapeptide that contains both L- and D-amino acids. These oligopeptide subunits are covalently cross-linked by peptide bridges to form the network of the peptidoglycan sacculus.

Lysozymes are glycosidases, which hydrolyse the glycosidic bond between C-1 of NAM and C-4 of NAG of the bacterial peptidoglycan. The bond between C-1 of NAG and C-4 of NAM is not attacked. Lysozymes are broadly expressed throughout all organisms, ranging from bacteria and bacteriophages over fungi and plants to vertebrates.

As determined from their amino acid composition, catalytic activities and immunological cross-reactivities (Jollès and Jollès, 1984), lysozyme structures fall into four different classes of endo-*N*-acetylmuramidases. The prototypes of these classes are HEWL, goose egg-white lysozyme (GEWL), bacteriophage T4 lysozyme (T4L), and *Chalaropsis* (Ch) lysozyme.

A classification of glycosyl hydrolases (GH) into families based on amino acid similarities was established by Henrissat (Henrissat, 1991; Henrissat and Bairoch, 1993, 1996) and is available on the CAZy web server (Henrissat, 1998). The latest update (Nov., 2003) contained 91 families. This classification was designed to allow reflections on structural features, enzymatic mechanism and the evolutionary relationship of these enzymes. In agreement with the classification of Jollès and Jollès (1984), HEWL-, GEWL-, T4- and Ch-type lysozymes belong to 4 separate glycosyl hydrolase families, family 22, 23, 24 and 25, respectively.

Occasionally a fifth class of plant lysozymes is mentioned, with hevamine, a chitinase from the latex rubber tree *Hevea brasiliensis*, as the archetype. However, the alleged muramidase activity of these enzymes is restricted to a very narrow pH range (pH 4.5 – 5.0), while they display chitinase activity over a much broader pH-range (Beintema and Terwisscha van Scheltinga, 1996). Recent studies on the cleavage specificity of hevamine for peptidoglycan revealed that hevamine cleaves the bond between the C-1 of NAG and C-4 of

NAM. Therefore, the enzyme exhibits (beside its chitinase activity) β -*N*-acetylglucosamidase activity and cannot be classified as lysozyme (Bokma *et al.*, 1997).

1.3 Catalytic mechanisms of lysozymes

Hydrolysis of the β -1,4-glycosidic bond between NAM and NAG can proceed via two major mechanisms resulting in either an overall retention, or an inversion of the anomeric configuration (Rye and Withers, 2000; Zechel and Withers, 2001; Vasella *et al.*, 2002).

As already mentioned, HEWL served as model enzyme for studies of the catalytic mechanism of glycosyl hydrolases operating under retention of configuration. Based on structural data, Phillips (1966) proposed the following mechanism for HEWL. The enzyme possesses six binding sites, labelled A to F for binding the distinct glucose units of the substrate. The scissile bond is located between sites D and E and binding of the substrate in subsite D is only possible if the glycoside is distorted into a half chair or skew boat conformation. Due to this conformational change the anomeric carbon (C1) is in plane with C2, C4, C5 and the pyranosidic oxygen (Stryer, 1996). A proton from Glu35 attacks the glycosidic oxygen, leading to the cleavage of the C1-O bond and formation of an oxocarbenium-ion intermediate. This inherently unstable intermediate is stabilised by electrostatic interactions with the deprotonated Asp52 as well as by partial charge delocalisation to the oxygen in the pyranose ring. Due to steric hindrance, the nucleophilic attack by a water molecule can only proceed from the β side of the anomeric centre, which results in a net retention of the anomeric conformation completing this S_N1 -like reaction sequence. The instability of the proposed oxocarbenium-ion intermediate (Amyes and Jencks, 1989) together with subsequently investigated mechanisms of other retaining glycosyl hydrolases (Davies *et al.*, 1998) fuelled a debate whether the reaction really follows this route. Only recently it could be demonstrated that the catalysis by HEWL, in contrast to the mechanism proposed by Phillips, proceeds via a covalent intermediate (Vocadlo *et al.*, 2001). Electrospray Ionisation Mass spectroscopy in conjunction with X-ray crystallographic structure determination of a trapped glycosyl-enzyme intermediate of an E35Q mutant lysozyme demonstrated unequivocally that HEWL follows the common mechanism of retaining β -glycosidases.

Figure 1.3-1 depicts the reaction sequence established for glycosyl hydrolases operating under retention of configuration. In general, the catalytic machinery of these

enzymes involves two carboxylate side chains. The glycosidic oxygen of the substrate is protonated by a carboxylic acid, i.e. Glu35 in HEWL, resulting in the cleavage of the scissile bond. The so formed oxocarbenium-ion intermediate immediately forms a covalent ester bond with the second carboxylate side chain, i.e. Asp52 in HEWL (see first transition state in Fig. 1.3-1). The ester is then hydrolysed by a water molecule attacking from the β -side of the anomeric carbon completing the second S_N2 reaction (see second transition state in Fig. 1.3-1). This double Walden-inversion at C1 gives rise to an overall net retention of the anomeric configuration.

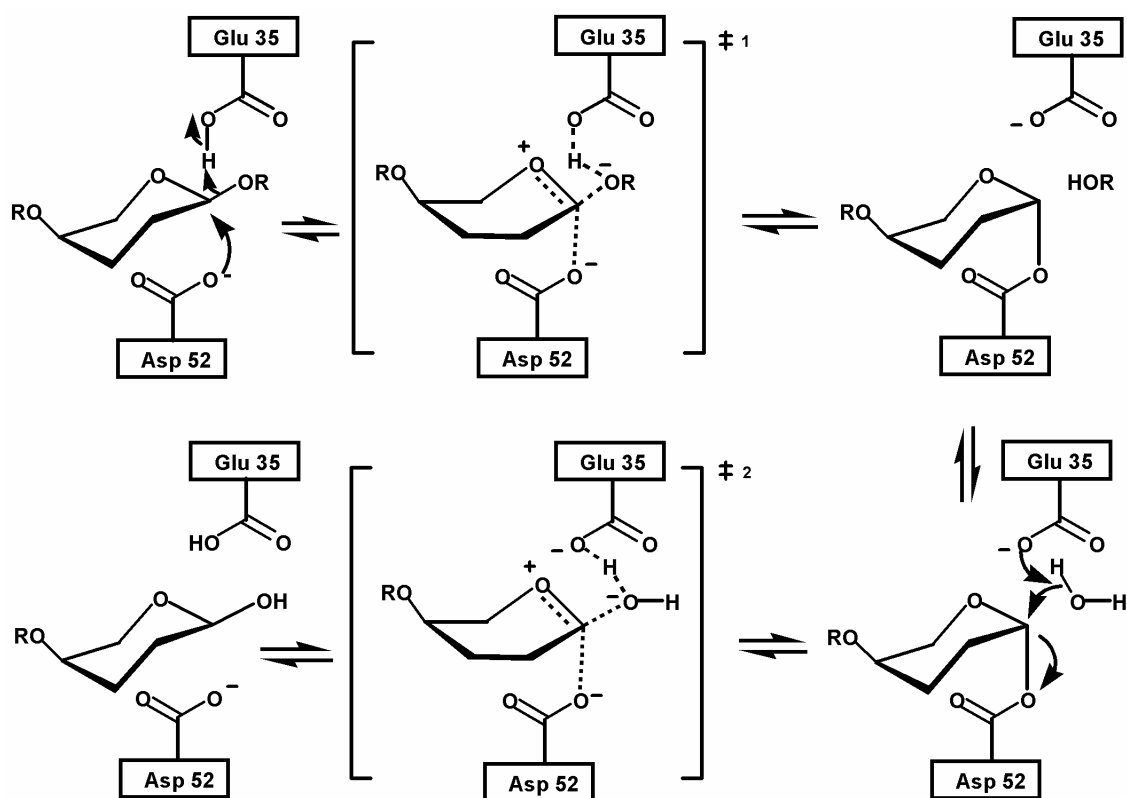


Figure 1.3-1 Reaction sequence of retaining glycosyl hydrolases, e.g. HEWL.

Lysozymes from different organisms not only differ in their overall structure but also in their reaction mechanisms. While HEWL is a retaining enzyme, GEWL and T4L hydrolyse the substrate under inversion of configuration (Kuroki *et al.*, 1993, 1999). Substrate binding and activation by inverting enzymes is very similar to that of retaining enzymes (see Fig. 1.3-

2). However, the reaction mechanism is very different and, most importantly, does not involve a covalent intermediate. The glycosidic bond is cleaved due to protonation by a carboxylic acid (Glu73 in GEWL, Glu11 in T4L). A water molecule, held in the spatial vicinity of the anomeric carbon atom, is activated by hydrogen bonding to a carboxylate group, e.g. Asp20 in T4L, which serves as the catalytic base. The water attacks the anomeric carbon from the α -side in a single displacement mechanism resulting in inversion of configuration (Kuroki *et al.*, 1993).

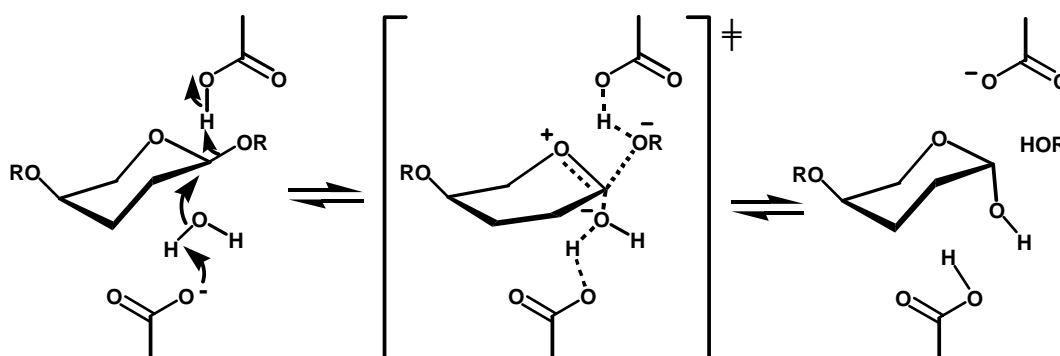


Figure 1.3-2 Reaction sequence of inverting glycosyl hydrolases.

It is important to mention that GEWL lacks the apparent analogue to Asp52 in HEWL or Asp20 in T4L (Weaver *et al.*, 1995). Therefore, it has been proposed that functional groups of the substrate, i.e. carboxylic groups of the peptide component, stabilise the transition state (Matsumura and Kirsch, 1996; Rye and Withers, 2000).

The catalytic mechanism of Ch-type lysozymes is unknown up to now. Based on chemical modification experiments, the catalytic activity of the *Chalaropsis* lysozyme is at least partially attributed to residues Asp6 and Glu33 (Fouche *et al.*, 1978).

1.4 *Chalaropsis*-type lysozymes

The family of Ch-type lysozymes was named after the first member to be described, the muramidase from the fungus *Chalaropsis* (Hash and Rothlauf, 1967; Felch *et al.*, 1975). This fungal enzyme was the first microbial lysozyme to have its amino acid sequence determined. It is comprised of 211 residues with a calculated molecular weight of 22,4 Da

(Felch *et al.*, 1975). Like all lysozymes, the muramidase from *Chalaropsis* hydrolyses the β -1,4-glycosidic bond in cell-wall murein. Differences in substrate specificity between the *Chalaropsis* enzyme and HEWL have been already described in 1967. While HEWL is a weak chitinase, chitin is neither a substrate nor an inhibitor of the *Chalaropsis* lysozyme (Hash and Rothlauf, 1967; Fouche and Hash, 1978). However, the *Chalaropsis* enzyme possesses a feature which makes it superior to HEWL, i.e., it has the ability to cleave *O*-acetylated peptidoglycans. *O*-acetylation of the bacterial peptidoglycan takes place at the C6 hydroxyl group of the *N*-acetylmuramic acid moiety resulting in a 2,6-diacetylmuramic acid derivative (Clarke and Dupont, 1992). Spontaneous *O*-acetylation of the cell wall of a *Micrococcus* was reported when cultured in the presence of HEWL, which rendered it insensitive to HEWL (Brumfitt *et al.*, 1958; Brumfitt, 1959).

In general it can be said that in contrast to the more 'classical' lysozymes of the HEWL, GEWL and T4-types, Ch-type lysozymes display both β -1,4 *N*-acetyl as well as β -1,4 *N*,6-*O*-diacetylmuramidase activity and therefore possess the ability to cleave 6-*O*-acetylated peptidoglycans such as present in the cell wall of *Staphylococcus aureus*, which are not hydrolysed by the other lysozymes (Fig. 1.4-1).

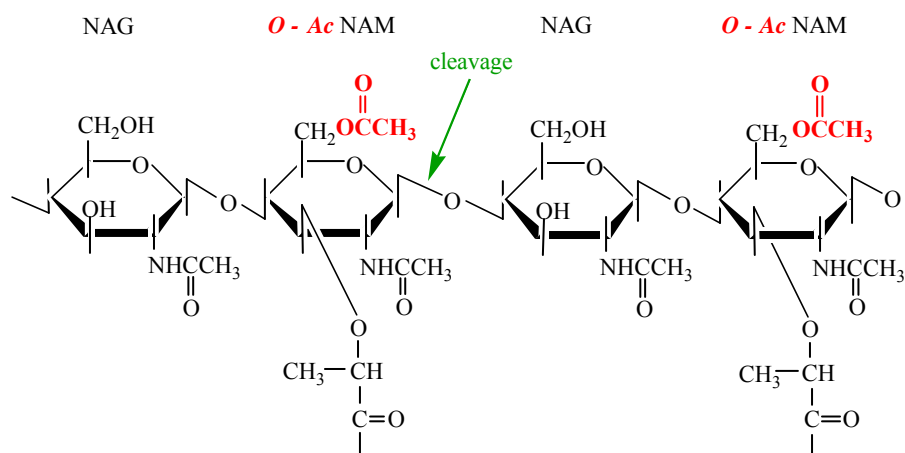


Figure 1.4-1 The substrate of the Ch-type lysozymes is *N*,6-*O*-diacetylated peptidoglycan. The 6-*O*-acetylation, marked in red, renders cell walls insensitive to HEWL, GEWL and T4-type lysozymes.

Although first discovered in a fungus, the majority of the Ch-type lysozymes known today has been found in bacteriophages and in Gram-positive bacteria. Lysozymes expressed by bacteriophages play an important role for the release of virions at the end of an infection

cycle by lysing the host cell walls. Ch-type phage lysozymes have been reported in *Streptococci* and *Lactobacilli* (Ronda *et al.*, 1981; Mata *et al.*, 1986). Also the phages MV1 and Φ adh, which infect the Gram-positive *Lactobacillus gasseri*, encode muramidases with significant sequence similarity of their N-terminal domain to the *Chalaropsis* lysozyme (Fastrez, 1996).

One of the best-studied phage lysozymes is Cpl-1, a 39 kDa muramidase found in phage Cp-1-infected *Streptococcus pneumoniae*. A special feature of this Gram-positive bacterium is the presence of choline in the teichoic acids of the pneumococcal cell wall, a prerequisite for the enzymatic activity of Cpl-1 (Fastrez, 1996). The nucleotide sequence of the *cpl* gene implies that this muramidase is composed of two modules serving different functions. The first module is responsible for the catalytic activity and shows a high similarity to the muramidase from *Chalaropsis*, while the second one is binding to the choline residues in the cell wall. Asp10 and Glu37, which are equivalent to Asp6 and Glu33 of the *Chalaropsis* enzyme, have been identified as potential catalytic residues. A similar modular organisation of genes has been observed for other Cp phages. In addition to a muramidase domain with clear sequence similarity to the lysozyme from *Chalaropsis*, some of these phage lysozymes carry additional domains for binding to the bacterial cell wall in order to enable efficient substrate recognition and lysis (Fastrez, 1996).

The role of bacterial lysozymes is not fully understood. These enzymes are likely involved in the regulation of peptidoglycan synthesis. Due to their ability to cleave specific bonds in the peptidoglycan sacculus, they allow new peptidoglycan subunits to be incorporated into the cell wall, thereby remodelling its shape and surface area. The way in which bacteria control these potentially suicidal enzymes is still a matter of debate. It has been suggested that the regulation of bacterial lysozymes is interlinked with other metabolic pathways of the cell such as stringent response and heat shock control (Tuomanen and Tomasz, 1986; Young *et al.*, 1989).

Ch-type muramidases have been reported in *Streptomyces* (e.g. *S. coelicolor* and *S. globisporus*), *Clostridium acetabutylicum* and *Lactobacillus acidophilus*. Several of these bacteria express more than one lysozyme which target the same bond in the peptidoglycan. For example, a bacteriolytic enzyme extract consisting of two different muramidases (named M1 and M2) has been obtained from *Streptomyces globisporus* (Höltje, 1996). Named “Mutanolysin”, this enzyme mixture is widely used for the lysis and study of bacterial cell walls. Lichenstein *et al.* succeeded in cloning and sequencing of the M1 enzyme (Lichenstein

et al., 1990). Beside an unusually long leader peptide of 77 amino acids, the enzyme comprises 217 amino acid residues and exhibits 61 % sequence identity with the lysozyme from fungus *Chalaropsis*.

Until recently, very little structural information for Ch-type lysozymes was available. Circular dichroism studies of *Chalaropsis* lysozyme in the far UV-region showed that the secondary structure is high in β -sheet content. With a structural composition of 21 % α -helix, 63% β -sheet and 16 % loop regions it is very different from that of HEWL with 28 % α -helix, 10% β -sheet and 62 % loop regions as determined from X-ray crystallography (Chang *et al.*, 1979). The lysozyme functionality of Ch-type enzymes seems to be realised by a completely different protein architecture.

Crystallisation has been reported for a number of Ch-type lysozymes, including the enzymes from *Chalaropsis* (Hash and Rothlauf, 1967; Lyne *et al.*, 1990), *Streptomyces erythraeus* (Morita *et al.*, 1978; Harada *et al.*, 1981), *Streptomyces globisporus* (Harada *et al.*, 1989; Shiba *et al.*, 2000), *Streptomyces coelicolor* (Hilgenfeld *et al.*, 1992) and from phage Cp-1-infecting *Streptococcus pneumoniae* (Monteroso *et al.*, 2002).

The only X-ray crystallographic report available for a Ch-type lysozyme so far describes a low-resolution model for the muramidase from *Streptomyces erythraeus*, which was determined in the absence of a complete amino acid sequence for the enzyme (Harada *et al.*, 1981). This structural model comprises a polypeptide chain of approximately 180 amino acid residues folded into three domains. In the present work it will be shown that it is very likely not correct. (Two years after the results of this work were published (Rau *et al.*, 2001) the structure of the Ch-type phage lysozyme Cpl-1 was reported (Hermoso *et al.*, 2003), showing the same overall fold as Cellosyl.)

1.5 Cellosyl – a Ch-type lysozyme from *Streptomyces coelicolor*

Cellosyl is a very basic (pI > 10), 23 kD-muramidase excreted by *Streptomyces coelicolor* “Müller”, a Gram-positive bacterium found in soil. It shares 61 % sequence identity with the lysozyme from fungus *Chalaropsis*. As is typical for Ch-type lysozymes, Cellosyl possesses both β -1,4-*N*-acetylmuramidase and β -1,4-*N*,6-*O*-diacetylmuramidase, i.e. staphylolytic, activity (Bräu *et al.*, 1991). Since *Staphylococcus aureus* is an important food pathogen, this enzymatic activity makes Cellosyl very interesting for prevention of food spoilage and food-borne disease. The enzyme was therefore tested in food preservation, with

very promising results in cheese manufacturing (prevention of late blowing caused by anaerobic spore formers such as *Clostridium tyrobutyricum*). Based on these results, Cellosyl appears to be an attractive replacement for nitrate (a potential source of the formation of carcinogenic nitrosamines) in food preservation (Bräu *et al.*, 1991; Hughey and Johnson, 1987).

A surprising feature is the total identity of Cellosyl to the M1 muramidase of *Streptomyces globisporus* with respect to their amino-acid composition. Bräu *et al.* reported that the *cel* gene of *S. coelicolor* and the *acm* gene of *S. globisporus* show exactly the same sequence in the coding region and differ only slightly in the adjacent 5' and 3'-sequences. The authors concluded that either both strains are closely related or that the *cel* and *acm* genes descended from a bacteriophage which exists as a lyogenised phage or defective prophage in both strains (Bräu *et al.*, 1991).

1.6 Aim of the project

The different classes of lysozyme share no significant similarities in their amino-acid composition. However, the three-dimensional structures of HEWL, GEWL and T4-type lysozymes show some intriguing, albeit distant, similarities and it is believed that they are a result of divergent evolution from a common ancestor (Grütter *et al.*, 1983; Weaver *et al.*, 1985). In contrast, no detailed structural information has been available so far for the Ch-type lysozymes.

The aim of this project was to elucidate the three-dimensional structure of Cellosyl by X-ray crystallography since it is an attractive candidate for comparative structural studies of the lysozyme family. The hypothesis that the Ch-type muramidases are structurally unrelated to the HEWL, GEWL and T4-type lysozymes and therefore form an independent class without evolutionary relationship to the other lysozyme structures will be investigated. Furthermore, it is hoped to obtain new insights into the architecture of the enzyme's catalytic site and that the knowledge of the three-dimensional structure can reveal information on the enzymatic mechanism.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1 Proteins

Table 2.1.1-1 Proteins and their manufactures

Protein	Supplier
Cellosyl (<i>Streptomyces coelicolor</i>)	Hoechst AG (Frankfurt)
Lysozyme (chicken egg-white)	Merck (Darmstadt)
Bovine albumin	Sigma (Steinheim)
γ -Globulin	Bio-Rad (München)
Low molecular weight protein standards: Phosphorylase b (rabbit muscle) Albumin (bovine serum) Ovalbumin (chicken egg) Carbonic anhydrase (bovine erythrocyte) Trypsin inhibitor (soybean) α -Lactalbumin (bovine milk)	Pharmacia Biotech (Freiburg)

2.1.2 Carbohydrates

Table 2.1.2-1 Carbohydrates and their manufactures

Carbohydrate	Manufacturer
Sucrose	Merck (Darmstadt)
N,N',N''-Triacetyl-Chitotriose	CALBIOCHEM® (Schwalbach)
N,N',N'',N''',N''''',N''''''-Hexaacetyl-Chitohexose	CALBIOCHEM® (Schwalbach)
Chitosan Polysaccharide Mixture (Crab Shell)	CALBIOCHEM® (Schwalbach)

2.1.3 Chemicals

Table 2.1.3-1 Chemicals and their manufactures

Chemical substance	Manufacturer
PEG 4000, 6000, 8000	Fluka (Steinheim)
HEPES	Merck (Darmstadt)
TRIS	Merck (Darmstadt)
Ammonium sulphate	Merck (Darmstadt)
Osmium(III)-chloride	Aldrich (Steinheim)
Mercury(II)-chloride	Aldrich (Steinheim)
Yttrium (III)-chloride	Aldrich (Steinheim)
Ytterbium(III)-chloride	Aldrich (Steinheim)
Potassium tetrachloroplatinat	Aldrich (Steinheim)
cis-Platinum	Aldrich (Steinheim)
β -mercaptoethanol	Merck (Darmstadt)
Sodium azide	Merck (Darmstadt)
Highly liquid paraffin oil and all other chemicals not separately stated	Merck (Darmstadt)

2.1.4 Crystallisation screens

Table 2.1.4-1 Crystallisation screens

Screen	Manufacturer
Ammonium sulphate screen	Hampton Research (Laguna Niguel, USA)
Crystal screen 1	Hampton Research (Laguna Niguel, USA)
Crystal screen 2	Hampton Research (Laguna Niguel, USA)

2.1.5 Dialysing tools, assays, crystallisation materials and cryo-tools

Table 2.1.5-1 Equipment and manufactures

Equipment	Manufacturer
Phastgel homogenous 12.5 gel beds and SDS buffer strips	Pharmacia (Freiburg).

Bio-Rad Protein Assay	Bio-Rad (München)
Ultrafree [®] -MC filter units	Millipore (Bedford, USA)
Slide-A-Lyzer dialysis cassettes	Pierce (Rockford, USA).
Dialysis membranes and sample tubes	Roth (Karlsruhe).
High vacuum grease	Dow Corning (Midland, USA)
Magnetic base crystal caps,	Hampton Research (Laguna Niguel, USA)
Mounted cryoloops,	
24-well Linbro plates and VDX plates	
22 mm circular siliconised coverslips	
Crystal storage vials	
Cryo canes	
Magnetic crystal wands	

2.1.6 Laboratory equipment and synchrotron facilities

Table 2.1.6-1 Equipment and Manufacturers

Equipment	Manufacturer
Centrifuge- Heraeus Labofuge 400R	Heraeus Instruments (Hanau)
Centrifuge- Heraeus Biofuge plus	
Spectrophotometer- UV Vis Spekol	Zeiss (Jena)
Analytical balance- Sartorius BP 210 D	Sartorius (Goettingen)
Table balance- Sartorius portable PT2100	
pH meter- CG 840 Schott	Schott (Mainz)
Water purification- Milli-Qplus 185	Millipore (Eschborn)
Pharmacia LKB PhastSystem	Pharmacia (Freiburg)
Microscope- Olympus SZH10 binocular	Olympus (Hamburg)
Microscope- Zeiss Stemi 1000 binocular	Zeiss (Jena)
X-ray generator- rotating anode Nonius FR591	Nonius (Delft, The Netherlands)
Image plate detector- Mar 300	Mar Research (Hamburg)
Image plate detector- Mar 345	
Cryostat- Oxford controller 600 series	Oxford Cryo (Oxford, UK)
Air stream cooler- FTS TC-84	FTS systems (Stone Ridge, USA)
Goniometer head- Charles Supper Standard	Charles Supper (Troy, USA)
Microscope- Leica MZ 8 binocular	Leica (Bensheim)
Indy workstation	SGI (Mountain View, USA)
Onyx graphics workstation	

O2 graphics workstation	SGI (Mountain View, USA)
Indigo2 graphics workstation	
Synchrotron- EMBL Hamburg beamlines X11, BW7B	DESY (Hamburg)
Synchrotron- Elettra Light Source beamline 5.2 R	ELETTRA (Trieste, Italy)
Synchrotron- ESRF Grenoble beamline BM30	ESRF (Grenoble, France)

2.2. Methods

2.2.1 Determination of protein purity

The purity of the protein samples was determined by denaturing polyacrylamide gel electrophoresis in presence of the detergent sodium dodecyl sulphate (SDS-PAGE) (Shapiro *et al.*, 1967) using the Pharmacia LKB PhastSystem. Cellosyl (10 mg/ml) was mixed with the sample buffer in a ratio of 1:10, the latter containing 8 M urea, 2 % (w/v) SDS, 5 % β -mercaptoethanol as reducing agent, 10 % (w/v) sucrose and 0.004 % (w/v) bromphenol blue. The mixture was heated for ca. 5 min at 100 °C to ensure complete denaturation of the protein sample. The protein standards of known molecular weight were treated the same way. The Cellosyl samples and the protein standards were applied to the gel and stained after electrophoresis with Coomassie[™] Brilliant Blue.

Under given conditions, the electrophoretic mobility of a protein depends solely on its size (Reynolds and Tanford, 1970). The integrity of the Cellosyl sample was determined by comparing its migration distance in the gel with the following protein standards: phosphorylase b (97000 Da), albumin (66000 Da), ovalbumin (45000 Da), carbonic anhydrase (30000 Da), trypsin inhibitor (20100 Da) and α -lactalbumin (14400 Da). Cellosyl was only used for crystallisation trials if a single characteristic band at a molecular weight of 23000 Da was observed by SDS-PAGE.

2.2.2 Determination of protein concentration

The protein quantity in solution was determined by the Bradford method (Bradford, 1976) using the Bio-Rad Protein-Assay. The method is based on a shift in the absorption maximum of Coomassie™ brilliant blue from 465 nm to 595 nm upon binding to arginine and aromatic residues of the protein sample. The increase in absorption at 595 nm is a measure for the protein concentration in the protein solution under investigation. For this very sensitive test, 1 µl of a Cellosyl solution was brought up to a volume of 800 µl with water and 200 µl concentrated Bio-Rad agent was added. The mixture was allowed to incubate for 5 min at room temperature before measuring the absorption at 595 nm. A blank buffer solution (800 µl water / 200 µl Bio-Rad agent) served as reference. The protein quantity was determined from a calibration curve prepared with bovine serum albumin and γ -globulin in a concentration range of 10 – 350 µg/ml.

2.2.3 Dialysis

Dialysis was used for exchange of the storage buffer (10 mM KH_2PO_4 at pH 6.0) of the Cellosyl samples prior to crystallisation trials. The protein solution was injected into the Slide-A-Lyzer® dialysis cassette with a syringe and the cassette was clasped in a buoy and suspended in the dialysis buffer containing an aqueous solution of 20 mM Tris at pH 7.0. After dialysing for 20 h, the protein sample was again removed with a syringe from the cassette.

2.2.4 Sample concentration

Cellosyl samples were concentrated using an Ultrafree®-MC 1000 NMWL filter unit. The samples were centrifuged at 12000 rpm until the required concentration was achieved (Bradford test).

2.2.5 Crystallisation

The basic principles and variables influencing the crystallisation of biological macromolecules are comprehensively described by McPherson (1982). Crystallisation is a self-ordering process where molecules go from a supersaturated solution into the solid state

and, despite intensive research, it remains mostly a trial-and-error procedure. Supersaturation of the Cellosyl solution was obtained by addition of a precipitant, i.e. PEG and ammonium sulphate. There are several methods to slowly increase the protein and precipitant concentration, e.g. liquid-liquid or vapour diffusion. In this work, all crystallisation trials, unless otherwise stated, were carried out using the hanging-drop vapour diffusion method. 2-4 μ l of protein were mixed with 2 μ l reservoir solution and pipetted onto a siliconised cover slip. This cover slip was inverted over a well of the crystallisation tray which was pre-filled with 1 ml of a well-defined precipitant solution, and sealed with grease to create a closed system. Because the precipitant concentration in the drop is lower than in the reservoir, the volume of the drop decreases due to vapour diffusion. This leads to an increase of the protein and precipitant concentration, and if the conditions are sensibly chosen, crystal growth can be observed. The exact crystallisation conditions for the two crystal forms of Cellosyl are described in chapter 3.2.

2.2.6 Heavy-atom and polysaccharide soaks

The incorporation of heavy atoms into the crystal lattice is crucial for phase determination by single (SIR) and multiple isomorphous replacement (MIR) techniques, and the method of multiple wavelength anomalous dispersion (MAD). Heavy-atom derivatives were prepared by adding aqueous solutions of heavy-atom compounds directly to the drops containing preformed crystals. The crystals were allowed to soak for a minimum period of 72 h at room temperature. A variety of heavy-atoms were screened, including mercury chloride (HgCl_2), osmium chloride (OsCl_3), cis-Platinum ($\text{Pt}(\text{NH}_3)\text{Cl}_2$), potassium tetrachloroplatinate (K_2PtCl_4), yttrium chloride (YCl_3) and ytterbium chloride (YbCl_3). Double-derivatives were prepared by addition of two heavy-atom compounds to a drop, e.g. HgCl_2 and OsCl_3 . The concentration of the heavy-atom compound in the crystallisation drops varied from 2.5 mM to 10 mM. The crystals were stable over the whole concentration range.

Chitin with a defined number of polysaccharide moieties, i.e. N,N',N''-Triacetyl-Chitotriose, N,N',N'',N''',N''''',N''''''-Hexaacetyl-Chitohexose, and a chitosan polysaccharide mixture was used to prepare substrate / inhibitor complexes. The respective polysaccharide was dissolved in water to give a 10 mM stock solution and added to the crystals so that a final concentration range from 0.5 mM to 2 mM in the crystallisation drops was covered.

2.2.7 Cryocooling

Cryocooling techniques are routinely used in macromolecular crystallography to preserve crystals and reduce radiation damage during X-ray data collection. To protect Cellosyl crystals from the damaging effect of ice formation they were treated with highly liquid paraffin oil as a cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). Prior to application, the paraffin oil was dried in a rotary vacuum centrifuge at 333 K for 1 h and allowed to adjust to the same temperature as the crystals. The crystals were taken up in a cryoloop and slid through the oil before they were flash-cooled in the nitrogen stream.

2.2.8 Data acquisition and processing

Native as well as heavy-atom derivative data were collected using CuK_α radiation and a MAR 300 mm imaging plate detector with 1° rotation per image. The benefits of synchrotron radiation were exploited for the collection of MAD data and high-resolution native data.

All diffraction data were analysed using the *HKL* package (Otwinowski and Minor, 1997). First, the original unprocessed diffraction patterns were visualised with the program *XdisplayF*. Lattice type, unit cell parameters and the crystal orientation were determined with the autoindexing routine of *DENZO*, usually using the first oscillation image. Only if problems occurred, e.g. due to a small crystalline impurity, a different image (i.e. at 60°) was used. The refinement of crystal and camera parameters and the integration of the diffraction maxima were also performed with the program *DENZO*. Correction of the polarisation of the X-ray beam and the air absorption of the scattered X-rays were only included in the integration of data collected at a synchrotron radiation source.

Scaling, post-refinement of the crystal parameters, merging and statistical analysis of all data sets collected were performed with the program *SCALEPACK* (Otwinowski and Minor, 1997). The maximum resolution to which a crystal diffracted was determined by analysing the ratio of the measured intensity to its standard deviation, $I/\sigma(I)$. All data at a high-resolution limit were omitted if their mean value of $I/\sigma(I)$ in the resolution bin was less than 2. Other criteria to assess the quality of the data measured were the completeness of the dataset, the redundancy of the data, and the merging *R*-factor (R_{merge}) (see equation {1}) which is a measure of how well multiple observations of the same reflection and its symmetry related reflections merge.

$$R_{\text{merge}} (\%) = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad \{1\}$$

Since R_{merge} is dependent on the redundancy of the measured data (R_{merge} is increasing the more often a given reflection is measured), the redundancy-independent R -factor ($R_{\text{r.i.m.}}$ {2}) and the precision indicating R -factor ($R_{\text{p.i.m.}}$ {3}) were calculated with the program *Rmerge* (Weiss and Hilgenfeld, 1997; Weiss, 2001).

$$R_{\text{r.i.m.}} (\%) = \frac{\sum_{hkl} \sqrt{\frac{N}{N-1}} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad \{2\}$$

$$R_{\text{p.i.m.}} (\%) = \frac{\sum_{hkl} \sqrt{\frac{1}{N-1}} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad \{3\}$$

Further data reduction was carried out using the CCP4 suite of programs (CCP4, 1994). The merged data from *SCALEPACK* were converted into an *mtz* format by the program *SCALEPACK2MTZ* (CCP4, 1994). *TRUNCATE* (French and Wilson, 1978) was used to calculate mean amplitudes from the averaged intensities and to put them on an absolute scale using estimated scale factors from the Wilson plot (Wilson, 1942).

2.2.8.1 Native data collection on the monoclinic crystal form

Data were collected at 100 K with a MAR 300 mm imaging plate detector and CuK_α radiation, using highly liquid paraffin oil as cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). At a crystal-to-detector distance of 100 mm, the crystal still diffracted beyond the edge of the plate. Due to the set-up of the cryosystem, the detector could not be moved closer and a data set to a maximum resolution of 1.65 Å was collected.

Integration, scaling and analysis of the diffraction data were performed with the above-described routine. The crystal could be unambiguously assigned to the monoclinic space group C2 with cell dimensions $a = 111.15 \text{ \AA}$, $b = 38.22 \text{ \AA}$, $c = 51.04 \text{ \AA}$, $\beta = 108.21^\circ$.

2.2.8.2 Native data collection on the hexagonal crystal form

A single data set was collected at beamline BM30 of the ESRF (Grenoble) with an incident wavelength of 0.9797 Å and a crystal-to-detector distance of 330 mm. The crystal was cooled in a stream of nitrogen gas at 100 K using highly liquid paraffin oil as cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). The crystal diffracted to a maximum resolution of 2.3 Å, but unfortunately the diffraction pattern was very anisotropic.

A total of 180 images were collected with a 1° rotation of Φ and an exposure time of 15 seconds per image. The unprocessed diffraction patterns were visualised with the program *XdisplayF* and autoindexed using the program *DENZO* (Otwinowski and Minor, 1997). Autoindexing was quite difficult since a small crystalline impurity caused a weak second lattice on some images. Performing the peak picking manually circumvented this problem. Additionally, determination of the distance from the edge of the data to the beam spot (x beam, y beam) caused problems since the beam was not in the centre of the diffraction image. The default beam-centre value is the half of the film width, which would be 150 mm in the x- and y-directions for a 300 mm image plate. The actual position of 148 mm and 144 mm in the x- and y-directions, respectively, could not be determined by the automatic refinement procedure and was therefore worked out manually. A hexagonal lattice and unit cell parameters of $a = b = 124.59$ Å and $c = 143.18$ Å could be determined.

Since the mosaicity of the crystals was relatively high (0.72), the reflections were scaled again with the program *SCALA* (CCP4, 1994). In contrast to the program *SCALEPACK*, *SCALA* employs a different scaling algorithm, which includes a correction of the tails for partial reflections to improve poor partial bias. However, the deviation in the scaling statistics between both programs was negligible.

2.2.8.3 MAD data collection

An osmium-derivatised Cellosyl crystal was selected for MAD data collection, since osmium absorbs X-ray's in an energy range which is easily accessible at a synchrotron radiation source. The absorption of X-ray energy causes a difference in the intensity of the Bijvoet pairs $|F(+)|^2$ and $|F(-)|^2$, which can be used for phase determination. This difference is very small, making out only approximately 3% of the total scattering. The anomalous scattering contributions f' (real part) and f'' (imaginary part) of an heavy atom are wavelength-dependent and can be determined from theoretical plots. Figure 2.2.8.3-1 depicts

the anomalous scattering contributions f' and f'' for osmium in dependence of the wavelength.

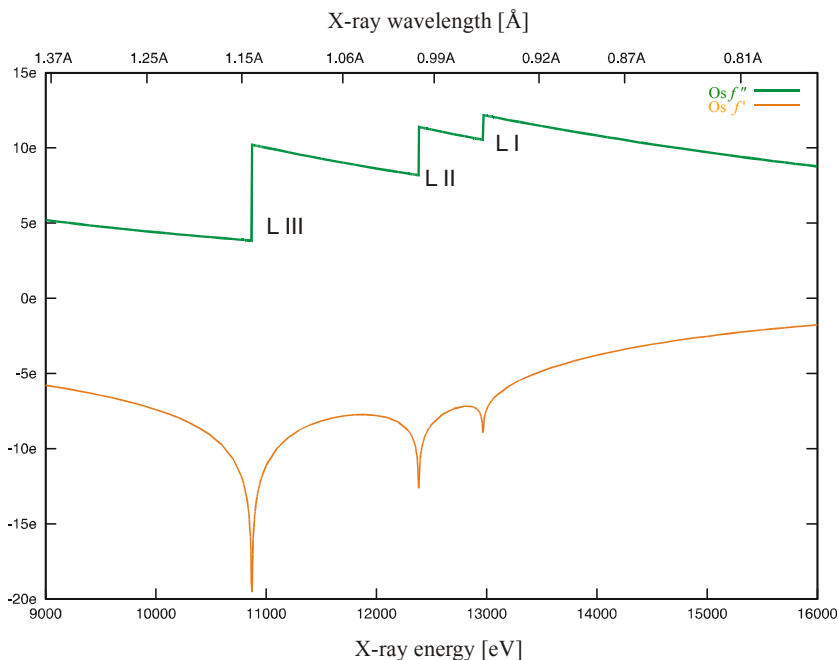


Figure 2.2.8.3-2 Theoretical plots of f' and f'' over the three L-absorption edges of osmium. The theoretical absorption edges are at the energy of 10.871 keV (LIII), 12.385 keV (LII) and 12.968 keV (LI). Data for this plot was obtained from Merritt (1996).

However, these plots do not take the effect of neighbouring atoms on the heavy-atom scattering values into account. Since the effect can be quite pronounced close to an absorption edge, the scattering behaviour of the crystal was determined by recording a fluorescence scan around the LIII absorption edge of osmium. Based on this scan, diffraction data were collected at three different wavelengths:

- $\lambda_1 = 21.140868 \text{ \AA}$ (10.8676 keV), with maximal f'' and the largest difference between the Bijvoet pairs $|F(+)|^2$ and $|F(-)|^2$
- $\lambda_2 = 1.141099 \text{ \AA}$ (10.8654 keV), at the inflection point of the absorption edge, with minimal f'
- $\lambda_3 = 1.142182 \text{ \AA}$ (10.8551 keV), at the low energy side of the absorption edge, where the absorption of the heavy atom and therefore f'' is small and the dispersive difference between f' of λ_2 and λ_3 is maximal

All data sets were collected on the same crystal in order to reduce systematic errors and obtain more accurate estimates of the intensity difference between the Bijvoet pairs. The crystal diffracted to a resolution of 1.8 at a crystal-to-detector distance of 200 mm. In order to obtain highly redundant data, 249 images were collected at each wavelength with a $\Delta\Phi$ rotation angle of 1° per image. After each pass, Φ was reset to zero and exactly the same range was measured again.

All three data sets were integrated with the program *DENZO*. Data reduction and scaling was performed with the program *SCALEPACK* (Otwinowski and Minor, 1997). A monoclinic unit cell of $a = 111.14 \text{ \AA}$, $b = 37.75 \text{ \AA}$, $c = 50.99 \text{ \AA}$, $\beta = 108.21^\circ$ was determined. The keyword 'ANOMALOUS ON' was activated in the *SCALEPACK* input file to merge the Bijvoet pairs separately and output them as $I(+)$ and $I(-)$ for each reflection.

The *SCALEPACK* output files for all three data sets were transmuted into *CNS* reflection file format and *CNS* (Brünger *et al.*, 1998) was used for further data reduction. All three data sets were merged into a single reflection file and placed on a common scale, using the data set collected at $\lambda 3$ as the reference set.

2.2.8.4 Data collection on heavy-atom derivatised crystals

Diffraction data were collected on an individual osmium derivative, as well as on an osmium-mercury double derivative of the monoclinic Cellosyl crystals, using CuK_α radiation. The data were recorded at cryogenic temperatures, using highly liquid paraffin oil as cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). Both crystals diffracted to a maximum resolution of 1.8 \AA . As can be seen from Figure 2.2.8.4-1, the anomalous scattering contributions of osmium and mercury are significant using CuK_α radiation. Therefore, Bijvoet pairs were merged separately in *SCALEPACK*. The cell dimensions of both crystals were isomorphous, with $a = 111.26 \text{ \AA}$, $b = 37.75 \text{ \AA}$, $c = 51.22 \text{ \AA}$, $\beta = 108.31^\circ$ for the osmium derivative and $a = 111.28 \text{ \AA}$, $b = 38.33 \text{ \AA}$, $c = 50.98 \text{ \AA}$, $\beta = 108.23^\circ$ for the osmium-mercury double derivative.

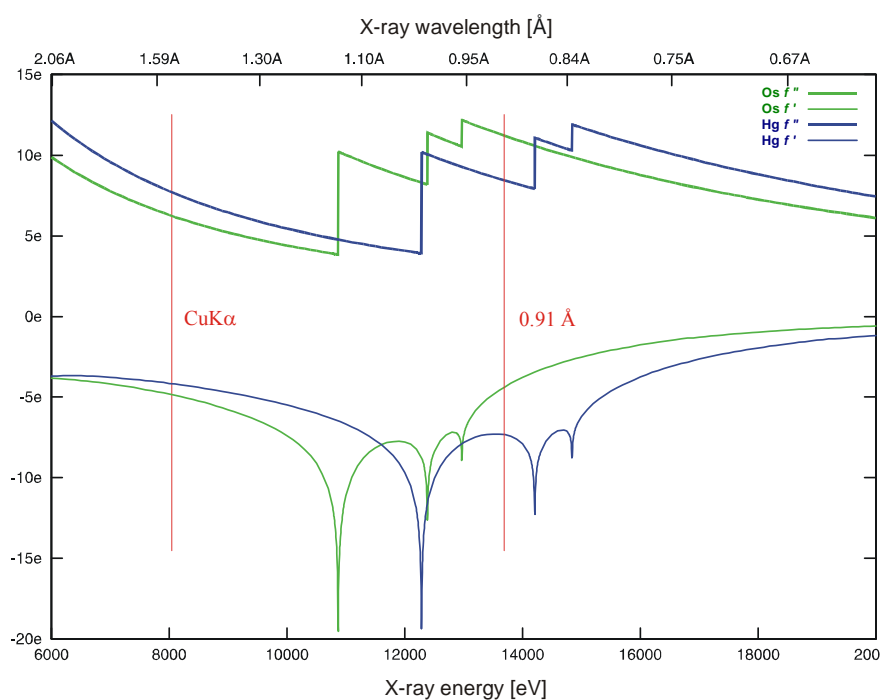


Figure 2.2.8.4-1 Theoretical plots of f' and f'' over the three L-absorption edges of osmium and mercury. Vertical red bars indicate the X-ray energy of the incident beam, chosen for data collection. Data for this plot was obtained from Merritt (1996).

A further data set was collected on a mercury derivatised crystal at the beamline X11 of the EMBL Outstation at DESY, Hamburg, at an incident wavelength of 0.913 Å. The crystal diffracted to a maximum resolution of 1.55 Å using cryogenic temperatures of 100 K. As described for the other heavy-atom derivatives, 'ANOMALOUS ON' was activated in the SCALEPACK input file, and $I(+)$ and $I(-)$ were recorded for each reflection. Cell dimensions of $a = 111.24$ Å, $b = 38.51$ Å, $c = 50.80$ Å, $\beta = 108.18^\circ$ were determined.

2.2.8.5 Collection and processing of atomic-resolution data

At first, a single data set was collected at the multipole wiggler beamline BW7B at the EMBL Hamburg Outstation equipped with a MAR 345 mm image plate detector. Using an incident wavelength of 0.8445 Å and a crystal-to-detector distance of 120 mm, the crystal diffracted to a maximum resolution of 0.92 Å. The reduction of the crystal-to-detector distance from 120 mm to 90 mm, after 56 images, increased the maximum resolution limit from 0.92 Å to 0.82 Å resolution. The overall scanning range of the Φ angle comprised 135°. Images were recorded in 0.5° oscillation steps.

In order to avoid oversaturation, a second and a third data set were recorded at a lower X-ray dose to a maximum resolution of 1.84 Å and 1.24 Å, respectively (see Table 2.2.8.5-1).

Table 2.2.8.5-1 Data collection parameters for the three passes on a single Cellosyl crystal.

pass	frames	Φ_{start} [°]	Φ_{end} [°]	oscil. steps [°]	dose	detector dist. [mm]	max. res. [Å]
1	1 - 56	0	28	0.5	10000	120	0.92
	57 - 270	27	134	0.5	10000	90	0.82
2	1 - 180	0	180	1	2000	180	1.24
3	1 - 88	0	176	2	400	180	1.84

All diffraction data were analysed using the HKL package (Otwinowski and Minor, 1997). The original unprocessed diffraction patterns were examined with the program XdisplayF. A monoclinic lattice type, unit cell parameters of $a = 111.1$ Å, $b = 38.5$ Å, $c = 51.0$ Å, $\alpha = \gamma = 90^\circ$, $\beta = 108.2^\circ$ and the crystal orientation were determined with the autoindexing routine of DENZO using a single oscillation image.

DENZO was further used for the refinement of crystal and detector parameters and for the integration of the diffraction maxima of all three data sets. To correct for the polarisation of the X-ray beam and the air absorption of the scattered X-rays, a polarisation value of 0.9 and an air absorption length of 5234 mm were included in the refinement.

Further, shadows of the beamstop, and also the shadow of the cryo nozzle in the high-resolution data set, had to be omitted from the images. The MAR IP 345 mm detector coordinate system has its origin in the upper left corner with the x-axis pointing downward and the y-axis to the right. In this coordinate system a circle was defined giving the centre and the radius in mm (e.g.: ignore circle 172.5 172.6 5.85) to mask out the shadow of the beamstop. Masking out the shadow of the cryo nozzle was more complicated. Even though the shadow resembled a half circle on the edge of the image plate the 'ignore circle' keyword could not be used since it was not possible to define a centre of the circle outside the diffraction image. Therefore two irregular quadrangles, defined by four x, y coordinates, were designed to mask out this shadow (e.g.: ignore quadrilateral 114.0 36.6 87.9 44.1 59.1 42.6 87.7 4.1).

Scaling, postrefinement of the crystal parameters, merging and statistical analysis were conducted with the program SCALEPACK (Otwinowski and Minor, 1997). The maximum usable resolution was defined by a value for $I/\sigma > 2.0$. During the high-resolution pass, the maximum usable resolution decreased although the data were collected at 100 K. This is indicative of radiation damage. Therefore the maximum resolution was gradually reduced for processing.

Based on these scaling statistics, the high-resolution data set was divided into 4 batches, Φ ranging from 1 - 28°, 28.5 - 63.5°, 64 - 83.5° and 84 - 135°, which were integrated to a maximum resolution of 0.92 Å, 0.83 Å, 0.86 Å and 0.90 Å, respectively. The integration procedure of DENZO was repeated for all batches to the respective resolution limits using the postrefined crystal orientation, unit cell and mosaicity from the first run of SCALEPACK. Subsequently the data of all three data sets were merged in the resolution range of 99.0 - 0.83 Å. The data were divided into 40 resolution shells for the calculation of statistics, the outermost shell including data from 0.84 – 0.83 Å. The monoclinic space group C2 with unit cell parameters of $a = 111.19$ Å, $b = 38.45$ Å, $c = 50.93$ Å, $\alpha = \gamma = 90.0^\circ$ and $\beta = 108.33^\circ$ was determined for the crystal.

Subsequent data reduction was carried out using the CCP4 suite of programs (CCP4, 1994), i.e. conversion of the merged data from SCALEPACK into an mtz format was performed with the program SCALEPACK2MTZ (CCP4, 1994) and the program TRUNCATE was used to calculate mean amplitudes from the averaged intensities and to put them on an absolute scale using estimated scale factors from the Wilson plot (French and Wilson, 1978). The approximate overall temperature factor estimated from the Wilson plot (Wilson, 1942) was 6.5 \AA^2 .

2.2.9 Phase determination

X-rays scattered by the electron shell of the protein atoms in a crystal contain the complete information of the three-dimensional structure of the protein. This information is encoded in the amplitude and the phase of the scattered X-rays. The wavelength of the scattered X-ray is the same as the one of the incident beam and the amplitude can be determined from the relative intensity of the reflections. However, the phase cannot be directly measured, thus an important part of the information about the three-dimensional electron density distribution is lost. To obtain phases is one of the most difficult tasks in macromolecular crystallography and known as the “phase problem”.

In the present work, three methods to obtain phase information were employed. The basic principles of these methods are described in the following paragraphs. A detailed description of the specific aspects for the respective experiments can be found in chapter 3 of this thesis.

2.2.9.1 Molecular replacement

Molecular replacement was performed using two different programs, *CNS* (Brünger *et al.*, 1998) and *EPMR* (Kissinger *et al.*, 1999). A requirement for the molecular replacement procedure is a search model with a known three-dimensional structure. This could be, for example, the crystal structure of a protein with high sequence similarity or the structure of the same protein in a different space group. Because of the differences in space group and orientation between the search model and the unknown crystal structure, the search model has to be correctly placed into the new unit cell. Then, structure factors can be calculated from the correctly orientated search model and the phases of these calculated structure factors serve as initial phase estimates for the new protein structure (Drenth, 1999).

The program *CNS* employs a separate search for the orientation (rotation search) and for the location (translation search) of the search model in the new unit cell. The rotation search is performed by a correlation of the Patterson maps computed from the observed structure factor amplitudes (F_o) of the investigated protein and of the calculated structure factor amplitudes (F_c) computed from the coordinates of the search model. If the rotation function is determined, the search model is rotated accordingly and then translated through the unit cell by monitoring the correlation of the structure factor amplitudes.

The program *EPMR* employs an evolutionary optimisation algorithm, which allows a direct search for the three rotational and translational parameters. This six-dimensional search starts with arbitrary molecular replacement solutions by random generation of values for the orientation and position of the search model in the unit cell. A stochastic ranking is used to choose the solutions, which survive into the next round. The orientation and location of the surviving members of the population is maintained and used to generate offspring by applying normally distributed random mutations to the orientation of the parent solution in order to create the next population. This process is reiterated for a fixed number of generations, after which the solution with the highest correlation coefficient between the observed and calculated structure factors is chosen for a conjugate gradient optimisation procedure (Powell, 1977). This very sensible and reliable procedure allows the rapid and highly automated

solution of molecular replacement problems with single or multiple molecules in the asymmetric unit (Kissinger *et al.*, 1999).

Once the search model is placed in the new unit cell, an electron density map can be calculated using model phases and the observed structure factor amplitudes. Ideally, difference Fourier synthesis should show the regions of disagreement between the search model and the new structure, which can now be determined through rebuilding and refinement.

2.2.9.2 Multiple wavelength anomalous dispersion

A further attempt to obtain phases was carried out using the method of multiple wavelength anomalous dispersion. In crystal structure determination, it is usually assumed that the intensities of the reflections (h, k, l) and $(-h, -k, -l)$ are equal (Friedel's law). The reflections (h, k, l) and $(-h, -k, -l)$ are called a Friedel or a Bijvoet pair. However, the presence of heavy atoms in the crystal can lead to the abolishment of Friedel's law. Heavy atoms can absorb X-rays of specific wavelength. This leads to a participation of the inner electrons in the scattering process and gives rise to a phase shift and to a difference in the intensities of (h, k, l) and $(-h, -k, -l)$. This disproportion between the Friedel mates is called anomalous scattering or anomalous dispersion (Rhodes, 2000). The anomalous scattering power of a heavy atom is dependent on the wavelength of the incident X-ray beam. Taking advantage of tuneable synchrotron radiation sources, this wavelength dependency can be exploited to locate the anomalous scattering atoms within the unit cell and calculate the corresponding phase angle. Location of the heavy-atom sites and MAD phasing were performed using the programs *CNS* (Brünger *et al.*, 1998) and *SHARP* (De LaFortelle and Bricogne, 1997).

2.2.9.3 Multiple isomorphous replacement with anomalous scattering

The major breakthrough in the elucidation of the three-dimensional structure of Cellosyl was achieved employing the method of multiple isomorphous replacement with anomalous scattering (MIRAS). Heavy atom containing protein crystals, which are isomorphous to the native crystals, are essential for this method. That implies that binding of the heavy atom to the protein does neither alter its structure nor its crystal packing. Non-isomorphism is reflected in altered cell constants or a change in crystal symmetry and renders these crystals futile for this method. If native and heavy atom crystals are isomorphous, the

reflections occur under the same geometry. Because of their high number of electrons, heavy atoms scatter X-rays much stronger than the light atoms of amino acids, which leads to a significant change in the relative intensities of the reflections between the native and derivative crystals. This difference in relative intensities can be used to determine the position of the heavy atoms in the unit cell, facilitated by calculation of Patterson functions from the differences between structure factor amplitudes of the native and derivatised crystal (Rhodes, 2000).

From the Patterson vectors, it is possible to derive the position of the heavy atom in the unit cell. With this knowledge, initial phases can be estimated for the protein structure. Using only one heavy atom derivative, the phase estimates are vague and it is highly unlikely to obtain an interpretable electron density map. However, phase estimates can be improved by combination of the phase information of different heavy-atom derivatives.

R_{Cullis} is one of the most useful indicators of the validity of the heavy-atom model. Values of $R_{\text{Cullis}} < 0.6$ for centric reflections (eq. {4a}) are considered to be very good, and values < 0.9 are usable for phasing. If the R_{Cullis} for anomalous data (eq. {4b}) is less than 1, the heavy-atom data usually provide significant phase information (CCP4 tutorial).

$$R_{\text{cullis}}(\text{iso}) = \frac{\sum_{h,k,l} \|F_{PH} \pm F_P - F_H(\text{calc})\|}{\sum_{h,k,l} |F_{PH} \pm F_P|} \quad \{4a\}$$

$$R_{\text{cullis}}(\text{ano}) = \frac{\sum_{h,k,l} \|\Delta F_{PH}^{\pm}(\text{obs}) - |\Delta F_{PH}^{\pm}(\text{calc})|\|}{\sum_{h,k,l} \Delta F_{PH}^{\pm}(\text{obs})} \quad \{4b\}$$

, where $\Delta F_{PH}^{\pm}(\text{obs})$ is the structure factor amplitude difference between Bijvoet pairs and $\Delta F_{PH}^{\pm}(\text{calc}) = 2 \frac{f''}{f'} |F_H \sin(\alpha_{PH} - \alpha_H)|$.

Another measure of the quality of a heavy-atom derivative is its phasing power and the mean figure of merit (FOM) (Blow and Crick, 1959). The phasing power is defined as the mean heavy-atom amplitude divided by the residual lack of closure error:

$$\frac{\sum_{hkl} |F_H(\text{calc})|}{\sum_{hkl} |E|} \quad \{5\}$$

, where the lack of closure error, E, is the difference between the observed and calculated values of the structure factor amplitude of the heavy-atom derivative

The mean figure of merit is a measure of the precision of the "best" phase:

$$FOM = \frac{|F(h, k, l)_{best}|}{|F(h, k, l)|} \quad \{6\}$$

$$\text{with } F(h, k, l)_{best} = \frac{\sum_{\alpha} P(\alpha) F_{h,k,l}(\alpha)}{\sum_{\alpha} P(\alpha)}, \quad P(\alpha) = \text{total probability of phase angle } \alpha \text{ as derived from more than one derivative}$$

In this work, the structure of Cellosyl was determined using an osmium, a mercury and an osmium-mercury double derivative. In addition to the isomorphous difference to the native crystal, the anomalous scattering of the heavy atoms was used to obtain phases (see chapter 3.3.5).

2.2.10 Model building and electron-density maps

The electron density ρ can be calculated at each point (x, y, z) of the crystal from the structure factor amplitude $F(h, k, l)$, which is proportional to the square root of the measured intensity I for the reflection (h, k, l) , the phase $\alpha(h, k, l)$, and the volume V of the unit cell:

$$\rho(x, y, z) = \frac{1}{V} \sum F(h, k, l) \cdot e^{ia(h,k,l)} \cdot e^{-2\pi i(hx+ky+lz)} \quad \{8\}$$

In order to obtain a molecular model of Cellosyl, an initial electron density map was calculated using the measured intensities of the native data set and the phase information derived from the MIRAS experiment to 2 Å resolution. The initial phases were of very good quality. The electron density of the protein was clearly distinguishable from the bulk water and features like α -helices could be recognized. Model building was automatically performed with the program *ARP/wARP* (Perrakis *et al.*, 1999). The program was able to build all amino acids, except the C-terminal Ala217, and a majority of the side chains (see chapter 3.3.6) combining the initial model and the native data to 1.65 Å. The program *O* (Jones *et al.*, 1991) was used for visualisation, manual model building and verification.

In the process of improving the quality and resolution of the electron density map, the observed phases are replaced or combined with calculated phases from the initial model. This increases the danger of introducing bias from the model into the electron density. In order to minimize model bias, several difference Fourier syntheses containing the structure factor amplitude in form of $(n|F_{obs}| - |F_{calc}|)$ were calculated. The calculated structure-factor amplitudes are subtracted from multiples of the observed structure-factor amplitudes in each Fourier term. The resulting electron density map for $n=1$ is called a $\mathbf{F}_o - \mathbf{F}_c$ map (eq. {9}).

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} (|F_o| - |F_c|) \cdot e^{-2\pi i(hx+ky+lz) + i\alpha_{calc}} \quad \{9\}$$

The $\mathbf{F}_o - \mathbf{F}_c$ map reveals errors in the current model. The electron density is either positive or negative, depending on which of \mathbf{F}_o or \mathbf{F}_c is larger. Positive electron density implies that atoms of the model have to be moved towards that region, whereas in regions of negative density atoms have to be moved away, i.e. wrongly built amino-acid side chains, indicated by patches of negative density, can be moved into nearby regions of positive density and so adjusted to the right conformation (Drenth, 1999).

For $n=2$, the Fourier difference map is called $2\mathbf{F}_o - \mathbf{F}_c$ map (eq. {10}). The $2\mathbf{F}_o - \mathbf{F}_c$ map shows the electron density of the model and is solely positive if the model does not contain severe errors (Drenth, 1999).

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} (2|F_o| - |F_c|) \cdot e^{-2\pi i(hx+ky+lz) + i\alpha_{calc}} \quad \{10\}$$

SigmaA weighted maps reduce model bias even further, applying weighting to the observed structure factor amplitudes (Read, 1986). The equations for the (mF_o-DF_c)-map and ($2mF_o-DF_c$)-map are shown in {11} and {12}, respectively.

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} (m|F_o| - D|F_c|) \cdot e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}} \quad \{11\}$$

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} (2m|F_o| - D|F_c|) \cdot e^{-2\pi i(hx+ky+lz)+i\alpha_{calc}} \quad \{12\}$$

, where m is the figure of merit and D is the estimated error in the (partial) structure from the Luzzati plot (Luzzati, 1952).

SigmaA weighted maps were calculated using *SHELXPRO* (option 'M') from the *pdb* and *fcf* output files of the atomic resolution model of Cellosyl.

2.2.11 Structure refinement

After a model has been obtained from a MAD, MIRAS or molecular replacement experiment, the positional parameters and the temperature factor of each atom have to be refined in order to adjust the protein model to fit the observed data. A measure for the agreement between the structure factors calculated for the existing model and the observed structure factors is the crystallographic R -factor (eq. {13}, where k is a scale factor).

$$R = \frac{\sum_{hkl} || F_{obs} | - k | F_{calc} ||}{\sum_{hkl} | F_{obs} |} \quad \{13\}$$

Because the observation-to-parameter ratio in protein crystallography is rather low, there is the possibility to overfit data, especially at moderate resolution. Therefore, structure validation using the free R -value (eq. {14}) has been introduced by Brünger (1992). To compute R_{free} , data are divided into a working set and a test set ($hkl \subset T$). The test set comprises a random selection of about 5-10 % of the observed reflections. Only the working set is used to refine the model. The free R -value is computed from the test set, which is

omitted in the refinement process. If a structure is improved during refinement, both the R -value and R_{free} should decrease.

$$R_{free} = \frac{\sum_{hkl \subset T} \| |F_{obs}| - k |F_{calc}| \|}{\sum_{hkl \subset T} |F_{obs}|} \quad \{14\}$$

Throughout this work, different refinement programs were employed, i.e. *CNS* (Brünger *et al.*, 1998), *REFMAC* (CCP4, 1994; Murshudov *et al.*, 1997) and *SHELX-97* (Sheldrick and Schneider, 1997).

The *CNS* program suite was used to refine the low-resolution model of Cellosyl in the monoclinic crystal form. First, rigid body refinement was performed, where the Cellosyl molecule was regarded as a rigid entity and its position and orientation in the unit cell was optimised. The model was further improved by simulated annealing using torsion angle dynamics. In order to correct large errors in the initial model, the dynamic behaviour was simulated using a starting temperature of 5000 K. Then, the atomic positions were optimised by conjugate-gradient refinement. As last step of a refinement cycle, refinement of the thermal parameter B was carried out. The isotropic B -value reflects the thermal motion of the atoms in the unit cell and is defined as: $B = 8\pi \times \overline{u^2}$ (where $\overline{u^2}$ is the mean square displacement of the atom vibration).

The program *REFMAC*, which was used for the refinement of the Cellosyl model in the hexagonal crystal form, implements the maximum likelihood formalism. A special feature, which is included into *REFMAC5*, is the possibility to refine TLS parameters (Schomaker and Trueblood, 1968; 1998) to describe the anisotropic motion of rigid bodies. Even at moderate resolution, the mean square displacements of rigid bodies can be described since, in contrast to the large number of parameters needed for a full anisotropic refinement, only 20 parameters are required for each TLS group (Winn *et al.*, 2001).

The program *SHELX-97* was used to refine the structure of Cellosyl at 0.83 Å resolution. The program was originally designed for the refinement of small molecules but has been adapted for the refinement of biological macromolecules at atomic resolution. The program is based on a least-square refinement algorithm and is especially suitable for the refinement of disorder and anisotropic displacement parameters. To operate *SHELXL*, two input files were necessary, namely the *hkl* file which contains h , k , l , F and $\sigma(F)$ or better F^2

and $\sigma(F^2)$ and the *ins* file containing crystal characteristics and instructions. Both files were generated with the auxiliary program *SHELXPRO*, an interactive user-interface between *SHELX* and other programs. The *hkl* file was generated from the *CNS* format reflection file using the option 'H' in *SHELXPRO* and contained *h*, *k*, *l*, *F* and $\sigma(F)$. 1 % of the unique reflections (1767) covering the full resolution range (40.0 - 0.83 Å) were randomly selected and flagged for use as a reference set for calculation of free *R*-values (Brünger, 1992). The initial *ins* file was prepared using a low-resolution model of Cellosyl. All solvent atoms and ions were omitted from the model, before the *pdb* file was read into *SHELXPRO* (option '1'), which automatically generated standard Engh & Huber restraints (Engh and Huber, 1991), applied consistency checks and incorporated instructions needed for refinement.

Alternate sessions of model building using the program *O* (Jones *et al.*, 1991) and refinement were carried out in order to improve the quality of the structural model.

2.2.12 Validation of model quality

After the protein structures had been refined, the electron density maps were carefully inspected to assure the connectivity of the main-chain and the side-chain atoms and the correct interpretation of the side-chain electron density. The crystallographic *R*-factor gave a measure of how well the protein model fitted the observed data.

In order to examine the stereochemistry of the polypeptide chain, a Ramachandran plot (Ramachandran and Sasisekharan, 1968) was prepared for the atomic resolution model using the program *SHELXPRO* and for all other structures with the program *PROCHECK* (Laskowski *et al.*, 1993). The dihedral angles Φ and Ψ for each residue were plotted in a square matrix. For all structures, the majority of the Φ/Ψ values lay within the allowed regions. For residues in the generously allowed or disallowed regions, the respective electron density maps were carefully investigated to either explain the phenomenon or apply corrections to the model.

A Luzzati plot of the *R*-factor as a function of resolution gives estimates of the average error in the atomic coordinates for a refined structure assuming perfect data (Luzzati, 1952). R.m.s. deviations from ideality of bond length and angles as well as the Luzzati plot were determined with the program *CNS* (Brünger *et al.*, 1998) and for the high-resolution structure with the program *SHELXPRO* (Sheldrick and Schneider, 1997).

3. RESULTS

3.1 Preparation of Cellosyl

Cellosyl from *Streptomyces coelicolor* was expressed and purified over a decade ago at Hoechst AG (Birr *et al.*, 1989; Bräu *et al.*, 1991). The purity of these aged samples was investigated by SDS-PAGE analysis (Fig.3.1-1). Two samples showed a single band exhibiting a molecular weight of 23 kDa and were pure enough for crystallisation trials. One sample contained Cellosyl in a lyophilised state, while the second consisted of a solution of Cellosyl in the FPLC elution buffer.

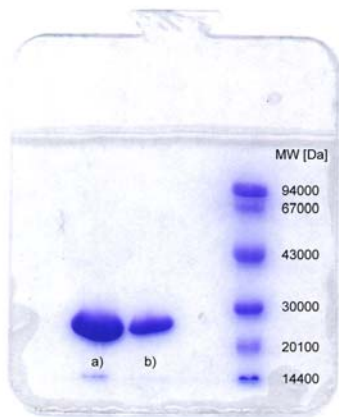


Figure 3.1-2

SDS-PAGE analysis of the original Cellosyl samples
a) Sample of lyophilised Cellosyl
b) Sample of Cellosyl in FPLC elution buffer

3.2 Crystallisation

3.2.1 Hexagonal crystals

Rod-shaped crystals of Cellosyl have been described in the literature. The protein was dissolved in a 10 mM acetate buffer at pH 6.1 and precipitated with a 50% ammonium sulphate solution in the same buffer (Hilgenfeld *et al.*, 1992).

Unfortunately, resolubilising the original lyophilised sample proved to be quite difficult. After several attempts, a buffer solution of 10 mM KH_2PO_4 at pH 6.0 was used with success. In order to sediment insoluble particles, the sample was centrifuged for 5 min at 12000 rpm and the protein dissolved in the supernatant was concentrated to 10 mg/ml at

12000 rpm using a centrifugal filter device. Crystal screening was performed with the Grid Screens from Hampton Research (Laguna Niguel, USA). None of these crystallisation conditions yielded suitable crystals for X-ray analysis.

Since proteins have very complex properties and their aggregation behaviour depends very much on the used buffer conditions (Jancaric *et al.*, 2004), Cellosyl was dialysed into a different buffer solution, containing 20 mM Tris at pH 7.0 and subsequently concentrated to 10 mg/ml. In the new buffer conditions, rod-shaped crystals were obtained at room temperature in a pH range from pH 7.0 to pH 8.0 and an ammonium sulphate concentration between 1.2 - 2.0 M. The crystals actually used for X-ray data collection were prepared by the hanging-drop vapour diffusion method, where hanging drops were comprised of 4 μ l protein solution and 2 μ l reservoir. The latter contained 1.6 M $(\text{NH}_4)_2\text{SO}_4$ in 0.1 M HEPES at pH 7.0. Crystals grew within two weeks and reached a maximum size of 0.25 x 0.25 x 1.2 mm (Fig.3.2-1). A primitive hexagonal Bravais lattice with cell dimensions of $a = b = 124.5 \text{ \AA}$ and $c = 143.1 \text{ \AA}$ was determined by initial X-ray analysis.

3.2.2 Monoclinic crystals

Nearly identical crystallisation conditions were applied to the second sample, which contained Cellosyl in the FPLC elution buffer. Interestingly, a completely different crystal form was obtained. The elution buffer contained a 10mM NaH_2PO_4 solution at pH 7.0 and an unknown concentration of NaCl. The protein concentration was only 2.3 mg/ml because part of the sample had already precipitated from the solution. Prior to crystallisation trials, the protein solution was concentrated to 10 mg/ml. Droplets were prepared by mixing 2 μ l protein and 2 μ l precipitant solution that contained 1.6 M $(\text{NH}_4)_2\text{SO}_4$ in 10 mM HEPES, pH 7.0. Very thin platelets appeared at room temperature after two weeks. Within two months they reached a size of $0.3 \times 0.3 \times 0.04 \text{ mm}$, which was sufficient for diffraction data collection. The crystals belong to the space group C2 with cell dimensions of $a = 111.5 \text{ \AA}$, $b = 38.2 \text{ \AA}$, $c = 51.0 \text{ \AA}$ and $\beta = 108.4^\circ$ (Fig.3.2-1).

Surprisingly, an amorphous-looking precipitate in the original container consisted of crystalline material. Investigation under the light microscope showed the same platelet-like crystals. These crystals diffracted to a maximum resolution of better than 1.0 \AA .

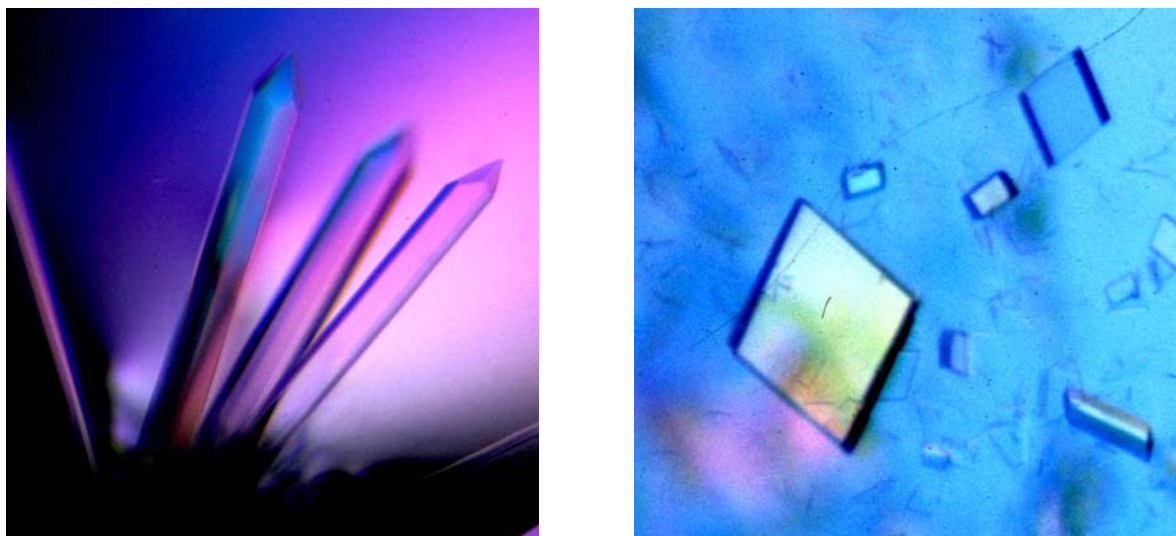


Figure 3.2-2 The two crystal forms of Cellosyl: hexagonal (left) and monoclinic (right)

3.3 Structure elucidation of the monoclinic crystal form

3.3.1 Native data collection

A native data set was collected using in-house facilities with a MAR 300 mm imaging plate detector and $\text{CuK}\alpha$ radiation. Prior to data collection, the crystal was flash-cooled in the liquid-nitrogen stream using highly liquid paraffin oil as cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). A diffraction image of the monoclinic crystal, diffracting to 1.65 Å resolution, is presented in Figure 3.3.1-1.

After autoindexing and integration of the diffraction data, the crystal could be unambiguously assigned to the monoclinic space group $C2$ with cell dimensions $a = 111.15$ Å, $b = 38.22$ Å, $c = 51.04$ Å, $\beta = 108.21^\circ$. The scaling statistics are shown in Table 3.3.1-1. The crystals contained one protein monomer (23 kDa) per asymmetric unit, corresponding to a Matthews coefficient (Matthews, 1968) of 2.18 Å³/Dalton and a solvent content of 43 %.

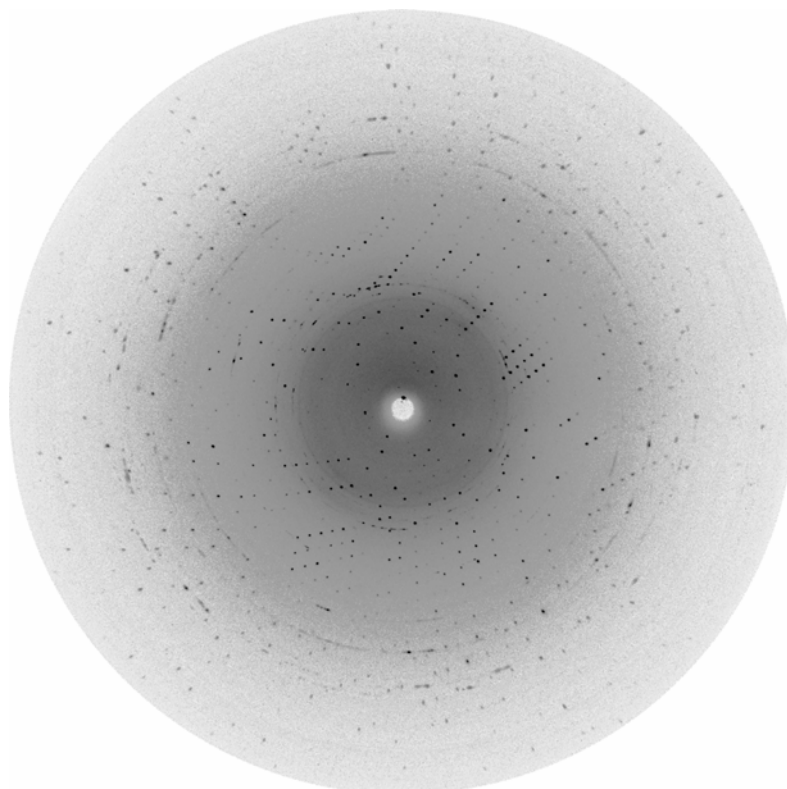


Figure 3.3.1-2 Diffraction image of a monoclinic Cellosyl crystal diffracting to 1.65 Å resolution.

Table 3.3.1-2 Scaling statistics of the monoclinic data set

resolution range [Å]	40-1.65
outermost shell [Å]	{1.71-1.65} ¹
total number of reflections	94865
number of unique reflections	24195
number of rejected reflections	135
completeness [%]	97.2 {92.3} ¹
redundancy	3.93
R _{merge} [%] ²	4.5 {15.7} ¹
R _{r.i.m} [%] ²	5.3 {18.0} ¹
R _{p.i.m} [%] ²	2.7 {9.1} ¹
I / σ	27.7 {8.5} ¹

¹ Values given in {} correspond to those in the outermost shell

² (Weiss and Hilgenfeld, 1997; Weiss, 2001), for formulas see chapter 2.2.8, 21

3.3.2 Initial attempts to solve the structure by molecular replacement

A crystallographic study of a C α backbone model of a bacterial lysozyme from *Streptomyces erythraeus* to a maximum resolution of 2.9 Å was published in 1981 (Harada *et al.*, 1981). The lysozyme from *Streptomyces erythraeus* is comprised of 185 amino acids, exhibiting a molecular weight of 20600 Da, and is therefore very similar in size to Cellosyl. Both enzymes share 49 % sequence identity (Fig. 3.3.2-1), which should result in a very similar fold (Chothia and Lesk, 1986). Therefore, the structure of the lysozyme from *Streptomyces erythraeus* was used as a search model for molecular replacement.

```
1 1 DTSGVQGIDVSHWQGSINWSSVKSGMSEFAYIKATEGTNYKDDRFSAANYTNAYNAGIIRG
2 1 --ATVAGIDVSGHQRNVDWQYWWNQKRFAYVKATEGTGYKNPYFAQQYNGSYNIGMIRG

1 61 AYHFARPNASSGTAQADYFASNGGGWSRDNRTLPGVLDIEHNPSGAMCYGLSTTQMRTWI
2 59 AYHFALPDRSSGAAQANYFVDNNGGGSWKDGKTLPGALDMEYNPYGGTCYGKTPAQMTAWI

1 121 NDFHARYKARTTRDVVIYTTASWNTCTGSWNGMAAKSPFWVAHWGVSAPTVPSPGFPTWT
2 119 KDFSDTYHARTGRWPVIYT--SWWSSCVN-GDFSSTNPLWWVARYASTVGKLEYNWGFHT

1 181 FWQYSATGRVGGVSGDVDRNKFNGSAARLLALANNTA
2 176 IWQYTSS-----PIDQNSFNGGYDRLQALANG--
```

Figure 3.3.2-3 Sequence comparison between Cellosyl (1) and *Streptomyces erythraeus* lysozyme (2)

Molecular replacement was carried out using the *CNS* program package (Brünger *et al.*, 1998). First, a polyalanine model of *Streptomyces erythraeus* lysozyme was prepared and the rotation search was performed in order to determine the correct rotational orientation in the monoclinic cell of Cellosyl. The rotation search was carried out in a resolution range of 15.0 - 4.0 Å using a minimum Patterson vector length of 4 Å. Two different search modes were applied; the direct rotation search and the real space rotation search. Both gave the same 'top' solution with a clear separation from the next peaks, which is a good indication of the right orientation (Table 3.3.2-1).

Table 3.3.2-4 Results of the rotation search

direct rotation search				real space rotation search			
theta 1	theta 2	theta 3	peak [σ]	theta 1	theta 2	theta 3	peak [σ]
332.71	48.46	123.41	8.4	334.60	48.46	125.29	6.0
345.27	48.46	110.85	5.2	347.16	48.46	112.74	4.3
344.07	55.38	119.07	4.7	173.19	45.00	209.19	3.9

The first ten solutions of the rotation function were applied to the search model, followed by analysis of the translation function within a resolution range of 15.0 – 4.0 Å. An optimisation of the rotational orientation of the search model by Patterson-correlation refinement before and after the translation search was incorporated in the program. Unfortunately, an outstanding solution could not be obtained since all translation functions had very similar correlation coefficients (cc) and packing values (pv) and vary only in the y-value. The first four solutions are depicted in Table 3.3.2-2.

Table 3.3.2-5 Rotation and translation function, which gave the highest score using the *CNS* program package.

	theta 1	theta 2	theta 3	x	y	z	cc	pv
1	332.18	49.23	124.99	8.39	1.19	20.38	0.538	0.5687
2	332.20	49.24	125.01	8.41	4.99	20.39	0.538	0.5684
3	332.58	49.42	124.61	8.38	1.08	20.69	0.537	0.5684
4	332.18	49.21	125.00	8.40	-8.44	20.35	0.538	0.5674

For the best solution, rigid-body refinement was carried out to optimise the position of the search model in the cell and resulted in a *R*-factor of 53.3 % and a free *R*-factor of 50.5 %. Subsequent to this, three rounds of conjugate gradient minimisation refinement were carried out, slowly increasing the maximum resolution from 3.5 to 2.3 Å. This lowered the *R*-factor to 48.0 % and the free *R*-factor to 51.1 %. Further simulated annealing and *B*-value refinement did not lead to an improvement. $2F_o-F_c$ and F_o-F_c electron density maps were calculated using phases from the model obtained by molecular replacement. The overall map looked promising, with the density clearly matching the model in some regions. However, in other regions, especially at the N- and C-terminus, obvious differences between the model and the electron density were observed. Unfortunately, rebuilding did not lead to an improvement in

refinement and the *R*-factor did not drop below 50 %. The same procedure was repeated for other solutions of the translation search, however a better solution could not be obtained.

Another attempt to solve the structure by molecular replacement was made using the program *EPMR* (Kissinger *et al.*, 1999). This program applies an evolutionary search algorithm, which allows direct optimisation of the three rotational and translational parameters. The polyalanine model of *Streptomyces erythraeus* lysozyme was again used as search model and the program was run with the observed structure factors for the monoclinic crystal in a resolution range from 15.0 - 3.5 Å. For the correct solution, a correlation coefficient of 0.5 or more can be expected. In this case, the correlation coefficient was as low as 0.15 and the *R*-factor higher than 60 %. Refinement of this solution did not lead to an improvement. Even if *EPMR* did not find the right solution it is interesting to note that the program generated the same solution of the rotation function as the program *CNS* (Table 3.3.2-3).

Table 3.3.2-6 Solution obtained with the program *EPMR*.

theta 1	theta 2	theta 3	x	y	z	cc	R [%]
332.46	48.71	124.53	3.81	47.28	35.15	0.15	60.9

One explanation for the failure of molecular replacement could be the poor quality of the search model. The Ramachandran plot (Ramachandran and Sasisekharan, 1968) of the search model is depicted in Figure 3.3.2-2. Only 50 % of all residues lie in the most favoured regions of the plot, indicating the low quality of the model.

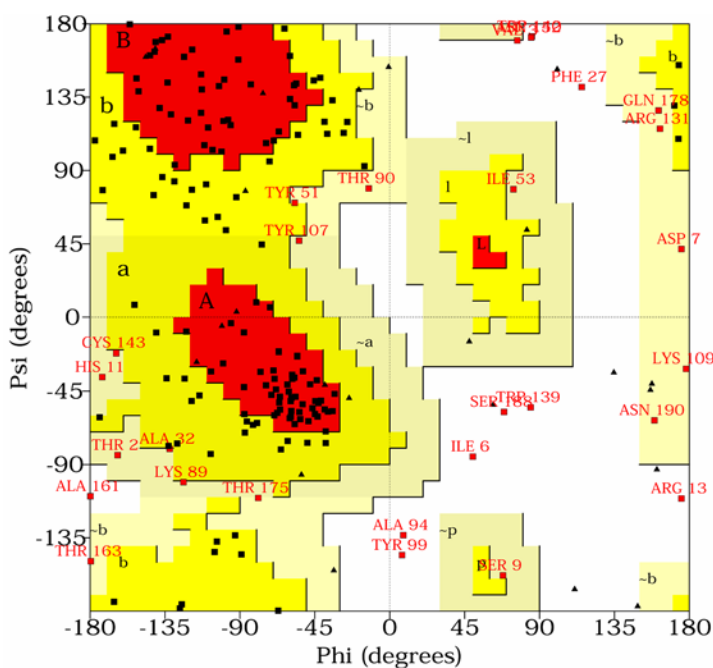


Figure 3.3.2-4 Ramachandran plot of the structure of *Streptomyces erythraeus* lysozyme (Harada *et al.*, 1981) generated with the program *PROCHECK* (Laskowski *et al.*, 1993).

3.3.3 Preparation of heavy-atom derivatives

Heavy-atom derivatives employed in phasing experiments were prepared using mercury(II)- and osmium(III)-chloride. An aqueous solution of the respective heavy-atom salt was added directly to the crystallisation drop containing pre-grown monoclinic crystals. Good results were obtained with a final concentration of 7.5 mM of the heavy-atom compound in the crystallisation drop. An osmium-mercury double derivative was obtained by adding both HgCl_2 and OsCl_3 to the crystallisation drops so that each metal ion was present in a concentration of 7.5 mM.

3.3.4 Initial attempts to solve the structure by MAD

3.3.4.1 Data collection

MAD-data collection was performed on an osmium-derivatised crystal at the synchrotron *ELETTRA* (Trieste). Prior to data collection, a fluorescence scan around the LIII absorption edge of osmium was recorded (Fig.3.3.4.1-1). Based on this scan, three wavelengths were chosen for data collection, according to the criteria described before (see chapter 2.2.8.3, 23). They are listed in Table 3.3.4.1-1.

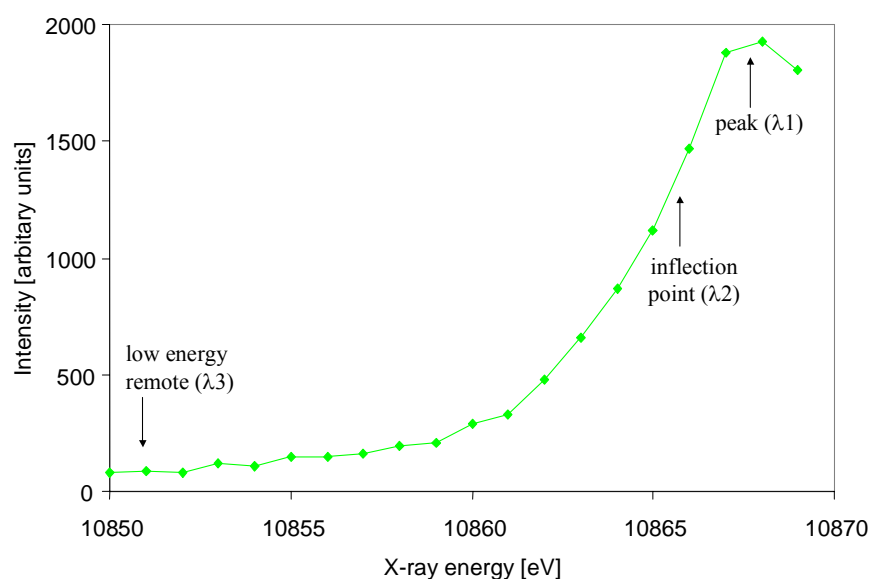


Figure 3.3.4.1-3 X-ray fluorescence scan of the Os-derivatised crystal. The X-ray energies utilized for MAD data collection are marked with arrows.

Table 3.3.4.1-2 Selected wavelengths for the MAD experiment around the LIII edge of osmium.

	energy [keV]	wavelength [Å]
λ1 (peak maximum)	10.8676	1.140868
λ2 (inflection point)	10.8654	1.141099
λ3 (low energy remote)	10.8551	1.142182

All three data sets were measured using the same crystal. In order to minimise the introduction of experimental errors, an identical set-up of the experiment was maintained for the data collection at all three wavelengths. Data were collected at 100 K using highly liquid paraffin oil as a cryoprotectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). The crystal diffracted to a resolution of 1.8 Å and a monoclinic unit cell of $a = 111.14$ Å, $b = 37.75$ Å, $c = 50.99$ Å, $\beta = 108.21^\circ$ was determined. The data collection and merging statistics are depicted in Table 3.3.4.1-2.

Table 3.3.4.1-2 Scaling statistics of the three passes on an Os-derivatised crystal

	λ_1	λ_2	λ_3
wavelength [Å]	1.1409	1.1411	1.1422
crystal-to-detector distance [mm]	200	200	200
resolution range [Å]	99.0-1.80	99.0-1.80	99.0-1.81
total number of reflections	96073	96220	95454
number of unique reflections	18897	18897	18579
number of rejected reflections	160	140	138
completeness [%]	96.9	96.9	97.0
redundancy	5.25	5.26	5.3
R_{merge} [%] ¹	5.8	5.0	4.5
$R_{\text{r.i.m}}$ [%] ¹	6.7	5.8	5.2
$R_{\text{p.i.m}}$ [%] ¹	2.9	2.5	2.3
I / σ	22.2	26.1	31.7

3.3.4.2 Location of the heavy-atom sites and MAD phasing

An anomalous difference Patterson map was calculated using the data set collected at the peak wavelength λ_1 . A plot of the Harker section at $y = 0.5$ is depicted in Figure 3.3.4.2-1a. The dispersive difference Patterson map calculated between the data of λ_2 and λ_3 is shown in Figure 3.3.4.2-1b. Both Patterson maps indicated a single osmium-binding site, which could be positioned at (9.29, 0.0, 9.66) using an automated search method.

¹ (Weiss and Hilgenfeld, 1997; Weiss, 2001), for formulas see chapter 2.2.8, 21

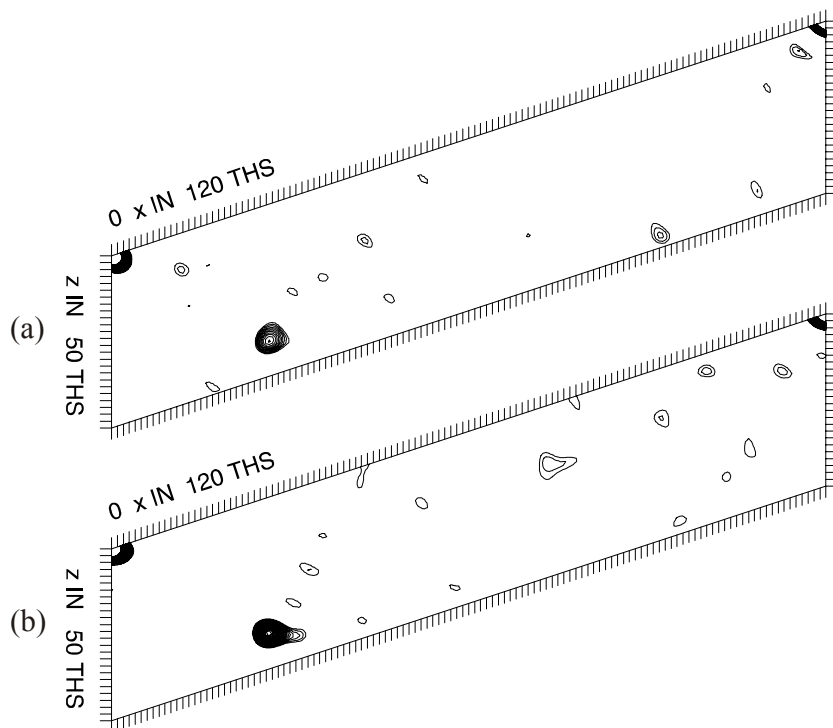


Figure 3.3.4.2-2 Harker section at $y = 0.5$ of the (a) anomalous difference Patterson map (λ_1) and (b) dispersive difference Patterson map ($\lambda_2 - \lambda_3$) at 4 Å.

Since the heavy-atom search is only based on amplitude information, the enantiomeric ambiguity remains to be resolved. Therefore, refinement of the osmium position and MAD phasing was performed with the original osmium configuration and with its inverse image. A distinct difference in the resulting electron density maps should be observed in the ideal case, where only the correct enantiomorph should result in an interpretable map. After phasing, the figure of merit was 0.43 for both hands and the electron density maps were nearly identical. The electron density was reinspected after density modification, including solvent flipping and density truncation (Abrahams and Leslie, 1996), with an estimated solvent content of 43%. Both electron density maps showed clear boundaries between protein and solvent regions but neither of the maps was of sufficient quality to start model building.

3.3.5 Structure elucidation by MIRAS

3.3.5.1 Data collection

The crystal structure of Cellosyl was determined by the method of MIRAS using a native crystal and three different heavy-atom derivatives including a mercury, an osmium and an osmium-mercury double derivative. In addition to the isomorphous difference between the native crystal and the heavy-atom derivatives, the effect of anomalous scattering of osmium and mercury was exploited for phase determination. The diffraction data of the osmium and the osmium-mercury double derivative were collected at 100 K using CuK α -radiation, while the mercury derivative diffraction data were collected using a synchrotron radiation source.

All crystals were assigned to the monoclinic space group C2 with isomorphous cell dimensions of approximately $a = 111 \text{ \AA}$, $b = 38 \text{ \AA}$, $c = 51 \text{ \AA}$, $\beta = 108^\circ$. The largest variation of 1 % was observed along the c axis between the native crystal and the osmium derivative. The data scaling and merging statistics are presented in Table 3.3.5.1-1.

Table 3.3.5.1- 2 Scaling and merging statistics of the heavy atom derivatives.

Data set	OsCl ₃	HgCl ₂	OsCl ₃ +HgCl ₂
Resolution range (Å)	40-1.86	40-1.55	40-1.86
Outermost shell (Å)	{1.97-1.86} ¹	{1.61-1.55} ¹	{1.93-1.86} ¹
Total number of reflections	68985	116778	123237
Number of unique reflections	17231	30038	17434
Number of rejected reflections	101	57	580
Completeness (%)	100.0 {100.0} ¹	100.0 {100.0} ¹	100.0 {100.0} ¹
Redundancy	4.00	3.89	7.07
R _{merge} (%) ²	3.9 {15.6} ¹	4.7 {13.2} ¹	5.1 {24.8} ¹
R _{r.i.m.} (%) ²	4.6 {14.5} ¹	5.5 {15.2} ¹	5.5 {26.9} ¹
R _{p.i.m.} (%) ²	2.3 {7.2} ¹	2.8 {7.7} ¹	2.1 {10.3} ¹
I/ σ	25.2 {5.3} ¹	24.9 {10.5} ¹	27.4 {5.3} ¹

¹ Values given in {} correspond to those in the outermost shell

² (Weiss and Hilgenfeld, 1997; Weiss, 2001), for formulas see chapter 2.2.8, 21

3.3.5.2 Location of the heavy atoms in the unit cell

Patterson maps were calculated and plotted for each derivative using the *CNS* program package (Brünger *et al.*, 1998). The data for the mercury derivative were collected on the high-energy side of the LI absorption edge, at a wavelength of 0.91 Å, thus leading to a high anomalous scattering contribution. This was exploited to calculate anomalous difference Patterson maps, in addition to the isomorphous Patterson maps between the mercury derivative and native data set.

Theoretical plots of f' and f'' show that mercury and osmium absorb X-rays at $\text{CuK}\alpha$ -radiation and therefore, the anomalous scattering contribution of the data sets, collected using in-house facilities, was used for the location of heavy atom sites and phase determination.

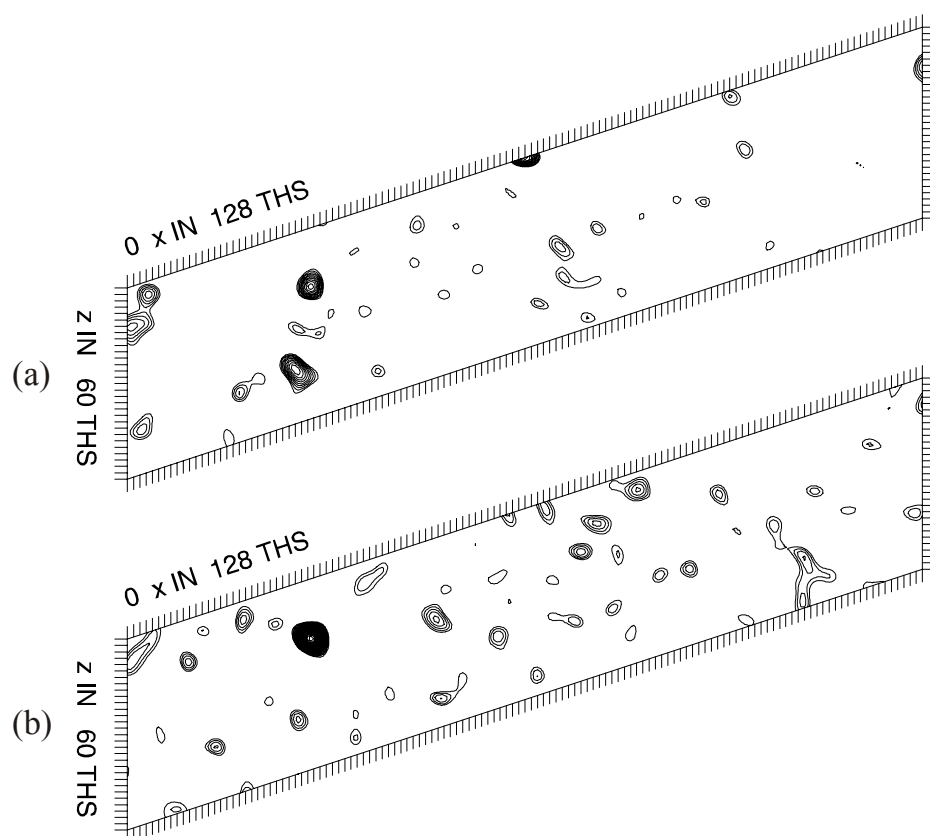


Figure 3.3.5.2-2 Isomorphous (a) and anomalous (b) difference Patterson map for the osmium-mercury double derivative at 3.5 Å.

Isomorphous and anomalous difference Patterson maps of the osmium-mercury double derivative are depicted in Fig. 3.3.5.2-1a) and b). As would be expected from theoretical plots of f' and f'' , the anomalous difference peak for mercury is somewhat higher, since the

anomalous scattering contribution of mercury is higher than for osmium using $\text{CuK}\alpha$ -radiation. It can be also seen that osmium and mercury bind at different locations in the asymmetric unit and therefore both metals provide independent information to phasing.

MIRAS phasing, including the location of the heavy-atom sites, was initially conducted using the automated structure solution program *SOLVE* (Terwilliger and Berendzen, 1999). Unfortunately, by the time the structure was solved, the version of *SOLVE* (version 1.15) could not refine f' and f'' for two different heavy atoms in one derivative but it was able to find two heavy atom sites $> 5\sigma$ for each derivative (Table 3.3.5.2-1).

Table 3.3.5.2-2 Initial heavy atom sites $> 5\sigma$ determined with the program *SOLVE*.

derivative	site	x	y	z	occupancy	B-factor [\AA^2]	height [σ]
osmium + mercury	1	0.615	-0.503	0.077	0.293	15.0	17.8
	2	0.609	0.432	0.186	0.389	41.5	14.5
mercury	1	0.376	0.369	0.424	0.417	50.6	14.2
	2	0.811	0.241	0.297	0.260	30.6	12.2
osmium	1	0.120	0.075	0.136	0.374	15.0	23.0
	2	0.609	0.454	0.182	0.130	15.0	7.0

3.3.5.3 Phase determination

Data collected on the osmium-mercury double derivative were used for the initial refinement of heavy-atom sites and phase determination using the program *SHARP* (De LaFortelle and Bricogne, 1997). The first round of refinement was composed of three big cycles. In the first cycle, scale factors of the derivative data set, together with isomorphous and anomalous lack-of-isomorphism (LOI) and occupancy parameters were refined, as they exert the strongest influence on the maximum-likelihood function. Scaling and LOI parameters of the native data set were excluded from the refinement because the native data set served as the reference data set. In cycles 2 and 3, the coordinates of the heavy-atom positions and isotropic B -values were included in the refinement. After convergence of the first round of refinement, isomorphous and anomalous residual maps (log-likelihood gradient maps) were inspected using the program *O* (Jones *et al.*, 1991) to probe for minor sites. Positive and negative density arranged around the heavy atom sites could be observed, indicating the anisotropic thermal motion of the heavy atoms. Otherwise, the heavy atom model seemed to be complete. Anisotropic B -value refinement together with refinement of the

anomalous scattering factors, f''_{Os} and f''_{Hg} , was performed in the second round of refinement, which converged with a figure of merit of 0.44 and 0.31 for the acentric and centric reflections, respectively. The resulting electron density map was solvent-flattened with the density-modification program *SOLOMON* (CCP4, 1994; Abrahams and Leslie, 1996) using an estimated solvent content of 43 %. Protein and solvent regions could be already distinguished in the solvent-flattened map but model building would have been very cumbersome at this stage.

Therefore, data subsequently collected on a mercury derivative were scaled with the osmium-mercury derivative and the native data set, with the latter serving as reference data set. The two heavy-atom sites determined for each derivative (see Table 3.3.5.2-1) were used in the maximum likelihood refinement to obtain better phases. Refinement of the heavy-atom parameters increased the figure of merit to 0.50 for acentric and to 0.37 for centric reflections. The solvent flattened electron density map showed clear molecular boundaries and looked interpretable but still left room for improvement.

Finally, a third derivative was included in phase determination. It contained a major osmium-binding site with 37% occupancy and a minor site with only 13 % occupancy (Table 3.3.5.2-1) that was initially excluded from the refinement. It was later added since there was clear electron density for it in the residual electron density maps. The last round, including refinement of the anomalous scattering factors f'' at the respective wavelengths and anisotropic B -value refinement, converged with a figure of merit of 0.62/0.54 for acentric/centric reflections. The experimental MIRAS phases extended to 2.0 Å and the electron density map, improved by density modification, was beyond expectations. The phasing statistics are depicted in Table 3.3.5.3-1.

Table 3.3.5.3-2 Phasing statistics for the three heavy atom derivatives

Data set	OsCl ₃	HgCl ₂	OsCl ₃ +HgCl ₂
Number of sites	2	2	2
R_{cullis} (iso): centr /acentr ¹	0.81/0.83	0.91 / 0.93	0.83 / 0.84
R_{cullis} (ano): acentr ¹	0.82	0.87	0.84
Phasing power (iso): centr/acentr ¹	1.23 / 1.49	0.84 / 0.94	0.92 / 1.37
Phasing power (ano) : acentr ¹	1.36	1.30	1.46
FOM centr / acentr ¹	0.54/0.62		

¹ For formulas see chapter 2.2.9.3, 30-31

3.3.6 Model building and refinement

Density-modified phases from *SHARP* were used as input to *ARP/wARP* (Perrakis *et al.*, 1999). The first cycle was performed in mode *warp* where the program built and refined a free-atom model using experimental phases up to 2.0 Å. This model served as input for the next cycle that was run in mode *warpNtrace*. The improved model was again used as input for another round of autobuilding, leading to a model consisting of 11 chains and 131 residues with a connectivity index of 0.81. Using this model together with the native data to 1.65 Å resolution, *ARP/wARP* was able to automatically trace the entire main-chain with the exception of the C-terminal residue, which exhibits poor electron density. Furthermore, *ARP/wARP* automatically placed all but five side chains into the density. Manual building of the remaining poorly fit side-chains was conducted in *O* (Jones *et al.*, 1991), although the majority of necessary adjustments involved side-chain flipping of the terminal χ angle for several Asn, Gln and His residues in order to satisfy hydrogen-bonding requirements. Refinement of the rebuilt model was conducted in *CNS* (Brünger *et al.*, 1998). Both overall *B*-value corrections and bulk solvent corrections were applied. Subsequent simulated annealing, conjugate gradient and atomic *B*-value refinement resulted in an *R*-factor of 27.3 % ($R_{\text{free}} = 30.8\%$).

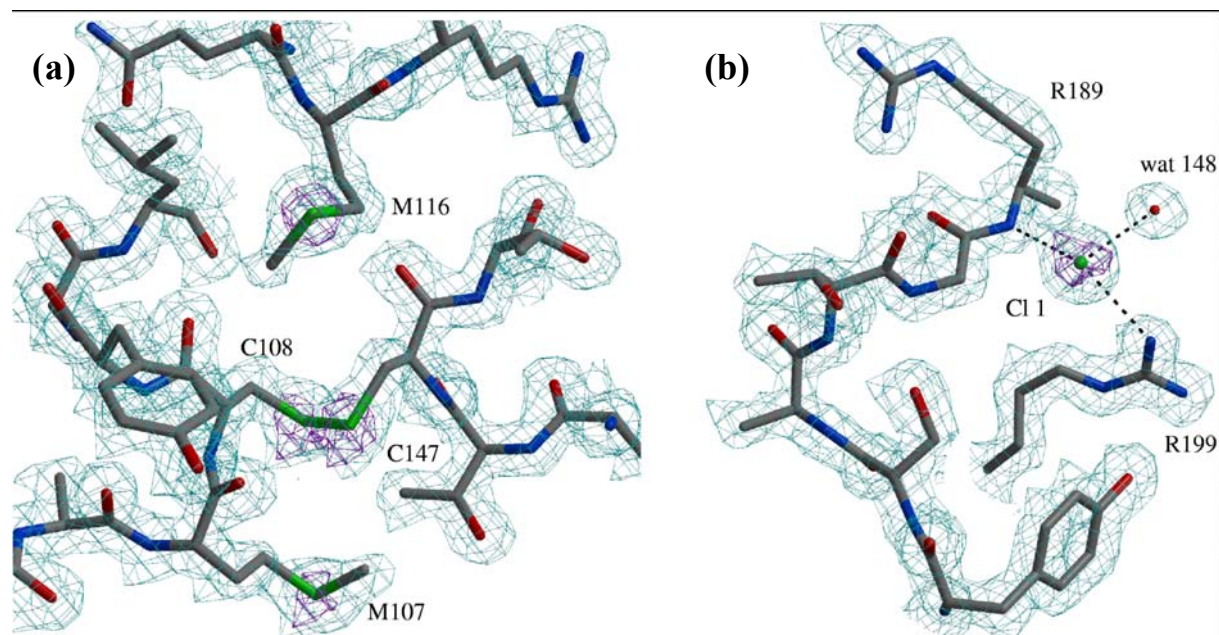


Figure 3.3.6-2 Representative section of the final electron density map with anomalous difference density (contoured 5 σ above the mean) coloured in magenta and the $2F_o - F_c$ map in turquoise (contoured at 1.3 σ); (a) anomalous density for sulphur atoms, (b) anomalous density for a chloride ion clearly distinguishing it from a neighbouring water molecule.

Since the crystals were grown in a buffer containing sodium chloride, anomalous Fourier maps were calculated to search for binding of chloride ions and distinguish them unambiguously from water molecules. Eight chloride ions could be located based on strong anomalous peaks ($> 5 \sigma$ above the mean) present in the primary solvent shell. Further, the positions of the sulphur atoms of the cysteine and methionine residues could be confirmed (Fig. 3.3.6-1).

The addition of 8 chloride ions and of 399 water molecules, along with further refinement, lowered the R -factor to a final value of 15.2 % and R_{free} of 18.4 % for all reflections between 50-1.6 Å. The refinement statistics are shown in Table 3.3.6-1.

Table 3.3.6-2 Refinement statistics of the Cellosyl structure at 1.65 Å resolution

total number of reflections used	24126 (97.4 %)
number of reflections in the working set	22912 (92.5 %)
number of reflections in the test set	1214 (4.9 %)
number of protein atoms	1671
number of water molecules	399
R -factor [%]	15.2
R_{free} [%]	18.4

3.3.7 Assessment of the model quality

There is clear electron density in the final map for all 217 amino-acid residues with the exception of the apparently highly mobile side chain of Arg208 and the carboxy-terminal residue Ala217. Further, the model contains 8 chloride ions and 399 water molecules. More than 20 % of all solvent molecules have B -values less than 15 Å².

At the end of the refinement, the R -factor and R_{free} had converged to 15.2 % and 18.4 %, respectively. The r.m.s. deviations from ideality of bond length and angles of the structure are 0.008 Å and 1.46° respectively, indicating ideal geometry for almost all residues. The overall error in the atomic coordinates based on the Luzzati plot (Luzzati, 1952) was estimated to be less than 0.14 Å.

The geometry of the model was good as assessed by the program *PROCHECK* (Laskowski *et al.*, 1993). Most of the residues (87%) in the polypeptide chain were found within the highly favourable regions and no residues lie in disallowed Ramachandran regions (Fig. 3.3.7-1).

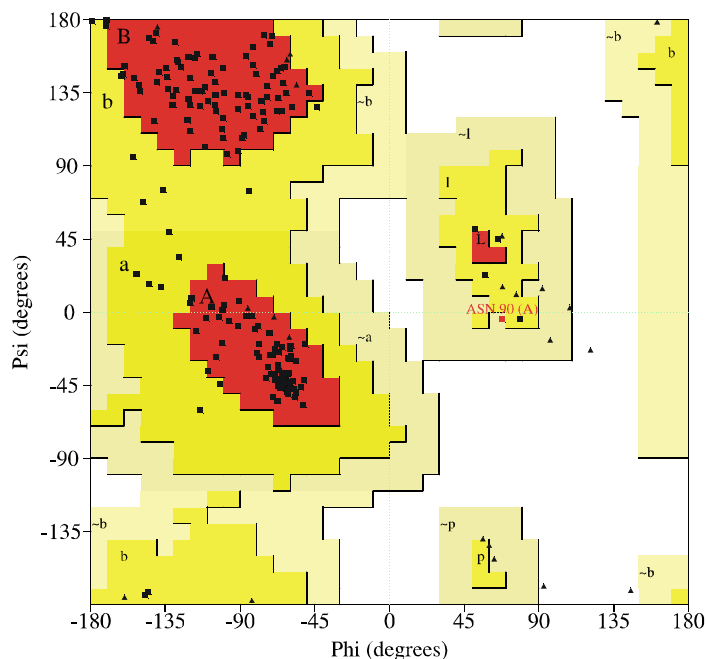


Figure 3.3.7-2 Ramachandran plot of the final structural model of Cellosyl generated with the program *PROCHECK* (Laskowski *et al.*, 1993).

Only Asn90, with Φ and Ψ angles of 67.8° and -4.2° , respectively, is found in the generously allowed regions of the Ramachandran plot, but as can be seen from Fig. 3.3.7-2, it fits the electron density well. According to the program *DSSP* (Kabsch and Sander, 1983), Asp89, Asn90 and Arg91 are located in a bend in the loop between α_3 and β_4 . The strained main-chain conformation of Asn90 is stabilised by a hydrogen-bonding network involving the side chain of Asp89, which is hydrogen bonded to the main chain of Arg91 (Asp89 OD1 – Arg91 N: 2.90 Å) and to the side chain of Thr92 (Asp89 OD1 – Thr92 OG1: 2.78 Å). Asn90 itself forms hydrogen bonds to Asn214 of helix α_6 .

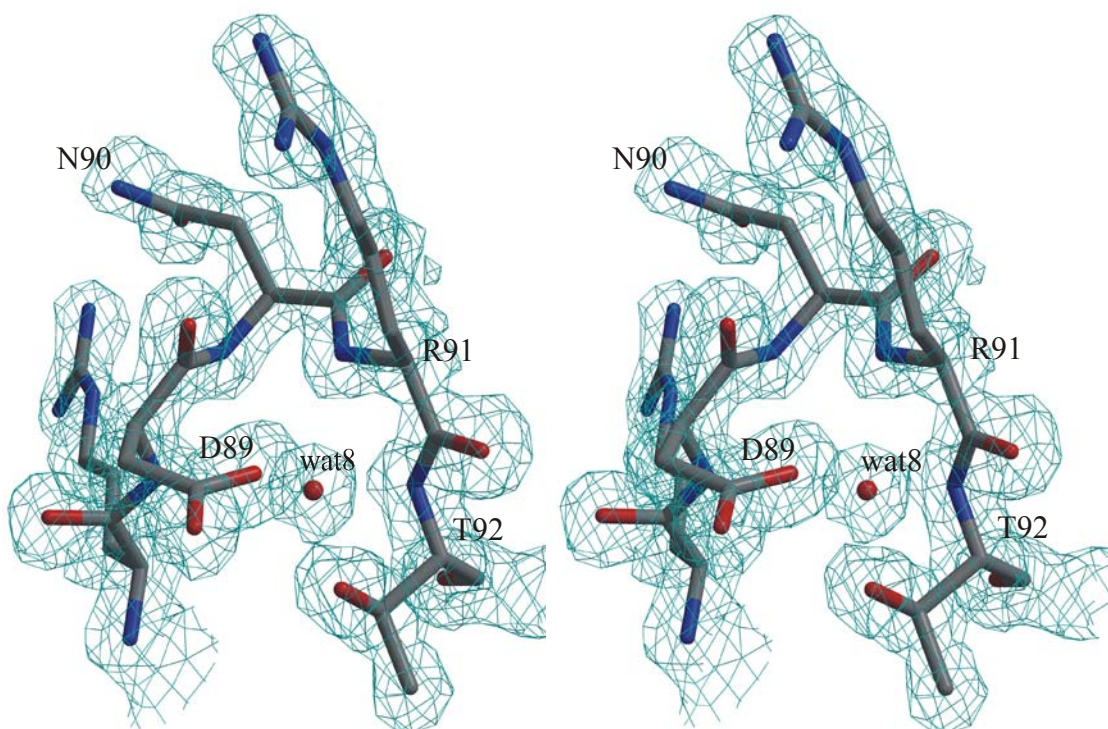


Figure 3.3.7-3 Stereo representation of the bend within the loop between $\alpha 3$ and $\beta 4$. Asn90 in the generously allowed region of the Ramachandran plot fits the density well. The $2F_o - F_c$ electron density map in turquoise is contoured at 1.3σ .

3.4 Atomic-resolution structure

3.4.1 Data acquisition and processing

Three data sets were collected on a single Cellosyl crystal at the multipole wiggler beamline BW7B at the EMBL Hamburg Outstation. The data were recorded at 100 K with an incident wavelength of 0.8445 Å using highly liquid paraffin oil as cryo-protectant (Riboldi-Tunnicliffe and Hilgenfeld, 1999). The first data set was collected to a maximum resolution of 0.82 Å. A second and a third data set were collected at a lower dose in order to avoid oversaturation to a maximum resolution of 1.84 Å and 1.24 Å, respectively.

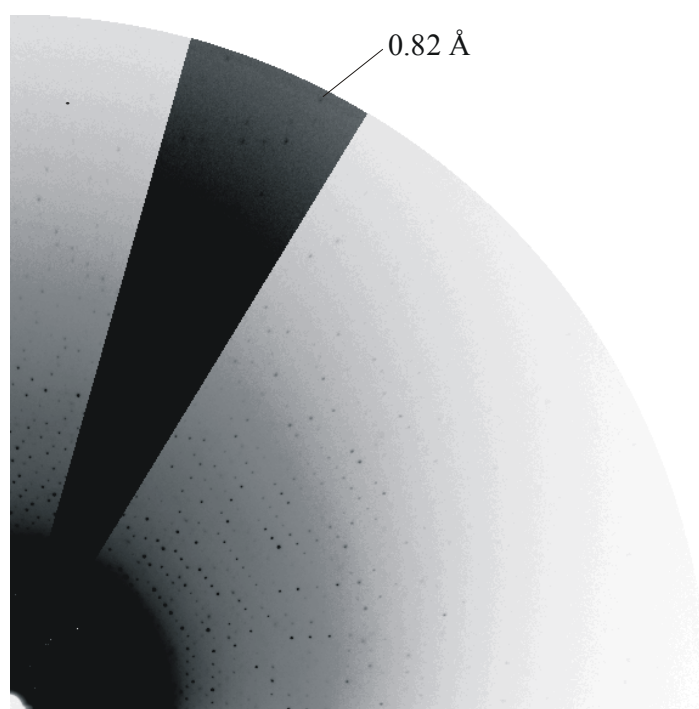


Figure 3.4.1-1 Diffraction image of a Cellosyl crystal. A wedge is contoured darker in order to show that the crystal diffracted to a maximum resolution of 0.82 Å.

Data collection and processing was very challenging (for a detailed description see chapter 2.2.8.5. The crystal belongs to the monoclinic space group $C2$ with unit cell parameters of $a = 111.19$ Å, $b = 38.45$ Å, $c = 50.93$ Å, $\alpha = \gamma = 90.0^\circ$ and $\beta = 108.33^\circ$. The merging statistics are listed in Table 3.4.1-3.

Table 3.4.1-3 Scaling and merging statistics of the three merged data sets from one

Cellosyl crystal.

resolution range [Å]	99.0 – 0.83
total number of reflections	674791
number of unique reflections	176077
number of rejected observations	3051
completeness [%]	91.5 {49.3} ¹
redundancy	3.83 {1.23} ¹
R_{merge} [%] ²	3.5 {29.2} ¹
$R_{\text{r.i.m.}}$ [%] ²	3.8 {41.2} ¹
$R_{\text{p.i.m.}}$ [%] ²	1.6 {29.1} ¹
I/σ	25.7 {2.01} ¹

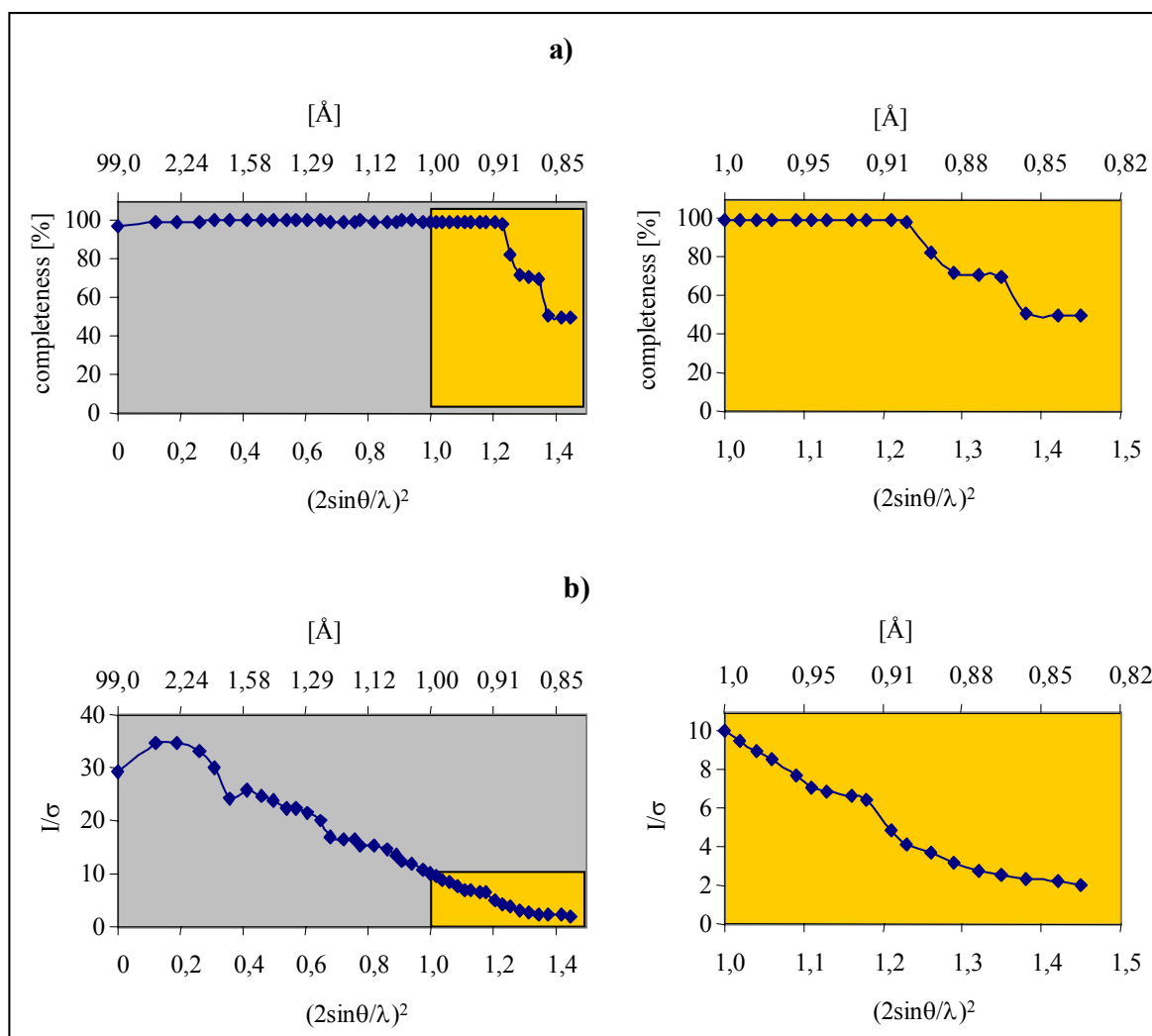


Figure 3.4.1-2 a) Completeness and b) I/σ as depending on the resolution

¹ Values given in {} correspond to those in the outermost shell (0.84 – 0.83 Å)

² (Weiss and Hilgenfeld, 1997; Weiss, 2001), for formulas see chapter 2.2.8, 21

3.4.2 Refinement of the high-resolution structure

The refinement of the high-resolution structure was performed with the program *SHELX-97*, which is best suited for the refinement of proteins at high resolution (Sheldrick and Schneider, 1997). The implemented least-squares refinement algorithm was originally designed for the refinement of small molecules and has been adapted to macromolecular structures, providing the opportunity to model atomic anisotropy, complicated disorder and twinning and to obtain a least-squares estimation of parameter errors.

The structure of Cellosyl (Rau *et al.*, 2001; PDB code: 1JFX) derived from an isomorphous crystal, previously refined at 1.65 Å resolution to an *R*-factor of 15.2 % ($R_{\text{free}} = 18.5$ %) (see chapter 3.3.5 and 3.3.6), was used as a starting model for *SHELXL* refinement. Each round of refinement consisted of 20 conjugate-gradient cycles. The first round was carried out over a resolution range from 40.0 – 1.0 Å using isotropic displacement parameters and converged with an *R* of 24.5 %. The increase in the *R*-factor by more than 9 % compared to the low-resolution model can be explained by a slight difference in the unit cell parameters of the two different crystals used and further by the omission of all solvent molecules from the model.

After each refinement cycle, SigmaA maps ($2mF_o - Df_c$ and $mF_o - DF_c$) were generated and inspected for model building and verification. The strongest Fourier peaks in the SigmaA difference map guided the building of well-defined water molecules and chloride ions. The latter had been unambiguously identified in the low-resolution model by their anomalous scattering contribution at CuK_α radiation. The building of 39 water molecules and 3 chloride ions in the second round of refinement led to a drop of the *R* and R_{free} of more than 3 % to 20.6 % and 21.3 %, respectively.

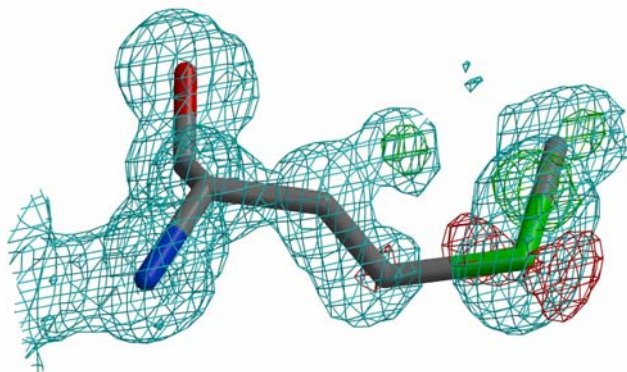
In the consecutive rounds of refinement, more water molecules and the remaining 5 chloride ions, also present in the low resolution model, were added which made the electron density maps progressively clearer and allowed further identification of solvent molecules and the improvement of the protein model. Since the 1.65 Å structure was build with high accuracy, only minor changes had to be applied to the protein model, e.g. flipping of the χ_2 angle of His12 and fitting the O_γ atom of Ser185 into the electron density. As in the low-resolution model, the side chain of Arg208 and the C-terminal residue Ala217 remained ill defined. A detailed inspection of the model showed that alternative side-chain conformations of the protein slowly became distinguishable, particularly a second side-chain conformation of Met107 was clearly evident at this stage (Fig. 3.4.2-1a). However, for other side-chains

such as Ser142 the electron density was not of sufficient quality to model alternative side-chain conformations. Therefore, the occupancy of these side chains was reduced to 60 %, which lead to a large improvement of the difference Fourier maps in the respective regions after the subsequent round of refinement and simplified the building of alternative side-chain conformations (Fig. 3.4.2-2). This strategy was subsequently used at other sites of the protein.

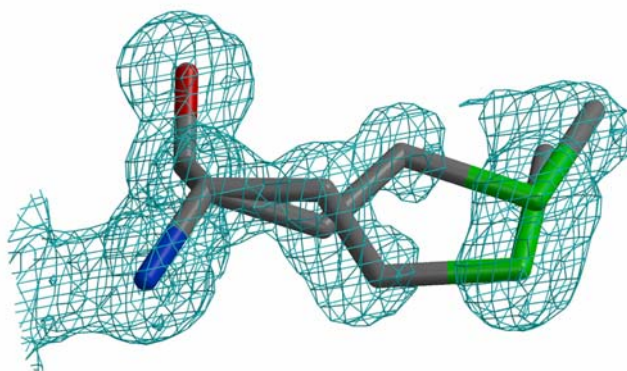
In the fifth round of refinement, the resolution limit was extended to 0.83 Å and all high-resolution reflections included in the refinement. The second conformation of Met107 was modelled and refined with complementary occupancies (Fig.3.4.2-1b,c). The resulting R and R_{free} were 17.31 and 17.79 %, respectively. In round six, alternative side-chain conformations of three more residues (Ser20, Arg88, Asn152) were introduced. At this stage of refinement, the difference Fourier maps revealed numerous patches of positive as well as negative density close to heavier atoms, i.e. Cl⁻ ions and S atoms (Fig.3.4.2-3). This distribution of electron density indicated that isotropic treatment of the atomic displacement parameters was no longer adequate at this resolution. Therefore, anisotropic displacement parameters were introduced for the chloride ions in round 7 of the refinement and for the whole protein and all solvent molecules in the subsequent run. This lowered the R and R_{free} by more than 4 % to a value of 12.51 % and 13.89 %, respectively. Alternating sessions of model building and refinement were carried out in order to facilitate electron density interpretation and improve the model.

Met107

(a)



(b)



(c)

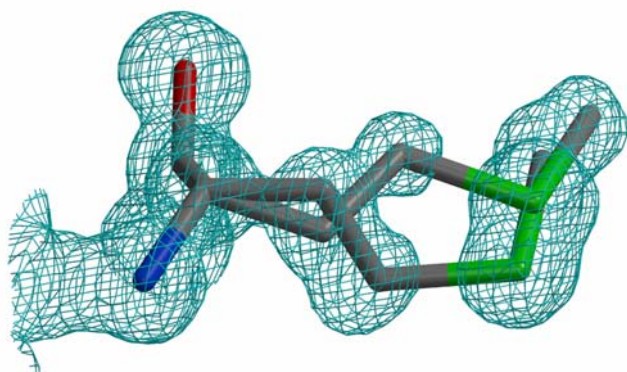
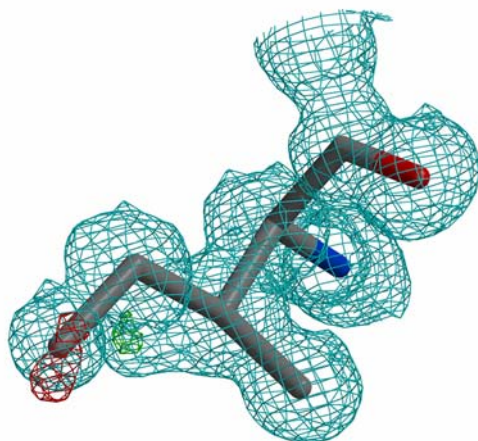


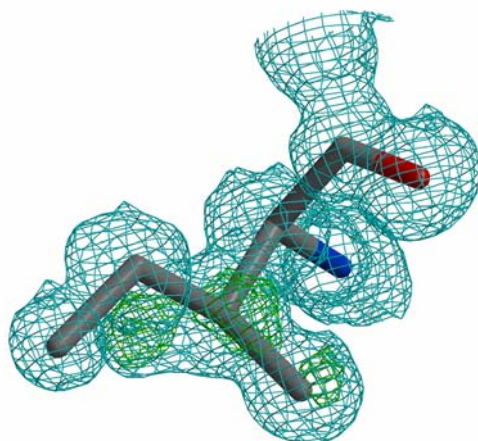
Figure 3.4.2-1 The electron density of Met 107, at different stages of refinement, is illustrated as a $2mF_o-DF_c$ map (turquoise, contoured at 1.0σ) and a mF_o-DF_c map (green, 4.0σ and red -3.0σ): (a) single conformation surrounded by difference density indicating a second side-chain conformation, (b) after modelling and isotropic refinement of the alternative side-chain conformation and (c) after refinement of anisotropic displacement parameters.

Ile99

(a)



(b)



(c)

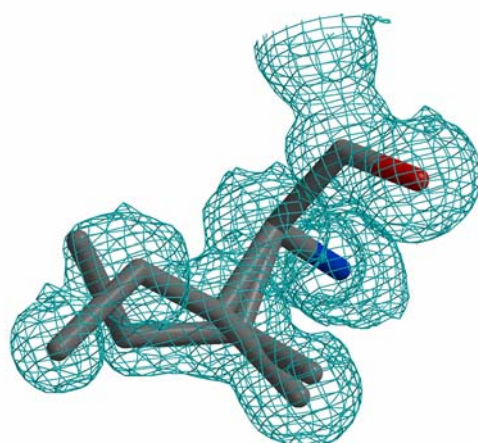


Figure 3.4.2-2 Improvement of the electron density map of Ile99 during refinement. Electron density maps are $2mF_o-DF_c$, coloured in turquoise (1.0σ), and mF_o-DF_c , in green (4.0σ) and red (-3.0σ): (a) patches of positive and negative difference density around the side-chain at 100 % occupancy (b) difference density showing up much clearer after side-chain was set to 60 % occupancy (c) double conformation built and refined to unity.

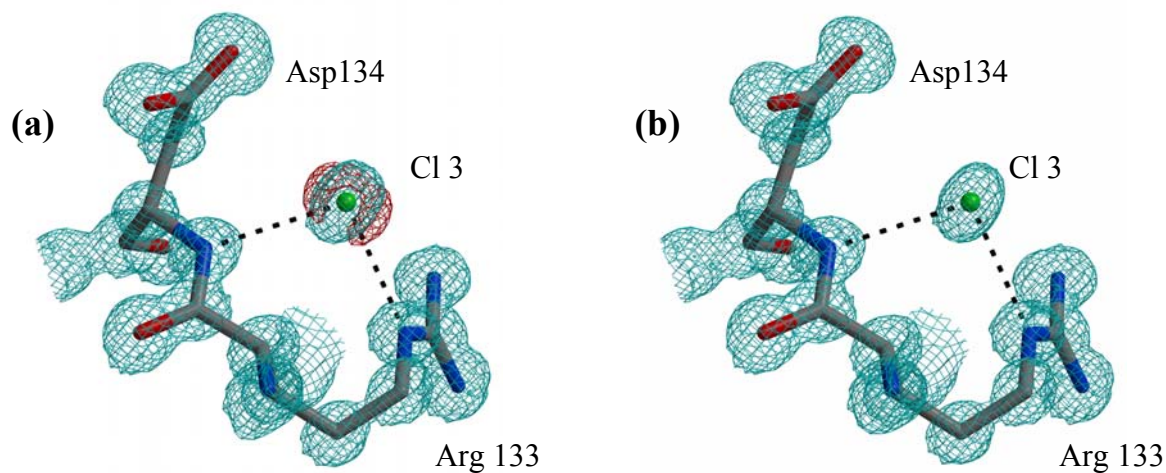


Figure 3.4.2-3 Example of a chloride ion (a) before and (b) after anisotropic refinement. The $2mF_o-DF_c$ map is coloured in turquoise and contoured at 1.3σ and the mF_o-DF_c in red contoured at -3.0σ .

After 11 rounds of refinement, the auxiliary program *SHELXWAT* was employed for automatic water building. The program, a simplified version of the *ARP/wARP* procedure (Perakakis *et al.*, 1999), searches for potential water molecules by iterative least-squares refinement, difference electron density calculation and rejection of waters with high displacement parameters (*SHELXL* Workshop Manual, 2000). 152 water molecules were automatically built in 10 iterative rounds and subsequently inspected manually. The occupancy of several water molecules was manually lowered to 50 % and water molecules were removed from the model if the respective $2mF_o-DF_c$ density was less than 0.7σ .

In the subsequent rounds of refinement, more alternative side-chain conformations were built following the above-described routine. For residue Ser24, the electron density could only be interpreted by modelling a triple conformation of the side chain. The occupancy of the respective side-chain conformations was estimated from the peak height of the $2mF_o-DF_c$ electron density map. Further, double main-chain conformations for 5 residues, i.e. Ser104, Ala141, Ser142, Ala156 and Lys157, were modelled and refined with complementary occupancies.

In round 30, hydrogen atoms were added, which lowered the R and R_{free} by 1.03 and 1.6 % to 9.72 and 10.53 %, respectively. The huge drop in R_{free} underlined the significance of

hydrogen contribution in the refinement of proteins at atomic resolution, which was also manifested in a noticeable improvement of the quality of the electron density maps. Thus, minor adjustments were made to the model and more solvent atoms were added. The refinement was terminated with an R -factor of 9.10 % and an R_{free} of 9.67 % for data with $F_o > 4\sigma$ and an R -factor of 9.63 % and an R_{free} of 10.38 % for all data. Although, there was still difference density over 5σ which could be assigned to the solvent shell, it did not seem meaningful to build more half-occupied water molecules. A schematic representation of the course of refinement is depicted in Fig. 3.4.2-4 and the refinement statistics are summarised in Table 3.4.2-1.

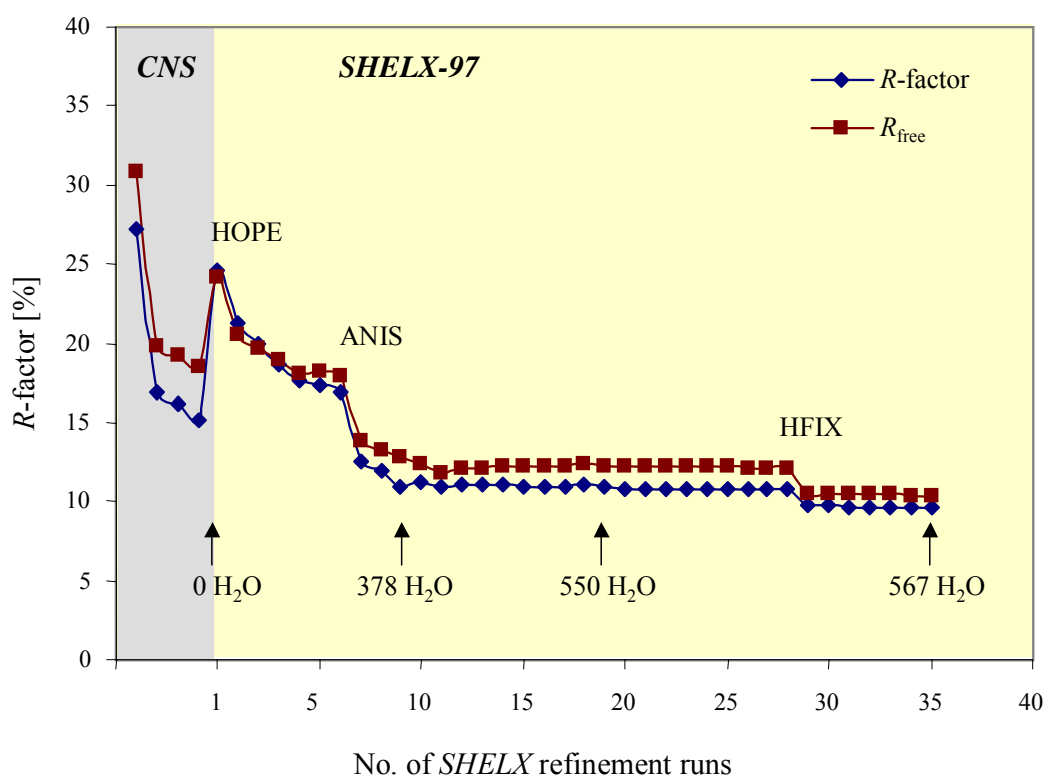


Figure 3.4.2-4 Schematic representation of the course of refinement of the high-resolution structure of Cellosyl.

Table 3.4.2-1 Refinement statistics of the high-resolution structure

resolution range [Å]	40 - 0.83
total number of reflections used	174310
number of reflections in the test set	1767
number of non-H protein atoms	1799
number of chloride ions	8
number of water molecules	567
<i>R</i> -factor [%]; $F_o > 4\sigma$ / all data	9.10 / 9.63
<i>R</i> _{free} [%]; $F_o > 4\sigma$ / all data	9.67 / 10.38

3.4.3 Validation of the model quality

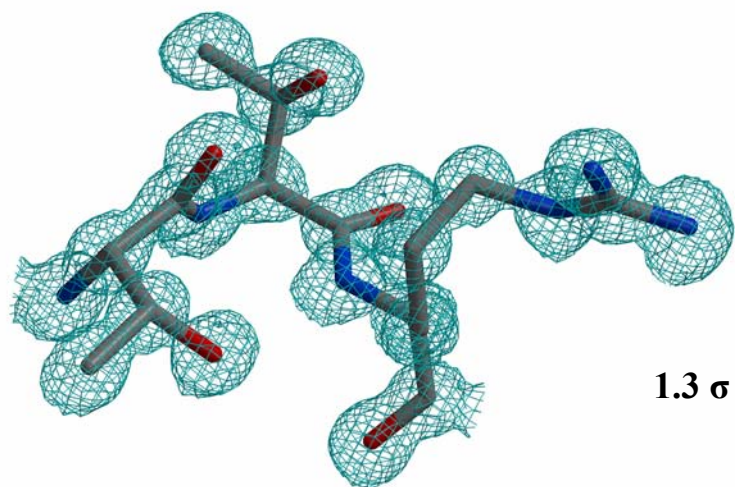
The three-dimensional structure of Cellosyl has been refined with high accuracy at 0.83 Å resolution to an *R*-factor < 10 % (see Table 3.4.2-1). For data with $F_o > 4\sigma(F_o)$, the *R* and *R*_{free} are as low as 9.10 % and 9.67 %, respectively.

The final refined protein model consists of 1799 non-H protein atoms from 216 amino acid residues. 28 amino acid residues were modelled in double or triple side-chain conformations and 5 residues possess a double main-chain conformation. The model is further comprised of 8 chloride ions (also present in the 1.65 Å structure) and 567 water molecules of which 364 are fully occupied.

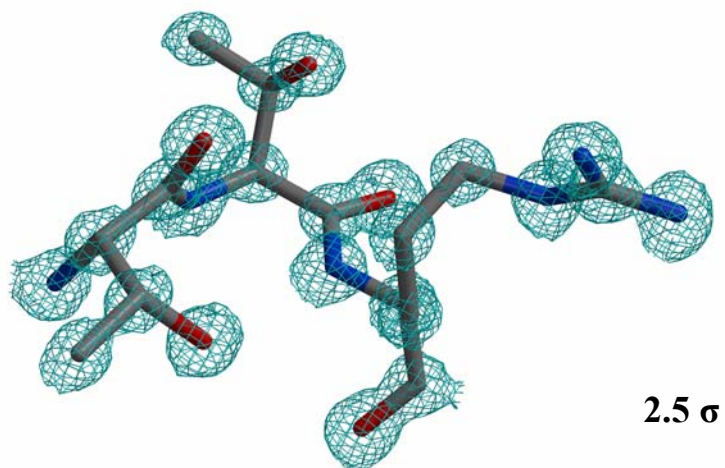
The refinement resulted in electron density maps of excellent quality with an r.m.s. deviation value of 0.07 eÅ⁻³ for the electron density synthesis with F_o - F_c coefficients. The $2mF_o$ - DF_c electron density is well defined for the whole protein model with exception of the C-terminal residue Ala217, which was omitted from the model. Continuous electron density can be observed for nearly all protein atoms above 1.5 σ level, including the residues adapting alternative conformations. Even at the 4 σ level, $2mF_o$ - DF_c density can be observed for the majority of the main-chain atoms (for example see Fig. 3.4.3-1c).

Thr132

(a)



(b)



(c)

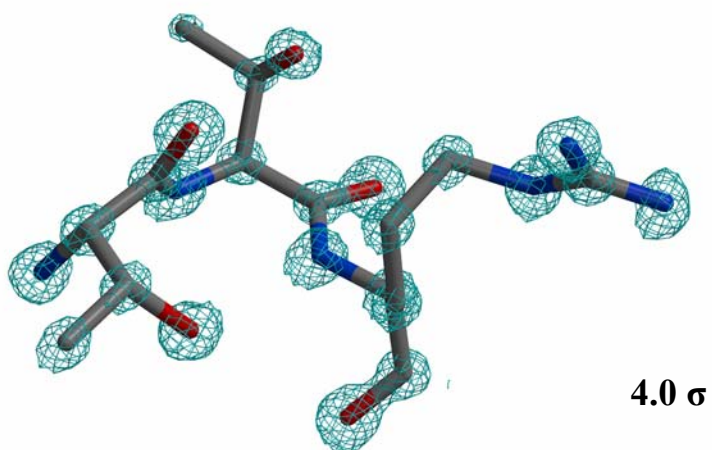


Figure 3.4.3-1 Representation of the final electron density map for residue Thr132, which is one of the Ramachandran plot outliers. The $2mF_o-DF_c$ electron density map is contoured at (a) 1.3 σ , (b) 2.5 σ and (c) 4.0 σ .

Besides the good fit of the model to the electron density maps, the quality of the model was assessed from the Ramachandran plot (Ramachandran and Sasisekharan, 1968), which is depicted in see Fig. 3.4.3-2.

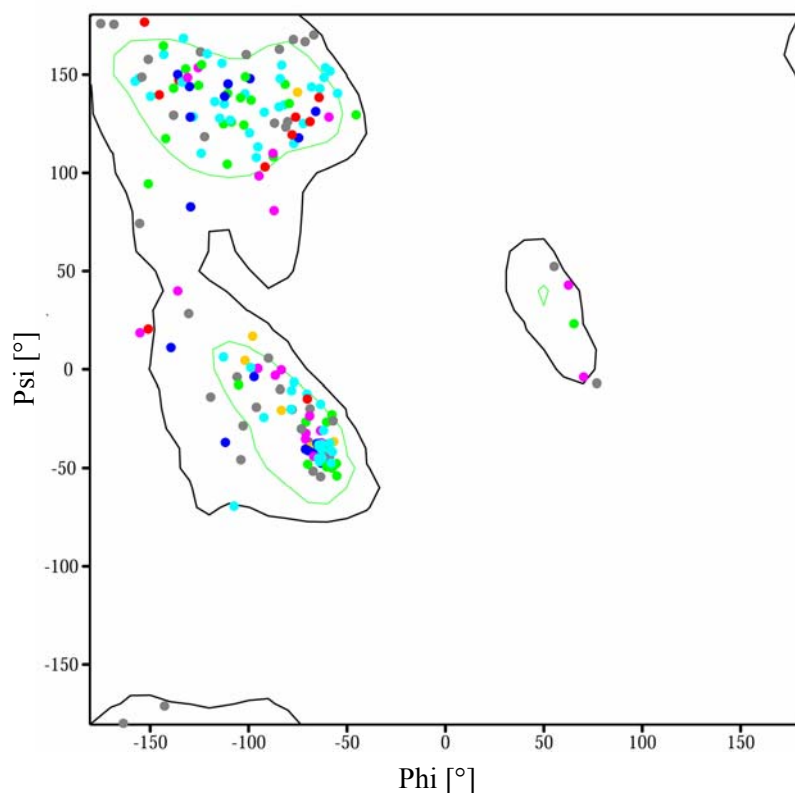


Figure 3.4.3-2 Ramachandran plot of the Cellosyl structure at 0.83 Å resolution prepared with the program *SHELXPRO*. The colours are coded according to residue characteristics: yellow = Cys, Met; green = Phe, Tyr, Trp, His; cyan = Ala, Leu, Ile, Val, Pro; red = Glu, Asp; blue = Arg, Lys; purple = Gln, Asn; gray = Ser, Thr.

Of the 191 standard residues (Gly were excluded), 153 (80.1 %) are located within the inner core region and 185 (96.9 %) are within the core region. Six Ramachandran outliers were detected, namely Ala34, Asn39, Asp89, Asn90, Ser104 and Thr132. According to the program *DSSP* (Kabsch and Sander, 1983), all these residues are located at special positions in loops, with exception of Ala34, which is located in strand β 2. For all six of these residues the electron density is very well defined. An example is depicted in Fig. 3.4.3-1, showing Thr132 located in a bend within the loop between α 4 and β 5. It can be clearly seen that even at the 4σ level, all atoms perfectly match the electron density.

The r.m.s. deviation of the model from ideality is 0.015 Å for bond distances and 0.031 Å for bond-angle distances.

A Luzzati plot for the final model is presented in Fig. 3.4.3-3, with the R -factor [$F_0 > 4\sigma(F_0)$] represented as a function of the resolution. The plot gives estimates of the average error in the atomic coordinates for the refined Cellosyl structure assuming perfect data (Luzzati, 1952).

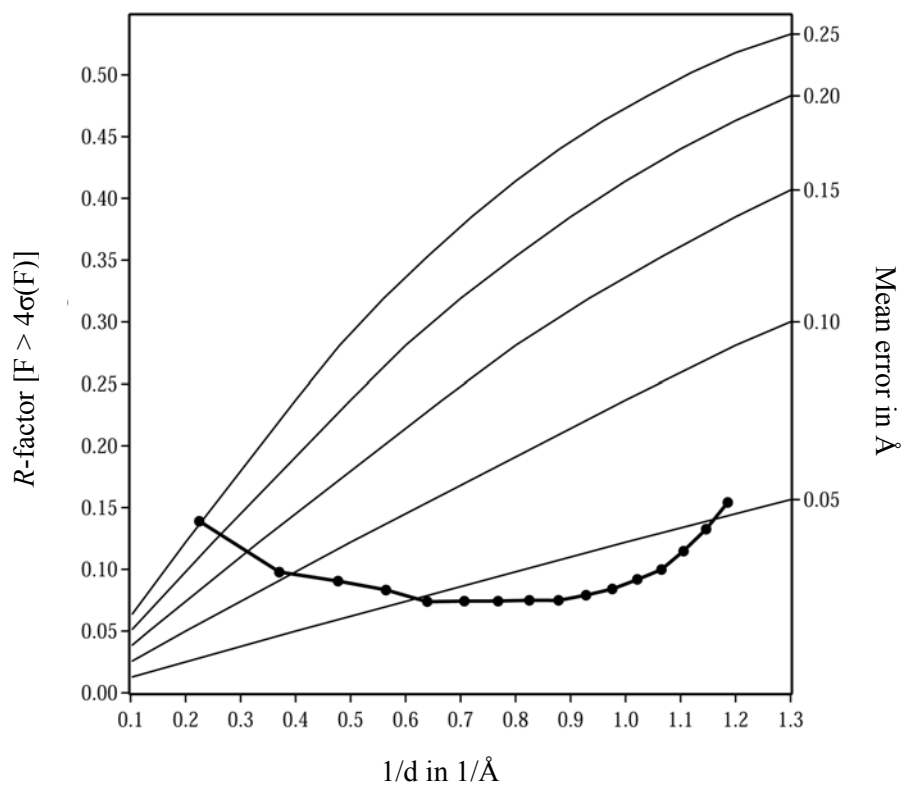


Figure 3.4.3-3 Luzzati plot (Luzzati, 1952) of the final model of Cellosyl at 0.83 Å resolution.

3.5 Structure elucidation of the hexagonal crystal form

3.5.1 Data collection and space group determination

X-ray data were collected on a flash-frozen crystal at the beamline BM30 of the ESRF (Grenoble). Using an incident wavelength of 0.9797 Å and a crystal-to-detector distance of 330 mm, the crystal diffracted to a maximum resolution of 2.3 Å.

In order to distinguish between a trigonal and a hexagonal crystal system, both of them have the same cell geometry of $a = b$, $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$, data were scaled in space group P1. Pictures of the hkl planes along the l axis (see Fig. 3.5.1-1) were carefully inspected: Hexagonal symmetry is clearly visible along the l axis, therefore a trigonal crystal system could be excluded. This was confirmed by analysing the systematic absences. Only every sixth reflection along the screw axis could be measured. This reflection condition is valid for the space groups P6₁, P6₅, P6₁22 and P6₅22. It is impossible to distinguish between the space groups P6₁ and P6₅ or between P6₁22 and P6₅22 by just merging and scaling the intensities, therefore data were scaled in the space groups P6₁ and P6₁22 and the χ^2 (goodness of fit) values were compared (Table 3.5.1-1). The χ^2 was about the same for P6₁ and P6₁22, indicating that the crystals belong to space group P6₁22 or P6₅22.

Table 3.5.1-2 Scaling statistics of space groups P6₁ / P6₅ and P6₁22 / P6₅22

	P6 ₁ / P6 ₅	P6 ₁ 22 / P6 ₅ 22
resolution range [Å]	99.0 - 2.32	99.0 - 2.32
total number of reflections	412060	412271
number of unique reflections	54011	28813
number of rejected reflections	788	824
completeness	99.1	99.2
redundancy	7.6	14.3
χ^2	1.033	1.028
R_{merge} [%] ¹	5.4	5.5
$R_{\text{r.i.m}}$ [%] ¹	7.4	7.4
$R_{\text{p.i.m}}$ [%] ¹	2.5	1.8
I / σ	19.2	26.1

¹ (Weiss and Hilgenfeld, 1997; Weiss, 2001), for formulas see chapter 2.2.8, 21

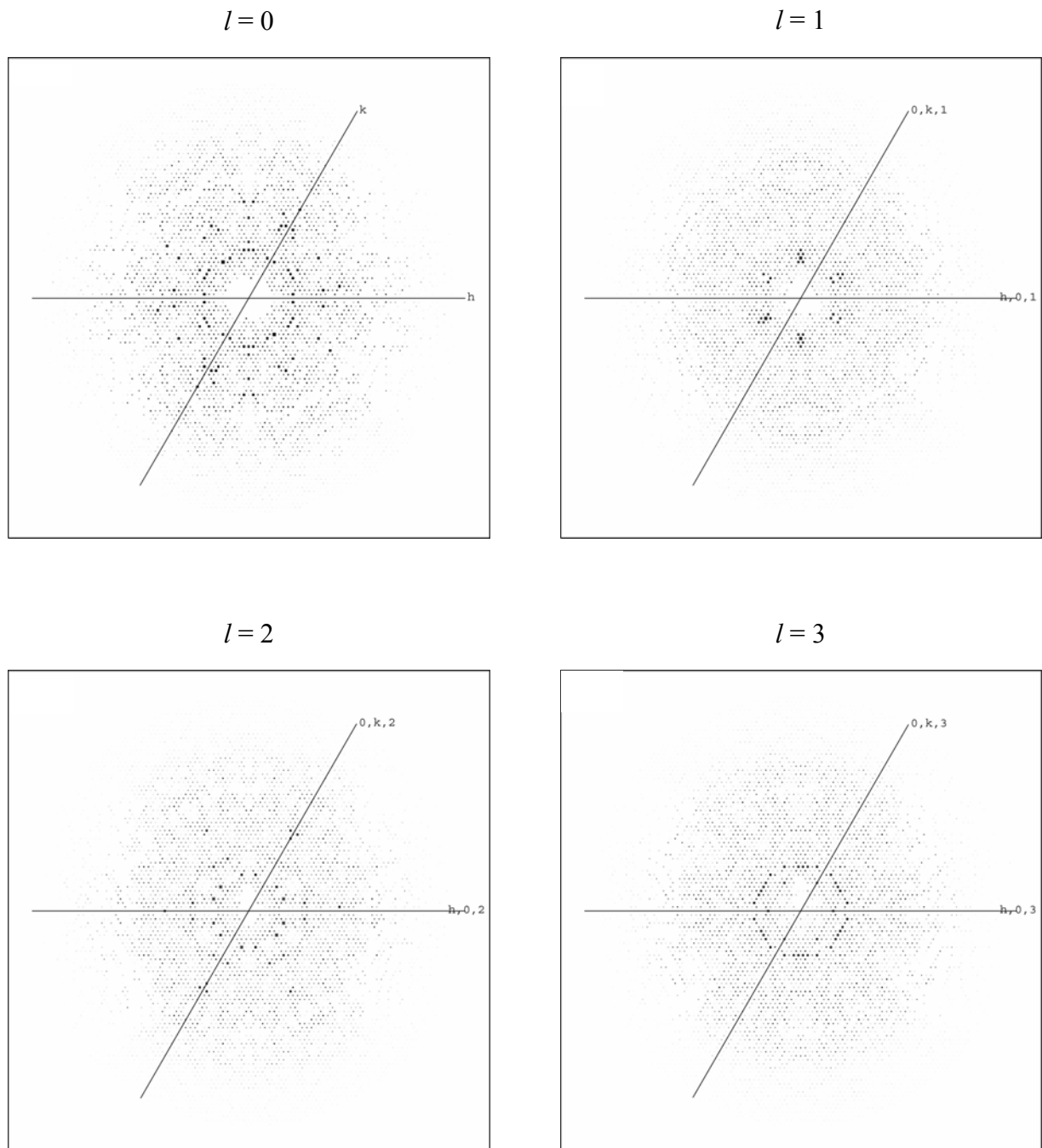


Figure 3.5.1-2 The hkl planes of the hexagonal crystal at $l = 0, 1, 2$ and 3 created with the program *HKLVIEW* (CCP4, 1994).

This was supported by analysis of $R_{p.i.m.}$ (Weiss and Hilgenfeld, 1997; Weiss 2001) for both space groups. The observable redundancy in space group $P6_122$ is by a factor of 2 higher than in $P6_1$. Therefore, the $R_{p.i.m.}$ for $P6_122$ should be smaller than for $P6_1$ by the square root of 2, if only statistical deviations are taken into account. The value of $R_{r.i.m.}$ should be the same, which can be clearly seen in Table 3.5.1-1. In this case, the space group with the higher symmetry is very likely the correct one but unambiguous evidence can only be given by the refined structure.

3.5.2 Structure solution by molecular replacement

Structure solution of Cellosyl in the hexagonal crystal form was approached by the method of molecular replacement. Searches were performed in the space groups $P6_122$ and $P6_522$. As a prerequisite, the number of molecules in the asymmetric unit had to be determined. Therefore, Matthews coefficients (Matthews, 1968) and the estimated solvent content for one to four molecules in the asymmetric unit were calculated and are listed in Table 3.5.2-1.

Table 3.5.2-3 Matthews coefficient and solvent content of the hexagonal crystal in the space groups $P6_122$ and $P6_522$.

molecules / au	Matthews coefficient [$\text{\AA}^3/\text{Da}$]	estimated solvent content [%]
1	6.78	81.88
2	3.39	63.75
3	2.26	45.63
4	1.70	27.50

Since the distribution range of the solvent content of protein crystals is rather large (Kantardjieff and Rupp, 2003), the molecular replacement search was carried out for one, two and three molecules in the asymmetric unit for both space groups employing the 1.65 \AA structure of Cellosyl as a search model. Beforehand all solvent molecules were omitted from the model and the molecular replacement was performed in a resolution range from 40.0 – 3.5 \AA using the program *EPMR* (Kissinger *et al.*, 1999). The results are summarised in Table 3.5.2-2. Assessing correlation coefficient (cc) and *R*-factor, it is evident that the best result was obtained in spacegroup $P6_122$ with two molecules in the asymmetric unit.

Table 3.5.2-4 Correlation coefficients (cc) and *R*-factors of the molecular replacement solutions in space groups P6₁22 and P6₅22. The best result is highlighted in grey.

molecules / asymmetric unit	P6 ₁ 22		P6 ₅ 22	
	cc	<i>R</i> -factor [%]	cc	<i>R</i> -factor [%]
1	0.263	57.0	0.153	63.2
2	0.446	48.9	0.175	62.8
3	0.376	53.4	0.193	62.7

3.5.3 Refinement

Rigid-body refinement followed by simulated annealing, conjugate-gradient and atomic *B*-value refinement was performed on the crude hexagonal model with the *CNS* program package (Brünger *et al.*, 1998) and resulted in a surprisingly low *R*-factor of 25.39 % and an *R*_{free} of 28.02 %. The calculated 2F_o-F_c electron density map fitted the model well with the exception of several side chains, which had to be manually placed in the electron density. The map was also of sufficient quality to initialise the building of solvent molecules. Unexpectedly, the addition of waters led to a slight increase of the *R*-factor as well as the *R*_{free} of about 1.5 % and 0.5 %, respectively. This phenomenon was very puzzling since all built side chains and water molecules fitted the electron density perfectly. The addition of more waters increased the *R*-factor even more.

This problem was circumvented by conjugate-gradient minimisation of the crude hexagonal model omitting experimental energy terms in order to reduce phase bias introduced by the search model. Subsequent refinement resulted in an *R*-factor and *R*_{free} of 35.33 % and 46.10 % respectively, both noticeably higher than for the model described above. For a second time, side-chains were manually modelled into the electron density. As before, the primary solvent shell was clearly evident in the 2F_o-F_c as well as the F_o-F_c electron density maps. 142 water molecules could be modelled on the basis of distance criteria, electron density and shape using the automatic solvent building procedure of *ARP/wARP* (Perrakis *et al.*, 1999) which lowered the *R*-factor and *R*_{free} considerably to 22.0 % and 27.9 %. Inspection of the F_o-F_c electron density maps revealed that large peaks of positive difference density covered several water molecules. These water molecules were subsequently omitted from the structure and the electron density at the respective positions was thoroughly inspected after the next round of refinement. Instead of water molecules, 20 sulphate ions were added to the model, based on the peak height of the F_o-F_c difference peaks, the shape of the difference

density and the particular environment (Fig. 3.5.3-1). The occupancy of the sulphate ions was adjusted in the following rounds of refinement based on the peak height of the $2F_o - F_c$ electron density.

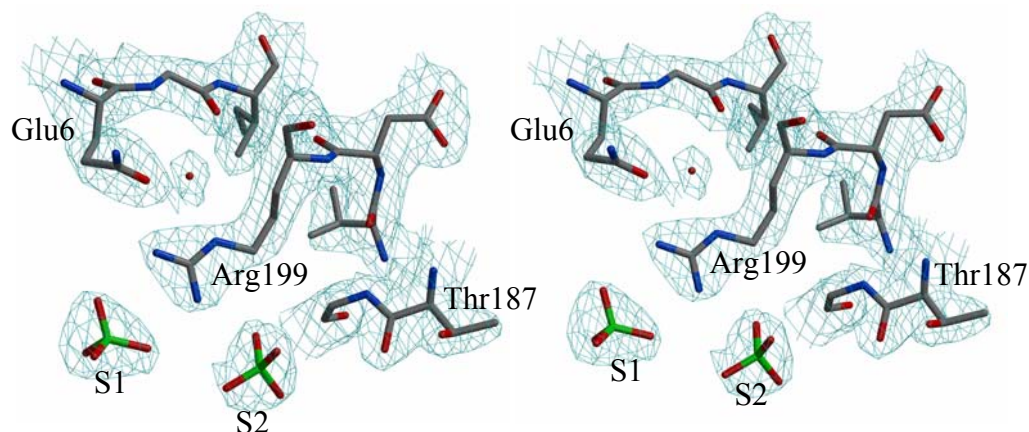


Figure 3.5.3-1 Stereographic depiction of a representative section of the final electron density map covering two sulphate ions. The $2F_o - F_c$ electron density map is contoured at 1.2σ above mean.

After the model was roughly complete, TLS parameters were refined for each of the two molecules in the asymmetric unit to describe their anisotropic motion, which lowered the R -factor by 4.1 % to 18.1 % and the R_{free} by 3.4 % to 22.9 %. The introduction of noncrystallographic symmetry (NCS) restraints (tight restraints for the main-chain atoms and medium restraints for the side-chain atoms between residues A1 – A216 and B1 – B216) resulted in a slight increase of the R -factor of 0.25 % but lowered the free R -factor by 0.4 %.

The quality of the resulting electron density maps allowed modelling of alternate conformations for three residues, i.e. Glu14, Arg91 and Arg208 of monomer A. In contrast, the electron density of the respective side-chains Glu14 and Arg91 of monomer B displayed a well-defined single conformation. For residue Arg208 of Monomer B, there was no electron density detectable from atom C γ onwards.

The model was further improved by alternate sessions of water building and refinement, which converged with an R -factor of 17.57 % and an R_{free} of 22.50 %. The final model contained 434 amino acid residues (2 monomers of 217 residues each), 20 sulphate ions and 252 water molecules. The refinement statistics are summarised in Table 3.5.3-1.

Table 3.5.3-1 Refinement statistics of the Cellosyl structure derived from the hexagonal crystal form.

resolution range	90.0 – 2.32 Å
total number of reflections used	27276 (99.12 %)
number of reflections in the working set	25814 (94.02 %)
number of reflections in the test set	1462 (5.1 %)
number of protein atoms	3363
number of sulphate ions	20
number of water molecules	252
<i>R</i> -factor [%]	17.57
<i>R</i> _{free} [%]	22.50

3.5.4 Validation of the model quality

The hexagonal crystal form contains two copies of the monomer in the asymmetric unit (monomer A and B) with an estimated solvent content of 63.75 %. The electron density in the final electron density map is well defined for all residues with exception of the C-terminal Ala217 of monomer A. However, there is no electron density for the side-chain of residue Arg88 B from atom C δ onwards, while the respective residue in monomer A is well defined. The electron density around the flexible side-chain of Arg208 could be interpreted by modelling two conformations for monomer A, whilst for monomer B there is no electron density observable from atom C γ onwards. Besides the double conformation of Arg208 A, the structural model contains two other residues with alternative side-chain conformations, i.e. Glu14 A and Arg91 A. Additionally, the model includes 20 sulphate ions and 252 water molecules.

The final *R*-factor for the structure is 17.57 % and the *R*_{free} 22.50 % for all data to 2.32 Å resolution. The r.m.s. deviations from ideality for the bond lengths and angles for the refined atoms are 0.025 Å and 2.05 ° respectively, indicating a good geometry for almost all residues. The Luzzati plot based on the *R*-value (Luzzati, 1952) gives an estimate of error in the atomic positions of 0.228 Å.

The geometry of the model was further assessed with the program *PROCHECK* (Laskowski *et al.*, 1993). 88 % of all residues are located within the highly favourable regions of the Ramachandran plot and none are located in the disallowed regions (see Fig. 3.5.4-1).

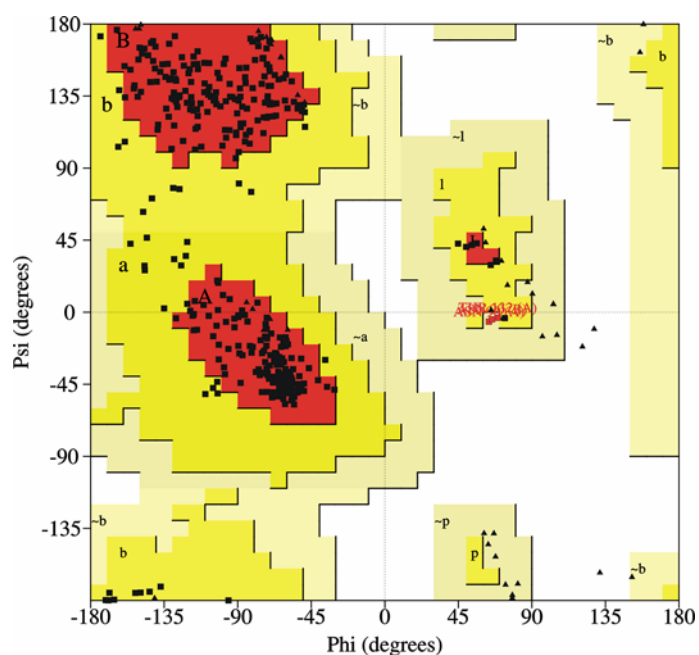


Figure 3.5.4-1 Ramachandran plot (Ramachandran and Sasisekharan, 1968) of the final structural model of Cellosyl derived from the hexagonal crystals generated with the program *PROCHECK* (Laskowski *et al.*, 1993) Glycins are marked as black triangles.

Three residues, i.e. Asn90 A, Asn90 B and Thr132 A, were found in the generously allowed regions. However, all three residues fit the electron density very tightly. It is interesting to mention, that in the structure of the monoclinic crystal form both residues, Asn90 and Thr132, were also found in the generously allowed regions of the Ramachandran plot due to their special position in loop regions (see chapters 3.3.7 and 3.4.3).

4. DISCUSSION

4.1 Overall structure of the monoclinic crystal form at 1.65 Å

The structure of Cellosyl comprises a single domain, shaped into a flattened ellipsoid with dimensions of 45 Å × 35 Å × 25 Å. At first sight its β/α -barrel fold resembles a TIM barrel, a structural motif first found in triosephosphate isomerase (Banner *et al.*, 1975) and later in many other enzymes (for review see Reardon and Farber, 1995; Nagano *et al.*, 2002). In a regular TIM barrel, the β -strands and α -helices alternate in a way that a regular 8-fold repeat of a $\beta\alpha$ -supersecondary structure element can be observed. The twisted, all parallel β -strands are stabilised by hydrogen bonds, with $\beta 8$ hydrogen bonded to $\beta 1$. The α -helices which connect the parallel β -strands, are located on the outside of the barrel. They are parallel to themselves but antiparallel to the strands. Often the barrels are preceded, interrupted or followed by additional domains.

The structure of Cellosyl is similar to a TIM barrel but shows very distinct features distinguishing it from the typical $(\beta/\alpha)_8$ fold. The enzyme is composed of eight β -strands and six α -helices (Fig. 4.1-1 and 4.1-2). As in regular TIM barrels, the first five β -strands and α -helices alternate, however the fifth α -helix in Cellosyl is followed by strands $\beta 6$ to $\beta 8$ which are connected by loops lacking any helices. Helix $\alpha 6$ is located at the carboxy-terminus of the polypeptide chain, sitting at the bottom (N-terminal end) of the barrel. All β -strands are arranged parallel to one another, except strand $\beta 8$ which, very unusually, is in an antiparallel orientation with respect to the other strands.

The β -strands are three to ten residues long. Strand $\beta 2$ forms a short β -hairpin (hydrogen bonding with residues 40 and 41) at its carboxy-terminal end. Strands $\beta 7$ and $\beta 8$ show a tight hydrogen bonding pattern and, with 10 residues each, are the longest strands in the barrel. The α -helices vary in length from 1.5 turns ($\alpha 1$) to 5 full turns ($\alpha 4$). The loop connecting $\beta 1$ and $\alpha 1$ contains a short 3_{10} helix. The loop linking the antiparallel strands $\beta 7$ and $\beta 8$ consists of only two amino acid residues. This is considerably shorter than the connections between the parallel β -strands, which are between 17 and 37 residues long (including the α -helices).

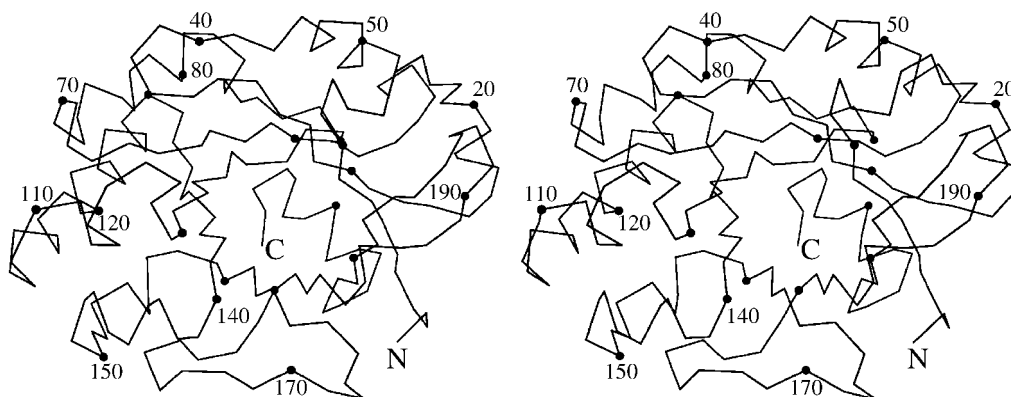


Figure 4.1-3 Stereo C α -trace of Cellosyl. The amino- and carboxy-termini are labelled and every tenth residue is highlighted with a black circle.

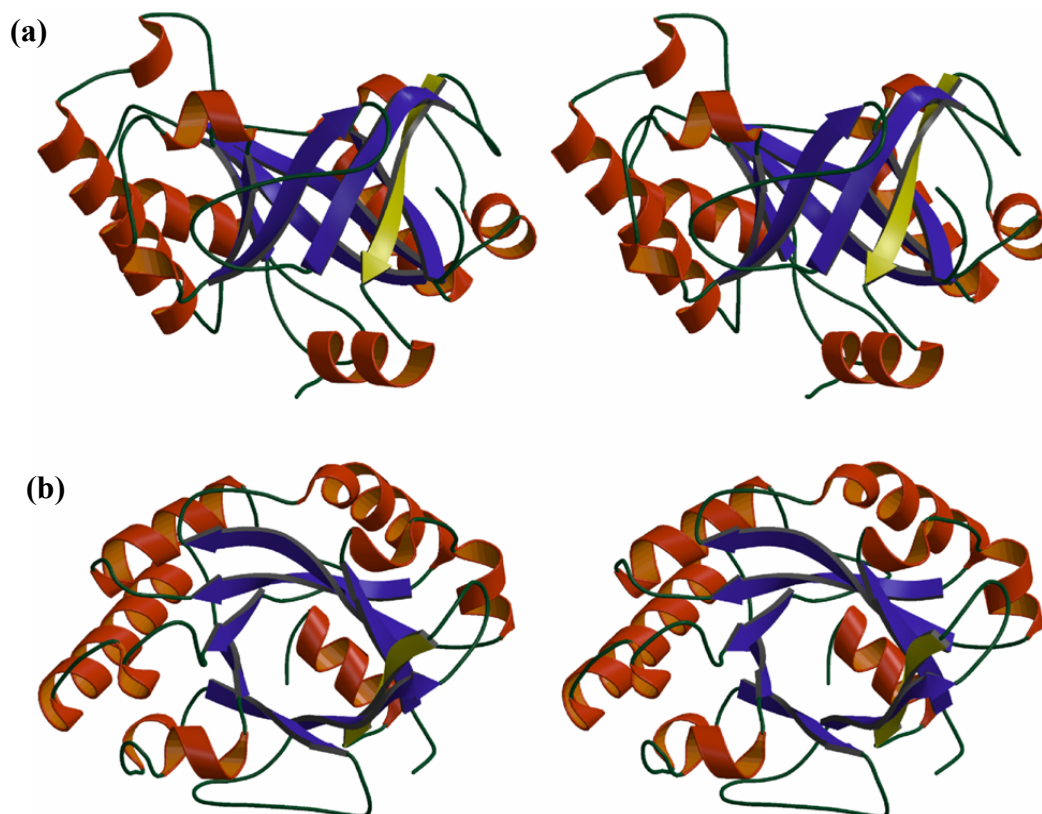


Figure 4.1-4 Stereographic ribbon representation of the overall fold of Cellosyl prepared with *BOBSCRIPT* (Esnouf, 1997); a) side view of the barrel, b) view inside the barrel, rotated by 90° with respect to the top view. Helices are depicted in red, loops in green, the parallel strands in blue and the antiparallel β -strand is highlighted in yellow.

The only disulfide bond of Cellosyl involves Cys108 in the loop between β 4 and α 4, and Cys147 at the carboxy-terminal end of α 5. No cis-peptide bond was found in the entire structure.

The primary solvent shell of the model contains 8 chloride ions, which were located on the basis of strong anomalous peaks (see chapter 3.3.6). All chloride anions are involved in a hydrogen-bonding network with at least two different hydrogen bond donors. N-H donor functions originating from the main-chain amide as well as from N-H functions of side-chains are always complemented by water-based O-H donor atoms. The coordination of the chloride ions is listed in Table 4.1.1 in the appendix.

4.2 Overall structure of the hexagonal crystal form at 2.32 Å

The crystal structure derived from the hexagonal crystal form contains 2 molecules in the asymmetric unit. As expected, both monomers exhibit the same unusual β/α -barrel fold, including the antiparallel arrangement of strand β 8, as the structure of Cellosyl in the monoclinic crystal form. A least-squares superposition of the monoclinic structure with each of the two monomers of the hexagonal structure resulted in an overall r.m.s. displacement between the 217 equivalent C α -atoms of 0.36 Å and 0.48 Å for monomer A and B, respectively.

The two monomers in the hexagonal crystal are related by a non-crystallographic symmetry relationship whereby the C-terminal ends of the β strands of both barrels face each other. The loops following β 1- β 4 and as well as the loop between β 6 and β 7 are involved in the intermolecular contacts. A hydrogen-bonding network, including several water molecules and 2 sulphate ions, stabilises this arrangement. A superposition of the two independent monomers in the asymmetric unit gives an r.m.s. deviation of 0.278 Å for the 217 equivalent C α -atoms. By omitting the C-terminal residue Ala217, which is poorly defined in the electron density of both monomers, the r.m.s deviation is as small as 0.152 Å.

There are 20 sulphate ions located in the primary solvent shell of the protein (Fig.4.2-1). Nine sulphate ions of monomer A can be superimposed with sulphate ions of monomer B. Only two sulphates, one in each monomer, have no equivalent in the other monomer. Table 4.2.1, showing the coordination of the sulphate ions, is presented in the appendix. All sulphate anions form hydrogen bonds to the protein, preferably to the N-H functions of the arginine side chains.

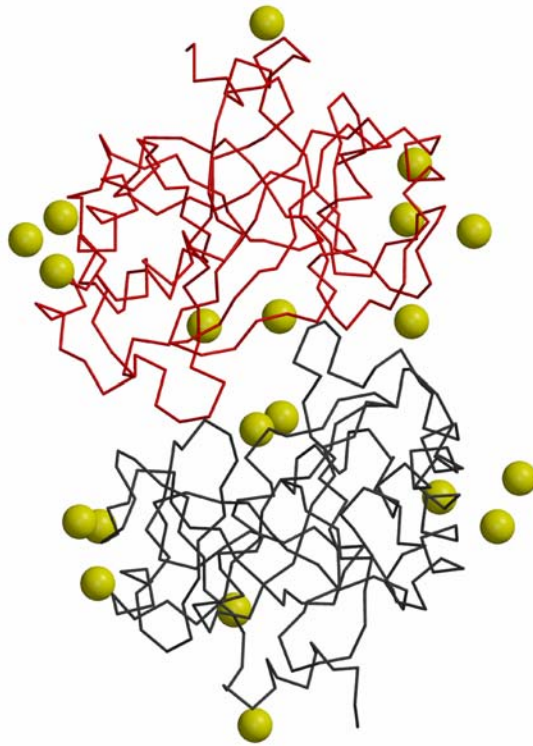


Figure 4.2-1 Backbone of Cellosyl structure derived from hexagonal crystals. Monomer A is depicted in red and monomer B in dark grey. The sulphate ions are shown as yellow spheres.

4.3 Atomic-resolution structure of Cellosyl at 0.83 Å resolution

4.3.1 Advantages and biological relevance of macromolecular crystal structures at atomic resolution

Over the last decade, the study of protein structures at atomic resolution has gained tremendous interest. The collection of atomic resolution data presents a challenging task due to the nature of protein crystals. The large size and the flexibility of protein molecules and the high solvent content within the crystals contribute to data weakness at high resolution. Until recently, only a small number of atomic-resolution protein structures were published (Dauter *et al.*, 1995; 1997). With developments in crystallographic methods, i.e. cryogenic cooling techniques, improvements in detector technology, and increasing availability of high-intensity synchrotron radiation sources, the number of atomic resolution data sets has steadily increased. A generally accepted definition of atomic resolution has been proposed by

Sheldrick (1990): The data should extend to a resolution higher than 1.2 Å with at least 50 % of the measured reflections in the outermost resolution shell having intensities $> 2\sigma(I)$.

One of the benefits of atomic resolution is the high number of measured reflections compared to the number of refined parameters which is crucial for a comprehensive least-squares refinement with anisotropic atomic temperature factors. The introduction of anisotropic displacement parameters allows a more accurate definition of the atomic positions. Therefore, atomic-resolution structures provide a much more detailed model of the protein under investigation and reveal features that are not normally identifiable at moderate resolution. Hydrogen atoms become visible in the difference Fourier maps allowing the identification of protonation states and yielding reliable views on biologically relevant hydrogen bonds. Multiple conformations can be identified for a greater portion of residues and modelled with partial occupancies. Further, a broader and more detailed picture of the solvent structure is obtained. Ordered water molecules emerge from the solvent continuum and the organisation of solvent beyond the first hydration shell can be studied. Atomic-resolution structures open the possibility to describe active sites and catalytic residues in greater detail, including the interactions with bound substrates or inhibitors, and might be of importance for detailed clarification of the mechanism of action of macromolecules.

The increasing number of macromolecular crystal structures solved at atomic resolution yields a more precise picture of the geometric and conformational properties of proteins in general. This store of knowledge enables the validation of the parameters employed in refinement. Target libraries for chemical and stereochemical parameters can be improved and applied for the refinement and validation of structures at low resolution. Furthermore, new structural features can be detected in atomic resolution structures, e.g. the presence of a very short hydrogen bond (less than 2.45 Å) has been reported (Wang *et al.*, 1997).

4.3.2 The overall structure

The three-dimensional structure of Cellosyl has been determined to atomic resolution using synchrotron radiation and new cryogenic cooling techniques (Riboldi-Tunnicliffe and Hilgenfeld, 1999). Model refinement with data extending to 0.83 Å resolution has been performed with anisotropic-displacement parameters, and the refinement has converged with excellent statistics, i.e. an *R*-factor of 9.63 %. This makes the structure of Cellosyl one of the highest resolution TIM-barrel fold structures determined to date, which is important

considering the fact that roughly 10 % of all enzymes of known structure assume the TIM-barrel fold (Copley and Bork, 2000; Gerlt, 2000).

As expected, the crystallographic studies of Cellosyl at atomic resolution revealed the same overall fold as the 1.65 Å resolution structure, which was used as initial model for the refinement. The r.m.s. deviation between the 216 equivalent C α -atoms of both structures is as low as 0.08 Å. However, compared to the lower-resolution model several new details have been revealed, e.g. an anisotropic description of the position of each atom in the structure, multiple side-chain conformations and alternate solvent networks.

4.3.3 Alternate conformations and disordered residues

The current structure has revealed several regions exhibiting multiple conformations, including 27 double side-chain, 1 triple side-chain and 5 dual main-chain conformations. Thus, about 13 % of the total residues adopt alternate conformations. A similar percentage of dynamic residues has been reported in other atomic-resolution structures (Walsh *et al.*, 1998; Esposito *et al.*, 2000). As depicted in Figure 4.3.3-1, the alternate conformations are widely distributed over the surface of the molecule.

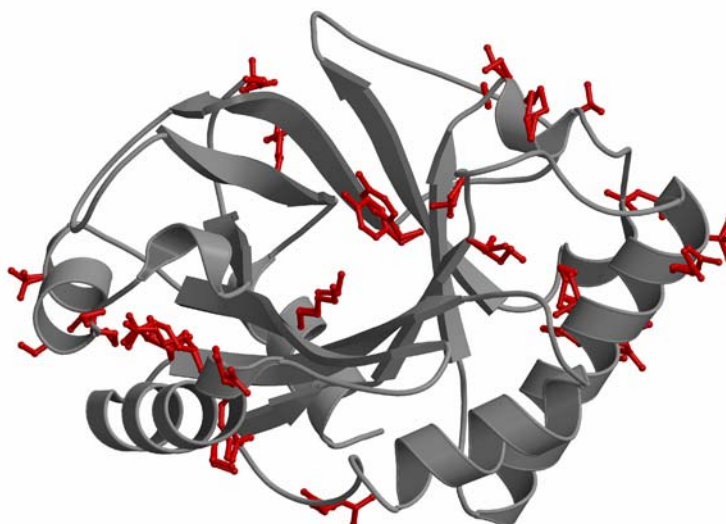


Figure 4.3.3-1 Overall fold of the high-resolution model of Cellosyl with the alternative side-chain conformations depicted in red as ball-and-stick.

Serine residues seem to be particularly prone to adopt multiple conformations. 6 of the 15 serine residues of Cellosyl were observed in dual conformations. Ser24 could be unambiguously modelled in three different conformations (Fig. 4.3.3-2). Side-chain occupancies of 50, 30 and 20 % were estimated from the peak heights of the $2mF_o-DF_c$ electron density map. All side-chain conformations interact with at least two water molecules of the first solvent shell.

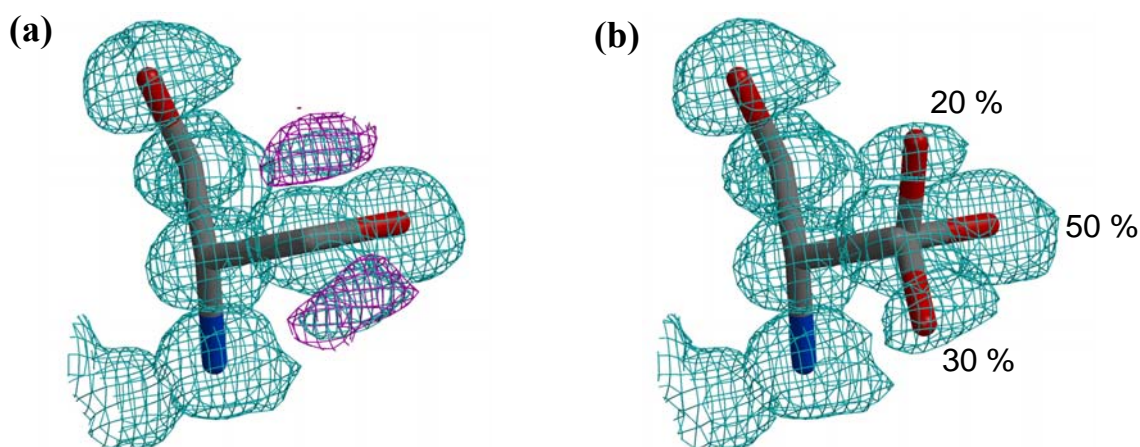


Figure 4.3.3-2 Depiction of residue Ser24: (a) The mF_o-DF_c electron density map, coloured in magenta, indicates a triple side-chain conformation. (b) Three alternative conformations were successfully modelled into the electron density. The occupancies of the respective side-chain conformations were estimated from the peak heights of the $2mF_o-DF_c$ electron density map, which is coloured in cyan.

Among the hydrophobic residues, Ile99 represents an interesting case. Two alternative side-chain conformations could be unambiguously modelled with a very clear $2mF_o-DF_c$ electron density map for all atoms. The occupancy of the bifurcated side chain is 70 % for the major and 30 % for the minor conformation. Ile99 is completely buried in the interior of the protein and surrounded by the side chains of Pro67, Leu97, Met116, Ile120, Trp143 and Trp144. There are no steric restrictions on the position of the side chain; all distances to neighbouring atoms exceed 3.3 Å.

Tyr138, a residue lining the active site, exhibits a major and a minor side-chain conformation. Because of its special location it will be discussed in more detail in the description of the active site.

Both, the N- and C-terminal residues of the structural model of Cellosyl are either dynamic or disordered. The C-terminal disorder is confined to residue Ala217 for which no interpretable electron density could be observed. The electron density map for the preceding residue Thr216 is well defined and a single conformation of the residue was modelled. However, the carbonyl oxygen of Thr216 is, with a ratio of 0.15 (a value of 1.0 describes a perfectly isotropic atom), one of the most anisotropic atoms within the structure, and the mean anisotropy for the whole residue is only 0.29. The N-terminal residue Asp1 could be confidently modelled in two alternative conformations with 50 % occupancy each. Both side-chain orientations are directed towards the solvent. However, the OD1 atom of the first conformation is stabilised by a hydrogen bond to the NZ atom of Lys201 (2.60 Å), while the OD1 atom of the second conformation is interacting with the hydroxyl group of the neighbouring Ser2 (2.62 Å).

5 residues show dual conformations of the main chain. One of them, Ser104, is located in the loop between β_4 and α_4 , forming a bend together with residues Pro103 and Gly105. Ser104 is positioned at the tip of this bend and the disorder is restricted to this residue. The backbone of Ser104 points out towards the solvent and both conformations interact with water molecules of the first solvent shell. Two further residues, Ala156 and Lys157, are also positioned in a surface loop. According to the program *DSSP* (Kabsch and Sander, 1983), they are part of a 5 residue long turn (residue 153-157) within this loop between α_5 and β_6 . A dual main-chain fragment was also observed at the C-terminal end of helix α_5 , where Ala141 and Ser142 were modelled in two alternate conformations.

4.3.4 Solvent region

The final model comprises 567 water molecules, of which 364 are fully and 199 half occupied. Four water molecules were refined with the same occupancy as an alternative side-chain conformation within hydrogen-bonding distance. Most of the solvent structure is very well ordered, with a *B*-value as low as 5.57 Å² for the partially occupied water 8. Thirty-nine water molecules have *B*-values lower than 10 Å². Several of them are completely buried within the protein. For example, a very intriguing hydrogen-bonding pattern evolves around water molecule 4. Wat4 possesses no hydrogen bonds to other water molecules. It serves as hydrogen bond donor for the main-chain oxygens of Ala65 (2.90 Å) and Asp98 (2.95 Å) and as hydrogen bond acceptor for the main-chain nitrogen of Glu100 (2.90 Å) and the phenolic oxygen of Tyr109 (3.02 Å). Therefore, the surrounding protein framework saturates all

intermolecular contacts of wat4. In addition to the very interesting interactions of wat4 itself, it might also serve as a structurally stabilising component for the alignment of the catalytically active carboxylates Asp98 and Glu100 through fixation of the main chain oxygen of Asp98 and the nitrogen of Glu100.

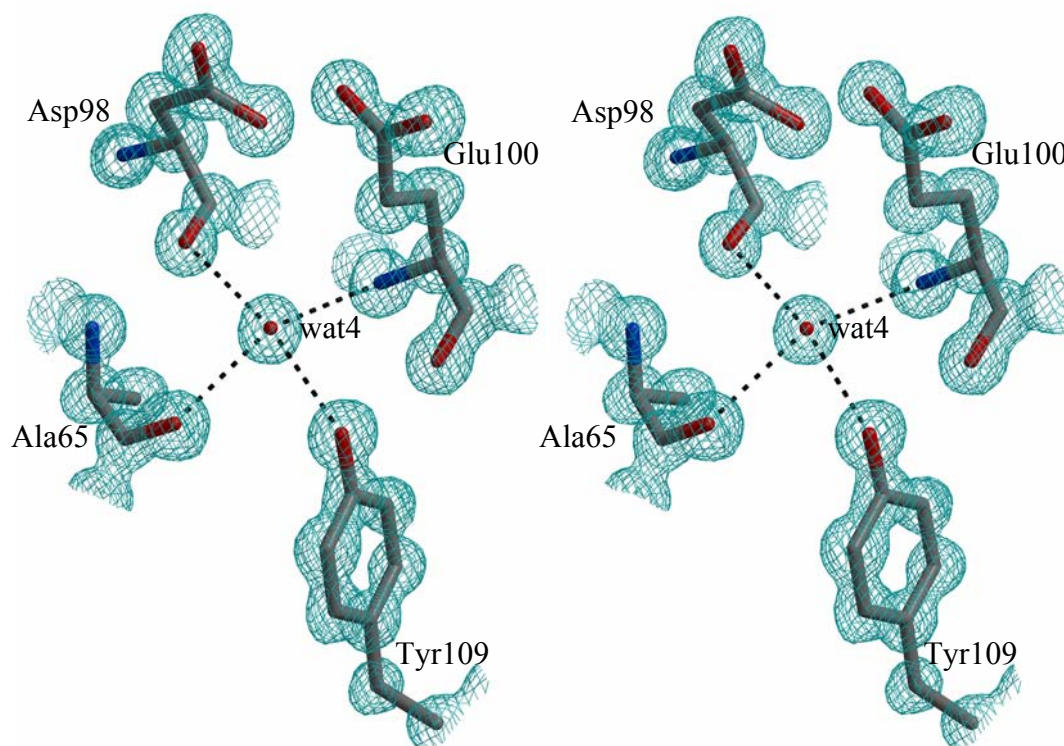


Figure 4.3.4-1 Stereographic depiction of the completely buried water 4. The $2mF_o$ -DFc electron density map is contoured at 1.5σ above mean.

Like the 1.65 \AA model, the high-resolution structure of Cellosyl contains 8 chloride ions. Their positions as well as their interactions with the protein framework are conserved between the two structures.

4.3.5 Anisotropic displacement parameters

Anisotropy is defined as the ratio of the minimum and maximum Eigenvalues of the anisotropic displacement parameter (ADP) matrix (Trueblood *et al.*, 1996). The ratio is 1.0 for a perfectly isotropic (spherical) atom, while it approaches 0 for an extremely non-spherical atom. Introduction of ADPs into the refinement of the atomic-resolution structure of Cellosyl resulted in a drop in the *R*-factor by more than 4 %. This confirms that this structure, like other protein structures refined at high resolution, is better described in terms of anisotropic displacement.

The ADPs of the atomic resolution structure of Cellosyl have been analysed using the web-based program *PARVATI* (Merritt, 1999). The anisotropy for the protein atoms ranges from 0.1 to 0.95, showing the typical Gaussian distribution of anisotropy in protein structures. It could be observed that the atoms became increasingly non-spherical as their distance from the centre of mass increased. The mean anisotropy for all protein atoms is 0.56 with a standard deviation of 0.162. A thermal ellipsoid representation of four residues containing highly anisotropic atoms is shown in Fig. 4.3.5-1.

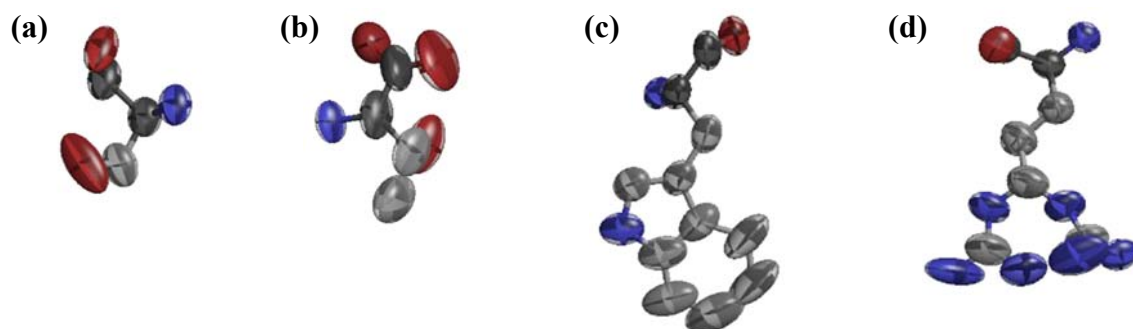


Figure 4.3.5-1 Illustration of the anisotropic displacement parameters for a selection of residues containing atoms with an anisotropy < 0.2 : (a) Ser70, (b) Thr216, (c) Trp165 and (d) Arg88, the latter was modelled with two alternative side-chain conformations.

Although the solvent molecules were refined with the ISOR restraint (to keep them roughly isotropic) they turned out to be rather anisotropic: the mean anisotropy factor of water was 0.33 ($\sigma = 0.14$) and of the eight chloride ions 0.39 ($\sigma = 0.19$).

The mean *B*-value and the mean anisotropy of the displacement parameters averaged for all main-chain and side-chain atoms are depicted in Figure 4.3.5-2. The mean *B*-value for the protein chain and the solvent atoms are 8.44 \AA^2 and 25.45 \AA^2 , respectively.

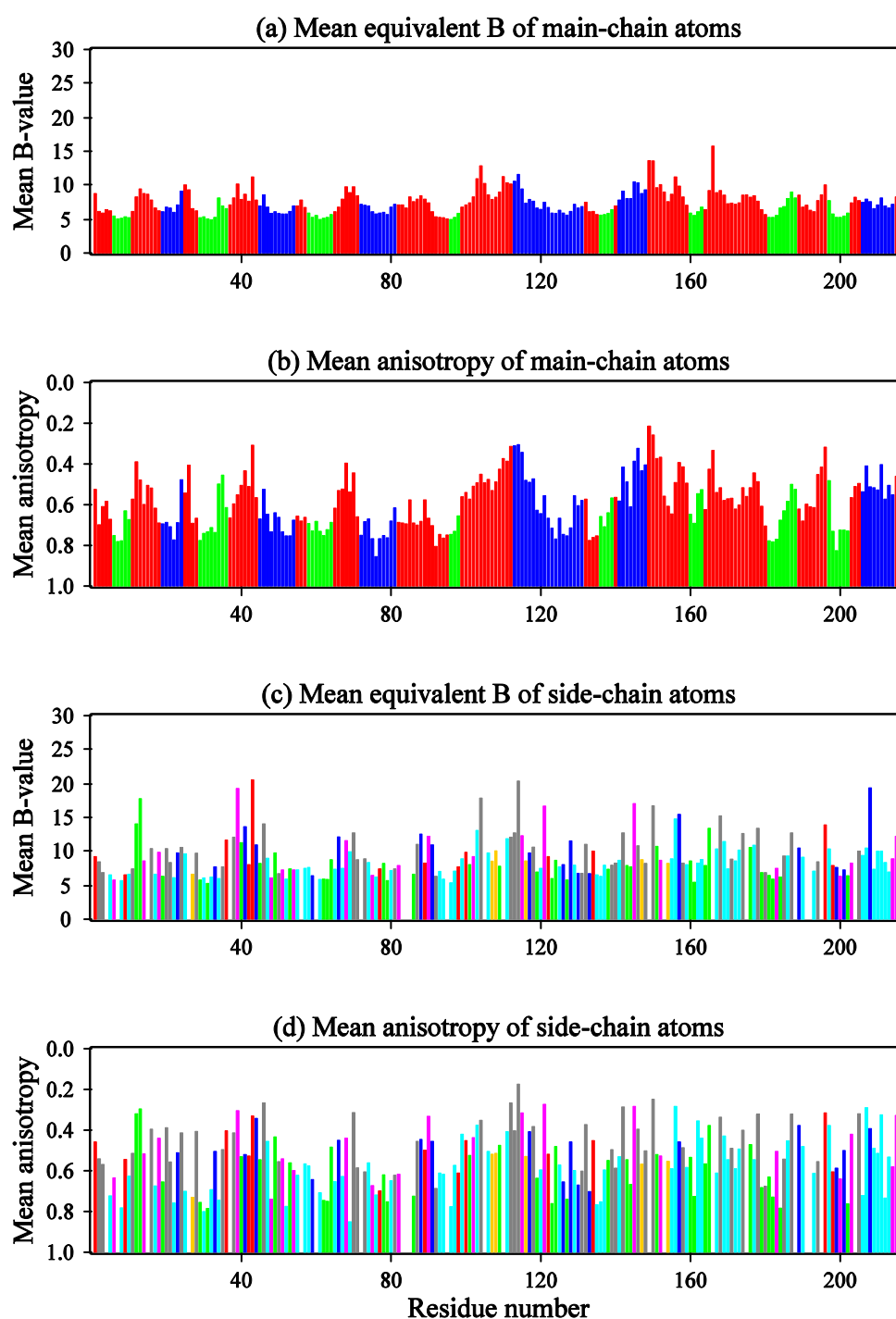


Figure 4.3.5-2 (a) Mean equivalent B of main-chain atoms. (b) Mean anisotropy of main-chain atoms. For (a) and (b), the colours are coded according to secondary structure: blue = alpha helix, green = beta-strand, red = other. (c) Mean equivalent B of side-chain atoms. (d) Mean anisotropy of side-chain atoms. For (c) and (d), the colours are coded according to residue characteristics: yellow = Cys, Met; green = Phe, Tyr, Trp, His; cyan = Gly, Ala, Leu, Ile, Val, Pro; red = Glu, Asp; blue = Arg, Lys; purple = Gln, Asn; grey = Ser, Thr.

4.4 Active site and mechanism

In all known ‘conventional’ β/α -barrel enzymes, the active site is located at the carboxy-terminal end of the β -barrel. In spite of the presence of one antiparallel strand ($\beta 8$) in the barrel, the same can be expected for Cellosyl. Indeed, a prominent, long groove, very likely the substrate-binding site, is located on the carboxy-terminal face of the Cellosyl β -barrel. This groove culminates in a deep hole of highly negative electrostatic potential (Fig. 4.4-1), which was identified as the catalytic site of the enzyme.

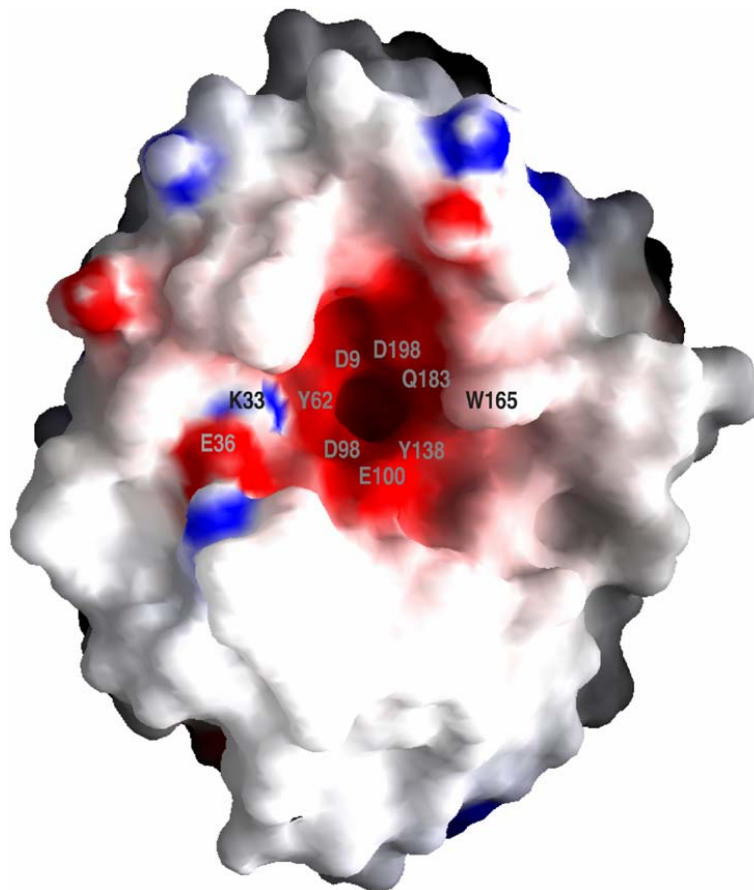


Figure 4.4-1 Charge distribution on the surface of Cellosyl. Positively and negatively charged electrostatic potentials are indicated in blue and red, respectively. The surface was calculated using a probe radius of 1.4 Å, and the potential displayed on a scale ranging from $-17.2 k_B T$ to $+19.8 k_B T$ (k_B = Boltzmann constant). Important residues are labelled. The figure was generated with the program *GRASP* (Nicholls *et al.*, 1991).

In order to cleave the β -1,4-glycosidic bond via general acid catalysis, glycosyl hydrolases generally employ a pair of carboxylic acids at the active site: one functioning as proton donor and one acting as a nucleophile/base (McCarter and Withers, 1994; Rye and Withers, 2000). The residues Glu36 and Asp9 (using the Cellosyl numbering) have been proposed as catalytic residues for the enzymatic mechanism of fungus *Chalaropsis* lysozyme (Fouche and Hash, 1978).

In Cellosyl, Glu36 is located at the tip of strand β 2 but it is not part of the active site. However, its carboxylate side-chain is oriented towards the substrate-binding groove. Also, it lacks an acidic or neutral hydrogen-bonding residue in its immediate spatial neighbourhood; it forms a strong salt-bridge (2.67 Å) with the conserved Lys33. Glu36 is very likely involved in substrate binding, but not in substrate cleavage. This is supported by the finding that mutation of Glu36 in the pneumococcal Ch-type lysozymes does not lead to dramatic reduction of the activity (Sanz *et al.*, 1992). On the other hand, mutation studies of Asp9 on the Ch-lysozyme yielded 2.2%, 1.7% and 0.2% of the wild-type activity for mutants D9N, D9E and D9A, respectively (Sanz *et al.*, 1992), underlining the critical role of Asp9.

The crystal structure of Cellosyl revealed new insights on the active-site architecture. Two pairs of carboxylic residues, Asp9 – Asp198 and Asp98 – Glu100, line opposite faces of the central hole (Fig.4.4-1). Asp9, Asp98 and Glu100 are strictly conserved within the Ch-type lysozyme family (Fig.4.4-2). Asp198, however, is only partially conserved. Judging from the active-site geometry, Asp9, Asp98 and Glu100 are candidates for the catalytically active residues. The critical role of Asp9 for the enzymatic mechanism has already been established by mutation studies, as discussed above. Superposition of Cellosyl with related TIM-barrel structures of chitinases, e.g. hevamine of family GH-18, showed that their proton donor residue is placed in an equivalent position at the carboxy-terminal end of β 4 and matches the Glu100 residue of Cellosyl. These observations were confirmed by a very recent study of the crystal structure of the lysozyme from phage Cp-1 (Hermoso *et al.*, 2003). Side-directed mutagenesis of Glu100 (using the Cellosyl numbering) led to an extinction of the activity and demonstrated the importance of this residue in the catalytic mechanism. Therefore, it seems likely that Glu100 acts as the proton donor and Asp9 acts as the nucleophile/base.

As already mentioned, both of the proposed catalytically active residues are involved in a hydrogen-bonding interaction with an aspartate residue in close spatial vicinity (Fig. 4.4-3). Asp98 and Asp198 seem to play an important role in positioning Glu100 and Asp9, respectively, into the correct orientation for effective substrate hydrolysis.

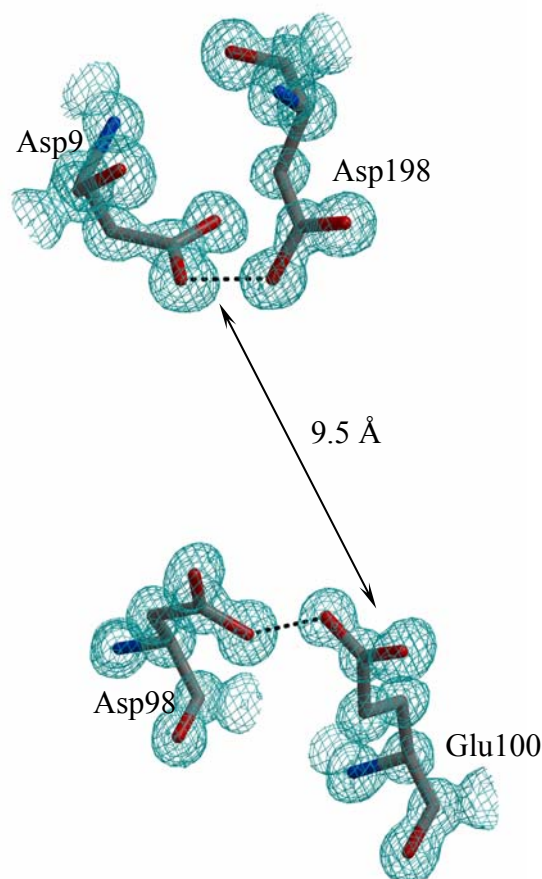


Figure 4.4-3 Depiction of the two pairs of carboxylic residues lining opposite faces of the active site cleft. The proposed catalytic residues Asp9 and Glu100 are 9.5 Å apart (average of the four possible distances between the oxygen atoms), indicating a mechanism under inversion of configuration. The $2mF_o-DFc$ electron density map is contoured at 1.5 σ above mean.

But besides the stabilising effect, the strength of both hydrogen bonds (with distances Asp9 OD1 - Asp198 OD2 of only 2.55 Å and Asp98 OD1 - Glu100 OE1 of only 2.50 Å) indicates the presence of low-barrier hydrogen bonds (LBHB). The energetic profile of a LBHB is ideally suited for fast and therefore efficient proton trafficking within these pairs of catalytically active carboxylates. The enormous importance of LBHBs lies in the accelerating effect on protonation/deprotonation reactions. LBHBs involved in acid/base catalysis were reported for several enzymes. It is believed that they accelerate the catalytic reaction by at least five orders of magnitude (Cleland *et al.*, 1998). Unfortunately, the protons of these carboxylic side chains could not be modelled, since they could not be unambiguously differentiated from the background noise of the mF_o-DFc electron-density map.

The active site of glycosidases is commonly rich in aromatic residues, i.e. tryptophan and tyrosine. The hydrophobic character and hydrogen-bonding capability of these residues seems to be favourable for the binding of polysaccharide substrates (McCarter and Withers, 1994). The same is true for the active site of Cellosyl. The central hole is approximately 8 - 9 Å deep. The aromatic side chains of Tyr31, Tyr62, Phe64, Tyr138, and Trp161 mainly form the walls, while Val96 is at the very bottom of the hole. Hydrophobic residues are located at those positions throughout the Ch-lysozyme family. Two of the tyrosines, Tyr62 and Tyr138, are absolutely conserved. The atomic-resolution structure (0.83 Å) of Cellosyl revealed that the side chain of Tyr138 exhibits dual conformations. The major side-chain conformation is orientated parallel to Tyr62 at a distance of 7.03 Å. Interestingly, a water molecule (wat44) is sandwiched between their phenyl rings, almost halfway down the deep hole. The distance between the water oxygen and the centre of the aromatic ring of Tyr138 is 3.32 Å, suggesting a possible O-H... π interaction. The corresponding distance to the parallel phenyl ring, Tyr62, is too large (3.72 Å) for a hydrogen bond-like interaction. Instead, wat44 forms H-bonds with the carboxylic group of the residue Asp98 (2.98 Å) and with water wat85 (2.79 Å). However, in the minor side-chain conformation of Tyr138, its χ_2 angle is rotated by 54° with respect to the major conformation. Therefore, this side chain is orientated nearly coplanar to the phenyl moiety of Trp161 with a distance of 3.85 Å, indicating a weak π -stacking interaction. The dual conformation of the side chain of Tyr138 also resulted in a formation of an alternative solvent network within the active-site cleft.

As already described in the introduction (chapter 1.3), hydrolysis of the polysaccharide substrate can occur via two major mechanisms, giving rise to either an overall retention or an inversion of the anomeric configuration. At first glance, retaining and inverting enzymes have similar active-site architectures, with two essential carboxylic acids facing each other at the substrate-binding site. However, the distance between the carboxylate residues of inverting enzymes is approximately 10 Å, and thus significantly larger than that for retaining enzymes in which the two carboxylic acid residues are approximately 5.5 Å apart. The greater distance for the inverting enzymes is probably required for the mechanism, where in addition to the substrate the nucleophilic water molecule has to be placed between the two carboxylates. The distance between Asp9 and Glu100 in the Cellosyl structure is 9.5 Å (Fig.4.4-3), indicating that the enzymatic mechanism proceeds under inversion of configuration via a single displacement mechanism.

4.5 Relationship to other lysozyme structures

As outlined in the introduction, four different classes of lysozymes can be distinguished at present (Jollès and Jollès, 1984). It is believed that the three classes represented by HEWL, GEWL, and T4 lysozyme are the result of divergent evolution from a common ancestor (Weaver *et al.*, 1985). Despite the fact that they do not share any statistically significant sequence identity, their three-dimensional structures show some intriguing, albeit distant, similarities: The active site is located in a crevice between two domains which are connected by a long α -helix. A glutamic acid residue, proposed to be essential for general acid catalysis, is located at the carboxy-terminal end of an α -helix in the amino-terminal half of each of the proteins (see Fig.4.5-1).

The β/α barrel fold of Cellosyl is completely unrelated to the tertiary structures of HEWL, GEWL and T4 lysozyme and defines a fourth family, that of the Ch-type lysozymes. A previous structural report for a Ch-type lysozyme describes a low-resolution model for the muramidase from *Streptomyces erythraeus* (Harada *et al.*, 1981). This structural model comprises a polypeptide chain of approximately 180 amino acid residues folded into three domains and is very different from the structure presented in this work. It is conceivable that the apparent structural differences are due to the relatively low resolution (2.9 Å) of that crystallographic study which was carried out in 1981, i.e. before the advent of molecular graphics, and in the absence of a complete amino-acid sequence. However, given the significant sequence similarity (49% identity) between the two *Streptomyces* lysozymes this structural model is most probably incorrect. A recent publication of the three-dimensional structure of a further Ch-type lysozyme confirmed this assumption (Hermoso *et al.*, 2003). The authors describe the pneumococcal lysozyme from phage Cp-1 consisting of two domains. In addition to the catalytic module, the Cpl-1 possesses a choline-binding module, with both domains joined together by an acidic linker. Superposition of the catalytic domain of Cpl-1 onto Cellosyl resulted in an r.m.s. deviation of 2.4 Å for 174 equivalent C α -atoms. The catalytic domain of Cpl-1 folds, exactly like Cellosyl, into a irregular $(\beta/\alpha)_5\beta_3$ barrel and also contains the antiparallel strand β_8 .

Thus, the three-dimensional structure of Cellosyl represents the first detailed model of a Ch-type lysozyme, and given the high sequence similarity (47 %) to the lysozyme of fungus *Chalaropsis*, it is conceivable that all members of this family share the same unusual β/α fold described in this work.

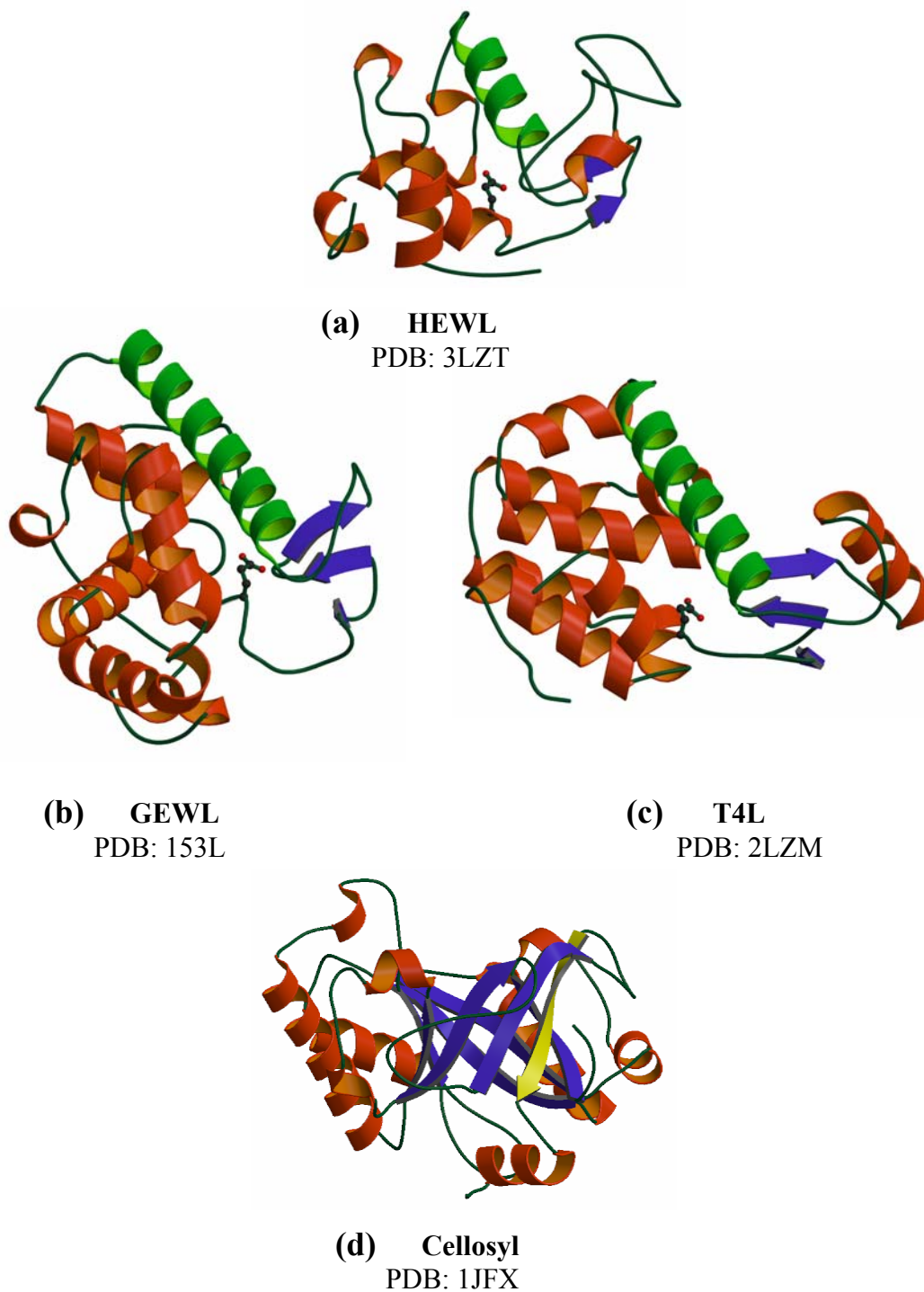


Figure 4.5-1 Comparison of the three-dimensional structures of (a) HEWL, (b) GEWL, (c) T4L, and (d) Cellosyl. The structures (a)-(c) were superposed with the program *DALI* (Holm and Sander, 1993; 1996). The completely different structure of Cellosyl is depicted from the side view of the barrel.

4.6 Relationship to other β/α barrel enzymes

A search for structural similarities using the program *DALI* (Holm and Sander, 1993; 1996) shows that the highest similarity to Cellosyl is displayed by the Ch-type lysozyme from phage Cp-1 which is discussed in the previous chapter. Apart from Cpl-1, which also belongs to GH family 25, the search yielded chitinases from family 18 and glucanases from family 5 as closest (albeit rather distant) relatives. Among these enzymes, the closest relative of Cellosyl is a chitinase from fungus *Coccidioides immitis* (CiX1) (PDB code: 1D2K) (Hollis *et al.*, 2000), with an r.m.s. deviation of 3.5 Å for 181 equivalent C α atoms. A least-squares superposition of Cellosyl and CiX1 is depicted in Figure 4.6-1. CiX1 belongs to the GH-18 family and possesses an irregular $(\beta/\alpha)_8$ -barrel topology in which helix $\alpha 1$ is absent and strand $\beta 2$ is followed by two helices. β -strands 3 through 8 are followed by α -helices as in a conventional TIM barrel. Furthermore, the barrel contains two insertions: strand $\beta 2$ is followed by a two-stranded antiparallel β -sheet and the connection between $\beta 7$ and $\alpha 7$ contains an additional subdomain with a modified Greek key motive. In contrast to Cellosyl, CiX1 contains 4 cis-peptide bonds.

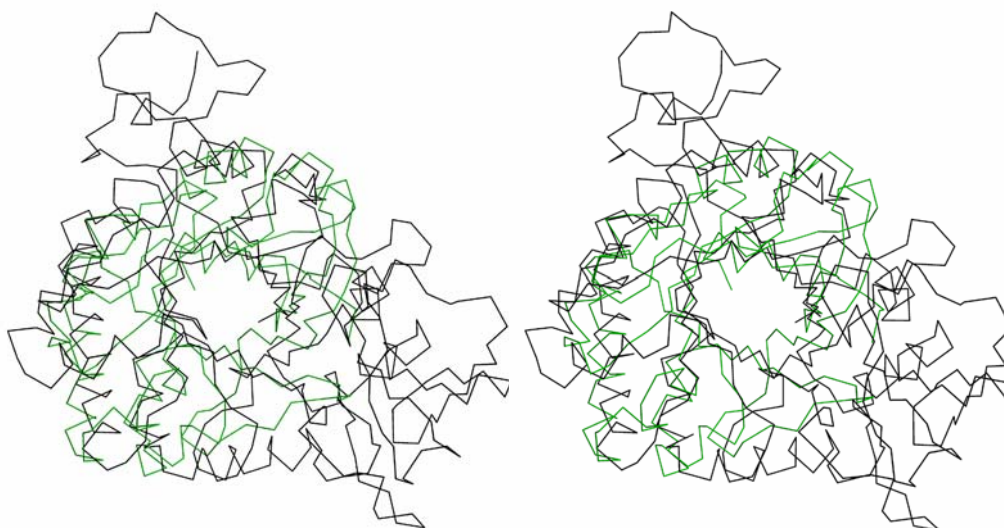


Figure 4.6-1 A stereo representation of the superposition of Cellosyl and CiX1. Cellosyl is coloured in green, while CiX1 is coloured in black. The insertion of two additional domains into the β/α barrel of CiX1 is clearly visible.

The next relative listed by *DALI* is the β -mannase from *Thermomonospora fusca* with an r.m.s. deviation of 3.7 Å for 168 residues (PDB code: 1BQC) (Hilge *et al.*, 1998). The family 5 glycosyl hydrolase exhibits the classical $(\beta/\alpha)_8$ -barrel fold with two short, additional strands at the N-terminus forming the bottom of the barrel. It is interesting to mention that despite the functional differences between the members of family GH-5, GH-18 and Cellosyl, the catalytically active proton-donor residue is located at the carboxy-terminal end of strand β_4 , and spatially matches the residue Glu100 of Cellosyl.

As mentioned above, a very unique structural feature of Cellosyl is the antiparallel orientation of strand β_8 in the barrel. The only other enzyme known to have an 8-standed β/α -barrel fold with an antiparallel β -strand is enolase (Lebioda *et al.*, 1989). Topology diagrams of a conventional TIM barrel and the two irregular barrels of Cellosyl and enolase are depicted in Figure 4.6-2.

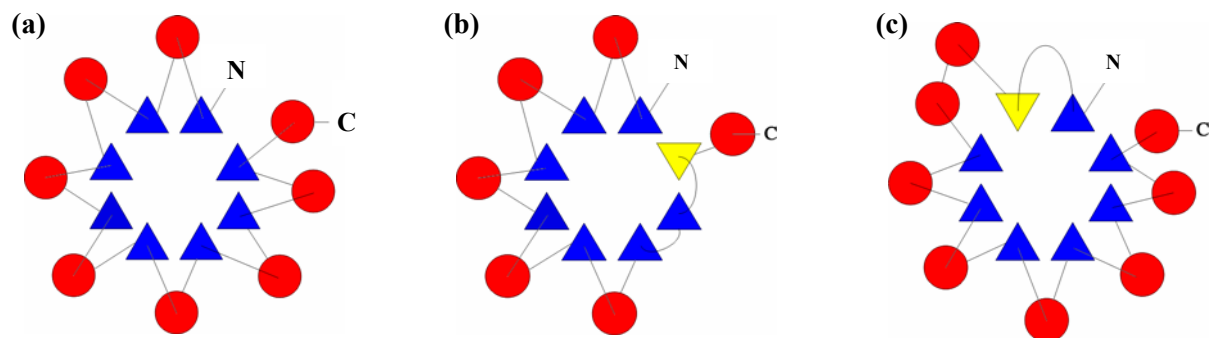


Figure 4.6-2 Topology diagrams of (a) a conventional TIM barrel, (b) Cellosyl and (c) enolase. α -helices are depicted as red circles and β -strands as blue triangles. The antiparallel β -strand of Cellosyl and enolase is highlighted in yellow.

Enolase is a glycolytic enzyme which catalyses the dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate. The protein consists of a smaller N-terminal domain and a larger C-terminal domain which is an 8-fold β/α -barrel with a $\beta\beta\alpha\alpha(\beta\alpha)_6$ topology. In the enolase barrel, the second β -strand is antiparallel to the other strands and the first α -helix is antiparallel to the other α -helices. The unusual antiparallel orientation of a β -strand in a β/α -barrel does not seem to lead to a reduced number of hydrogen bonds made by main-chain amides and carbonyls in the strand: The antiparallel strand β_2 in enolase forms 10 main-chain hydrogen bonds with its neighbours. In Cellosyl, seven H-bonds are formed between strands

$\beta 8$ and $\beta 7$, and three such interactions between strands $\beta 8$ and $\beta 1$. The number of hydrogen bonds made by strand $\beta 8$ of Cellosyl therefore agrees well with the average number (6 - 14) for strands $\beta 2$ to $\beta 6$ which constitute the all-parallel part of the barrel.

4.7 Evolution of $(\beta/\alpha)_8$ -barrel structures

The β/α -barrel scaffold is one of the most versatile and most frequently encountered protein folds in nature. Roughly 10 % of all enzymes with known three-dimensional structure possess a $(\beta/\alpha)_8$ -barrel domain (Copley and Bork, 2000; Gerlt, 2000). Proteins of this architecture are nearly always enzymes and play an important role in molecular and energy metabolism. They catalyse a vast variety of different reactions, functioning as oxidoreductases, transferases, hydrolases, lyases or isomerases (Nagano *et al.*, 2002).

The evolution of $(\beta/\alpha)_8$ -barrels has been discussed for many years and is still the subject of vigorous debate (Lesk *et al.*, 1989; Farber and Petsko, 1990; Brändén, 1991; Reardon and Farber, 1995; Copley and Bork, 2000; Nagano *et al.*, 2002; Gerlt and Raushel, 2003). The question if they arose by convergent evolution to a stable fold, by divergent evolution from a common ancestor, or even by divergent evolution from several ancestors that had similar β/α folds, remains unresolved.

The absence of significant sequence similarity, as well as the large range of reactions catalysed, suggests the possibility that $(\beta/\alpha)_8$ barrels have developed by convergent evolution. Analyses of the geometry of these proteins and the packing arrangements of the hydrophobic residues in the centre of the β -barrel were performed to elucidate the origin of this fold (Lesk *et al.*, 1989). The authors identified three layers of side chains forming the interior of the barrel and suggested two different packing types, which led to the conclusion that not all proteins with a $(\beta/\alpha)_8$ barrel domain evolved from a common ancestor. However, this analysis was restricted to three enzymes only. More recent studies on residue packing within the interior of the barrel revealed that even within homologous families, the number of layers may differ and no evidence for two different packing types was found (Nagano *et al.*, 1999).

Farber and Petsko (1990), and more recently, Reardon and Farber (1995), suggested divergent evolution of $(\beta/\alpha)_8$ barrels from a common ancestor. Although these enzymes catalyse a great variety of different reactions, the active site is always located at the carboxy-

terminal end of the β -sheets. Furthermore, barrel shape and domain composition allowed grouping these enzymes into distinct families and more of the data available could be explained by divergent than by convergent evolution.

No less than 25 superfamilies of $(\beta/\alpha)_8$ -barrels were defined in the recent release of the SCOP (structure comparison of proteins) database¹ which clusters proteins into families and superfamilies based on a hierarchical level, describing near and distant evolutionary relationships (Lo Conte *et al.*, 2000; 2002). Copley and Bork (2000) analysed 23 of these TIM-barrel superfamilies and concluded that at least 12 derived from a common ancestor. Thornton and co-workers reported similar results (Nagano *et al.*, 2002), analysing TIM-barrel folds deposited in the CATH database² (Pearl *et al.*, 2000). Structural and functional comparisons of the 21 homologous TIM-barrel superfamilies indicated a common ancestry for 17 of them. The question if the remaining superfamilies are the result of convergent evolution or if all $(\beta/\alpha)_8$ -barrels diverged from a common ancestor remains unanswered.

Very intriguing results were reported for two enzymes, HisA and HisF, catalysing successive reactions in the histidine biosynthesis pathway. An internal duplication in both HisA and HisF genes was revealed and led to the conclusion that HisA and HisF originated from the duplication of a smaller ancestral gene (Fani *et al.*, 1994). With the knowledge of amino acid sequence and tertiary structure, it was suggested that both enzymes have evolved from a common half-barrel ancestor by twofold gene duplication, followed by gene fusion (Lang *et al.*, 2000). This was supported by the observation that the N- and C-terminal $(\beta/\alpha)_4$ -units in HisF build stable folded substructures and could be assembled into the fully active complex upon co-expression *in vivo* or joint refolding *in vitro* (Höcker *et al.*, 2001). Moreover, observations have been made leading to the suggestion that β/α barrels assembled from $(\beta/\alpha)_2$ -subdomain structures to create an 8-fold barrel (Nagano *et al.*, 2002; Gerlt and Raushel, 2003).

Little attention has been paid to β/α barrel enzymes which differ from the typical TIM-barrel fold. The structure of Ch-type lysozymes, i.e. Cellosyl, should be of relevance for the ongoing debate on the evolution of β/α barrels because of the presence of the antiparallel strand β_8 in the otherwise parallel barrel. The presence of an antiparallel β -strand in a β/α -barrel has been also reported for yeast enolase (Lebioda *et al.*, 1989). Enolase has been grouped into one superfamily with muconate-lactonizing enzyme (MLE) and mandelate racemase (MR) because they share the same two-domain structure and have a common

¹ <http://scop.mrc-lmb.cam.ac.uk/scop>

² <http://www.biochem.ucl.ac.uk/bsm/cath>

mechanism of catalysis (Babbitt and Gerlt, 1997; Lo Conte *et al.*, 2000). While MLE and MR possess the normal TIM-barrel fold, the structure of enolase has the $\beta\beta\alpha\alpha(\beta\alpha)_6$ topology as discussed above (see chapter 4.6). That the second β -strand of the barrel of enolase is arranged in an antiparallel fashion is, unfortunately, hardly mentioned at all in the literature. With the two novel structures of Ch-type lysozymes (Rau *et al.*, 2001; Hermoso *et al.*, 2003), further enzymes have been shown to have an antiparallel β -strand in the barrel. Has this ‘irregularity’ evolved by chance? Are these enzymes derived from a common ancestor? Would that ancestor be different from that of the other TIM barrels?

Overlapping enolase and Cellosyl by an automated 3D structure-comparing program (Holm and Sander, 1993; 1996) yielded a perfect overlap of the two antiparallel β -strands. If the two enzymes share a common ancestor, the question arises why the location of the antiparallel β -strand is in position 2 for enolase and in position 8 for Cellosyl. It seems possible that the enzymes have evolved by circular permutation of the N- and C-termini. Circular permutations within β/α -barrel enzymes have been reported in literature (Sergeev and Lee, 1994; Jia *et al.*, 1996) and were also suggested for the enolase superfamily (Copley and Bork, 2000; Nagano *et al.*, 2002). Taking these studies into account, Cellosyl and enolase may have derived from a common ancestor through cyclic permutation of N- and C-termini in the barrel. Further, it is quite possible that similar to the proposal by Lang *et al.* (2000), this common ancestor itself evolved by gene fusion from two independent half-barrels, an unconventional one comprising the secondary structure elements around the antiparallel strand (like β_6 - β_7 - β_8 - α_6 - β_1 in Cellosyl), and a canonical one comprising alternating β/α units. In any case, the structures of Cellosyl and enolase suggest that β/α barrels divergently evolved from more than one common ancestor. Moreover, it is becoming clear now that the structural and functional plasticity of the β/α barrel fold goes beyond variations on the all-parallel β -barrel theme.

5. SUMMARY

Lysozymes can be classified on the basis of their sequence homology into 4 distinct groups: (i) chicken, (ii) goose, (iii) phage and (iv) *Chalaropsis*-type lysozymes. The three-dimensional structures of chicken-, goose- and phage-type lysozymes are already known and extensively studied but no detailed structure was known for any *Chalaropsis*-type lysozyme.

Cellosyl is a bacterial lysozyme from *Streptomyces coelicolor*. Being a member of the *Chalaropsis*-type lysozymes, it differs from other lysozymes in molecular weight, amino-acid composition and substrate specificity. The protein possesses both β -1,4-*N*-acetylmuramidase and β -1,4-*N*,6-*O*-diacetylmuramidase activity, which is of considerable interest since Cellosyl is able to degrade cell walls of *Staphylococcus aureus* and other bacteria which are not hydrolysed by chicken-, goose- and phage-type lysozymes.

Cellosyl could be crystallised in two different crystal forms: Rod-shaped crystals which were assigned to the hexagonal space group $P6_122$, and platelet-like crystals belonging to the monoclinic space group $C2$. The three-dimensional structure of the monoclinic crystal form of Cellosyl was solved with the MIRAS technique using individual osmium and mercury derivatives as well as an osmium-mercury double derivative. The structure was initially determined to 1.65 Å and refined to an *R*-factor of 15.2 %. The enzyme is comprised of a single domain that folds into an irregular TIM-barrel. All β -strands of the $(\beta/\alpha)_5\beta_3$ barrel are arranged parallel to one another, except strand β_8 which is in antiparallel orientation to the neighbouring strands β_1 and β_7 .

The crystal structure of Cellosyl in the hexagonal crystal form was determined by the method of molecular replacement using the structure of the monoclinic crystal form as search model. The hexagonal structure was refined to a resolution of 2.32 Å and an *R*-factor of 17.6 %. As expected, both structures exhibit the same overall fold. However, in contrast to the monoclinic crystals, the hexagonal crystals contain two copies of the monomer in the asymmetric unit.

A very important part of this work was the elucidation of the three-dimensional structure of Cellosyl at atomic resolution. Only in rare cases it is feasible to collect data beyond 1.0 Å resolution from protein crystals. However, it was possible to collect a complete data set on the monoclinic crystal form to 0.83 Å resolution using a synchrotron radiation X-ray source. The large number of experimental observations allowed refinement of the structure with anisotropic displacement parameters, which lowered the crystallographic *R*-

factor by more than 4 %. The refinement converged with excellent statistics, e.g. an *R*-factor of 9.63 %. Thus, the crystal structure of Cellosyl is among the highest resolution TIM-barrel fold structures determined to date, which is important considering the fact that roughly 10 % of all known enzyme structures assume the TIM-barrel fold. The structure revealed a number of interesting features. 13 % of the residues adopt multiple conformations, including Ser24 whose side chain could be unambiguously modelled in three alternative conformations.

Moreover, the high-resolution structure yields very detailed insight into the active site architecture of the enzyme. The substrate-binding groove of Cellosyl leads into a deep hole of highly negative electrostatic potential, which is lined by two pairs of carboxylic acids. Within these pairs, the proposed catalytically active residues Glu100 and Asp9 are stabilised by low barrier hydrogen bonds to Asp98 and Asp198, respectively. The distance of 9.5 Å between Glu100 and Asp9 suggests a mechanism involving inversion of the anomeric configuration.

The structure of Cellosyl exhibits a new lysozyme fold and represents the first structurally characterised member of the *Chalaropsis*-type lysozyme family. Very likely, all lysozymes of this family, ranging from bacteria, over bacteriophages and fungi, have the same overall fold. Whilst β/α barrels are widely believed to have arisen through divergent evolution from one common ancestor, the highly unusual structure of Cellosyl suggests that there may have been more than one such ancestor. Besides the Ch-type lysozymes, the only other β/α barrel known to have an antiparallel strand in the eight-stranded sheet is enolase. Perhaps Cellosyl and enolase share a common ancestor, from which they evolved through a series of mutation and cyclic permutation events.

6. ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurde die röntgenkristallographische Untersuchung von Cellosyl, einem Lysozym aus dem Bakterium *Streptomyces coelicolor*, beschrieben. Lysozyme können auf Grund ihrer Aminosäuresequenz in 4 unterschiedliche Familien eingeordnet werden: (i) Hühner-, (ii) Gänse-, (iii) Phagen- and (iv) *Chalaropsis*-Typ Lysozyme. Die dreidimensionalen Strukturen von Hühner-, Gänse-, und Phagentyp Lysozymen sind bekannt und gut erforscht, aber es gibt noch keine detaillierte Beschreibung der Struktur eines *Chalaropsis*-Typ Lysozymes.

Cellosyl ist ein bakterielles Lysozym aus der *Chalaropsis*-Typ-Familie und unterscheidet es sich von anderen Lysozymfamilien in Molekulargewicht, Aminosäurekomposition und Substratspezifität. Das Enzyme weist sowohl β -1,4-*N*-acetylmuramidase- als auch β -1,4-*N*,6-*O*-diacetylmuramidase-Aktivität auf. Damit ist Cellosyl in der Lage, Zellwände von Bakterien, wie z.B. *Staphylococcus aureus*, zu hydrolysieren, die gegenüber Hühner-, Gänse-, und Phagentyp Lysozymen resistent sind.

Es gelang Cellosyl mittels der „Hanging-Drop“ Methode, in Abhängigkeit der gewählten Kristallisationsbedingungen, in zwei unterschiedlichen Habitus zu kristallisieren. Von beiden Kristallformen wurden native Datensätze gesammelt. Die stiftförmigen Kristalle gehören zur hexagonalen Raumgruppe $P6_122$, während die plättchenförmigen Kristalle zur monoklinen Raumgruppe $C2$ gehören.

Zur Bestimmung der Phasen wurden die monoklinen Cellosylkristalle in drei unterschiedliche Schwermetallkomplexe überführt: jeweils individuelle Osmium- und Quecksilber-Derivate und ein gemischtes Osmium-Quecksilber-Doppelderivat. Die Phasen konnten mit der Methode des multiplen isomorphen Ersatzes, unter Ausnutzung der anomalen Streuung der Schweratome, bestimmt werden. Die Struktur wurde zu einer Auflösung von 1.65 Å und einem *R*-Faktor von 15.2 % verfeinert.

Das Enzym besteht aus einer einzelnen Domäne, welche sich zu einem irregulären TIM-Barrel faltet. Alle β -Stränge des $(\beta/\alpha)_5\beta_3$ -Fasses sind parallel zueinander angeordnet, mit Ausnahme von β -Strang 8, welcher antiparallel zu den Strängen β_1 and β_7 ausgerichtet ist.

Die Struktur von Cellosyl in der hexagonalen Kristallform wurde mit der Methode des molekularen Ersatzes bestimmt, wobei die Struktur der monoklinen Kristallform als Suchmodell genutzt wurde. Das Strukturmodell der hexagonalen Kristalle wurde zu einer

Auflösung von 2.32 Å und einem *R*-Faktor von 17.6 % verfeinert. Wie vermutet zeigen, beide Strukturen dieselbe räumliche Faltung. Allerdings enthalten die hexagonalen Kristalle, im Gegensatz zu den monoklinen Kristallen, zwei Kopien des Monomers in der asymmetrischen Einheit.

Weiterhin wurde ein vollständiger nativer Datensatz der monoklinen Kristallform, mit einer atomaren Auflösung von 0.83 Å, an einer Synchrotronstrahlungsquelle gesammelt. Die große Anzahl der experimentell beobachteten Reflexe im Verhältnis zu den benötigten Parametern, erlaubte die Verfeinerung der Struktur mit anisotropen Auslenkungsparametern. Durch die exzellente Übereinstimmung des Modells mit den gemessenen Daten war es möglich, die Struktur bis zu einem *R*-Faktor von 9.63 % zu verfeinern. Die Struktur von Cellosyl ist somit eine der höchstaufgelötesten TIM-Barrel-Strukturen die bekannt ist. Dies ist von Bedeutung, da ungefähr 10 % aller bekannten Enzymstrukturen die TIM-Barrel-Faltung aufweisen. Die hochaufgelöste Struktur zeigt eine Reihe von sehr interessanten Eigenschaften. So liegen z.B. 13 % aller Reste in multiplen Konformationen vor, einschließlich Ser24, dessen Seitenkette eindeutig in drei verschiedenen Konformationen modelliert werden konnte.

Außerdem gibt die Struktur bei atomarer Auflösung detaillierten Einblick in die Architektur des aktiven Zentrums des Enzyms. Die Substratbindespalte von Cellosyl führt zu einem tiefen Loch von sehr negativen elektrostatischem Potenzial, an dessen oberen Rand zwei gegenüberliegende Paare von Carbonsäuren angeordnet sind. In diesen Paaren werden die mutmaßlichen aktiven Reste Glu100 und Asp9 durch „Low-barrier“-Wasserstoffbrückenbindungen zu Asp98 und Asp198 respektive, stabilisiert. Der räumliche Abstand von 9.5 Å zwischen Glu100 and Asp9 deutet auf einen invertierenden Mechanismus der Spaltung der glycosidischen Bindung hin.

Die Struktur von Cellosyl weist eine neue Faltung auf und repräsentiert das erste strukturell charakterisierte Mitglied der *Chalaropsis*-Typ-Lysozyme. Höchstwahrscheinlich haben alle Lysozyme dieser Familie, die in Bakterien, Bakteriophagen und Pilzen vorkommen, die gleiche generelle Faltung.

Während davon ausgegangen wird, dass sich β/α -Fässer durch divergente Evolution von einem gemeinsamen Vorfahren entwickelt haben, zeigt die ungewöhnliche Struktur von Cellosyl, dass es mehr als einen Vorfahren gegeben haben muss. Neben den *Chalaropsis*-Typ Lysozymen ist nur ein weiteres Enzym, Enolase, mit einem antiparallelen Strang im 8-blättrigen Faltblatt bekannt. Es ist deshalb möglich, dass Cellosyl und Enolase einen

gemeinsamen Vorfahren haben, aus dem sie sich durch eine Reihe von Mutationen und cyclischer Permutation entwickelt haben.

7. REFERENCES

- Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Cryst. D* **52**, 30-42
- Amyes, T.L., and Jencks W.P. (1989) Lifetimes of oxocarbenium ions in aqueous-solution from common ion inhibition of the solvolysis of α -azido ethers by added azide ion. *J. Am. Chem. Soc.* **111**, 7888-7900
- Babbitt, P.C., and Gerlt, J.A. (1997) Understanding enzyme superfamilies - chemistry as the fundamental determinant in the evolution of new catalytic activities. *J. Biol. Chem.* **272**, 30591-30594
- Banner, D.W., *et al.*, and Waley, S.G. (1975) Structure of chicken muscle triosephosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609-614
- Beintema, J.J., and Terwisscha van Scheltinga, A.C. (1996) Plant lysozymes. *Experientia Suppl.* **75**, 75-86 (In: *Lysozymes: Model Enzymes in Biochemistry and Biology*. ed. by P.Jollès: 1996 Birkhäuser Verlag Basel / Switzerland)
- Birr, E., Wohlleben, W., Aufderheide, K., Schneider, T., Pühler, A., Bräu, B., Marquardt, R., Wöhner, G., Präve, P., and Schlingmann, M. (1989) Isolation and complementation of mutants of *Streptomyces coelicolor* "Müller" DMS3030 deficient in lysozyme production. *Appl. Microbiol. Biotechnol.* **30**, 358-363
- Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C., and Sarma, V.R. (1965) Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**, 757-761
- Bloomfield, A.L. (1919) The fate of bacteria introduced in the upper air passages. *Bull. Johns Hopkins Hosp.* **30**, 317-322
- Blow, D.M., and Crick, F.H.C. (1959) The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **12**, 794-802
- Bokma, E., van Koningsveld, G.A., Jeronimus-Stratingh, M., and Beintema, J.J. (1997) Hevamine, a chitinase from the rubber tree *Hevea brasiliensis*, cleaves peptidoglycan between the C-1 of *N*-acetylglucosamine and C-4 of *N*-acetylmuramic acid and therefore is not a lysozyme. *FEBS Letters* **411**, 161-163

- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Anal. Biochem.* **72**, 248-254
- Brändén, C.I. (1991) The TIM barrel – the most frequently occurring folding motif in proteins. *Curr. Opin. Struct. Biol.* **1**, 978-983
- Bräu, B., Hilgenfeld R., Schlingmann, M., Marquardt, R., Birr, E., Wohlleben, W., Aufderheide, K., and Pühler, A. (1991) Increased yield of a lysozyme after self-cloning of the gene in *Streptomyces coelicolor* "Müller". *Appl. Microbiol. Biotechnol.* **34**, 481-487
- Brumfitt, W., Wardlaw, A.C., and Park, T.J. (1958) Development of lysozyme resistance in *Micrococcus lysodeikticus* and its association with an increased *O*-acetyl content of the cell wall. *Nature* **181**, 1783-1784
- Brumfitt, W. (1959) The mechanism of development of resistance to lysozyme by some Gram-positive bacteria and its results. *Br. J. Exp. Pathol.* **40**, 441-451
- Brünger, A.T. (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475
- Brünger, A.T., *et al.*, and Warren, G.L. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination. *Acta Cryst. D* **54**, 905-921
- Canfield, R.E. (1963) The amino acid sequence of egg-white lysozyme. *J. Biol. Chem.* **238**, 2698-2707
- CCP4 (1994) – see Collaborative Computational Project, Number 4 (1994)
- CCP4 tutorial <http://www.ccp4.ac.uk/dist/ccp4i/help/modules/appendices/irbathtutorial/irbath98.html>
- Chang, J-J., Holladay, L.A., and Hash, J.H. (1979) The *N,O*-diacetylmuramidase of *Chalaropsis* species. Circular dichroism studies. *J. Biol. Chem.* **254**, 7772-7777
- Chothia, C. and Lesk, A.M. (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5**, 823-826
- Clarke, A.J., and Dupont, C. (1992) *O*-acetylated peptidoglycan: its occurrence, pathological significance, and biosynthesis. *Can. J. Microbiol.* **38**, 85-91
- Cleland, W.W., Frey, P.A., and Gerlt, J.A. (1998) The low barrier hydrogen bond in enzymatic catalysis. *J. Biol. Chem.* **273**, 25529-25532
- Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**, 760-763

- Copley, R.R., and Bork, P. (2000) Homology among ($\beta\alpha$)₈ barrels: Implications for the evolution of metabolic pathways. *J. Mol. Biol.* **303**, 627-640
- Dauter, Z., Lamzin, V.S., and Wilson, K.S. (1995) Proteins at atomic resolution. *Curr. Opin. Struct. Biol.* **5**, 784-790
- Dauter, Z., Lamzin, V.S., and Wilson, K.S. (1997) The benefits of atomic resolution. *Curr. Opin. Struct. Biol.* **7**, 681-688
- Davies, G. J., Sinnott, M.L., and Withers, S.G. (1998) Glycosyl transfer. In: Comprehensive Biological Catalysis, London, Academic Press Ltd, 119-208
- De LaFortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472- 494
- Drenth, J. (1999) Principles of protein X-ray crystallography. Springer-Verlag New York
- Engh, R.A., and Huber, R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392-400
- Esnouf, R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132-134
- Esposito, L., Vitagliano, L., Sica, F., Sorrentio, G., Zagari, A., and Mazzarella, L. (2000) The ultrahigh resolution crystal structure of ribonuclease A containing an isoaspartyl residue: Hydration and stereochemical analysis. *J. Mol. Biol.* **297**, 713-732
- Fani, R., Liò, P., Chiarelli, I., and Bazzicalupo, M. (1994) The evolution of the histidine biosynthetic genes in prokaryotes: A common ancestor of the *hisA* and *hisF* genes. *J. Mol. Evol.* **38**, 489-495
- Farber, G.K., and Petsko, G.A. (1990) The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15**, 228-234
- Fastrez, J. (1996) Phage lysozymes. *Experientia Suppl.* **75**, 35-64 (In: Lysozymes: Model Enzymes in Biochemistry and Biology. ed. by P.Jollès: 1996 Birkhäuser Verlag Basel / Switzerland)
- Felch, J.W., Inagami, T., and Hash, J.H. (1975) The *N,O*-diacetylmuramidase of *Chalaropsis* species. V. The complete amino acid sequence. *J. Biol. Chem.* **250**, 3713-3720
- Fleming, A. (1922) On a remarkable bacteriolytic element found in tissues and secretions. *Proc. Roy. Soc. London Ser. B* **93**, 306-317
- Fleming, A. (1929) Arris and Gale Lecture on Lysozyme. A bacteriolytic ferment found normally in tissues and secretions. *The Lancet*, 217-220
- Fleming, A. (1932) Lysozyme. *Proc. Roy. Soc. Med. London* **26**, 71-84

- Fouche, P.B., and Hash, J.H. (1978) The *N,O*-diacetylmuramidase of *Chalaropsis* species. Identification of aspartyl and glutamyl residues in the active site. *J. Biol. Chem.* **253**, 6787-6793
- French, G.S., and Wilson, K.S. (1978) On the treatment of negative intensity observations. *Acta Cryst. A* **34**, 517-525
- Garcia, E., Garcia, J.L., Garcia, P., Arraras, A., Sanchez-Puelles, J.M., and Lopez, R. (1988) Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Proc. Natl. Acad. Sci. USA* **85**, 914-918
- Garcia, P., Garcia, J.L., Garcia, E., Sanchez-Puelles, J.M., and Lopez, R. (1990) Modular organisation of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. *Gene* **86**, 81-88
- Gerlt, J.A. (2000) New wine from old barrels. *Nature Struct. Biol.* **7**, 171-173
- Gerlt, J.A., and Raushel, F.M. (2003) Evolution of function in $(\beta/\alpha)_8$ -barrel enzymes. *Curr. Opin. Chem. Biol.* **7**, 252-264
- Ghuysen, J.-M. (1968) Use of bacteriolytic enzymes in determination of wall structure and their role in cell metabolism. *Bacteriol. Reviews* **32**, 425-464
- Grütter, M.G., Weaver, L.H., and Matthews B.W. (1983) Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? *Nature* **303**, 828-831
- Harada, S., Sarma, R., Kakudo, M., Hara, S., and Ikenaka, T. (1981) The three-dimensional structure of the lysozyme produced by *Streptomyces erythraeus*. *J. Biol. Chem.* **256**, 11600-11602
- Harada, S., Kitadokoro, K., Fujii, T., Kai Y., and Kasai N. (1989) Preliminary X-ray crystallographic study of lysozyme produced by *Streptomyces globisporus*. *J. Mol. Biol.* **207**, 851-852
- Hash, J.H., and Rothlauf, M.V. (1967) The *N,O*-diacetylmuramidase of *Chalaropsis* species. I. Purification and crystallization. *J. Biol. Chem.* **242**, 5586-5590
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309-316
- Henrissat, B., and Bairoch A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**, 781-788
- Henrissat, B., and Bairoch A. (1996) Updating of the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695-696
- Henrissat, B. (1998) CAZy – Carbohydrate-Active enZYmes. <http://afmb.cnrs-mrs.fr/CAZY>

- Hermoso, J.A., Monterroso, B., Albert, A., Galan, B., Ahrazem, O., Garcia, P., Martinez-Ripoll, M., Garcia, J.L., and Menendez, M. (2003) Structural basis for selective recognition of pneumococcal cell wall by modular endolysin from phage Cp-1. *Structure* **11**, 1239-1249
- Hilge, M., Gloor, S.M., Rypniewski, W., Sauer, O., Heightman, T.D., Zimmerman, W., Winterhalter, K., and Piontek, K. (1998) High-resolution native and complex structures of thermostable β -mannase from *Thermomonospora fusca* – substrate specificity in glycosyl hydrolase family 5. *Structure* **6**, 1433-1444
- Hilgenfeld, R., Liesum, A., Storm, R., and Plaas-Link, A. (1992) Crystallization of two bacterial enzymes on an unmanned space mission. *J. Cryst. Growth* **122**, 330-336
- Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A., and Sterner R. (2001) Dissection of a $(\beta/\alpha)_8$ -barrel enzyme into two folded halves. *Nat. Struc. Biol.* **8**, 32-36
- Höltje, J.-V. (1996) Bacterial lysozymes. *Experientia Suppl.* **75**, 65-74 (In: Lysozymes: Model Enzymes in Biochemistry and Biology. ed. by P.Jollès: 1996 Birkhäuser Verlag Basel / Switzerland)
- Hollis, T., Monzingo, A.F., Bortone, K., Ernst, S., Cox, R., and Robertus, J.D. (2000) The X-ray structure of a chitinase from the pathogenic fungus *Coicidioides immitis*. *Protein Science* **9**, 544-551
- Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123-138
- Holm, L., and Sander, C. (1996) Mapping the protein universe. *Science* **273**, 595-603
- Hughey, V.L. and Johnson, E.A. (1987) Antimicrobial activity of lysozyme against bacteria involved in food spoilage and food-borne disease. *Appl. Environ. Microbiol.* **53**, 2165-2170
- Jancarik, J., Pufan, R., Hong, C., Kim, S.H., and Kim R. (2004) Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. *Acta Cryst. D* **60**, 1670-1673
- Jia, J., Huang, W., Schorken, U., Sahm, H., Sprenger, G.A., Lindqvist, Y., and Schneider, G. (1996) Crystal structure of transaldolase B from *Escherichia coli* suggests a circular permutation of the α/β barrel within the class I aldolase family. *Structure* **4**, 715-724
- Jollès, J., Jauregui-Adell, J., Bernier, I., and Jollès, P. (1963) La structure chimique du lysozyme de blanc d'oeuf de poule : étude détaillée. *Biochim. Biophys. Acta* **78**, 668-689

- Jollès, P., and Jollès, J. (1984) What's new in lysozyme research? *Mol. Cell. Biochem.* **63**, 165-189
- Jollès, P. (1996) From the discovery of lysozyme to the characterization of several lysozyme families. *Experientia Suppl.* **75**, 3-5 (In: Lysozymes: Model Enzymes in Biochemistry and Biology. ed. by P.Jollès: 1996 Birkhäuser Verlag Basel / Switzerland)
- Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for building protein models into electron density maps and for the location of errors in these models. *Acta Cryst. A* **47**, 101- 119
- Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637
- Kantardjieff, K.A., and Rupp B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science* **12**, 1865-1871
- Kissinger, C.R., Gehlhaar, D.K., and Fogel D.B. (1999) Rapid automated molecular replacement by evolutionary search. *Acta Cryst. D* **55**, 484-491
- Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946-950
- Kuroki, R., Weaver, L.H., and Matthews, B.W. (1993) A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262**, 2030-2033
- Kuroki, R., Weaver, L.H., and Matthews, B.W. (1999) Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. *Proc. Natl. Acad. Sci. USA* **96**, 8949-8954
- Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **289**, 1546-1550
- Laschtschenko, P. (1909) Über die keimtötende und entwicklungshemmende Wirkung von Hühnereiweiß. *Z. Hyg. Infekt. Krankheiten* **64**, 419-426
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283-291

- Lebioda, L., Stec, B., and Brewer, J.M. (1989) The structure of yeast enolase at 2.25 Å resolution. An 8-fold β/α -barrel with a novel $\beta\beta\alpha\alpha(\beta/\alpha)_6$ topology. *J. Biol. Chem.* **264**, 3685-3693
- Lesk, A.M., Brändén, C.I., and Chothia, C. (1989) Structural principles of α/β barrel proteins: the packing of the interior of the sheet. *Proteins* **5**, 139-148
- Lichenstein, H.S., Hastings, A.E., Langley, K.E., Mendiaz, E.A., Rohde, M.F., Elmore, R., and Zukowski, M.M. (1990) Cloning and nucleotide sequence of the *N*-acetylmuramidase M1-encoding gene from *Streptomyces globisporus*. *Gene* **88**, 81-86
- Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E, Murzin, A.G., and Chothia, C. (2000) SCOP: a structural classification of proteins database. *Nucleic Acids Res.* **28**, 257-259
- Lo Conte, L., Brenner, S.E, Hubbard, T.J., Chothia, C., and Murzin, A.G. (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids Res.* **30**, 264-267
- Luzzati, V. (1952) Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Cryst.* **5**, 802-810
- Lyne, J.E., Carter, D.C., He, X.-M., Stubbs, G., and Hash, J.H. (1990) Preliminary crystallographic examination of a novel fungal lysozyme from *Chalaropsis*. *J. Biol. Chem.* **265**, 6928-6930
- Martin, A.C., Lopez, R., and Garcia, P. (1996) Analysis of the complete nucleotide sequence and functional organization of the genome of *Streptococcus pneumoniae* bacteriophage Cp-1. *J. Virol.* **70**, 3678-3687
- Mata, M., Trautwetter, A., Luthaud, G., and Ritzenthaler, P. (1986) Thirteen virulent and temperate bacteriophages of *Lactobacillus bulgaris* and *Lactobacillus lactis* belong to a single DNA homology group. *Appl. Environ. Microbiol.* **52**, 812-818
- Matsumura, I., and Kirsch, J.F. (1996) Is aspartate 52 essential for catalysis by chicken egg white lysozyme? The role of substrate-assisted hydrolysis. *Biochemistry* **35**, 1881-1889
- Matthews, B.W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497
- McCarter, J.D., and Withers, S.G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885-892
- McPherson, A. (1982) The preparation and analysis of protein crystals. Wiley, New York
- Merritt, E.A. (1996) Anomalous Scattering Coefficients. http://www.bmsc.washington.edu/scatter/AS_form.html

- Merritt, E.A., and Bacon D.J. (1997) Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505-524
- Merritt, E.A. (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Cryst. D* **55**, 1109-1117
- Monterroso, B., Albert, A., Martinez-Ripoll, M., Garcia, P., Garcia, J.L., Menendez, M., and Hermoso, J.A. (2002) Crystallization and preliminary X-ray diffraction studies of the complete modular endolysin from Cp-1, a phage infecting *Streptococcus pneumoniae*. *Acta Cryst. D* **58**, 1487-1489
- Morita, T., Hara, S., and Matsushima, Y. (1978) Purification and characterization of lysozyme produced by *Streptomyces erythraeus*. *J. Biochem.* **83**, 893-903
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the Maximum-Likelihood Method. *Acta Cryst. D* **53**, 240-255
- Nagano, N., Hutchinson, E.G., and Thornton, J.M. (1999) Barrel structures in proteins - automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci.* **8**, 2072-2084
- Nagano, N., Orengo, C.A., and Thornton, J.M. (2002) One fold with many functions: The evolutionary relationship between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321**, 741-765
- Nicholls, A., Sharp, K.A., and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326
- Pearl, F.M., Lee, D., Bray, J.E., Sillitoe, I., Todd, A.E., Harrison, A.P., Thornton J.M., and Orengo, C.A. (2000) Assigning genomic sequences to CATH. *Nucl. Acid Res.* **28**, 277-282
- Perrakis, A., Morris, R.M., and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458-463
- Phillips, D.C. (1966) The three-dimensional structure of an enzyme molecule. *Sci. Am.* **215**, 78-90
- Powell M.D.J. (1977) Restart procedures for conjugate gradient method. *Math. Programming* **12**, 241-254
- Ramachandran, G.N., and Sasisekharan, V. (1968) Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283-438

- Rau, A., Hogg, T., Marquardt, R., and Hilgenfeld, R. (2001) A new lysozyme fold. Crystal structure of the muramidase from *Streptomyces coelicolor* at 1.65 Å resolution. *J. Biol. Chem.* **276**, 31994-31999
- Read, R.J. (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140-149
- Reardon, D., and Farber, G.K. (1995). The structure and evolution of α/β barrel proteins. *FASEB J.* **9**, 497-503
- Rettger, L.F., and Sperry, J.A. (1912) The antiseptic and bactericidal properties of egg white. *J. Med. Res.* **26**, 55-64
- Reynolds J.A., Tanford C. (1970) The gross conformation of protein-sodium dodecyl sulfate complexes. *J. Biol. Chem.* **245**, 5161-5165
- Riboldi-Tunnicliffe, A. and Hilgenfeld, R. (1999) Cryocrystallography with oil - an old idea revived. *J. Appl. Cryst.* **32**, 1003-1005
- Rhodes, G. (2000) Crystallography made crystal clear. A guide for users of macromolecular models. 2nd edition, Academic Press, 107-127
- Ronda, C., Lopez, R., and Garcia, E. (1981) Isolation and characterization of a new bacteriophage, Cp-1, infecting *Streptococcus pneumoniae*. *J. Virol.* **40**, 551-559
- Rye, C.S., and Withers, S.G. (2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573-580
- Salton, M.R.J. (1952) Cell wall of *Micrococcus lysodeikticus* as the substrate of lysozyme. *Nature* **170**, 746-747
- Sanz, J.M., Garcia, P., and Garcia, J.L. (1992) Role of Asp-9 and Glu-36 in the active site of the pneumococcal CPL1 lysozyme: an evolutionary perspective of lysozyme mechanism. *Biochemistry* **31**, 8495-8499
- Schomaker, V. and Trueblood, K.N. (1968) On the rigid body motion of molecules in crystals. *Acta Cryst. B* **24**, 63-76
- Schomaker, V. and Trueblood, K.N. (1998) Correlation of internal torsional motion with overall molecular motion in crystals. *Acta Cryst. B* **54**, 507-514
- Sergeev, Y., and Lee, B. (1994) Alignment of β -barrels in $(\beta/\alpha)_8$ proteins using hydrogen-bonding pattern. *J. Mol. Biol.* **244**, 168-182
- Shapiro A.L., Vinuela E., Maizel J.V. (1967) Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochem. Biophys. Res. Commun.* **28**, 815-820

- Sheldrick, G.M. (1990) Phase annealing in *SHELX-90*: direct methods for larger structures. *Acta Cryst. A* **46**, 467-473
- Sheldrick, G.M., and Schneider, T.R. (1997) *SHELXL*: high-resolution refinement. *Methods Enzymol.* **277**, 319-343
- SHELX Workshop, St. Paul ACA meeting (2000)
www.scl.kyoto-u.ac.jp/scl/appli/appli_manual/shelx/aca2000.pdf
- Shiba, T., Harada, S., Sugawara, H., Naitow, H., Kai, Y., and Satow, Y. (2000) Crystallization and preliminary X-ray analysis of a bacterial lysozyme produced by *Streptomyces globisporus*. *Acta Cryst. D* **56**, 1462-1463
- Simpson, A.J.G., *et al.*, and Setubal, J.C. (2000) The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature* **406**, 151-157
- Stryer, L. (1996) Biochemie, Spektrum Akademischer Verlag Heidelberg·Berlin·Oxford, 223-228
- Terwilliger, T.C., and Benrendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Cryst. D* **55**, 849-861
- Trueblood, K.N., Bürgi, H.-B., Burzlaff, H., Dunitz, J.D., Grammacioli, C.M., Schulz, H.H., Shmueli, U., and Abrahams, S.C. (1996) Atomic displacement nomenclature. Report of a subcommittee on atomic displacement parameter nomenclature. *Acta Cryst. A* **52**, 770-781
- Tuomanen, E., and Tomasz, A. (1986) Induction of autolysis in nongrowing *Escherichia coli*. *J. Bacteriol.* **167**, 1077-1080
- Vasella, A., Davies, G.J., and Böhm, M. (2002) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **6**, 619-629
- Vocadlo, D.J., Davies, G.J., Laine, R., and Withers, S.G. (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* **412**, 835-838
- Walsh, M.A., Schneider, T.R., Sieker, L.C., Dauter, Z., Lamzin, V.S., and Wilson, K.S. (1998) Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Cryst. D* **54**, 522-546
- Wang, Z., Luecke, H., Yao, N., and Quioco, F.A. (1997) A low energy short hydrogen bond in very high resolution structures of protein receptor-phosphate complexes. *Nat. Struct. Biol.* **4**, 519-522
- Weaver, L.H., Grütter, M.G., Remington, S.J., Gray, T.M., Isaacs, N.W., and Matthews, B.W. (1985) Comparison of goose-type, chicken-type, and phage-type lysozymes illustrates

- the changes that occur in both amino acid sequence and three-dimensional structure during evolution. *J. Mol. Evol.* **21**, 97-111
- Weaver, L.H., Grütter, and Matthews, B.W. (1995) The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the “goose-type” lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* **245**, 54-68
- Weiss, M.S., and Hilgenfeld, R. (1997) On the use of the merging *R* factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **30**, 203-205
- Weiss, M.S. (2001) Global indicators of X-ray data quality. *J. Appl. Cryst.* **34**, 130-135
- Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001) Use of TLS parameters to model anisotropic displacement in macromolecular refinement. *Acta Cryst. D* **57**, 122-133
- Wilson, A.J.C. (1942) Determination of absolute from relative X-ray intensity. *Nature* **150**, 151-152
- Young, K.D., Anderson, R.J., and Hafner, R.J. (1989) Lysis of *Escherichia coli* by bacteriophage ϕ X174 E protein: inhibition of lysis by heat shock. *J. Bacteriol.* **171**, 4334-4341
- Zechel, D.L., and Withers, S.G. (2001) Dissection of nucleophilic and acid-base catalysis in glycosidases. *Curr. Opin. Chem. Biol.* **5**, 643-649

8. APPENDIX

Table 8.-1 Hydrogen-bonding interactions of the chloride ions within the structure of Cellosyl in the monoclinic crystal form. The table was generated with the program *CONTACT* (CCP4, 1994).

chloride ion	target atoms	H-bonding distance [Å]
Cl1	Arg189 N	3.18
	Arg199 NH1	3.30
Cl2	Wat148	3.14
	Trp179 NE1	3.22
	Wat 16	3.10
Cl3	Wat 47	3.18
	Wat 149	3.42
	Wat341	3.25
	Arg133 NE	3.03
	Arg133 NH2	3.48
Cl4	Asp134 N	3.12
	Asn215 ND2	3.50
	Wat 43	3.18
	Wat196	3.36
Cl5	Arg199 NH1	3.30
	Ala207 N	3.20
	Wat148	3.34
	Wat164	2.66
	Wat297	2.95
Cl6	Asn152 N	3.07
	Wat96	3.39
	Wat131	3.01
	Wat173	3.48
Cl7	Wat317	3.43
	Asn102 ND2	3.27
	Wat78	3.07
Cl8	Wat347	3.19
	Ala74 O	3.41
	Ser70 OG	3.27
	Wat70	2.99
Cl9	Wat313	2.95
	Wat14	3.25
	Wat107	3.14
	Wat250	2.98
	Wat382	3.00

Table 8.-2 Hydrogen-bonding interactions of the sulphate ions within the structure of Cellosyl in the hexagonal crystal form. The table was generated with the program *CONTACT* (CCP4, 1994).

sulphate ions		target atoms	H-bonding distance [Å]
SO4 1	O3	Arg199A NH2	2.82
	O4	Arg199A NH1	2.66
SO4 2	O2	Arg189A N	3.49
	O3	Arg199A NH1	3.51
	O4	Arg199A NH1	2.71
	O4	Arg189A N	2.75
SO4 3	O4	Arg189A O	3.20
	O3	Gly192A N	2.98
	O4	Gly192A N	3.16
SO4 4	O1	Ser142A OG	3.20
	O3	Wat141	2.80
	O4	Wat141	3.24
	O1	Ser142A OG	3.20
	O3	Wat141	2.80
	O3	Thr140A OG1	3.19
	O4	Wat141	3.24
	O3	Thr140A OG1	3.19
SO4 5	O1	Wat36	2.98
	O2	Ser11A OG	3.16
	O2	Wat36	2.62
	O3	Tyr62A OH	2.79
	O4	Lys33A NZ	2.42
SO4 6	O4	Gly192B N	3.51
SO4 7	O2	Arg 126A NH1	3.19
	O4	Arg126A NH1	2.81
SO4 8	O4	Arg117A NH1	3.30
SO4 9	O2	Arg199B NH2	2.81
	O3	Arg199B NH1	3.23
SO4 10	O1	Arg126A NH2	2.71
	O2	Arg126A NH2	2.73
	O3	Arg126A NH2	3.30
SO4 11	O2	Arg199B NH1	2.69
	O2	Arg189B N	2.98
	O3	Arg189B N	3.29
SO4 12	O1	Ser142B OG	3.42
	O2	Thr140B OG1	2.77
	O2	Wat82	2.71
	O3	Wat82	3.31
SO4 13	O2	Arg126B NE	2.89
	O2	Arg126B NH2	3.11
	O3	Arg126B NH2	2.80
	O4	Arg126B N	3.24
	O4	Arg126B NE	3.17

	O4	Arg126B NH2	3.49
	O4	Wat124	2.62
SO4 14	O2	Wat54	2.95
	O3	Arg117B NH1	3.36
	O3	Arg117B NH2	3.04
	O4	Arg117B NH2	3.17
	O4	Wat54	3.02
SO4 15	O1	Tyr62B OH	2.62
	O2	Lys33B NZ	2.54
	O4	Wat128	2.95
	O4	Wat118	2.70
SO4 16	O1	Arg44A NH1	2.91
	O1	Arg44A NH2	2.38
	O2	Wat191	2.98
	O4	Arg 44A NH1	2.70
	O4	Wat191	3.00
SO4 17	O1	Arg126B NH1	3.23
	O4	Arg126B NH2	2.68
SO4 18	O1	Ser205A OG	2.56
	O1	Arg208A NH1	2.47
	O2	Gly 204A N	3.41
	O2	Arg208A NE	2.87
	O2	Arg208A NH2	3.20
	O2	Arg208A NH1	2.67
	O2	Wat195	3.28
	O3	Wat195	3.46
	O4	Arg208A NE	3.45
	O4	Ser205A N	3.11
	O4	Ser205A OG	3.13
SO4 19	O1	Thr216A O	3.27
	O2	Arg88A NH1	2.82
	O2	Trp179B NE1	3.09
	O2	Wat74	2.76
	O3	Trp179B NE1	2.78
	O4	Wat127	3.28
SO4 20	O1	Arg91A NE	2.54
	O1	Arg91A NH1	3.18
	O1	Arg91A NH2	3.45
	O1	Ser205B OG	2.33
	O2	Ser205B OG	2.77
	O2	Arg208B NH1	3.00
	O3	Asn90A ND2	2.86
	O3	Arg208B NH1	2.20

Acknowledgements

First, I'd like to express my gratitude to Prof. Dr. Rolf Hilgenfeld, for providing me with the opportunity to work in his multinational research group on a both challenging and stimulating project.

I am grateful to Dr. Manfred S. Weiss for his advice and assistance. Discussing problems with you opened new horizons and was very motivating and inspiring. I'd like to thank Dr. Jeroen Mesters, especially for his instant help whenever computer problems occurred (and also for rewriting my input files after you deleted them).

Special thanks to my friends and fellow students in the department of structural biology and crystallography for keeping the spirits high and creating a vibrant atmosphere. Besides your help with scientific problems, I'd like to thank Alan for welcoming me with a hot cup o' tea every morning, Santosh for introducing me to certain aspects of Indian culture, e.g. 'palm', Andre and Tom for discussions beyond science and Tanis for operating a Pringles lid out of my eye and for proofreading this manuscript.

I am also thankful to all technicians, especially Barbara Härtl, for their generous assistance.

I am very grateful to my parents for their constant help and encouragement. Last but not least I'd like to thank my husband Sven for his support and comprehension and his imperturbable love. It's all your fault! This work is dedicated to you.

CURRICULUM VITAE

Name	Astrid Rau, geb. Skirl
Geburtsdatum und -ort	01.06.1974, Saalfeld
Familienstand	verheiratet, 2 Kinder (07.10.2001, 25.08.2003)
Schulische Bildung	
09/1980 – 08/1982	Besuch der POS „Adolf-Reichwein I“ in Jena
09/1982 – 08/1990	Besuch der POS „Adolf Reichwein II“ in Jena - Schule mit erweitertem Russischunterricht
09/1990 – 06/1992	Abitur am Sportgymnasium Jena (Spezialsport: Judo) Abschluss: Abitur (Note: 2,0)
Studium	
10/1992 – 09/1994	Grundstudium Chemie an der Friedrich-Schiller- Universität Jena
10/1994 – 09/1995	2 Auslandssemester im Studiengang Chemie an der Dublin City University, Dublin, Irland
10/1995 – 09/1996	Hauptstudium Chemie an der Friedrich-Schiller- Universität Jena
10/1996 - /1997	2 Auslandssemester zur Anfertigung der Diplomarbeit an der Dublin City University, Dublin, Irland in den Arbeitsgruppen von Prof. Dr. J. G. Vos (DCU, Dublin) und Prof. Dr. D. Walther (FSU, Jena) Thema: Metal complexes as model compounds for natural photosynthesis – Ruthenium polypyridyl complexes with oxalamidine and related ligands (Note: 1,4)
09.09.1997	Abschluss als Diplomchemikerin nach 10 Semestern mit dem Prädikat sehr gut
10/1997 – 03/1998	Weiterbildung Umweltsicherung an der FSU Jena

Promotionsstudium

04/1998 - 09/2001	Doktorandin am Institut für Molekulare Biotechnologie Jena in der Arbeitsgruppe von Prof. Dr. R. Hilgenfeld Thema: The crystal structure of bacterial lysozyme at atomic resolution
10/2001 - 03/2002	Erziehungsurlaub
04/2002 - 07/2003	Weiterführung der Promotionsarbeit am Institut für Molekulare Biotechnologie Jena in der Arbeitsgruppe von Prof. Dr. R. Hilgenfeld
08/2003 – 07/2004	Erziehungsurlaub und Niederschrift der Promotionsarbeit

Berufliche Laufbahn

08/2004 – 09/2004	Mitarbeit in der Bioconcept GmbH
seit 11/2004	Product Development Manager – Protein Crystallography in der Jena Bioscience GmbH

Publikationen

Inorganica Chimica Acta 303 (2000), 206 - 214	M. Ruben, S. Rau, <u>A. Skirl</u> , K. Krause, H. Görls, D. Walther, J. G. Vos: "Novel polypyridyl ruthenium (II) complexes containing oxalamidines as ligands"
The Journal of Biological Chemistry 276 (2001), 31994 - 31999	<u>A. Rau</u> , T. Hogg, R. Marquardt, R. Hilgenfeld: "A new lysozyme fold – Crystal structure of the muramidase from <i>Streptomyces coelicolor</i> at 1,65 Å"

Poster

2 nd Symposium on Protein Structure Analysis for Biomedical Research (1999) Murnau	<u>A. Skirl</u> , R. Hilgenfeld: "Crystallization and Structure Determination of Cellosyl from <i>Streptomyces coelicolor</i> "
DGK Jahrestagung (2001) Bayreuth	<u>A. Rau</u> , T. Hogg, R. Hilgenfeld: "The Crystal Structure of Cellosyl from <i>Streptomyces coelicolor</i> "

Vorträge

Heart of Europe Bio-
Crystallography Meeting (2000)
Poznan

A. Rau*, R. Hilgenfeld: "Crystallization and Structure
Determination of Cellosyl from *Streptomyces
coelicolor*"

European Crystallography
Meeting (2001) Krakau

A. Rau, T. Hogg, M. S. Weiss*, R. Hilgenfeld: "A new
lysozyme fold at ultra-high resolution: 0.83 Å crystal
structure of the muramidase from *Streptomyces
coelicolor*"

Besondere Kenntnisse /

Fertigkeiten

Sprachen

englisch, sehr gute Kenntnisse
russisch, gute Kenntnisse

Computer

z.B. Word, MS Excel, Powerpoint
z.B. UNIX, CS ChemOffice
gängige Internetanwendungen

Jena, den 10.12.2004

Astrid Rau

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

Jena, den 06.12.2004

Astrid Rau

Erklärung zur Bewerbung

Ich erkläre, dass ich mich mit der vorliegenden Arbeit an keiner anderen Hochschule um den akademischen Grad doctor rerum naturalium beworben habe und dass ich weder früher noch gegenwärtig die Eröffnung eines Verfahrens zum Erwerb des o.g. akademischen Grades an einer anderen Hochschule beantragt habe.

Jena, den 06.12.2004

Astrid Rau

