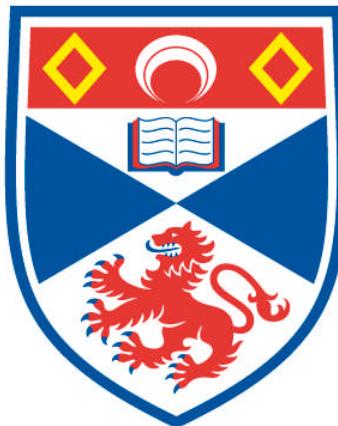


**MECHANISMS OF SILICATE POLYMERISATION,
CARBOHYDRATE EPIMERISATION AND
METALLOPROTEASE INHIBITION**

Thomas Kowatz

**A Thesis Submitted for the Degree of PhD
at the
University of St Andrews**



2009

**Full metadata for this item is available in
Research@StAndrews:FullText
at:**

<http://research-repository.st-andrews.ac.uk/>

**Please use this identifier to cite or link to this item:
<http://hdl.handle.net/10023/771>**

This item is protected by original copyright

**This item is licensed under a
Creative Commons Licence**

**Mechanisms of silicate polymerisation,
carbohydrate epimerisation and
metalloprotease inhibition**

Thomas Kowatz

**A thesis submitted for the degree of Doctor of
Philosophy**

**University of St Andrews
July 2009**

Declaration

I, Thomas Kowatz certify that this thesis, which is approximately 70,000 words and which has been written by me, is a record of work carried out by myself and that it has not been submitted in any previous application for a degree.

Date

Signature of candidate

I was admitted as a research student in January 2005 and as a candidate for the degree of Ph.D. in January 2006; the higher study for which this is a record was carried out at the University of St Andrews between 2005 and 2009.

Date

Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the resolution and regulations appropriate for the degree of Ph.D. in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date

Signature of supervisor

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis.

We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to all or part of printed copy but embargo of all or part of electronic publication of thesis for a period of 1 year on the following ground:

publication would preclude future publication

Date Signature of candidate Signature of supervisor

Abbreviations

Amino acids

Ala (A) – alanine	Leu (L) – leucine
Arg (R) – arginine	Lys (K) – lysine
Asn (N) – asparagine	Met (M) – methionine
Asp (D) – aspartic acid	Phe (F) – phenylalanine
Cys (C) – cysteine	Pro (P) – proline
Gln (Q) – glutamine	Ser (S) – serine
Glu (E) – glutamic acid	Thr (T) – threonine
Gly (G) – glycine	Tyr (Y) – tyrosine
His (H) – histidine	Trp (W) – tryptophan
Ile (I) – isoleucine	Val (V) – valine

4-BPA – 4-biphenylacetyl rings

ADAM – a disintegrin and metalloproteinase

ADAMTS – ADAM with thrombospondin fingerprint sequences

ADP – adenosine diphosphate

ADP-D,D-Hep – ADP-D-glycero-D mannoheptose

ADP-L,D-Hep – ADP-L-glycero-D mannoheptose

AGME (HldD) – ADP- β -L-glycero-D-mannoheptose 6-epimerase

bp – base pair(s)

ΔH – enthalpy

ΔS – entropy

DGD – dTDP-D-glucose 4,6 dehydratase

dTDP – deoxythymidine diphosphate

DTT – Dithiothreitol
ECM – extracellular matrix
EDTA – ethylenediaminetetraacetic acid
ER – Endoplasmic Reticulum
 F_o - F_c – Maximum likelihood / σ_A -weighted F_o - F_c electron density map
 $2F_o$ - F_c – Maximum likelihood / σ_A -weighted $2F_o$ - F_c electron density map
GMER – GDP-4-keto-6-deoxy-D-mannose epimerase/reductase
Hep – Heptose
HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag – Histidine tag
Hpx – hemopexin (haemopexin)
IPTG – Isopropyl β -D-thio-galactopyranoside
ITC – Isothermal Titration Calorimetry
 K_a – binding constant
kbp – kilobase pair(s)
 K_d ($1/K_a$) – dissociation constant
kDa – kiloDalton
 K_i – inhibition constant
L – liter
LBHB – low-barrier hydrogen bond
M – molar
ml – milliliter(s)
 μ l – microlitre(s)
mM – millimolar
 μ M – micromolar
MMP – matrix metalloproteinase
MT-MMPs –membrane-type matrix metalloproteinases
MWCO – Molecular Weight Cut-Off
n – stoichiometry

NAD⁺ – nicotinamide adenine dinucleotide (oxidised form)
NADH – nicotinamide adenine dinucleotide (reduced form)
NADP⁺ – nicotinamide adenine dinucleotide phosphate (oxidised form)
NADPH – nicotinamide adenine dinucleotide phosphate (reduced form)
NCS – Non-crystallographic symmetry
nM – nanomolar
NMR – Nuclear Magnetic Resonance Spectroscopy
PAGE – Polyacrylamide Gel Electrophoresis
PCR – Polymerase Chain Reaction
Pdb (PDB) Protein data bank
PEG – Polyethyleneglycol
PISA – Protein Interfaces, Surfaces and Assemblies
 pK_a – logarithmic measure of acid dissociation constant
PPi – pyrophosphate
RHTS – Recombinant human truncated stromelysin
RmlD – dTDP-6-deoxy-L-*lyxo*-4-hexulose reductase
rmsd – root mean square deviation
SCD – catalytic domain of human stromelysin-1
SDRs – Short-chain dehydrogenases/reductases
SDS – Sodium Dodecyl Sulfate
SH – thiol group
S-S – disulfide bond
SSPF – Scottish Structural Proteomics Facility
TEOS – tetraethoxysilane
TEV – Tobacco Etch Virus
TIMP – Tissue Inhibitor of Metalloproteinases
U – Unit(s)
UDP – Uridine diphosphate
UDP-GlcNAc – UDP-*N*-acetylglucosamine

UDP-ManNAc – UDP-*N*-acetyl- mannosamine

UGE – UDP-Glucose 4-epimerase or UDP-Galactose 4-epimerase

Acknowledgements

First I would like to thank my supervisor, Jim Naismith, for all his help during my PhD and for introducing me to protein crystallography.

I would like to thank my parents and my girlfriend Li (Honey) for being here for me and their support during my PhD.

My thanks also go to our collaborators Michael Fairhead and Chris van der Walle (4SER project), James Morrison and Martin Tanner (AGME project) and Erin Wilfong and Eric Toone (SCD project).

Finally I want to thank my colleagues and friends Gregor, Lester, Kannan, Changjiang, Louise, Stefan, Huanting, Ken, Muse, Kerstin, Wenjian, Jane, Simon, Arif, Stephen and Thomas for their help and spending a nice time in St Andrews.

Abstract

In biotechnology and drug delivery silica materials are of interest but the controlled generation of silicic acid is difficult. To get more insight into the molecular mechanisms that control biosilification, it is important to study the proteins involved in this process. The sponge protein silicatein α synthesises part of the axial filament in the spicules which *in situ* polymerises silicic acid. It has been demonstrated that the polymerisation of siloxanes such as for example tetraethoxysilane (TEOS) can be carried out by both wild type and recombinant silicatein α . Unfortunately, it has not been possible yet to get reasonable amounts of wild type or recombinant silicatein α to perform biophysical studies. The human cysteine protease cathepsin L has almost 50 % identical residues with silicatein α . To get more insight into the mechanism of silica polymerisation, cathepsin L mutants were generated by our collaborators. Those mutants show sequence features and activity specific for silicatein α . The X-ray structure of one of those mutants (mutant 4SER) to 1.5 Å has allowed us to propose a new chemical mechanism for the catalysis of silicic acid polymerisation.

ADP- β -D-glycero-D-mannoheptose and ADP- β -L-glycero-D-mannoheptose are interconverted by the SDR-enzyme ADP- β -L-glycero-D-mannoheptose 6-epimerase (AGME). This epimerisation reaction is the final reaction in the biosynthetic route of the precursor of heptose. Heptose is a part of the inner core of the lipopolysaccharide in Gram-negative bacteria. In mutants which do not have heptose, nonpolar compounds can penetrate more easily through the outer membrane. These mutants also show less pathogenicity. As a consequence the lipopolysaccharide biosynthetic pathway represents an interesting target for antimicrobial compounds. The crystal structure of AGME in complex with ADP- α -glucose has already been solved. From this structure a catalytic mechanism for this enzyme has been proposed with Tyr140 and Lys178 operating as acid/base residues.

The disordered nature of the nucleotide sugar's glucose moiety in the previous structure due to the wrong configuration of the sugar has hindered assignment of a mechanism. The determination of the X-ray structure of AGME Y140F in complex with a substrate in the β -manno configuration (ADP- β -mannose) to 2.4 Å resolution has given new insight into the mechanism of this SDR enzyme. A mechanism is proposed with only Tyr140 operating as catalytic acid/base.

Initially it was thought that MMP-3 participates in the synovitis cascade. Glycoproteins, several parts of the ECM such as fibronectin and laminin and also collagens and procollagens are targets of this matrixin. Furthermore MMP-3 can undergo autocatalysis and is also able to cleave a range of other members of the matrixin family. Matrixins also play an important role in diseases such as cancer, rheumatoid arthritis and osteoporosis. This makes them targets for inhibitor design. Many structures of matrix metalloproteinases, such as stromelysin-1, in complex with various inhibitors have already been solved. The structures of the catalytic domain of MMP-3 in complex with two nonpeptide inhibitors are discussed.

List of Figures

Chapter 1

Figure 1. (after Sajid and McKerrow, 2002) Representation of the binding of a peptide substrate to the catalytic site of a cysteine protease. Amino acids from the peptide are indicated by P and the substrate binding sites of the protease by S. The thiol nucleophile of the catalytic cysteine is indicated by SH, and the corresponding amide bond that is attacked before hydrolysis (scissile bond) is marked with an arrow. The COOH side of the peptide ($P'1, P'2, P'n$) and corresponding substrate binding sites ($S'1, S'2, S'n$) are called prime side, whereas the NH_2 side of the peptide ($P1, P2, Pn$) and corresponding substrate binding sites ($S1, S2, Sn$) are named non-prime side. (A). Representation of the SH of Cys and the imidazole ring of His (B1), and the equilibrium with the thiolate-imidazolium charge relay. The delocalised electron density cloud is also shown (B2).

Figure 2. Ribbon diagram of the structure of human cathepsin S (Pdb accession code 1ms6). Secondary structural elements of the left (L) domain are shown in green, and secondary structural elements of the right (R) domain in cyan. This figure has been produced using PyMol (DeLano, 2007).

Figure 3. Sequence alignment of papain and various cathepsins (based on Turk *et al.*, 2000) The key active sites residues C25 and H159 are highlighted in yellow and marked with asterisks. Other amino acid residues, conserved in all 9 enzymes are shown in red. The sequences are taken from the PROTEIN DATA BANK (PDB): papain from *Carica papaya* (Pdb accession code 9pap); cathepsin L from *Homo sapiens* (Pdb accession code 1mhw); cathepsin V mutant (N108Q, N178D) from *Homo sapiens* (Pdb accession code 1fh0); cathepsin K from *Homo sapiens* (Pdb accession code 1yt7); cathepsin S from *Homo sapiens* (Pdb accession code 1nqc); cathepsin H from *Sus scrofa* (Pdb accession code 8pch); cathepsin C from *Rattus norvegicus* (Pdb accession code 1jqp); cathepsin B from *Homo sapiens* (Pdb accession code 1huc); cathepsin X from *Homo sapiens* (Pdb accession code 1ef7). This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 4. Stereo view showing residues Met75-Gln79 of the prosegment of procathepsin L (Pdb accession code 1cs8), dipping into the substrate binding cleft of the mature part of procathepsin L (Coulombe *et al.*, 1996). Secondary structure elements of the propeptide and the mature part of procathepsin L are shown in cyan and green, respectively. The catalytic active site residues Cys25 (its sulfur is oxidised to sulfonate) and His163, as well as the residues Met75-Gln79 of the prosegment are shown as sticks (carbons in magenta). This figure has been produced using PyMol (DeLano, 2007).

Figure 5. Stereo views of the prosegment of procathepsin L (A) (based on Coulombe *et al.*, 1996) (Pdb accession code 1cs8), α -helices are shown in cyan and loops in magenta, and procathepsin L (B). The prosegment is shown in cyan and cathepsin L in green. The catalytic active site residues Cys25 (its sulfur is oxidised to sulfonate; see Pdb accession code 1cs8 and Chowdhury *et al.*, 2002) and His163, as well as the residues His140-Asp155 of the prosegment binding loop (PBL) (Coulombe *et al.*, 1996) are shown in magenta. This figure has been produced using PyMol (DeLano, 2007).

Figure 6. Stereo view of procathepsins L (Pdb accession code 1cs8) (A) and B (Pdb accession code 1pbh) (B), based on Coulombe *et al.*, 1996. The prosegments of both enzymes are shown in cyan and cathepsins L and B in green. This figure has been produced using PyMol (DeLano, 2007).

Figure 7. Ribbon diagrams of papain from *Carica papaya* (A) (Pdb accession code 9pap), α -helices are shown in cyan β -strands in red and loops in magenta, and human cathepsin L (B) (Pdb accession code 3bc3). α -helices are shown in cyan, β -strands in magenta and loops in pink. (C) C α tracing of cathepsin L (magenta) and papain (green) after their superposition. The disordered loop region (Phe172-Asn179) in cathepsin L (Fujishima *et al.*, 1997) is shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 8. (based on Chowdhury *et al.*, 2008) Location of the subsites in the active site cleft of cathepsin L (Pdb accession code 3bc3). Carbons are shown in green, oxygens in red, nitrogens in blue and sulfurs in yellow. This figure has been produced using PyMol (DeLano, 2007).

Figure 9. (A) Close up view of the active site of cathepsin L (Pdb accession code 3bc3). The residues forming the oxyanion hole and Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 10. (A) Close up view and surface representation of the binding site for a substrate's or an inhibitor's P1 side chain in papain (Pdb accession code 9pap). The same conserved region is also shown in cathepsin L (B) (Pdb accession code 3bc3). The sulfur of Cys25 is oxidized in both shown structures, to sulfonate in papain (see Pdb accession code 9pap; Kamphuis *et al.*, 1984) and to sulfinate in cathepsin L (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). Conserved residues are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). This figure has been produced using PyMol (DeLano, 2007).

Figure 11. (A) Stereo view of the active site cleft of cathepsin L (Pdb accession code 3bc3). Residues forming the continuous wall on the right hand side of the S2 subsite (Fujishima *et al.*, 1997), Met70 at the bottom of the pocket, and the active site residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His163 are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 12. (A) Stereo view of the S3 subsite of cathepsin L (based on Fujishima *et al.*, 1997) (Pdb accession code 3bc3). Residues at the centre of the S3 subsite, surrounding amino acids, and the active site residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). and His163 are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 13. (A) Cathepsin B in complex with CA030 (Pdb accession code 1csb; based on Turk *et al.*, 1998) (Stereo view).The inhibitor CA030, residues, important for the binding of CA030, the active site residues Cys29 and His199, and residues of the S1' subsite and covering its top are shown as sticks (carbons in magenta (CA030) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 14. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The inhibitor, S1'-residues interacting with the inhibitor's Cys side chain (Cys) and Cys25 with its oxidised sulfur (to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 15. (A) Cathepsin B in complex with CA030 (Pdb accession code 1csb) (Stereo view). The inhibitor, residues of the conserved loop surface from Gln23 to Ser28 (Turk *et al.*, 1998), amino acids of the occluding loop from Cys108 to Cys119, and the catalytic active site residues Cys29 and His199 are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 16. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The inhibitor, S2'-residues interacting with the inhibitor's 4-biphenylacetyl rings (4-BPA), and the catalytic residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His163 are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 17. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The first inhibitor, 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide, the second inhibitor molecule, missing the Tyr residue and the N-(2-phenylethyl)amide, residues interacting with the second inhibitor's biphenyl moiety (BP2), and the catalytic amino acids Cys 25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His 163 are shown as sticks (carbons in magenta (inhibitor molecule 1), orange (inhibitor molecule 2) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Figure 18. Pictures of adult sponges (left) (taken from <http://upload.wikimedia.org/wikipedia/commons/6/62/SpongeColorCorrect.jpg> and a sponge larva (right) (taken from http://www.biology.ualberta.ca/CMD/Pics/_LeysLarva.jpg).

Figure 19. (taken from Shimizu *et al.*, 1998) Scanning electron micrographs of (A) isolated silica spicules (x130) and (B) axial filaments (x 1,000) from *Tethya aurantia*.

Figure 20. Sequence alignment of mature human cathepsin L (UniProtKB/TrEMBL entry P07711) with mature silicatein α from the sponge *Tethya aurantia* (O76238) (based on Shimizu *et al.*, 1998) and also mature silicatein α from the sponges *Lubomirskia baicalensis* (Q2PC18), *Suberites domuncula* (Q2MEV3) and *Hymeniacidon perleve* (Q2HYF6) (based on Fairhead *et al.*, 2008). Identical amino acids in all five proteins are marked with asterisks. Catalytic triad amino acid residues of the active site of cathepsin L and corresponding residues in the silicateins are highlighted in yellow. Cysteine residues involved in disulfide bridges in cathepsin L are highlighted in green. Hydroxy amino acid residues in cathepsin L and the silicateins are highlighted in turquoise. The 4 amino acid loop in cathepsin L (Smith and Gottesman, 1989; Fairhead *et al.*, 2008) that is missing in the silicateins (Fairhead *et al.*, 2008) is highlighted in grey. The sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 21. (A) The chemical reaction catalysed by the organic filament of the sponge spicule (based on Fairhead *et al.*, 2008) and (B) the chemical structure of tetraethoxysilane (based on Cha *et al.*, 1999).

Figure 22. Sequence alignment of the 4SER chimera (Pdb accession code 2vhs) with mature human cathepsin L (UniProtKB/TrEMBL entry P07711). Identical amino acids in both proteins are marked with asterisks. Catalytic triad amino acids of the active site of cathepsin L and corresponding residues in 4SER are highlighted in yellow, and positions corresponding to the sequence of cathepsin L are marked with asterisks. Residues, flanking the catalytic cysteine and histidine in cathepsin L and the corresponding mutations in the chimera are highlighted in cyan. Other mutated amino acids and deletions are highlighted in green. The 4 amino acid loop in cathepsin L (Smith and Gottesman, 1989; Fairhead *et al.*, 2008), which is missing in 4SER (Fairhead *et al.*, 2008) is highlighted in grey. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 23. 0.5° oscillation X-ray image of the crystal of 4SER collected to a higher resolution pass to 1.5 Å (A) and a lower resolution pass to 2.0 Å (B) at ID29, ESRF.

Figure 24. The asymmetric unit of 4SER showing four monomers. The C α traces of the monomers A, B, C and D are shown in green, cyan, magenta and red, respectively. This figure has been produced using PyMol (DeLano, 2007).

Figure 25. C α tracing of the four monomers A (green), B (cyan), C (magenta) and D (red) after their superposition. This figure has been produced using PyMol (DeLano, 2007).

Figure 26. (A) Ribbon diagram of the structure of the 4SER chimera (Pdb accession code 2vhs). α -helices (in cyan) and β -sheets (in red) are labelled. Loops are shown in magenta. The disordered loop between β -strands 7 and 8 is shown as dotted lines. (B) Stereo view of the connection of the heavy and light chain of 4SER (shown as C α traces in magenta and green, respectively) via the S-S bridge between the heavy chain residue Cys156 and Cys209 (shown as sticks, carbons in cyan). The disordered loop region is shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 27. Stereo view of C α tracing of chain A of 4SER (in green, Pdb accession code 2vhs) and chain A of human cathepsin L (in magenta, Pdb accession code 1mhw) after their superposition. Mutated residues in the chimera are shown as sticks (carbons in cyan) and are labelled. The disordered loop region in both structures is shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 28. (A) Ribbon diagram of the 4SER chimera, showing the “‘V’ shaped active site cleft” (Turk *et al.*, 1998) with the catalytic residues Ser25 and His163 (Turk *et al.*, 1998) (in sticks, carbons in green) in its centre. The third residue of the catalytic triad, Asn 187 (Fairhead *et al.*, 2008), is also shown (in sticks, C-atoms in green) α -helices are shown in cyan, β -sheets in red and loops in magenta. The disordered loop region is shown as dashed lines. (B) Stereo view of the active site residues Ser25 and His163 (in sticks, carbons in cyan) and their flanking residues (in sticks, C-atoms in green). The third amino acid of the catalytic triad, Asn187 (Fairhead *et al.*, 2008), is also shown (in sticks, carbons in cyan). This figure has been produced using PyMol (DeLano, 2007).

Figure 29. Final 2F_o-F_c at 1 σ (blue) and unbiased F_o-F_c maps at 3 σ (green) around the sulfates at the active site (1) and about 17 Å distant from Ser25 (2) of each monomer of 4SER (Stereo view). The SO₄²⁻ molecules are modelled in the 2F_o-F_c densities. This figure has been produced using PyMol (DeLano, 2007).

Figure 30. Stereo view showing one of the sulfate ions (shown as sticks) at a crystal contact between monomer A (shown as C α traces in green) and a crystallographically related monomer (shown as C α traces in cyan). This figure has been produced using PyMol (DeLano, 2007).

Figure 31. The mechanism of silica condensation.

Figure 32. (after Cha *et al.*, 1999) Proposed reaction mechanism of tetraethoxysilane polymerisation carried out by silicatein α , based on the well studied catalytic reaction performed by Ser /His and Cys / His active-site proteases (Cha *et al.*, 1999; Zhou *et al.*, 1999).

Figure 33. (taken from Fairhead *et al.*, 2008) Silica condensing activity of the mutants C2S, AS2, AS2H, LOOP, 2SER, 4SER (see Table 1) generated by our collaborators. Protein in 0.1 M sodium phosphate buffer, pH 7 was mixed with sodium silicate at a concentration of 4.5 mM. The samples were then incubated at room temperature for one day (Fairhead *et al.*, 2008). After the samples had been centrifuged the precipitated silica was then collected and assayed using the Merck spectroquant silicon assay kit (Müller *et al.*, 2003; Fairhead *et al.*, 2008).

Figure 34. (A) Stereo view of the active site of the C25S procathepsin L mutant (Pdb accession code 1cjl). Active site and adjacent residues are shown as sticks. The distance of about 3.3 Å between the OG atom of Ser25 and ND1 of His163 is shown as dashed lines. (B) Stereo view of the active site of monomer C of 4SER (Pdb accession code 2vhs) showing the close contacts (<3.5 Å) as dashed lines. Active site and adjacent residues are shown as sticks, and waters as red spheres. The distance between the OG atom of Ser25 and ND1 of His163 is also shown as dotted lines (after Fairhead *et al.*, 2008). (C) As comparison the catalytic triad (His57, Asp102, Ser195) (shown as sticks) (reviewed in Polgár, 2005) of the serine protease chymotrypsin (Pdb accession code 4cha) is also shown. OG of Ser195 and NE2 of His57 are within H-bonding distance (Polgár, 2005) (2.8 Å) (dashed lines). This figure has been produced using PyMol (DeLano, 2007).

Figure 35. Close up view of the active site of 4SER (residues in sticks) with tetraethoxysilane (carbons in cyan, silicon atom in magenta) modelled at the same position as the sulfate ion (sulfur in yellow). Contacts between CAE of TEOS and OG of Ser25, CAA of TEOS and ND1 of His163, and also CAA and O of Asn162 are shown as dashed lines. The direction of rotation around the silicon and oxygen OAI is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007).

Figure 36. (A) (after Fairhead *et al.*, 2008) The proposed chemical mechanism for the condensation of Si(OH)₄. We favour a mechanism in which His163 abstracts H⁺ from silicic acid; the extensive water network could permit the shuttling of protons. (B) The other more unlikely mechanism of deprotonation of His163 and subsequent nucleophilic attack of the OH-group of the second silicic acid molecule on the first Si(OH)₄ (Fairhead *et al.*, 2008).

Chapter 2

Figure 37. (after Tanner, 2002) There are three distinct routes to stereochemical inversion. The top route (1) shows the cut of a C-H bond. This could be via deprotonation (amino acid racemases (Tanner, 2002), RmlC (Dong *et al.*, 2003) or H⁺ transfer (UDP galactose 4-epimerase (Frey, 1996; Tanner, 2002), ADP-L-glycero-D-mannoheptose 6-epimerase (Read *et al.*, 2005). The central route (2) describes cleavage of a C-O bond. An example is the epimerisation of 3-hydroxyacyl-CoA. This employs non-stereospecific dehydration/hydration to invert stereochemistry (crotonase reactions) (Yang and Elzinga, 1993; Yang *et al.*, 1991; Smeland *et al.*, 1991; Wakil, 1956). The bottom route (3) involves C-C bond cleavage. A good example is L-ribulose 5-phosphate 4-epimerase (Tanner, 2002).

Figure 38. (after Allard *et al.*, 2001) The epimerisation reaction: “the inversion of configuration of an asymmetrically substituted carbon in a sugar” (Allard *et al.*, 2001).

Figure 39. H⁺-abstraction and readdition involving transient oxidation (after Tanner, 2002).

Figure 40. Ribbon diagram of 3 β /17 β hydroxysteroid dehydrogenase (3 β /17 β HSD) from *Comamonas testosteroni* (Pdb accession code 1hxh). α -helices are shown in cyan, β -strands in magenta, and loops in pink. The α/β sandwich and the β -sheet in the centre, decorated by several α -helices represent the characteristic Rossmann fold (Oppermann *et al.*, 2003; Jörnvall *et al.*, 1995; Rossmann *et al.*, 1975). The β -strand in the centre is shown in yellow and is boxed.

Figure 41. Reaction mechanism of 3 β /17 β hydroxysteroid dehydrogenase with NADH and steroid substrate (5 α -androstane, 3-one,17ol). At the beginning of the catalytic reaction an H⁺ is transferred from the side chain hydroxyl of Tyr 151 to C=O of the substrate. Then H⁺ is transferred to C3 of 5 α -androstane, 3-one,17ol. Finally an H⁺ relay is built up including Lys155, Asn111, H₂O and the 2'-hydroxyl of the ribose. ARPP: adenosine ribose pyrophosphate part of the cofactor (after Filling *et al.*, 2002).

Figure 42. Structural sequence alignment of members from all five families of SDRs (Kallberg *et al.*, 2002) Conserved sequence motifs (see Table 10) in classical SDRs (3 β /17 β -hydroxysteroid dehydrogenase from *Comamonas testosteroni*, Pdb accession code 1hxh; 3 α ,20 β -hydroxysteroid dehydrogenase from *Streptomyces exfoliatus*, Pdb accession code 2hsd) are highlighted in yellow, the corresponding motifs in the representative of the extended SDR family (UDP-galactose 4-epimerase from *Homo sapiens*, Pdb accession code 1ek6) are highlighted in green, in intermediate SDRs (alcohol dehydrogenase from *Drosophila lebanonensis*, Pdb accession code 1a4u; alcohol dehydrogenase from *Drosophila melanogaster*, Pdb accession code 1mg5) in turquoise, in divergent SDRs (enoyl reductase from *Escherichia coli*; Pdb accession code, 1qsg; enoyl-ACP reductase from *Mycobacterium tuberculosis*; Pdb accession code 1bvr) in grey, and in one member of the complex SDR family (6-methylsalicylic acid synthase from *Penicillium patulum* (*Penicillium griseofulvum*), UniProtKB/TrEMBL entry P22367) in pink. Residues conserved in all five families are marked with an asterisk. The positions of secondary structural elements in classical SDRs are marked on top of each motif, positions of α -helices and β -strands, as they occur in the extended family are shown in red at the bottom of each motif. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 43. (taken from Persson *et al.*, 2003) The two levels of classification within the superfamily of SDR enzymes. Enzymes belonging to the superfamily of SDRs are divided into five families. Classical and extended SDRs are further divided into seven and three subfamilies, respectively, based on their amino acids associating to the cofactor (Persson *et al.*, 2003; Kallberg *et al.*, 2002).

Figure 44. Sequence alignment of various classical and extended SDRs, based on their coenzyme specificity (based on Kallberg *et al.*, 2002). The “key position” numbers refer to $3\alpha,20\beta$ -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) as a member of the classical family (blue asterisks) and UDP-galactose 4-epimerase (UniProtKB/Swiss-Prot Q14376 (GALE_HUMAN)) as a representative of the extended family (red asterisks) (Kallberg *et al.*, 2002). Conserved residues in all eight proteins are marked with asterisks. The conserved TG_{xxx}GhG motif of the classical SDRs (Kallberg *et al.*, 2002) is highlighted in yellow, the counterpart in the extended family TG_{xx}GhaG (Kallberg *et al.*, 2002) in green. “Key position” residues are highlighted in grey in the classical family and in turquoise in the extended one. $3\alpha,20\beta$ -hydroxysteroid dehydrogenase from *Streptomyces exfoliatus* (Pdb accession code 2hsd), subfamily cD1d (Kallberg *et al.*, 2002); 17-beta-hydroxysteroid dehydrogenase 4 (UniProtKB/Swiss-Prot P51659 (DHB4_HUMAN)) from *Homo sapiens*, subfamily cD1d (Persson *et al.*, 2003); 17-beta-hydroxysteroid dehydrogenase 14 (UniProtKB/Swiss-Prot Q9BPX1(DHB14_HUMAN)) from *Homo sapiens*; subfamily cD1d; Corticosteroid 11-beta-dehydrogenase isozyme 2 (UniProtKB/Swiss-Prot P80365 (DHI2_HUMAN)) from *Homo sapiens*; subfamily cD3 (Persson *et al.*, 2003); Estradiol 17-beta-dehydrogenase 12 (UniProtKB/Swiss-Prot Q53GQ0(DHB12_HUMAN)) from *Homo sapiens*, subfamily cP2; Carbonyl reductase (Pdb accession code 1cyd) from *mus musculus*, subfamily cP3 (Kallberg *et al.*, 2002); UDP-galactose 4-epimerase (UniProtKB/Swiss-Prot Q14376 (GALE_HUMAN)) from *Homo sapiens*, subfamily eD1 (Kallberg *et al.*, 2002); R-specific alcohol dehydrogenase (UniProtKB/Swiss-Prot Q84EX5 (Q84EX5_LACBR)) from *Lactobacillus brevis*, subfamily eP1 or cP2. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 45. (after Morrison and Tanner, 2007) Biosynthetic pathway of ADP-L-glycero-D-mannoheptose. The reaction carried out by ADP- β -L-glycero-D-mannoheptose 6-epimerase is surrounded with a box.

Figure 46. Ribbon diagram of AGME from the *E. coli* strain K12 (Pdb accession code 1eq2) (based on Deacon *et al.*, 2000) α -helices are shown in cyan, β -strands in magenta, and loops in pink. The seven parallel β -sheets in the centre, decorated by seven α -helices on either side represent the characteristic Rossmann fold. The carboxyterminal domain is composed of a small α/β domain, which plays a role in the interaction with the substrate. NADP and the inhibitor ADP-glucose are shown as sticks (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

Figure 47. (A) Stereo view of C α tracing of AGME from *E. coli* (in cyan, Pdb accession code 1eq2) and UGE from *E. coli* (in green, Pdb accession code 1nah) after their superposition (based on Deacon *et al.*, 2000). (B) Ribbon diagrams of both structures after their superposition (stereo view). This figure has been produced using PyMol (DeLano, 2007).

Figure 48. (after Morrison *et al.*, 2005 and Morrison and Tanner, 2007) The dismutation of ADP- β -D-manno-hexodialdose employed by AGME.

Figure 49. (after Morrison and Tanner, 2007) Proposed two-base mechanism for the reaction carried out by AGME. B₁ and B₂ are catalytic site acid/base amino acids. The hydrated form of ADP- β -D-mannohexodialdose 1 bound in the catalytic region is boxed.

Figure 50. Structures of ADP- α -D-glucose, ADP- β -L-glycero-D-manno-heptose and ADP- β -D-mannose.

Figure 51. PCR product of *hldD* from pET-30 Xa/LIC. Lane 1: 1 kbp ladder; lane 2: amplified 956 bp PCR product

Figure 52. Restriction analysis with *Nco*I and *Xho*I confirming the insertion of *hldD* into pEHISTEV. Lane 1: “clone 1”; lane 2: “clone 2”; lane 3: “clone 3”; lane 4: 1 kbp DNA ladder

Figure 53. SDS-PAGE showing hexa-histidine tagged AGME and AGME Y140F expressed from pET-30 Xa/LIC (A and B), and hexa-histidine tagged AGME expressed from pEHISTEV (C). (A) Lane 1: Mark 12TM Protein Standard; lane 2: whole cell extract of AGME (expression for two hours at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant; (B) lane1: Mark 12TM Protein Standard : lane 2: whole cell extract of AGME Y140F (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME Y140F (expression for three hours at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME Y140F (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant; (C) lane 1: Mark 12TM Protein Standard; lane 2: whole cell extract of AGME (expression for two hours at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant.

Figure 54. SDS-PAGE showing all the fractions obtained during the first purification step. Lane 1: Mark 12TM Protein Standard : lane 2: supernatant after first lysis; lane 3: cell extract after second lysis; lane 4: supernatant after second lysis; lane 5: final supernatant; lane 6: Flowthrough; lanes 7-9: wash fractions; lanes 10-13: eluted AGME containing fractions

Figure 55. SDS-PAGE showing the fractions of AGME eluted from the HiTrapTM Butyl FF column. Lane 1: Mark 12TM Protein Standard : lane 2: AGME sample before its application to the column; lanes 3-10: eluted AGME fractions

Figure 56. SDS-PAGE showing the fractions of AGME eluted from the HiPrepTM 16/60 SephacrylTM S-200 size exclusion column. Lane 1: Mark 12TM Protein Standard : lane 2: AGME sample before its application to the column; lanes 3-8: eluted AGME fractions

Figure 57. SDS-PAGE showing concentrated AGME before being stored at -80 °C. Lane 1: Mark 12TM Protein Standard : lane 2: AGME at 4.1 mg/ml; lane 3: AGME at 7 mg/ml

Figure 58. Crystals of AGME grown in presence of ADP-β-mannose (5-6.6 mM). 1: condition # 46 from JCSG⁺ screen; 2: condition # 74 from Classics screen; 3: condition # 2 from JCSG⁺ screen; 4: condition # 78 from JCSG⁺ screen; 5: condition # 5 from Classics screen

Figure 59. Optimised crystals of AGME grown in presence of 6.6 mM ADP-β-mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

Figure 60. Optimised crystals of reconstituted AGME grown in presence of 2.5 mM ADP-β-mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

Figure 61. Optimised crystals of AGME Y140F grown in the presence of 2.5 mM ADP-β-mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

Figure 62. Crystals of AGME Y140F grown in presence of ADP-β-mannose (2.5 mM). 1: condition # 67 from PEGs screen; 2: condition # 54 from PEGs screen; 3: condition # 76 from PEGs screen; 4: condition # 82 from Classics screen 5: condition # 73 from Classics screen

Figure 63. Optimised crystals of AGME Y140F grown in presence of 1 mM ADP- β -mannose under the condition 2 M (NH₄)₂SO₄, 0.1 M HEPES-Na, pH 7.1, 2 % PEG 400 (A) and 2 M (NH₄)₂SO₄, 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400, 10 mM Spermidine (B).

Figure 64. ITC analysis of AGME/ADP- β -mannose (A) and AGME Y140F/ADP- β -mannose (B) interactions. The titration profiles of ADP- β -mannose into AGME and AGME Y140F are shown in the top panels of (A) and (B). The integrated binding isotherms of the AGME/ADP- β -mannose and AGME Y140F/ADP- β -mannose titrations and experimental fits are demonstrated in the bottom panels. Each experiment was performed once.

Figure 65. Stereo view of the buried N-terminus of AGME (monomer A is shown) (Pdb accession code 1eq2). The aminoterminal β -sheet (β 1) is shown in red. The other secondary structure elements are shown in green. α -helices and β -strands surrounding β 1 are labelled.

Figure 66. Diffraction pattern of one of the AGME Y140F crystals grown from 0.2 M Ca-acetate, 0.1 M Na-cacodylate, pH 6.5 and 35 % PEG 300.

Figure 67. 0.25° oscillation X-ray images of the crystal of AGME at the start of the collection (A) and after about 30° of collected data (B).

Figure 68. 0.1° oscillation X-ray image of the crystal of AGME Y140F collected in-house to 2.8 Å resolution.

Figure 69. 0.5° oscillation X-ray image of the crystal of AGME Y140F collected to 2.4 Å resolution at BM14, ESRF.

Figure 70. The asymmetric unit of AGME showing two pentamers. C α traces of the ten monomers (A-J) are shown in different colours.

Figure 71. Unbiased F_o-F_c at 2.5 σ (green) and final 2F_o-F_c at 1 σ (blue) maps around the NADP⁺ of monomers A-J of AGME. The NADP⁺-molecule is modelled in the 2F_o-F_c density. Both electron densities for the cofactor are shown in stereo view in monomer A. Both maps were calculated after one run of restrained refinement. The weak electron density at the the nicotinamide part of NADP⁺ (F_o-F_c maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

Figure 72. Structural sequence alignment of AGME from *E. coli* K-12 (Pdb accession code 1eq2) with the structurally most related enzymes, UDP-galactose 4-epimerase from *E. coli* (Pdb accession code 1nah) and *Homo sapiens* (UniProtKB/TrEMBL entry Q14376), dTDP-D-glucose 4,6 dehydratase from *E. coli* (Pdb accession code 1bxk) and GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from *E. coli* (Deacon *et al.*, 2000) (Pdb accession code 1bws). 3 α ,20 β -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) as a member of the classical SDR family is also shown as comparison to the five enzymes (Kallberg *et al.*, 2002). Conserved sequence motifs (see Tables 12 and 14) in AGME are highlighted in yellow, the corresponding motifs in the two UDP-galactose 4-epimerases, in dTDP-D-glucose 4,6 dehydratase and in GDP-4-keto-6-deoxy-D-mannose epimerase/reductase are highlighted in green. Motifs in 3 α ,20 β -hydroxysteroid dehydrogenase are highlighted in turquoise. Residues conserved in all six SDR enzymes are marked with an asterisk. The positions of secondary structural elements in the five extended SDR enzymes are marked on top of each motif, positions of α -helices and β -strands, as they occur in 3 α ,20 β -hydroxysteroid dehydrogenase are shown in red at the bottom of each motif. The crucial residues (Ser116, Tyr140 and Lys144) in the catalytic region of AGME (Deacon *et al.*, 2000) are labelled as well. Residues at “key positions” based on coenzyme specificity are highlighted in magenta. The Asp residue at the end of β 2 in AGME (Kallberg *et al.*, 2002) is highlighted in grey. Numbers (*36) refer to 3 α ,20 β -hydroxysteroid dehydrogenase and human UDP-galactose 4-epimerase(*33) (Kallberg *et al.*, 2002). This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 73. (A) C α traces of the monomers (A-J) forming two pentamers in the 2.4 Å structure of AGME Y140F. (B) C α traces of the monomers (A-T) forming four pentamers (green, blue, red and magenta) in the 2.8 Å structure of AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Figure 74. Stereo view of the C α tracing of the ten monomers of the 2.4 Å (A) and the twenty monomers of the 2.8 Å (B) AGME Y140F structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

Figure 75. Unbiased F_o-F_c at 3 σ (green) and final 2F_o-F_c at 1 σ (blue) maps around the NADP⁺ of monomers A-J of the 2.4 Å structure of AGME Y140F. The NADP⁺-molecule is modelled in the 2F_o-F_c density. Both electron densities for the cofactor of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. The weak electron density at the the nicotinamide part of NADP⁺ (F_o-F_c maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

Figure 76. Unbiased F_o-F_c at 3 σ (green) and final 2F_o-F_c at 1 σ (blue) maps around ADP- β -mannose of monomers A-J of the 2.4 Å structure of AGME Y140F. The sugar molecule is modelled in the 2F_o-F_c density. Both electron densities for the nucleotide sugar of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. This figure has been produced using PyMol (DeLano, 2007).

Figure 77. Unbiased F_o-F_c at 3 σ (green) and final 2F_o-F_c at 1 σ (blue) maps around the NADP⁺ of monomers A-T of the 2.8 Å structure of AGME Y140F. The NADP⁺-molecule is modelled in the 2F_o-F_c density. Both electron densities for the cofactor of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. The weak electron density at the the nicotinamide part of NADP⁺ (F_o-F_c maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

Figure 78. Unbiased F_o-F_c at 3 σ (green) and final 2F_o-F_c at 1 σ (blue) maps around ADP- β -mannose of monomers A and C-T and ADP of monomer B of the 2.8 Å structure of AGME Y140F. The sugar molecule and ADP are modelled in the 2F_o-F_c density. Both electron densities for the nucleotide sugar of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. This figure has been produced using PyMol (DeLano, 2007).

Figure 79. Stereo view of the ten superimposed NADP and ADP- β -mannose molecules of 2.4 Å AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Figure 80. Stereo view of the twenty superimposed NADP and nineteen ADP- β -mannose molecules of 2.8 Å AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Figure 81. Stereo view of one superimposed NADP and ADP- β -mannose molecule from monomer A of the 2.4 Å (both molecules in green) and 2.8 Å structure (both molecules in magenta) of AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Figure 82. Stereo view of the superimposed NADP-molecules from monomer A of 2.4 Å AGME Y140F and monomer A of wildtype AGME (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

Figure 83. Close up view of the superimposed nicotinamide moieties from monomer A of 2.4 Å AGME Y140F, monomer A of wildtype AGME (Pdb accession code 1eq2) (Deacon *et al.*, 2000), monomer A of UGE (1ek6) (Thoden *et al.*, 2000) and monomer A of DGD (Pdb accession code 1bxk). This figure has been produced using PyMol (DeLano, 2007).

Figure 84. Stereo view of the superposition of NADP and ADP- β -mannose molecules from monomer A of the 2.4 Å resolution AGME mutant (carbons in green) with the NADP and ADP- α -glucose-moieties from monomers B (C-atoms in cyan) and D (carbons in magenta) of AGME (Pdb accession code 1eq2) (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

Figure 85. Stereo view of the adenine binding site in AGME Y140F. Hydrogen bonds are shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 86. Stereo view of the ribose binding site in AGME Y140F. H-bonds between atoms of the pentose and amino acids Ser180, Ala182 and His187 are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 87. Close up view of the pyrophosphate binding site in AGME Y140F. Hydrogen bonds between atoms of the PP_i and amino acids Thr81, Arg209 and Asn169 are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 88. Stereo view of the interactions of O2, O3, O4 and O6 of the mannose moiety of ADP- β -mannose (carbons in cyan) in AGME Y140F with closely positioned residues (C-atoms in green). H-bonds are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 89. Stereo view of superimposed ADP- β -mannose (carbon atoms in green) of AGME Y140F and ADP- α -glucose (carbons in cyan) from monomer D of AGME (Pdb accession code 1eq2). Residues close to the hexose moieties of both molecules, and also the highly conserved Lys144 (Deacon *et al.*, 2000) are also shown (C-atoms of residues of AGME Y140F are shown in green and carbons of AGME in cyan). The exactly same positions of the active site residues (Ser116, 140Phe/Try140 and Lys144 (Deacon *et al.*, 2000)) in both structures indicate that the mutation does not perturb the catalytic region of AGME. This figure has been produced using PyMol (DeLano, 2007).

Figure 90. Stereo view of the distances between C6 of the glucose moiety of ADP- α -glucose and C4 of NADP in monomers B (4 Å) (A) and D (3.4 Å) (B) of AGME (Pdb accession code 1eq2) and between C6 of the mannose part of ADP- β -mannose and C4 of NADP (2.9 Å) (C) of 2.4 Å AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Figure 91. Stereo view of the distance between O6 of the mannose moiety of ADP- β -mannose and NZ of Lys178 of AGME (7.7 Å) (A) and after a 120° rotation about the C-5"-C-6" bond (6.2 Å) (B). Tyr140 has been modelled based on the position of 140Phe. This figure has been produced using PyMol (DeLano, 2007).

Figure 92. Stereo view of the distances between Tyr140 and Lys 144 and the 2'OH and 3'OH of the ribose of NADP (red dotted lines), the distances between Tyr140 and Lys144 and Tyr140 and N1N of the nicotinamide (black dashed lines), and the intramolecular H-bond of the cofactor (blue dotted line). Amino acids, NADP and ADP- β -mannose are shown as sticks .This figure has been produced using PyMol (DeLano, 2007).

Figure 93. Stereo views showing the orientation of the hydrogens of C-6" of the mannose moiety in AGME (A) and C-4" of the rhamnose moiety in RmlD (Pdb accession code 1kc3) (B) in respect to C-4" of the nicotinamide ring of the coenzymes. Sugar moieties, cofactors (with hydrogens) and the conserved Tyr are shown as sticks. Distances between C-6"(mannose) / C-4" (rhamnose) and C-4" of the niacinamides are 2.9 Å and 3.1 Å, respectively. This figure has been produced using PyMol (DeLano, 2007).

Figure 94. Stereo view showing the distances (dashed lines) between Tyr140 and Ser116 and between both residues and O6 of the sugar moiety of ADP- β -mannose in AGME. This figure has been produced using PyMol (DeLano, 2007).

Figure 95. Stereoview of ADP- β -mannose and Tyr140 from AGME. ADP-L,D-Hep is shown as comparison. The C-6"-OH in all molecules is boxed. This figure has been produced using PyMol (DeLano, 2007).

Figure 96. Stereo view of ADP- β -mannose after a 120° rotation about the C-5"-C-6" bond and Tyr140 of AGME. ADP-D,D-Hep is shown as comparison. The C-6"-OH in all molecules is boxed This figure has been produced using PyMol (DeLano, 2007).

Figure 97. Stereo view of the contacts between C-6" of the sugar moiety of ADP- β -mannose (after a 120° rotation about the C-5"-C-6" bond) and C-4" of NADP⁺ and O6 of mannose and O of Tyr140 (2.1 Å). Hydrogens are shown in white, directions of proton and hydride transfers as arrows, and distances as dashed lines This figure has been produced using PyMol (DeLano, 2007).

Figure 98. Stereo view of the same site as demonstrated in Figure 97. Hydrogens are shown in white. The required little movement of Tyr140 (only 8.3°) to accommodate a proper position for adding H⁺ to the C=O group of the intermediate (Morrison and Tanner, 2007) is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007).

Figure 99. Stereo view of the active site of AGME. Tyr140 and atoms of the mannose moiety of ADP- β -mannose are shown as sticks. The required movement of the tyrosine residue into the other pocket of the catalytic region to be in an appropriate position to add H⁺ the C=O group at position 6 of the rotated keto-intermediate (Morrison and Tanner, 2007) is shown as arrow. This movement would not cause any clashes of Tyr140 with any amino acids of the protein. Ser116, as most closely located residue, is now 2.4 Å from Tyr140. This figure has been produced using PyMol (DeLano, 2007).

Figure 100. Stereo view of contacts between C-6" of the sugar moiety of ADP- β -mannose and C-4" of NADPH and the location of the rotated intermediate in respect to the moved Tyr140. Hydrogens are shown in white, directions of proton and hydride transfers as arrows. This figure has been produced using PyMol (DeLano, 2007)

Figure 101. Stereo views of the interactions of ADP-D,D-Hep (A) and ADP-L,D-Hep (B) with residues of AGME. The nucleotide sugars, NADP and amino acids are shown as sticks. The 35° movement of Tyr140 is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007)

Figure 102. Proposed one-base mechanism for the reaction catalysed by HldD. Tyr140 is the catalytic site acid/base amino acid.

Chapter 3

Figure 103. (taken from Page-McCaw *et al.*, 2007) Possible modes of matrix metalloproteinase action. Members of this enzyme family have the ability to cut parts of the extracellular matrix, which leads to more space for cells or tissues to migrate (a). Specific autocrine or paracrine signalling molecules can also be produced by cleavage carried out by matrix metalloproteinases (b). Matrix metalloproteinases are also able to directly control the composition of the tissue of the epithelium by cutting junctions between cells or the basement membrane (c). The action of cryptic signalling molecules can also be activated or altered by this family of proteinases which leads to various processes in cells (d). Deactivation or change of the function of signalling compounds which leads to modifications in cell death, proliferation, cell motility or differentiation can also be accomplished by matrix metalloproteinases (Page-McCaw *et al.*, 2007).

Figure 104. Structural sequence alignment of various MMPs (MMP-1, UniProtKB/TrEMBL entry P03956; MMP-2, UniProtKB/TrEMBL entry P08253; MMP-3, UniProtKB/TrEMBL entry P08254; MMP-7, UniProtKB/TrEMBL entry P09237; MMP-9, UniProtKB/TrEMBL entry P14780; MMP-10, UniProtKB/TrEMBL entry P09238; MMP-11, UniProtKB/TrEMBL entry P24347; MMP-12, UniProtKB/TrEMBL entry P39900; MMP-14, UniProtKB/TrEMBL entry P50281) from *Homo sapiens*. The conserved “Cys switch” fingerprint (Nagase *et al.*, 2006) sequence is highlighted in yellow, the Zn²⁺ associating motif in green, and the Met residue, 8 amino acids after the Zn²⁺ interacting cluster (Nagase *et al.*, 2006) in turquoise. Residues conserved in all nine matrix metalloproteinases are highlighted in magenta and marked with asterisks. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Figure 105. (taken from Nagase *et al.*, 2006) Organisation of the various domains of matrix metalloproteinases. sp: signal sequence; pro: prosegment; cat: region responsible for catalysis; FNII: fibronectin type II motif; L1: linker or “hinge-region”; Hpx: hemopexin region; CysR: cysteine rich; Ig: immunoglobulin part; L2: second connection; Mb: plasma membrane; TM: transmembrane domain; GPI: glycosylphosphatidylinositol anchor; Cy: tail located in the cytoplasm.

Figure 106. (taken from Page-McCaw *et al.*, 2007). Different levels of regulation of the actions of matrix metalloproteinases. The actions of matrixins can be controlled at the levels of (1) “RNA transcription” (Page-McCaw *et al.*, 2007), (2) translation, (3) “secretion, intracellular trafficking, (4) subcellular or extracellular localisation, (5) activation of the zymogen form, (6) expression of their endogenous protein inhibitors, such as tissue inhibitors of metalloproteinases (TIMPs)” (Page-McCaw *et al.*, 2007); Gomis-Rüth *et al.*, 1997) “and α2-macroglobulin” (Page-McCaw *et al.*, 2007; Sottrup-Jensen *et al.*, 1989);, “and (7) protease degradation” (Page-McCaw *et al.*, 2007).

Figure 107. Stereo ribbon diagram of the prosegment of pro-MMP-3 (Pdb accession code 1slm). This figure has been produced using PyMol (DeLano, 2007).

Figure 108. (A) Stereo view of Cα tracing of the prodomains of pro-MMP-3 from *Homo sapiens* (in green, Pdb accession code 1slm) and human pro-MMP-1 (in magenta, Pdb accession code 1su3) after their superposition. (B) Stereo ribbon diagrams of both structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

Figure 109. Stereo view of the region around the catalytic Zn^{2+} in pro-stromelysin-1 (Pdb accession code 1slm). Secondary structural elements of the prosegment and the catalytic domain are coloured in green and cyan, respectively. Residues from each of the two domains ligated with Zn^{2+} (grey sphere) are shown as sticks. This figure has been produced using PyMol (DeLano, 2007).

Figure 110. (A) Stereo view of the structure of RHTS (Pdb accession code 1cqr) (based on Chen *et al.*, 1999). α -helices are shown in cyan, β -sheets in red and loops in magenta (flexible loop from amino acids 210-234 (Chen *et al.*, 1999) labelled as L)). (B) C- α tracing of the catalytic domain of stromelysin-1 from *Homo sapiens* (in cyan, Pdb accession code 1cqr) and porcine MMP-1 (in magenta, Pdb accession code 1fbl) after their superposition. The rms deviation between both structures is 0.49 Å for 141 $C\alpha$ positions. (C) Ribbon stereo diagrams of both structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

Figure 111. Close up view of the interactions between NE2 of His201, His205 and His211 and the catalytic Zn^{2+} in RHTS (Chen *et al.*, 1999) (Pdb accession code 1cqr). The three histidines are shown as sticks, and the catalytic zinc as grey sphere and distances as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 112. Close up view of the interactions of NE2 of His151 and His166, ND1 of His179 and OD2 of Asp153 with the structural Zn^{2+} in RHTS (Pdb accession code 1cqr) (Chen *et al.*, 1999). The three histidines and the aspartic acid are shown as sticks, the catalytic zinc as grey sphere and distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 113. (A) Close up views of the interactions of OD1 of Asp158, the O-atoms of Gly159, Gly161 and Val163, OD2 of Asp181 and OE2 of Glu184 with Ca1 (Chen *et al.*, 1999; Dhanaraj *et al.*, 1996), (B) the interactions of OD1 of Asp107 and Asp182, the O-atoms of Asp182 and Glu184 and the two H_2O with Ca2 (Chen *et al.*, 1999), and (C) the interactions of OD1 of Asp177, O of Asp141, Gly173 and Asn175 and the two H_2O with Ca3 in RHTS (Chen *et al.*, 1999) (Pdb accession code 1cqr). Ligating residues are shown as sticks, the calcium ion as yellow sphere, waters as red spheres and distances as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 114. (based on Finzel *et al.*, 1998) A surface diagram to show the location of the subsites S1, S2, S3, S4, S1', S2' and S3' in the catalytic domain of stromelysin-1 (Pdb accession code 1usn) (shown as surface). Carbons are shown in green, oxygens in red and nitrogen in blue. The catalytic zinc ion is shown as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 115. (A) (based on Nagase *et al.*, 2006; Gomis-Rüth *et al.*, 1997) Stereo ribbon diagram of the catalytic region of matrixin 3 (cyan) complexed with TIMP-1 (magenta) (Pdb accession code 1uea). Amino acids Cys1-Val4 and Glu67-Cys70 of the inhibitor are shown as spheres (based on Gomis-Rüth *et al.*, 1997 and Nagase *et al.*, 2006). (B) (based on Gomis-Rüth *et al.*, 1997) Stereo ribbon diagram of the residues in the vicinity of the catalytic Zn^{2+} . The three conserved histidines of MMP-3 and N and O of Cys1 of the inhibitor interact with the catalytic Zn^{2+} (grey sphere) (Gomis-Rüth *et al.*, 1997). Residues of the matrixin (carbons in green) and TIMP-1 (carbons in yellow) are shown as sticks. This figure has been produced using PyMol (DeLano, 2007).

Figure 116. Surface representation of the interactions of Ro-26-2812 with subsites of the catalytic domain of stromelysin-1 (Pdb accession code 1c3i) The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 117. (based on Steele *et al.*, 2000) Stereo view of close interactions of Ro-26-2812 with residues of the catalytic domain of stromelysin-1 (Pdb accession code 1c3i). Amino acids (carbons in green) and the inhibitor (C-atoms in magenta) are shown as sticks, the catalytic Zn²⁺ as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 118. Stereo view showing the binding mode of the Cys-Gly-Val sequence of the propeptide to the catalytic domain stromelysin-1 (Pdb accession code 1slm). Amino acids (carbons in green) and the tripeptide (C-atoms in magenta) are shown as sticks, the catalytic Zn²⁺ as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 119. (after Becker *et al.*, 1995; Esser *et al.*, 1997) Chemical structures of the N-carboxyalkyl inhibitors I (Becker *et al.*, 1995; Chapman *et al.*, 1993) (left) and L-764,004 (Esser *et al.*, 1997) (right).

Figure 120. (after Esser *et al.*, 1997) Overview of the structure-activity relationship for P1'-lemonene-ethyl carboxylate blockers of stromelysin-1.

Figure 121. (after Pavlovsky *et al.*, 1999) Nonpeptide inhibitors I-IV co-crystallised with the catalytic domain of MMP-3. K_i-values for stromelysin-1 are also shown (Pavlovsky *et al.*, 1999).

Figure 122. Surface representation of the interactions of inhibitor II with subsites of the catalytic domain of stromelysin-1 (Pdb accession code 1caq). The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 123. Chemical structures of methoxyphenyl (A), phenyl (B) and 4-phenoxyphenyl (C).

Figure 124. (after Kohno *et al.*, 2006) Chemical structure of SM-25453.

Figure 125. (after Pikul *et al.*, 2001) Chemical structure of 4e.

Figure 126. (after Cheng *et al.*, 1999) Chemical structure of compound 14.

Figure 127. Chemical structures of piperazine based inhibitors (A) (after Cheng *et al.*, 2000), one 2,4,6-pyrimidine trione (B) (after Dunten *et al.*, 2001) and two thiadiazole inhibitors (C and D) (after Finzel *et al.*, 1998). An IC₅₀ value for the 2,4,6-pyrimidine trione (Dunten *et al.*, 2001) and K_i values for both thiadiazole inhibitors for three different matrixins (Finzel *et al.*, 1998) are also presented.

Figure 128. (based on Dunten *et al.*, 2001) Stereo view of the interactions of N3, O2 and O4 of compound 2 with the catalytic Zn²⁺ of the catalytic domain of stromelysin-1 (Dunten *et al.*, 2001) (Pdb accession code 1g4k). The three histidines ligating with Zn²⁺ (carbons in green) and the inhibitor (C-atoms in magenta) are shown as sticks, the catalytic Zn²⁺ as orange sphere and distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 129. (after Finzel *et al.*, 1998) Surface representation of the interactions of inhibitors PNU-142372 (carbons in green) and PNU-141803 (C-atoms in magenta) with subsites of the catalytic domain of stromelysin-1 (Pdb accession codes 1usn and 2usn). The inhibitors are shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 130. Chemical structure of inhibitor 3 and functions of its diverse moieties (after MacPherson *et al.*, 1997).

Figure 131. Chemical structure of inhibitor 70 (CGS 27023A) (MacPherson and Parker, 1995; MacPherson *et al.*, 1997). K_i values for various matrixins are also shown (after MacPherson *et al.*, 1997).

Figure 132. Chemical structures of inhibitors based on CGS 27023A. The various moieties of CGS 27023A are shown in different colours (hydroxamic acid in red, p-methoxybenzenesulfonamide in blue, pyridine in magenta, isopropyl in green).

Figure 133. SDS-PAGE showing lyophilised SCD dissolved in H₂O. Lane 1: Mark12TM Protein Standard; lanes 2-4: SCD in H₂O in increasing dilutions.

Figure 134. (A) SDS-PAGE showing the fractions of SCD eluted from the HiPrepTM 16/60 SephadrylTM S-200 size exclusion column. Lane 1: Mark12TM Protein Standard; lane 2: SCD before its application to the column; lanes 3-9: eluted SCD fractions; (B) Size exclusion chromatogram of SCD.

Figure 135. SDS-PAGE showing concentrated SCD. Lane 1: Mark12TM Protein Standard; lane 2: SCD at 11.5 mg/ml

Figure 136. Crystals of SCD grown in presence of 2mM of compound 3 under the condition 0.1 M (NH₄)₂SO₄, 0.1 M Na-cacodylate, pH 6.5, 29% PEG 8000 (A) and in presence of 2mM of compound 4 under the condition 0.1 M (NH₄)₂SO₄, 0.1 M Na-cacodylate, pH 6.5, 26% PEG 8000 (B).

Figure 137. 0.5° oscillation X-ray image of the co-crystal of SCD collected in-house to 2.5 Å resolution.

Figure 138. 0.5° oscillation X-ray image of the co-crystal of SCD collected in-house to 2.4 Å resolution.

Figure 139. (A) Stereo view of the Cα tracing of SCD complexed with inhibitor 3 (cyan) and chain A of the native catalytic domain of stromelysin-1 (green) (Chen *et al.*, 1999) (Pdb accession code 1cqr). (B) Stereo view of the Cα tracing of SCDs complexed with inhibitors 3 (cyan) and 4 (magenta). This figure has been produced using PyMol (DeLano, 2007).

Figure 140. Interactions of both SCDs with the catalytic (left) and structural (right) zinc ions. Carbons of amino acids and Zn²⁺ are shown in green for the SCD structure complexed with inhibitor 3 and in cyan for the SCD structure in complex with inhibitor 4. This figure has been produced using PyMol (DeLano, 2007).

Figure 141. Interactions of both SCDs with the first (A), second (B) and third (C) calcium ion. Carbons of amino acids and waters are shown in green for the SCD structure complexed with inhibitor 3 and in cyan for the SCD structure in complex with inhibitor 4. Calcium ions are shown in magenta. This figure has been produced using PyMol (DeLano, 2007).

Figure 142. Stereo view of superimposed sulfate ion and α-helix 1 of SCD structures in complex with inhibitors 3 (green) 4 (cyan) and inhibitor I (magenta) (Pdb accession code 1b8y) (Pavlovsky *et al.*, 1999). Lys110 is shown as sticks and close interactions (< 3.5 Å) between sulfate O-atoms of SCD complexed with inhibitor 3 and N of Lys110 as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 143. Unbiased F_o-F_c at 3σ (green) and final 2F_o-F_c at 1σ (blue) maps around compound 3 of the 2.4 Å structure of SCD (A) and unbiased F_o-F_c at 3σ (green) and final 2F_o-F_c at 1σ (blue) maps around compound 4 of the 2.5 Å structure of SCD (B). Both inhibitor molecules are modelled in the 2F_o-F_c density. Both maps were calculated at final stages of refinement before addition of the compounds. This figure has been produced using PyMol (DeLano, 2007).

Figure 144. Stereo view of superimposed inhibitors 3 (carbons in magenta) and 4 (C-atoms in yellow) and adjacent residues (shown as green sticks in the 2.4 Å SCD structure and as cyan sticks in the 2.5 Å SCD structure). The zinc ions located at the exactly same position in both structures are shown as sphere(s) in green (higher resolution structure) and cyan (lower resolution structure). This figure has been produced using PyMol (DeLano, 2007).

Figure 145. Close up view of the interactions of the hydroxamic acid moiety of inhibitor 3 (carbons in cyan) with the catalytic zinc (grey sphere) and residues of SCD (shown as sticks with C-atoms in green). Distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 146. Stereo views of the interactions of the sulfonamide moiety (A) and the methoxy group (B) of inhibitor 3 with residues of SCD. Carbons of the inhibitor and residues of SCD are shown in cyan and green, respectively are The catalytic zinc is shown as grey sphere, distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 147. Surface representation of the interactions of inhibitor 3 with subsites of the catalytic domain of the 2.4 Å SCD structure. The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Figure 148. (after Wilfong, 2008) Isothermal titration calorimetry (ITC) data for CGS 27023A and inhibitors 2-4.

Figure 149. Stereo view of the interactions of the hydroxamic acid moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 150. Stereo view of the interactions of the sulfonamide moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 151. Stereo view of the interactions of the methoxy phenyl moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Figure 152. Close up view of the interactions of the pyridine ring of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in cyan) (Pdb accession code 1bm6). The superimposed residues and inhibitor 3 of the 2.4 Å resolution SCD structure are also shown as sticks (C-atoms in green and orange, respectively). Zinc ions in the X-ray and NMR structure are shown as green and cyan spheres, respectively. This figure has been produced using PyMol (DeLano, 2007).

Figure 153. Surface representation of the interactions of inhibitors CGS 27023A (carbons in magenta) and 3 with subsites of the catalytic domain of the SCD structure. The surface of the 2.4 Å SCD crystal structure is shown. This figure has been produced using PyMol (DeLano, 2007).

Figure 154. Ribbon diagram showing the 2.4 Å SCD structure complexed with inhibitor 3 (secondary structure elements in green, C-atoms of the inhibitor in orange) superimposed with the NMR structure of SCD in complex with inhibitor 1 (NMR structure with the lowest energy of the ensemble of the twenty structures deposited is used) (secondary structure elements in cyan, carbons of the inhibitor in magenta) (Pdb accession code 1bm6) (Li *et al.*, 1998) (Rmsd = 1.55 Å for 139 C α positions). Both loop regions, amino acids 221-231 in the NMR structure, and residues 221-224 in the crystal structure are shown in magenta and orange, respectively. Additionally disordered parts of the entire flexible loop (amino acids 210-234) (Chen *et al.*, 1999) in the X-ray structure are boxed. This figure has been produced using PyMol (DeLano, 2007).

Figure 155. (after Chen *et al.*, 1999) Chemical structure of PGV-25727.

Figure 156. Stereo view of superimposed Tyr223 of SCD complexed with inhibitor 3 (carbons in green $2F_o - F_c$ density at 1σ), SCD in complex with inhibitor PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999), native SCD (carbons in magenta) (Pdb accession code 1cqr) (Chen *et al.*, 1999) and the NMR structure of SCD complexed with CGS 27023A (C-atoms in yellow) (Pdb accession code 1bm6) (Li *et al.*, 1998). This figure has been produced using PyMol (DeLano, 2007).

Figure 157. (A) Superposition of Tyr223 residues and inhibitors of SCD structures complexed with the nonpeptide inhibitors 3 (carbons in green), PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999) and I (carbons in blue) (Pdb accession code 1b8y) (Pavlovsky *et al.*, 1999) and the N-carboxyalkyl peptide inhibitor I (C-atoms in orange) (Pdb accession code 1sln) (Becker *et al.*, 1995). (B) Ribbon diagram of the same four superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed. This figure has been produced using PyMol (DeLano, 2007).

Figure 158. (A) Superposition of Tyr223 residues and inhibitors of SCD structures complexed with the nonpeptide inhibitors PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999), PNU-142372 (carbons in green) (Pdb accession code 1usn) (Finzel *et al.*, 1998), PNU-141803 (C-atoms in magenta) (Pdb accession code 2usn) (Finzel *et al.*, 1998) and the native SDR structure (carbons in yellow) (Pdb accession code 1cqr) (Chen *et al.*, 1999). (B) Ribbon diagram of the same four superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed. This figure has been produced using PyMol (DeLano, 2007).

Figure 159. (A) Superposition of Tyr223 residues and the same *N*-carboxyalkyl peptide inhibitor of the crystal (C-atoms in green) (Pdb accession code 1sln) (Becker *et al.*, 1995) and NMR (C-atoms in cyan) (Pdb accession code 2srt) (Gooley *et al.*, 1994) structures of SCD. (B) Ribbon diagram of the same superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed and shown in different colours (magenta in the crystal and yellow in the NMR structure). This figure has been produced using PyMol (DeLano, 2007).

List of Tables

Chapter 1

Table 1. (after Fairhead *et al.*, 2008) Mutant constructs made

Table 2. Output from POINTLESS (Evans, 2006)

Table 3. (after Fairhead *et al.*, 2008, Supplementary Material (ESI)) Data collection statistics for the 4SER crystal. Data for the highest resolution shell are shown in brackets.

Table 4. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(1.5) = probability at 1.5 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of $2.08 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a 40.9% solvent content at a probability of 0.81 at 1.5 Å and a total probability of 0.53 (shown in orange), suggested 4 molecules of the 4SER chimera in the asymmetric unit.

Table 5. The top solution from PHASER.

RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes (unfavourable molecular contacts in crystal packing)

Table 6. TLS groups employed throughout the refinement of 4SER.

Table 7. Refinement statistics of 4SER (after Fairhead *et al.*, 2008, Supplementary Material (ESI)). Values for R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Table 8. Output from MOLPROBITY (Davis *et al.*, 2007)

Table 9. Interface areas between the monomers of 4SER in the asymmetric unit. The Complexation Significance Scores (CSS) for each interface are also shown.

Chapter 2

Table 10. (after Persson *et al.*, 2003, Kallberg *et al.*, 2002) Conserved sequence motifs in classical and extended SDRs

Table 11. (after Persson *et al.*, 2003; Kallberg *et al.*, 2002) “Key positions” for assignments of cofactor preference of classical (A) and extended short chain dehydrogenases/reductases (B). Numbers refer to $3\alpha,20\beta$ -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) in A and UDP-galactose 4-epimerase (Pdb accession code 1ek6) in B.

Table 12. Details of the PCR reaction for the amplification of *hldD*. The reaction was carried out in the GeneAmp® PCR System 2400 thermal cycler (Applied Biosystems)

Table 13. Reagents used in the restriction digest of amplified *hldD* and pEHISTEV

Table 14. Details of the PCR reactions for the amplification of *hldD* cloned into pEHISTEV. Reactions were carried out in the GeneAmp® PCR System 2400 thermal cycler (Applied Biosystems)

Table 15. Reagents used in the restriction digest

Table 16. Overview of performed expression trials. Conditions of best soluble AGME expression are highlighted.

Table 17. ITC results of AGME and AGME Y140F with ADP- β -mannose

Table 18. Data collection statistics for the AGME crystal.

Table 19. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.93) = probability at 2.93 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of $2.75 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a 55.26 % solvent content at a probability of 0.16 at 2.93 Å and a total probability of 0.14 (shown in orange), suggested 10 molecules of AGME in the asymmetric unit.

Table 20. The top solution from PHASER. RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes (unfavourable molecular contacts in crystal packing)

Table 21. Data collection statistics for the AGME Y140F crystal. Data for the highest resolution shell are shown in brackets.

Table 22. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.80) = probability at 2.8 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of $2.82 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a 56.48 % solvent content at a probability of 0.06 at 2.8 Å and a total probability of 0.06 (shown in orange), suggested 20 molecules of AGME in the asymmetric unit.

Table 23. The top solution from PHASER. RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes (unfavourable molecular contacts in crystal packing)

Table 24. TLS groups employed throughout the refinement of AGME Y140F. “Residues” 400 and 401 are referred to NADP and ADP- β -mannose, respectively.

Table 25. Refinement statistics of AGME Y140F. R_{work} and R_{free} in the highest resolution shell are shown in brackets

Table 26. Output from MOLPROBITY (Davis *et al.*, 2007)

Table 27. Data collection statistics for the AGME Y140F crystal. Data for the highest resolution shell are shown in brackets.

Table 28. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.36) = probability at 2.36 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of $2.77 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a 55.59 % solvent content at a probability of 0.13 at 2.36 Å and a total probability of 0.13 (shown in orange), suggested 10 molecules of AGME Y140F in the asymmetric unit.

Table 29. The top solutions from Molrep.

Table 30. TLS groups employed throughout the refinement of AGME Y140F. “Residues” 400 and 401 are referred to NADP and ADP- β -mannose, respectively.

Table 31. Refinement statistics of AGME Y140F. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Table 32. Output from MOLPROBITY (Davis *et al.*, 2007)

Table 33. Overview of average B values (\AA^2) of protein, ligands, diverse ligand moieties and H_2O in the mutant structures and the already published AGME structure. Differences over the monomers are given in brackets.

Chapter 3

Table 34. Output from phenix.xtriage (Zwart *et al.*, 2005) showing pseudo translational symmetry and reindexing parameters.

Table 35. Data collection statistics for the SCD complex crystal. Data for the highest resolution shell are shown in brackets.

Table 36. TLS groups employed throughout the refinement of SCD. “Residues” 1-5 are referred to the metal ions Zn^{2+} and Ca^{2+} , “residue” 6 to compound 4.

Table 37. Refinement statistics of SCD complexed with inhibitor 4. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Table 38. Output from MOLPROBITY (Davis *et al.*, 2007)

Table 39. Data collection statistics for the SCD complex crystal. Data for the highest resolution shell are shown in brackets.

Table 40. TLS groups employed throughout the refinement of SCD. “Residues” 1-5 are referred to the metal ions Zn^{2+} and Ca^{2+} , “residue” 6 to compound 3.

Table 41. Refinement statistics of SCD complexed with inhibitor 3. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Table 42. Output from MOLPROBITY (Davis *et al.*, 2007)

Table 43. Temperature factors (\AA^2) of the first, second and third calcium ions (Ca1-Ca3) in both SCD complex structures compared to two other inhibited structures of the catalytic domain of MMP-3.

Contents

Chapter 1

THE STRUCTURE OF A SILICATEIN α-CATHEPSIN L CHIMERA	1-57
Introduction	1-28
1.1 Enzymatic peptide bond hydrolysis with emphasis on cysteine proteases	1-2
1.2 Cathepsins	3-21
1.2.1 Roles of cathepsins	6
1.2.2 Cathepsin L	7-21
1.2.2.1 Procathepsin L	7-11
1.2.2.1.1 Structure of procathepsin L	7-11
1.2.2.2 The structure of cathepsin L	11-21
1.2.2.2.1 Subsites of cathepsin L	12-21
1.2.2.2.1.1 S1 binding site	13-14
1.2.2.2.1.2 S2 binding site	14-15
1.2.2.2.1.3 S3 binding site	15-16
1.2.2.2.1.4 S1' binding site	16-18
1.2.2.2.1.5 S2' binding site	19-20
1.2.2.2.1.6 S3' binding site	20-21
1.3 Silicate formation in sponges	22-28
1.3.1 Silicatein α , a member of the cathepsin L subfamily	24-25
1.3.2 Enzymatic properties of silicateins	26-28
1.4. Project aims	28-29

Molecular biology, expression, purification and crystallisation of silicatein α-cathepsin L chimeras	30-31
Summary	30
1.5 Molecular biology, protein expression and purification	30
1.6 Assay for detection of silica condensation from cathepsin L mutants	31
1.7 Crystallisation of the 4SER mutant	31
Data collection, Processing, Molecular Replacement and Refinement of the 4SER chimera	32-40
Summary	32
1.8 Data collection and processing	32-34
1.9 Structure solution and refinement	35-39
1.9.1 Refinement	36-39
Discussion	39-40
Structure of the 4SER chimera	40-49
Summary	40
1.10 Overall structure	41-43
1.11 Comparison of 4SER to cathepsin L	44-45
1.12 Active site of 4SER	45-46
1.13 Sulfate ions in the structure of 4SER	46-49
Discussion	49-56
Future aims	57
Chapter 2	
THE STRUCTURE OF AGME Y140F IN COMPLEX WITH ADP-β- MANNOSE	58-186
Introduction	58-82
2.1 Epimerases and Racemases	58-60

2.2 Short-chain dehydrogenases/reductases (SDRs)	61-82
2.2.1 Reaction mechanism of SDR enzymes: 3 β /17 β hydroxysteroid dehydrogenase from <i>Comamonas testosteroni</i> as an example	63-64
2.2.2 Families of SDR enzymes	64-73
2.2.3 ADP-L-glycero-D-mannoheptose 6-epimerase (AGME)	73-82
2.2.3.1 Biosynthetic pathway of ADP-L-glycero-D-mannoheptose	73-75
2.2.3.2 AGME, a member of the SDR superfamily	76-79
2.2.3.3 Reaction mechanism of AGME	79-82
2.3 Project aims	82-83

Molecular biology, expression, purification crystallisation and Isothermal Titration Calorimetry (ITC) of AGME and mutant Y140F **84-114**

Summary	84
2.4 Molecular biology	84-89
2.4.1 Sub-cloning of <i>hldD</i> into pEHISTEV	84-89
2.5 Protein expression	90-93
2.6 Purification of recombinant AGME and AGME Y140F	94-99
2.6.1 Preparation and reconstitution of apo-AGME	98-99
2.6.1.1 Preparation of apo-AGME (based on Gomi <i>et al.</i> , 1989)	98
2.6.1.2 Reconstitution of apo-AGME (based on Ni <i>et al.</i> , 2001)	99
2.7 Crystallisation	99-106
2.7.1 Crystallisation of wildtype AGME	100-102
2.7.2 Crystallisation of wildtype AGME reconstituted with NADP ⁺	103
2.7.3 Crystallisation of AGME Y140F	103-106
2.8 Isothermal Titration Calorimetry (ITC) of AGME and AGME Y140F with ADP- β -mannose	107-109
2.8.1 Preparation of samples	107
2.8.2 ITC experiments	107-109
Discussion	109-114
Bioinformatics studies	114

Data collection, Processing, Molecular Replacement of AGME and AGME Y140F and Refinement of AGME Y140F **115-137**

Summary	115
2.9 Data collection, processing and structure solution of AGME (IO2, Diamond, Oxford)	116-119
2.9.1 Data collection and processing	116-117
2.9.2 Structure solution	118-119
2.10 Data collection, processing, structure solution and refinement of the AGME Y140F complex (in-house)	119-125
2.10.1 Data collection and processing	119-121
2.10.2 Structure solution	121-123
2.10.3 Refinement	123-125
2.11 Data collection, processing, structure solution and refinement of the AGME Y140F complex (BM14, ESRF, Grenoble)	125-132
2.11.1 Data collection and processing	125-127
2.11.2 Structure solution	127-129
2.11.3 Refinement	129-132
Discussion	133-137
AGME (IO2, Diamond, Oxford)	133-136
AGME Y140F (in-house and at BM14, ESRF)	137
2.12 Sequence alignment between AGME and closely related short chain dehydrogenases/reductases	138-141
Structures of AGME Y140F in complex with ADP-β-mannose	142-173
Summary	142
2.13 Overall structures	142-145
2.14 Active site of AGME Y140F	145-173
2.14.1 NADP ⁺ and ADP- β -mannose molecules in both mutant structures	145-162
2.14.2 NADP-binding site	162-166
2.14.3 Substrate-binding site	166-173

Discussion	174-185
Future aims	186

Chapter 3

STRUCTURES OF THE CATALYTIC DOMAIN OF STROMELYSIN-1 (SCD) IN COMPLEX WITH NONPEPTIDE INHIBITORS **187-261**

Introduction	187-217
---------------------	----------------

3.1 Matrix metalloproteinases	187-217
3.1.1 Structure of matrix metalloproteinases	189-191
3.1.2 Domains of matrix metalloproteinases	192
3.1.3 Stromelysin-1 (MMP-3)	193-217
3.1.3.1 Structure of pro-stromelysin-1	193-196
3.1.3.2 Structure of the catalytic domain of stromelysin-1	196-197
3.1.3.2.1 Location of metal ion sites	198-200
3.1.3.2.2 The active site and subsites of the catalytic domain of MMP-3	201-202
3.1.3.3 Inhibitors of stromelysin-1	203-217
3.2 Project aims	218-219

Purification and crystallisation of the catalytic domain of human stromelysin-1 (SCD) **219-224**

Summary	219
3.3 Purification	219-222
3.3.1 Size exclusion chromatography	220-222
3.4 Crystallisation	222-224
Discussion	224

Data collection, Processing, Molecular Replacement and Refinement of the catalytic domain of human MMP-3 complexed with inhibitors 3 and 4 225-237

Summary	225
3.5 Data collection, processing, structure solution and refinement of SCD in complex with inhibitor 4	226-231
3.5.1 Data collection and processing	226-228
3.5.2 Structure solution	228
3.5.3 Refinement	229-231
3.6 Data collection, processing, structure solution and refinement of SCD in complex with inhibitor 3	232-236
3.6.1 Data collection and processing	232-233
3.6.2 Structure solution	234
3.6.3 Refinement	234-236
Discussion	237
Structures of SCD in complex with inhibitors 3 and 4	238-260
Summary	238
3.7 Overall structures	238-240
3.8 Ion binding sites and coordination	240-243
3.9 Interactions with inhibitors 3 and 4	243-249
3.10 Common features and differences in inhibitor binding between the crystal structure of SCD in complex with inhibitor 3 and the NMR structure of SCD complexed with CGS 27023A (Inhibitor 1)	250-260
Future aims	261
References	262-280
Appendix	281-287
Publications	
Fairhead, M.; Johnson, K. A.; Kowatz, T. ; McMahon, S. A.; Carter, L. G.; Oke, M.; Liu, H.; Naismith, J. H.; van der Walle, C. F., Crystal structure and silica condensing activities of silicatein alpha-cathepsin L chimeras. <i>Chem Commun (Camb)</i> 2008 , (15), 1765-7.	

CHAPTER 1

THE STRUCTURE OF A SILICATEIN α -CATHEPSIN L CHIMERA

Introduction

1.1 Enzymatic peptide bond hydrolysis with emphasis on cysteine proteases

The cleavage of amide bonds in proteins is catalysed by ubiquitous peptide hydrolase enzymes (proteases). It is assumed that without involvement of proteases several hundred years might be required for cleavage of a peptide bond. A protease, however, would be able to cleave about a million amide bonds per second. The molecular weights of proteolytic enzymes vary between monomers of 10 kDa to polymers of several hundred kDa. Endoproteases cleave amide bonds within a protein whereas exopeptidases act at the N- or C-terminal end of a polypeptide. Proteases have been grouped based on the catalytic mechanism utilised to hydrolyse peptides. The most important groups are denoted serine, aspartate, threonine, metallo and cysteine proteases. Before hydrolysing an amide bond, a protease has to interact with its protein or peptide substrate. The efficiency of the interaction is controlled by the substrate binding sites of the protease and by the chemical nature of the peptide which binds to the active site. Even though only one amide bond is cut during the catalytic process, various amino acids flanking the cleavage site are crucial for the specificity of a protease (Figure 1A). The mechanism employed by cysteine proteases is similar to the one of serine proteases, but due to an additional shell of electrons in the S of the SH-group cysteine proteases are the better nucleophiles. Cysteine proteases need a crucial Cys in the catalytic site to perform hydrolysis. The nucleophilicity of the mercaptan group is further increased by a closely positioned active site His which operates as an H⁺-donor/general base (Sajid and McKerrow, 2002). The thiol group of the side chain of the Cys residue, and the imidazole ring of His lead to a “thiolate-imidazolium charge relay diad” (Sajid and McKerrow, 2002) (Figure 1B).

This is often stabilised by an Asn residue which displays high conservation. A Gln which is also highly conserved creates the oxyanion hole which is essential in establishing an electrophilic centre for the stabilisation of the tetrahedral intermediate during hydrolysis (Sajid and McKerrow, 2002). Stabilisation of the “charge relay system” (Sajid and McKerrow, 2002) is accomplished by the chemical milieu of the catalytic region (Sajid and McKerrow, 2002). The mechanism of hydrolysis has been well studied. Temporary binding of protease and substrate takes place and unstable tetrahedral intermediates are generated before the enzyme becomes active again (Sajid and McKerrow, 2002).

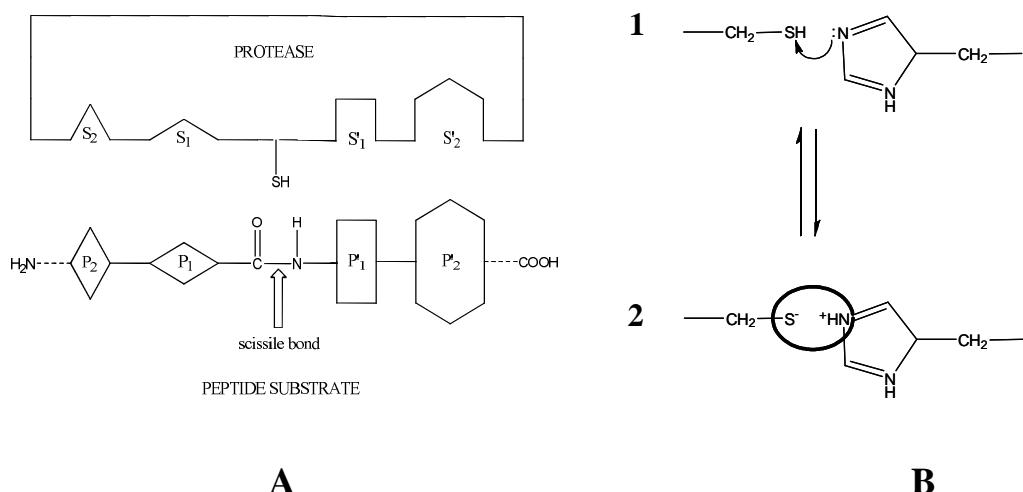


Figure 1. (after Sajid and McKerrow, 2002) Representation of the binding of a peptide substrate to the catalytic site of a cysteine protease. Amino acids from the peptide are indicated by P and the substrate binding sites of the protease by S. The thiol nucleophile of the catalytic cysteine is indicated by SH, and the corresponding amide bond that is attacked before hydrolysis (scissile bond) is marked with an arrow. The COOH side of the peptide (P'1, P'2, P'n) and corresponding substrate binding sites (S'1, S'2, S'n) are called prime side, whereas the NH₂ side of the peptide (P1, P2, Pn) and corresponding substrate binding sites (S1, S2, Sn) are named non-prime side. (A). Representation of the SH of Cys and the imidazole ring of His (B1), and the equilibrium with the thiolate-imidazolium charge relay. The delocalised electron density cloud is also shown (B2).

1.2 Cathepsins, members of the papain superfamily

Cysteine proteases are ubiquitous. The majority of these enzymes belong to the papain family. Members of this family are papain, plant proteases, cruzipain and lysosomal cysteine proteases, also called cathepsins. In *Homo sapiens* there are eleven cathepsins known so far denoted B, H, L, S, C, K, O, F, V, X and W. Initially it was thought that the main function of cathepsins is the promiscuous degradation of proteins inside the lysosomal compartment. Later, however, it was also demonstrated that they predominantly occur in the tissue and participate in highly specialized processes in the cell (Turk *et al.*, 2000). In general cathepsins show their best activity in a slightly acidic milieu (Turk *et al.*, 2000). The majority of cathepsins are endopeptidases (Turk *et al.*, 2000; Turk *et al.*, 2001). Examples for endoproteinases are the lysosomal cysteine proteases S and L (Dahl *et al.*, 2001; Turk *et al.*, 2000; Turk *et al.*, 2001). Most cathepsins occur as monomers apart from cathepsin C which is composed of four subunits. The molecular weight of these enzymes varies between 20 and 30 kDa. The structure of cathepsins and papain-like proteases in general consists of two domains. Whereas the left domain is composed of three α -helices, the right domain consists of five or six β -strands (Mc Grath, 1999). “The domains separate on the top in a ‘V’ shaped active site cleft” (Turk *et al.*, 1998) (Figure 2).

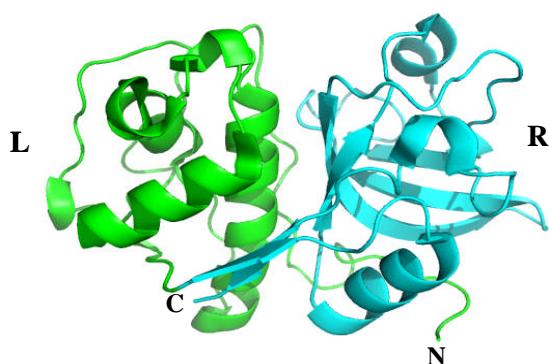


Figure 2. Ribbon diagram of the structure of human cathepsin S (Pdb accession code 1ms6). Secondary structural elements of the left (L) domain are shown in green, and secondary structural elements of the right (R) domain in cyan. This figure has been produced using PyMol (DeLano, 2007).

The left domain is usually stabilised by two S-S bridges, the right domain by only one S-S bond. The active site residue Cys25 is situated in the left domain, whereas His159 (papain numbering) is located in the right domain (Mc Grath, 1999). The roles of the Cys and His residues in cysteine proteases have already been discussed in Section 1.1. An amino acid sequence alignment of papain and various cathepsins from *Homo sapiens* (Turk *et al.*, 2000) shows the high conservation of both catalytic residues (Figure 3).

<pre> 9PAP -----IPEYVDWR-QKG-AVTPVKNQGS-----CCSCWAFSAVVTIEGIKIRT-G---NLNQYSEQELLDCDRR-----SYGCNGGYPWSALQLVAQYG 1MHW -----APRSVDWR-EKG-YVTPVKNQQG-----CCSCWAFSATGALEGQMFRKT-G---RLISLSEQNVLDCSGP-QGNEGCNGGLMDYAFQYVQDNG 1FH0 -----LPKSVDWR-KKG-YVTPVKNQQK-----CCSCWAFSATGALEGQMFRKT-G---KLVLSEQNVLDCSRP-QGNQGCNGGFMARAFQYVKENG 1YT7 -----APDSVDYR-KKG-YVTPVKNQQG-----CCSCWAFSSVGALEGQLKKKT-G---KLLNLSPQNVLDCVSE-----NDGCGGGYMTNAFQYVQKNR 1NQC -----LPDSVDWR-EKG-CVTEVKYQGS-----CCACWAFSAVGALAEQLKLKT-G---KLVTLSAQNLVDCDSTEKYGNKGCGNGFMTTAFQYIIDNK 8PCH -----YPPSMDWR-KKGN-FVSPVKNQGS-----CCSCWTFSTTGALESAVAIAT-G---KMLSLAEQQLVDCAQN-FNNHGCQGGLPSQAFEYIRYNK 1JQP QQILSLPESWDWRNVRGINFVSPVRNQES-----CCSCYSFASLGMLEARIRILT-NNSQTPILSPQEVVSCPY-----AQGCDGGFPYLIAKGYAQDF 1HUC -----VSVEVSAEDLLTC-----CCSMCGDGCGNGGPAEAWNFWTRKGLVSGGLYESHVGCRPY-----SIPPCEHHVNGS 1EF7 -----LPKSWDWRNVGDGVNYASITRNQHIPQY-----CCSCWAHASTSAMADRINIKRKGAPSTLLSVQNVIDCG-----NAGSCEGGNDLSVWDYAHQH </pre>	25 **	*
<pre> 9PAP -IHRYNTYPYEGVQRYCRSR-----EKGPYAAKTDGVRQVQPYN-----QGALLYSIAN-QPVSVVLQAAGKDFQLYRGGIFVGPCG-N----- 1MHW GLDSEESYPYEATEESCKYN-----PKY-SVANDTGFDIPK-Q-----EKALMKAVATVGPISVAIDAGHESFLFYKEGIYFEPDCSS-----</pre>	*	
<pre> 1FH0 GLDSEESYPYVAVDECKYR-----PEN-SVAQDTGFTVVAPGK-----EKALMKAVATVGPISVAMDAHGSSFQFYKSGIFYFEPDCSS-----</pre>	*	
<pre> 1YT7 GIDSEDAYPYVGQEESCMYN-----PTG-KAAKCRGYREIPEGN-----EKALKRAVARVGPVSAIDASLTSFQFYSKGVYYDESCNSD-----</pre>	*	
<pre> 1NQC GIDSDASYPYKAMDQKCQYD-----SKY-RAATCSKYTELPYGR-----EDVLKEAVANKGPVSVGVDARHPSFFLYRSGVYYEPSCTQN-----</pre>	*	
<pre> 8PCH GIMGEDTYPYKGDDHCKFQ-----PDK-AIAFKDVANITMND-----EEAMVEAVALYNPVSFAFEVTN-DFLMYRKGIYSSTSCHKTP-D</pre>	*	
<pre> 1JQP GVVEENCFPYTADCPK-----ENCLRYYSSEYYVGGFYGGCNEALMKLELVKHGPMAVAFEV-HDDFLHYHSGIYHHTGLSDPFNPFE-----</pre>	*	
<pre> 1HUC RPPCTGEGDTPKCSKCEPG-----YSPTYKQDKHYGNSYSVSNSEKDMAEIYKNGPVEGAFSVS-DFLLYKSGVYQHVTGEMMG-----</pre>	*	
<pre> 1EF7 GIPDETCNNYQAKDQECDKFNQCGTCNEFKECHAIRNYTLWRVGDYGSLSGREKMMAEIYANGPISCGIMATER-LANYTGGIYAEYQDTTY-----</pre>	*	*

Figure 3. (continues on next page)

9PAP	KVD A AAAV G YGPN———YILIKNSWGTGWG E NGYIRIKRG T GNSYGVCGLYTSSFYPVKN———
1MHW	DMD G VLVV G Y F EST———
1FH0	NLD G V L V G Y G FE G ANS D NS K YWL V KNSW G P E WG S NGY V KIA K D K NN H CG I ATA A SP N V———
1YT7	NLN G VL A AVL A V G Y G I Q KGN———KHW I IKNSW G EN W GN K GY I LMARN K NN A CG I AN L ASFP K M———
1NQC	-VN G VL V V G Y G D L NG———KEY W L V KNSW G H N F E EG Y I R M A N K G N H C I A S F P S Y P E I ———
8PCH	KVN G VL A AVL A V G Y G E E NG I P———Y W I V KNSW G P Q W G M N GY F LER G K-NMC G LAAC A S Y P I PL V ———
1JQP	LTN G AVLL V G Y G K D P V TG-LDY W I V KNSW G S Q W G E S GY F R I R R G-T D E C A I E S I A MA I P I P K L———
1HUC	-G A IRIL G W G V E NG T P-Y W L V ANS W NTD W GD N GFF K ILRG Q D-H C G I E S E V V A G I P R TD———
1EF7	-IN G V V S V A G W G I S D G -TEY W I V R N SW G E P W G ER G W L R I V T ST Y KDG K G A R Y N L A I E H C T FG D P I V———

159 * *** ***

Figure 3. Sequence alignment of papain and various cathepsins (based on Turk *et al.*, 2000) The key active sites residues C25 and H159 are highlighted in yellow and marked with asterisks. Other amino acid residues, conserved in all 9 enzymes are shown in red. The sequences are taken from the PROTEIN DATA BANK (PDB): papain from *Carica papaya* (Pdb accession code 9pap); cathepsin L from *Homo sapiens* (Pdb accession code 1mhw); cathepsin V mutant (N108Q, N178D) from *Homo sapiens* (Pdb accession code 1fh0); cathepsin K from *Homo sapiens* (Pdb accession code 1yt7); cathepsin S from *Homo sapiens* (Pdb accession code 1nqc); cathepsin H from *Sus scrofa* (Pdb accession code 8pch); cathepsin C from *Rattus norvegicus* (Pdb accession code 1jqp); cathepsin B from *Homo sapiens* (Pdb accession code 1huc); cathepsin X from *Homo sapiens* (Pdb accession code 1ef7). This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Similar to most other proteolytic enzymes, lysosomal cysteine proteases are synthesised as proenzymes which do not show any activity. Activation of these proenzymes takes place by removal of the aminoterminal proregion (Turk *et al.*, 2001). Cleavage of this prosegment can be carried out by either other proteolytic enzymes such as cathepsin D or pepsin, or by self cleavage at acidic pH (Turk *et al.*, 2000; Turk *et al.*, 2001). Rowan and coworkers reported such an autocatalytic activation on procathepsin B (Rowan *et al.*, 1992). The proregions have several functions, such as blocking proteolytic activity (Carmona *et al.*, 1996; Coulombe *et al.*, 1996), stabilisation of the protease under nonacidic conditions and folding of the enzyme (Mc Grath, 1999).

1.2.1 Roles of cathepsins

Gene knock-out experiments showed that lysosomal cysteine protease have specific functions in an organism. While lysosomal cysteine proteases such as V, K and S only occur in specific tissues, cathepsins H, L, O, B, F, X and C for instance are widely distributed. However, the latter lysosomal cysteine proteases can also have specific functions (Turk *et al.*, 2001). Details about functions of various cathepsin have been reviewed in Mc Grath, 1999. Lysosomal cysteine proteases have also been linked to various diseases, such as cancer (Yagel *et al.*, 1989), arthritis (Esser *et al.*, 1994), muscular dystrophy (Katunuma and Kominami, 1987) or osteopetrosis (Saftig *et al.*, 1998). Another example of a disease associated with cathepsins, pycnodysostosis, is an “autosomal recessive skeletal disorder” (Polymeropoulos *et al.*, 1995) which was described for the first time in 1962 by Maroteaux and Lamy (Maroteaux and Lamy, 1962; Polymeropoulos *et al.*, 1995). The main characteristic features of this disease are short stature, increased bone density, loss of phalanges, abnormal teeth and extremely fragile bones (Polymeropoulos *et al.*, 1995). Pycnodysostosis was found to be linked to a mutation in the cathepsin K gene (Gelb *et al.*, 1996; Turk *et al.*, 2001). Papillon-Lefevre syndrome (keratosis palmoplantaris with periodontopathia) is also “an autosomal recessive disorder” (Toomes *et al.*, 1999) and characterised by severe periodontitis which agonises patients (Gorlin *et al.*, 1964; Hart and Shapira, 2000; Toomes *et al.*, 1999). Palmoplantar keratosis affects the skin of patients and usually occurs between the age of 1 and 3. The reason for this disease is a mutation in the gene encoding cathepsin C (Toomes *et al.*, 1999).

1.2.2 Cathepsin L

1.2.2.1 Procathepsin L

There are two subfamilies within the papain superfamily (Karrer *et al.*, 1993, Coulombe *et al.*, 1996). In mammals this superfamily can be divided into cathepsin L-like and cathepsin B-like enzymes. Most enzymes are members of the cathepsin L-like subfamily. There are not many distinctions between the groups apart from some short parts (Coulombe *et al.*, 1996). Members of the cathepsin B-like subfamily contain an additional loop structure at around amino acid 95 (cathepsin L numbering), the occluding loop (Musil *et al.*, 1991; Coulombe *et al.*, 1996), that plays an important role in the exopeptidase activity specific for members of the cathepsin-B like subfamily. The prosegments of both groups, compared to the mature enzymes, are not very homologous to each other and differ in length. In cathepsin L-like enzymes this region has 30-40 amino acids more than the one in cathepsin B-like enzymes (Coulombe *et al.*, 1996).

1.2.2.1.1 Structure of procathepsin L

The mature form of procathepsin L (cathepsin L) displays a fold that is common to members of the papain superfamily (Baker and Drenth, 1987; Coulombe *et al.*, 1996). It consists of two domains forming an active-site cleft that comprises Cys25 and His163 which are crucial for catalysis. Cys25 is situated in the first domain, predominantly composed of α -helices, while His163 is in the other one, predominantly composed of β -strands (Coulombe *et al.*, 1996).

Essentially no distinctions were observed in the mature region of procathepsin L compared to the mature form of cathepsin L (Fujishima *et al.*, 1997).

The structure of the mature domains of procathepsin L (Pdb code 1cs8; Coulombe *et al.*, 1996) is essentially homologous to that of cathepsin L in complex with the MHC class II-associated p41 li fragment, with an rmsd of 0.31 Å between equivalent C α atoms. (Gunçar *et al.*, 1999). The rms deviation between equivalent C α atoms in the cathepsin L structure in the chagasin complex (Ljunggren *et al.*, 2007) and procathepsin L (Coulombe *et al.*, 1996) is 0.35 Å. Both structures only differ in the flexible loop region (residues Glu173-Asn180) (Ljunggren *et al.*, 2007). The prosegment consists of a globular domain, represented by the first about 80 amino acids, and an extended carboxyterminal-terminal part with around 20 residues. The globular domain is connected to Ala1 of cathepsin L by this extended part. The globular domain consists of three α -helices and loops. A short loop connects the α 1-helix with the long α 2-helix in the centre. (Coulombe *et al.*, 1996).

At the end of the helix α 2 the main chain flips round and parallels the helix adopting an extended conformation pointing towards the active site. The chain forms α 3 which is also part of the active site (amino acids 75-79) (Coulombe *et al.*, 1996) (Figure 4).

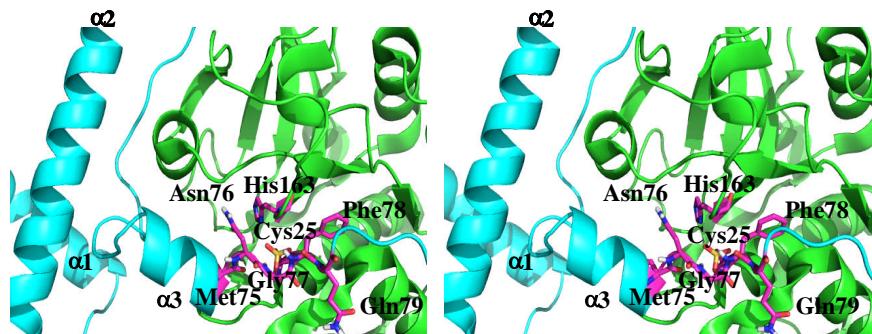


Figure 4. Stereo view showing residues Met75-Gln79 of the prosegment of procathepsin L (Pdb accession code 1cs8), dipping “into the substrate binding cleft” (Coulombe *et al.*, 1996) of the mature part of procathepsin L (Coulombe *et al.*, 1996). Secondary structure elements of the propeptide and the mature part of procathepsin L are shown in cyan and green, respectively. The catalytic active site residues Cys25 (its sulfur is oxidised to sulfonate) and His163, as well as the residues Met75-Gln79 of the prosegment are shown as sticks (carbons in magenta). This figure has been produced using PyMol (DeLano, 2007).

The interactions between the prosegment and cathepsin L mainly occur along the cleft which is involved in the interactions with the substrate and along the surface that is built up by the loop built up residues by His140-Asp155 (Coulombe *et al.*, 1996) (Figure 5). The amino acids Phe71 and Met75 to Gln79 of the prosegment play a role in the interactions between the proregion and the active site region of cathepsin L. The proregion is orientated in the active site in a way that is reverse to that of a natural substrate. (Coulombe *et al.*, 1996).

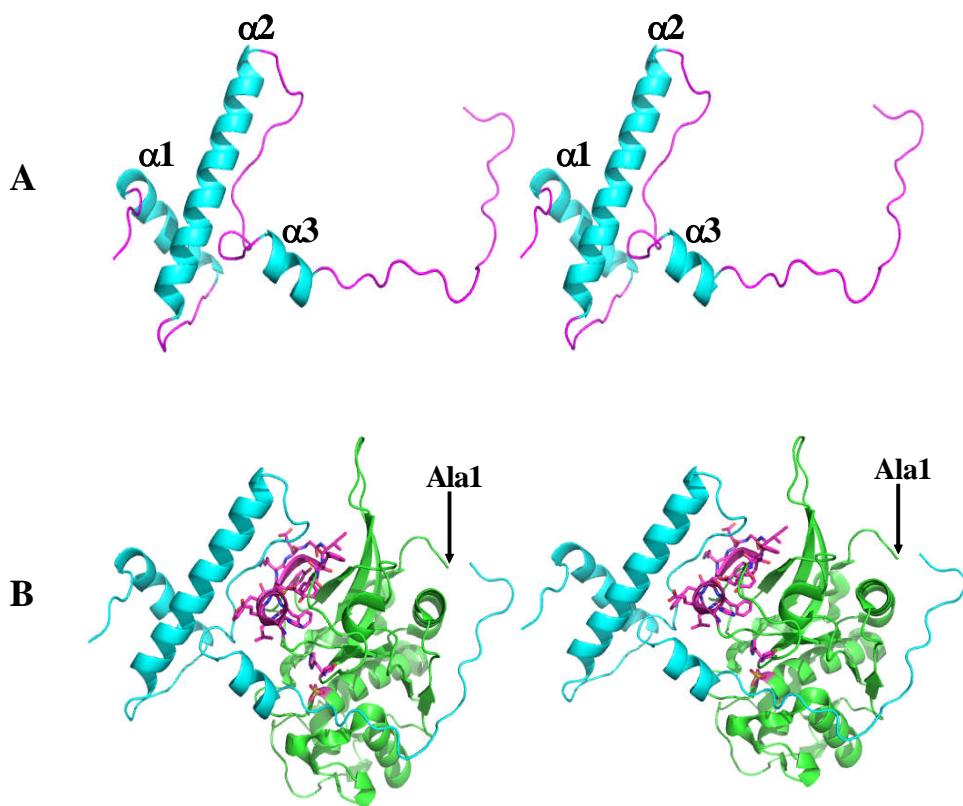


Figure 5. Stereo views of the prosegment of procathepsin L (A) (based on Coulombe *et al.*, 1996) (Pdb accession code 1cs8), α -helices are shown in cyan and loops in magenta, and procathepsin L (B). The prosegment is shown in cyan and cathepsin L in green. The catalytic active site residues Cys25 (its sulfur is oxidised to sulfonate; see Pdb accession code 1cs8 and Chowdhury *et al.*, 2002) and His163, as well as the residues His140-Asp155 of the prosegment binding loop (PBL) (Coulombe *et al.*, 1996) are shown in magenta. This figure has been produced using PyMol (DeLano, 2007).

Even though there is about 25 % sequence identity between cathepsin L and cathepsin B, the prosegments of procathepsin L and B are not very homologous. However, there is still an overall similarity in the structures of the prosegments of both enzymes (Figure 6). The most remarkable distinctions can be observed at the aminoterminus. Procathepsin B lacks helix α 1 and the N-terminal part of α 2, that do not show any contacts with cathepsin L at all. As a consequence the prosegment of procathepsin B is about 30 amino acids shorter. The main interactions between the prosegment and cathepsin B also apply to the proform of cathepsin L as well. The prosegment binding loop is important for the interaction between the prosegment and the protease in both cathepsins (Coulombe *et al.*, 1996).

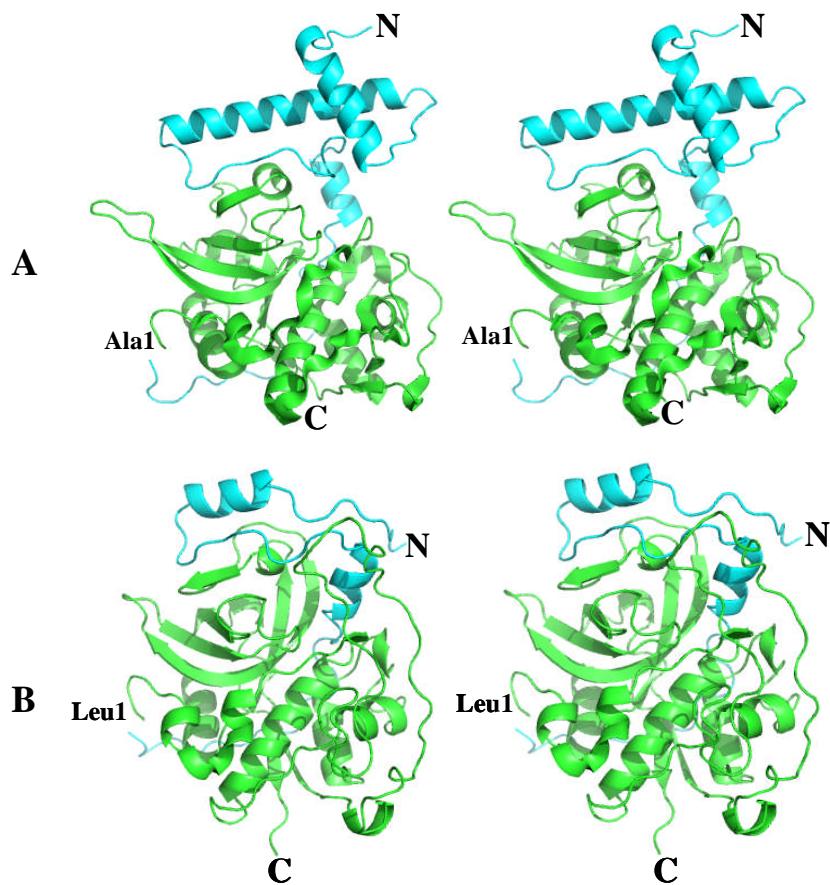


Figure 6. Stereo view of procathepsins L (Pdb accession code 1cs8) (A) and B (Pdb accession code 1pbh) (B), based on Coulombe *et al.*, 1996. The prosegments of both enzymes are shown in cyan and cathepsins L and B in green. This figure has been produced using PyMol (DeLano, 2007).

The fact that the prosegment blocks the activity of procathepsins L and B by interacting with the active site in a reverse orientation with respect to a substrate or an inhibitor seems to apply to all members of the papain superfamily (Coulombe *et al.*, 1996).

1.2.2.2 The structure of cathepsin L

The overall structure of human cathepsin L displays similarity to that of papain (Figure 7, A and B). The Rmsd between both structures is 0.82 Å for 183 C α atoms (Figure 7, C). The only significant difference lies in the loop region between Phe172 and Asn179 of cathepsin L (Fujishima *et al.*, 1997).

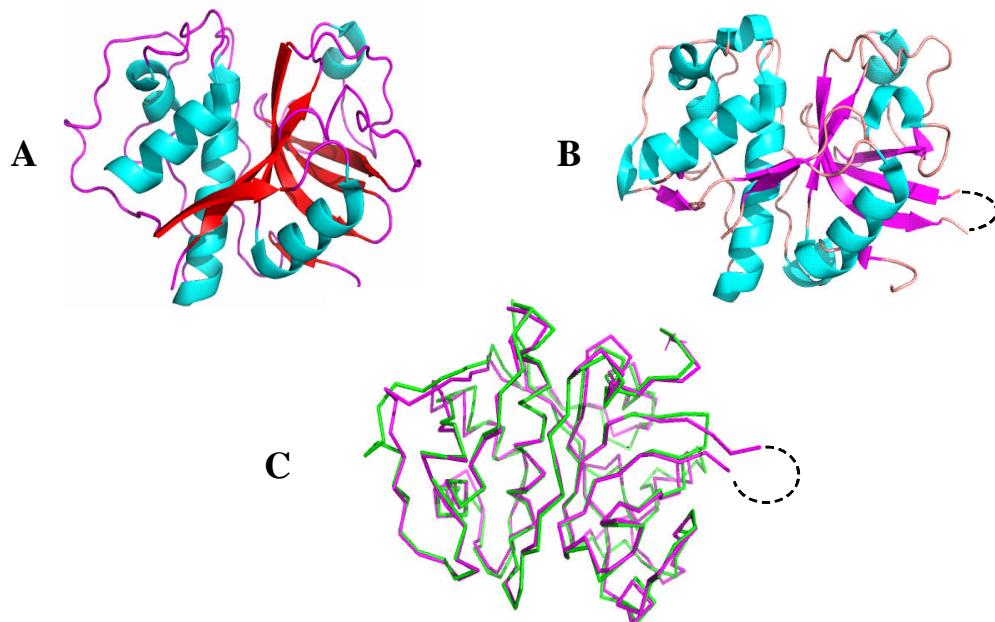


Figure 7. Ribbon diagrams of papain from *Carica papaya* (A) (Pdb accession code 9pap), α -helices are shown in cyan, β -strands in red and loops in magenta, and human cathepsin L (B) (Pdb accession code 3bc3). α -helices are shown in cyan, β -strands in magenta and loops in pink. (C) $C\alpha$ tracing of cathepsin L (magenta) and papain (green) after their superposition. The disordered loop region (Phe172-Asn179) in cathepsin L (Fujishima *et al.*, 1997) is shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1 Subsites of cathepsin L

It has been proposed by Schechter and Berger, 1967, that there are seven subsites in papain denoted S4-S3', but there is actually no proof that all those subsites really exist (Chowdhury *et al.*, 2008). The definition of substrate-binding sides has been revised by Turk *et al.*, 1998. Inhibitors binding to cathepsin L are usually substrate analogues which contain groups that establish a covalent bond with Cys25 (Chowdhury *et al.*, 2008). However, the propeptide in procathepsin L does not show any covalent interactions with the mature enzyme at all (Chowdhury *et al.*, 2002; Coulombe *et al.*, 1996).

Consequently, various noncovalent blockers of cathepsin L have been synthesised to mimic the binding mode of the proenzyme (Chowdhury *et al.*, 2002; Chowdhury *et al.* 2008). Those inhibitors are peptide-derived and show interactions with the S' and S substrate binding sites of the protease's catalytic region (Chowdhury *et al.*, 2002). The locations of these subsites are shown in Figure 8 (based on Chowdhury *et al.*, 2008).

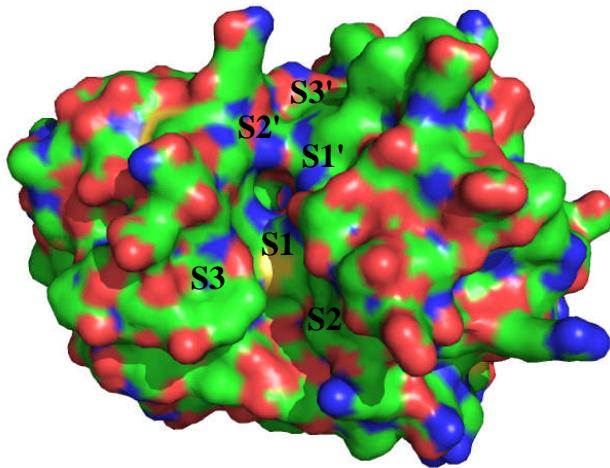


Figure 8. (based on Chowdhury *et al.*, 2008) Location of the subsites in the active site cleft of cathepsin L (Pdb accession code 3bc3) (Surface representation). Carbons are shown in green, oxygens in red, nitrogens in blue and sulfurs in yellow. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.1 S1 binding site

In papain the P1 C=O group of a substrate or an inhibitor points into the oxyanion hole that is formed at the aminoterminus of the central helix between H⁺ of the Cys25 amide and the amide of the side chain of Gln19. The P1 main chain amide is directed to C=O of Asp158 (Turk *et al.*, 1998). In cathepsin L the oxyanion hole is composed of the amino acids Gln19, Trp189, His163 and Cys25 (Figure 9) (Fujishima *et al.*, 1997).

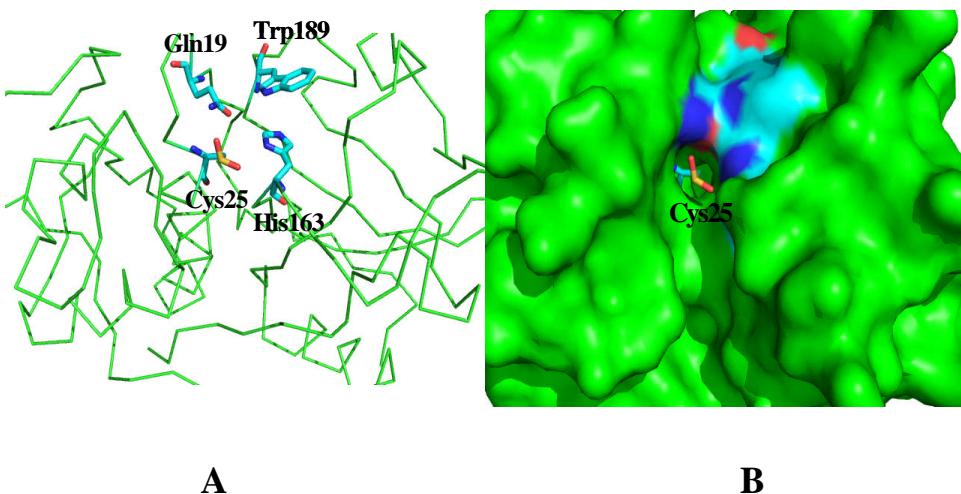


Figure 9. (A) Close up view of the active site of cathepsin L (Pdb accession code 3bc3). The residues forming the oxyanion hole and Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

In papain an inhibitor's P1 side chain binds to the region which is located between the loop from Ser21 to Gly23. This loop extends into the central α -helix on one side and a turn built up by Cys63-Gly66 on the other side. Both chains are connected at the top with the S-S bond between Cys22 and Cys63 (Turk *et al.*, 1998) (Figure 10). The P1 side chain points along the subsite S1 (Turk *et al.*, 1998).

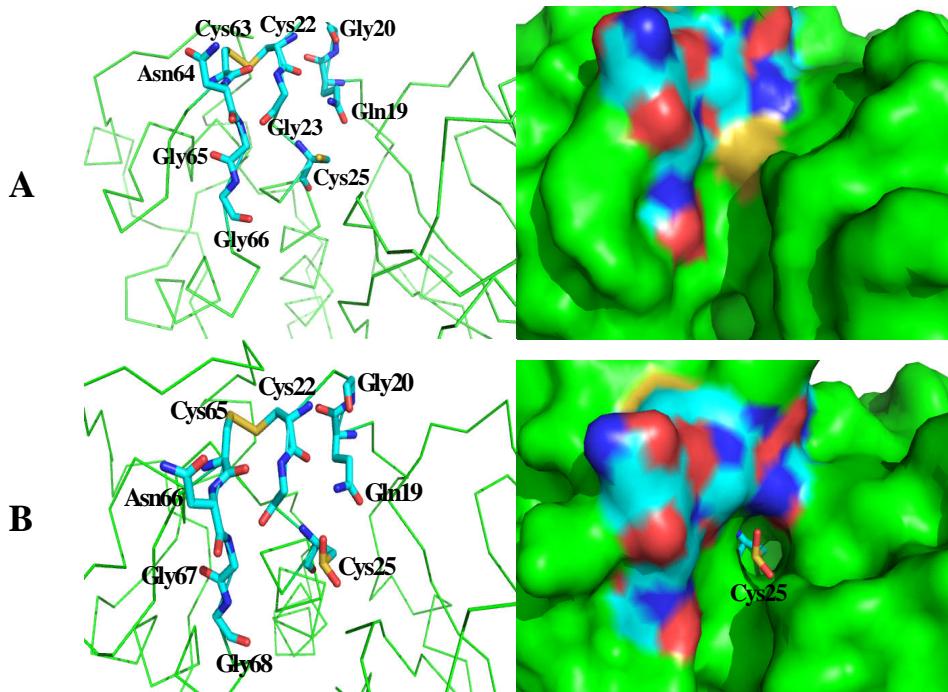


Figure 10. (A) Close up view and surface representation of the binding site for a substrate's or an inhibitor's P1 side chain in papain (Pdb accession code 9pap). The same conserved region is also shown in cathepsin L (B) (Pdb accession code 3bc3). The sulfur of Cys25 is oxidized in both shown structures, to sulfonate in papain (see Pdb accession code 9pap; Kamphuis *et al.*, 1984) and to sulfinate in cathepsin L (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). Conserved residues are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.2 S2 binding site

The S2 substrate binding site is possibly the best defined subsite. (Figure 11) (Turk *et al.*, 1998). In cathepsin L the S2 substrate binding site is deep and hydrophobic. Cysteine proteases that prefer bulky and hydrophobic residues at the P2 position of a substrate or an inhibitor should share these features. However, there are distinctions in the shape of this subsite in cathepsins L and B.

If one looks down the active site cleft of cathepsin L with the S1' and S2' subsites up and the S1-S3 subsites down, Asp162, Met161, Asp160 and Ala214 build a continuous wall (Figure 11) on the right hand side of the S2 subsite that makes the pocket narrow.

No equivalent amino acid for Met161 of cathepsin L is present in cathepsin B, which leads to a more open S2 substrate binding site (Fujishima *et al.*, 1997). The side chain of Met70 (Pro76 in cathepsin B) at the bottom of the pocket sits across the cleft (Figure 15) resulting in a shallower site in cathepsin L (Fujishima *et al.*, 1997).

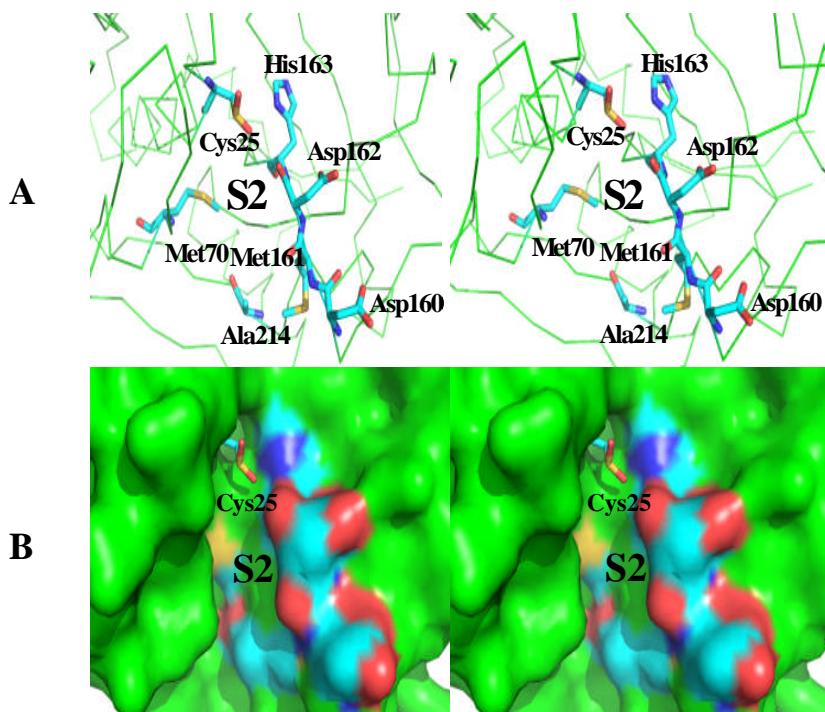


Figure 11. (A) Stereo view of the active site cleft of cathepsin L (Pdb accession code 3bc3). Residues forming the continuous wall on the right hand side of the S2 subsite (Fujishima *et al.*, 1997), Met70 at the bottom of the pocket, and the active site residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His163 are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.3 S3 binding site

The S3 substrate binding site of cathepsin L is located in the centre of the left wall of the cleft (Figure 12). Amino acids Gly67 and Gly68 are located in the centre of this substrate binding site. Both residues are decorated by the side chains of Asn66, Glu63 and Leu69 and the O-atom of C=O of Gly61. The S3 subsites of the cathepsins L and B show more similarity than other parts of the catalytic region.

However, the subsite in cathepsin B is narrower than that in cathepsin L. Therefore, a larger moiety of a substrate or an inhibitor may be accommodated by cathepsin L in this region (Fujishima *et al.*, 1997).

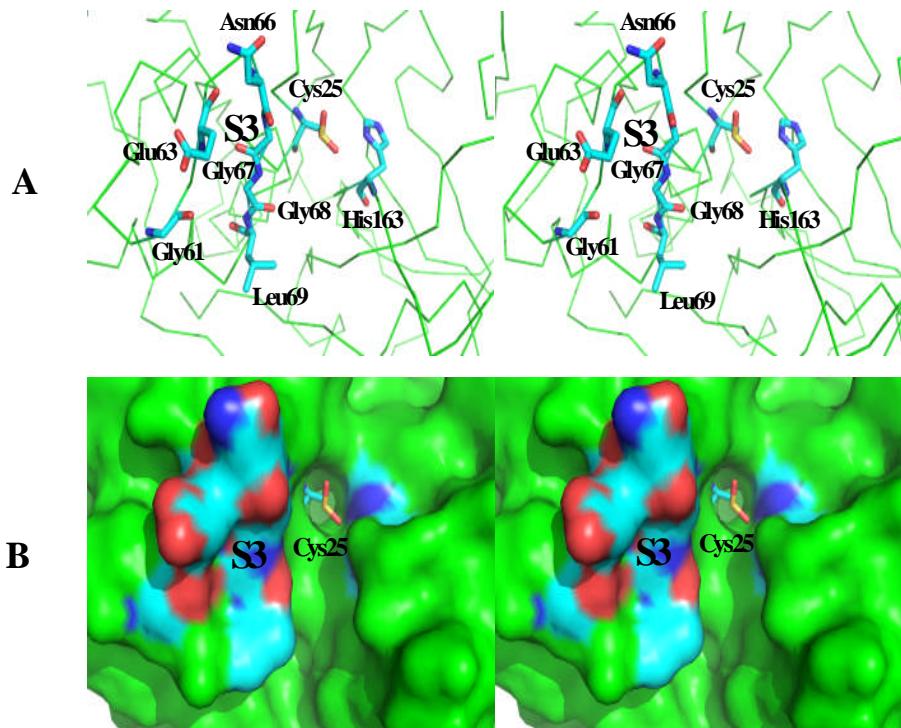


Figure 12. (A) Stereo view of the S3 subsite of cathepsin L (based on Fujishima *et al.*, 1997) (Pdb accession code 3bc3). Residues at the centre of the S3 subsite, surrounding amino acids, and the active site residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008), and His163 are shown as sticks (carbons in cyan, nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.4 S1' binding site

Turk *et al.*, 1998 illustrated the structure of this region with cathepsin B complexed with the inhibitor CA030 (Pdb accession code 1csb; Turk *et al.*, 1995). The amino group of CA030 isoleucine points to the C=O of Gly198, while its C=O group participates in an H-bond interaction with the NE1 atom of Trp221 (177 in papain).

The side chain is positioned between the nonpolar residues Leu181 and Val176 and is also close to Met196. (Turk *et al.*, 1998).

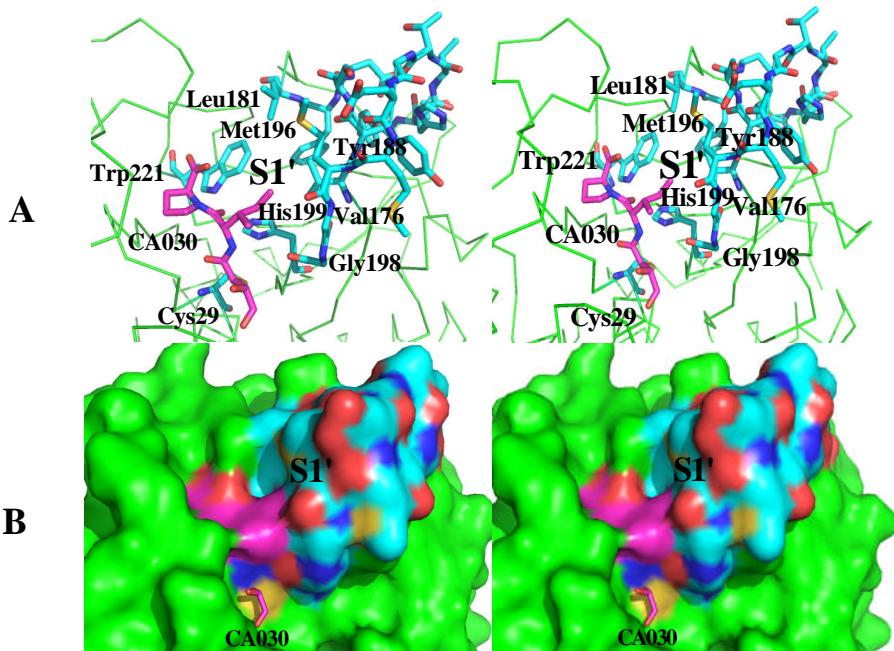


Figure 13. (A) Cathepsin B in complex with CA030 (Pdb accession code 1csb; based on Turk *et al.*, 1998) (Stereo view). The inhibitor CA030, residues important for the binding of CA030, the active site residues Cys29 and His199, and residues of the S1' subsite and covering its top are shown as sticks (carbons in magenta (CA030) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

Polar, as well as small, nonpolar side chains of an inhibitor's P1 group are accepted by papain (Schuster *et al.*, 1992; Turk *et al.*, 1998). It has been demonstrated that larger aromatic and nonpolar amino acids are accepted by cathepsin B and not cathepsins L and S and papain (Menard *et al.*, 1993; Turk *et al.*, 1998). Small, neutral polar P1' groups are favoured by cathepsin L. Generally cathepsin S and papain accept small and uncharged amino acids. Cathepsin B has the largest S1' subsite compared to the other members of the papain superfamily and its surface comprises nonpolar amino acids (Turk *et al.*, 1998).

In the complex structure of cathepsin L with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide, an inhibitor that binds noncovalently and similar to the propeptide to cathepsin L, the cysteine side chain of the inhibitor shows interactions with the side chains of the residues Ala138, Asp162, His163, and Trp189 of the S1' subsite (Figure 14). Its S- atom is in contact with a second inhibitor molecule, which only displays electron density for the 4-biphenylacetyl-Cys-D-Arg moieties. (Chowdhury *et al.*, 2002).

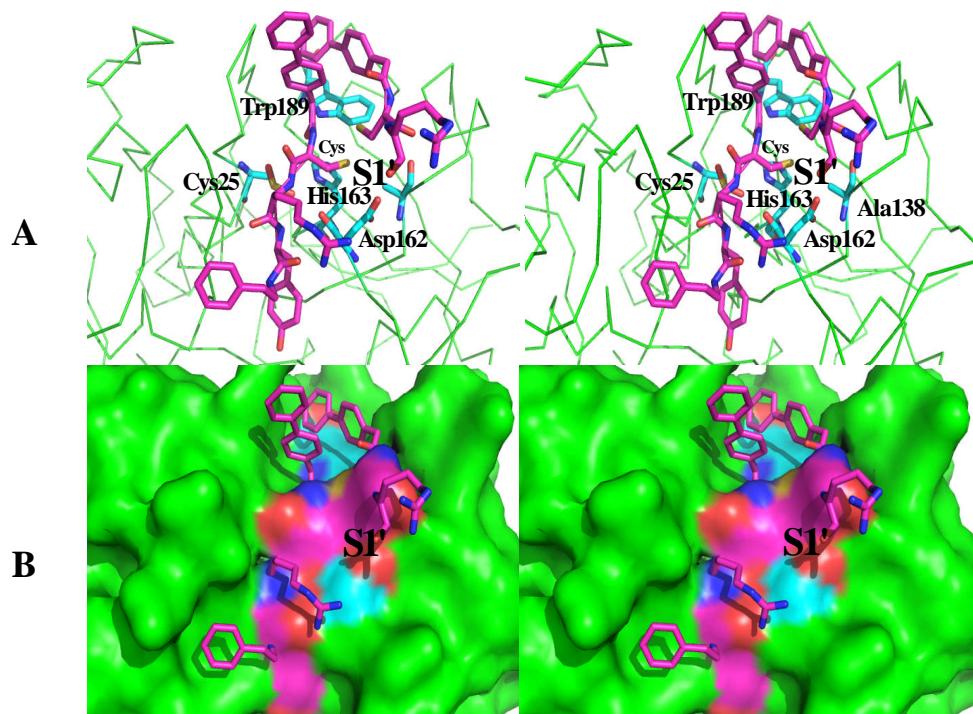


Figure 14. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The inhibitor, S1'-residues interacting with the inhibitor's Cys side chain (Cys) and Cys25 with its oxidised sulfur (to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008). are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.5 S2' binding site

Cathepsin B in complex with CA030 shows that the P2' side chain group binds to a conserved loop surface composed of Gln23-Ser28 (Turk *et al.*, 1998) (Figure 15). Cathepsins L and S, as well as papain have a similar binding profile (Brömmel and Kirschke, 1993; Turk *et al.*, 1998). In contrast the associations of cathepsin B with the same ligands are different. It has been suggested that interactions with the occluding loop above the S2' subsite and His110 and His111 at the bottom of the loop are the reason for these differences (Figure 15) (Turk *et al.*, 1998).

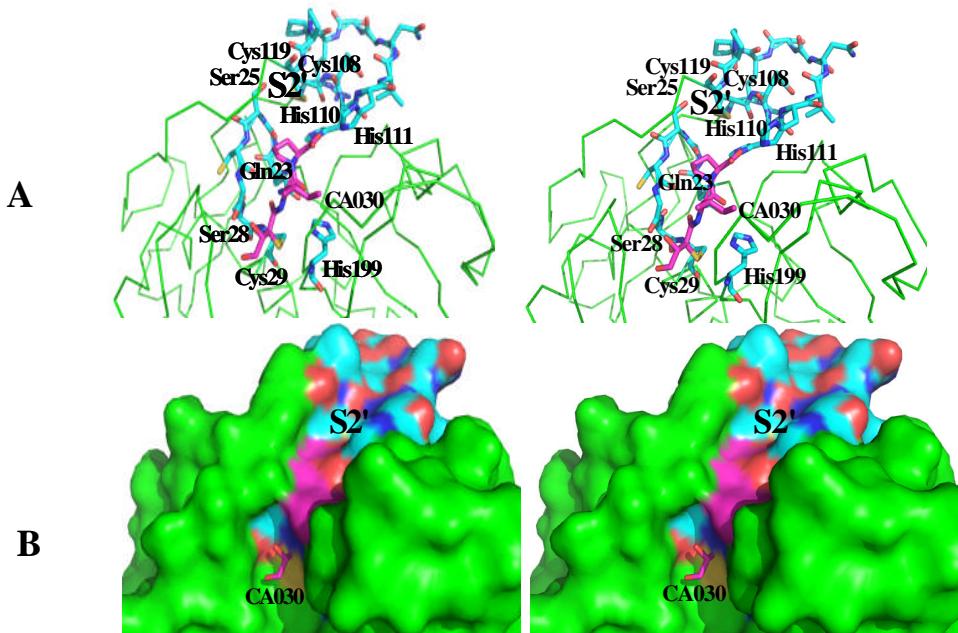


Figure 15. (A) Cathepsin B in complex with CA030 (Pdb accession code 1csb) (Stereo view). The inhibitor, residues of the conserved loop surface from Gln23 to Ser28 (Turk *et al.*, 1998), amino acids of the occluding loop from Cys108 to Cys119, and the catalytic active site residues Cys29 and His199 are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

In the complex structure of cathepsin L with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) the 4-biphenylacetyl groups point into the S2' substrate binding site. They show interactions with Gln21, Gly23, and the main chain of Cys22 (Figure 16) (Chowdhury *et al.*, 2002)

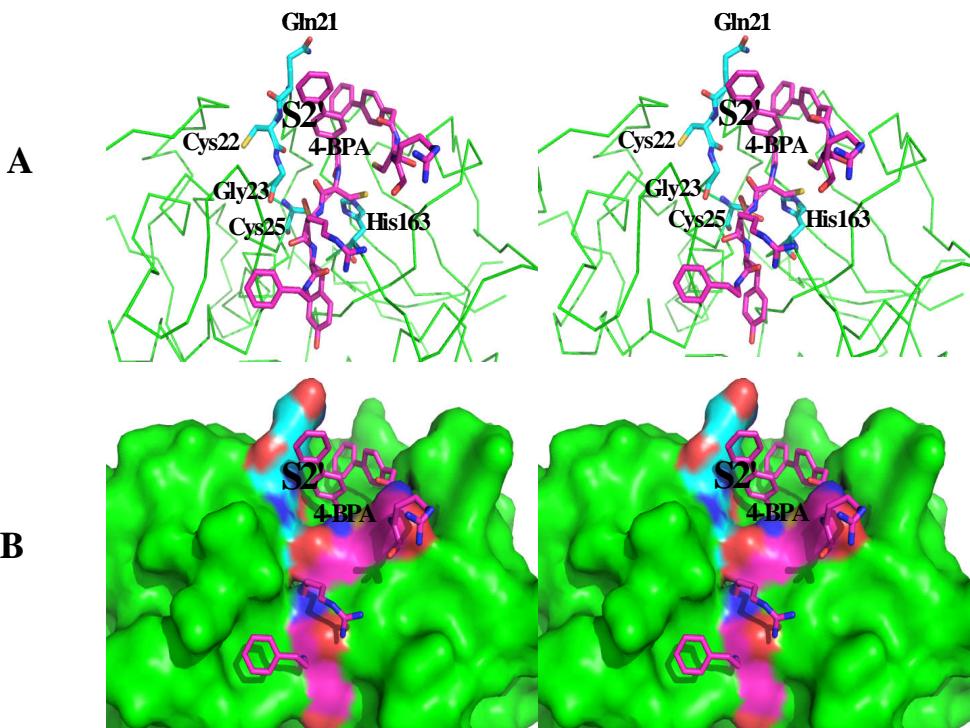


Figure 16. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The inhibitor, S2'-residues interacting with the inhibitor's 4-biphenylacetyl rings (4-BPA), and the catalytic residues Cys25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His163 are shown as sticks (carbons in magenta (inhibitor) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

1.2.2.2.1.6 S3' binding site

Turk *et al.*, 1998 did not see any evidence for an S3' subsite in their revised definition of substrate binding sites in cysteine proteases (Chowdhury *et al.*, 2008). However, in the structure of cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Chowdhury *et al.*, 2002), a dimerisation of the inhibitor took place and part of the second monomer is situated in a location denoted S3' substrate binding site of cathepsin L (Figure 17) (Chowdhury *et al.*, 2008). For the carboxyterminal part of this second inhibitor molecule consisting of the Tyr and the N-(2-phenylethyl)amide moiety, no electron density was observed.

In the visible part, the Cys and D-Arg residues show much higher B factors than the enzyme and the first inhibitor molecule. Only the phenyl benzene group of the second inhibitor molecule packs well between Asn18, Leu144 and Trp189 on one side and Gly20 and the diphenyl group of the first inhibitor molecule on the other side (Figure 17) (Chowdhury *et al.*, 2002). More inhibitors have been synthesised to further study the interactions with the S' subsites (Chowdhury *et al.*, 2008).

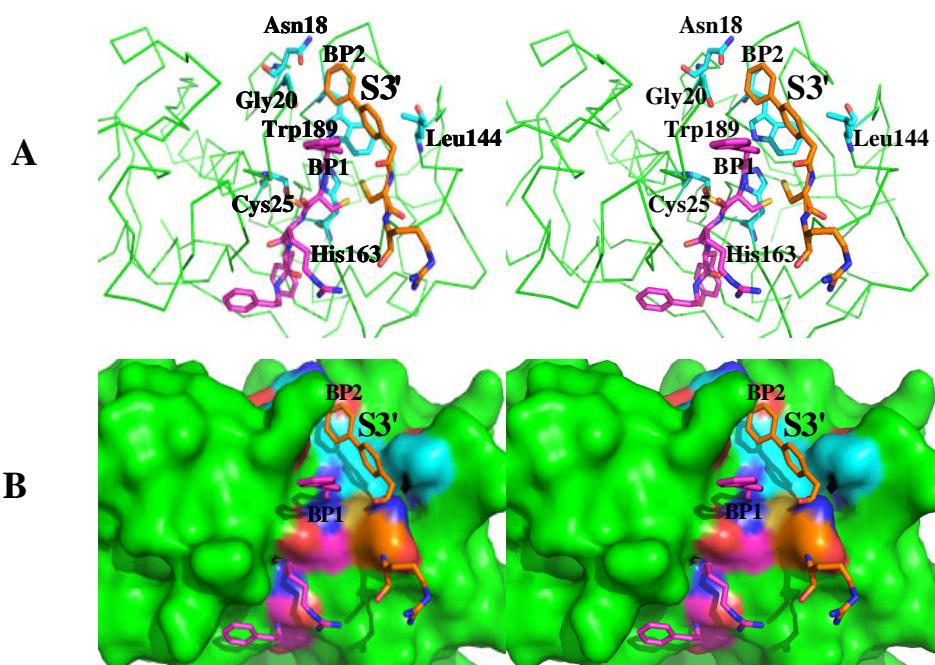


Figure 17. (A) Cathepsin L in complex with 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide (Pdb accession code 1mhw) (Stereo view). The first inhibitor, 4-biphenylacetyl-Cys-D-Arg-Tyr-N-(2-phenylethyl)amide, the second inhibitor molecule, missing the Tyr residue and the N-(2-phenylethyl)amide, residues interacting with the second inhibitor's biphenyl moiety (BP2), and the catalytic amino acids Cys 25 (its sulfur is oxidised to sulfinate) (Chowdhury *et al.*, 2002; Chowdhury *et al.*, 2008) and His 163 are shown as sticks (carbons in magenta (inhibitor molecule 1), orange (inhibitor molecule 2) or cyan (amino acids), nitrogens in blue, oxygens in red and sulfurs in yellow). (B) Surface representation of the same region. This figure has been produced using PyMol (DeLano, 2007).

1.3 Silicate formation in sponges

During the last 3.5 billion years, firstly prokaryotic and later on higher organisms developed strategies to synthesise minerals, predominantly biominerals consisting of CaCO_3 and $\text{Ca}_3(\text{PO}_4)_2$. Only some organisms such as porifera (sponges), radiolarians and some algae use Si(OH)_4 for the formation of biominerals. However, only sponges form biosilica enzymatically (Müller *et al.*, 2007). At the beginning of the life cycle of a sponge (Figure 18), the fertilized egg becomes a larva (Müller *et al.*, 2007). During and after the larvae (Figure 18) have been settled, firstly choanocyte chambers are built that open this body cavity to the external environment (Müller *et al.*, 2007). They are built in membrane-enclosed vesicles within specialized cells called sclerocytes (Simpson, 1984; Shimizu *et al.*, 1998). In the centre of these siliceous spicules there is an organic axial filament of about $2\mu\text{m}$ in diameter (Bütschli, 1901; Müller *et al.*, 2007).



Figure 18. Pictures of adult sponges (left) (taken from <http://upload.wikimedia.org/wikipedia/commons/6/62/SpongeColorCorrect.jpg>) and a sponge larva (right) (taken from <http://www.biology.ualberta.ca/CMD/Pics/LeysLarva.jpg>).

The size of these spicules varies from $1\ \mu\text{m}$ to $3\ \text{m}$. (Schulze, 1904; Müller *et al.*, 2007).

The shallow-water Pacific sponge *Tethya aurantia* has spicules (Figure 19, A) around 2 mm length by 30 µm wide (Shimizu *et al.*, 1998). The silica spicules contribute to 75% of the dry weight of *Tethya aurantia*, the axial filament (Figure 19, B) is only about 0.1% of the mass of each spicule. About 91% of the filament consists of protein (Shimizu *et al.*, 1998).

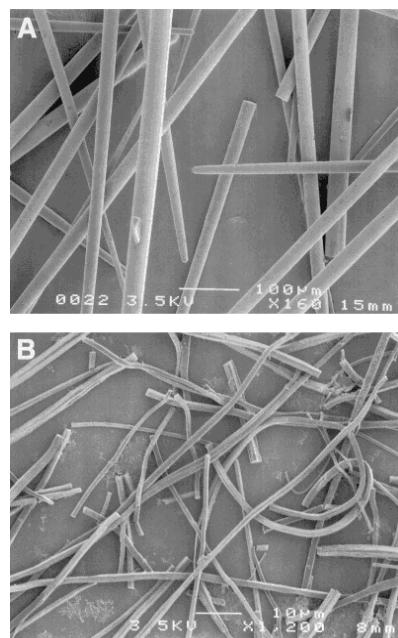


Figure 19. (taken from Shimizu *et al.*, 1998) Scanning electron micrographs of (A) isolated silica spicules (x130) and (B) axial filaments (x 1,000) from *Tethya aurantia*.

To get more insight into the molecular mechanisms that control biosilification, it is very important to study the proteins of the axial filaments (Shimizu *et al.*, 1998). The proteins that had been isolated from the these filaments of the spicules of *Tethya aurantia* by Shimizu *et al.*, have been named silicateins (for *silica* proteins). Electrophoretical analysis of protein revealed the presence of three proteins, silicatein α , β and γ in a ratio 12:6:1, respectively. The amino acid sequence of all three silicateins is also very similar (Shimizu *et al.*, 1998).

1.3.1 Silicatein α , a member of the cathepsin L subfamily

It has been reported that the sequence of silicatein α displays high similarity to members of the papain superfamily (Berti and Storer, 1995; Shimizu *et al.*, 1998). Sequence comparison of silicatein α from *Tethya aurantia* with human cathepsin L (Figure 20) showed almost 50 % identical residues in both enzymes. Silicatein α , like cathepsin L, contains the six Cys residues that create S-S bonds in the protease. That is why it can be assumed that there is similarity between the 3-D structures of silicatein α and cathepsin L. The prosegments of silicatein α and cathepsin L from *Homo sapiens* and sponge also display high similarity, which suggests a common ancestry of both enzymes. The fact that both, silicatein α and cathepsin L, occur in subcellular compartments, also supports this suggestion. It is very interesting that the Cys (SH) at the active site of cathepsin L is changed to Ser (OH) in silicatein α (Shimizu *et al.*, 1998). It has been demonstrated by mutagenesis that this Ser is crucial for the activity of silicatein α (Zhou *et al.*, 1999; Fairhead *et al.*, 2008). The other amino acids in the catalytic triad of the cysteine protease, His163 and Asn187, display similar conservation in silicatein α (Fairhead *et al.*, 2008). Cathepsin L contains a loop of four amino acids (Smith and Gottesman, 1989; Fairhead *et al.*, 2008) which is missing in silicatein α (Fairhead *et al.*, 2008). Silicatein α does not contain as many charged residues as cathepsin L, however, it contains more hydroxyl residues than the cysteine protease (20.3% and 17.5%, respectively). Very specific for silicatein α are clusters of the hydroxyl residues Tyr, Ser and Thr. Examples for such clusters in silicatein α are the motifs Ser-Ser-Arg-Cys-Ser-Ser-Ser, Ser-Ser-Cys-Thr-Tyr, Ser-Xaa-Ser-Xaa-Ser and Ser-Tyr (Shimizu *et al.*, 1998).

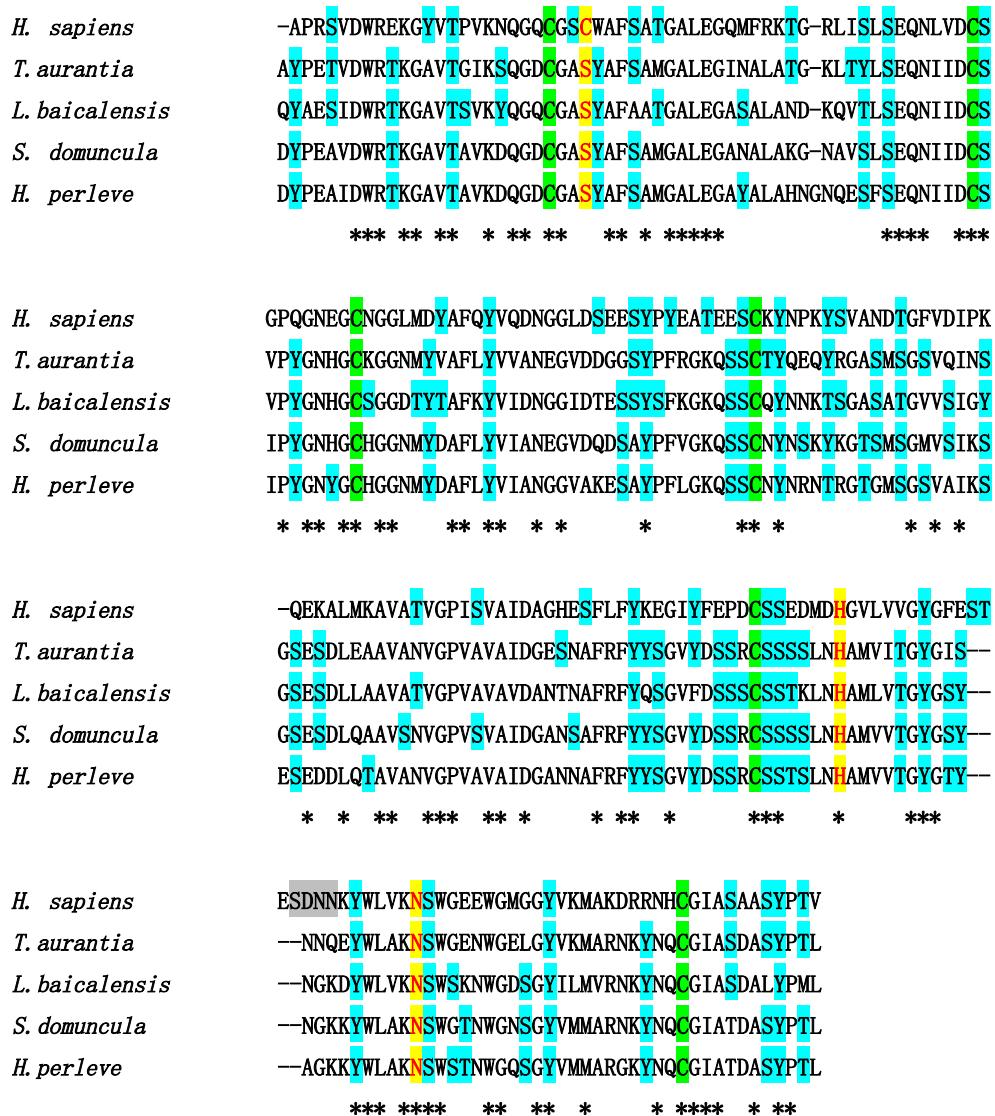


Figure 20. Sequence alignment of mature human cathepsin L (UniProtKB/TrEMBL entry P07711) with mature silicatein α from the sponge *Tethya aurantia* (O76238) (based on Shimizu *et al.*, 1998) and also mature silicatein α from the sponges *Lubomirskia baicalensis* (Q2PC18), *Suberites domuncula* (Q2MEV3) and *Hymeniacidon perleve* (Q2HYF6) (based on Fairhead *et al.*, 2008). Identical amino acids in all five proteins are marked with asterisks. Catalytic triad amino acid residues of the active site of cathepsin L and corresponding residues in the silicateins are highlighted in yellow. Cysteine residues involved in disulfide bridges in cathepsin L are highlighted in green. Hydroxy amino acid residues in cathepsin L and the silicateins are highlighted in turquoise. The 4 amino acid loop in cathepsin L (Smith and Gottesman, 1989; Fairhead *et al.*, 2008) that is missing in the silicateins (Fairhead *et al.*, 2008) is highlighted in grey. The sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

1.3.2 Enzymatic properties of silicateins

Silicatein α synthesises part of the axial filament in the spicules which *in situ* polymerises silicic acid (Figure 21A) (Weaver and Morse, 2003; Fairhead *et al.*, 2008). It has been demonstrated that the polymerisation of siloxanes such as for example tetraethoxysilane (TEOS) can be carried out by both wild type and recombinant silicatein α (Figure 21B) (Shimizu *et al.*, 1998; Cha *et al.*, 1999, Fairhead *et al.*, 2008).

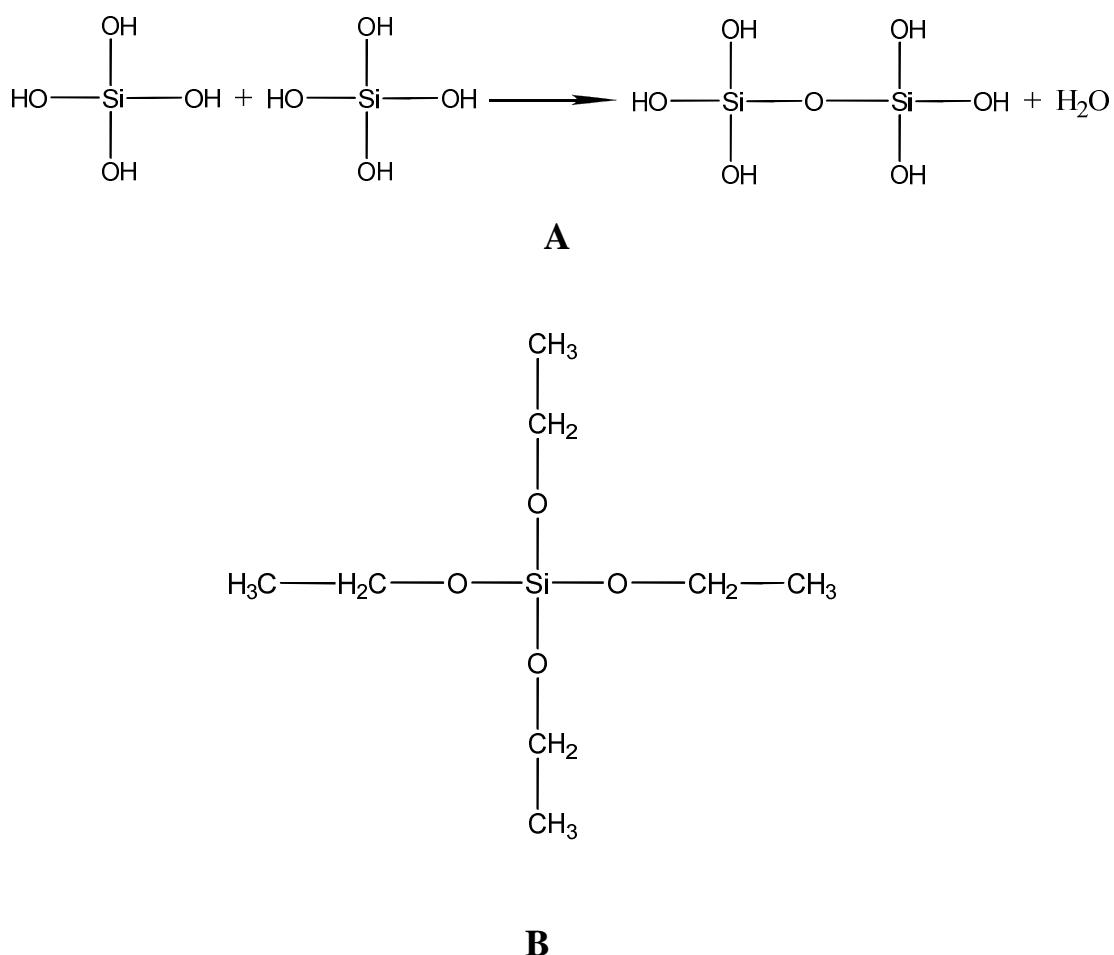


Figure 21. (A) The chemical reaction catalysed by the organic filament of the sponge spicule (based on Fairhead *et al.*, 2008) and (B) the chemical structure of tetraethoxysilane (based on Cha *et al.*, 1999).

Cha and coworkers showed that the three silicateins, α , β and γ in *T. aurantia* accelerate the *in vitro* condensation of silica from TEOS at neutral pH (Cha *et al.*, 1999). Usually, under these circumstances, either an acid or a base catalyst is needed for this reaction (Iler, 1979; Cha *et al.*, 1999). Thermally denatured silicatein subunits do not show any activity indicating the requirement of the native 3-D conformation of the proteins. It has been shown that silicatein α , because of its high content in the silicatein filaments (about 70 % of their mass), is able to accelerate this condensation process in absence of the silicateins β and γ . The silicateins are also able to catalyse the polymerisation of silicon alkoxides and organically modified silicon alkoxides to polymerised silica or silsesquioxanes $(RSiO_{3/2})_n$ at neutral pH. The condensation of organically substituted triethoxysilane precursors ($R-Si(OEt)_3$); R represents $-C_6H_5$, $-CH_3$, etc.) can also be catalysed by active silicatein filaments (Cha *et al.*, 1999). It has also been demonstrated that silicatein α can lead to the deposition of compounds at surfaces, such as l-lactide, ZrO_2 , TiO_2 and $Ti_3(PO_4)_4$ (Curnow *et al.*, 2006; Tahir *et al.*, 2005; Fairhead *et al.*, 2008). It has been shown by Müller *et al.*, 2007, that in the freshwater sponge *L. baicalensis* the filaments did have silica forming properties. Furthermore, it was demonstrated that the specific cathepsin L substrate Rhodamine 10 + bis-(CBZ-L-Phe-L-Arg) (Leytus *et al.*, 1983; Müller *et al.*, 2007) can be digested by the silicateins, suggesting that they can also act as proteases (Müller *et al.*, 2007). The serine, replaced by the catalytic cysteine in cathepsin L that has been proposed to be a very good nucleophile necessary for the digestion of OR- substrates (Cha *et al.*, 1999; Müller *et al.*, 2007), can act like the SH-group as a nucleophile in combination with the H^+ donor histidine in proteolytic cleavage reactions as well. This discovery of both functions, silica condensation and protein cleavage, suggests that there are further modifications in the silicateins, that might play a role in the reaction process. It was also reported that silicatein is a phosphoprotein (putative protein kinase C, casein kinase II and a tyrosine kinase-phosphorylation site(s) and is also glycosylated (Müller *et al.*, 2007).

It was shown that that the 26 kDa silicatein isoform shows glycosylation (Müller *et al.*, 2007). This finding raises the question of whether silicatein γ (named after Shimizu *et al.*, 1998) is actually a different, glycosylated form of the silicateins α or β (Müller *et al.*, 2007).

1.4. Project aims

In biotechnology and drug delivery (Giri *et al.*, 2007; Fairhead *et al.*, 2008) silica materials are of interest but the controlled generation of silicic acid is quite difficult (Fairhead *et al.*, 2008). To get more insight into the molecular mechanisms that control biosilification, it is very important to study the proteins of the axial filaments (Shimizu *et al.*, 1998). Unfortunately, it has not been possible yet to get reasonable amounts of wild type or recombinant silicatein α to perform biophysical studies (Shimizu *et al.*, 1998; Tahir *et al.*, 2004; Curnow *et al.*, 2006; Fairhead *et al.*, 2008).

Cathepsin L mutants able to condense silica from solution have been generated by our collaborators at the University of Strathclyde, Glasgow. Those mutants (Table 1) show sequence features that are specific for silicatein α (Fairhead *et al.*, 2008).

Table 1. (after Fairhead *et al.*, 2008) Mutant constructs made

Cathepsin L construct ^a	Mutations	Match to silicatein α
C2S	C25S	Catalytic serine
AS2	S24A, W26Y	Residues flanking catalytic serine
AS2H	M161L, D162N, G164A, V165M	Residues flanking catalytic histidine
LOOP	¹⁷³ ESTESDNN ¹⁸⁰ to ISSNNQ	To replicate loop
2SER	E159S, D160S	Match serines
4SER	E153S, P154S	Match serines

^aThe mutations are additive from top to bottom,
that is why 4SER contains all the mutations

The mechanism of silicate polymerisation is not fully understood (Fairhead *et al.*, 2008). Therefore, the determination of the crystal structure of the mutant 4SER (Table 1) which displays 91 % sequence identity to human cathepsin L (Figure 22) should provide more insight into the catalytic mechanism of silicic acid polymerisation (Fairhead *et al.*, 2008).

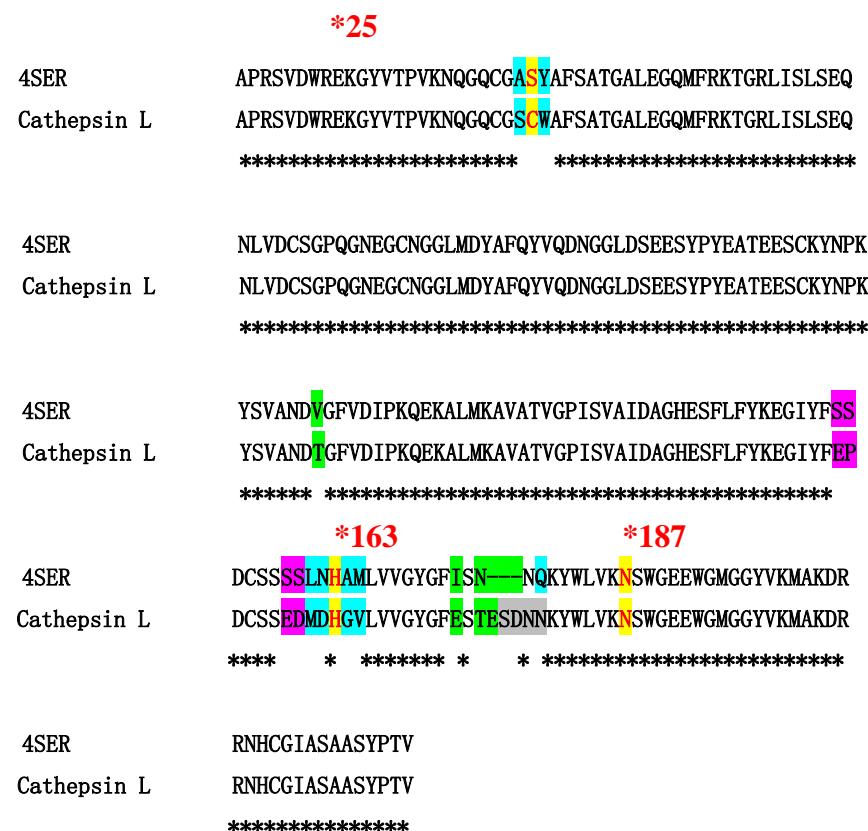


Figure 22. Sequence alignment of the 4SER chimera (Pdb accession code 2vhs) with mature human cathepsin L (UniProtKB/TrEMBL entry P07711). Identical amino acids in both proteins are marked with asterisks. Catalytic triad amino acids of the active site of cathepsin L and corresponding residues in 4SER are highlighted in yellow, and positions corresponding to the sequence of cathepsin L are marked with asterisks. Residues, flanking the catalytic cysteine and histidine in cathepsin L and the corresponding mutations in the chimera are highlighted in cyan. Other mutated amino acids and deletions are highlighted in green. The 4 amino acid loop in cathepsin L (Smith and Gottesman, 1989; Fairhead *et al.*, 2008), which is missing in 4SER (Fairhead *et al.*, 2008) is highlighted in grey. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

Molecular biology, expression, purification and crystallisation of silicatein α -cathepsin L Chimeras

Summary

This section describes molecular biological methods, protein expression and purification of the diverse mutants, the assay performed on the various cathepsin L mutants to detect silica condensation activity and crystallisation of the 4SER mutant. Cloning, introduction of mutations, expression and purification, and the assay were carried out by our collaborators at the University of Strathclyde, Glasgow. Crystallisation of the 4SER mutant was carried out in The Scottish Structural Proteomics Facility (SSPF), University of St. Andrews.

1.5 Molecular biology, protein expression and purification

Mutants were generated using a combination of the QuikchangeTM mutagenesis technique and conventional PCR (Fairhead *et al.*, 2008). The mutant enzymes were then expressed as the proenzyme and purified from *Pichia pastoris* (Fairhead and van der Walle, 2008; Fairhead *et al.*, 2008). Generation of the processed cathepsin L-silicatein α chimeras was then carried out via published procedures (Menard *et al.*, 1998; Fairhead *et al.*, 2008).

1.6 Assay for detection of silica condensation from cathepsin L mutants

Protein in 0.1 M sodium phosphate buffer, pH 7 was mixed with sodium silicate at a concentration of 4.5 mM. The samples were then incubated at room temperature for one day (Fairhead *et al.*, 2008). After the samples had been centrifuged the precipitated silica was then collected and assayed using the Merck spectroquant silicon assay kit (Müller *et al.*, 2003; Fairhead *et al.*, 2008). This assay was performed by our collaborators in Glasgow.

1.7 Crystallisation of the 4SER mutant

The enzyme was screened for crystallisation conditions against a range of commercially available 96-condition screens such as JCSG⁺, Classics and PEGs (GE Healthcare) and Wizard 1 and 2 (Emerald Biosystems) (Supplementary Material (ESI), Fairhead *et al.*, 2008). Conditions that yielded crystals were optimised by employing stochastic screens using in-house software (Rupp *et al.*, 2002; Fairhead *et al.*, 2008, Supplementary Material (ESI)). The crystal used for data collection was grown by the hanging drop vapour diffusion method using protein at a concentration of 10 mg/ml. Protein was mixed with mother liquor containing 17.4% PEG 3350, 0.1 M sodium acetate, pH 4.5 and 0.1 M lithium sulfate in a 2:1 ratio (2 μ l : 1 μ l). Before data collection, the crystal was cryoprotected in a solution containing 20 % PEG 3350, 0.1 M sodium acetate, pH 4.5, 0.1 M lithium sulfate and 20% PEG 400. The crystal was flash frozen in liquid nitrogen and taken to beamline ID29 at the European Synchrotron Radiation Facility (ESRF), Grenoble. (Fairhead *et al.*, 2008, Supplementary Material (ESI)).

Data collection, Processing, Molecular Replacement and Refinement of the 4SER chimera

Summary

A complete X-ray dataset was collected at the European Synchrotron Radiation Facility (ESRF) beamline ID29 to a resolution of 1.5 Å. Crystals were obtained as described in Section 1.7. The diffraction data were processed and the structure factor amplitudes calculated.

The sequence of the 4SER chimera is almost identical to that of cathepsin L (Fairhead *et al.*, 2008). It was decided to use cathepsin L as a model to calculate a set of initial phase estimates using the technique of Molecular Replacement (MR). After the structure had been solved using PHASER (McCoy *et al.*, 2005, Storoni *et al.*, 2004), the protein was modified manually to change the sequence to 4SER. Full high-resolution anisotropic refinement, TLS parameters and Non crystallographic symmetry (NCS) restraints were carried out in all steps of refinement using the programme REFMAC5 (Murshudov *et al.*, 1999). The final model was checked and validated using PROCHECK (Laskowski *et al.*, 1993) and MOLPROBITY (Davis *et al.*, 2004). The final structure and experimental data were deposited with the Pdb (Boutselakis *et al.*, 2003), accession code 2vhs (Fairhead *et al.*, 2008, Supplementary Material (ESI)).

1.8 Data collection and processing

Before data collection crystals were cryoprotected in a solution containing 20% PEG 3350, 0.1 M sodium acetate, pH 4.5, 0.1 M lithium sulfate and 20% PEG 400.

Diffraction data were collected at -173.15 °C at a wavelength of 1.27 Å on ID29 at the ESRF. (Fairhead *et al.*, 2008, Supplementary Material (ESI)). A high resolution pass to 1.5 Å resolution was collected in 0.5° oscillations with a 0.25s exposure. A lower resolution pass to 2 Å in 0.5° oscillations with a 0.1s exposure that provided measurements of data overloaded on the first pass was collected as well (Figure 23). Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). All the data collection, the indexing and merging of the data were carried out by people from the SSPF. The crystal belonged to the space group P1 (Table 2) with unit cell parameters $a = 56.8 \text{ \AA}$, $b = 58.1 \text{ \AA}$, $c = 70.2 \text{ \AA}$, $\alpha = 105.7^\circ$, $\beta = 105.0^\circ$, $\gamma = 105.1^\circ$. The data have an overall completeness of 92% and 74% in the highest resolution shell from 1.55 – 1.5 Å resolution. Data collection and processing statistics are presented in Table 3.

Table 2. Output from POINTLESS (Evans, 2006)

Best Solution: space group P 1
Reindex operator: [h,k,l]
Laue group probability: 0.918
Systematic absence probability: 1.000
Total probability: 0.918
Space group confidence: 0.877
Laue group confidence: 0.877

Table 3. (after Fairhead *et al.*, 2008, Supplementary Material (ESI)) Data collection statistics for the 4SER crystal. Data for the highest resolution shell are shown in brackets.

Data collection	4SER
λ (Å)	1.27
Resolution	28 – 1.5
Last shell (Å)	(1.55 - 1.5)
Spacegroup	P1
Cell (Å, °)	$a = 56.8, b = 58.1, c = 70.2$ $\alpha = 105.7, \beta = 105.0, \gamma = 105.1$
Unique reflections	108723
Average redundancy	1.9 (1.7)
I/ σ	20 (2.1)
Completeness (%)	92 (74)
R_{merge}^*	0.039 (0.293)

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

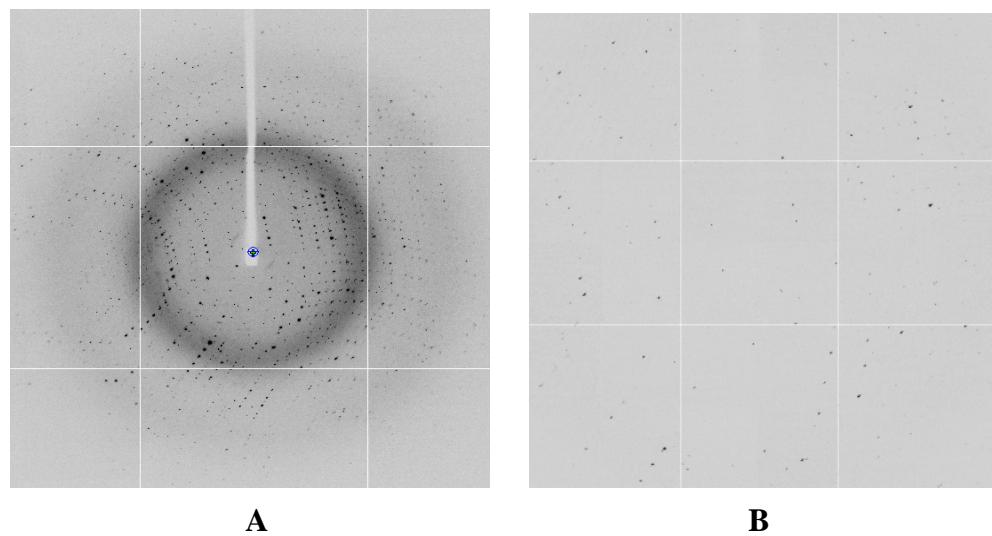


Figure 23. 0.5° oscillation X-ray image of the crystal of 4SER collected to a higher resolution pass to 1.5 Å (A) and a lower resolution pass to 2.0 Å (B) at ID29, ESRF.

1.9 Structure solution and refinement

The structure of the 4SER chimera was solved by molecular replacement using Phaser (McCoy *et al.*, 2005; Storoni *et al.*, 2004). A monomer of cathepsin L (PDB accession code 1mhw) showing an almost identical sequence to the chimera (Fairhead *et al.*, 2008) was used as a search unit. A solution was found for 4 monomers in the asymmetric unit consistent with the Matthews coefficient calculation (Table 4) (Matthews, 1968) and a resolution range between 28 and 2.5 Å.

Table 4. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(1.5) = probability at 1.5 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of 2.08 Å³Da⁻¹, corresponding to a 40.9% solvent content at a probability of 0.81 at 1.5 Å and a total probability of 0.53 (shown in orange), suggested 4 molecules of the 4SER chimera in the asymmetric unit.

Nmol/asym	Matthews Coeff.	% solvent	P(1.50)	P(tot)
1	8.32	85.22	0.00	0.00
2	4.16	70.45	0.00	0.02
3	2.77	55.67	0.17	0.44
4	2.08	40.90	0.81	0.53
5	1.66	26.12	0.01	0.00
6	1.39	11.34	0.00	0.00

Clear rotation and translation solutions were found for each of the 4 monomers. No unfavourable molecular contacts were observed for the final solution in the crystal packing (Table 5).

The Molecular Replacement solution was then applied to the coordinates of the search model to work as the starting model in the refinement described in the following section.

Table 5. The top solution from PHASER.

RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes
(unfavourable molecular contacts in crystal packing)

Solution	RFZ	TFZ	PAK	RFZ	TFZ	PAK	RFZ	TFZ	PAK	RFZ	TFZ	PAK
1	25.6	100	0	26	28.4	0	20.5	80.2	0	27.6	75.2	0
Monomer 1			Monomer 2			Monomer 3			Monomer 4			

1.9.1 Refinement

Refinement is performed to minimise the difference between experimentally observed structure factors (F_{obs}) and model-derived structure factors (F_{calc}). Thereby refinement of the coordinates x, y and z and the temperature factor (B-factor) of each atom in the model is carried out. For the judgement of the resultant model various qualitative and quantitative factors are used. Normally, the standard crystallographic Refinement factors R_{work} and R_{free} are compared and one checks that all bond-lengths, angles and temperature factors are within normal ranges. REFMAC5 (Murshudov *et al.*, 1999) was used to refine to structure of the chimera. Data in the range 63 – 1.5 Å were used. The protein was manually modified to change the sequence to 4SER and rebuilding of several amino acid residues were performed in COOT (Emsley and Cowtan, 2004). TLS parameters (Table 6), anisotropic B-thermal factors and NCS restraints were employed in all cycles of refinement. Water molecules and sulfate ions were also added in COOT (Emsley and Cowtan, 2004) (Fairhead *et al.*, 2008, Supplementary Material (ESI)).

Table 6. TLS groups employed throughout the refinement of 4SER.

TLS						
RANGE 'A' 1.' 'A 220.' ALL						
ORIGIN -4.431 -27.884 17.511						
T	-0.1207	-0.1366	-0.1425	-0.0046	-0.0032	-0.0076
L	2.0108	1.4785	1.2081	-0.3484	-0.3503	-0.1250
S	-0.1014	0.1305	0.0815	0.0730	-0.0943	-0.0881
	0.0356	0.0996				
TLS						
RANGE 'B' 1.' 'B 220.' ALL						
ORIGIN 11.117 -8.108 1.610						
T	-0.1060	-0.1252	-0.1227	-0.0169	0.0098	-0.0033
L	2.9163	1.2744	1.1387	0.1873	-0.3902	-0.2063
S	0.0662	-0.1209	-0.0503	-0.2041	0.0963	-0.0461
	0.0421	-0.0531				
TLS						
RANGE 'C' 1.' 'C 220.' ALL						
ORIGIN 22.774 7.357 32.647						
T	-0.1276	-0.1383	-0.1558	-0.0104	-0.0220	-0.0078
L	1.3113	1.5386	1.0548	-0.1951	-0.3159	-0.1475
S	0.0283	-0.0353	0.0089	0.0049	0.0970	-0.0279
	-0.0129	-0.0588				
TLS						
RANGE 'D' 1.' 'D 220.' ALL						
ORIGIN 7.192 -12.379 48.556						
T	-0.1168	-0.1331	-0.1624	-0.0121	-0.0119	-0.0145
L	1.6144	1.4337	1.0725	0.0494	0.0517	-0.3844
S	-0.0310	0.0141	0.0417	0.0125	-0.0472	-0.0362
	0.0545	0.0530				

The R_{work} and R_{free} after the last round of refinement were 18.8 % and 22 %, respectively (Table 7). The quality of the structure was judged with PROCHECK (Laskowski *et al.*, 1993) and MOLPROBITY (Table 8) (Davis *et al.*, 2004).

Finally the 4SER structure and the experimental data were deposited with the Pdb (Boutselakis *et al.*, 2003) and have the accession code 2vhs (Fairhead *et al.*, 2008, Supplementary Material (ESI)).

Table 7. Refinement statistics of 4SER (after Fairhead *et al.*, 2008, Supplementary Material (ESI)). Values for R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Refinement	4SER
R _{work} * %	18.8 (26.4)
R _{free} * %	22 (32.1)
Number of atoms (average B value (Å ²))	
protein (overall)	6636 (24.4)
protein (chain A)	1661 (24.4)
protein (chain B)	1657 (24.4)
protein (chain C)	1661 (24.4)
protein (chain D)	1657 (24.4)
water (overall)	1172 (36.6)
SO ₄ ²⁻	40 (36.9)
NCS deviations (rmsd (Å))	
(chain A - chain B)	0.07
(chain A - chain C)	0.08
(chain A - chain D)	0.07
rmsd bonds (Å) / angles (°)	0.006 / 0.985
Ramachandran favoured (%)	98
PDB accession code	2vhs

$$* R_{\text{work}} = \sum_{\text{hkl}} | |F_{\text{obs}} - |F_{\text{calc}}| | / \sum_{\text{hkl}} |F_{\text{obs}}|$$

* For R_{free} the same calculation applies, but for only 5-10 % of reflections which are not included in the refinement.

Table 8. Output from MOLPROBITY (Davis *et al.*, 2007)

All-Atom Contacts	Clashscore, all atoms:	4.59	94th percentile* (N=598, 1.25Å - 1.75Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Rotamer outliers	2.31%	Goal: <1%
	Ramachandran outliers	0.00%	Goal: <0.2%
	Ramachandran favored	97.99%	Goal: >98%
	C_B deviations >0.25Å	0	Goal: 0
	MolProbity score	1.51	84th percentile* (N=4836, 1.25Å - 1.75Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.00%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

Discussion

A crystal of the 4SER chimera diffracted to a resolution of 1.5 Å at the ESRF beamline ID29. A complete data set was obtained at this resolution combining data to 1.5 Å from a high resolution pass and to 2 Å from a lower resolution pass. This was necessary to provide the data that were overloaded in the higher resolution pass. After an appropriate starting orientation had been given, a 180° rotation range at 0.5° oscillations was sufficient for the space group P1 to get a completeness of almost 100%. The presence of more than one lattice in the diffraction pattern due to multiple crystals caused some problems in indexing and integrating the data. For Molecular Replacement a monomer of cathepsin L was used as a search unit finding 4 monomers in the asymmetric unit. Both sequences, the one from the 4SER chimera and Cathepsin L are almost identical apart from some mutations in the chimera (Fairhead *et al.*, 2008). In PHASER the Z-scores for the rotation and translation functions for each of the four monomers are very high (Table 5) indicating that the correct solution was found. Z-scores more than 10 normally indicate a correct solution.

Refinement of the model against the data was straightforward. High resolution data have a hugely increased observation to parameter ratio. This means that normally in a 1.5 Å resolution structure it is not necessary to “add” observations in the form of restraints. However, application of NCS restraints improved the values for both R_{work} and R_{free} by about 1 %. The geometry of the final 4SER model agreed well with the average geometry of well-refined structures as seen by the rmsd values of bond lengths and angles (Rhodes, 2000). Deviations of bond lengths around 0.01 and angle deviations of about 1.6° are expected for good quality models. Rmsd bonds and angles of the chimera at the end of the refinement process are within this range, 0.006 Å and 0.985°, respectively (Table 7). A clashscore of all atoms of 4.59 (94rd percentile) and a MOLPROBITY score of 1.51 (84th percentile) (Table 8) also indicate the good quality of this structure.

Structure of the 4SER chimera

Summary

The structure of the 4SER chimera, solved to 1.5 Å resolution, is described in this section. The structure is composed of 9 α -helices and 10 β -strands. It is highly similar to cathepsin L. Like cathepsin L, the chimera comprises two domains, resembling a closed book with the spine in front (Turk et al., 1998). Both “domains separate on the top in a ‘V’ shaped active site cleft” (Turk et al., 1998). The catalytic cysteine (C25) in cathepsin L has been mutated to serine. The other amino acids in the catalytic triad of cathepsin L (His163 and Asn187) are the same in the 4 SER chimera. There is a sulfate ion at the active site of each of the four monomers due to the presence of lithium sulfate in the crystallisation condition. A second sulfate ion is located 17 Å distant from Ser25. The structure provides insights into the catalytic mechanism of silicic acid condensation (Fairhead *et al.*, 2008).

1.10 Overall structure

Unsurprisingly, because of their almost identical sequences (Fairhead *et al.*, 2008), the chimera shares the same fold as cathepsin L. The asymmetric unit of 4SER is composed of four monomers (Figure 24).

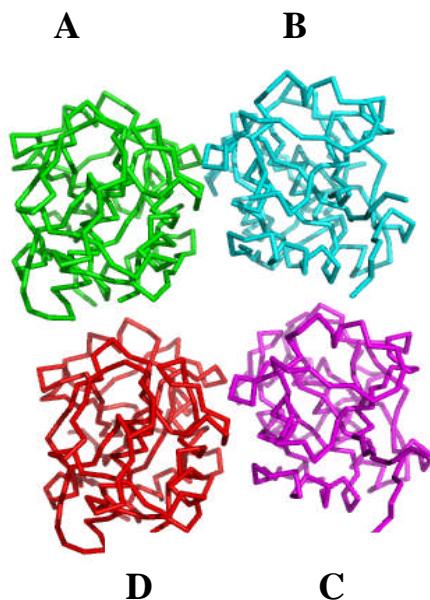


Figure 24 The asymmetric unit of 4SER showing four monomers. The C α traces of the monomers A, B, C and D are shown in green, cyan, magenta and red, respectively. This figure has been produced using PyMol (DeLano, 2007).

The four subunits are essentially identical. Superposition of the monomers B, C and D to A shows a rmsd of 0.07 Å, 0.08 Å, 0.07 Å, respectively. All four superimposed monomers are shown as C α traces in Figure 25.

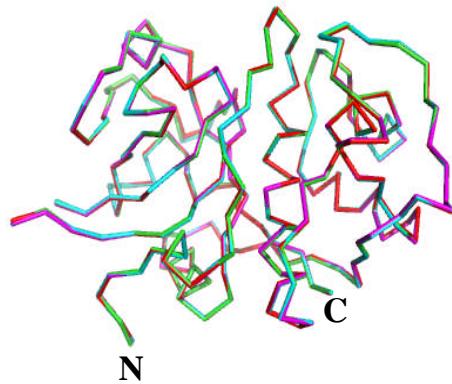


Figure 25. Ca tracing of the four monomers A (green), B (cyan), C (magenta) and D (red) after their superposition. This figure has been produced using PyMol (DeLano, 2007).

The interfaces between the subunits in the asymmetric unit were analysed using the programme PISA (Krissinel and Henrick, 2005). The interface areas between the monomers are shown in Table 9.

Table 9. Interface areas between the monomers of 4SER in the asymmetric unit. The Complexation Significance Scores (CSS) for each interface are also shown.

Monomers	Interface area (\AA^2)	CSS
B and A	708.9	0
D and C	703.8	0
B and D	484.8	0
A and C	482.4	0
D and A	320.6	0
C and B	314.2	0
B and C	271.3	0
A and D	269.3	0

The analysis did not reveal any specific associations between the monomers that could cause any formation of stable quaternary structures. Therefore it suggests that there is no complexation in solution. This result confirms the fact that cathepsins, except the tetrameric cathepsin C usually occur as monomers (Mc Grath, 1999; Turk *et al.*, 2001).

The 4SER chimera from amino acid Ala1-Val220 consists of 9 α -helices and 10 β -strands (Figure 26A). The heavy aminoterminal chain, composed of the amino acids Ala1-Ser174 and the light carboxyterminal chain (Gln180-Val220) are connected via an S-S bridge between Cys156 and Cys209 (Figure 26B). Like in cathepsin L (Fujishima *et al.*, 1997), the loop region between amino acids Ser174 and Gln180 is disordered (Figure 26).

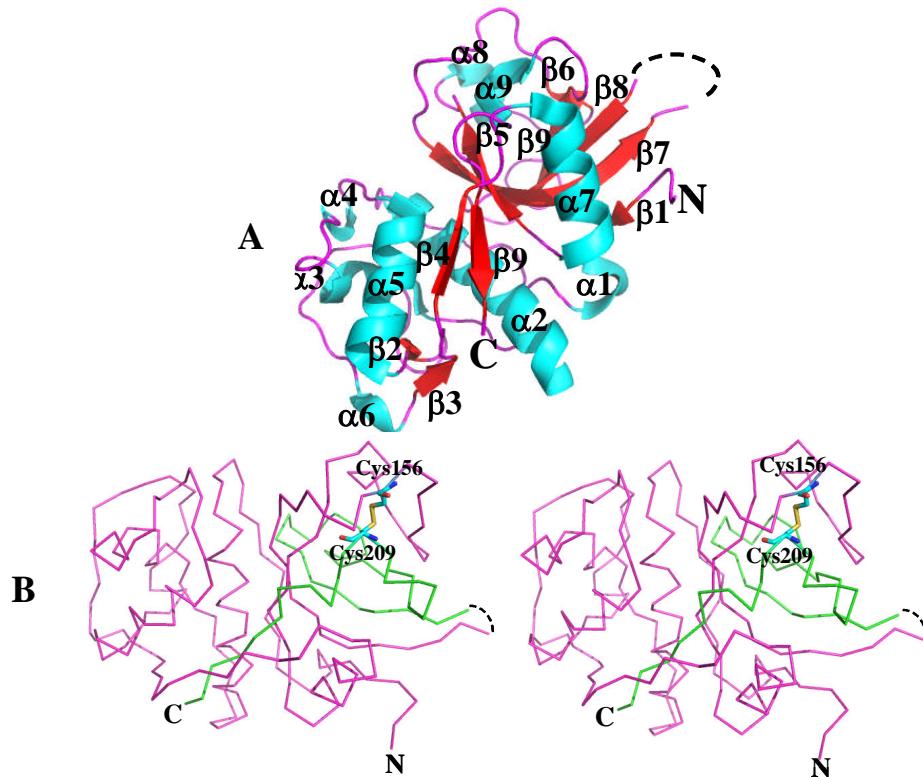


Figure 26. (A) Ribbon diagram of the structure of the 4SER chimera (Pdb accession code 2vhs). α -helices (in cyan) and β -sheets (in red) are labelled. Loops are shown in magenta. The disordered loop between β -strands 7 and 8 is shown as dotted lines. (B) Stereo view of the connection of the heavy and light chain of 4SER (shown as C α traces in magenta and green, respectively) via the S-S bridge between the heavy chain residue Cys156 and Cys209 (shown as sticks, carbons in cyan). The disordered loop region is shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

1.11 Comparison of 4SER to cathepsin L

The differences between 4SER and cathepsin L only lie in the mutations and deletions introduced into the chimera. The catalytic cysteine (Cys25) in cathepsin L has been mutated to serine in the chimera. The other amino acids in the catalytic triad of cathepsin L (His163 and Asn187) have not been changed in the 4 SER chimera. The residues flanking the catalytic Ser25 and His163 have also been substituted; Ser24 and Trp26 in cathepsin L to Ala24 and Tyr26 in 4SER and Met161, Asp162, Gly164 and Val165 in the cysteine protease to Leu161, Asn162, Ala164 and Met165 in the chimera. The amino acids Glu153, Pro154, Glu159 and Asp160 in cathepsin L have all been replaced by serines in 4SER (Fairhead *et al.*, 2008) to create the serine rich region Ser-Ser-Asp-Cys-Ser-Ser-Ser-Ser, which almost completely matches the cluster Ser-Ser-Arg-Cys-Ser-Ser-Ser-Ser, present in silicatein α of *Tethya aurantia* (Shimizu *et al.*, 1998). Instead of the loop in cathepsin L (Smith and Gottesman, 1989; Fairhead *et al.*, 2008), that consists of the amino acids Ser177 to Asp180 and the flanking residues Glu173 to Glu176, 4SER contains the sequence Ile-Ser-Asn-Asn-Gln of silicatein α (Fairhead *et al.*, 2008). The substitution of Thr110 in cathepsin L to valine disrupts the glycosylation site in the cysteine protease (Fairhead and van der Walle, 2008).

The structure of the 4SER chimera is, unsurprisingly, very similar to cathepsin L (Fairhead *et al.*, 2008). The overall rms deviation between both structures is only 0.25 Å for 188 Ca positions considered equivalent (Figure 27).

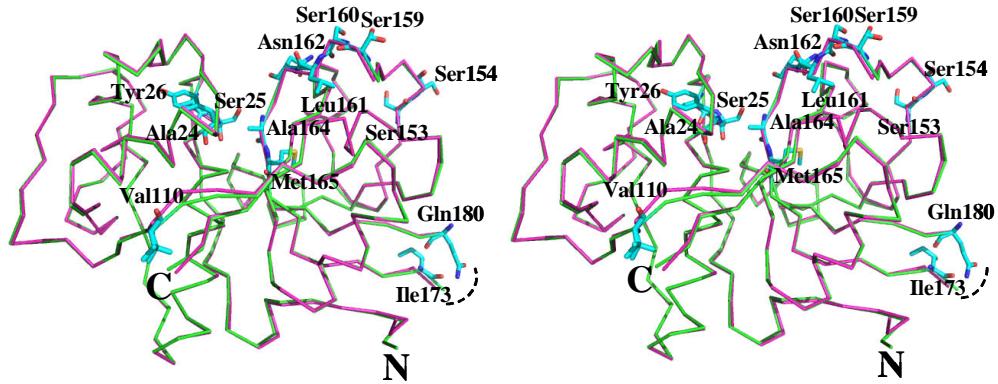


Figure 27. Stereo view of Ca tracing of chain A of 4SER (in green, Pdb accession code 2vhs) and chain A of human cathepsin L (in magenta, Pdb accession code 1mhw) after their superposition. Mutated residues in the chimera are shown as sticks (carbons in cyan) and are labelled. The disordered loop region in both structures is shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

1.12 Active site of 4SER

The chimera is composed, like cathepsin L and other enzymes of the papain superfamily, of two domains, resembling a closed book with the spine in front (Turk *et al.*, 1998). The domains show a separation “on the top in a ‘V’ shaped active site cleft”. (Turk *et al.*, 1998). In its centre there are the catalytic residues Ser25 and His163, one from each domain (Turk *et al.*, 1998). Ser25 is part of the central α -helix (α 2), His163 is right at the beginning of β 7, which almost spans to the end of the heavy chain of 4SER. The amino acids, flanking both catalytic residues in the chimera (Ala24, Tyr26, Leu161, Asn162, Ala164 and Met165) correspond to those in silicatein α (Fairhead *et al.*, 2008) (Figure 28A and B).

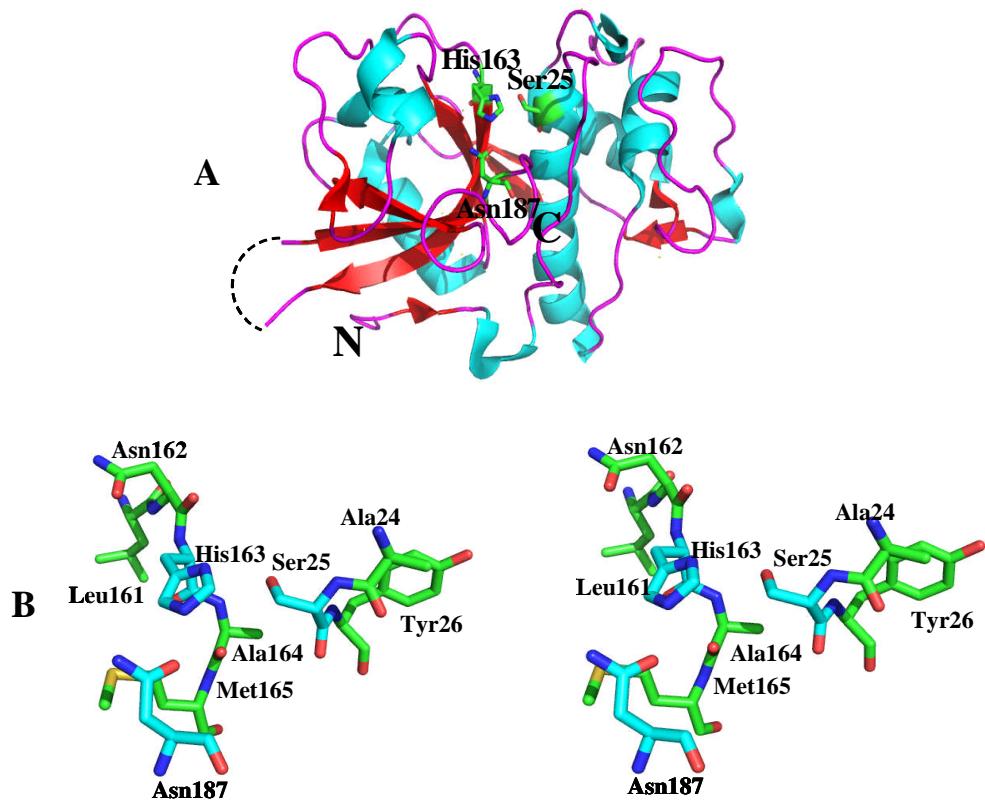


Figure 28. (A) Ribbon diagram of the 4SER chimera, showing the “V” shaped active site cleft” (Turk *et al.*, 1998) with the catalytic residues Ser25 and His163 (Turk *et al.*, 1998) (in sticks, carbons in green) in its centre. The third residue of the catalytic triad, Asn 187 (Fairhead *et al.*, 2008), is also shown (in sticks, C-atoms in green) α -helices are shown in cyan, β -sheets in red and loops in magenta. The disordered loop region is shown as dashed lines. (B) Stereo view of the active site residues Ser25 and His163 (in sticks, carbons in cyan) and their flanking residues (in sticks, C-atoms in green). The third amino acid of the catalytic triad, Asn187 (Fairhead *et al.*, 2008), is also shown (in sticks, carbons in cyan). This figure has been produced using PyMol (DeLano, 2007).

1.13 Sulfate ions in the structure of 4SER

There is clear electron density for two sulfate ions in each of the four monomers (Figure 29) due to the presence of lithium sulfate in the crystallisation solution. One molecule is in close contact to Ser25 and His163, the second one is at a crystal contact 17.2 Å distant from Ser25 (Fairhead *et al.*, 2008) (average distance over the four monomers) (Figure 30).

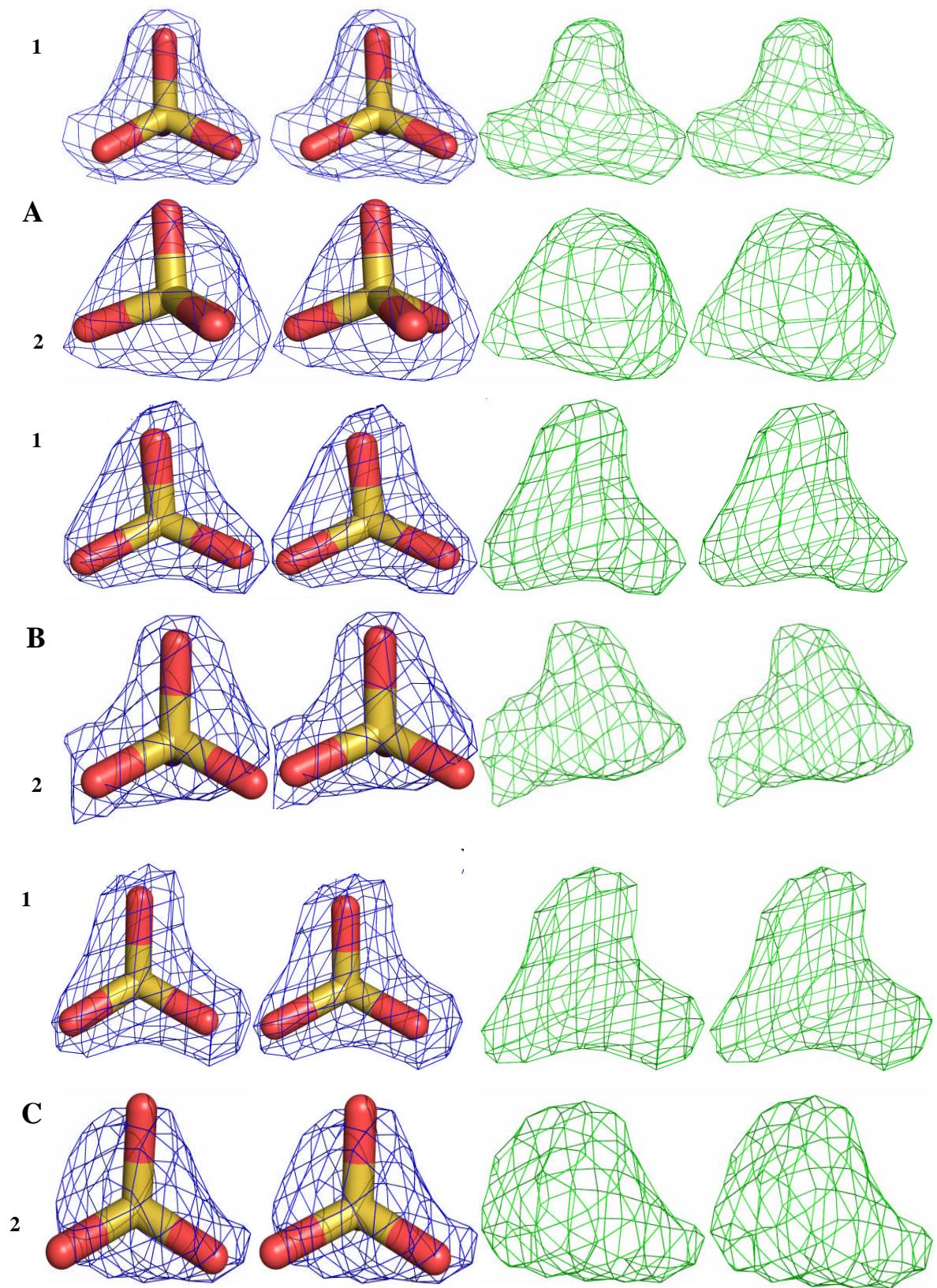


Figure 29. (continues on next page)

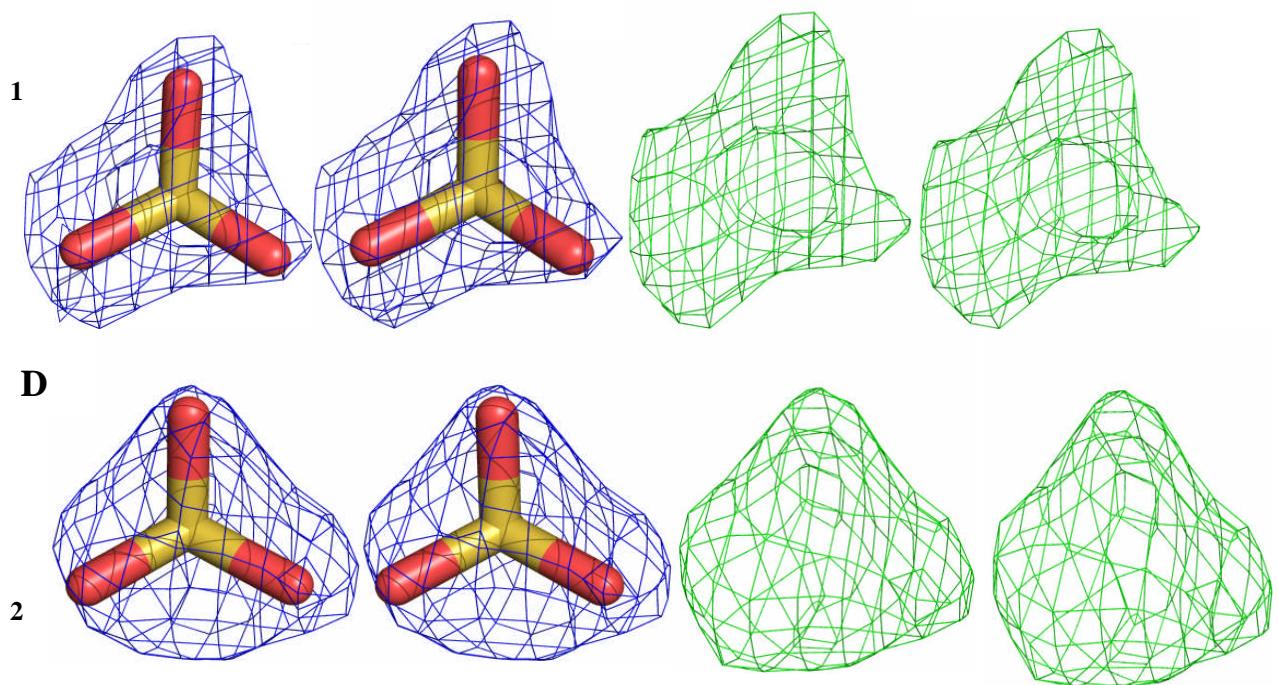


Figure 29. Final $2F_o - F_c$ at 1σ (blue) and unbiased $F_o - F_c$ maps at 3σ (green) around the sulfates at the active site (1) and about 17 \AA distant from Ser25 (2) of each monomer of 4SER (Stereo view). The SO_4^{2-} molecules are modelled in the $2F_o - F_c$ densities. This figure has been produced using PyMol (DeLano, 2007).

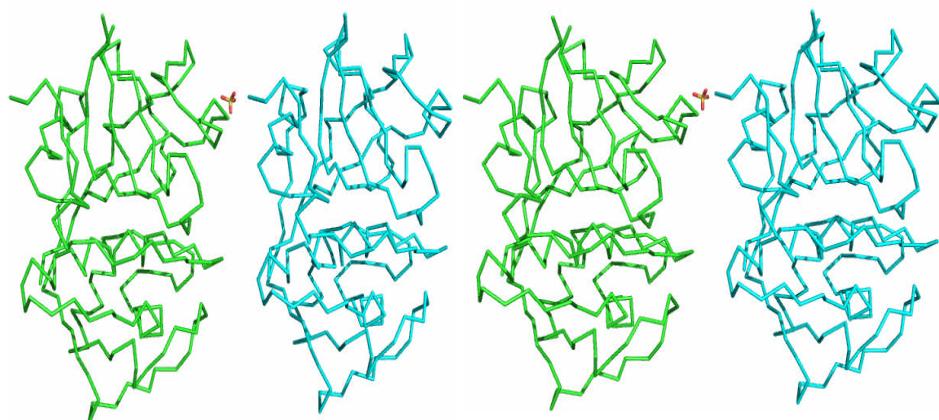


Figure 30. Stereo view showing one of the sulfate ions (shown as sticks) at a crystal contact between monomer A (shown as $\text{C}\alpha$ traces in green) and a crystallographically related monomer (shown as $\text{C}\alpha$ traces in cyan). This figure has been produced using PyMol (DeLano, 2007).

The importance of the sulfate at the active site of 4SER will be one of the main subjects in the following discussion.

Discussion

Various human cathepsin L mutants have been generated by our collaborators at the University of Strathclyde, Glasgow, that increasingly match the sequence, specific for silicatein α from the marine sponge *Tethya aurantia* (Table 1) (Fairhead *et al.*, 2008). Our collaborators also assayed those cathepsin L mutants for silica condensation activity (Figure 31) (detected by the precipitation of silicate).

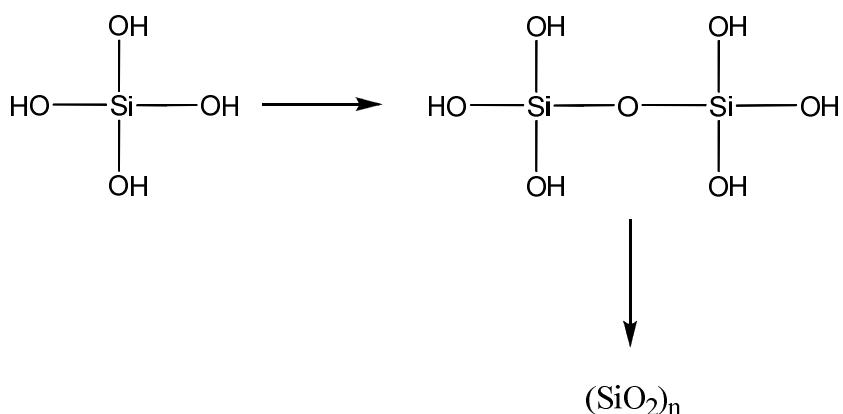


Figure 31. The mechanism of silica condensation.

The mutant C2S, in which the catalytic Cys25 has been replaced by serine was essentially unable to condense silica (Fairhead *et al.*, 2008) (Figure 33). This substitution would have been expected to be enough from the already proposed mechanism of silicatein α (Cha *et al.*, 1999; Fairhead *et al.*, 2008) (Figure 32).

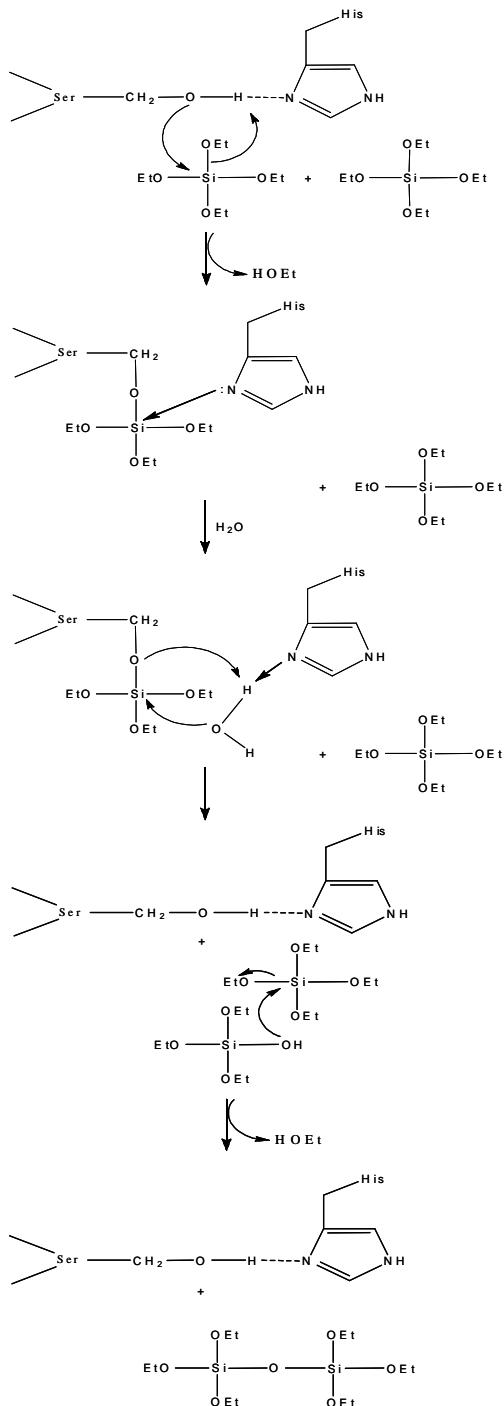


Figure 32. (after Cha *et al.*, 1999) Proposed reaction mechanism of tetraethoxysilane polymerisation carried out by silicatein α , based on the well studied catalytic reaction performed by Ser /His and Cys / His active-site proteases (Cha *et al.*, 1999; Zhou *et al.*, 1999).

For this mechanism it has been proposed that H-bonding between N of the imidazole ring of His165 and OH of Ser26 increases the nucleophilicity of the O-atom of the serine which enforces its attack on the Si of tetraethoxysilane.

Due to this nucleophilic attack $\text{CH}_3\text{-CH}_2\text{-OH}$ is released and a covalent enzyme-O-Si intermediate is generated. Hydrolysis of this first RO^- bond is completed by the addition of H_2O . The disiloxane product is then finally generated by the condensation reaction. The starting point for this process is that the released Si-O^- performs a nucleophilic attack on the silicon of the second Si(OEt)_4 (Cha *et al.*, 1999; Zhou *et al.*, 1999).

Our collaborators in Glasgow observed that it was necessary to change the residues flanking Ser25 to detect any significant silica precipitation (Figure 33). Additional substitutions to cathepsin L to increase its match to silicatein α did not improve the ability of these mutants to condense silica (Fairhead *et al.*, 2008) (Figure 33).

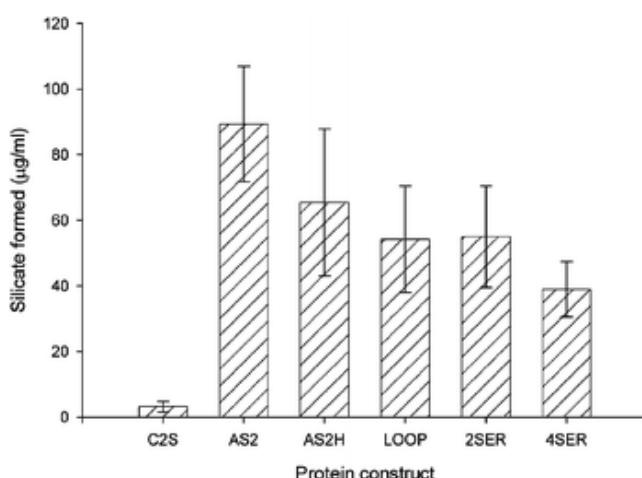


Figure 33. (taken from Fairhead *et al.*, 2008) Silica condensing activity of the mutants C2S, AS2, AS2H, LOOP, 2SER, 4SER (see Table 1) generated by our collaborators. Protein in 0.1 M sodium phosphate buffer, pH 7 was mixed with sodium silicate at a concentration of 4.5 mM. The samples were then incubated at room temperature for one day (Fairhead *et al.*, 2008). After the samples had been centrifuged the precipitated silica was then collected and assayed using the Merck spectroquant silicon assay kit (Müller *et al.*, 2003; Fairhead *et al.*, 2008).

Moreover, many of these additional substitutions actually showed a decrease in their ability to condense silica (Figure 33) (Fairhead *et al.*, 2008). Therefore, it is possible that these amino acids play other roles in the sponge spicule. They could participate in interactions of silicatein α with silicatein β and galectin (Curnow *et al.*, 2005; Schröder *et al.*, 2006; Fairhead *et al.*, 2008). The mutants were more or less able to precipitate silica from the natural substrate sodium silicate, but not from TEOS. To elucidate the mechanism of catalysis of silica polymerisation the crystal structure of the 4SER chimera was solved to 1.5 Å resolution (Fairhead *et al.*, 2008). The C25S mutant of procathepsin L (Pdb accession code 1cjl) cannot function as a protease, which indicates that this mutation from Cys (SH) to Ser (OH) caused a significant decrease in nucleophilicity. In 4SER an increase of the distance between the OG atom of Ser25 and ND1 of His163 to 3.6 Å (Fairhead *et al.*, 2008) (average distance over the four monomers) can be observed (Figure 34), which further lowers the nucleophilicity of Ser25. Mutation of the residues flanking Ser25 perturbs a cluster of amino acids that are located behind Ser25. The substitution S24A leads to a removal of an H-bond and the replacement of Trp26 with Tyr decreases a hydrophobic interaction. The substitutions are both to amino acids smaller than in the wildtype protein. As a consequence Ser25 would be able to decrease its distance to His163 and the volume of the catalytic region would be increased (Fairhead *et al.*, 2008).

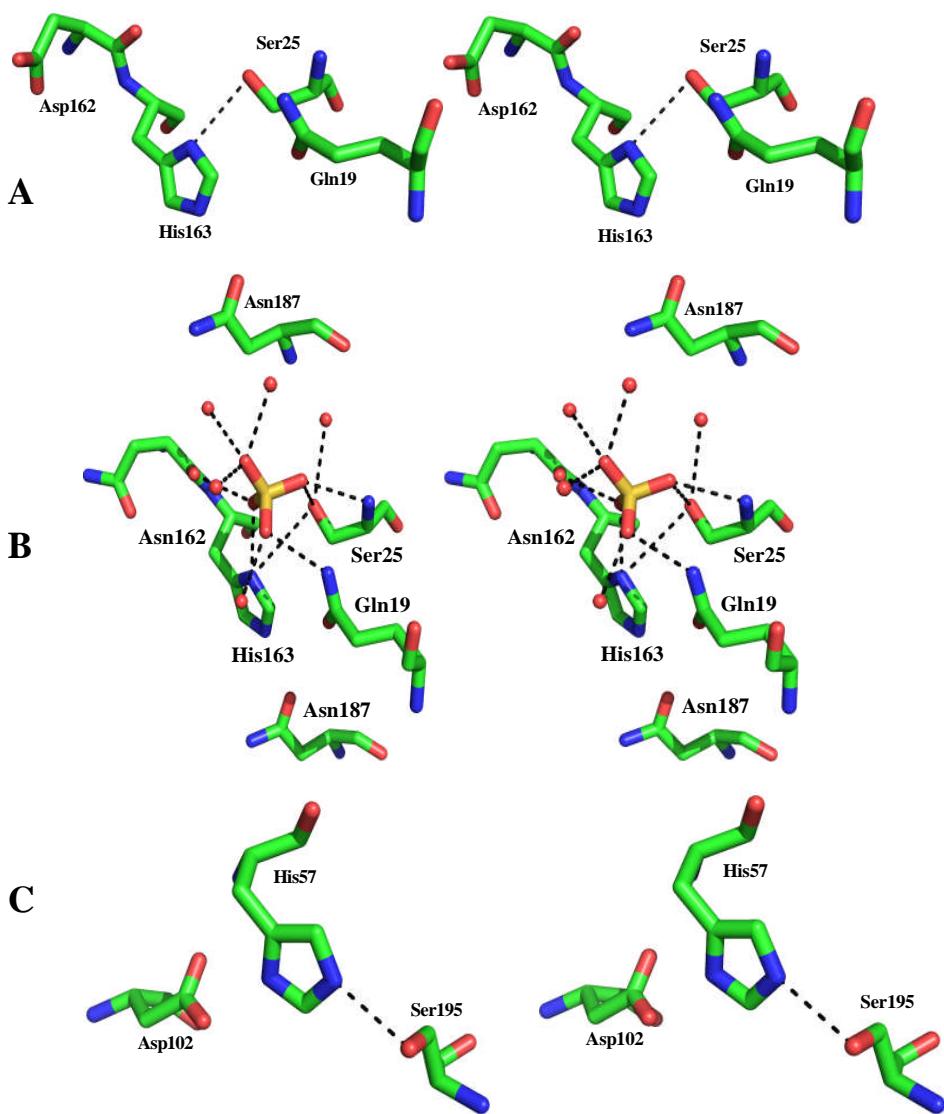


Figure 34. (A) Stereo view of the active site of the C25S procathepsin L mutant (Pdb accession code 1cjl). Active site and adjacent residues are shown as sticks. The distance of about 3.3 Å between the OG atom of Ser25 and ND1 of His163 is shown as dashed lines. (B) Stereo view of the active site of monomer C of 4SER (Pdb accession code 2vhs) showing the close contacts (<3.5 Å) as dashed lines. Active site and adjacent residues are shown as sticks, and waters as red spheres. The distance between the OG atom of Ser25 and ND1 of His163 is also shown as dotted lines (after Fairhead *et al.*, 2008). (C) As comparison the catalytic triad (His57, Asp102, Ser195) (shown as sticks) (reviewed in Polgár, 2005) of the serine protease chymotrypsin (Pdb accession code 4cha) is also shown. OG of Ser195 and NE2 of His57 are within H-bonding distance (Polgár, 2005) (2.8 Å) (dashed lines). This figure has been produced using PyMol (DeLano, 2007).

At a concentration above 100 ppm, spontaneous condensation of silicic acid takes place, because the concentration of nucleophilic ionised molecules is sufficient. There are two ways that may explain how enzyme catalysis can take place. The first one is that one molecule of deprotonated $\text{Si}(\text{OH})_4$ is stabilised at the active site (the nucleophile), which will then undergo a reaction with a second molecule of silicic acid. The alternative is the stabilisation of a protonated $\text{Si}(\text{OH})_4$ molecule (the electrophile), which will then undergo a reaction with a second molecule of silicic acid as well. The SO_4^{2-} at the active site suggests that the former mechanism may operate. The crystal structure shows that it would be difficult to place the $\text{Si}(\text{OEt})_4$ at the active site without causing any conformational change of this region (Fairhead *et al.*, 2008). Modelling of tetraethoxysilane at the same position as the sulfate ion in the 4SER structure (Figure 39) shows that one of the $\text{CH}_3\text{-CH}_2$ groups would clash either with Ser25 (distance 1.9 Å), His163 (distance 1.3 Å) or Asn162 (distance 2.2 Å) depending on the performed rotation (Figure 35). This could be an explanation for the absence of activity of the chimera against tetraethoxysilane indicating a larger catalytic region in silicatein α than in 4SER. This larger active site could allow to fit the bulky $\text{Si}(\text{OEt})_4$ into it (Fairhead *et al.*, 2008).

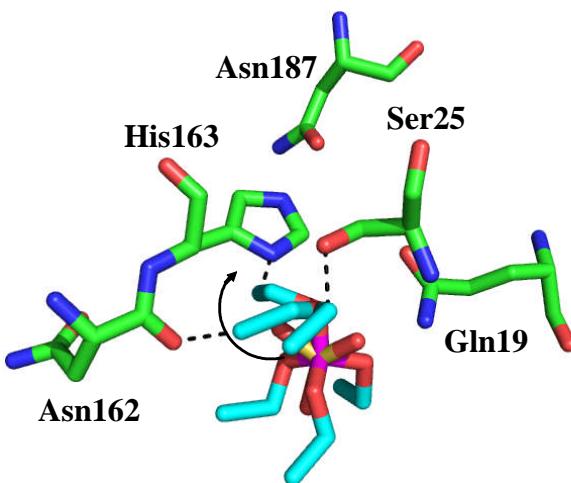


Figure 35. Close up view of the active site of 4SER (residues in sticks) with tetraethoxysilane (carbons in cyan, silicon atom in magenta) modelled at the same position as the sulfate ion (sulfur in yellow). Contacts between CAE of TEOS and OG of Ser25, CAA of TEOS and ND1 of His163, and also CAA and O of Asn162 are shown as dashed lines. The direction of rotation around the silicon and oxygen OAI is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007).

The tetrahedral SO_4^{2-} at the active site represents a good mimic of the tetrahedral $\text{Si}(\text{OH})_4$. The sulfate has an H-bond with Ser25. Consequently one could expect that Ser25 performs a nucleophilic attack on silicic acid to form a covalent enzyme intermediate (see Figure 32). With silicic acid, however, a high energy intermediate would not appear to be chemically likely and as a consequence a hydrolysis back to silicic acid would take place. There is also an H-bond between SO_4^{2-} and His163 (Fairhead *et al.*, 2008), the amino acid which increases the nucleophilicity of the nucleophile in proteases (Sajid and McKerrow, 2002; Fairhead *et al.*, 2008). The structure suggests a procedure in which His163 interacts with the deprotonated form of silicic acid and then stabilises it at the active site. The extensive network of H_2O molecules and H-bonds (Figure 34) would lead to significant shuttling of protons. As a consequence, the negative charge can move to another O-atom. This deprotonated protein bound can then perform a nucleophilic attack on $\text{Si}(\text{OH})_4$ solution species starting the condensation process (Figure 36A). The abstraction of a silicic acid OH-group caused by H^+ -abstraction and elimination of H_2O could also promote the reaction. However, there is no evidence for an acid. It is not likely either that at neutral pH silicic acid represents a sufficient base to first abstract H^+ from His163 (whose pK_a is probably less than 7 because of Asn187) and then for the OH-group of the second silicic acid molecule to carry out a nucleophilic attack (Figure 36B). We propose that Ser25 and the adjacent amino acid substitutions simply form a pocket of sufficient size. Then recognition of the silicic acid molecule would be possible in such a way that H^+ -abstraction by His163 can occur. The deprotonated silicic acid enzyme complex can be seen as a starting point for the polymerisation reaction. A high energy covalent intermediate would not have to be involved (Fairhead *et al.*, 2008). A specific silicic acid transporter in sponges (Schröder *et al.*, 2004; Fairhead *et al.*, 2008) indicates that the real substrate *in vivo* is $\text{Si}(\text{OH})_4$ and not high energy silicon alkoxides. That is why our proposed mechanism appears to be a reasonable model for the biological process (Fairhead *et al.*, 2008).

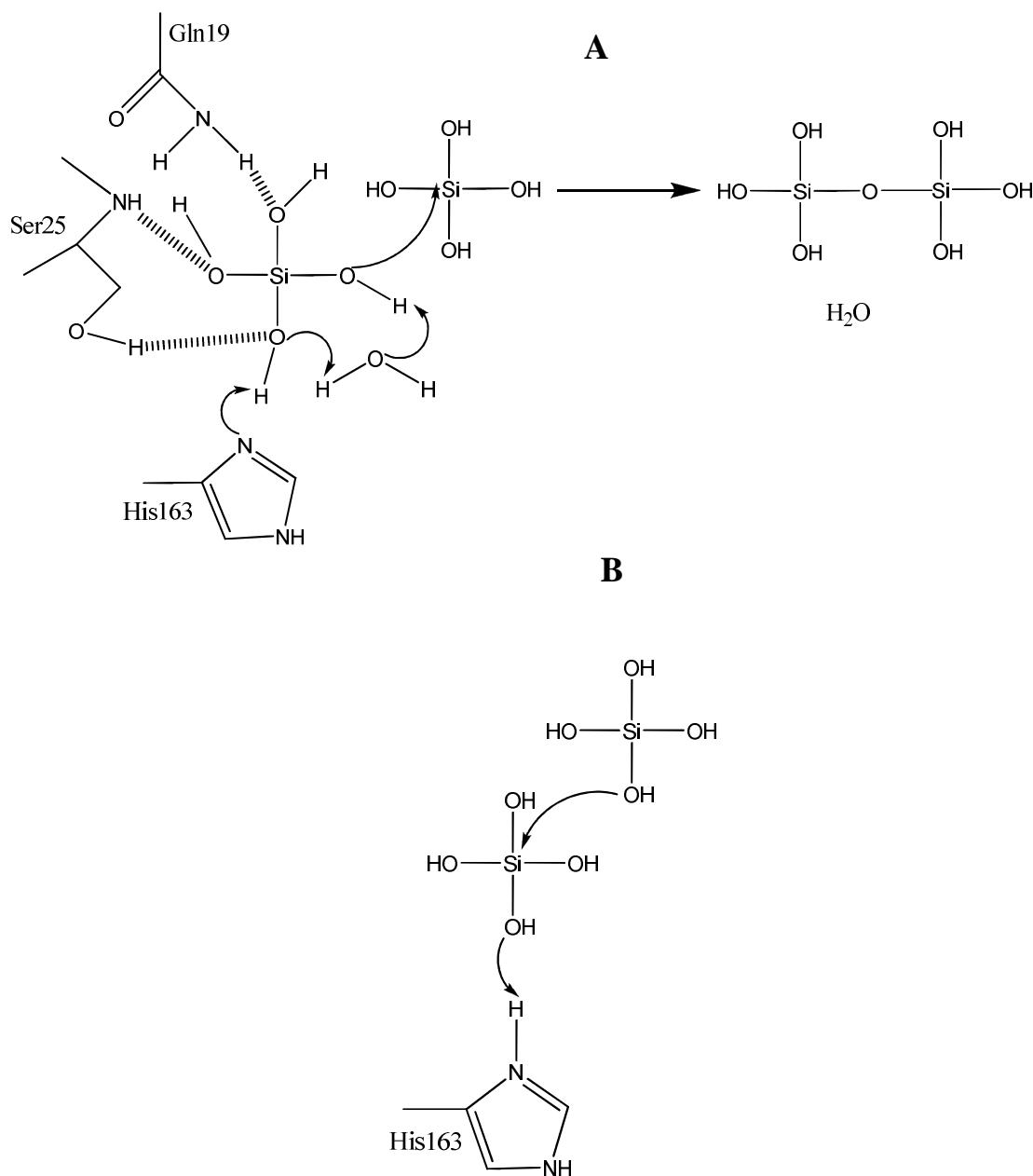


Figure 36. (A) (after Fairhead *et al.*, 2008) The proposed chemical mechanism for the condensation of $\text{Si}(\text{OH})_4$. We favour a mechanism in which His163 abstracts H^+ from silicic acid; the extensive water network could permit the shuttling of protons. (B) The other more unlikely mechanism of deprotonation of His163 and subsequent nucleophilic attack of the OH-group of the second silicic acid molecule on the first $\text{Si}(\text{OH})_4$ (Fairhead *et al.*, 2008).

Future aims

The structure of the 4SER chimera has been solved to 1.5 Å. The tetrahedral sulfate ion at the active site helped us to propose a mechanism for the polymerisation of the also tetrahedral silicic acid (Fairhead *et al.*, 2008). A structure in complex with silicic acid, the real substrate would definitely give even more insight into the mechanism of silica polymerisation. Further objectives could also include finding a method to express a higher amount of soluble silicatein α in order to produce enough for crystallisation. The structure of silicatein α in complex with silicic acid and the bulky Si(OEt)₄ would also contribute greatly to the elucidation of the mechanism of silica condensation.

CHAPTER 2

THE STRUCTURE OF AGME Y140F IN COMPLEX WITH ADP- β -MANNOSE

Introduction

2.1 Epimerases and Racemases

Inversion of stereochemistry in molecules such as amino acids or sugars is carried out by specific enzymes called epimerases and racemases. For the catalysis of this reaction the bond has to be cleaved and reformed stereospecifically. A C-H connection is the usual target for this reaction (Figure 37) (Tanner, 2002).

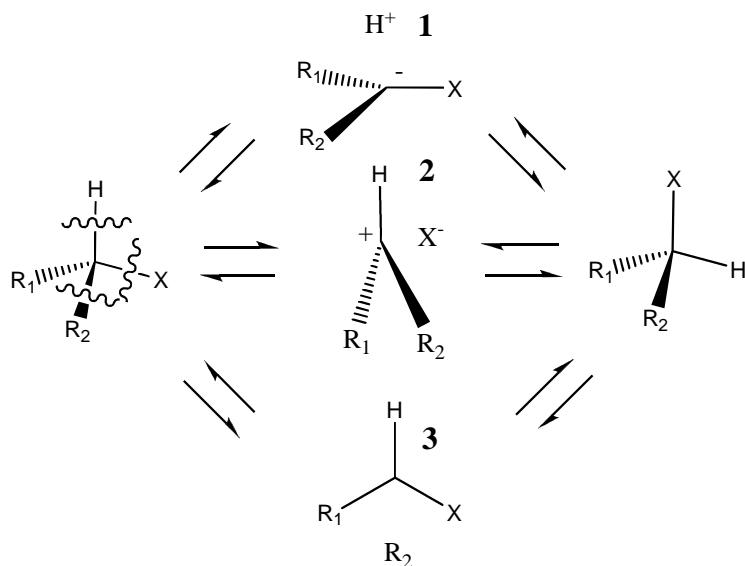


Figure 37. (after Tanner, 2002) There are three distinct routes to stereochemical inversion. The top route (1) shows the cut of a C-H bond. This could be via deprotonation using amino acid racemases (Tanner, 2002), RmlC (Dong *et al.*, 2003) or H^- transfer using UDP galactose 4-epimerase (Frey, 1996; Tanner, 2002), ADP-L-glycero-D-mannoheptose 6-epimerase (Read *et al.*, 2005). The central route (2) describes cleavage of a C-O bond. An example is the epimerisation of 3-hydroxyacyl-CoA. This employs non-stereospecific dehydration/hydration to invert stereochemistry (crotonase reactions) (Yang and Elzinga, 1993; Yang *et al.*, 1991; Smeland *et al.*, 1991; Wakil, 1956). The bottom route (3) involves C-C bond cleavage. A good example is L-ribulose 5-phosphate 4-epimerase (Tanner, 2002).

Epimerisation of carbohydrates is defined as an “inversion of stereochemistry at a stereogenic centre in a sugar” (Figure 38) (Samuel and Tanner, 2002). This can be seen as a reaction where an H-atom is abstracted from one face and returns to the opposite face of the C-atom in the centre.

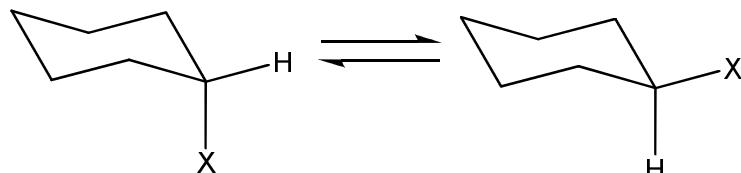


Figure 38. (after Allard *et al.*, 2001) The epimerisation reaction: “the inversion of configuration of an asymmetrically substituted carbon in a sugar” (Allard *et al.*, 2001).

Each stereogenic centre in a sugar can be targeted by an epimerase. A few of these enzymes are even able to carry out epimerisation at two stereocentres (Allard *et al.*, 2001). Any epimerases and racemases operate at stereogenic centres that are activated where abstraction and readdition of H^+ take place (Tanner and Kenyon, 1998; Tanner, 2002). Unactivated stereocentres are targeted by a smaller number of epimerases and racemases. In most of these reactions transient oxidation takes place whereby a coenzyme such as nicotinamide adenine dinucleotide is involved (Tanner, 1998; Tanner, 2001; Tanner, 2002). Activation of a stereogenic centre can be accomplished by oxidation of an adjacent OH-group. As a consequence the acidity of the $\alpha\text{-H}^+$ increases (Figure 39). Inversion of stereochemistry occurs by a mechanism of H^+ -abstraction and readdition and at the end the epimer is formed by reduction of the keto-intermediate (Tanner, 2002).

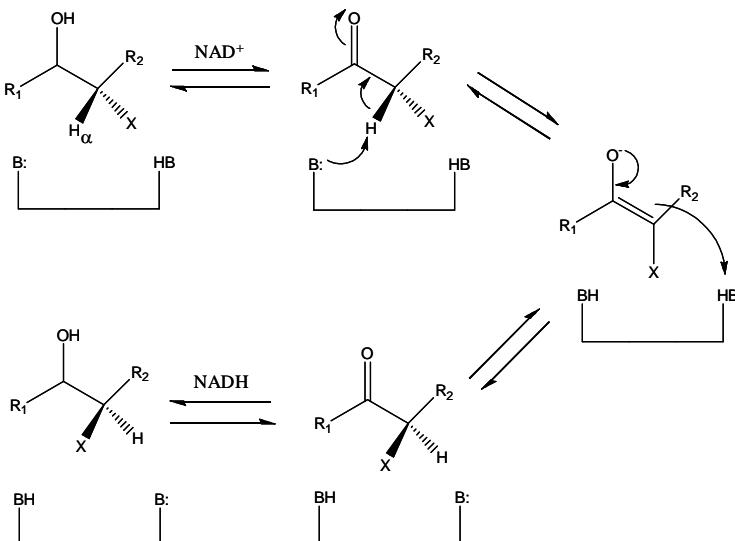


Figure 39. H⁺-abstraction and readdition involving transient oxidation (after Tanner, 2002).

There is a small number of epimerases which act at unactivated stereogenic centres and do not require a coenzyme (Tanner, 2002). Examples for such enzymes are L-ribulose 5-phosphate 4-epimerase and UDP-*N*-Acetylglucosamine 2-epimerase (Tanner, 2002; Allard *et al.*, 2001). L-ribulose 5-phosphate and D-xylulose 5-phosphate are interconverted in the presence of a zinc ion by L-ribulose 5-phosphate 4-epimerase, which occurs in bacteria (Tanner and Kenyon, 1998; Tanner, 2002). In this reaction inversion of stereochemistry takes place by cleavage of a C-C connection (reviewed in Allard *et al.*, 2001; Tanner *et al.*, 2002). UDP-*N*-acetylglucosamine and UDP-*N*- acetylmannosamine are reversibly interconverted by the bacterial UDP-*N*-Acetylglucosamine 2-epimerase (Kawamura *et al.*, 1979; Tanner, 2002). In this case epimerisation takes place by removing and readding uridine diphosphate.

2.2 Short-chain dehydrogenases/reductases (SDRs)

The family of short-chain dehydrogenases/reductases (SDRs) was described in the early 1980s. The first known enzymes of this protein family were an alcohol dehydrogenase from *Drosophila* and a bacterial ribitol dehydrogenase (Persson *et al.*, 2003; Jörnvall *et al.*, 1981). There are more than 3000 amino acid sequences of short-chain dehydrogenases/reductase in sequence databases, and around 70 members of this enzyme family in humans (Oppermann *et al.*, 2003; Kallberg *et al.*, 2002 and 2002a). Their definition and identification derives from some specific regions in their sequence. They represent oxidoreductases that differ from other families with similar functions such as medium-chain dehydrogenases/oxidoreductases (MDRs) (Oppermann *et al.*, 2003; Jörnvall *et al.*, 1995; Oppermann *et al.*, 1997; Filling *et al.*, 2002) or aldo-keto reductases (AKR) (Oppermann *et al.*, 2003; Jez *et al.*, 1997). SDR-enzymes operate on a variety of substrates such as xenobiotics, steroids, prostaglandins and alcohols (Persson *et al.*, 2003; Oppermann *et al.*, 2003). Pair-wise alignments of various short-chain dehydrogenases/reductases usually show 15-30 % identity (Oppermann *et al.*, 2003; Persson *et al.*, 2003; Kallberg *et al.*, 2002 and 2002a; Jörnvall *et al.*, 1995). However, proteins of this family have a highly conserved α/β sandwich fold, where a β -sheet in the centre is surrounded by several α -helices (Rossmann fold) (Figure 40) (Oppermann *et al.*, 2003; Jörnvall *et al.*, 1995; Rossmann *et al.*, 1975).

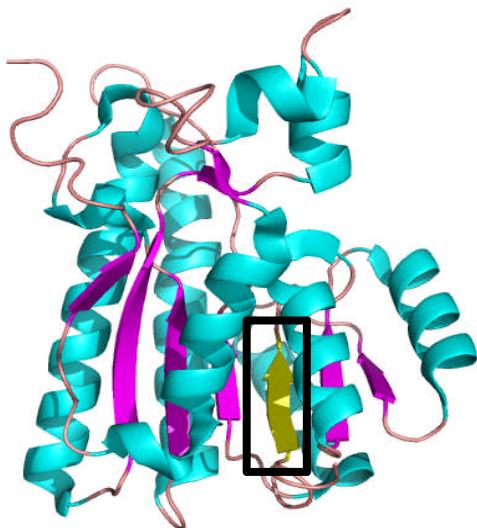


Figure 40. Ribbon diagram of $3\beta/17\beta$ hydroxysteroid dehydrogenase ($3\beta/17\beta$ HSD) from *Comamonas testosteronei* (Pdb accession code 1hxh). α -helices are shown in cyan, β -strands in magenta, and loops in pink. The α/β sandwich and the β -sheet in the centre, decorated by several α -helices represent the characteristic Rossmann fold (Oppermann *et al.*, 2003; Jörnvall *et al.*, 1995; Rossmann *et al.*, 1975). The β -strand in the centre is shown in yellow and is boxed.

The majority of enzymes from this family are 250-350 amino acids long (Oppermann *et al.*, 2003; Kallberg *et al.*, 2002; Persson *et al.*, 2003; Filling *et al.*, 2002). The aminoterminal part of short chain dehydrogenases/reductases is involved in the interaction with a cofactor, nicotinamide adenine dinucleotide or nicotinamide adenine dinucleotide phosphate. The carboxyterminal region interacts with substrate (Kallberg *et al.*, 2002; Oppermann *et al.*, 2003; Persson *et al.*, 2003; Filling *et al.*, 2002). Highly conserved motifs in SDRs are a glycine rich cluster at the aminotermminus, and the catalytic triad Ser, Tyr and Lys in the carboxyterminal part (Oppermann *et al.*, 2003; Filling *et al.*, 2002; Jörnvall *et al.*, 1995). A more detailed discussion on the functions and positions of those two and other conserved motifs in SDR enzymes is given in a recent review (Oppermann *et al.*, 2003).

2.2.1 Reaction mechanism of SDR enzymes: $3\beta/17\beta$ hydroxysteroid dehydrogenase from *Comamonas testosteroni* as an example

Filling and coworkers discovered that the catalytic triad Ser, Tyr and Lys of SDR enzymes (Oppermann *et al.*, 2003; Filling *et al.*, 2002; Jörnvall *et al.*, 1995) can be extended to a tetrad including a conserved Asn residue (Filling *et al.*, 2002; Oppermann *et al.*, 2003). The roles of the active site residues Ser138, Tyr151, and Lys155 in $3\beta/17\beta$ hydroxysteroid dehydrogenase had already been known before their studies (Filling *et al.*, 2002). Tyr151 acts as a base (Filling *et al.*, 2002; Jörnvall *et al.*, 1995), Ser138 plays an important role in the stabilisation of the substrate and Lys155 generates H-bonds with the ribose adjacent to the cofactor's niacinamide whereby the pK_a value of the side chain hydroxyl of Tyr151 is lowered. As a consequence H^+ -transfer can occur. In the crystal structure of bacterial $3\beta/17\beta$ hydroxysteroid dehydrogenase, it has been discovered that Asn111 interacts with Lys155 via H_2O . Because of this observation it has been proposed that this asparagine residue plays an essential role in the stabilisation of Lys155 and consequently an H^+ relay (Figure 41) occurs in the majority or even all short chain dehydrogenases/reductases with Tyr151, the 2' hydroxyl of ribose, cofactor, substrate, H_2O , Asn111 (or Ser in other SDRs) and Lys155 involved (Filling *et al.*, 2002). However, there is a possibility that there are differences in the active sites and catalysis. Crucial residues could be at other positions, substrates could be converted sequentially, or other amino acids such as His, Arg or Thr could play an essential role in catalysis (Oppermann *et al.*, 2003; Kallberg *et al.*, 2002; van Hylckama Vlieg *et al.*, 2001; Gourley *et al.*, 2001).

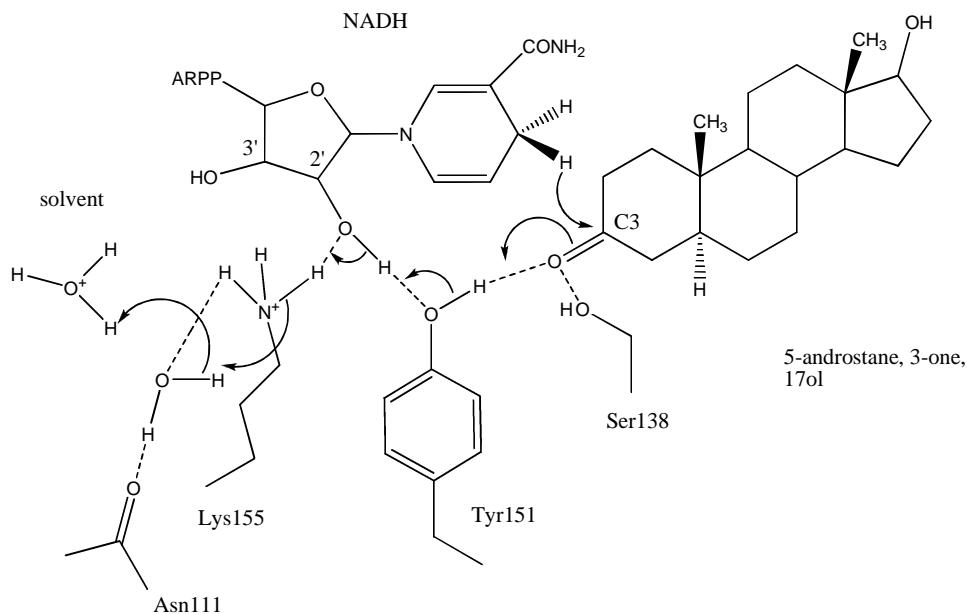


Figure 41. Reaction mechanism of $3\beta/17\beta$ hydroxysteroid dehydrogenase with NADH and steroid substrate (5α -androstane, 3-one, 17ol). At the beginning of the catalytic reaction an H^+ is transferred from the side chain hydroxyl of Tyr 151 to C=O of the substrate. Then H^- is transferred to C3 of 5α -androstane, 3-one, 17ol. Finally an H^+ relay is built up including Lys155, Asn111, H_2O and the 2'-hydroxyl of the ribose. ARPP: adenosine ribose pyrophosphate part of the cofactor (after Filling *et al.*, 2002).

2.2.2 Families of SDR enzymes

A full classification of SDR families and subfamilies was reported in two publications (Kallberg *et al.*, 2002; Persson *et al.*, 2003). The second paper largely supercedes the first one.

Short chain dehydrogenases/reductases have been classified into two families named “classical” and “extended” (Kallberg *et al.*, 2002; Jörnvall *et al.*, 1995).

Whereas members of the classical family contain about 250 amino acids, representatives of the extended family have about 350 amino acids (Kallberg *et al.*, 2002; Jörnvall *et al.*, 1995). Oxidoreductases and carbonyl reductases, for instance, belong to the classical family. Extended SDRs are for example epimerases and dehydratases (Figure 42). There are small distinctions in the conserved glycine-rich motif at the aminotermminus between these two families. Several amino acids are exactly at identical positions in the classical and extended SDRs whereas other sequence motifs show differences (Table 10) (Persson *et al.*, 2003). A more detailed analysis of conserved regions in short chain dehydrogenases/reductases revealed that apart from those already established families, SDRs can be further separated into three other families named “intermediate”, “divergent” and “complex” (Figure 43) (Kallberg *et al.*, 2002). Representatives of the intermediate family are alcohol dehydrogenases from *Drosophila* (Figure 42). Bacterial and herbal enoyl reductases (Figure 42) belong to the divergent family. A member of the complex family is for instance fatty acid synthase (Persson *et al.*, 2003).

A sequence alignment of representatives from all five families (Kallberg *et al.*, 2002) is shown in Figure 42. Conserved motifs (Table 10) (after Persson *et al.*, 2003) in all eight aligned enzymes are highlighted.

Table 10. (after Persson *et al.*, 2003) Conserved sequence motifs in classical and extended SDRs

Secondary structure element	Classical SDR motifs	Extended SDR motifs	Function (see references in Kallberg <i>et al.</i> , 2002 and Persson <i>et al.</i> , 2003)
$\beta_1 + \alpha_1$	TGxxxGhG	TGxxGahG	Structural role in cofactor binding site
	Dhx[cp]	DhxD	Adenine ring interaction of cofactor
	GxhDhhhNNAGh	[DE]xhhHxAA	Structural role in the stabilisation of the central β -sheet
	hNhxG	hNhhGTxxhhc	Part of catalytic region
	GxhhxhSSh	hhhxSSxxhaG	Part of catalytic region
	Yx[AS][ST]K	PYxx[AS]Kxxh[DE]	Part of catalytic region
	h[KR]h[NS]xhxPGxxxT	h[KR]xxNGP	Structural function, reaction direction

In the motifs, “a” stands for an aromatic amino acid, “c” for a charged amino acid, “h” for a hydrophobic amino acid, “p” for a polar amino acid and “x” for any amino acid. Other residues at a position in a motif are shown in brackets.

$\beta 1 + \alpha 1$

2HSD	-----NDLSGKTVII TGGARGLG AEAARQAVAAGAR--VVLADVL----
1HXH	-----TNRLQQGKVALV TGGASGVG LEVVKLLLGEAK--VAFSDIN----
1EK6	-----MAEKVLV TGGAGYIG SHTVLELLEAGYLPVIDNFHNAFRG
1A4U	-----MDLTNKNVIFV AALGGIG LDTRELVKRNLKNFVILDRAVE----
1MG5	-----SFTLTNKNVIFV AGLGGIG LDTSKELLKRDLKLNVLILDRIE----
1QSG	--GSHMGFLSGKRILV TGVASKLSIA YGIAQAMHREGAELAFTYQN----
1BVR	-----TGLLDGKRILVS GIITDSSIA FHIARVAQEQQGAQLVLTGFDR----
P22367 MSAS_PENPA	SEGPRLLPRPEGTYLI TGGLGVLC LEVADFVKEKGARRLLISRRALPPR

$\beta 3 + \alpha 3$

2HSD	-----DEEGAATARELGDA--RYQHLD DVTI EEDWQ-RVVAYARE
1HXH	-----EAAGQQLAAELGERS--MFVRHD DVSS EADWT-LVMAAVQR
1EK6	-----GGSLPESLRRVQELTGRSVEFEEM DILD QGALQRLFKKYSFM
1A4U	-----NPTALAEKAINPKVNITFHTY DVTW PVAESKKLLKKIFD
1MG5	-----NPAAIAELKAINPKVTVTFYPY DVTW PIAETTKLLKTIFA
1QSG	-----DKLKGRVEFAAQLGSDIVLQC DVAED AS-IDTMFAELGK
1BVR	-----LRLIQRTDRLPAKPLLELD DVQN EEHLASLAGRTEAIG
P22367 MSAS_PENPA	-QVSEDLQPTIAKIRLLESRGASVHVLPL DITK PDADVEQLTTALDRL

$\beta 4$ **$\alpha 4$**

2HSD	EFG SVVDGLVNNAGI ST-----GMFLETESVERFR---KVVE INLT
1HXH	RL GTLNVLVNNAGILL -----PGDMETGRLEDFS---RLLK INTE
1EK6	AVIHAGLKAVGESVQKPLDYRVNLT GTIQMLE IMKAHGKVNL VFSSSA
1A4U	QLK TVTDILINGAGILD -----DHQIERTIAINF---GLV NNTTA
1MG5	QLK TVDVILINGAGILD -----DHQIERTIAVNYT---GLV NNTTA
1QSG	VWPKFDGFVHSIGFAP-----GDQLDGDYVNAVTRGFKIAHDIS
1BVR	AGNKLDGVVHSIGFMP-----QTGMGINPFFDAPYADVSGKIHIS
P22367 MSAS_PENPA	SLP SVQGVVHAAGVLDN -----ELVMQTTTRDAFNRLAPKIAGALALH

$\alpha 4$ **$\beta 5$**

2HSD	GVFIGMKTVIPAMKDAG GGSIVNNISSA AGLMGLALTSS YGASK WGVRGLS
1HXH	SVFIGCQQGIAAMKETG G-SIINMASVSSW LPIEQYAG YSASK AAVSALT
1EK6	TVYGNPQYPLPDEAHTGGCTN PYGKS KFFIEEMIRDLCQADKTWNAVLL
1A4U	ILDFWDKR-----KGGPGGIIANICSVTGFNAIHQVPV YSASK AAVVSFT
1MG5	ILDFWDKR-----KGGPGGIIICNIG SVTGFNAIYQVPV YSGTK AAVVNFT
1QSG	SYSFVAMAKACRSM LNPGS ALLTLSYLAERAIPN YNVMGLAK ASLEANV
1BVR	AYSYASMAKALLP 1MNP PGGSIVGMDFDP-SRAMPAY YNWMTVAK SALESVN
P22367 MSAS_PENPA	EVFPKSVDFVMFSSC CNLVGTGQAS YGSGNA FLDTLATHRARLGAA

* **$\alpha 5$**

Figure 42. (continues on next page)

β6

2HSD	KLAALVELG-----TDR	IRVNSVHPGMTYT	PMTAETGIR--QGEGNYP
1HXH	RAAALSCRK-----QGYA	IRVNSIHPDGIYT	PMMQASLPKGVSKEMLVH
1EK6	RYFNPTGAHASGCCIGEDPQGIPNNLMPYVSQVAIGRREALNVFGNDYDTE		
1A4U	NSLAKLAP-----ITG	VTAYSINPGITRT	PLVHTFNSWLDVEPRVAE
1MG5	SSLAKLAP-----ITG	VTAYTVNPGITRT	TLVHKFNSWLDVEPVQVAE
1QSG	RYMANAMG-----PEG	VRVNAISAGPIRT	LAASGIKDFR---KMLAH
1BVR	RFVAREAG-----KYG	VRSNLVAAGPIRTL	AMSAIVGGALGEEAGAQ
P22367 MSAS_PENPA VSFQWTSWRGLGMGASTDFINAELESKGITDVTRDEAFAAWQHLAKYDMD			

Figure 42. Structural sequence alignment of members from all five families of SDRs (Kallberg *et al.*, 2002) Conserved sequence motifs (see Table 10) in classical SDRs ($3\beta/17\beta$ -hydroxysteroid dehydrogenase from *Comamonas testosteroni*, Pdb accession code 1hxh; $3\alpha,20\beta$ -hydroxysteroid dehydrogenase from *Streptomyces exfoliatus*, Pdb accession code 2hsd) are highlighted in yellow, the corresponding motifs in the representative of the extended SDR family (UDP-galactose 4-epimerase from *Homo sapiens*, Pdb accession code 1ek6) are highlighted in green, in intermediate SDRs (alcohol dehydrogenase from *Drosophila lebanonensis*, Pdb accession code 1a4u; alcohol dehydrogenase from *Drosophila melanogaster*, Pdb accession code 1mg5) in turquoise, in divergent SDRs (enoyl reductase from *Escherichia coli*; Pdb accession code, 1qsg; enoyl-ACP reductase from *Mycobacterium tuberculosis*; Pdb accession code 1bvr) in grey, and in one member of the complex SDR family (6-methylsalicylic acid synthase from *Penicillium patulum* (*Penicillium griseofulvum*), UniProtKB/TrEMBL entry P22367) in pink. Residues conserved in all five families are marked with an asterisk. The positions of secondary structural elements in classical SDRs are marked on top of each motif, positions of α -helices and β -strands, as they occur in the extended family are shown in red at the bottom of each motif. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

The alignment shows that $3\beta/17\beta$ -hydroxysteroid dehydrogenase and $3\alpha,20\beta$ -hydroxysteroid dehydrogenase (Kallberg *et al.*, 2002) contain the conserved aminoterminal glycine-rich Thr-Gly-X-X-Gly-h-Gly (see Table 10) specific for members of the family of classical SDRs (Kallberg *et al.*, 2002). In UDP-galactose 4-epimerase which is a representative of the extended family (Kallberg *et al.*, 2002), the same motif has a slightly different pattern (Thr-Gly-X-X-Gly-h-a-Gly) (Table 10) (Persson *et al.*, 2003). In both *Drosophila* alcohol dehydrogenases which are members of the intermediate family (Kallberg *et al.*, 2002) the unique aminoterminal Gly(Ala)-X-X-Gly-X-X-Gly(Ala) sequence cluster which displays similarity the glycine-rich region of extended SDRs can be found (Kallberg *et al.*, 2002). Additionally an Asn-Gly-Ala-Gly fingerprint sequence in β -strand 4 can be observed which is almost identical with the Asn-Asn-Ala-Gly motif (Table 10) of classical SDRs (Persson *et al.*, 2003).

Enoyl reductase and enoyl-ACP reductase contain the Gly-X-X-X-X-Ser-X-Ala motif at the aminotermminus typical for members of the divergent family. Additionally the conserved cluster at the active site also differs from the members of the other SDR families (Kallberg *et al.*, 2002). The Tyr is replaced by a Met whereas the Tyr is located “three positions upchain” (Kallberg *et al.*, 2002). 6-methylsalicylic acid synthase which belongs to the family of complex short chain dehydrogenases/reductases contains the specific active site cluster Tyr-X-X-X-Asn instead of the Tyr-X-X-X-Lys motif of classical, extended and intermediate SDR enzymes (Persson *et al.*, 2003).

Classical SDRs are most common, extended short chain dehydrogenases/reductases next and the intermediate third. The families with the least number of members are the divergent and complex families (Kallberg *et al.*, 2002).

Apart from this classification into five different SDR families, another classification of classical and extended short chain dehydrogenases based on their cofactor preference for nicotinamide adenine dinucleotide or nicotinamide adenine dinucleotide phosphate was performed (Figure 43 and Table 11) (Kallberg *et al.*, 2002). NAD(H)-interacting enzymes usually contain an aspartic acid or glutamic acid at end of β -strand 2 (Table 11) (Kallberg *et al.*, 2002; Wierenga *et al.*, 1986), whereas NADP(H) specific enzymes normally contain an arginine or lysine (Kallberg *et al.*, 2002; Tanaka *et al.*, 1996). One of these basic amino acids is in the glycine rich cluster, directly before the second glycine. The other basic amino acid can usually be found after the essential aspartic acid or glutamic acid of NAD(H)-binding enzymes (Table 11) (Kallberg *et al.*, 2002).

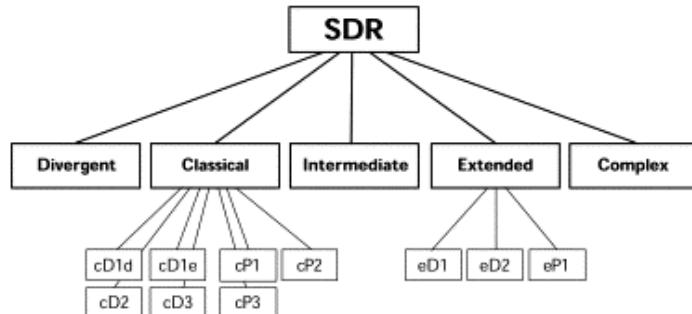


Figure 43. (taken from Persson *et al.*, 2003) The two levels of classification within the superfamily of SDR enzymes.

Enzymes belonging to the superfamily of SDRs are divided into five families. Classical and extended SDRs are further divided into seven and three subfamilies, respectively, based on their amino acids associating to the cofactor (Persson *et al.*, 2003; Kallberg *et al.*, 2002).

Table 11. (after Persson *et al.*, 2003; Kallberg *et al.*, 2002) “Key positions” for assignments of cofactor preference of classical (A) and extended short chain dehydrogenases/reductases (B). Numbers refer to $3\alpha,20\beta$ -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) in A and UDP-galactose 4-epimerase (Pdb accession code 1ek6) in B.

“Key positions“					
Subfamily	Cofactor	15	36	37	38
(A) Classical SDRs					
cD1d	NAD(H)		D		
cD1e	NAD(H)		E		
cD2	NAD(H)			D/E	
cD3	NAD(H)				D/E
cP1	NADP(H)		K/R		
cP2	NADP(H)			K/R	
cP3	NADP(H)	K/R		K/R	
		33	34	35	
(B) Extended SDR enzymes					
cD1	NAD(H)	D/E			
cD2	NAD(H)		D/E		
cP1	NADP(H)		K/R		

A sequence alignment of representatives from several subfamilies of classical and extended SDRs is shown in Figure 44, (based on Kallberg *et al.*, 2002). Residues at “key positions” specific for the respective subfamily (Table 11 and Figure 43) Persson *et al.*, 2003) in all seven aligned enzymes are highlighted.

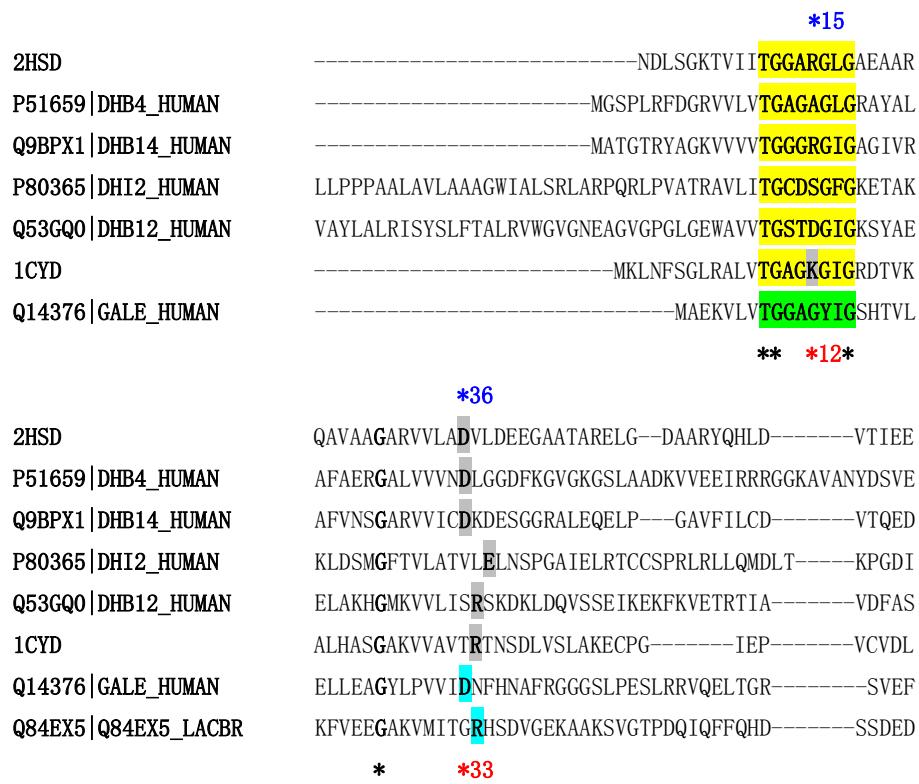


Figure 44. Sequence alignment of various classical and extended SDRs, based on their coenzyme specificity (based on Kallberg *et al.*, 2002). The “key position” numbers refer to $3\alpha,20\beta$ -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) as a member of the classical family (blue asterisks) and UDP-galactose 4-epimerase (UniProtKB/Swiss-Prot Q14376 (GALE_HUMAN)) as a representative of the extended family (red asterisks) (Kallberg *et al.*, 2002). Conserved residues in all eight proteins are marked with asterisks. The conserved TGxxxGhG motif of the classical SDRs (Kallberg *et al.*, 2002) is highlighted in yellow, the counterpart in the extended family TGxxGhaG (Kallberg *et al.*, 2002) in green. “Key position” residues are highlighted in grey in the classical family and in turquoise in the extended one. $3\alpha,20\beta$ -hydroxysteroid dehydrogenase from *Streptomyces exfoliatus* (Pdb accession code 2hsd), subfamily cD1d (Kallberg *et al.*, 2002); 17-beta-hydroxysteroid dehydrogenase 4 (UniProtKB/Swiss-Prot P51659 (DHB4_HUMAN)) from *Homo sapiens*, subfamily cD1d (Persson *et al.*, 2003); 17-beta-hydroxysteroid dehydrogenase 14 (UniProtKB/Swiss-Prot Q9BPX1(DHB14_HUMAN)) from *Homo sapiens*; subfamily cD1d; Corticosteroid 11-beta-dehydrogenase isozyme 2 (UniProtKB/Swiss-Prot P80365 (DHI2_HUMAN)) from *Homo sapiens*; subfamily cD3 (Persson *et al.*, 2003); Estradiol 17-beta-dehydrogenase 12 (UniProtKB/Swiss-Prot Q53GQ0(DHB12_HUMAN)) from *Homo sapiens*, subfamily cP2; Carbonyl reductase (Pdb accession code 1cyd) from *mus musculus*, subfamily cP3 (Kallberg *et al.*, 2002); UDP-galactose 4-epimerase (UniProtKB/Swiss-Prot Q14376 (GALE_HUMAN)) from *Homo sapiens*, subfamily eD1 (Kallberg *et al.*, 2002); R-specific alcohol dehydrogenase (UniProtKB/Swiss-Prot Q84EX5 (Q84EX5_LACBR) from *Lactobacillus brevis*, subfamily eP1 or cP2. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

The classical short chain dehydrogenases/reductases 3 α ,20 β -hydroxysteroid dehydrogenase and 17-beta-hydroxysteroid dehydrogenase 4 belong both to the cD1d subfamily (Kallberg *et al.*, 2002) as they are NAD(H) specific and contain an acidic amino acid, aspartic acid, at key position 36 (Persson *et al.*, 2003) (Table 11). Another member of this subfamily is 17-beta-hydroxysteroid dehydrogenase 14. Corticosteroid 11-beta-dehydrogenase isozyme 2 can be identified as a member of the cD3 subfamily of classical SDRs (Persson *et al.*, 2003) because it is NAD(H) dependent and contains an acid residue, glutamic acid, at key position 38 (Table 11) (Kallberg *et al.*, 2002). Estradiol 17-beta-dehydrogenase 12 can be classified into the cP2 subfamily of classical short chain dehydrogenases/reductases as it can be NADP(H) specific (see [http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-e+\[UNIPROT:DHB12_HUMAN\]+newId](http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-e+[UNIPROT:DHB12_HUMAN]+newId)) and contains a basic residue, arginine, at key position 37 (Table 11) (Persson *et al.*, 2003). Carbonyl reductase is an example for a representative of the cP3 subfamily of classical SDRs (Kallberg *et al.*, 2002). It prefers NADP(H) and there are basic amino acids at both key positions, a lysine at position 15 and an arginine at position 37 (Table 11) (Kallberg *et al.*, 2002). UDP-galactose 4-epimerase which is an NAD(H) dependent (Thoden *et al.*, 1996; Kallberg *et al.*, 2002) member of the extended family (Kallberg *et al.*, 2002) belongs to the eD1 subfamily as it contains an acidic amino acid, aspartic acid, at key position 33 (Table 11) (Kallberg *et al.*, 2002). R-specific alcohol dehydrogenase from *Lactobacillus brevis* could be classified into either the eP1 subfamily of extended short chain dehydrogenases/reductases or the eP2 subfamily of classical SDR enzymes. It contains the basic amino acid Arg (key position 37 in 3 α ,20 β -hydroxysteroid dehydrogenase and 34 in UDP-galactose 4-epimerase) (Table 11) and is NADP(H)-dependent (Schlieben *et al.*, 2005). Both features are specific for those subfamilies. It has been referred to as the extended SDR family (Schlieben *et al.*, 2005).

However, the glycine rich motif Thr-Gly-Gly-Thr-Leu-Gly-Ile-Gly resembles the conserved Thr-Gly-X-X-X-Gly-h-Gly motif, which occurs in the classical SDR family (Persson *et al.*, 2003), and also the subunit length of about 250 amino acids is a common feature of classical short chain dehydrogenases/reductases (Jörnvall *et al.*, 1995; Persson *et al.*, 2003). That is why this enzyme belongs to the cP2 subfamily of classical SDRs rather than to the extended family.

Members of the classical family have a higher preference for NADP(H) than NAD(H) correlated with higher preference for reduction. Most extended SDRs, unlike the classical ones, prefer NAD(H) as their coenzyme suggesting that the extended SDR-family probably comprises more dehydrogenases than reductases (Kallberg *et al.*, 2002).

2.2.3 ADP-L-glycero-D-mannoheptose 6-epimerase (AGME)

2.2.3.1 Biosynthetic pathway of ADP-L-glycero-D-mannoheptose

Two membranes occur in Gram-negative bacteria which divide the cytoplasm from the milieu outside of the cell. One of them is the cytoplasmic membrane and the other one the outer membrane (Deacon *et al.*, 2000; Osborn *et al.*, 1972). The cytoplasmic membrane is composed of two layers of phospholipids. The outer membrane contains two parts. In the outer part one can find a specific lipid called lipopolysaccharide. The bacterial lipopolysaccharide itself is composed of lipid A, the core and the O-antigen determinants.

The core domain is structured in two parts. The outer part contains six carbon sugars and *N*-acetylglucosamine, whereas the inner one contains the 7 C sugar L-glycero-D-mannoheptose (or heptose) and 2-keto-3-deoxyoctulosonic acid (Kneidinger *et al.*, 2002; Deacon *et al.*, 2000; Reeves *et al.*, 1996; Raetz and Whitfield, 2002). L-glycero-D-mannoheptose as part of the core domain of lipopolysaccharide shows a high degree of conservation among a great number of Enterobacteriaceae and other Gram-negative bacteria (Deacon *et al.*, 2000; Adams *et al.*, 1967). Lipopolysaccharide is involved in keeping the integrity of the structure of the outer membrane whereby it associates with other components (Kneidinger *et al.*, 2002; Deacon *et al.*, 2000). It also protects the organism against destructive intruders and a few bacteria infecting viruses (Deacon *et al.*, 2000; Coleman and Leive, 1979). In mutants which do not have heptose, nonpolar compounds can penetrate more easily through the outer membrane. These mutants also show less pathogenicity (Deacon *et al.*, 2000; Taylor, 1983). As a consequence the lipopolysaccharide biosynthetic pathway represents an interesting target for antimicrobial compounds (Deacon *et al.*, 2000).

ADP-D-glycero-D-mannoheptose and ADP-L-glycero-D-mannoheptose are interconverted in the final reaction in the biosynthetic route of the precursor of heptose (Figure 45) carried out by ADP- β -L-glycero-D-mannoheptose 6-epimerase (AGME) (Deacon *et al.*, 2000; Eidel and Osborn, 1974; Raetz, 1990; Coleman, 1983; Pegues *et al.*, 1990).

In the first reaction of this pathway an isomerisation of sedoheptulose 7-phosphate to D,D-Hep 7-phosphate catalysed by GmhA occurs. This sugar is then phosphorylated to give β -D,D-Hep 1,7-bisphosphate. This step is catalysed by HldE, an enzyme which is involved in two different reactions.

It displays kinase activity and also acts as an aldehylyltransferase in a later reaction in this pathway. Then cleavage of the phosphate group from the carbon at position 7 of β -D,D-Hep 1,7-bisphosphate takes place, carried out by the phosphatase GmhB, to form β -D,D-Hep 1-phosphate. ADP-D,D-Hep is then generated from β -D,D-Hep 1-phosphate by the action of again, HldE. The formation of ADP-L,D-Hep due to a reversible epimerisation reaction carried out by ADP- β -L-glycero-D-mannoheptose 6-epimerase represents the last step in the biosynthesis of the heptose precursor. L,D-Hep is then added to the growing lipopolysaccharide which is mediated by heptosyltransferases (Morrison and Tanner, 2007; Kneidinger *et al.*, 2002; Valvano *et al.*, 2002, Morrison *et al.*, 2005).

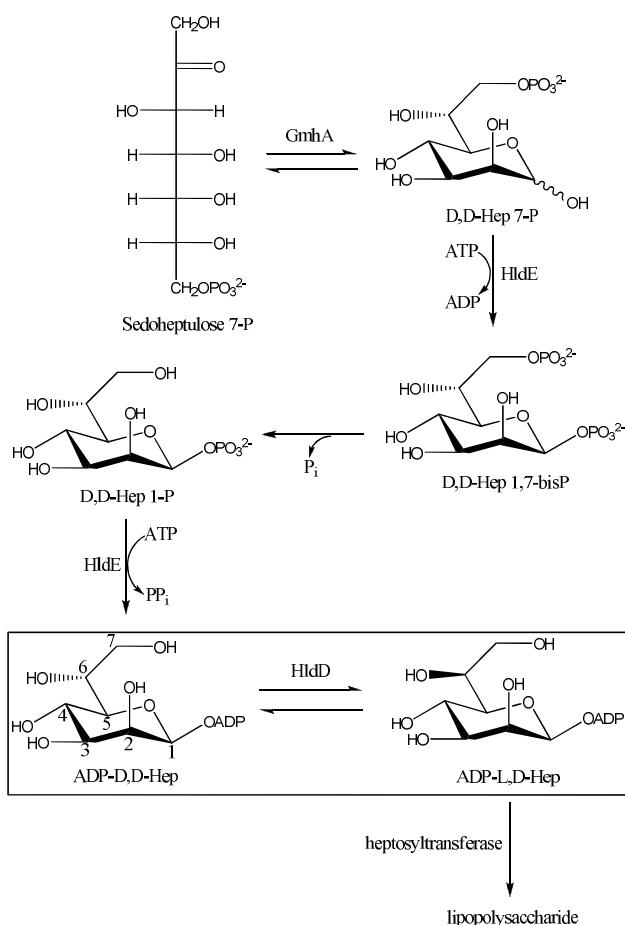


Figure 45. (after Morrison and Tanner, 2007) Biosynthetic pathway of ADP-L-glycero-D-mannoheptose. The reaction carried out by ADP- β -L-glycero-D-mannoheptose 6-epimerase is surrounded with a box.

2.2.3.2 AGME, a member of the SDR superfamily

In the early 1980s, activity of ADP- β -L-glycero-D-mannoheptose 6-epimerase was detected for the first time in extracts of the *Escherichia coli* strain K-12 (Deacon *et al.*, 2000; Coleman, 1983). About ten years later the *RfaD* gene which encodes AGME was cloned (Deacon *et al.*, 2000). Partial purification of the enzyme was then performed and it was demonstrated that its aminotermminus comprises the highly conserved motif Gly-Gly-X-Gly-X₂-Gly, involved in the interaction with ADP of the cofactor (Deacon *et al.*, 2000; Pegues *et al.*, 1990; Wierenga *et al.*, 1986). In 2001, it was demonstrated that AGME displays specificity to nicotinamide adenine dinucleotide phosphate. It is also able to be active in the presence of NAD instead of NADP, however, activity is then significantly lower. The apo-form of AGME is inactive (Deacon *et al.*, 2000; Ni *et al.*, 2001).

Due to the fact that AGME is present in various Gram-negative bacteria and not in mammals, it represents a target for the design of novel antimicrobial compounds. AGME is a pentameric enzyme which contains five identical subunits. Each of them is composed of two domains (Deacon *et al.*, 2000). The aminoterminal domain is composed of a modified Rossmann fold (Rossmann *et al.*, 1974) with seven parallel β -strands in the centre. These are decorated on either side by seven α -helices. This domain is involved in the tight association with the coenzyme. The smaller carboxyterminal domain contains three α -helices and two small parallel β -strands (Figure 46). This part of the structure is involved in the interaction with the substrate, which associates with residues of the cleft formed by both domains (Deacon *et al.*, 2000).

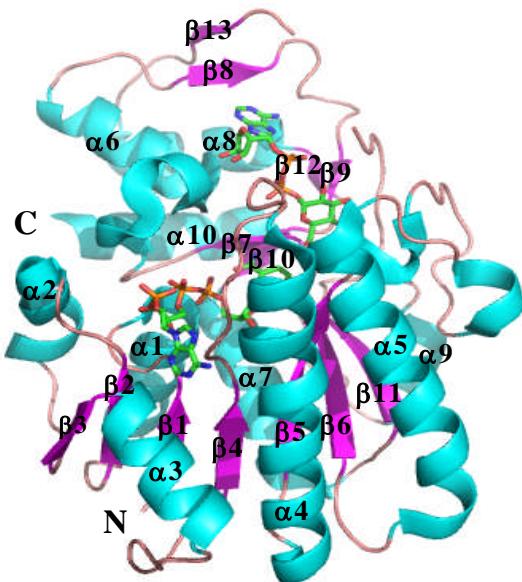


Figure 46. Ribbon diagram of AGME from the *E. coli* strain K12 (Pdb accession code 1eq2) (based on Deacon *et al.*, 2000) α -helices are shown in cyan, β -strands in magenta, and loops in pink. The seven parallel β -sheets in the centre, decorated by seven α -helices on either side represent the characteristic Rossmann fold. The carboxyterminal domain is composed of a small α/β domain, which plays a role in the interaction with the substrate. NADP and the inhibitor ADP-glucose are shown as sticks (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

AGME shows structural similarities to other members of the SDR family, such as UDP-galactose 4-epimerase (UGE, Pdb accession codes 1nah and 1nai) (Thoden *et al.*, 1996). The overall rms deviation between both structures is 1.88 Å for 238 Ca positions considered equivalent (Deacon *et al.*, 2000) (Figure 47).

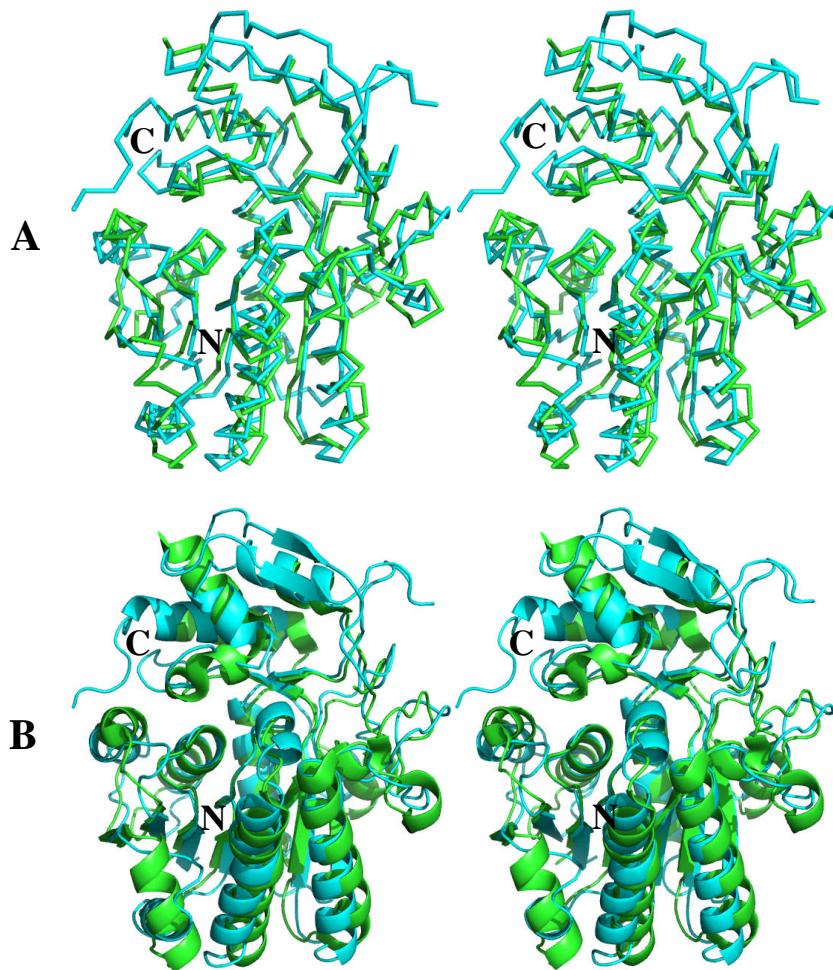


Figure 47. (A) Stereo view of $\text{C}\alpha$ tracing of AGME from *E. coli* (in cyan, Pdb accession code 1eq2) and UGE from *E. coli* (in green, Pdb accession code 1nah) after their superposition (based on Deacon *et al.*, 2000). (B) Ribbon diagrams of both structures after their superposition (stereo view). This figure has been produced using PyMol (DeLano, 2007).

Apart from UGE, AGME also displays similarity to dTDP-D-glucose 4,6 dehydratase (DGD, Pdb accession code 1bxk) with an rmsd of 1.85 Å for 237 $\text{C}\alpha$ positions (Deacon *et al.*, 2000), and GDP-4-keto-6-deoxy-D-mannose epimerase/reductase (GMER, Pdb accession code 1bws) (Rizzi *et al.*, 1998; Somers *et al.*, 1998) with an rms deviation of 2.14 Å for 226 $\text{C}\alpha$ positions (Deacon *et al.*, 2000).

The crucial amino acids in the catalytic region of UGE show a high degree of conservation in AGME. In ADP- β -L-glycero-D-mannoheptose 6-epimerase Ser116, Tyr140 and Lys144 represent the catalytic triad (Deacon *et al.*, 2000).

2.2.3.3 Reaction mechanism of AGME

It has been reported that ADP- β -L-glycero-D-mannoheptose 6-epimerase operates at an “unactivated” stereogenic centre which misses an acidic carbon-hydrogen connection (Tanner and Morrison, 2007; Morrison *et al.*, 2005; Samuel and Tanner, 2002; Tanner, 2002; Read *et al.*, 2004). This rules out a mechanism based on H⁺-abstraction and readdition at the site where inversion occurs (Morrison and Tanner, 2007; Morrison *et al.*, 2005; Read *et al.*, 2004). However, AGME shows a tight interaction with nicotinamide adenine dinucleotide phosphate and is a short chain dehydrogenase/reductase (Morrison and Tanner 2007; Morrison *et al.*, 2005; Read *et al.*, 2004; Ding *et al.*, 1994; Deacon *et al.*, 2000; Ni *et al.*, 2001). This suggests that during epimerisation the substrate is transiently oxidised (Morrison and Tanner, 2007; Morrison *et al.*, 2005; Read *et al.*, 2004). The fact that ADP- β -D-manno-hexodialdose **1** could be reduced to the alcohol ADP- β -mannose and oxidised to the acid ADP- β -mannuronate in a 1:1 ratio in the presence of ADP- β -L-glycero-D-mannoheptose 6-epimerase showed that in the reaction mechanism of AGME a redox step at the carbon at position 6 of a nucleotide sugar takes place (Figure 48) (Morrison *et al.*, 2005; Morrison and Tanner, 2007).

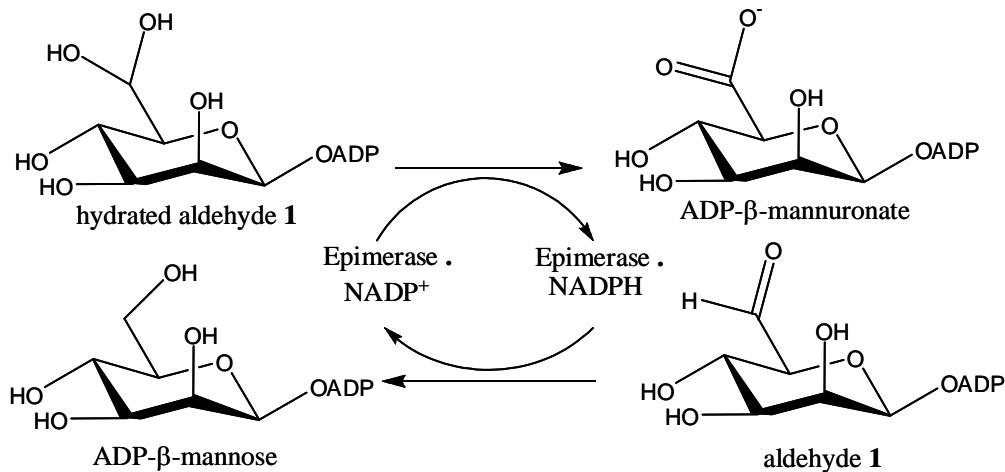


Figure 48. (after Morrison *et al.*, 2005 and Morrison and Tanner, 2007) The dismutation of ADP- β -D-manno-hexodialdose employed by AGME.

A mechanism for AGME has been proposed including transient oxidation of ADP-D,D-heptose at the carbon at position 6 to form a 6"-keto intermediate (Morrison and Tanner, 2007) (Figure 49). It has been suggested that two pockets have to be present, and each of them is able to bear either the OH- or HCOH-substituent at position 6. There is an acid/base amino acid at each site which has the ability to abstract a proton from the C-6"-OH of the substrate or the product in the oxidative part of the reaction. The C=O of the keto-intermediate undergoes a flip to the other site, caused by a rotation about the C-5"-C-6" bond (Morrison and Tanner, 2007; Morrison *et al.*, 2005) and the second acid/base adds H⁺ to the C=O in the reduction step on the opposite face. To accomplish H⁻-transfer during this reaction each of both epimers have to bind to the catalytic side in such a way that the H⁻ at C-6" of the sugar moiety is orientated toward the coenzyme (Morrison and Tanner, 2007; Morrison *et al.*, 2005).

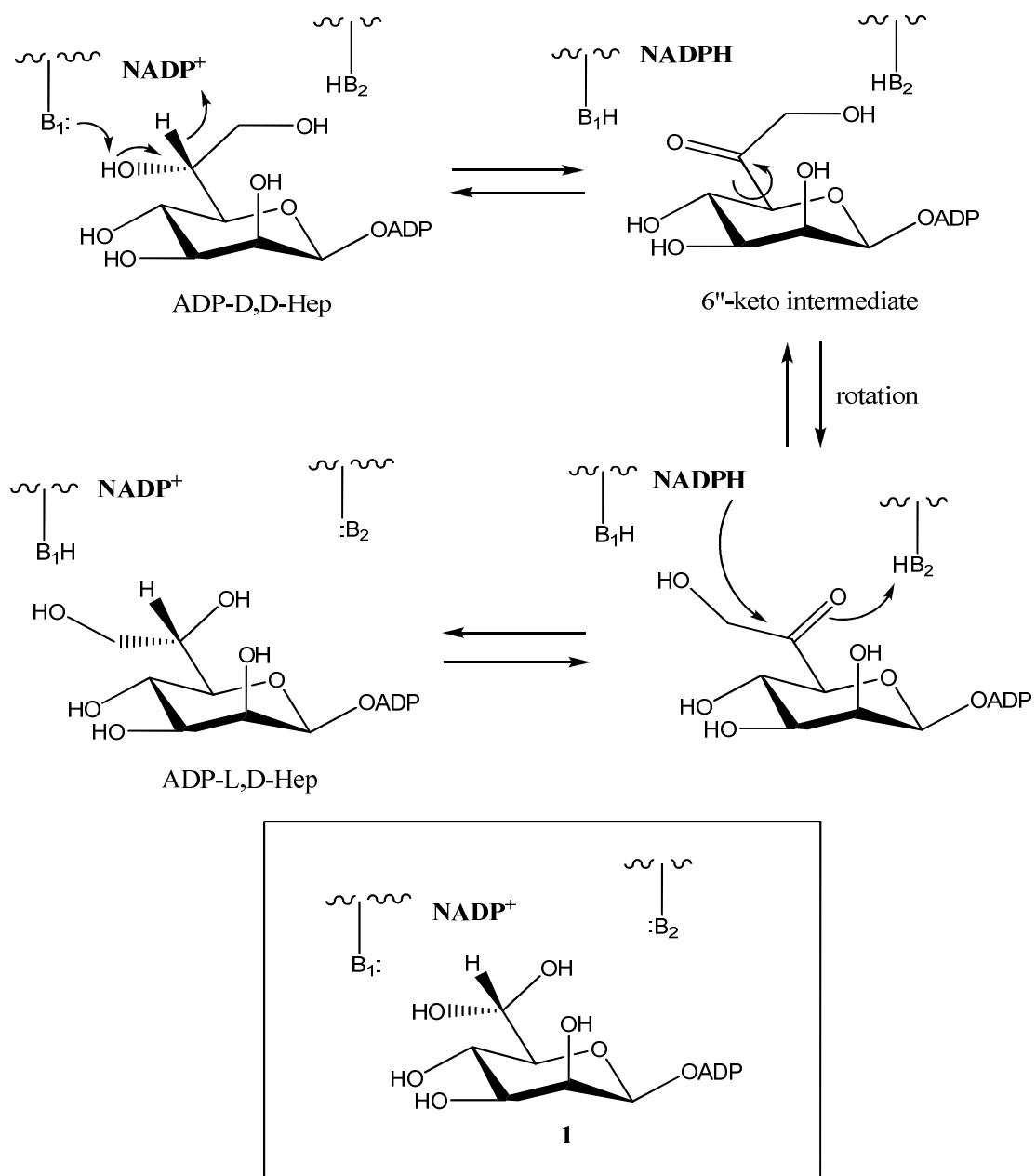


Figure 49. (after Morrison and Tanner, 2007) Proposed two-base mechanism for the reaction carried out by AGME. B₁ and B₂ are catalytic site acid/base amino acids. The hydrated form of ADP- β -D-mannohexodialdose 1 bound in the catalytic region is boxed.

To support this hypothesis, mutational studies were performed to find out which residues are candidates to function as the catalytic acid/base (Morrison and Tanner, 2007). These studies revealed that Tyr140 and Lys178 are the crucial amino acids, involved in catalysis as Y140F and K178 are inactive.

This indicates that “a non-stereospecific reduction of the ketone intermediate that leads to epimerization” (Morrison and Tanner, 2007) can only take place when both residues are present. Both mutants, however, were still able to dismutate ADP- β -D-manno-hexodialdose which shows that only one base is needed for the dismutation reaction (Morrison and Tanner, 2007). “One of the two stereoselective hydride transfers” (Morrison and Tanner, 2007) can still be carried out. The fact that the Y140F/K178M mutant could not even perform dismutation proved that at least one catalytic amino acid has to be present to perform H⁻-transfer (Morrison and Tanner, 2007).

It has been proposed that the Tyr always functions as the catalytic base which abstracts the H⁺ of the OH-group to oxidise an alcohol (or adds H⁺ to the C=O when reduction takes place) (Morrison and Tanner, 2007). It has been suggested that Tyr directly abstracts H⁺ from the substrate (Naismith, 2006; Morrison and Tanner, 2007), however, it has also been proposed that H⁺-shuttling, in which the catalytic Ser plays a role, may take place (Liu *et al.*, 1997; Berger *et al.*, 2001; Morrison and Tanner, 2007). In dTDP-6-deoxy-L-lyxo-4-hexulose reductase, for instance (Blankenfeldt *et al.*, 2002), H⁻ are transported to only one face of the C=O group. In epimerases, however, which prefer to transfer H⁻ directly, an H⁻ has to be transported to both faces of the C=O intermediate (Morrison and Tanner, 2007).

2.3 Project aims

Studies on the catalytic mechanism of ADP- β -L-glycero-D-mannoheptose 6-epimerase revealed that the amino acids Tyr140 and Lys178 operate as acid/base residues. However it still remains unclear which amino acid abstracts H⁺ from the OH-group of ADP-D-glycero-D-mannoheptose or ADP-L-glycero-D-mannoheptose. The crystal structure of AGME with ADP- α -D-glucose bound (Deacon *et al.*, 2000) suggests that Tyr140 abstracts H⁺ from ADP-L-glycero-D-mannoheptose, whereas Lys178 could perform H⁺-abstraction from ADP-D-glycero-D-mannoheptose.

However, it cannot be concluded from this structure, at which site oxidation takes place (Morrison and Tanner, 2007). The fact that the glucose moiety of the nucleotide sugar is not well defined in most of the subunits in the asymmetric unit of the AGME structure makes this difficult (Morrison and Tanner, 2007). Furthermore, the hexose is orientated differently in the catalytic sites of the monomers, in which ADP- α -D-glucose has been modelled in (Deacon et al., 2000; Morrison and Tanner, 2007). The reason for that is probably the difference in the stereochemistry between ADP- α -D-glucose and ADP- β -L-glycero-D-mannoheptose at C-1" and at C-2" as well (Figure 50) (Morrison and Tanner, 2007; Morrison *et al.*, 2005). Further experiments are therefore needed to elucidate the exact roles of Tyr140 and Lys178 in the epimerisation process of AGME (Morrison and Tanner, 2007). One approach to achieve this is to determine the crystal structure of HldD in complex with a sugar nucleotide in the β -manno configuration. One such a sugar is ADP- β -mannose (Figure 50), the reduced product (alcohol) in the dismutation reaction (Morrison and Tanner, 2007; Morrison *et al.*, 2005). This sugar, provided by our collaborators Martin Tanner and James Morrison at the University of British Columbia, Vancouver, Canada, was used in my crystallographic studies.

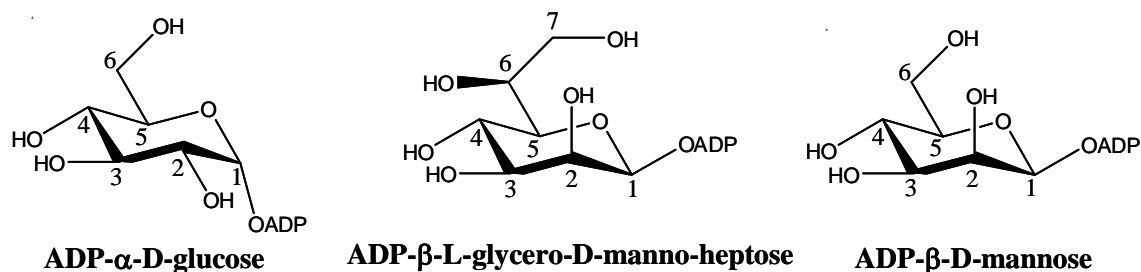


Figure 50. Structures of ADP- α -D-glucose, ADP- β -L-glycero-D-manno-heptose and ADP- β -D-mannose.

Molecular biology, expression, purification crystallisation and Isothermal Titration Calorimetry (ITC) of AGME and mutant Y140F

Summary

This section describes molecular biological methods, protein expression, purification, crystallisation and Isothermal Titration Calorimetry (ITC) of AGME and the AGME mutant Y140F.

2.4 Molecular biology

2.4.1 Sub-cloning of *hldD* into pEHISTEV

The *hldD* gene (UniProtKB/TrEMBL entry P67910) from *E. coli* K-12 W3110 comprising 933 bp, cloned into the pET-30Xa/LIC vector (Novagen) (see Appendix) (Morrison et al., 2005) was provided by our collaborators Martin Tanner and James Morrison from the University of British Columbia, Vancouver, Canada.

A design of forward and reverse primers was carried out to add the appropriate restriction endonuclease cleavage sites. This is necessary to clone *hldD* into the expression plasmid pEHISTEV (see Appendix) (Liu and Naismith, 2009).

An *Nco*I site was introduced at the start of *hldD* (forward primer 5' – GAGGGTCCCATGGTCATCGTTACC – 3'). Right at the end of the open reading frame, after the stop codon, an *Xho*I site was incorporated (reverse primer 5' – GGAGAGCTCGAGCCTTATGCGTC – 3'). A Polymerase Chain Reaction (PCR) was performed using 1 μM of each oligonucleotide and 1.5 U of Pfu DNA polymerase (Promega) (Table 12).

Table 12. Details of the PCR reaction for the amplification of *hldD*. The reaction was carried out in the GeneAmp® PCR System 2400 thermal cycler (Applied Biosystems)

Reagent	Amount/concentration of stock solution	final concentration/amount	PCR programme
sterile H ₂ O	38.5 μl		
Pfu buffer (+MgSO ₄)	5 μl	10x	2 min, 95 °C
dNTP Mix (dATP,dTTP, dGTP,dCTP)	1 μl	10 mM	30x [30 sec, 95 °C 30 sec, 55 °C 2 min/kb, 72 °C] 10 min, 72 °C
forward primer	1 μl	50 μM	1 μM
reverse primer	1 μl	50 μM	1 μM
template DNA	3 μl	89 ng/μl	267 ng
Pfu DNA Polymerase	0.5 μl	3 U/μl	1.5 U

Electrophoresis on a 1 % agarose gel confirmed the 956 bp PCR product (Figure 51).

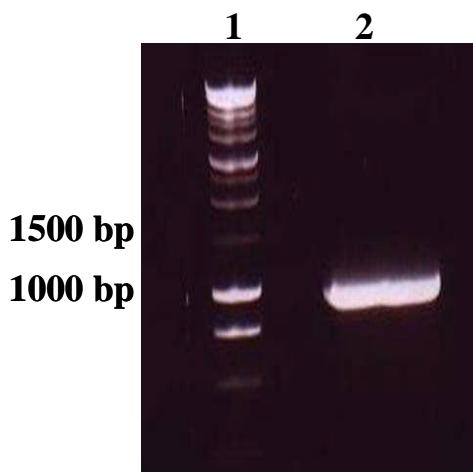


Figure 51. PCR product of *hldD* from pET-30 Xa/LIC. Lane 1: 1 kbp ladder; lane 2: amplified 956 bp PCR product

The amplified *hldD* fragment was excised from the 1 % agarose gel according to the protocol provided in the QIAquick™ Gel Extraction Kit (Qiagen). 50 µl sterile water was used to elute the DNA from the spin column. Then a restriction digest of eluted amplified *hldD* and the pEHISTEV plasmid (provided by Huanting Liu) using *Nco*I and *Xho*I (New England BioLabs® Inc.) was performed in a water bath for 3.5 hours at 37 °C (Table 13).

Table 13. Reagents used in the restriction digest of amplified *hldD* and pEHISTEV

Reaction 1:	Reaction 2:
2 µl buffer D (10x)	2 µl buffer D (10x)
0.2 µl acetylated BSA (10 µg/ml)	0.2 µl acetylated BSA (10 µg/ml)
15 µl amplified <i>hldD</i>	15 µl pEHISTEV
1.8 µl H ₂ O	1.8 µl H ₂ O
0.5 µl <i>Nco</i> I	0.5 µl <i>Nco</i> I
0.5 µl <i>Xho</i> I	0.5 µl <i>Xho</i> I

The double digested gene and vector as well were then solution cleaned, following the protocol provided in the QIAquick™ Gel Extraction Kit (Qiagen) to remove the restriction endonucleases and buffers. Ligating the *hldD* fragment into the pEHISTEV vector, which contains a kanamycin resistance marker was the next step. Ligation was carried out overnight at 4 °C with 3U of T4 DNA Ligase (Promega) using an insert to vector ratio of 3 : 1. TAM1 and JM109 competent *E. coli* cells were then transformed with the ligation mixture and plated onto agar plates containing 50 µg/ml kanamycin (Melford). Amplification reactions with the T7 forward and reverse primers were carried out for the identification of the colonies with the pEHISTEV plasmid and the inserted *hldD* gene. Colonies were picked from a kanamycin containing agar plate and suspended in 10 µl sterile water filled into PCR tubes. The samples were then boiled for 5 min at 100 °C and centrifuged. After that 5µl of sample were taken out of each of the tubes and transferred to sterile ones. The amplification reactions were performed using 0.5 µM of each oligonucleotide and 1.36 U of GoTaq® DNA-Polymerase (Promega) (Table 14).

Table 14. Details of the PCR reactions for the amplification of *hldD* cloned into pEHISTEV. Reactions were carried out in the GeneAmp® PCR System 2400 thermal cycler (Applied Biosystems)

“Master Mix” for 25 reactions				
Reagent	Amount/concentration of stock solution		final concentration/	PCR programme
sterile H ₂ O	139.1 µl			
GoTaq® DNA Polymerase	1.9 µl	5 U/µl	1.36 U	2 min, 95 °C
GoTaq® Green	70 µl	5x	1x	30x [20 sec, 95 °C 50 sec, 46 °C 1 min, 72 °C]
Reaction Buffer				
dNTP Mix (dATP,dTTP, dGTP,dCTP)	7 µl	10 mM	0.2 mM	5 min, 72 °C
T7 forward primer	3.5 µl	50 µM	0.5 µM	
T7 reverse primer	3.5 µl	50 µM	0.5 µM	
9 µl of “Master Mix” were added to each of the 22 PCR tubes				

Of 22 colonies screened 3 were confirmed as containing the 1 kb *hldD* fragment (“clones 1, 2 and 3”). Restriction analysis and DNA sequencing (University of Dundee Sequencing service) were also carried out to identify the presence of the *hldD* insert in pEHISTEV (Table 15 and Figure 52).

Table 15. Reagents used in the restriction digest

Reaction 1:	Reaction 2:	Reaction 3:
1 µl buffer D (10x)	1 µl buffer D (10x)	1 µl buffer D (10x)
0.1 µl acetylated BSA (10 µg/ml)	0.1 µl acetylated BSA (10 µg/ml)	0.1 µl acetylated BSA (10 µg/ml)
5 µl “clone 1”	5 µl “clone 2”	5 µl “clone 3”
1.9 µl H ₂ O	1.9 µl H ₂ O	1.9 µl H ₂ O
1 µl <i>Nco</i> I	1 µl <i>Nco</i> I	1 µl <i>Nco</i> I
1 µl <i>Xho</i> I	1 µl <i>Xho</i> I	1 µl <i>Xho</i> I

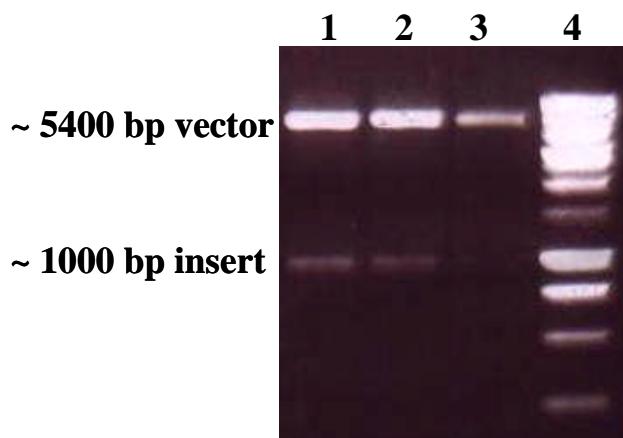


Figure 52. Restriction analysis with *Nco*I and *Xho*I confirming the insertion of *hldD* into pEHISTEV.

Lane 1: “clone 1” : lane 2: “clone 2”; lane 3: “clone 3”; lane 4: 1 kbp DNA ladder

The *hldD* gene was sequenced to make sure that no nucleotide substitutions had been introduced. The *hldD*-pEHISTEV construct was purified from 10 ml overnight cultures of the TAM1 or JM109 transformed *E. coli* cells grown in Luria-Bertani (LB) media (Formedium) containing 50 µg/ml kanamycin at 37 °C. The plasmid DNA was then purified according to the QIAprep Spin Miniprep Kit Protocol using a microcentrifuge (Qiagen).

2.5 Protein expression

E. coli cells BL21 (DE3), C43 (DE3), Rosetta (DE3) and BL21 (DE3) pLysS were transformed with *hldD*-pEHISTEV and *hldD*-pET-30 Xa/LIC vector DNA for overexpression of the *hldD* gene. Colonies of these cells were inoculated in various 10 ml overnight cultures of Luria Bertani media containing kanamycin and also chloramphenicol (Sigma) (only in the case of Rosetta (DE3) and BL21 (DE3) pLysS), at a final concentration of 50 µg/ml and 34 µg/ml, respectively. Growing of these starter cultures was performed overnight at 37 °C in a shaking incubator (Innova™ 4430, New Brunswick) at 200 rpm. The next day, inoculation of 10 ml of pre-warmed LB (supplemented with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol) with 100 µl overnight culture was carried out. These 10 ml cultures were then grown at 37 °C in a shaking incubator at 200 rpm. *E. coli* cells were grown until an optical density ($A_{600\text{nm}}$) of 0.4-0.8 (Helios Unicam Spectrophotometer). Then cells were induced with Isopropyl β-D-thio-galactopyranoside (IPTG) (Melford) in a range of concentrations between 0.2 and 2 mM and grown for a further 12 hours at 37 °C or at a lower temperature at 25 °C or 16 °C. In some experiments (see Table 16) tubes containing the growing cells were put on ice for 10 min (“cold shocked”) or in a waterbath at 42 °C for 10 min (“heat shocked”) and then cells were induced with IPTG. The media were then centrifuged for 20 min at 3000 rpm (Heraeus Labofuge 400, DJB labcare) at 4 °C and the supernatants discarded. Then the pelleted cells were resuspended in 500 µl BugBuster® Protein Extraction Reagent (Novagen). The resuspended cells were then spun at 3000 rpm for 25 min at 4 °C and the supernatants collected. Crude extracts and supernatants were mixed with sample buffer (Invitrogen NuPage™ LDS 4x sample buffer) and denatured for 5 min at 100 °C.

Then they were loaded onto a Sodium Dodecyl Sulfate (SDS) polyacrylamide gel (Invitrogen NuPage™ 4-12 % Bis-Tris Gels, 1 mm x 10, 12 or 17 wells) and run for 40 min at 200 V (1x MES buffer, Invitrogen NuPAGE MES SDS Running Buffer (for Bis-Tris Gels only) (20X)). SDS-PAGE gels were then stained in a Coomassie blue staining solution for approximately 10 min. For destaining, gels were then put into a beaker containing hot water and boiled for about 10 min in a microwave. After this procedure the gels were dried with the DryEASE Kit (Invitrogen). An overview of all expression trials carried out is demonstrated in Table 16. The highest amount of soluble AGME was observed when *E. coli* BL21 (DE3) [pET-30 Xa/LIC-*hldD*] and [pET-30 Xa/LIX-*hldD-YI40F*] (Figure 53A, B) (pET-30 Xa/LIC-*hldD-YI40F* was also provided by our collaborators in Canada) or *E.coli* BL21 (DE3) pLysS [pEHISTEV-*hldD*] (Figure 53C), grown in LB media until an $A_{600\text{nm}} = 0.7$, were induced with 0.2 mM IPTG. Then the temperature was decreased to 25 °C for soluble protein expression overnight. The same conditions for soluble expression were repeated in a large scale for purification.

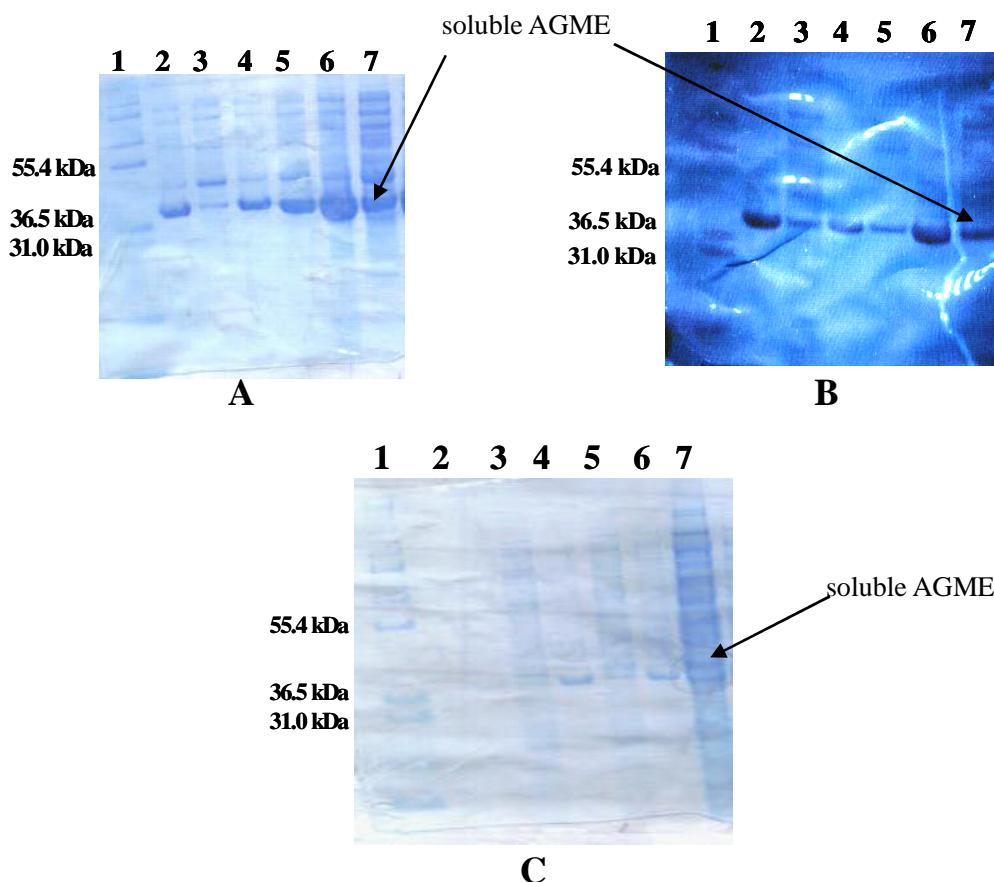


Figure 53. SDS-PAGE showing hexa-histidine tagged AGME and AGME Y140F expressed from pET-30 Xa/LIC (A and B), and hexa-histidine tagged AGME expressed from pEHISTEV (C). (A) Lane 1: Mark 12TM Protein Standard; lane 2: whole cell extract of AGME (expression for two hours at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant; (B) lane1: Mark 12TM Protein Standard : lane 2: whole cell extract of AGME Y140F (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME Y140F (expression for three hours at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME Y140F (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant; (C) lane 1: Mark 12TM Protein Standard; lane 2: whole cell extract of AGME (expression for two hours at 37 °C after induction with 0.2 mM IPTG); lane 3: corresponding supernatant; lane 4: whole cell extract of AGME (expression overnight at 37 °C after induction with 0.2 mM IPTG); lane 5: corresponding supernatant; lane 6: whole cell extract of AGME (expression overnight at 25 °C after induction with 0.2 mM IPTG); lane 7: corresponding supernatant.

Table 16. Overview of performed expression trials. Conditions of best soluble AGME expression are highlighted.

<i>E. coli</i> strain	temperature	expression vector	IPTG [mM]	heatshock (hs)	solubility: soluble (s) insoluble (i)
				cold shock (cs)	no shock (ns)
BL21 (DE3)	37 ° C	pEHISTEV	0.2	ns	i
BL21 (DE3)	37 ° C	pEHISTEV	0.5	ns	i
BL21 (DE3)	37 ° C	pEHISTEV	1	ns	i
BL21 (DE3)	37 ° C	pEHISTEV	2	ns	i
BL21 (DE3)	25 ° C	pEHISTEV	0.2	ns	i
BL21 (DE3)	25 ° C	pEHISTEV	0.5	ns	i
BL21 (DE3)	25 ° C	pEHISTEV	1	ns	i
BL21 (DE3)	25 ° C	pEHISTEV	2	ns	i
BL21 (DE3)	25 ° C	pEHISTEV	0.5	hs	i
BL21 (DE3)	25 ° C	pEHISTEV	0.5	cs	i
BL21 (DE3)	16 ° C	pEHISTEV	0.2	ns	i
BL21 (DE3)	16 ° C	pEHISTEV	0.5	ns	i
BL21 (DE3)	16 ° C	pEHISTEV	0.2	hs	i
BL21 (DE3)	16 ° C	pEHISTEV	0.2	cs	i
BL21 (DE3)	16 ° C	pEHISTEV	0.5	hs	i
BL21 (DE3)	16 ° C	pEHISTEV	0.5	cs	i
BL21 (DE3)	37 ° C	pET-30 Xa/LIC	0.2	ns	s
BL21 (DE3)	37 ° C	pET-30 Xa/LIC	0.5	ns	s
BL21 (DE3)	25 ° C	pET-30 Xa/LIC	0.2	ns	s
BL21 (DE3)	16 ° C	pET-30 Xa/LIC	0.2	ns	s
C43 (DE3)	37 ° C	pEHISTEV	0.2	ns	i
C43 (DE3)	37 ° C	pEHISTEV	0.5	ns	i
C43 (DE3)	37 ° C	pEHISTEV	1	ns	i
C43 (DE3)	37 ° C	pEHISTEV	2	ns	i
C43 (DE3)	25 ° C	pEHISTEV	0.2	ns	i
C43 (DE3)	25 ° C	pEHISTEV	0.5	ns	i
C43 (DE3)	25 ° C	pEHISTEV	1	ns	i
C43 (DE3)	25 ° C	pEHISTEV	2	ns	i
C43 (DE3)	25 ° C	pEHISTEV	0.5	ns	i
C43 (DE3)	25 ° C	pEHISTEV	0.5	ns	i
Ros. (DE3)	37 ° C	pEHISTEV	0.2	ns	i
Ros. (DE3)	37 ° C	pEHISTEV	0.5	ns	i
Ros. (DE3)	37 ° C	pEHISTEV	0.2	ns	i
Ros. (DE3)	16 ° C	pEHISTEV	0.5	ns	i
Ros. (DE3)	16 ° C	pEHISTEV	0.2	ns	i
BL21 (DE3) pLysS	37 ° C	pEHISTEV	0.5	ns	s
BL21 (DE3) pLysS	37 ° C	pEHISTEV	0.2	ns	s
BL21 (DE3) pLysS	25 ° C	pEHISTEV	0.2	ns	s
BL21 (DE3) pLysS	16 ° C	pEHISTEV	0.2	ns	s

2.6 Purification of recombinant AGME and AGME Y140F

About 30g of cell pellets, frozen at -80 °C after expression, were resuspended in 30-35 ml sample buffer containing 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM Imidazole, 1 % Triton X-100 and one Complete EDTA free Protease Inhibitor Cocktail Tablet (Roche). Then DNase was added to this buffer which leads to a degradation of the *E. coli* DNA. Lysis of the cells was performed by sonication (MSE Soniprep 150). The lysate was then centrifuged at 7,000 rpm for 20 min at 4 °C (Beckman Coulter™ Avanti™ J-20 XP). About 30 ml of supernatant were transferred to a 50 ml tube and kept on ice. Remaining pellets were resuspended in about 30 ml sample buffer and again sonicated. After that procedure, centrifugation at 7,000 rpm for 20 min at 4 °C was also performed again. This time the fraction was discarded. The supernatant and the previous supernatant fraction, kept on ice, were pooled and spun at 22,000 rpm for 35 min. The supernatant obtained after this centrifugation step was filtered (0.2 µM filter, Millipore) and loaded onto a 5 ml Ni²⁺ column (HisTrap HP™, Amersham Biosciences) pre-equilibrated with sample buffer. After the flow through had been collected about 10 column volumes (50 ml) of washing buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 20 mM Imidazole) were applied to the Ni²⁺ column. This step was performed to remove any proteins not bound to the column. The next step was the elution of the hexa-histidine tagged AGME with about 50 ml of elution buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM Imidazole). All the fractions obtained during this first purification step were analysed by SDS-PAGE (Figure 54).

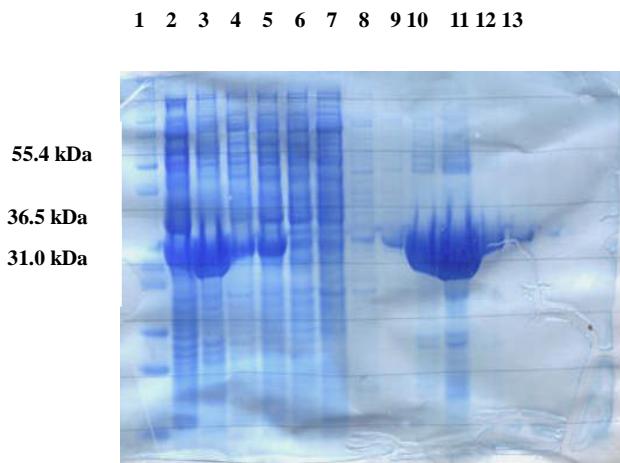


Figure 54. SDS-PAGE showing all the fractions obtained during the first purification step. Lane 1: Mark 12TM Protein Standard : lane 2: supernatant after first lysis; lane 3: cell extract after second lysis; lane 4: supernatant after second lysis; lane 5: final supernatant; lane 6: Flowthrough; lanes 7-9: wash fractions; lanes 10-13: eluted AGME containing fractions

The eluted AGME fractions (about 40 ml) were pooled and dialysed at 4 °C against 2 L buffer containing 50 mM NaCl and 250 mM Tris-HCl, pH 7.5 for 2 hours to remove the imidazole. The enzyme was then dialysed for further 2 hours against fresh buffer. A third dialysis step was performed overnight in another batch of fresh buffer. The next morning the dialysed protein was poured into a beaker, and ammonium sulfate, to give 25 % saturation was gradually added to the sample, which was kept at 4 °C, over a time of 1:30 hours to precipitate other contaminant proteins. After that the sample was centrifuged at 20,000 rpm for 30 min at 4 °C. The pellet was discarded and the remaining supernatant syringed through a 0.45 µM filter (Sartorius Stedim Biotech GmbH (SSB)). AGME was then applied to a 5 ml hydrophobic column (HiTrapTM Butyl FF, Amersham Biosciences). Buffer A (250 mM Tris-HCl, 50 mM NaCl pH 7.5, ammonium sulfate, 25 % saturation) and buffer B (250 mM Tris-HCl, pH 7.5, 50 mM NaCl) were used in this purification step performed on an ÄKTAXpress (Amersham Biosciences). First the column was pre-equilibrated with buffer A.

After the protein sample had been applied to the column elution of AGME over a gradient during which the concentration of buffer B increased (0-30 % buffer B, length: 5 column volumes; 30-100 % buffer B, length: 25 column volumes) was performed. Eluted enzyme fractions were analysed by SDS-PAGE (Figure 55).

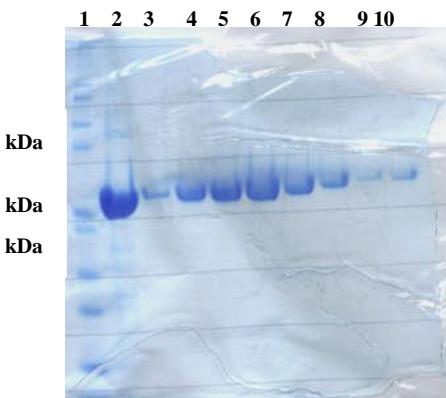


Figure 55. SDS-PAGE showing the fractions of AGME eluted from the HiTrap™ Butyl FF column. Lane 1: Mark 12™ Protein Standard : lane 2: AGME sample before its application to the column; lanes 3-10: eluted AGME fractions

Fractions containing pure AGME (about 170 ml) were pooled and dialysed against 5 L buffer C (20 mM Tris-HCl, pH 7.0, 50 mM NaCl, 1 mM Dithiothreitol (DTT) (Melford)) for 2-3 hours at 4 °C. Then a second dialysis step was carried out against 5 L fresh buffer F at 4 °C overnight. The next morning the dialysed enzyme was concentrated to about 5 mg/ml (NanoDrop 3.1.0) using Ultrafiltration Membranes YM10 (NMWL 10K) or YM30 (NMWL 30K) (Millipore) for the third and final purification step. Concentrated and filtered AGME samples (about 5 ml) were then loaded onto a HiPrep™ 16/60 Sephadryl™ S-200 size exclusion column (Amersham Biosciences). This step to separate AGME from further contaminant proteins was performed again on the ÄKTAXpress (Amersham Biosciences). AGME fractions eluted in buffer C were then analysed by SDS-PAGE (Figure 56).

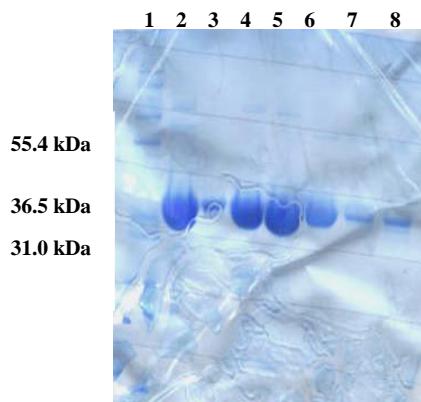


Figure 56. SDS-PAGE showing the fractions of AGME eluted from the HiPrep™ 16/60 Sephadryl™ S-200 size exclusion column. Lane 1: Mark 12™ Protein Standard : lane 2: AGME sample before its application to the column; lanes 3-8: eluted AGME fractions

Fractions containing pure enzyme were then pooled and concentrated (VIVASPIN 20, 10,000 or 30,000 MWCO PES, Sartorius Stedim Biotech GmbH (SSB)) to 4.1 and 7 mg/ml (Figure 57). 100 or 200 µl AGME were then aliquoted into PCR tubes, snap frozen in liquid nitrogen and stored at -80 °C until further use.

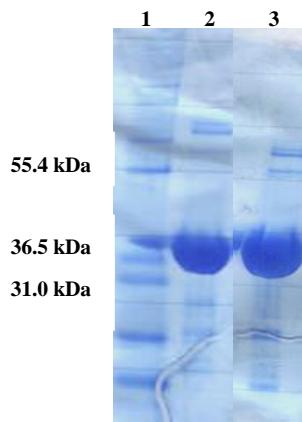


Figure 57. SDS-PAGE showing concentrated AGME before being stored at -80 °C. Lane 1: Mark 12™ Protein Standard : lane 2: AGME at 4.1 mg/ml; lane 3: AGME at 7 mg/ml

2.6.1 Preparation and reconstitution of apo-AGME

2.6.1.1 Preparation of apo-AGME (based on Gomi *et al.*, 1989)

Gomi and coworkers performed a treatment of rat liver S-adenosylhomocysteine with acidic $(\text{NH}_4)_2\text{SO}_4$ to separate this enzyme from its cofactor NAD. The intention of this experiment was to obtain the apoenzyme for further exploring the associations of this protein with its cofactor (Gomi *et al.*, 1989). Ni *et al.*, 2001 also used the same method to obtain apo-AGME. After reconstitution of AGME with either NAD^+ , NADP^+ or FAD they confirmed that AGME prefers NADP^+ as cofactor. (Ni *et al.*, 2001) The same methods of apoenzyme preparation and reconstitution with cofactor was also applied to my AGME sample to get the holoenzyme with homogeneous cofactor (NADP^+) content.

Firstly, after ADP- β -L-glycero-D-mannoheptose 6-epimerase in 20 mM Tris-HCl, pH 7.0, 50 mM NaCl and 1 mM DTT, stored at -80 °C, had been thawed on ice, saturated acidic $(\text{NH}_4)_2\text{SO}_4$ (adjusted to pH 3.0 with concentrated H_2SO_4) was added to the sample in a ratio of 3:1. Then the mixture was stirred for about half an hour at 4 °C. After this procedure centrifugation at 22,000 rpm for 20 min at 4 °C was performed. Then the supernatant was discarded and the pellet carefully resuspended in 20 mM Tris-HCl, pH 7.0, 50 mM NaCl and 5 mM DTT. After the enzyme had been mixed with acidic $(\text{NH}_4)_2\text{SO}_4$ for a second time, stirred and spun again, the pellet was gently resuspended in neutral $(\text{NH}_4)_2\text{SO}_4$ containing 5 mM DTT. This mixture was then centrifuged at 25,000 rpm for 30 min at 4 °C. The supernatant was discarded and the precipitated protein kept on ice overnight. The following day the pellet was gently dissolved in 20 mm Tris-HCl, pH 7.0, concentrated, filtered and applied to a HiPrep™ 26/10 Desalting column (Amersham Biosciences). This step, performed on the ÄKTAxpress, (Amersham Biosciences) removes any traces of ammonium sulfate and DTT.

2.6.1.2 Reconstitution of apo-AGME (based on Ni *et al.*, 2001)

The desalted apo-enzyme was then concentrated again. Incubation of 100 or 200 µM NADP⁺ (Sigma) at room temperature for one hour was then carried out. After that the reconstituted protein was dialysed against 1 L of 20 mM Tris-HCl buffer pH 7.0 overnight at 4 °C to remove any unbound cofactor. The next morning the dialysed sample was concentrated to about 5 mg/ml, filtered and loaded onto a HiPrepTM 16/60 SephacrylTM S-200 size exclusion column. This step was carried out to ensure a complete separation of holo- and apo-HldD. Finally the reconstituted AGME was concentrated to about 4.1 mg/ml, aliquoted into PCR tubes, snap frozen in liquid nitrogen and again stored at -80 °C until further usage.

2.7 Crystallisation

Prior to crystallisation attempts, AGME samples at concentrations of 4.1mg/ml and/or 7 mg/ml which had been stored at -80 °C were thawed on ice and spun for 5 min at 4 °C. ADP-β-D-mannose was then added to final concentrations between 1 and 6.6 mM. The samples were then incubated for 16 hours on ice or 2 hours at room temperature. After that they were centrifuged at either 4 °C (MSE Micro Centaur) or room temperature (Heraeus Biofuge Pico) before use.

2.7.1 Crystallisation of wildtype AGME

The crystallisation experiments were carried out using “grid screens” based on the condition (2 M ammonium sulfate, 0.1 M HEPES-Na pH 7.5, 2 % PEG 400) in which well diffracting AGME crystals in complex with ADP- α -D-glucose had been obtained (Ding *et al.*, 1999). The sitting-drop vapour diffusion method was applied using Douglas Instruments 96-well crystallisation plates (Douglas Instruments) at room temperature. Protein and mother liquor were mixed in a ratio of 1:1 (1 μ l + 1 μ l, reservoir volume 100 μ l). After the plates had been sealed they were incubated at 20 °C. However, none of the conditions yielded any crystals. For that reason AGME was screened for other crystallisation conditions against the commercially available 96-condition screens, JCSG⁺ and Classics (GE Healthcare), using a nanodrop crystallisation robot (Cartesian Honeybee X8+1, Genomic Solutions) on sitting drop plates (Hampton Research CrystalQuickTM 96 wells, 4 μ l square drop well, 3 drop well). The 288 wells on those plates allowed screening of two different concentrations (well # 1, 4.1 mg/ml; well # 2, 7 mg/ml) and also of a distinct drop size (well # 3, 4.1 mg/ml in a protein:mother liquor ratio of 2:1) per crystallisation condition. Initial crystals which appeared within one week were obtained under various different conditions. Staining of those crystals with Izit Crystal Dye (Hampton Research) confirmed that they were protein. The most promising looking crystallisation conditions (Figure 58) were used for further optimisation experiments based on different concentrations and pH values of the components of the mother liquor (see Appendix).

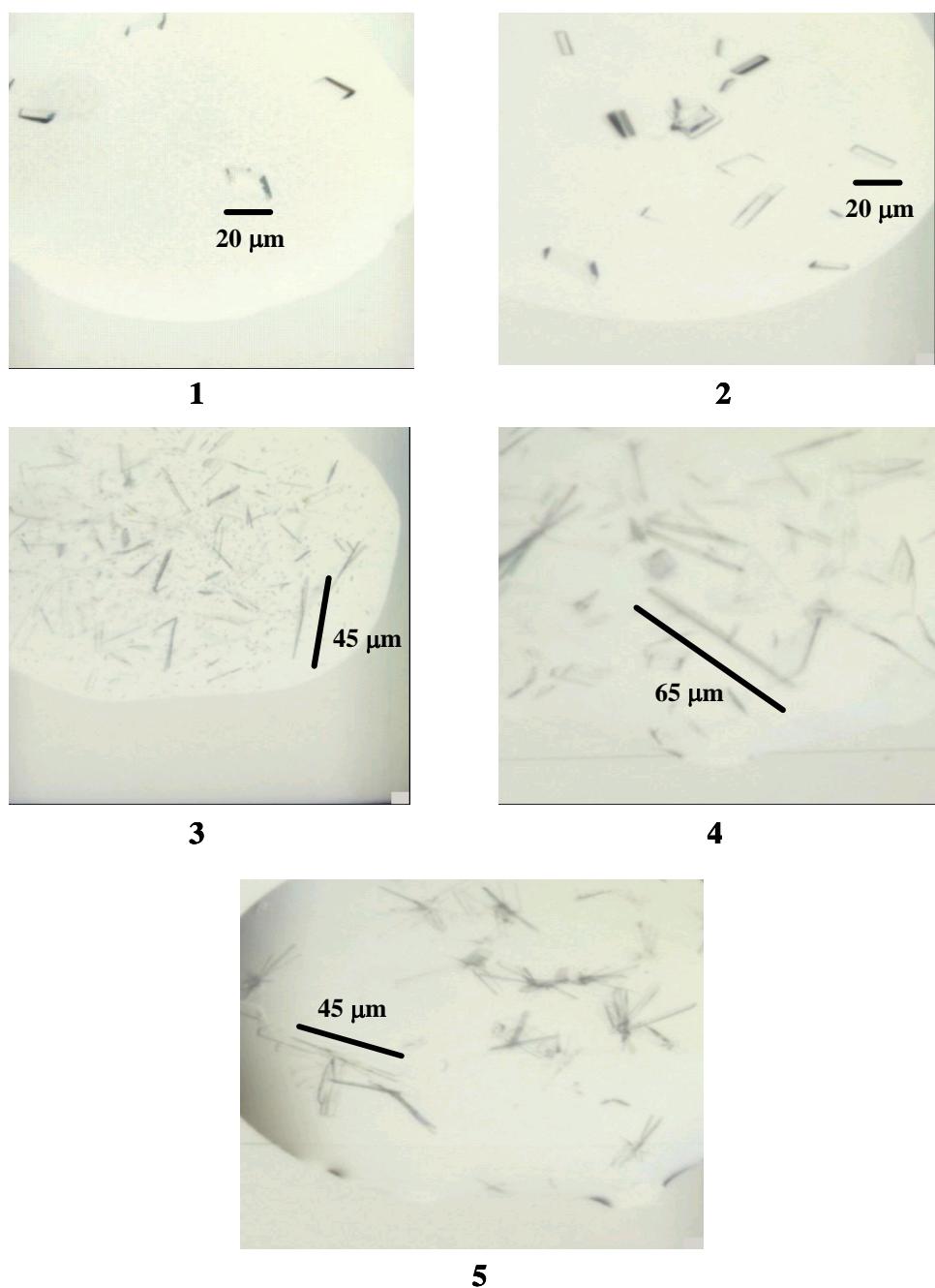


Figure 58. Crystals of AGME grown in presence of ADP- β -mannose (5-6.6 mM). 1: condition # 46 from JCSG⁺ screen; 2: condition # 74 from Classics screen; 3: condition # 2 from JCSG⁺ screen; 4: condition # 78 from JCSG⁺ screen; 5: condition # 5 from Classics screen

Crystals large enough for an X-ray diffraction experiment were obtained from the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300 (starting condition: JCSG⁺, condition # 46, see Appendix) (Figure 59) after approximately two weeks.

AGME at a concentration of 4.1 mg/ml was incubated with 10 mM ADP- β -mannose on ice overnight. Drops were made up manually by mixing 1.5 μ l protein solution with 0.75 μ l mother liquor (ratio 2:1) and the crystallisation plates were then incubated at 20 °C.

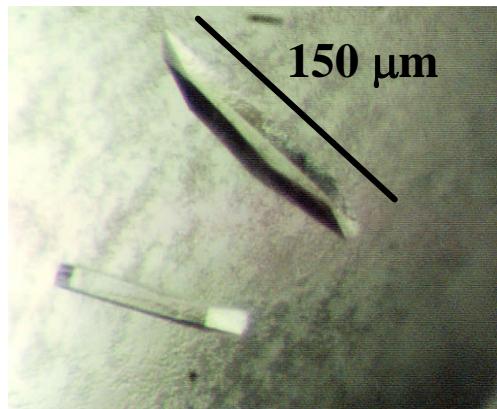


Figure 59. Optimised crystals of AGME grown in presence of 6.6 mM ADP- β -mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

Cryoprotection of these crystals was not necessary due to the presence of a cryoprotectant at a relatively high concentration (35 % PEG 300) in the mother liquor. These crystals diffracted to about 3 Å resolution in-house. The “grid screen” based on the condition 0.2 M CaCl₂, 0.1 M HEPES-Na, pH 7.5 and 28 % PEG 800 (Classics # 74) only yielded fragile, poorly diffracting crystals. No improvement of the crystal quality could be achieved with the other “grid screens” (see Appendix) either. Only small, needle-shaped crystals (Figure 58, 3-5) which diffracted very poorly appeared after 1-3 days.

2.7.2 Crystallisation of wildtype AGME reconstituted with NADP⁺

Based on the results obtained from the crystallisation experiments carried out with non-reconstituted AGME a “grid screen” (see Appendix) around condition # 46 of the JCSG⁺ screen was set up again. The protein solution was incubated with 5 mM ADP-β-mannose for about 2 hours at room temperature. Again, drops were made up manually by mixing 1.5 µl protein solution with 0.75 µl mother liquor (ratio 2:1). Crystallisation plates were then incubated at 20 °C. The largest crystals were again obtained under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300 (Figure 60) after 1-2 weeks.

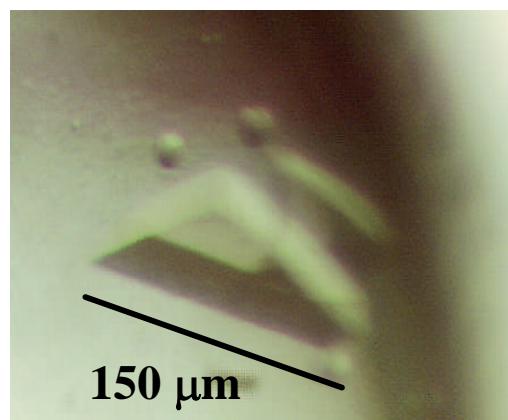


Figure 60. Optimised crystals of reconstituted AGME grown in presence of 2.5 mM ADP-β-mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

2.7.3 Crystallisation of AGME Y140F

The same “grid screen” (see Appendix) around condition # 46 of the JCSG⁺ screen was set up to grow crystals of the mutant protein. AGME Y140F at a concentration of 4.1 mg/ml was incubated with 2 or 5 mM ADP-β-mannose for about 2 hours at room temperature.

Drops were made up again manually by mixing protein solution with mother liquor in a ratio of 2:1. Crystallisation plates were then incubated at 20 °C. The best crystals were grown again from 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300 (Figure 61) and appeared after around 3 weeks.

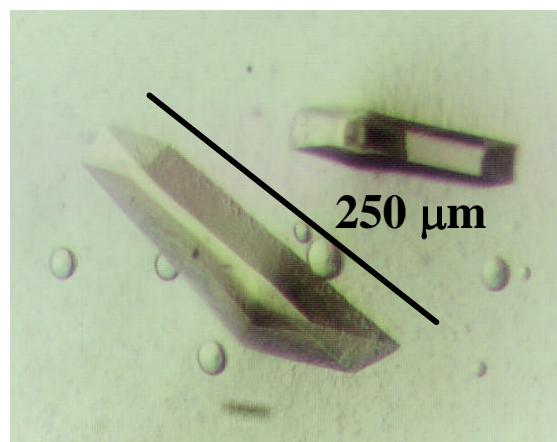


Figure 61. Optimised crystals of AGME Y140F grown in the presence of 2.5 mM ADP- β -mannose under the condition 0.2 M Ca-Acetate, 0.1 M Na-cacodylate pH 6.5 and 35 % PEG 300.

Additionally the mutant was also screened for new crystallisation conditions against the commercially available 96-condition screens, JCSG⁺, Classics, PEGs and pH clear (GE Healthcare), using the nanodrop crystallisation robot (Cartesian Dispensing Robot Honeybee 963, Genomic Solutions) on sitting drop plates (InnovaplateTM SD-2 Crystallography Plate, Innovadyne Technologies, Inc.). The 192 wells on those plates allowed screening of two different concentrations (well # 1, 4.1 mg/ml; well # 2, 7 mg/ml) per crystallisation condition. Initial crystals which appeared in a time range between 2 days and 2 weeks were obtained under lots of different conditions. The most promising looking crystallisation conditions (Figure 62) were again used for further optimization experiments based on different concentrations and pH values of the components of the mother liquor (see Appendix).

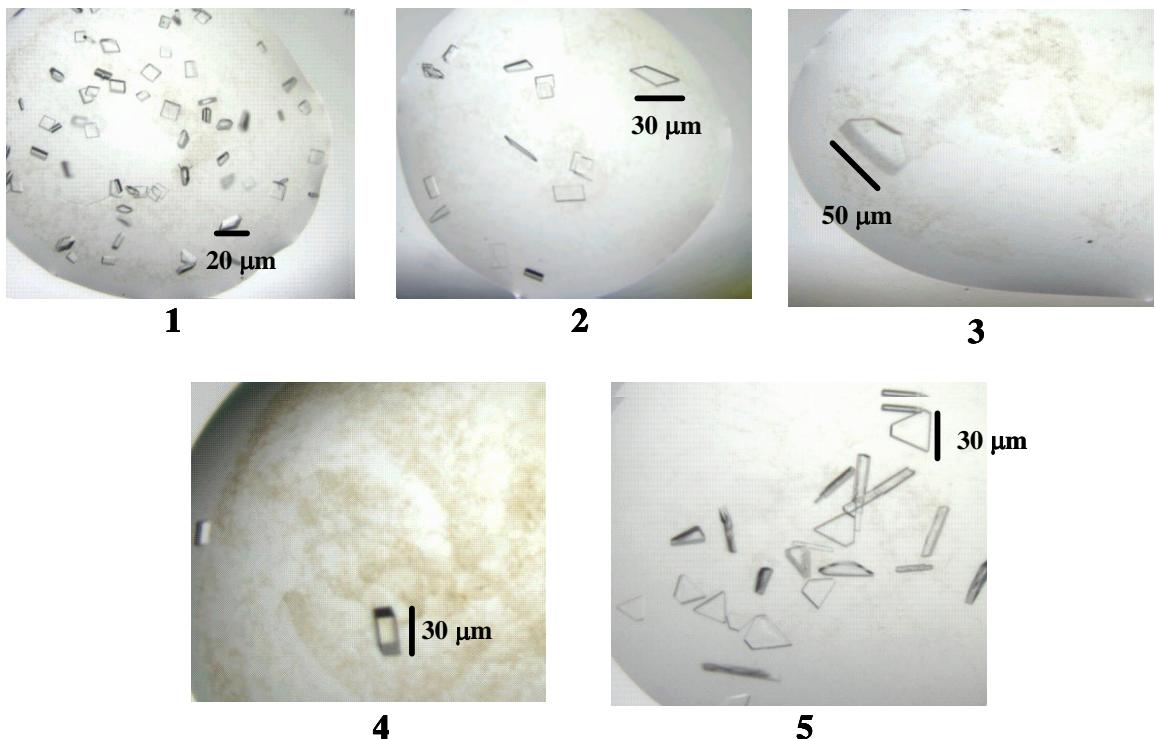


Figure 62. Crystals of AGME Y140F grown in presence of ADP- β -mannose (2.5 mM). 1: condition # 67 from PEGs screen; 2: condition # 54 from PEGs screen; 3: condition # 76 from PEGs screen; 4: condition # 82 from Classics screen; 5: condition # 73 from Classics screen

Good quality crystals were only grown from conditions of the “grid screen” around condition # 73 of the Classics screen. The best diffracting crystals appeared over the course of 1-2 weeks in drops containing 2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES-Na, pH 7.1, 2 % PEG 400 (Figure 63A) and 2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400, 10 mM Spermidine (Figure 63B). AGME Y140F at a concentration of 4.1 mg/ml was incubated with 2 mM ADP- β -mannose at room temperature for about 2 hours. Drops were made up manually by mixing 1.5 μl protein solution with 1.5 μl mother liquor (ratio 1:1) and the crystallisation plates were then incubated at 20 °C.

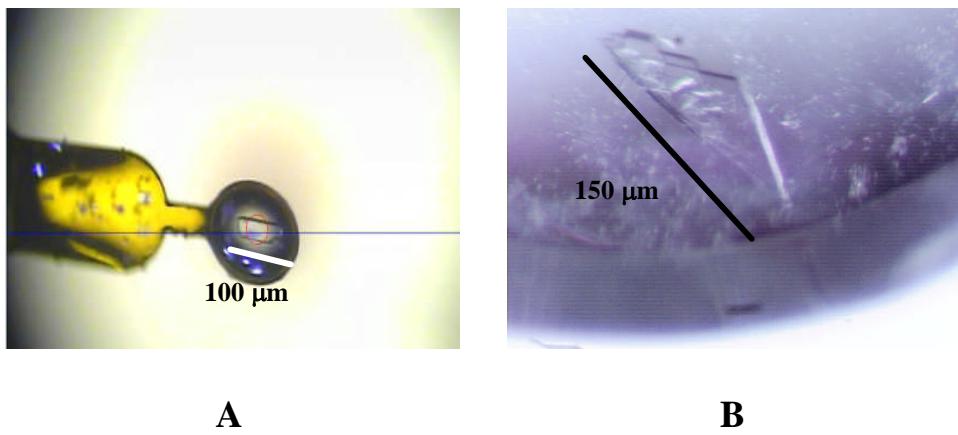


Figure 63. Optimised crystals of AGME Y140F grown in presence of 1 mM ADP- β -mannose under the condition 2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES-Na, pH 7.1, 2 % PEG 400 (A) and 2 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400, 10 mM Spermidine (B).

Cryoprotection of these crystals was performed in mother liquor containing 2 mM ADP- β -mannose and 15 % glycerol before being flash frozen in liquid nitrogen. The “grid screens” around the other initial hits (crystallisation conditions PEGs # 67, # 54 and # 76 and Classics condition # 82), however, only yielded crystals which diffracted poorly.

2.8 Isothermal Titration Calorimetry (ITC) of AGME and AGME Y140F with ADP- β -mannose

2.8.1 Preparation of samples

2 ml of ADP- β -L-glycero-D-mannoheptose 6-epimerase at a concentration of 77 μ M or ADP- β -L-glycero-D-mannoheptose 6-epimerase Y140F at a concentration of 100 μ M were dialysed in Thermo Scientific Slide-A-Lyzer Dialysis Cassettes (7 K MWCO or 10 K MWCO) for about 5 hours in 1 L buffer containing 20 mM Tris-HCl, pH 7.0 and 50 mM NaCl buffer at 4 °C. Then another dialysis step was carried out in exchanged, fresh buffer overnight at 4 °C. ADP- β -mannose was dissolved in the same buffer to a final concentration of 2 mM, which equates to 26 times higher than that of the wildtype and 20 times higher than that of the mutant enzyme. Adjustment of the pHs of the compound and the protein to the exactly same value of that of the dialysis buffer and filtering of all the solutions were performed to decrease the noise from the heat of dilution in the following ITC experiments.

2.8.2 ITC experiments

ITC was performed at 25 °C using the MicroCal™ VP-ITC system. ADP- β -mannose solution was titrated into the enzyme solution. 35 injections each with a volume of 6 μ l were carried out. As a control the same number of titrations into buffer was performed as well. When the titrant is added to the protein solution energy in form of heat is absorbed or released. Absorption of energy occurs when the reaction is endothermic and release of energy takes place in an exothermic reaction (Pierce *et al.*, 1999). In both ITC experiments, determination of entropy (ΔS), enthalpy (ΔH), the binding constant (K_a) and stoichiometry (n) was carried out (Pierce *et al.*, 1999).

Analysis of the raw data was performed with the OriginTM ITC (Version 5.1 – MicrocalTM) software. Subtraction of the blank titrations was carried out to correct this data and the integrated binding isotherm was adjusted to a one site binding model. The isothermal titration calorimetric analysis of the AGME/ADP- β -mannose and AGME Y140F/ADP- β -mannose interactions are shown in Figure 64A and B.

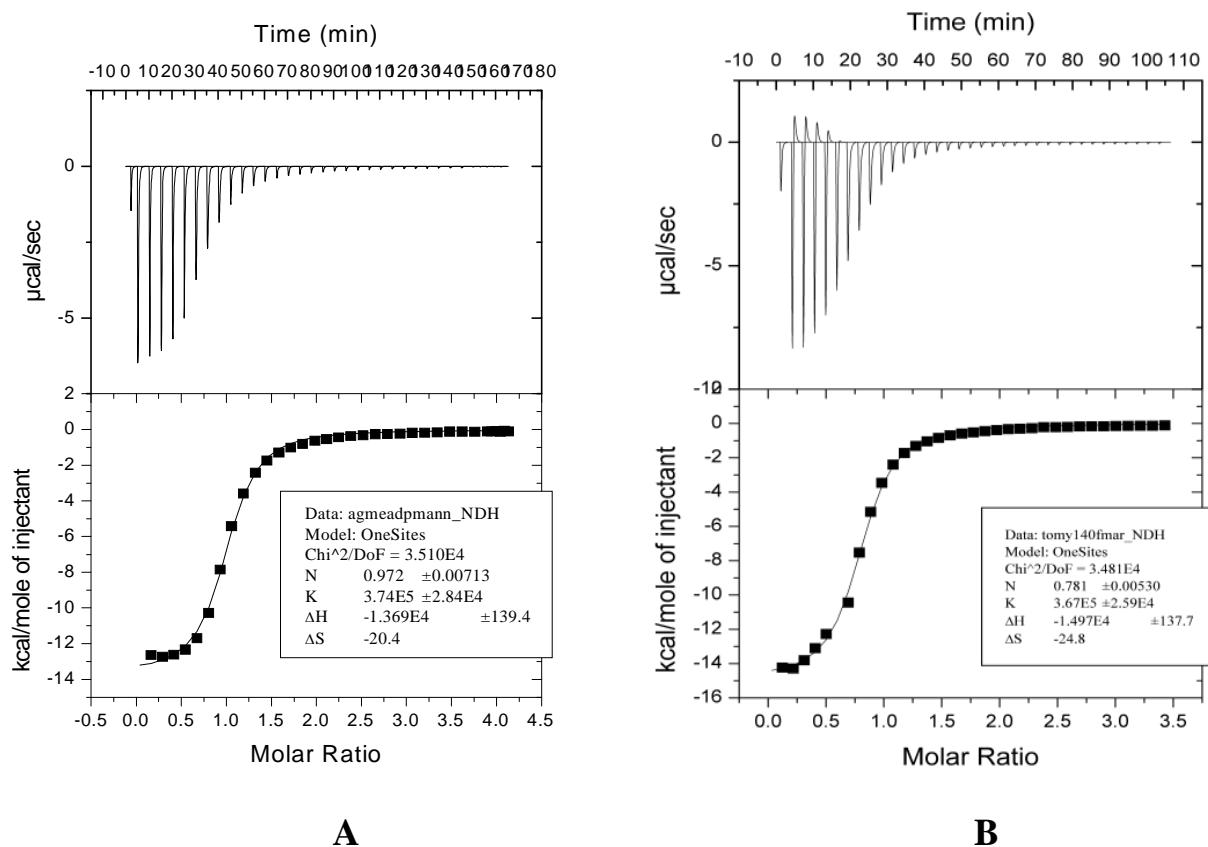


Figure 64. ITC analysis of AGME/ADP- β -mannose (A) and AGME Y140F/ADP- β -mannose (B) interactions. The titration profiles of ADP- β -mannose into AGME and AGME Y140F are shown in the top panels of (A) and (B). The integrated binding isotherms of the AGME/ADP- β -mannose and AGME Y140F/ADP- β -mannose titrations and experimental fits are demonstrated in the bottom panels. Each experiment was performed once.

The dissociation constants ($K_d = 1/K_a$) for ADP- β -mannose are 2.7 μM in both experiments (Table 17) These values show that the nucleotide sugar binds tightly to the active site in both enzymes.

The enthalpy (ΔH) (Table 17) is negative which means that heat was released upon titration of the compound into the protein solution (exothermic reaction). The additional heat absorption (endothermic reaction) during the first four titrations of ADP- β -mannose into AGME Y140F (Figure 64, B, top panel) is hard to explain. It could be possible that there was a slight difference in the pH values of ligand and protein. The stoichiometry (n) is, as expected, close to 1.0 in both cases (Table 17). A reaction stoichiometry of 1.0 means that one molecule of ligand binds to one active site of a protein's subunit. Both values are slightly below 1.0 which means that the concentrations of either the enzymes or the compound have not been accurate.

Table 17. ITC results of AGME and AGME Y140F with ADP- β -mannose

Protein interaction	Stoichiometry	K _d (μM)	ΔH (kCal/mole)
AGME/ADP- β -mannose	0.97	2.7	-1.369
AGME Y140F/ADP- β -mannose	0.78	2.7	-1.497

Discussion

The *hldD* gene from *E. coli* K-12 W3110 was cloned into pEHISTEV. Soluble expression was obtained when *E. coli* BL21 (DE3)pLysS cells were transformed with pEHISTEV-*hldD*. *E. coli* BL21 (DE3)pLysS is normally used to express a gene of interest efficiently in the presence of a T7 promotor. This *E. coli* strain has a plasmid named pLysS. The open reading frame which encodes the T7 lysozyme is located on this plasmid. Any low level expression of the gene of interest controlled by the T7 promotor is inhibited by T7 lysozyme.

Upon addition of IPTG, however, T7 lysozyme does not have any effect on higher expression levels any more (description after http://www.promega.com/catalog/products.aspx?categoryname=productleaf_1005). Expression trials with other *E. coli* strains such as BL21 (DE3), C43 (DE3) or Rosetta (DE3) (see Table 16) were unsuccessful. pEHISTEV is a pET based vector, contains a tobacco etch virus (TEV) protease cleavage site and expresses an aminoterminal hexahistidine tagged protein. This His-tag can be cleaved when TEV protease is added. The hexahistidine sequence at a protein's aminoterminus can facilitate its purification (Liu and Naismith, 2009). Both vectors pEHISTEV and pET-30 Xa/LIC contain such a tag. Unlike the tobacco etch virus protease recognition site in pEHISTEV, a factor Xa cleavage site is present in pET-30 Xa/LIC. Our collaborators were unable to cleave the tag. In our case, the idea was to perform a simple three step purification of AGME cloned into pEHISTEV (two times Ni-NTA column and gelfiltration as the third purification step). First the His-tagged enzyme was loaded onto a Ni-NTA column. Only His-tagged AGME would bind, other *E. coli* proteins would get through the column. After elution of the bound protein with a buffer containing a high concentration of imidazole (300-400 mM) and subsequent dialysis in an imidazole free buffer, TEV protease was added to remove the His-tag from the protein. However, the TEV digest of hexahistidine tagged AGME performed over 2 days under optimal conditions (dialysis in 20 mM Tris-HCl, 300 mM NaCl, 0.5 mM EDTA and 1 mM DTT, pH 7.5 at room temperature) was not successful. The most plausible explanation for that could be a buried cleavage site (Figure 65). Another protein expressed in the same construct was used as a control and was successfully cleaved when incubated with the same batch of TEV protease. Consequently, AGME was not further purified.

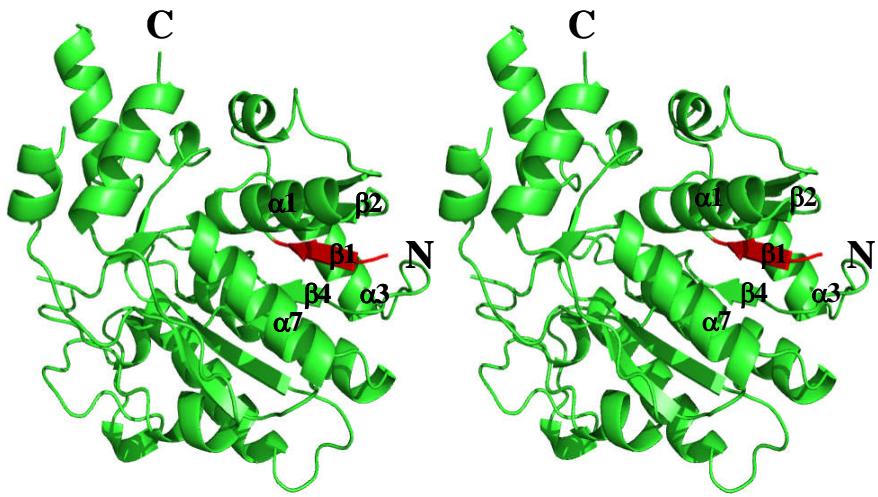


Figure 65. Stereo view of the buried N-terminus of AGME (monomer A is shown) (Pdb accession code 1eq2). The aminoterminal β -sheet (β 1) is shown in red. The other secondary structure elements are shown in green. α -helices and β -strands surrounding β 1 are labelled.

Next, expression of the pET-30 Xa/LIC-HldD construct in *E. coli* BL21 (DE3) was tested as well. The level of soluble AGME expression from pET-30 Xa/LIC was much higher than that expressed from pEHISTEV in *E. coli* BL21 (DE3) pLysS (Figure 53). As a consequence a different purification protocol without any TEV cleavage involved was designed (see Section 2.6). The first attempt to co-crystallise AGME with ADP- β -mannose was a “grid screen” around the already published condition, 2 M ammonium sulfate, 0.1 M HEPES-Na, pH 7.5 and 2 % PEG 400 which yielded high quality crystals of AGME complexed with ADP- α -glucose (Ding *et al.*, 1999; Deacon *et al.*, 2000). However, this screen did not yield any crystals. Reasons for that could be differences in the AGME constructs and the purification protocols. Whereas our AGME sample contains an aminoterminal His-tag, AGME which gave crystals in complex with ADP- α -glucose (Ding *et al.*, 1999; Deacon *et al.*, 2000) was untagged. In our case, AGME was purified using a Ni-NTA affinity column, followed by hydrophobic interaction chromatography and finally gel filtration. Ding and coworkers, however, performed hydrophobic interaction chromatography, affinity chromatography (blue Sepharose CL-6B column) (Ding *et al.*, 1994) and ion exchange chromatography as last step to obtain pure protein (Ding *et al.*, 1999).

Furthermore, the protein sample was desalted against H₂O (Ding *et al.*, 1999) whereas in our case AGME was in 20 mM Tris-HCl, pH 7.0, 50 mM NaCl and 1mM DTT before applied to crystallisation trials (see Section 2.6). As a consequence AGME was screened against a range of commercially available crystallisation screens. Initial crystals were obtained in many conditions. The most promising ones were used to design further “grid screens”. Crystals suitable for reasonable diffraction studies were grown in 0.2 M Ca-acetate, 0.1 M Na-cacodylate, pH 6.5 and 35 % PEG 300. Larger and thicker crystals than in the initial condition (0.2 M Ca-acetate, 0.1 M Na-cacodylate, pH 6.5 and 40 % PEG 300) were obtained. In this case a decrease of the number of nucleation events in the drop could be achieved by a lower concentration of the precipitant, PEG 300. A dataset of one crystal grown from this optimised condition and in the presence of 6.6 mM ADP-β-mannose was collected at IO2 at the Diamond synchrotron, Oxford (see next section). The structure was solved and there was clear density for NADP⁺ in the electron density map, however, no nucleotide sugar was observed in this structure. This was a very surprising result as ITC data of AGME with ADP-β-mannose showed tight binding of the ligand to the protein ($K_d = 2.7 \mu\text{M}$). Additional soaking of AGME co-crystals with 20 mM of compound in a time course from 10 min to 6 hours did not turn out to be successful either. The next approach was to denature AGME and to reconstitute it with NADP⁺. Usually a mixture of NADP⁺ and NADPH is present in AGME after its purification. (Morrison and Tanner, 2007) Therefore a homogeneous cofactor content in the enzyme could support growth of AGME crystals complexed with ADP-β-mannose. Crystals were obtained in the same condition as the unreconstituted enzyme had been crystallised in. The structure of one such crystal was solved, but again, there was no sign of a bound nucleotide sugar. A plausible explanation for this negative result could be that the compound had been turned over during the crystallisation process. However, it is not clear why at least ADP is not bound in the structure, as the adenine has strong van der Waals interactions with the enzyme (Deacon *et al.*, 2000).

An explanation that binding of ADP- β -mannose to AGME was observed in the ITC experiment but not in the crystal structures could be the fact that titration of a ligand into an enzyme solution, as performed in ITC (Pierce *et al.*, 1999), is a very fast process and binding is only detected by absorption or release of heat (Pierce *et al.*, 1999). Crystallisation, however, is a much slower process. Consequently, a substrate has already been turned over by the active enzyme before crystallisation has even taken place. For that reason only an empty active site could then be observed in a crystal structure. To exclude a possible turnover of the nucleotide sugar, co-crystallisation of the AGME mutant Y140F which only displays 0.08 % epimerase activity compared to the wildtype enzyme (Morrison and Tanner, 2007) with ADP- β -mannose was attempted next. ITC was also performed which confirmed a tight interaction of the mutant with the compound, similar to the wildtype enzyme ($K_d = 2.7 \mu\text{M}$). Crystals (Figure 61) were again obtained under the condition 0.2 M Ca-acetate, 0.1 M Na-cacodylate, pH 6.5 and 35 % PEG 300. However, problems appeared during data collection of these crystals. Many spots were overlapped because of the very high mosaic spread (Figure 66).

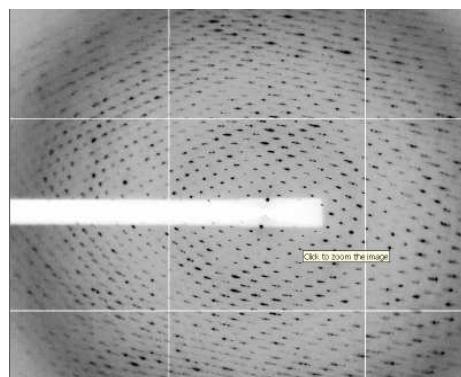


Figure 66. Diffraction pattern of one of the AGME Y140F crystals grown from 0.2 M Ca-acetate, 0.1 M Na-cacodylate, pH 6.5 and 35 % PEG 300.

Collecting data at even 0.1° oscillations did not show any improvement. Capillary mounting of those crystals at room temperature and also de- and rehydration experiments which can help to decrease the mosaicity were not successful. Collecting data at lower resolutions which could also aid to reduce overlaps did not work either. In the worst cases even autoindexing of the data failed. To screen for a new crystallisation condition AGME Y140F was rescreened. Fortunately, optimisation of one of many different conditions obtained proved to be successful. Two of the datasets, one collected in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku), and the other one at BM14 at the ESRF, Grenoble, are discussed in the following section. Interestingly, the initial condition this “grid screen” was based on is exactly the same one from which well diffracting AGME crystals in complex with ADP- α -glucose were obtained (Ding *et al.*, 1999; Deacon *et al.*, 2000) No wildtype protein crystals with ADP- β -mannose could be grown from this condition.

Bioinformatics studies

Sequence alignments between AGME and other short chain dehydrogenases were performed using the programme ClustalW (Chenna *et al.*, 2003). Sequence identities between the aligned proteins were calculated using the sequence alignment programme BLAST (Altschul *et al.*, 1990).

Data collection, Processing, Molecular Replacement of AGME and AGME Y140F and Refinement of AGME Y140F

Summary

One complete X-ray dataset of AGME was collected to 2.9 Å at the Diamond synchrotron, Oxford. Two complete datasets of AGME Y140F co-crystals were collected as well, one in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku) to 2.8 Å, and the other one at BM14 at the ESRF, Grenoble to 2.4 Å. The diffraction data were processed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). The structure of AGME in complex with ADP- α -glucose has already been solved (Deacon *et al.*, 2000). For that reason it was decided to use this structure as a model to calculate a set of initial phase estimates using the technique of Molecular Replacement (MR). After the structures had been solved using PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004) or Molrep (Vagin and Teplyakov, 1997), incorrect amino acid rotamers were manually modified in COOT (Emsley and Cowtan, 2004) and additionally the Tyr at position 140 in the mutant was changed to Phe. Isotropic refinement, TLS parameters and Non crystallographic symmetry (NCS) restraints were carried out in all steps of refinement using the programme REFMAC5 (Murshudov *et al.*, 1999). The final model was checked and validated using MOLPROBITY (Davis *et al.*, 2004).

2.9 Data collection, processing and structure solution of AGME (IO2, Diamond, Oxford)

2.9.1 Data collection and processing

Diffraction data from one AGME crystal to 2.9 Å were collected at -173.15 °C at a wavelength of 1.06 Å on IO2 at the Diamond synchrotron, Oxford. Images in 0.25° oscillations (Figure 67) with a 2s exposure were collected. Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor et al., 2002). The crystal belonged to the spacegroup C2 with unit cell parameters $a = 160 \text{ \AA}$, $b = 115.6 \text{ \AA}$, $c = 218.6 \text{ \AA}$, $\alpha = 90^\circ$, $\beta = 101.9^\circ$, $\gamma = 90^\circ$. The data have an overall completeness of 95.7 %. Data collection and processing statistics are presented in Table 18.

Table 18. Data collection statistics for the AGME crystal.

Data collection	AGME
λ (Å)	1.06
Resolution (Å)	50 – 2.9
Spacegroup	C2
Cell (Å,°)	$a = 160, b = 115.6, c = 218.6$ $\alpha = 90, \beta = 101.9, \gamma = 90$
Unique reflections	80387
Average redundancy	5
I/σ	23
Completeness (%)	95.7
R_{merge}^*	0.043

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

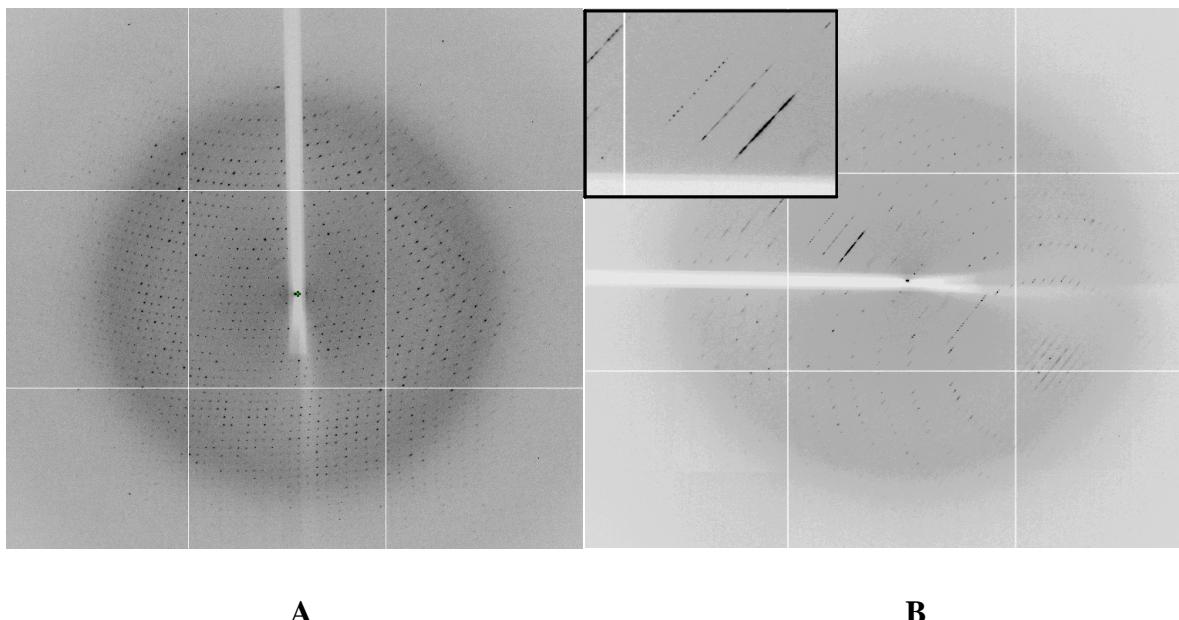


Figure 67. 0.25° oscillation X-ray images of the crystal of AGME at the start of the collection (A) and after about 30° of collected data (B).

2.9.2 Structure solution

The structure of AGME was solved by molecular replacement using PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004). A pentamer of AGME (Pdb accession code 1eq2) was used as a search unit. A solution was found for 2 pentamers in the asymmetric unit consistent with the Matthews coefficient calculation (Table 19) (Matthews, 1968) and a resolution range between 19.9 and 2.9 Å.

Table 19. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.93) = probability at 2.93 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of $2.75 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a 55.26 % solvent content at a probability of 0.16 at 2.93 Å and a total probability of 0.14 (shown in orange), suggested 10 molecules of AGME in the asymmetric unit.

	Nmol/asym	Matthews Coeff.	%solvent	P(2.93)	P(tot)
1	27.47	95.53	0.00	0.00	
2	13.74	91.05	0.00	0.00	
3	9.16	86.58	0.00	0.00	
4	6.87	82.10	0.00	0.00	
5	5.49	77.63	0.00	0.00	
6	4.58	73.16	0.01	0.00	
7	3.92	68.68	0.02	0.01	
8	3.43	64.21	0.04	0.03	
9	3.05	59.73	0.09	0.08	
10	2.75	55.26	0.16	0.14	
11	2.50	50.78	0.22	0.21	
12	2.29	46.31	0.22	0.23	
13	2.11	41.84	0.15	0.18	

Clear rotation and translation solutions were found for each of the two pentamers. The top solution from PHASER is shown in Table 20.

Table 20. The top solution from PHASER. RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes (unfavourable molecular contacts in crystal packing)

Solution	RFZ	TFZ	PAK	RFZ	TFZ	PAK
1	36.9	37.1	0	47.8	78.5	0

Pentamer 1 Pentamer 2

2.10 Data collection, processing, structure solution and refinement of the AGME Y140F complex (in-house)

2.10.1 Data collection and processing

Before data collection crystals were cryoprotected in a solution containing 2 M ammonium sulfate, 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400, 10 mM Spermidine, 2 mM ADP- β -mannose and 15 % glycerol. Diffraction data were collected at -173.15 °C at a wavelength of 1.542 Å in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku). Data from one crystal to 2.8 Å resolution were collected in 0.1° oscillations (Figure 68) with a 60s exposure. Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). The crystal belonged to the space group P2₁ with unit cell parameters a = 138 Å, b = 162 Å, c = 185 Å, α = 90°, β = 101°, γ = 90°.

The data have an overall completeness of 99 % and 100 % in the highest resolution shell from 2.85 – 2.8 Å resolution. Data collection and processing statistics are shown in Table 21.

Table 21. Data collection statistics for the AGME Y140F crystal. Data for the highest resolution shell are shown in brackets.

Data collection	AGME Y140F
λ (Å)	1.542 Å
Resolution	35 – 2.8
Last shell (Å)	(2.85-2.8)
Spacegroup	P2 ₁
Cell (Å, °)	a = 138, b = 162, c = 185 $\alpha = 90, \beta = 101, \gamma = 90$
Unique reflections	197672
Average redundancy	3.6 (3.5)
I/σ	15.3 (1.8)
Completeness (%)	99 (100)
R _{merge} *	0.092 (0.808)

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

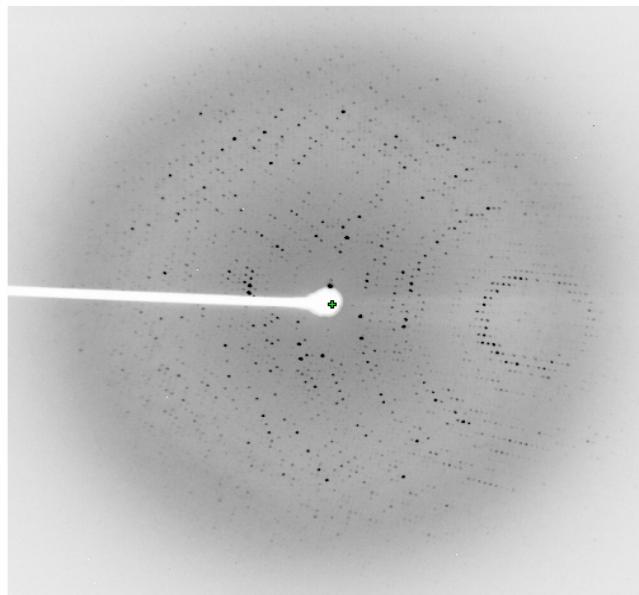


Figure 68. 0.1° oscillation X-ray image of the crystal of AGME Y140F collected in-house to 2.8 Å resolution.

2.10.2 Structure solution

The structure of AGME Y140F was solved by molecular replacement using PHASER (McCoy et al., 2005; Storoni et al., 2004). A pentamer of AGME (Pdb accession code 1eq2) was used as a search unit. A solution was found for 4 pentamers in the asymmetric unit consistent with the Matthews coefficient calculation (Table 22) (Matthews, 1968) and a resolution range between 35 and 2.8 Å.

Table 22. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.80) = probability at 2.8 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of 2.82 Å³ Da⁻¹, corresponding to a 56.48 % solvent content at a probability of 0.06 at 2.8 Å and a total probability of 0.06 (shown in orange), suggested 20 molecules of AGME in the asymmetric unit.

Nmol/asym	Matthews Coeff.	%solvent	P(2.80)	P(tot)
15	3.77	67.36	0.01	0.01
16	3.53	65.19	0.01	0.01
17	3.32	63.01	0.02	0.02
18	3.14	60.84	0.03	0.03
19	2.97	58.66	0.05	0.04
20	2.82	56.48	0.06	0.06
21	2.69	54.31	0.08	0.08
22	2.57	52.13	0.10	0.09
23	2.46	49.96	0.11	0.11
24	2.35	47.78	0.11	0.11
25	2.26	45.61	0.11	0.11
26	2.17	43.43	0.10	0.10
27	2.09	41.25	0.08	0.08

Clear rotation and translation solutions were found for each of the four pentamers (Table 23). Only one unfavourable molecular contact was observed for the final solution in the crystal packing. The Molecular Replacement solution was then applied to the coordinates of the search model to work as the starting model in the refinement described in section 2.10.3.

Table 23. The top solution from PHASER. RFZ = Z-score after fast rotation function, TFZ = Z-score after fast translation function, PAK = number of clashes (unfavourable molecular contacts in crystal packing)

Solution	RFZ	TFZ	PAK	RFZ	TFZ	PAK	RFZ	TFZ	PAK	RFZ	TFZ	PAK
1	41.3	54.7	0	29.1	61.3	0	31.4	83.6	0	24.1	67.0	1
Pentamer 1			Pentamer 2			Pentamer 3			Pentamer 4			

2.10.3 Refinement

REFMAC 5 (Murshudov *et al.*, 1999) was used to refine the structure of AGME Y140F in complex with ADP- β -mannose. Data in the range from 182.6– 2.8 Å were used. Incorrect rotamers and the mutation at position 140 of the enzyme were modified manually in COOT (Emsley and Cowtan, 2004). TLS parameters (Table 24), isotropic B-thermal factors and NCS restraints were employed in all cycles of refinement. Water, cofactor (NADP $^{+}$) and ligand (ADP- β -mannose) molecules were also added in COOT (Emsley and Cowtan, 2004).

Table 24. TLS groups employed throughout the refinement of AGME Y140F. “Residues” 400 and 401 are referred to NADP and ADP- β -mannose, respectively.

TLS A-T chain1
RANGE 'A 1.' 'A 106.' ALL
RANGE 'A 400.' 'A 400.' ALL
TLS A-T chain2
RANGE 'A 107.' 'A 197.' ALL
TLS A-T chain3
RANGE 'A 198.' 'A 307.' ALL
RANGE 'A 401.' 'A 401.' ALL

The R_{work} and R_{free} after the last round of refinement were 24.5 % and 27.3 % respectively (Table 25).

Table 25. Refinement statistics of AGME Y140F. R_{work} and R_{free} in the highest resolution shell are shown in brackets

Refinement	AGME Y140F
R _{work} * %	24.5 (45.9)
R _{free} * %	27.3 (52.1)
Number of atoms (average B value (Å ²))	
protein (overall)	48800 (35.7)
protein (chain A) protein (chain K)	2440 (35.7) 2440 (35.7)
protein (chain B) protein (chain L)	2440 (35.7) 2440 (35.7)
protein (chain C) protein (chain M)	2440 (35.7) 2440 (35.7)
protein (chain D) protein (chain N)	2440 (35.7) 2440 (35.7)
protein (chain E) protein (chain O)	2440 (35.7) 2440 (35.7)
protein (chain F) protein (chain P)	2440 (35.7) 2440 (35.7)
protein (chain G) protein (chain Q)	2440 (35.7) 2440 (35.7)
protein (chain H) protein (chain R)	2440 (35.7) 2440 (35.7)
protein (chain I) protein (chain S)	2440 (35.7) 2440 (35.7)
protein (chain J) protein (chain T)	2440 (35.7) 2440 (35.7)
water (overall)	958 (27.8)
NADP ⁺	960 (36.1)
ADP-β-mannose	722 (37.6)
NCS deviations (rmsd (Å))	
(chain A - chain B) (chain A - chain I) (chain A - chain P)	0.06 0.09 0.07
(chain A - chain C) (chain A - chain J) (chain A - chain Q)	0.08 0.05 0.09
(chain A - chain D) (chain A - chain K) (chain A - chain R)	0.08 0.09 0.06
(chain A - chain E) (chain A - chain L) (chain A - chain S)	0.07 0.11 0.09
(chain A - chain F) (chain A - chain M) (chain A - chain T)	0.05 0.06 0.12
(chain A - chain G) (chain A - chain N)	0.05 0.05
(chain A - chain H) (chain A - chain O)	0.09 0.08
rmsd bonds (Å) / angles (°)	0.009 / 1.191
Ramachandran favoured (%)	96.4

$$* R_{\text{work}} = \sum_{\text{hkl}} | |F_{\text{obs}} - |F_{\text{calc}}| | / \sum_{\text{hkl}} |F_{\text{obs}}|$$

* For R_{free} the same calculation applies, but for only 5-10 % of reflections which are not included in the refinement.

The quality of the structure was judged using MOLPROBITY (Table 26) (Davis *et al.*, 2004).

Table 26. Output from MOLPROBITY (Davis *et al.*, 2007)

All-Atom Contacts	Clashscore, all atoms:	8.03	99th percentile* (N=141, 2.55Å - 3.05Å)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Rotamer outliers	2.08%	Goal: <1%
	Ramachandran outliers*	0.41%	Goal: <0.2%
	Ramachandran favored	96.44%	Goal: >98%
	Cβ deviations >0.25Å	0	Goal: 0
	MolProbity score	1.91	99th percentile* (N=4482, 2.55Å - 3.05Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.00%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

* Ramachandran outliers: Phe168 (chains I, J, K, M, R), Arg174 (chains A-T)

2.11 Data collection, processing, structure solution and refinement of the AGME Y140F complex (BM14, ESRF, Grenoble)

2.11.1 Data collection and processing

Before data collection crystals were cryoprotected in a solution containing 2 M ammonium sulfate, 0.1 M HEPES-Na, pH 7.1, 2 % PEG 400, 2 mM ADP-β-mannose and 15 % glycerol.

Diffraction data were collected at -173.15 °C at a wavelength of 0.954 Å on BM14 at the ESRF. Data from one crystal to 2.4 Å resolution were collected in 0.5° oscillations (Figure 69) with a 10s exposure. Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). The crystal belonged to the space group C2 with unit cell parameters $a = 342$ Å, $b = 61$ Å, $c = 192$ Å, $\alpha = 105.7^\circ$, $\beta = 105.0^\circ$, $\gamma = 105.1^\circ$. The data have an overall completeness of 100 % and 98 % in the highest resolution shell from 2.39 – 2.35 Å resolution. Data collection and processing statistics are shown in Table 27.

Table 27. Data collection statistics for the AGME Y140F crystal. Data for the highest resolution shell are shown in brackets.

Data collection	AGME Y140F
λ (Å)	0.954
Resolution	50 – 2.35
Last shell (Å)	(2.39 -2.35)
Spacegroup	C2
Cell (Å, °)	$a = 342$, $b = 61$, $c = 192$ $\alpha = 105.7$, $\beta = 105.0$, $\gamma = 105.1$
Unique reflections	162824
Average redundancy	1.9 (1.7)
I/ σ	27 (5.2)
Completeness (%)	100 (98)
R_{merge}^*	0.044 (0.192)

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

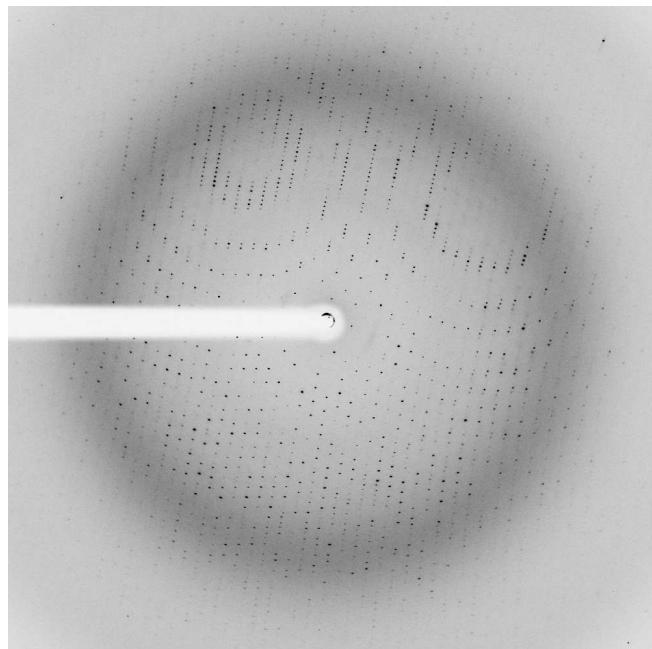


Figure 69. 0.5° oscillation X-ray image of the crystal of AGME Y140F collected to 2.4 \AA resolution at BM14, ESRF.

2.11.2 Structure solution

The structure of AGME Y140F complexed with ADP- β -mannose was solved by molecular replacement using Molrep (Vagin and Teplyakov, 1997). A monomer of AGME (Pdb accession code 1eq2) was used as a search unit. A solution was found for 10 monomers in the asymmetric unit consistent with the Matthews coefficient calculation (Table 28) (Matthews, 1968) and a resolution range between 39 and 2.5 \AA .

Table 28. Matthews coefficient and solvent content calculated from the unit cell and the molecular weight of the molecules in the unit cell. Nmol/asym = number of molecules in the asymmetric unit, %solvent = % solvent content, P(2.36) = probability at 2.36 Å resolution, P(tot) = probability across all resolution ranges. A Matthews coefficient of 2.77 Å³ Da⁻¹, corresponding to a 55.59 % solvent content at a probability of 0.13 at 2.36 Å and a total probability of 0.13 (shown in orange), suggested 10 molecules of AGME Y140F in the asymmetric unit.

	Nmol/asym	Matthews Coeff.	%solvent	P(2.36)	P(tot)
1	27.68	95.56	0.00	0.00	
2	13.84	91.12	0.00	0.00	
3	9.23	86.68	0.00	0.00	
4	6.92	82.24	0.00	0.00	
5	5.54	77.80	0.00	0.00	
6	4.61	73.35	0.00	0.00	
7	3.95	68.91	0.01	0.01	
8	3.46	64.47	0.03	0.03	
9	3.08	60.03	0.07	0.07	
10	2.77	55.59	0.13	0.13	
11	2.52	51.15	0.20	0.20	
12	2.31	46.71	0.24	0.23	
13	2.13	42.27	0.19	0.19	

Clear rotation and translation solutions were found for each of the 10 monomers (Table 29). No unfavourable molecular contacts were observed for the final solution in the crystal packing. The Molecular Replacement solution was then applied to the coordinates of the search model to work as the starting model in the refinement described in section 2.11.3.

Table 29. The top solutions from Molrep.

Rf/sig = Z-score after fast rotation function, Tf/sig = Z-score after fast translation function

Rf/sig	Tf/sig								
11	5.8	8.5	8.7	9.2	11	9.1	14.5	9.8	17.7

Monomer 1 Monomer 2 Monomer 3 Monomer 4 Monomer 5

Rf/sig	Tf/sig								
8.6	17.5	9.5	36.2	8.4	30.6	6.8	40.3	7.3	56.8

Monomer 6 Monomer 7 Monomer 8 Monomer 9 Monomer 10

2.11.3 Refinement

REFMAC 5 (Murshudov *et al.*, 1999) was used to refine the structure of AGME Y140F in complex with ADP- β -mannose. Data in the range from 38.7 – 2.36 Å were used. Incorrect rotamers and the mutation at position 140 of the enzyme were modified manually in COOT (Emsley and Cowtan, 2004). TLS parameters (Table 30), isotropic B-thermal factors and NCS restraints were employed in all cycles of refinement. Water, cofactor (NADP^+) and ligand (ADP- β -mannose) molecules were also added in COOT (Emsley and Cowtan, 2004).

Table 30. TLS groups employed throughout the refinement of AGME Y140F. “Residues” 400 and 401 are referred to NADP and ADP- β -mannose, respectively.

TLS	A-J chain1
RANGE	'A 1.' 'A 106.' ALL
RANGE	'A 400.' 'A 400.' ALL
TLS	A-J chain2
RANGE	'A 107.' 'A 197.' ALL
TLS	A-J chain3
RANGE	'A 198.' 'A 307.' ALL
RANGE	'A 401.' 'A 401.' ALL

One molecule of glycerol was also included in the final model in each of the 10 subunits. An additional chloride ion was also fitted in. The R_{work} and R_{free} after the last round of refinement were 18.9 % and 20.9 %, respectively (Table 31).

Table 31. Refinement statistics of AGME Y140F. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Refinement	AGME Y140F			
R _{work} * %	18.9 (23.4)			
R _{free} * %	20.9 (27.1)			
Number of atoms (average B value (Å ²))				
protein (overall)	24411 (33)			
protein (chain A)	2440 (33)			
protein (chain B)	2440 (33)			
protein (chain C)	2440 (33)			
protein (chain D)	2440 (33)			
protein (chain E)	2440 (33)			
protein (chain F)	2440 (33)			
protein (chain G)	2451(33)			
protein (chain H)	2440 (33)			
protein (chain I)	2440 (33)			
protein (chain J)	2440 (33)			
water (overall)	1549 (40)			
NADP ⁺ glycerol	480 (34.8)	60 (51.3)		
ADP-β-mannose chloride	380 (36.3)	1 (44.4)		
NCS deviations (rmsd (Å))				
(chain A - chain B) (chain A - chain I)	0.06 0.05			
(chain A - chain C) (chain A - chain J)	0.10 0.08			
(chain A - chain D)	0.07			
(chain A - chain E)	0.09			
(chain A - chain F)	0.05			
(chain A - chain G)	0.05			
(chain A - chain H)	0.07			
rmsd bonds (Å) / angles (°)				
0.005 / 0.956				
Ramachandran favoured (%)				
97.5				

$$* R_{\text{work}} = \sum_{\text{hkl}} | | F_{\text{obs}} - | | F_{\text{calc}} | | / \sum_{\text{hkl}} | | F_{\text{obs}} | |$$

* For R_{free} the same calculation applies, but for only 5-10 % of reflections which are not included in the refinement.

The quality of the structure was judged using MOLPROBITY (Table 32) (Davis *et al.*, 2004).

Table 32. Output from MOLPROBITY (Davis et al., 2007)

All-Atom Contacts	Clashscore, all atoms:	10.5	93rd percentile* (N=336, 2.11Å - 2.61Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Rotamer outliers	1.10%	Goal: <1%
	Ramachandran outliers	0.66%	Goal: <0.2%
	Ramachandran favored	97.48%	Goal: >98%
	Cβ deviations >0.25Å	0	Goal: 0
	MolProbity score	1.68	98th percentile* (N=8043, 2.11Å - 2.61Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.00%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

* Ramachandran outliers: Phe168 and Arg174 (chains A-J)

Discussion

AGME (IO2, Diamond, Oxford)

A complete data set from a crystal of AGME was collected to a resolution of 2.9 Å at the Diamond beamline IO2. After an appropriate starting orientation had been given enough data were collected to get a completeness of 95.7 % for the space group C2. The presence of lots of overlapping reflections (Figure 67B) caused problems in indexing the data. However, structure solution was straightforward.

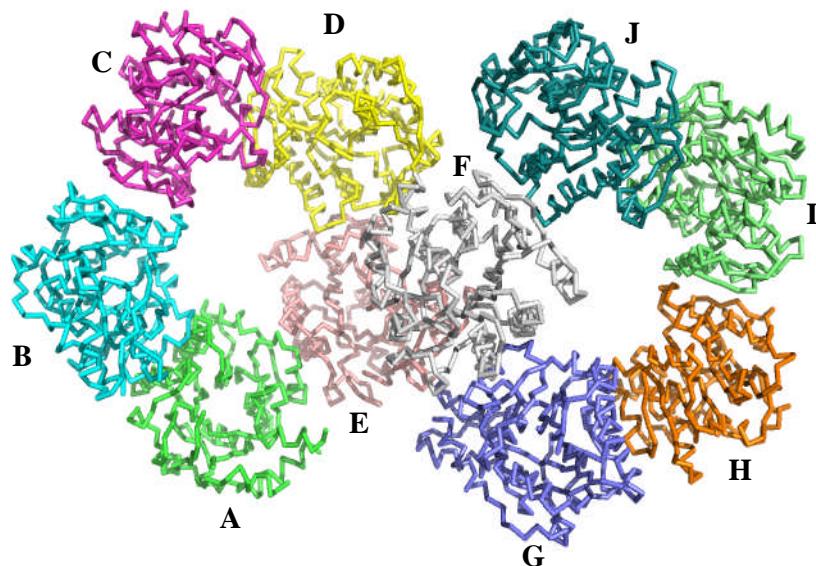


Figure 70. The asymmetric unit of AGME showing two pentamers. $\text{C}\alpha$ traces of the ten monomers (A-J) are shown in different colours.

The R_{work} and R_{free} after only one run of restrained refinement in REFMAC5 were 25.7 % and 29.7 %, respectively. These values represent a reasonably good agreement between F_{obs} and F_{calc} at a resolution of 2.9 Å, and are strong evidence for a correct solution.

In the electron density maps clear densities for NADP⁺ could be observed in each of the 10 monomers. However, there was no sign of any bound nucleotide sugar. After the cofactor molecules had been fitted into the appropriate electron densities and an additional run of refinement had been carried out, R_{work} and R_{free} decreased to 24.4 % and 28.4 %, respectively. The structure was not further refined because of the missing ADP-β-mannose in each of the monomers. A plausible explanation for that could be that the compound was turned over by the active enzyme during crystallisation. Another reason could be that the quality of the data was not very good. Data quality assessment by phenix.xtriage (Zwart *et al.*, 2005) revealed possible twinning of the data. However, the good R_{factor} values and also the clear densities of the tightly bound coenzyme in all ten monomers (Figure 71) support the former assumption.

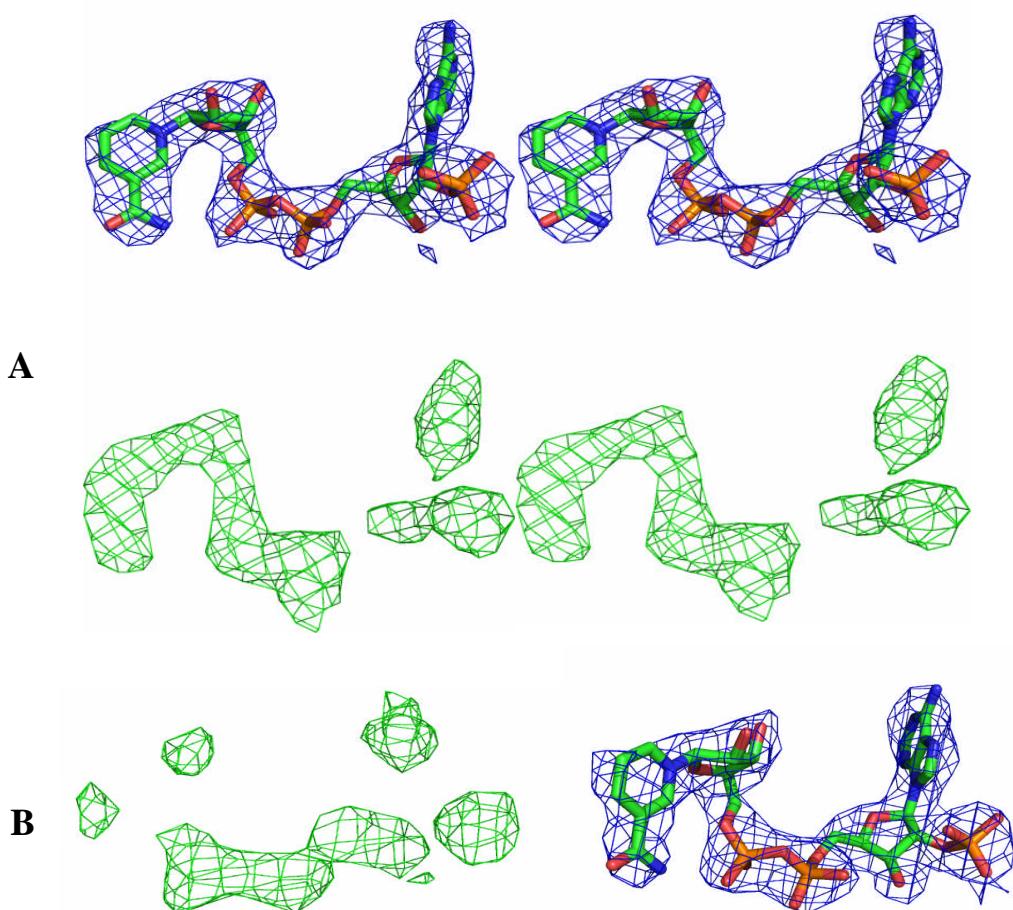


Figure 71. (continues on next page)

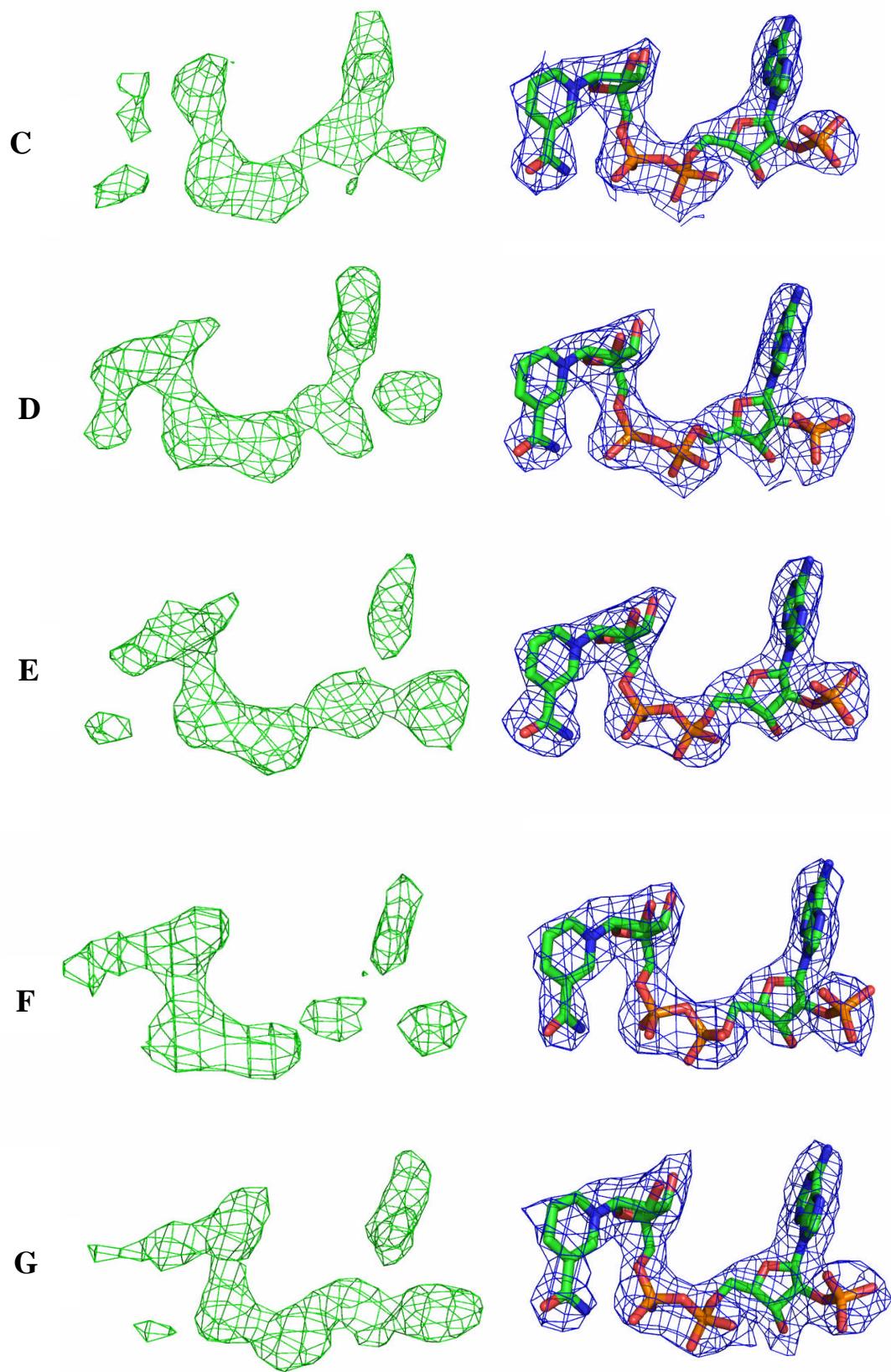


Figure 71. (continues on next page)

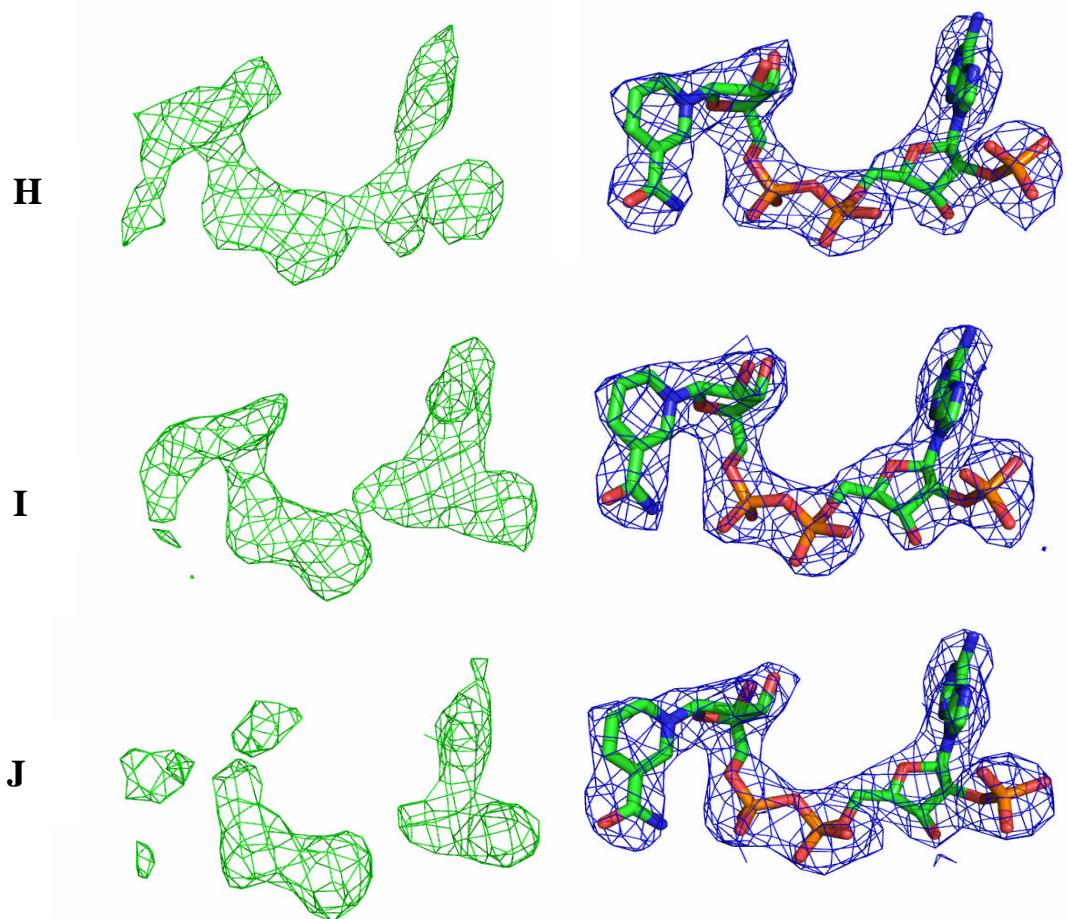


Figure 71. Unbiased $F_o - F_c$ at 2.5σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around the NADP^+ of monomers A-J of AGME. The NADP^+ -molecule is modelled in the $2F_o - F_c$ density. Both electron densities for the cofactor are shown in stereo view in monomer A. Both maps were calculated after one run of restrained refinement. The weak electron density at the the nicotinamide part of NADP^+ ($F_o - F_c$ maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

AGME Y140F (in-house and at BM14, ESRF)

Data of crystals of AGME Y140F complexed with ADP- β -mannose were collected to resolutions of 2.4 and 2.8 Å at BM14, ESRF, and in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku), respectively. The crystal which diffracted to 2.4 Å resolution belonged to the space group C2, whereas the crystal which diffracted to lower resolution belonged to the space group P2₁. AGME in complex with ADP- α -glucose also crystallised in space group P2₁. This is not surprising, as the crystallisation condition is the same, except variations in the concentration of the additive spermidine. However, the unit cell parameters were different. (Ding *et al.*, 1999; Deacon *et al.*, 2000). Structure solution (Tables 23 and 29) and refinement of both structures were straightforward. The clashscores and the MOLPROBITY scores (Tables 26 and 32) indicate the good quality of both structures.

2.12 Sequence alignment between AGME and closely related short chain dehydrogenases/reductases

An alignment between AGME and the structurally most closely related short chain dehydrogenases/reductases UGE, DGD and GMER (Deacon *et al.*, 2000) is shown in Figure 72. AGME shares 24% and 22% sequence identity to *E. coli* and human UGE, respectively. There are 21% identical amino acids between AGME and GMER, and 22% identical residues between AGME and the classical SDR enzyme 3 α ,20 β -hydroxysteroid dehydrogenase. These numbers agree very well with the identities published on pair-wise alignments of various short-chain dehydrogenases/reductases which usually show 15-30 % identity (Oppermann *et al.*, 2003; Persson *et al.*, 2003; Kallberg *et al.*, 2002; Jörnvall *et al.*, 1995). There is higher sequence identity between AGME and dTDP-D-glucose 4,6 dehydratase though (36%).

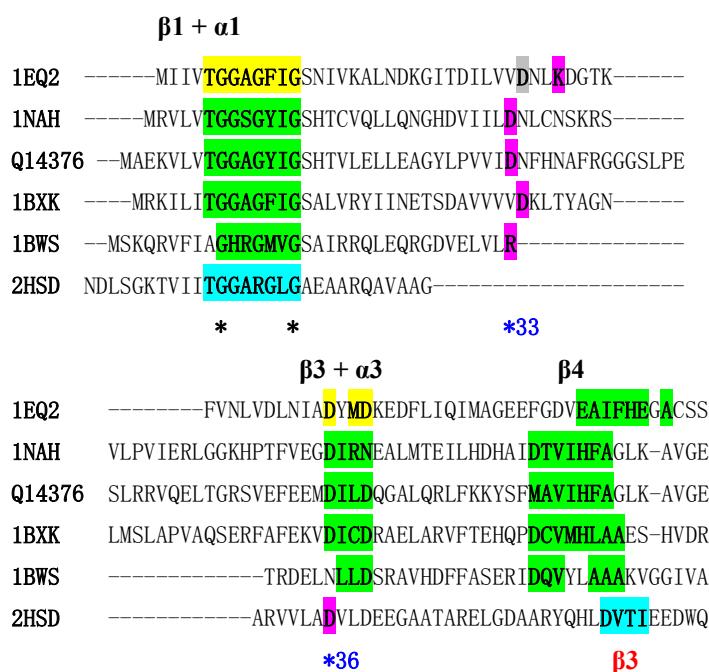


Figure 72. (continues on next page)

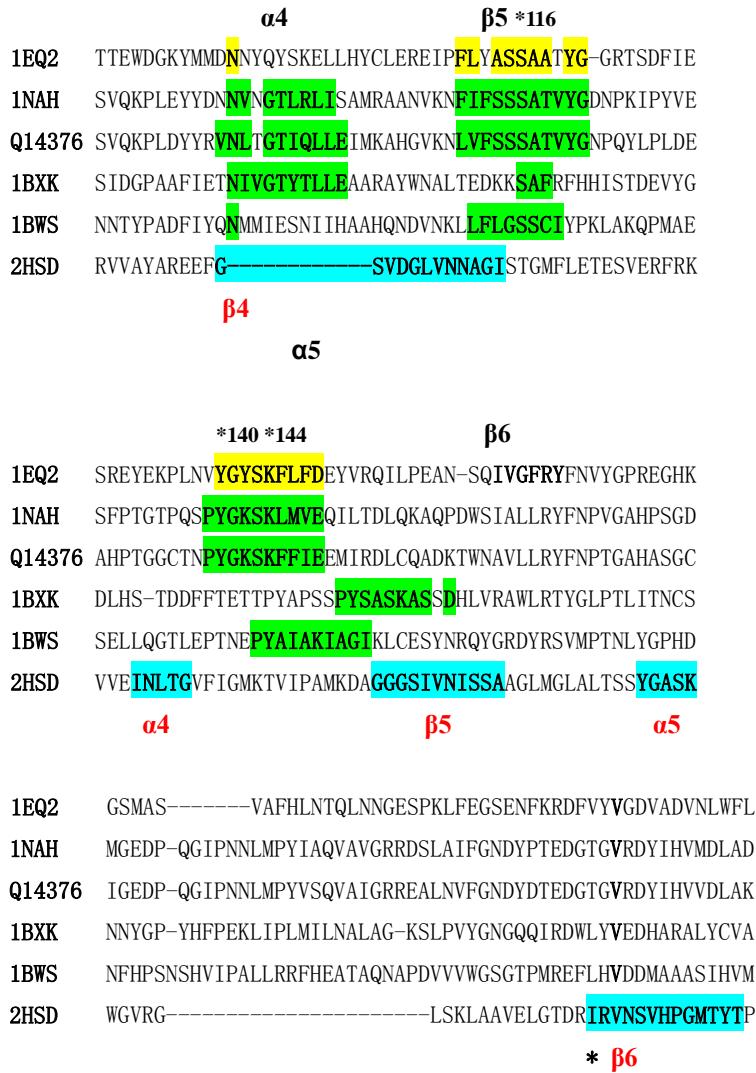


Figure 72. Structural sequence alignment of AGME from *E. coli* K-12 (Pdb accession code 1eq2) with the structurally most related enzymes, UDP-galactose 4-epimerase from *E. coli* (Pdb accession code 1nah) and *Homo sapiens* (UniProtKB/TrEMBL entry Q14376), dTDP-D-glucose 4,6 dehydratase from *E. coli* (Pdb accession code 1bxk) and GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from *E. coli* (Deacon *et al.*, 2000) (Pdb accession code 1bws). 3 α ,20 β -hydroxysteroid dehydrogenase (Pdb accession code 2hsd) as a member of the classical SDR family is also shown as comparison to the five enzymes (Kallberg *et al.*, 2002). Conserved sequence motifs (see Table 10) in AGME are highlighted in yellow, the corresponding motifs in the two UDP-galactose 4-epimerases, in dTDP-D-glucose 4,6 dehydratase and in GDP-4-keto-6-deoxy-D-mannose epimerase/reductase are highlighted in green. Motifs in 3 α ,20 β -hydroxysteroid dehydrogenase are highlighted in turquoise. Residues conserved in all six SDR enzymes are marked with an asterisk. The positions of secondary structural elements in the five extended SDR enzymes are marked on top of each motif, positions of α -helices and β -strands, as they occur in 3 α ,20 β -hydroxysteroid dehydrogenase are shown in red at the bottom of each motif. The crucial residues (Ser116, Tyr140 and Lys144) in the catalytic region of AGME (Deacon *et al.*, 2000) are labelled as well. Residues at “key positions” based on coenzyme specificity are highlighted in magenta. The Asp residue at the end of β 2 in AGME (Kallberg *et al.*, 2002) is highlighted in grey. Numbers (*36) refer to 3 α ,20 β -hydroxysteroid dehydrogenase and human UDP-galactose 4-epimerase(*33) (Persson *et al.*, 2003). This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

The sequence alignment demonstrated above shows that AGME is a member of the extended family of short chain dehydrogenases/reductases (Kallberg *et al.*, 2002). It is 310 residues long (Deacon *et al.*, 2000) and contains the fingerprint sequence Thr-Gly-X-X-Gly-a-h-Gly (Table 10) which is typical for members of the extended family of short-chain dehydrogenases/reductases (Kallberg *et al.*, 2002). The motif Asp-h-x-Asp in β 3 (Table 10) (Persson *et al.*, 2003), which is said to be involved in the association of the adenine base of the NADP (Kallberg *et al.*, 2002 and references therein) is also very conserved in AGME, the only difference is that instead of a hydrophobic amino acid after the first Asp, one can find a polar, aromatic residue (Tyr) in AGME (Figure 72). The cluster Asp/Glu-X-h-h-His-X-Ala-Ala (Table 10) which comprises β 4 in the extended SDR family (Persson *et al.*, 2003), also displays high conservation in ADP- β -L-glycero-D-mannoheptose6-epimerase (Figure 72). The motif h-Asn-h-h-Gly-Thr-X-X-h-h-c (Table 10) (Kallberg *et al.*, 2002), which is referred to be part of the catalytic region in extended SDRs (Kallberg *et al.*, 2002 and references therein) does not appear in AGME. However the Asn, which has been reported to create the tetrad Asn-Ser-Tyr-Lys in the majority of short chain dehydrogenases/reductases (Filling *et al.*, 2002; Oppermann *et al.*, 2003), is conserved (Figure 72) and one can suggest that apparently there is a catalytic tetrad in AGME as well. The conserved motif h-h-h-X-Ser-Ser-X-X-h-a-Gly (Table 10) of β 5 of members of the extended SDR-family (Persson *et al.*, 2003) is also present in AGME. Only the hydrophobic residue at position 3 in this cluster is replaced by an aromatic, polar amino acid (Tyr), and another polar residue (Thr) replaces the hydrophobic residue at position 8 in this motif (Figure 72). AGME also contains the highly conserved sequence arrangement Pro-Tyr-X-X-Ala/Ser-Lys-X-X-h-Asp/Glu of α 5 (Table 10), typical for members of the extended family of SDRs (Persson *et al.*, 2003) with the exception that the first residue (Pro) is missing (Figure 72). The motif h-Lys/Arg/X-X-Asn-Gly-Pro of β 6 (Table 10) (Kallberg *et al.*, 2002) is absent in each of the five extended SDRs (Figure 72).

Unsurprisingly, because of its high degree of homology to human UGE, UGE from *E. coli* belongs to the same subfamily of extended SDRs (eD1) (Kallberg *et al.*, 2002) (Table 11). The sequence alignment shown in Figure 72 also identifies bacterial, such as human dTDP-D-glucose 4,6 dehydratase (Persson *et al.*, 2003), as a member of the eD1 subfamily of extended short chain dehydrogenases/reductases (Kallberg *et al.*, 2002; Persson *et al.*, 2003) (Table 11). It contains an acidic residue (Asp) right after the end of β 2 and is NAD(H)-dependent (Kallberg *et al.*, 2002; Gross *et al.*, 2001).

AGME represents a special case. The last amino acid of β 2 is an aspartic acid (Figure 72) (Kallberg *et al.*, 2002) which is specific for SDRs interacting with NAD(H) (Kallberg *et al.*, 2002). However, AGME is an NADP(H)-specific member (Deacon *et al.*, 2000; Ni *et al.*, 2001; Kallberg *et al.*, 2002) of the extended SDR-family (Kallberg *et al.*, 2002). However, it can be seen from the X-ray structure of AGME in complex with ADP- α -glucose (Deacon *et al.*, 2000) that the aspartic acid occurs, unlike in other enzymes which are NAD(H) specific, in a “more open conformation” (Kallberg *et al.*, 2002) and for that reason there is more space to allow nicotinamide adenine dinucleotide phosphate to bind (Kallberg *et al.*, 2002; Deacon *et al.*, 2000; Ni *et al.*, 2001). Apart from that there is also a basic residue, a lysine (Figure 72) in the vicinity of β 2 (in the short three amino acids long loop between β 2 and α 2) which would classify it as a member of the cP1 subfamily of extended SDRs (Table 11) (Persson *et al.*, 2003). However, this residue occurs at key position 37 of UGE from *Homo sapiens* instead of 34 (Kallberg *et al.*, 2002) and for that reason a strict classification of this epimerase is not possible. GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from *E. coli* is similar, it is also a NADP(H)-binding member (Rizzi *et al.*, 1998) of extended SDRs (Persson *et al.*, 2003) and has a basic residue (Arg) (Figure 72) in the proximity of the second β -sheet of its structure, but this amino acid is located at key position 33 (Table 11) (Kallberg *et al.*, 2002) of UGE. Human GDP-4-keto-6-deoxy-D-mannose epimerase/reductase has not been classified into one of the subfamilies either (Persson *et al.*, 2003).

Structures of AGME Y140F in complex with ADP- β -mannose

Summary

This section describes the two solved structures of AGME Y140F complexed with ADP- β -mannose. First an analysis of both overall structures, bound cofactors and ligands is performed. Then a more detailed description of the active site is carried out based on the higher resolution structure. Differences and common features to the already solved wildtype structure in complex with ADP- α -glucose are also discussed. Finally it is attempted to elucidate the exact roles of Tyr140 and Lys178 in the epimerisation process of this enzyme as the nucleotide sugar in the structure is not disordered and only one bound conformation is observed.

2.13 Overall structures

Unsurprisingly, because of their essentially identical sequence, the mutant protein shares the same fold as wildtype AGME. The asymmetric unit of the 2.4 Å resolution AGME Y140F structure is composed of 10 monomers (Figure 73A), the one of the 2.8 Å resolution mutant structure consists of 20 subunits. (Figure 73B)

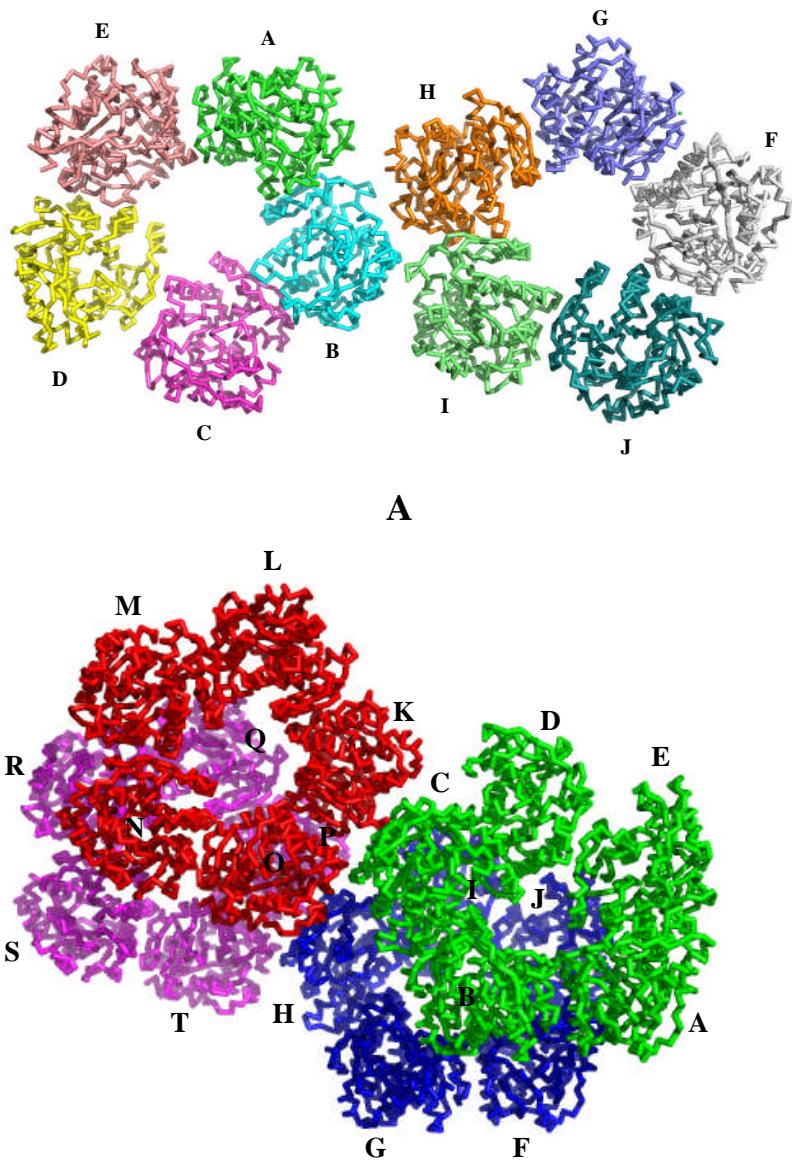


Figure 73. (A) C_α traces of the monomers (A-J) forming two pentamers in the 2.4 Å structure of AGME Y140F. (B) C_α traces of the monomers (A-T) forming four pentamers (green, blue, red and magenta) in the 2.8 Å structure of AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

In the 2.4 Å resolution structure the ten monomers are essentially identical. Superposition of monomers B-J to A (Figure 74A) shows an average Rmsd of 0.06 Å (see Table 31). In the 2.8 Å resolution structure the Rmsd values are also very low.

If one superimposes subunits B-T to A (Figure 74B), the average Rmsd is 0.07 Å (see Table 25) In the structure of AGME complexed to ADP- α -glucose (Deacon *et al.*, 2000) the Rmsd values from superposition of the same nine monomers to A are slightly higher. They lie in a range between 0.12 Å and 0.21 Å.

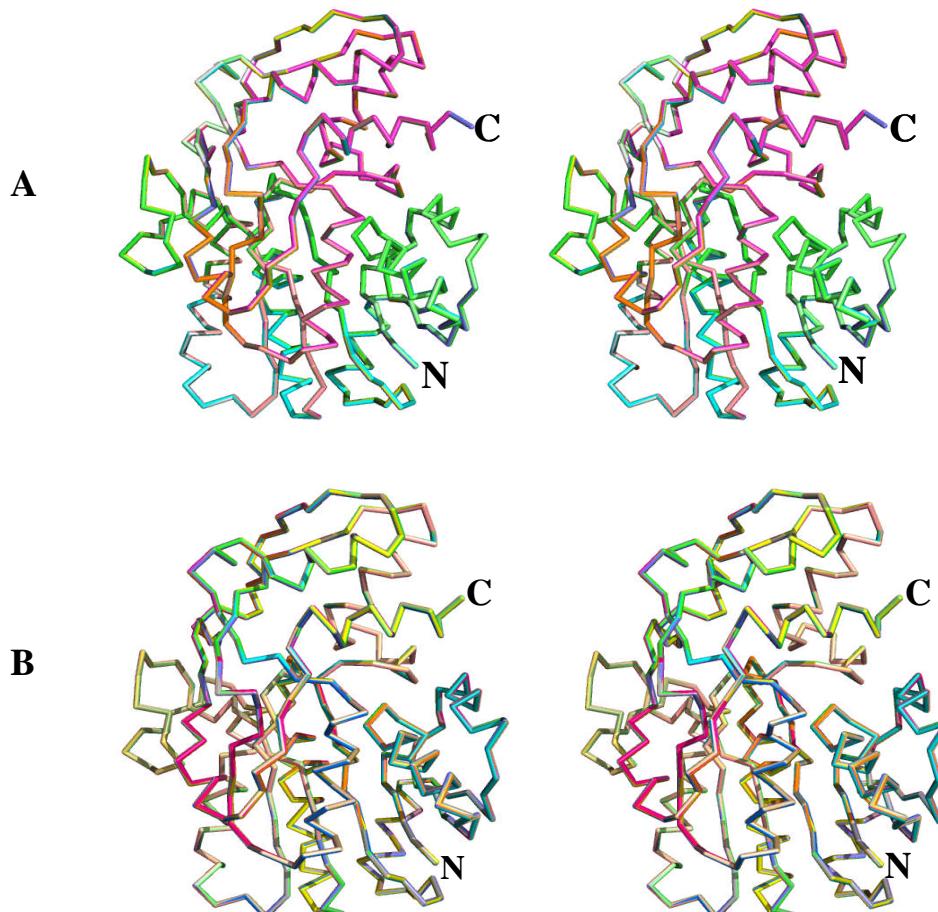


Figure 74. Stereo view of the Ca tracing of the ten monomers of the 2.4 Å (A) and the twenty monomers of the 2.8 Å (B)

AGME Y140F structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

The ten subunits of the higher resolution AGME Y140F structure essentially contain all amino acids. In one monomer only two residues are missing at the carboxyterminus (amino acids 309 and 310), whereas the other nine subunits lack three residues (amino acids 308-310).

No disordered regions have been observed in this structure unlike in Deacon and coworkers' structure in which most of the monomers lack additional residues (amino acids 265-271 are missing in three monomers, one subunit lacks residues 262-271, residues 195-207 and 251-271 are completely disordered in another monomer, and one subunit lacks amino acids 194-206 and 249-271) (Deacon *et al.*, 2000). Most of the twenty monomers of the lower resolution mutant structure have several amino acids such as Arg, Lys and Glu with disordered side chains. In general there is sufficient electron density for the main chain in those regions. Only in two chains several parts of the carboxyterminal main chain (amino acids 194-196, 199-201, 205, 251, 253-255, 257, 262 and 271 in chain G, and residues 194-195, 201, 204, 206, 254, 256-257, 262 and 272 in chain O) are disordered. In all twenty subunits the last three amino acids (308-310) are missing as well.

2.14 Active site of AGME Y140F

2.14.1 NADP⁺ and ADP-β-mannose molecules in both mutant structures

There is clear electron density for one molecule of cofactor and one of nucleotide sugar in each of the ten monomers of the 2.4 Å structure (Figures 75 and 76).

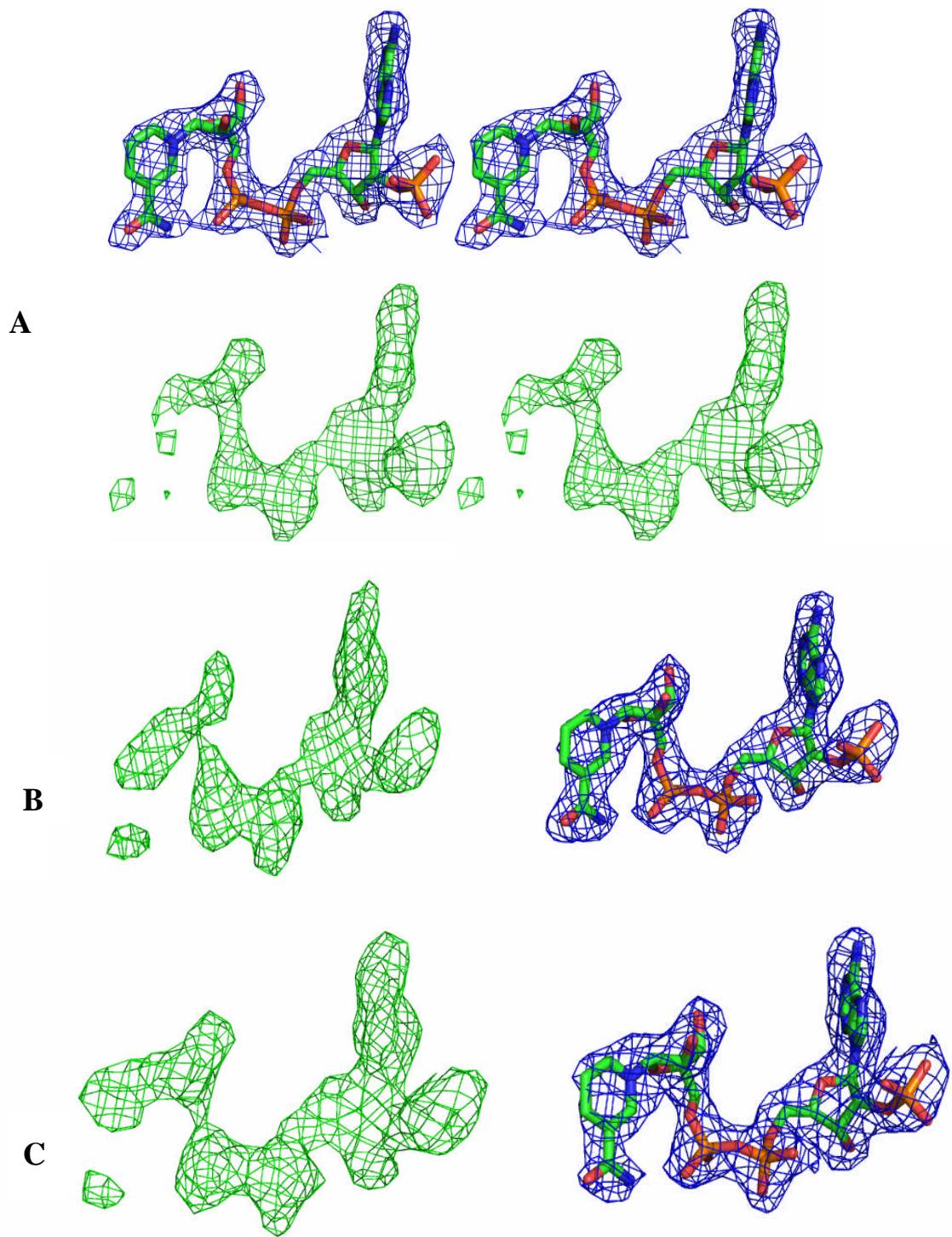


Figure 75. (continues on next page)

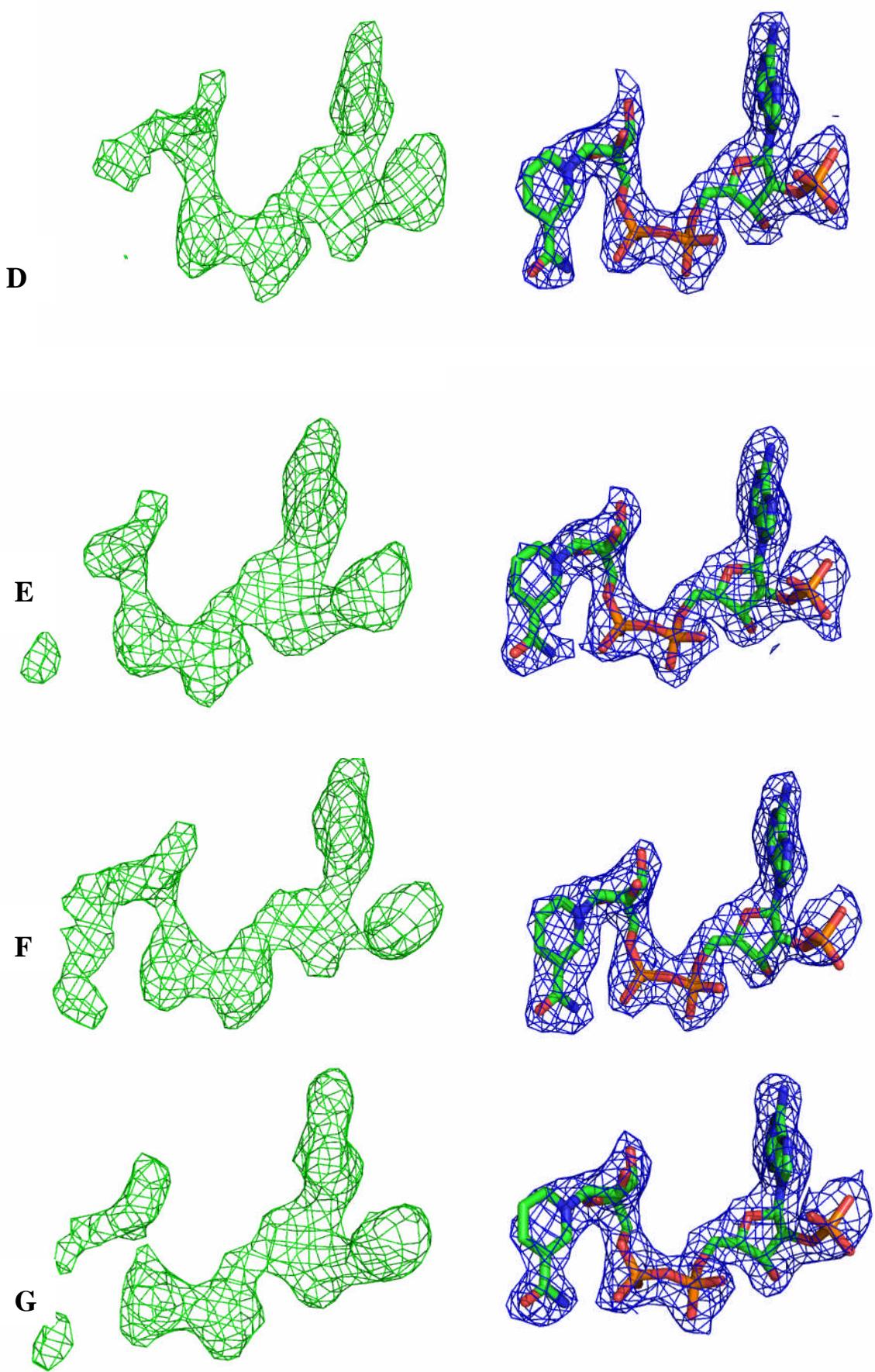


Figure 75. (continues on next page)

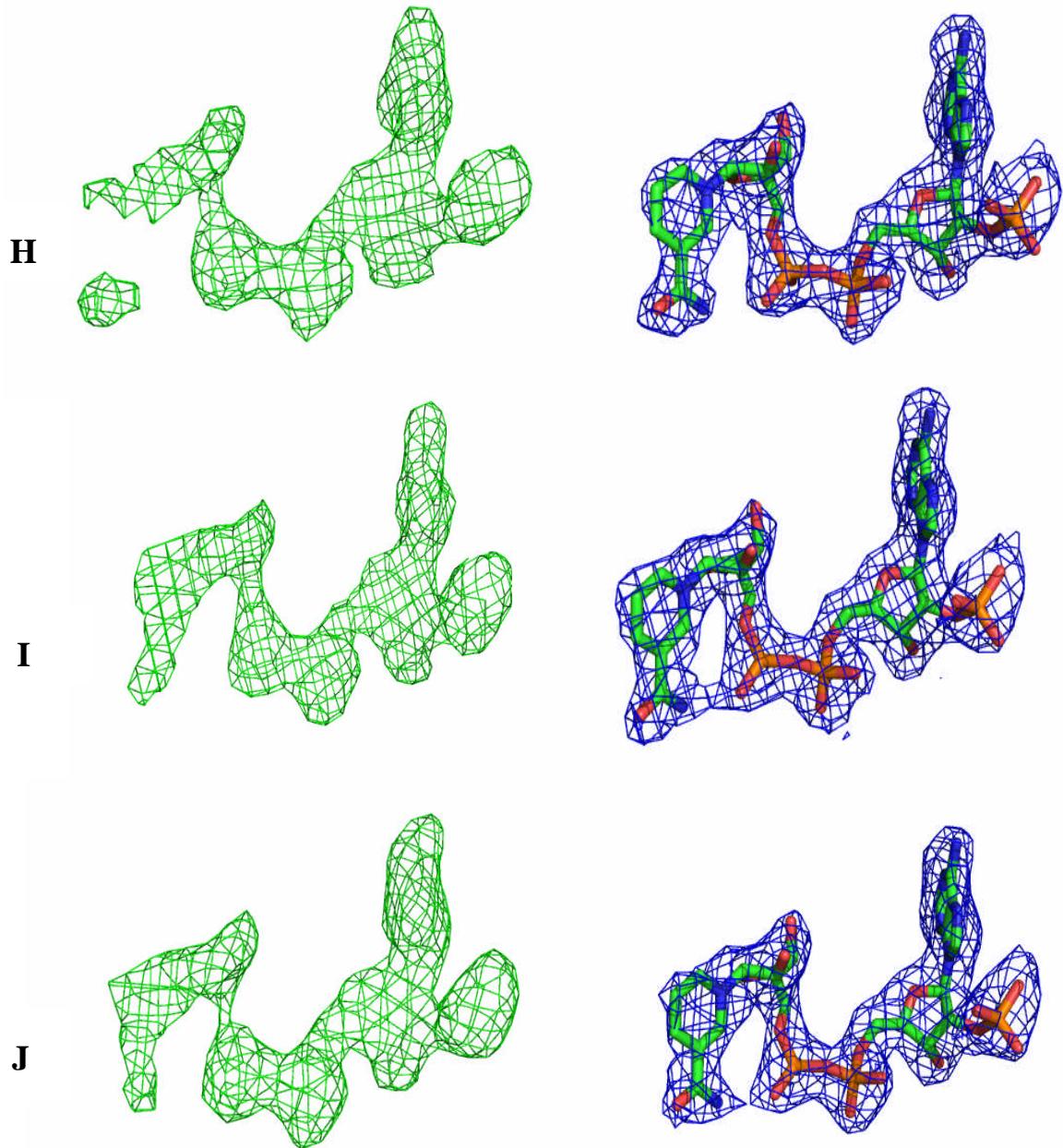


Figure 75. Unbiased F_o-F_c at 3σ (green) and final $2F_o-F_c$ at 1σ (blue) maps around the NADP^+ of monomers A-J of the 2.4 Å structure of AGME Y140F. The NADP^+ -molecule is modelled in the $2F_o-F_c$ density. Both electron densities for the cofactor of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. The weak electron density at the the nicotinamide part of NADP^+ (F_o-F_c maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

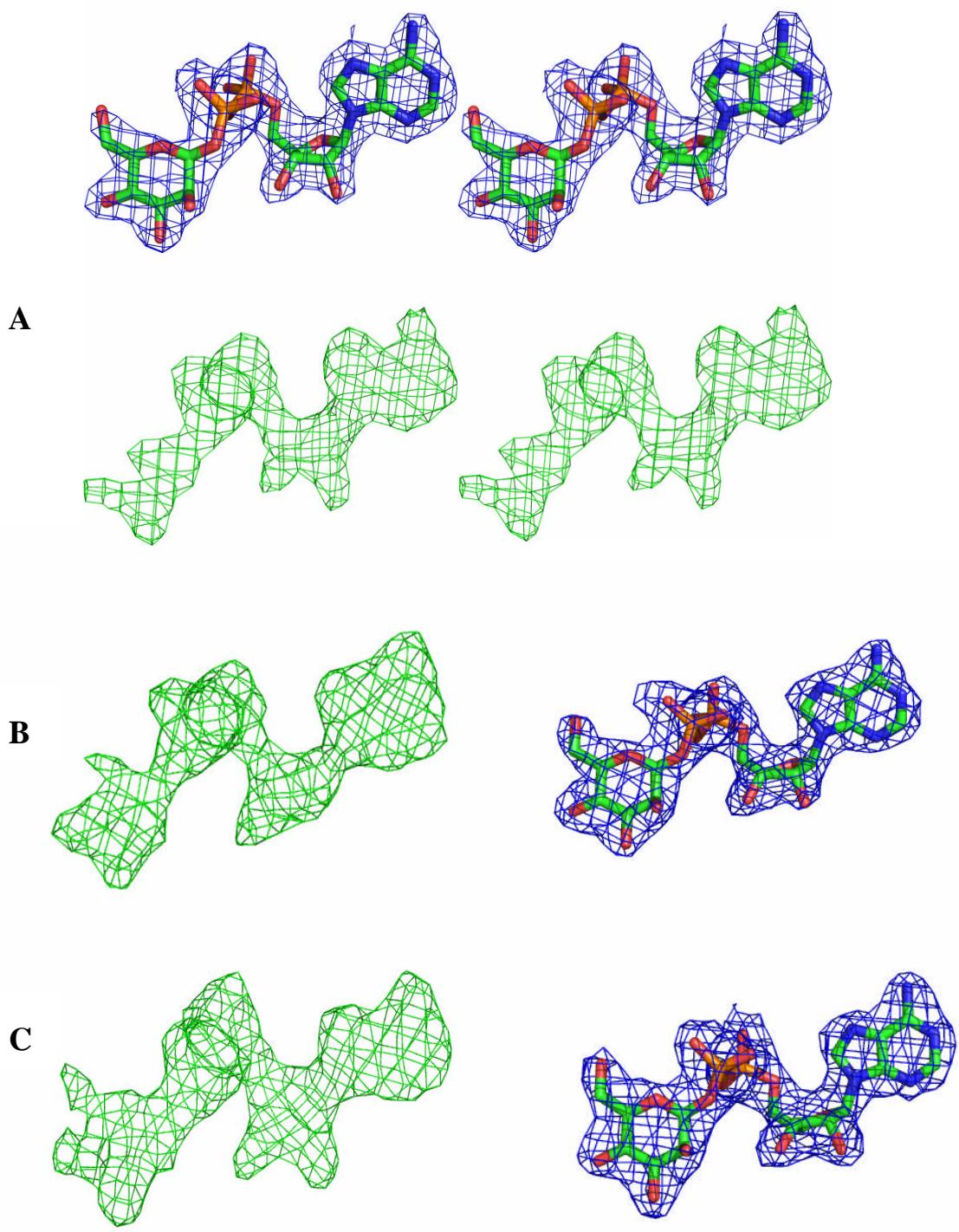


Figure 76. (continues on next page)

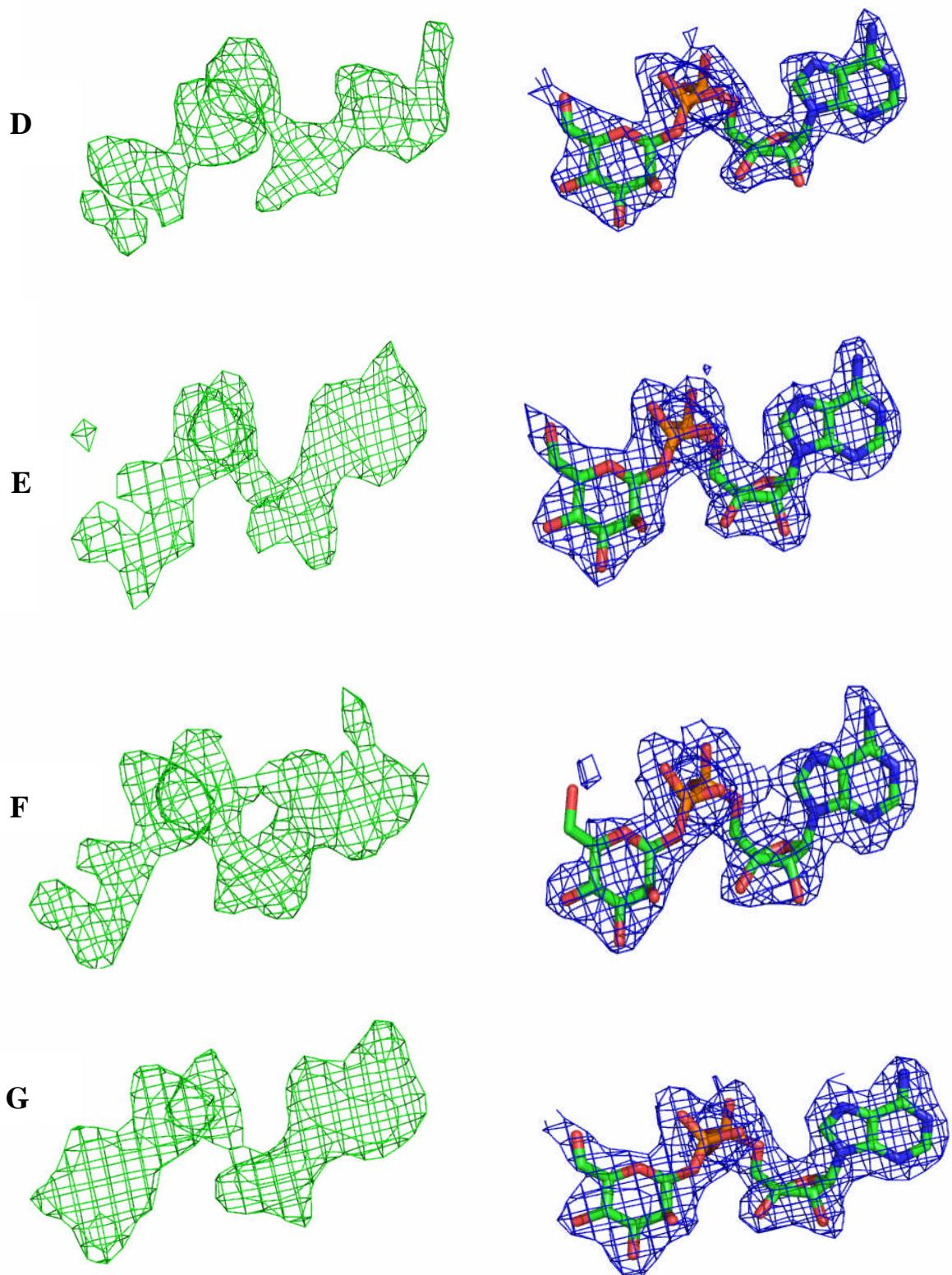


Figure 76. (continues on next page)

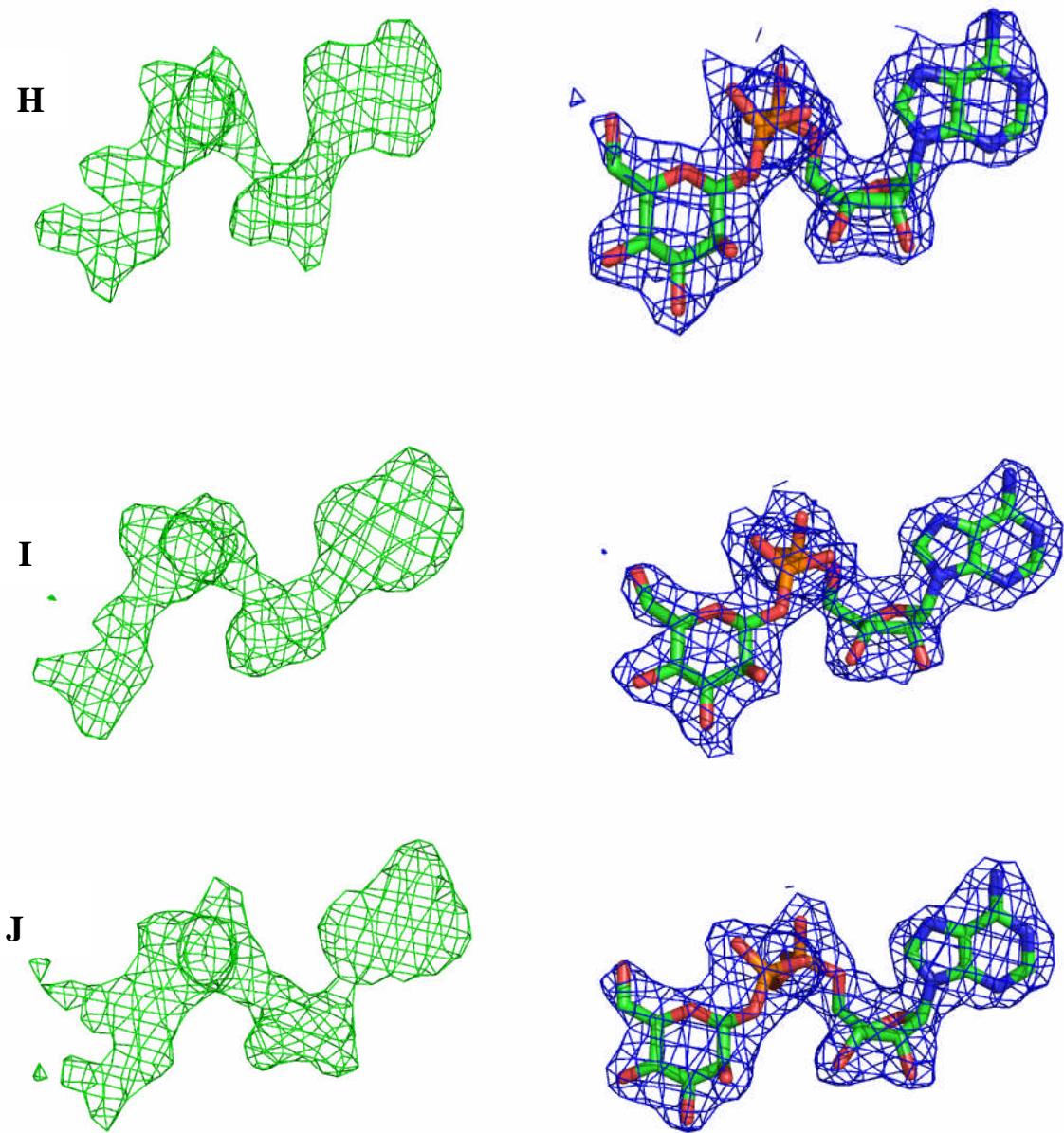


Figure 76. Unbiased $F_o - F_c$ at 3σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around ADP- β -mannose of monomers A-J of the 2.4 \AA structure of AGME Y140F. The sugar molecule is modelled in the $2F_o - F_c$ density. Both electron densities for the nucleotide sugar of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. This figure has been produced using PyMol (DeLano, 2007).

There is clear electron density for one molecule of cofactor in each of the 20 monomers of the 2.8 Å structure (Figure 77). In 19 subunits enough electron density was present to fit ADP- β -mannose. In one monomer (B) only ADP could be modelled in (Figure 78).

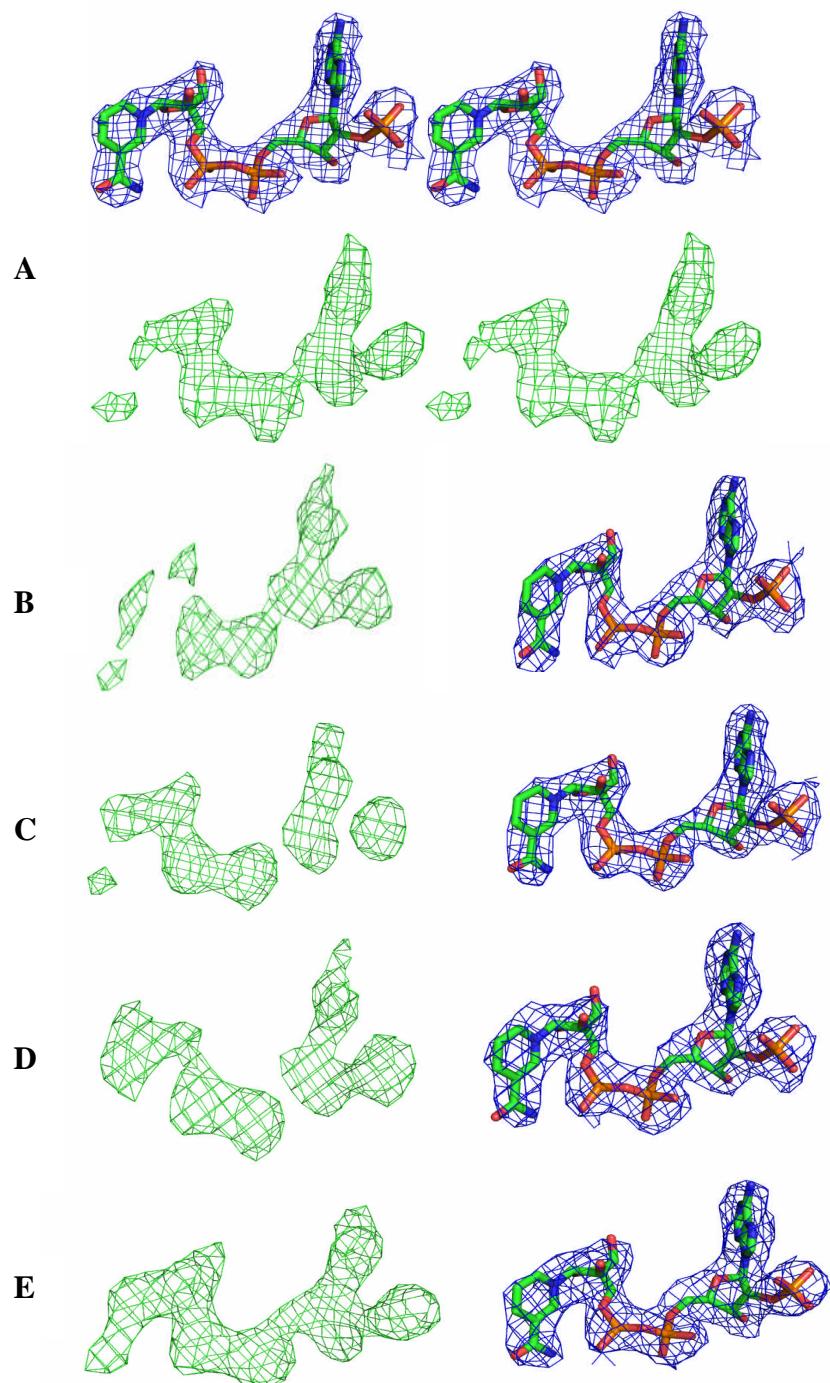


Figure 77. (continues on next page)

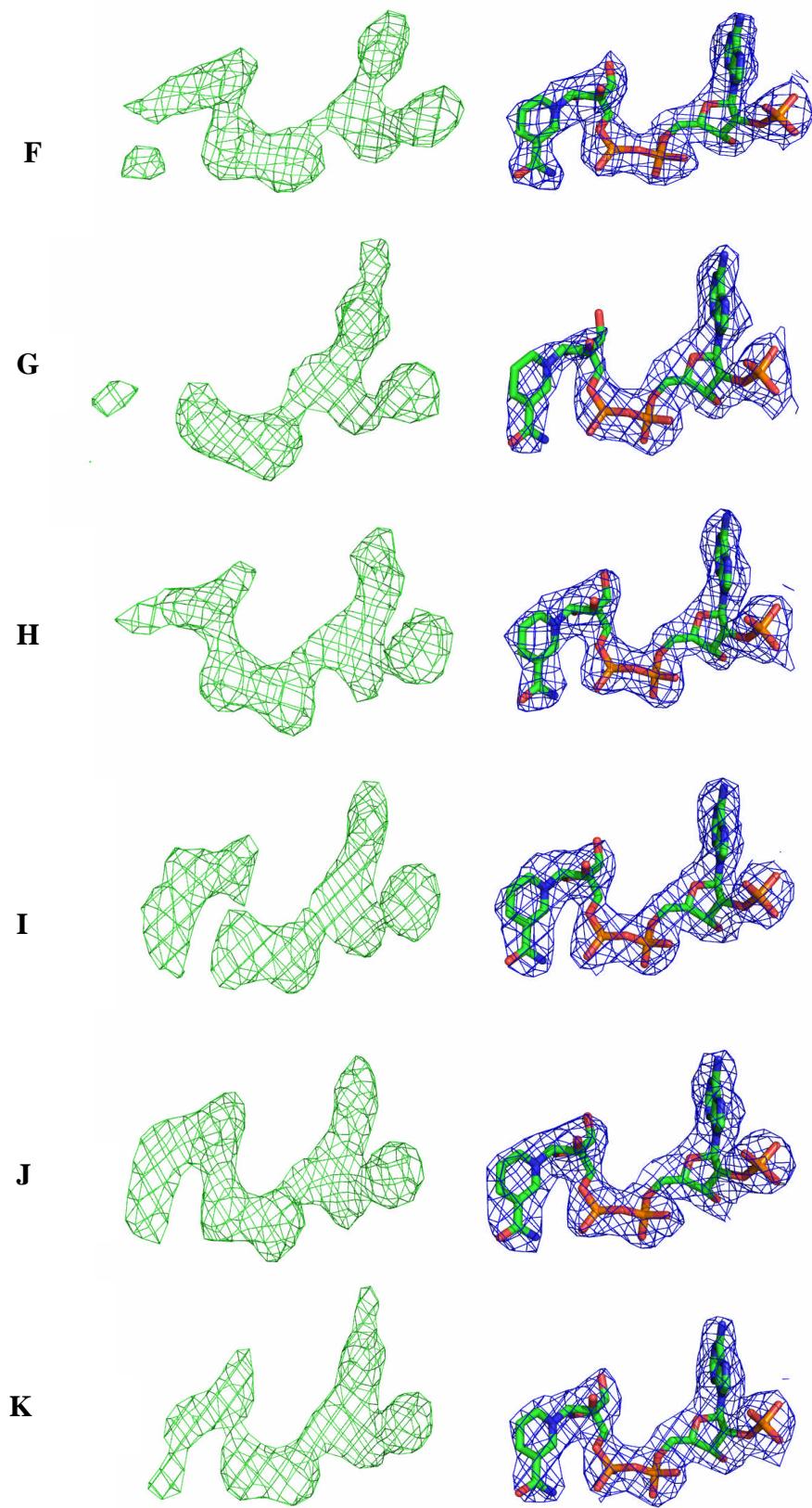


Figure 77. (continues on next page)

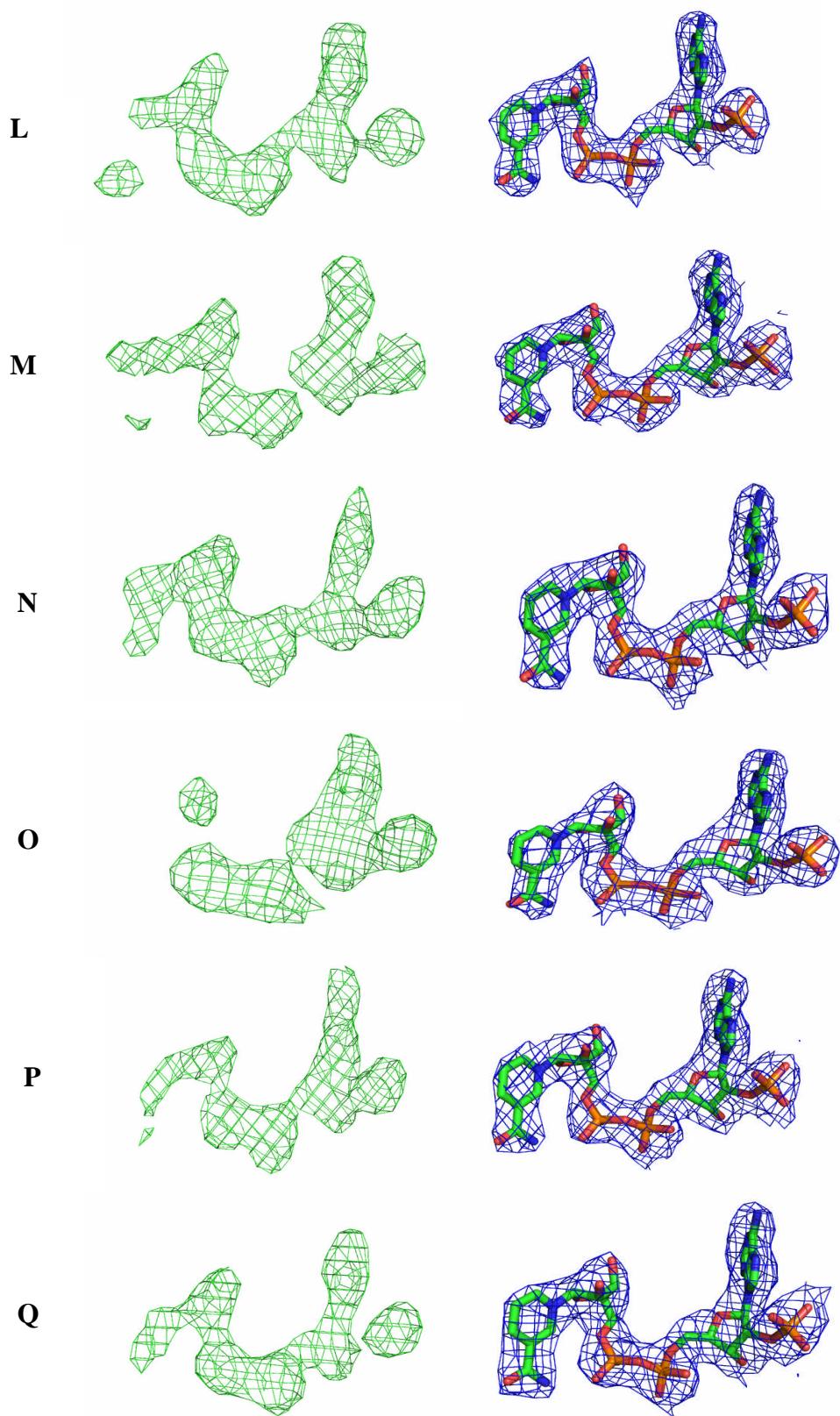


Figure 77. (continues on next page)

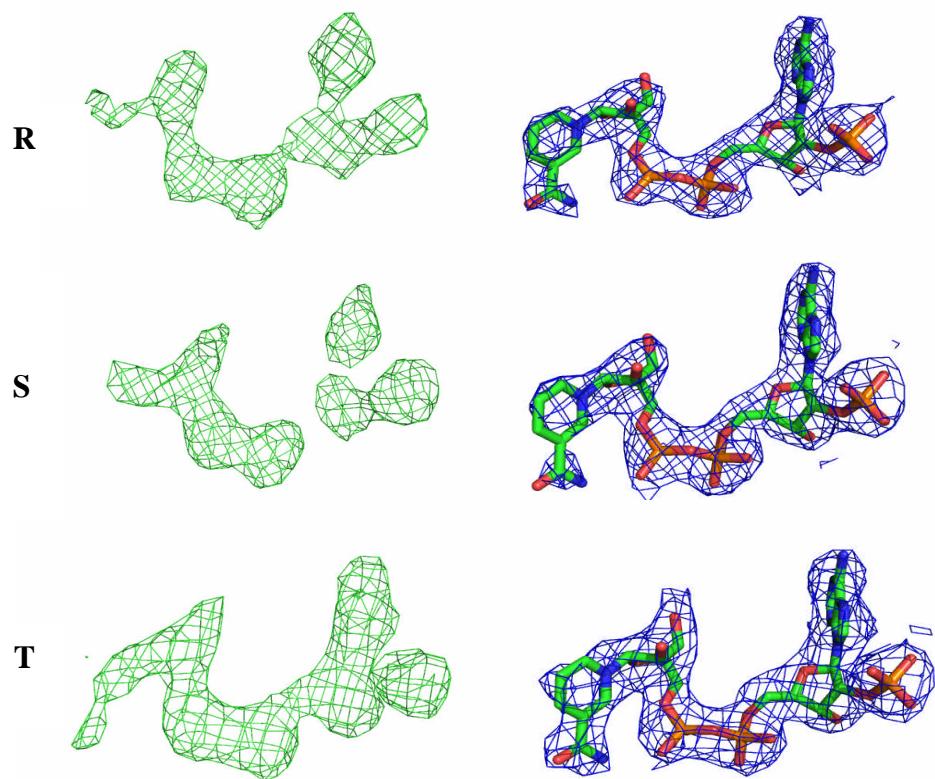


Figure 77. Unbiased $F_o - F_c$ at 3σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around the NADP^+ of monomers A-T of the 2.8 Å structure of AGME Y140F. The NADP^+ -molecule is modelled in the $2F_o - F_c$ density. Both electron densities for the cofactor of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. The weak electron density at the nicotinamide part of NADP^+ ($F_o - F_c$ maps) shows that this moiety of the molecule is flexible. This figure has been produced using PyMol (DeLano, 2007).

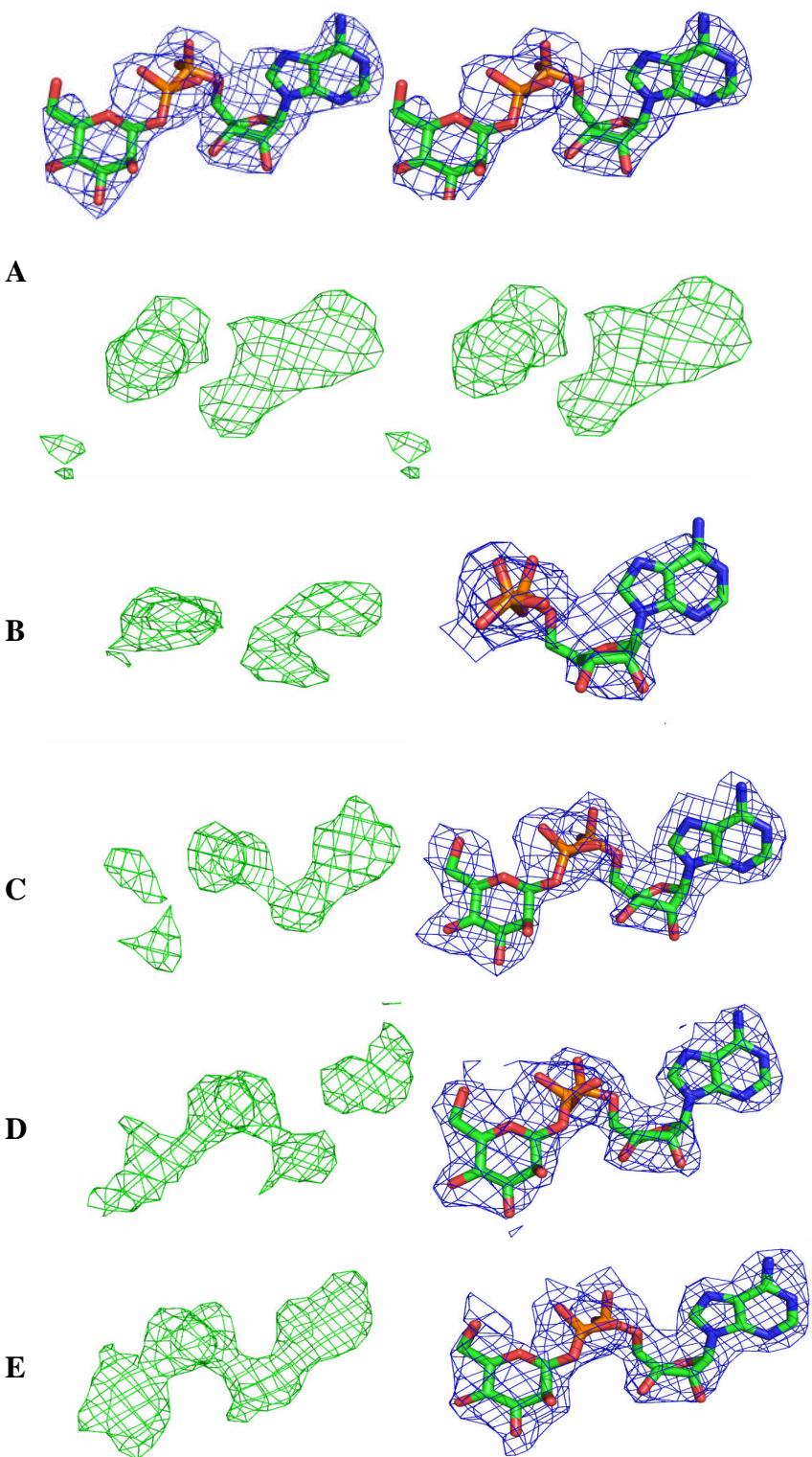


Figure 78. (continues on next page)

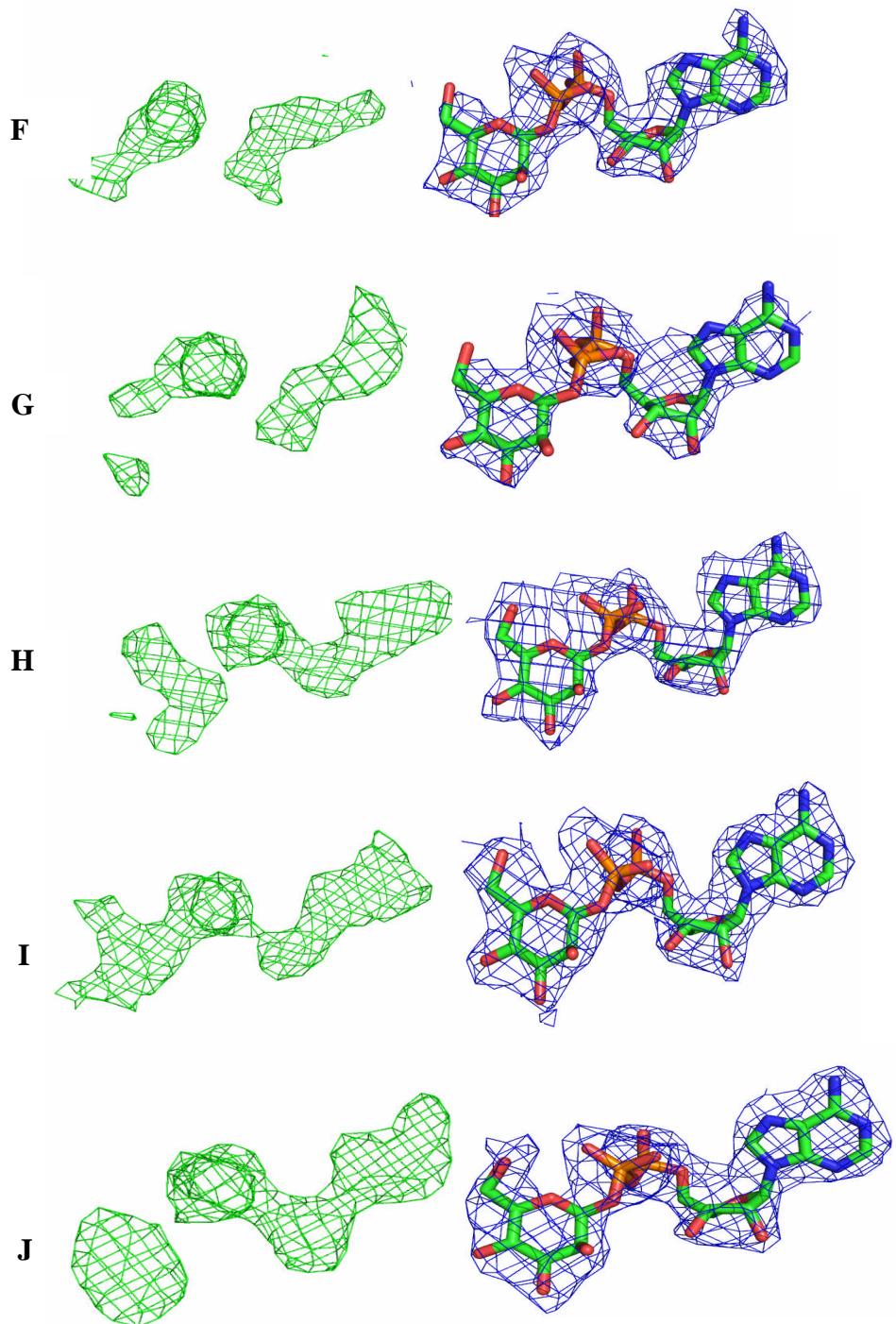


Figure 78. (continues on next page)

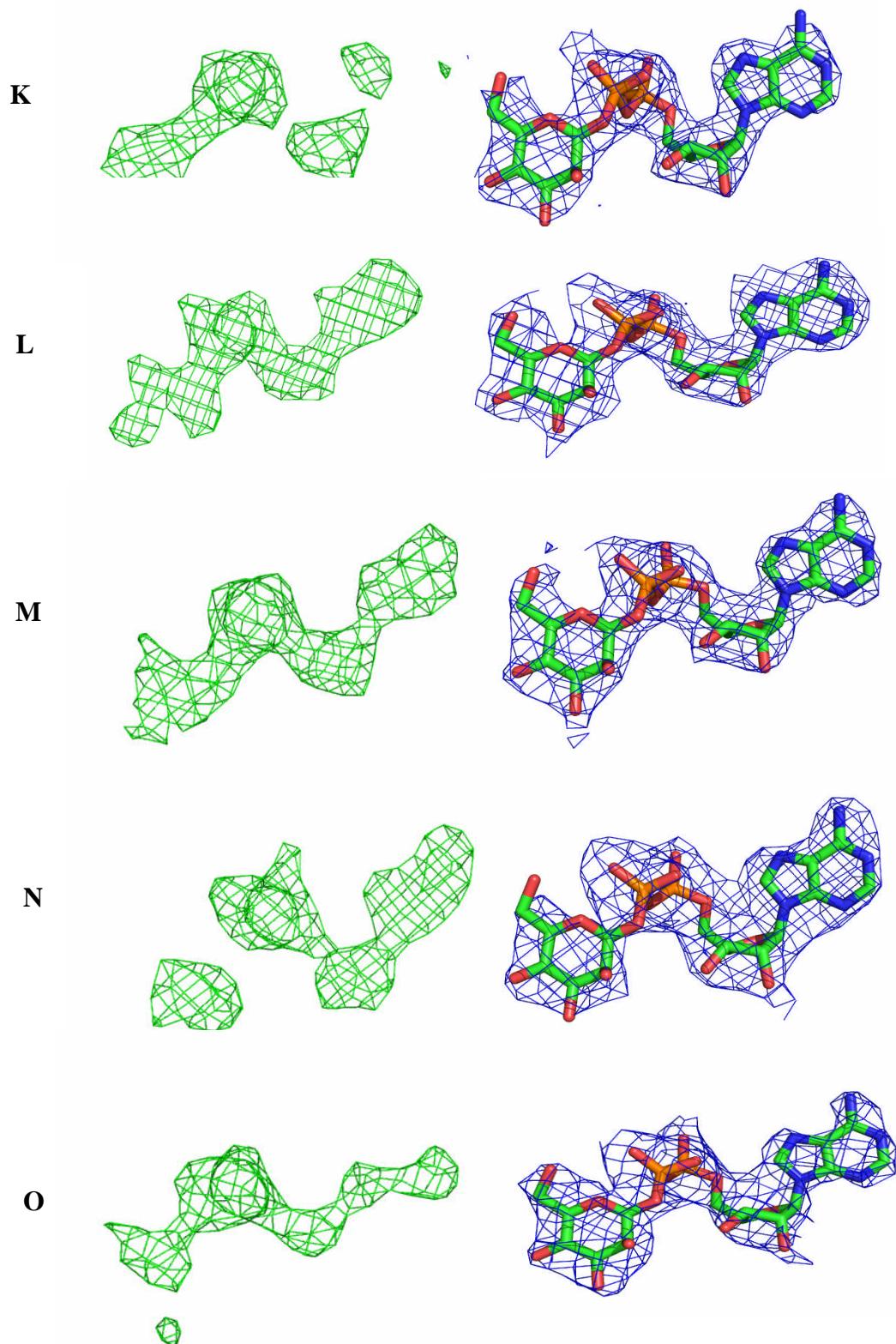


Figure 78. (continues on next page)

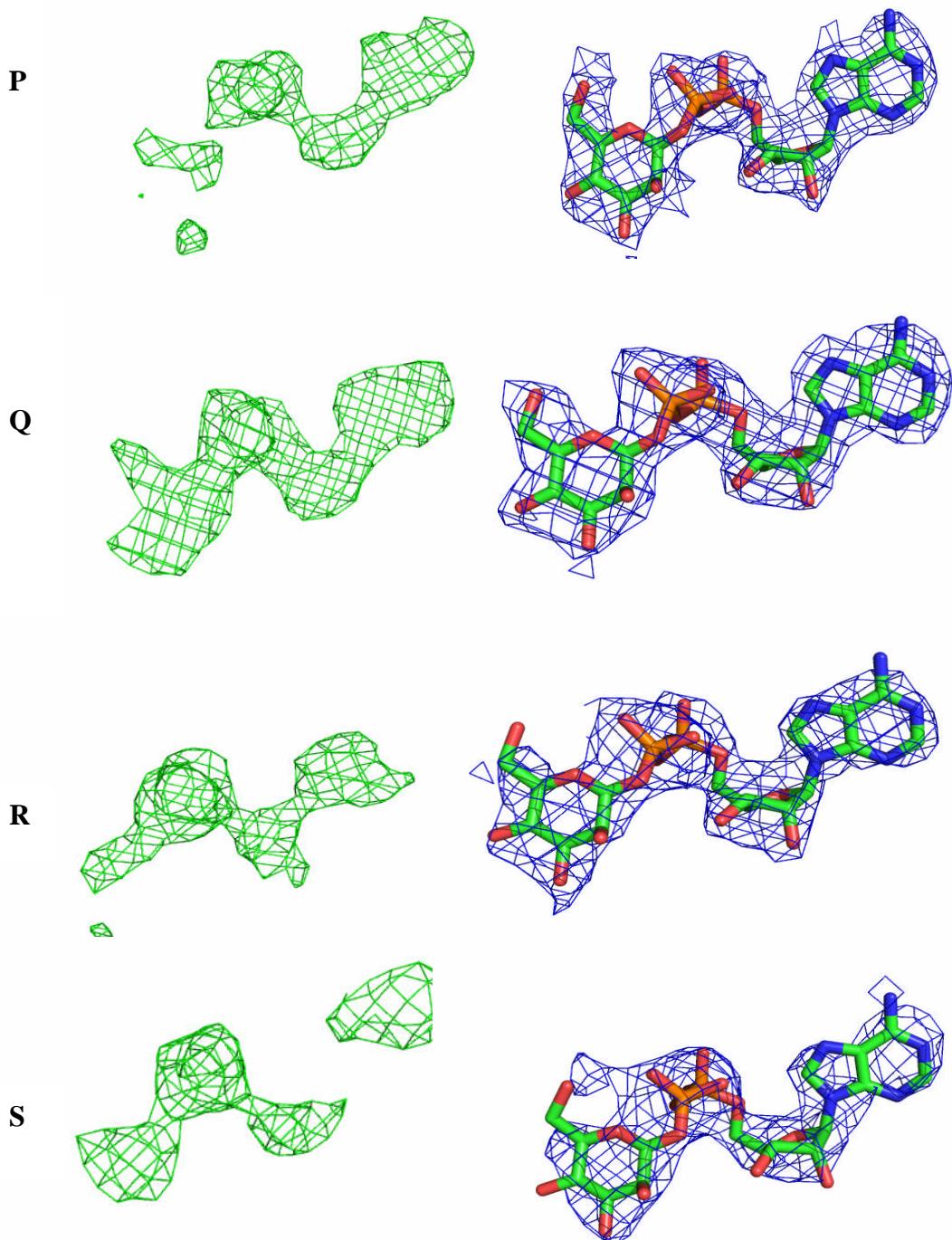


Figure 78. (continues on next page)

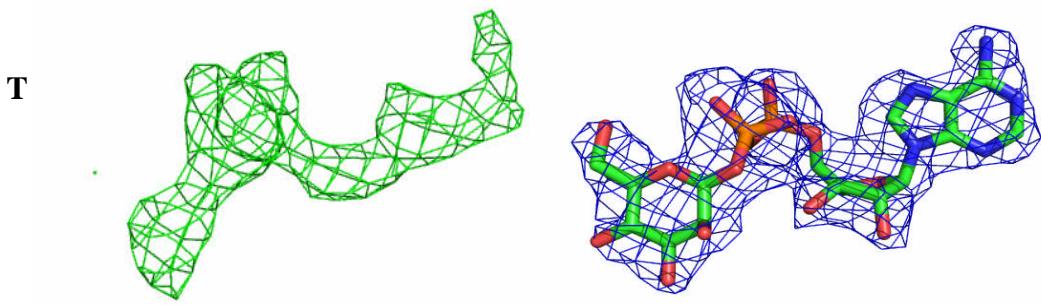


Figure 78. Unbiased $F_o - F_c$ at 3σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around ADP- β -mannose of monomers A and C-T and ADP of monomer B of the 2.8 \AA structure of AGME Y140F. The sugar molecule and ADP are modelled in the $2F_o - F_c$ density. Both electron densities for the nucleotide sugar of monomer A are shown in stereo view. Both maps were calculated at final stages of refinement before modelling of the ligands. This figure has been produced using PyMol (DeLano, 2007).

In the 2.4 \AA resolution structure, cofactor and ADP- β -mannose molecules are essentially located at the same positions in all ten monomers (Figure 79). The Rmsd values for the superimposed NADP molecules are in the range between $0.014\text{-}0.022 \text{ \AA}$ and for the nucleotide sugars they vary between 0.084 and 0.121 \AA . However, these Rmsd values are lower than the estimated overall coordinate error (see Murshudov and Dodson, 1997) The ESU (Estimated Standard Uncertainties) based on maximum likelihood are 0.14 \AA which means that these root mean square deviations are within structural error.

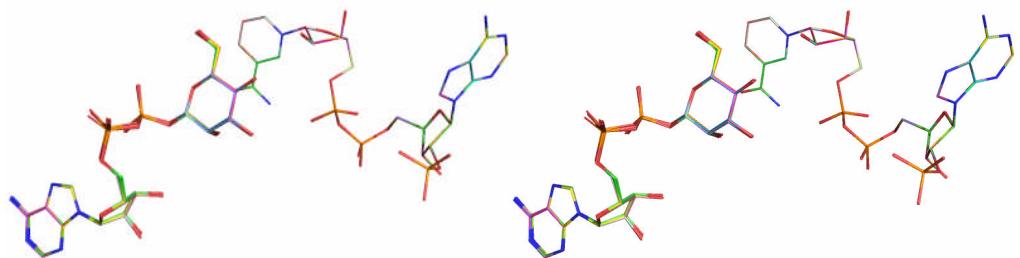


Figure 79. Stereo view of the ten superimposed NADP and ADP- β -mannose molecules of 2.4 \AA AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

In each of the twenty subunits of the lower resolution structure the ligands are essentially positioned in the same locations as well (Figure 80). The Rmsd values for the twenty superimposed coenzyme molecules differ in a range between 0.023-0.049 \AA . The Rmsds for the nineteen superimposed nucleotide sugars (monomer B with only ADP-fitted has not been considered) are slightly higher (between 0.201 and 0.319 \AA).

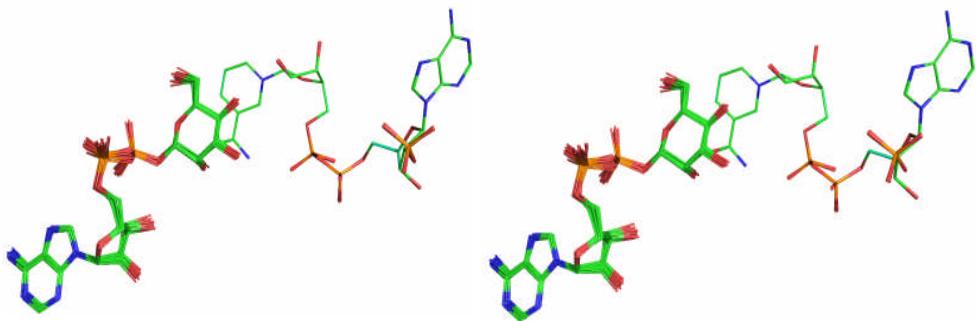


Figure 80. Stereo view of the twenty superimposed NADP and nineteen ADP- β -mannose molecules of 2.8 \AA AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Superposition of the ligand molecules from one subunit of the higher resolution and one monomer of the lower resolution structure (Figure 81) also shows that the positions of the coenzymes and nucleotide sugars in the active site are essentially the same in both structures. The Rmsd values for the aligned coenzymes and nucleotide sugars are 0.11 Å and 0.23 Å, respectively. However, these Rmsd values are lower than the estimated overall coordinate error of the structure (see Murshudov and Dodson, 1997). The ESU (Estimated Standard Uncertainties) based on maximum likelihood are 0.31 Å .

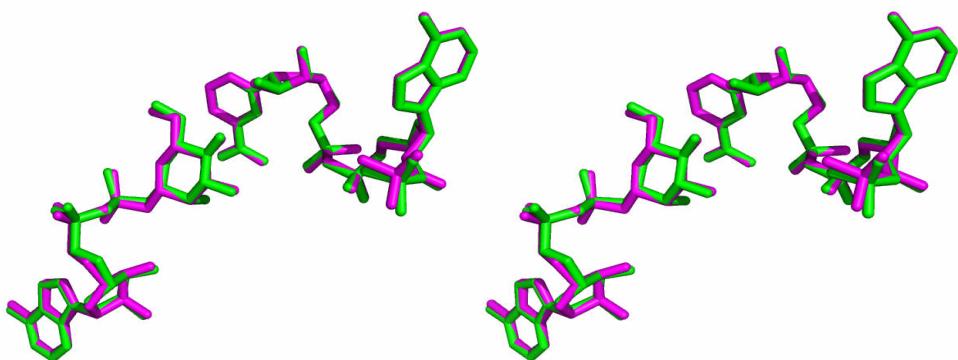


Figure 81. Stereo view of one superimposed NADP and ADP- β -mannose molecule from monomer A of the 2.4 Å (both molecules in green) and 2.8 Å structure (both molecules in magenta) of AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

2.14.2 NADP-binding site

As in the ten subunits of the AGME crystal structure in complex with ADP- α -glucose (Deacon *et al.*, 2000), the NADP-molecules are essentially located at the same positions in all ten or twenty monomers of AGME Y140F (Figures 79 and 80).

If one superimposes the cofactor molecules from one subunit of the mutant enzyme on one from that of the ADP- α -glucose bound structure (Figure 82) one can see very clearly that the ligands in both monomers also share more or less the same position.

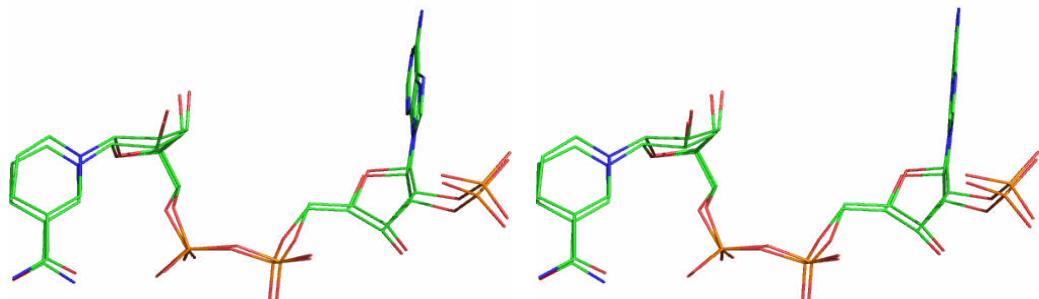


Figure 82. Stereo view of the superimposed NADP-molecules from monomer A of 2.4 Å AGME Y140F and monomer A of wildtype AGME (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

Contacts of the cofactor with the epimerase are discussed in detail by Deacon *et al.*, 2000. However, there are some notable distinctions between the mutant and Deacon and coworkers' structure. Closer inspection of the nicotinic acid amide moiety revealed that the N7 and the O7 atoms of the carboxamide group are positioned differently. Comparison to the NAD-molecules of the structurally most related SDR-enzymes UDP-galactose 4-epimerase (Pdb accession code 1ek6) (Thoden *et al.*, 2000) and dTDP-D-glucose 4,6 dehydratase (Pdb accession code 1bxk) (Thoden *et al.*, unpublished) (Deacon *et al.*, 2000) shows that both atoms of the niacinamide are in exactly the same positions as in AGME Y140F.

It has been reported that in UGE (Thoden *et al.*, 2000; Thoden *et al.*, 1996), DGD (RmlB) (Allard *et al.*, 2001a; Allard *et al.*, 2004) and also RmlD (Blankenfeldt *et al.*, 2002), the nicotinic acid amide of the nicotinamide adenine dinucleotide phosphate builds up an intramolecular H-bond between the amide and the α -PO₄³⁻ of the adjacent ribose (Thoden *et al.*, 2000; Thoden *et al.*, 1996; Allard *et al.*, 2001a; Allard *et al.*, 2004 ; Blankenfeldt *et al.*, 2002) to occur in a *syn* conformation with the B face being orientated to the sugar (Thoden *et al.*, 2000; Thoden *et al.*, 1996; Blankenfeldt *et al.*, 2002; Deacon *et al.*, 2000). For that reason it can be assumed that the same intramolecular H-bond occurs in the nicotinamide of AGME Y140F as well. However, one explanation for the different orientation of the carboxamide in Deacon and coworkers' structure could be that in most of the ten monomers of their structure the carbonyl is in H-bonding distance to Lys178. Due to the fact that the C=O represents a proton acceptor and the lysine usually carries a positive charge (NH₃⁺) (Blankenfeldt *et al.*, 2002) a hydrogen bond could be formed between both groups and the amide could become protonated. That is why it can be concluded that both orientations of the carboxamide of the nicotinamide are acceptable. Figure 83 shows a superposition of the nicotinamide moieties in all four structures.

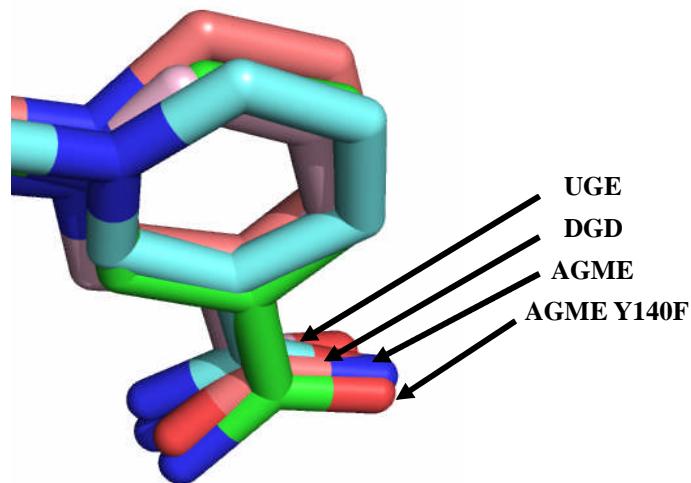


Figure 83. Close up view of the superimposed nicotinamide moieties from monomer A of 2.4 Å AGME Y140F, monomer A of wildtype AGME (Pdb accession code 1eq2) (Deacon *et al.*, 2000), monomer A of UGE (1ek6) (Thoden *et al.*, 2000) and monomer A of DGD (Pdb accession code 1bxk). This figure has been produced using PyMol (DeLano, 2007).

There is no clear evidence that Cys78, which occurs in an oxidised state in the ten monomers of the AGME structure (Deacon *et al.*, 2000) is also oxidised in the ten or twenty subunits of AGME Y140F. In the wildtype structure a direct coordination occurs in a few subunits whereas in the other monomers Cys78 coordinates via H₂O (Deacon *et al.*, 2000). The average B factors of the atoms in the cofactor molecules of the 2.4 Å and 2.8 Å mutant structures are 34.8 Å² and 36.1 Å². These values are slightly higher than the average atomic B factor of the NADP-molecules in Deacon and coworkers' structure which is 28.2 Å². In the structure of AGME complexed to ADP- α -glucose, the average B factor of the atoms of the niacinamide moiety (33.2 Å²) is significantly higher than that of adenine (24.2 Å²) which has strong interactions with residues of the enzyme indicating that the nicotinamide is more flexible (Deacon *et al.*, 2000).

The average atomic temperature factors of the nicotinic acid amide and adenine moieties in both mutant structures, however, do not differ much (Table 33). Furthermore, whereas the average B factors of the nicotinic acid amide part differ between the monomers in Deacon and coworkers' structure (Deacon *et al.*, 2000), essentially no differences between these values are observed in the ten subunits of the 2.4 Å and twenty monomers of the 2.8 Å mutant structure (Table 33).

Table 33. Overview of average B values (\AA^2) of protein, ligands, diverse ligand moieties and H_2O in the mutant structures and the already published AGME structure. Differences over the monomers are given in brackets.

average B value (\AA^2)	2.4 \AA mutant structure	2.8 \AA mutant structure	AGME structure (Deacon)
protein (mainchain)	32.8	35.6	29.1
protein (sidechain)	33.3	35.8	30.5
NADP	34.8 (-)	36.1 (-)	28.2 (26-34.8)
nicotinamide ring	35.4 (-)	36.7 (36.7-36.8)	33.2 (35.5-53.5)
adenine base	33.2 (-)	35.8 (35.7-35.8)	24.2 (16.9-26.2)
ADP-$\beta(\alpha)$-mannose (glucose)	36-37	37.6 (37.4-37.9)	55.4 (45.4-61.8)
mannose (glucose)	41-42	41-43	72 (62.5-80.1)
adenine base	32.6 (32.1-32.9)	35.2 (34.4-35.4)	45 (26 -51)
water	40	28	34

2.14.3 Substrate-binding site

In all monomers of the 2.4 \AA and in nineteen subunits of the 2.8 \AA resolution mutant structure, ADP- β -mannose is bound to the active site in one single conformation (Figures 79-81). In the subunits of AGME structure complexed to ADP- α -glucose (Deacon *et al.*, 2000), however, the sugar moiety is disordered. For that reason it was only possible to fit the whole nucleotide sugar into three of the ten monomers of the asymmetric unit. Moreover, two completely distinct orientations of the glucose part can also be observed in these three subunits (Deacon *et al.*, 2000) (Figure 84).

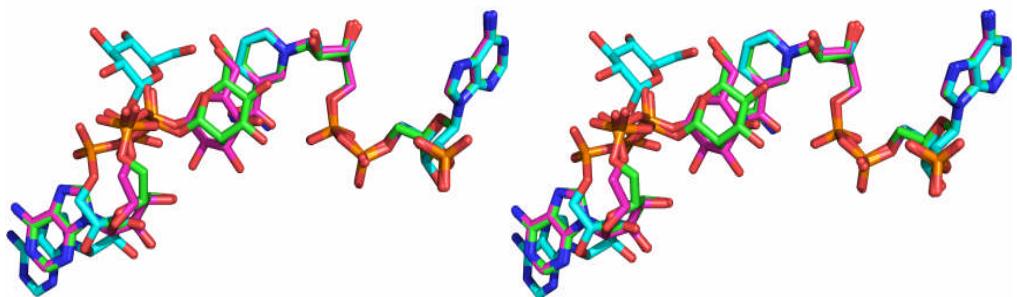


Figure 84. Stereo view of the superposition of NADP and ADP- β -mannose molecules from monomer A of the 2.4 \AA resolution AGME mutant (carbons in green) with the NADP and ADP- α -glucose-moieties from monomers B (C-atoms in cyan) and D (carbons in magenta) of AGME (Pdb accession code 1eq2) (Deacon *et al.*, 2000). This figure has been produced using PyMol (DeLano, 2007).

The average B-factor of ADP- α -glucose (55.4 \AA^2) is significantly higher than that of the coenzyme (28.2 \AA^2) and these values also show a great difference between the ten subunits (Table 33). In the three ADP- α -glucose molecules the average B-factors of the sugar moieties are much higher than those of the adenine parts, 71.6 \AA^2 and 45.2 \AA^2 , respectively (Deacon *et al.*, 2000). The high average temperature factor of 72 \AA^2 at the hexose position indicates that this part is very flexible due to the disordered electron density and the two different conformations. In both mutant structures, however, the average atomic temperature factors of ADP- β -mannose are only slightly higher than those for the cofactor (Table 33). There, the B-factors of the nucleotide sugar only vary slightly from subunit to subunit (Table 33). Unlike in the AGME structure complexed with ADP- α -glucose (Deacon *et al.*, 2000), the difference between the average temperature factors of the mannose and adenine moieties of the nucleotide sugar in both mutant structures is small (Table 33).

The relatively low average B-factors of the mannose moiety in the mutants compared to the high temperature factor of the glucose part in Deacon and coworkers' structure can be explained by the fact that the electron density for the mannose is well defined.

Given the similarity between the two mutant structures and within each monomer, the A subunit of the 2.4 Å resolution structure is discussed. Any significant differences amongst subunits or between the 2.4 and 2.8 Å structures are noted.

The adenine moiety of ADP-β-mannose interacts, like the base of ADP-α-glucose in Deacon and coworkers' structure, particularly with hydrophobic residues such as Val184, Leu200, Phe201 and Phe243 (Deacon *et al.*, 2000) (Figure 85).

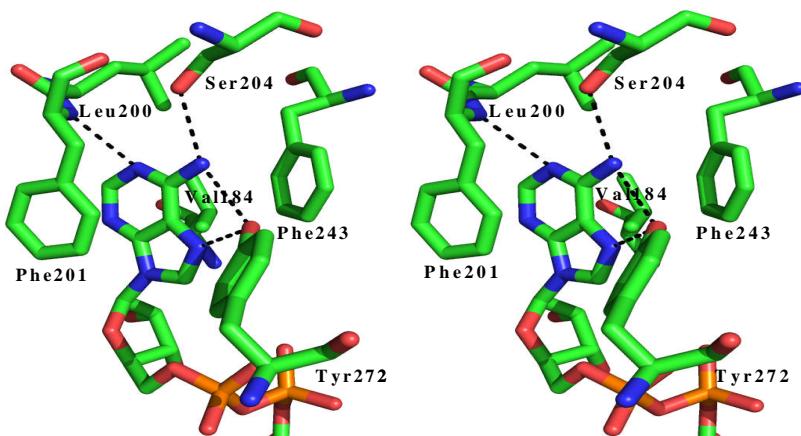


Figure 85. Stereo view of the adenine binding site in AGME Y140F. Hydrogen bonds are shown as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

In general the ribose part of ADP-β-mannose also shows the same associations with the protein as the pentose moiety of the adenosine diphosphate part of ADP-α-glucose (see Figure 5 (b), Deacon *et al.*, 2000) (Figure 86).

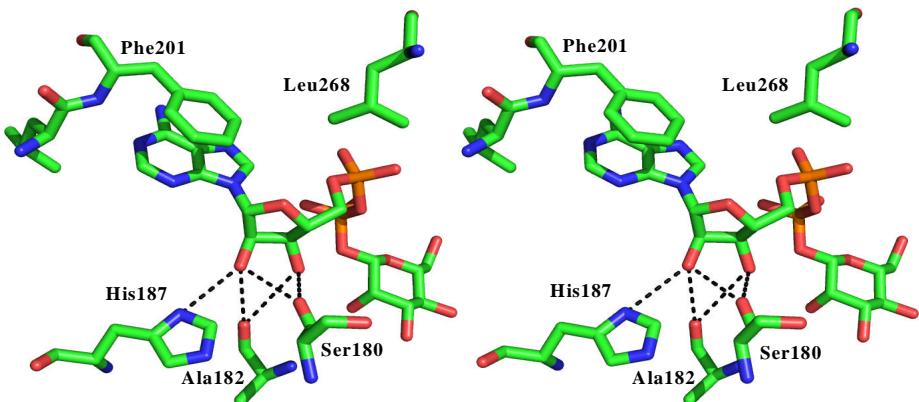


Figure 86. Stereo view of the ribose binding site in AGME Y140F. H-bonds between atoms of the pentose and amino acids Ser180, Ala182 and His187 are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

The distances of amino acids of AGME Y140F which are in close contact to the pyrophosphate moiety of ADP- β -mannose are also essentially identical to those responsible for the binding of the PP_i of ADP- α -glucose (see Figure 5 (b), Deacon *et al.*, 2000) (Figure 87)

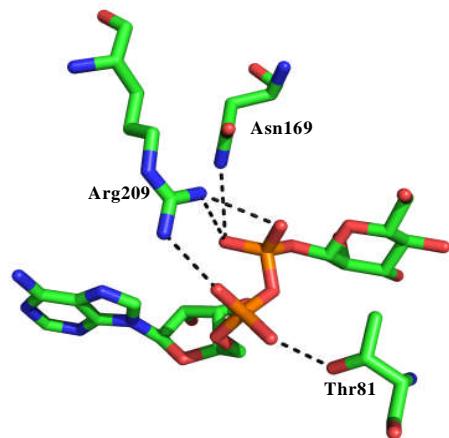


Figure 87. Close up view of the pyrophosphate binding site in AGME Y140F. Hydrogen bonds between atoms of the PP_i and amino acids Thr81, Arg209 and Asn169 are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

The most significant differences between the bound nucleotide sugars in the mutant and Deacon and coworkers' structures lie in the sugar moieties. Whereas two different orientations of the glucose can be observed in AGME (Deacon *et al.*, 2000), only one conformation of the mannose part occurs in AGME Y140F. The oxygen at position 6 of mannose is in close contact with three residues, Ser116, Ala118 and 140Phe (Figure 88). The distances between O6 of mannose and CZ of 140Phe differ from 4.2 to 4.4 Å over nine monomers. Only in subunit F the contact is not so close (4.8 Å). The oxygen at position 6 of the hexose moiety is in hydrogen bonding distance to OG of Ser116. The distances differ from 2.5 to 2.7 Å among the ten subunits. CB of Ala118 is the third residue which is located in the vicinity of O6 of mannose. The distances from monomer to monomer vary also only slightly (3.8-4.2 Å). Two residues are located close to O2 of mannose, Lys178 and Met181 (Figure 88). Distances between O2 and NZ of Lys178 range between 3.2 and 3.4 Å over the ten subunits, and O of Met181 is positioned 3.2-3.3 Å from O2. The oxygen at position 3 of the hexose moiety is hydrogen bonded to Ser79 and NZ of Lys178. O of Met181 is also closely located to O3 of the mannose moiety (Figure 88). The distances from monomer to monomer vary only slightly between 3.8 and 4.0 Å. O of Ser79 is also in H-bonding distance to O4 of the sugar. The second amino acid closely located to this oxygen is 140Phe (Figure 88). Distances between O4 and CZ of 140Phe range from 3.5 to 3.7 Å over the ten monomers.

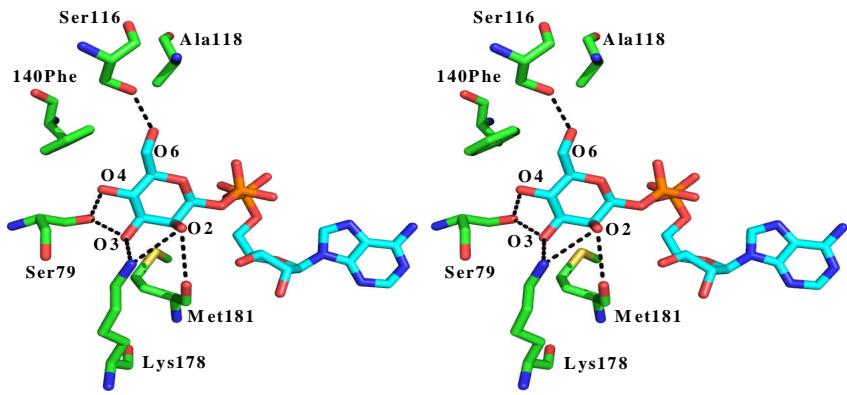


Figure 88. Stereo view of the interactions of O₂, O₃, O₄ and O₆ of the mannose moiety of ADP- β -mannose (carbons in cyan) in AGME Y140F with closely positioned residues (C-atoms in green). H-bonds are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

If one compares the positions of mannose from AGME Y140F and glucose from monomer D of Deacon and coworkers' structure in the active sites, similarities and also differences between both moieties can be observed (Figure 89).

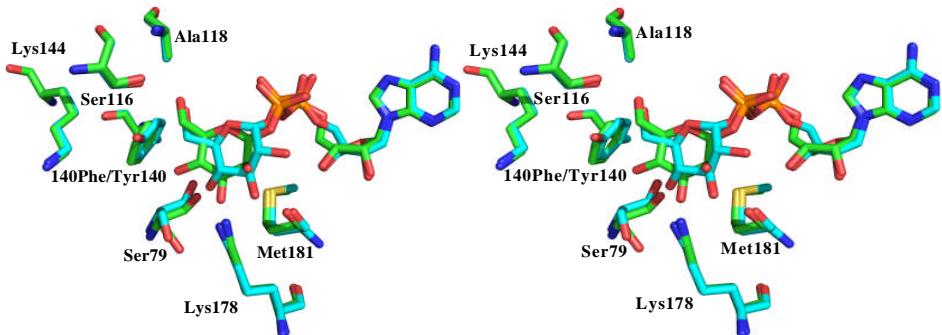


Figure 89. Stereo view of superimposed ADP- β -mannose (carbon atoms in green) of AGME Y140F and ADP- α -glucose (carbons in cyan) from monomer D of AGME (Pdb accession code 1eq2). Residues close to the hexose moieties of both molecules, and also the highly conserved Lys144 (Deacon *et al.*, 2000) are also shown (C-atoms of residues of AGME Y140F are shown in green and carbons of AGME in cyan). The exactly same positions of the active site residues (Ser116, 140Phe/Try140 and Lys144 (Deacon *et al.*, 2000)) in both structures indicate that the mutation does not perturb the catalytic region of AGME. This figure has been produced using PyMol (DeLano, 2007).

The distance of 3.8 Å between the oxygen at position 6 of glucose and CZ of Tyr140 hardly differs from that of O6 of mannose to CZ of 140Phe (4.2-4.4 Å). The same oxygen is positioned 3.3 Å from the hydroxyl group of the Tyr140 side chain. OG of Ser116 is also located within H-bonding distance (3.0 Å) to O6 of the glucose moiety. The distance of 4.4 Å to CB of Ala118 is also similar to the distance between O6 of mannose and CB of Ala118 in the mutant structure (3.8-4.2 Å). O5 is also located at almost the same position in both hexoses (Figure 89). However, due to the fact that ADP- α -glucose and ADP- β -mannose display a diverse stereochemistry at C-1" and C-2" (Morrison and Tanner, 2007 Morrison *et al.*, 2005), the other atoms of both sugars occupy different positions (Figure 89). Whereas the glucose moiety in monomer D of AGME is located at a relatively similar position as the mannose part in AGME Y140F, the orientation of glucose in monomers B and F of AGME is somewhat different (Deacon *et al.*, 2000) (see Figure 84).

Investigation of the contact between C-6" of the mannose/glucose moieties and C-4" of NADP also revealed that the distances between the carbons at positions 6 of the two completely different orientated Glc molecules and 4 of the cofactor are longer (around 4 Å in monomer B, 4.1 Å in monomer F) (Deacon *et al.*, 2000) (Figure 90A and B) than that between C-6" of mannose and C-4" of the coenzyme in the mutant structure (2.9-3 Å over the ten subunits) (Figure 90C).

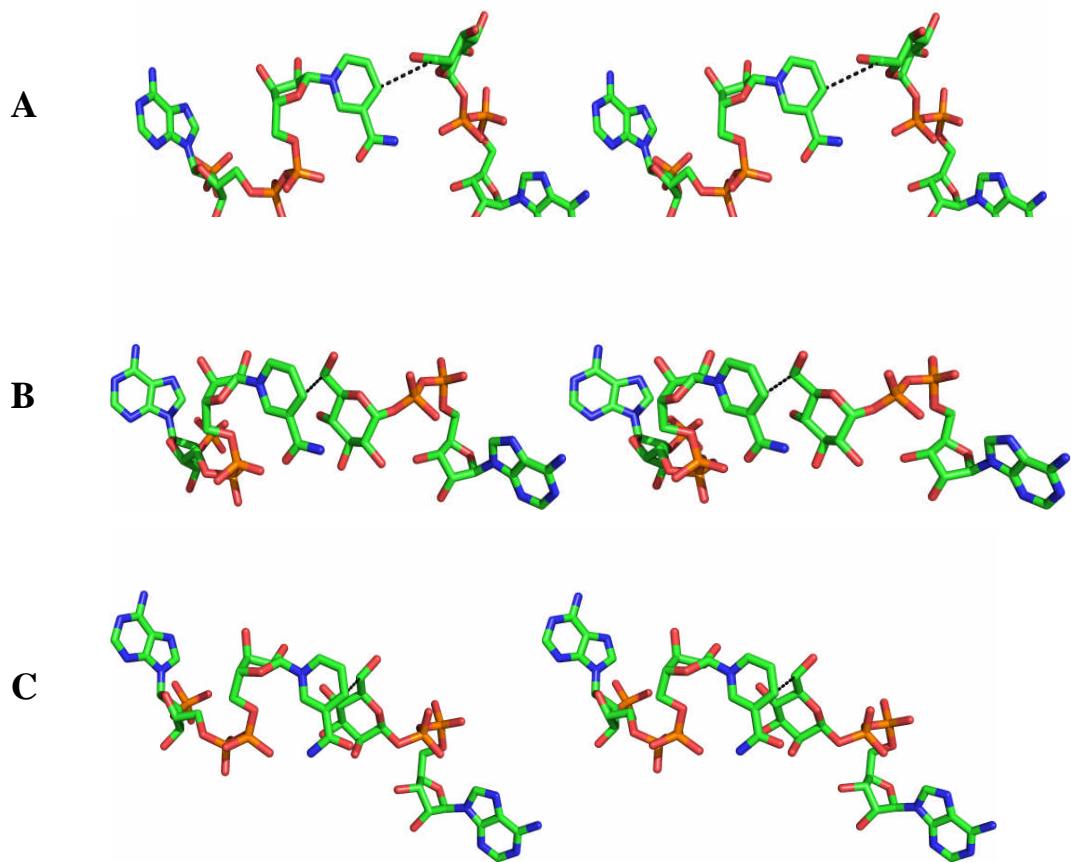


Figure 90. Stereo view of the distances between C6 of the glucose moiety of ADP- α -glucose and C4 of NADP in monomers B (4 Å) (A) and D (3.4 Å) (B) of AGME (Pdb accession code 1eq2) and between C6 of the mannose part of ADP- β -mannose and C4 of NADP (2.9 Å) (C) of 2.4 Å AGME Y140F. This figure has been produced using PyMol (DeLano, 2007).

Discussion

The structure of AGME in complex with ADP- β -mannose displays significant differences to the wildtype enzyme when bound to ADP- α -glucose (Deacon *et al.*, 2000). In the mutant only one orientation of the sugar moiety is observed whereas the hexose moiety of ADP- α -glucose in Deacon and coworkers' structure has two completely different orientations. For that reason it has been difficult to propose a mechanism for AGME based on structural data and biochemical analysis. It has been suggested that Tyr140 and Lys178 could act as the catalytic bases (Morrison and Tanner, 2007). However, the structure of AGME Y140F complexed with a substrate in the β -manno configuration clearly shows that Lys178 is located too far (7.7 Å) from O6 of mannose, the position at which H⁺-abstraction and -addition takes place (Figure 91A) (Morrison and Tanner, 2007). If one rotates the C-5"-C-6" bond by 120° so that O6 is orientated in the same way as it would be in the other epimer, the distance would still be too long (6.2 Å) for Lys178 to deprotonate the hydroxyl group (Figure 91B). This indicates that this Lys residue does not operate directly as a catalytic base. The fact that the K178M mutant only displays 0.1% epimerase activity compared to the wildtype protein (Morrison and Tanner, 2007) suggests that it plays some other important role in the catalytic process.

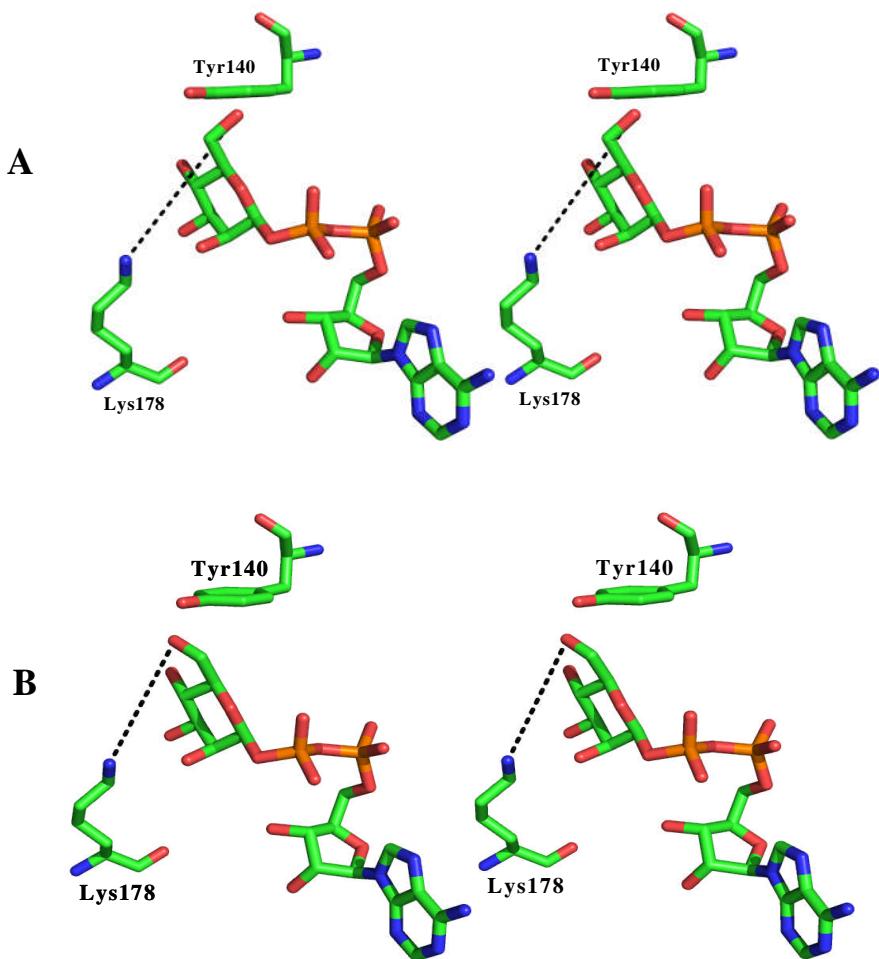


Figure 91. Stereo view of the distance between O6 of the mannose moiety of ADP- β -mannose and NZ of Lys178 of AGME (7.7 \AA) (A) and after a 120° rotation about the C-5''-C-6'' bond (6.2 \AA) (B). Tyr140 has been modelled based on the position of 140Phe. This figure has been produced using PyMol (DeLano, 2007).

It has been proposed that the Tyr residue of the catalytic triad always functions as the catalytic base which abstracts the H⁺ of the OH-group to oxidise an alcohol (or reduces the C=O group) (Morrison and Tanner, 2007). It has been suggested that, in Tyr directly abstracts H⁺ from the substrate (Naismith, 2006; Morrison and Tanner, 2007).

However, it has also been proposed, that H⁺-shuttling, in which the catalytic Ser plays a role, may take place (Liu *et al.*, 1997; Berger *et al.*, 2001; Morrison and Tanner, 2007).

A one base mechanism has been proposed for dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) from *Salmonella enterica* serovar Typhimurium. This member of the family of short chain dehydrogenases/reductases is essentially the best example to study the reaction mechanism of SDR-enzymes. Unlike the majority of other representatives of sugar metabolising enzymes which accomplish reactions in addition to simply transferring H⁻ to or from the coenzyme (Blankenfeldt *et al.*, 2002), RmlD just performs H⁻-addition to position 4 (Blankenfeldt *et al.*, 2002) of the substrate dTDP-6-deoxy-L-lyxo-4-hexulose (Schnaitman and Klena, 1993; Reeves, 1994; Graninger *et al.*, 1999; Blankenfeldt *et al.*, 2002). Amino acids which play a crucial role in H⁻-transfer are Tyr128, Lys132 and Thr104 (which is often replaced by Ser in other SDRs) (Blankenfeldt *et al.*, 2002). Thr104, Tyr128 and Lys132 represent the catalytic triad (Jörnvall *et al.*, 1995; Oppermann *et al.*, 2003; Filling *et al.*, 2002; Blankenfeldt *et al.*, 2002) in RmlD (Blankenfeldt *et al.*, 2002). The catalytic acid/base residue is Tyr128. The pK_a value of the oxygen of the tyrosine's phenolic side chain has been decreased to <7 due to NH₃⁺ of Lys132 (Blankenfeldt *et al.*, 2002) and the positively charged nitrogen of the niacinamide (Liu *et al.*, 1997; Thoden *et al.*, 2000; Blankenfeldt *et al.*, 2002) (3.6 Å between OH of Tyr128 and N1N of NADPH). This lysine is not in H-bonding distance with Tyr128 (5.3 Å between OH of Tyr128 and NZ of Lys132), however, it forms two H-bonds with O2D and O3D of the ribose moiety of the coenzyme. The distances are 3.2 and 3.1 Å, respectively. O_η of Tyr128 also forms a H-bond with O2D (3.1 Å) of the pentose. The nicotinic acid amide of the nicotinamide adenine dinucleotide phosphate also builds up an intramolecular H-bond between the amide and the α-PO₄³⁻ (2.7 Å between N7N of the pyridine-3-carboxamide and O2N of the α-phosphate) to occur in a *syn* conformation with the B face being orientated to the sugar (Blankenfeldt *et al.*, 2002).

In AGME complexed with ADP- β -mannose it is proposed that Tyr140 functions as the direct catalytic acid/base residue. Like in the structure of RmlD in complex with the product dTDP-L-rhamnose (Blankenfeldt *et al.*, 2002) the positively charged groups, NH₃⁺ of Lys144 and N1N of the cofactors moiety could also contribute to a decrease of the pK_a value of O_η of the conserved Tyr. The distance between OH of Tyr140 and N1N of the nicotinamide is 3.9 Å. NZ of Lys144 and the hydroxyl group of Tyr140 are not located in hydrogen bond distance either (5 Å), but the side chain nitrogen of Lys144 is also in close contact with the 2'OH and 3'OH of the cofactor's ribose (3.9 and 2.9 Å, respectively) (Figure 92). Another parallel to the RmlD complex structure is the presence of an H-bond between the conserved Tyr (Tyr140) and 2'OH of the pentose moiety of the coenzyme (2.8 Å) (Figure 92). The orientation of the nicotinic acid amide part is also in a *syn* conformation forming a short intramolecular H-bond (Blankenfeldt *et al.*, 2002) (Figure 92).

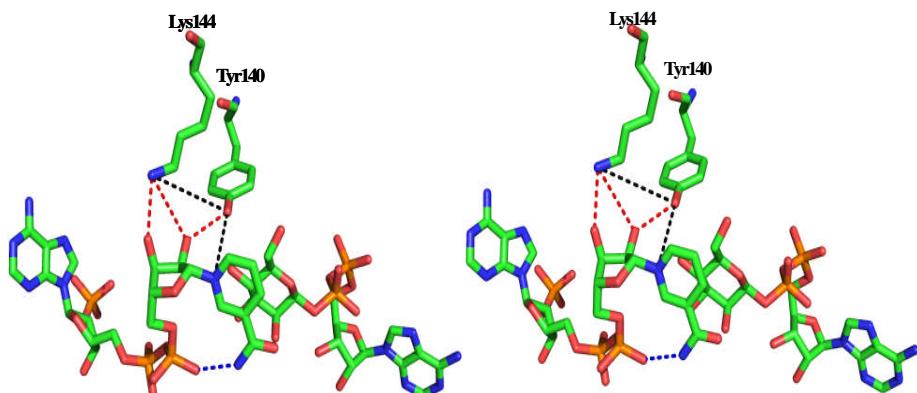


Figure 92. Stereo view of the distances between Tyr140 and Lys 144 and the 2'OH and 3'OH of the ribose of NADP (red dotted lines), the distances between Tyr140 and Lys144 and Tyr140 and N1N of the nicotinamide (black dashed lines), and the intramolecular H-bond of the cofactor (blue dotted line). Amino acids, NADP and ADP- β -mannose are shown as sticks .This figure has been produced using PyMol (DeLano, 2007).

H^- -transfer between C-6" of the sugar moiety and C-4" of the niacinamide ring requires that the C-6"-H of the hexose points directly to the carbon at position 4 of the nicotinamide (Morrison *et al.*, 2005; Morrison and Tanner, 2007). Both structures, RmlD in complex with dTDP-L-rhamnose, and AGME with ADP- β -mannose bound fulfil this requirement (Figure 93A and B).

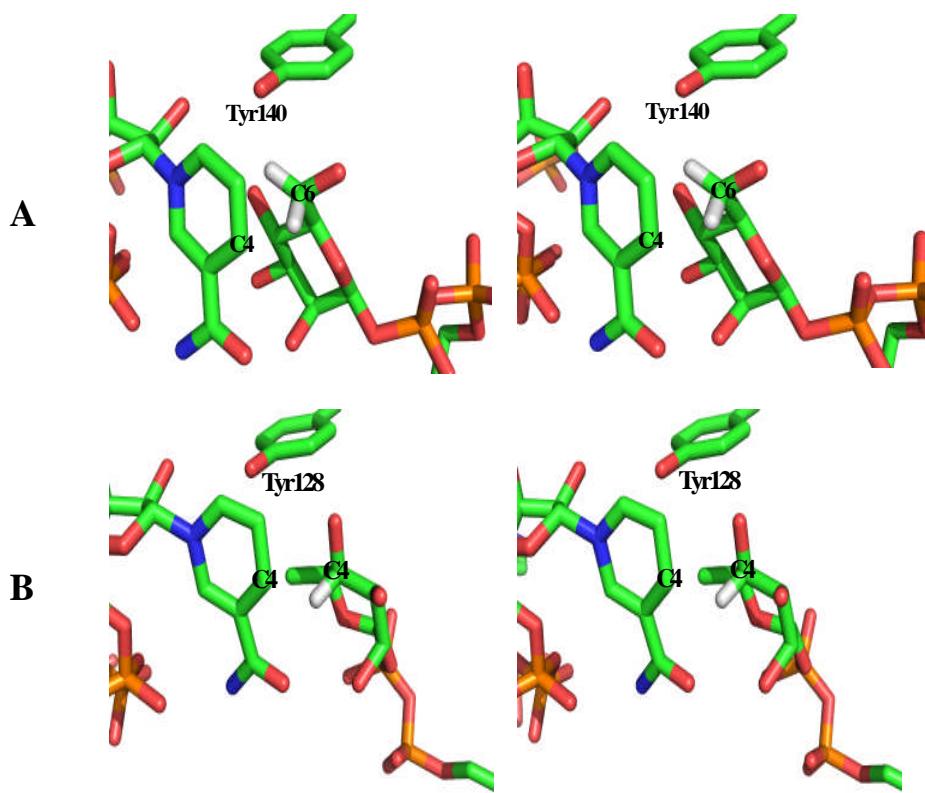


Figure 93. Stereo views showing the orientation of the hydrogens of C-6" of the mannose moiety in AGME (A) and C-4" of the rhamnose moiety in RmlD (Pdb accession code 1kc3) (B) in respect to C-4" of the nicotinamide ring of the coenzymes. Sugar moieties, cofactors (with hydrogens) and the conserved Tyr are shown as sticks. Distances between C-6"(mannose) / C-4" (rhamnose) and C-4" of the niacinamides are 2.9 Å and 3.1 Å, respectively. This figure has been produced using PyMol (DeLano, 2007).

The Ser or Thr residue of the catalytic triad of short chain dehydrogenases (Jörnvall *et al.*, 1995; Blankenfeldt *et al.*, 2002) plays a role in short H-connections with the C=O or OH of the substrate in SDR/substrate/cofactor complexes, however its specific function is the subject of debate (Blankenfeldt *et al.*, 2002). Firstly, it was proposed that this amino acid functions as H⁺-shuttle (Naismith, 2006; Liu *et al.*, 1997; Blankenfeldt *et al.*, 2002) because in abortive complexes it was observed that the Tyr and the sugar moiety are not located closely enough to carry out H⁺-transfer without the assistance of Ser (Thoden *et al.*, 1996a, Blankenfeldt *et al.*, 2002).

In the ternary AGME complex, the OH-group of Tyr140 is located more closely to O6 of the mannose moiety (3.3 Å) (Figure 94). The distance between OG of Ser116 and O6 of the hexose is 2.5 Å (Figure 94), similar to that between the corresponding atoms of the threonine and rhamnose in Blankenfeldt and coworkers' structure. The most significant difference between both complex structures lies in the distances between the Tyr and Thr/Ser. In AGME OH of Tyr140 and OG of Ser116 are only located 3.2 Å from each other (Figure 94) which would permit H⁺-shuttling. However, due to the close distance between Tyr140 and O6 of mannose a proton shuttling mechanism, involving another residue, would not be necessary. In non-abortive complexes, for instance in the structure of RmlB from *Streptococcus suis* in complex with dTDP-xylose (Allard *et al.*, 2002; Blankenfeldt *et al.*, 2002) or *Arabidopsis thaliana* SQD1 with UDP-glucose bound (Mulichak *et al.*, 1999), a closer contact between Tyr and the sugar (about 2.5 Å) takes place. In both cases it is very possible that a low-barrier H-bond (LBHB) between Tyr and the sugar is part of the mechanism. In such a LBHB, one H⁺ is split between the oxygens of the Tyr and the sugar (Blankenfeldt *et al.*, 2002).

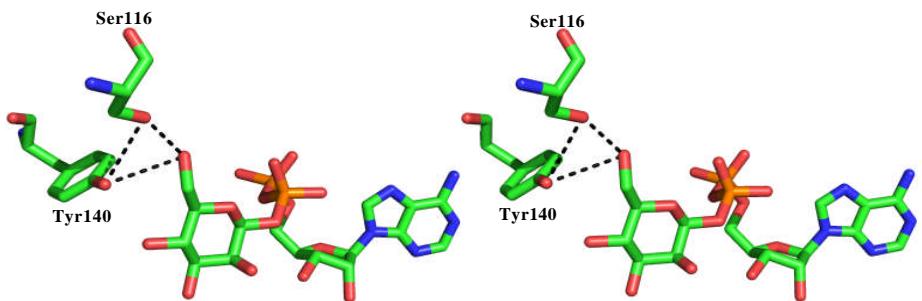


Figure 94. Stereo view showing the distances (dashed lines) between Tyr140 and Ser116 and between both residues and O6 of the sugar moiety of ADP- β -mannose in AGME. This figure has been produced using PyMol (DeLano, 2007).

No energy is needed then to move H^+ from one oxygen to the other (Cleland *et al.*, 1998; Blankenfeldt *et al.*, 2002). A LBHB connection between Tyr and the sugar would be very advantageous because it abolishes the barrier for H^+ -addition to a C=O connection or H^+ -abstraction from an OH-group (Blankenfeldt *et al.*, 2002). It is possible that in AGME Ser116 fulfils the role of “fine-tuning” (Blankenfeldt *et al.*, 2002) the transition state pK_a.

We propose there is only one base, Tyr140. An inspection of the crystal structure of AGME Y140F in complex with ADP- β -mannose reveals that the oxygen at position 6 of the mannose moiety displays an orientation similar to the natural product ADP-L-glycero-D-mannoheptose (Figure 95).

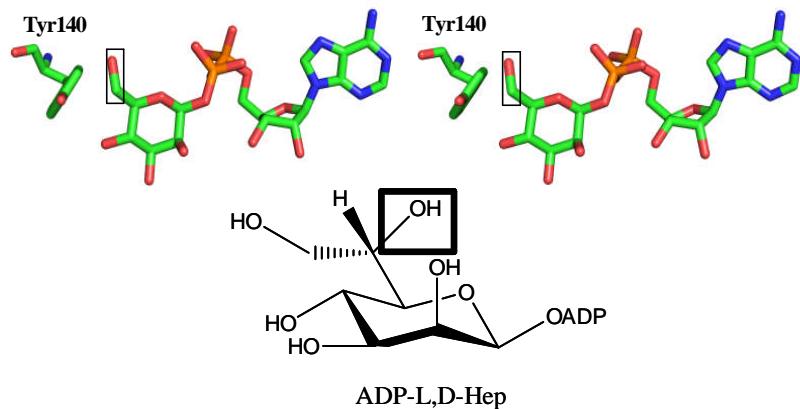


Figure 95. Stereoview of ADP- β -mannose and Tyr140 from AGME. ADP-L,D-Hep is shown as comparison. The C-6''-OH in all molecules is boxed. This figure has been produced using PyMol (DeLano, 2007).

If one rotates the C-5''-C-6'' bond of ADP- β -mannose by 120°, the orientation of O6 would mimic that seen in the natural substrate ADP-D-glycero-D-mannoheptose (Figure 96).

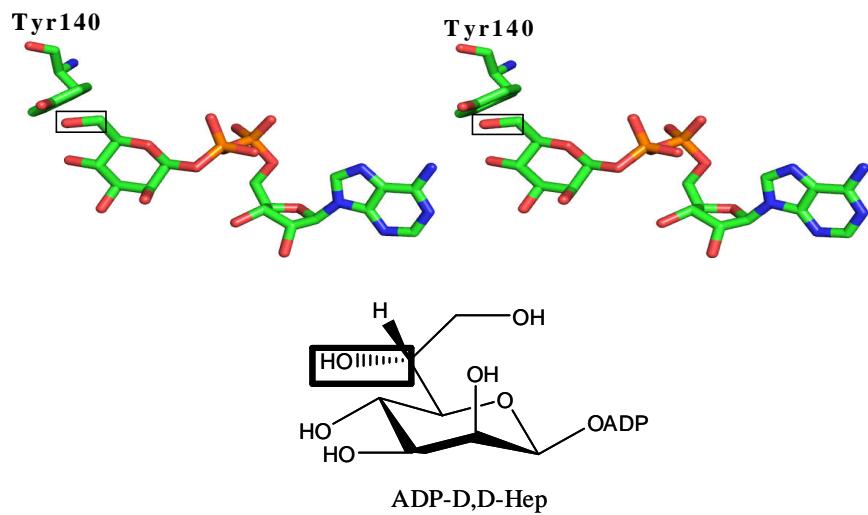


Figure 96. Stereo view of ADP- β -mannose after a 120° rotation about the C-5''-C-6'' bond and Tyr140 of AGME. ADP-D,D-Hep is shown as comparison. The C-6''-OH in all molecules is boxed. This figure has been produced using PyMol (DeLano, 2007).

First, a transient oxidation of ADP-D,D-Hep at C-6" takes place to form the 6"-ketone (Morrison and Tanner, 2007). Tyr140 (as tyrosinate) abstracts H⁺ from C-6-OH" of ADP-D,D-Hep which occurs together with H⁻ removal and transfer to NADP⁺ (Figure 97) to give NADPH and the keto-intermediate (Morrison and Tanner, 2007). The distance between OH of the phenolic side chain of Tyr 140 and O6 of the mannose moiety of ADP-β-mannose is only 2.1 Å (Figure 97) indicating that the tyrosine residue is appropriately located to perform oxidation of the substrate.

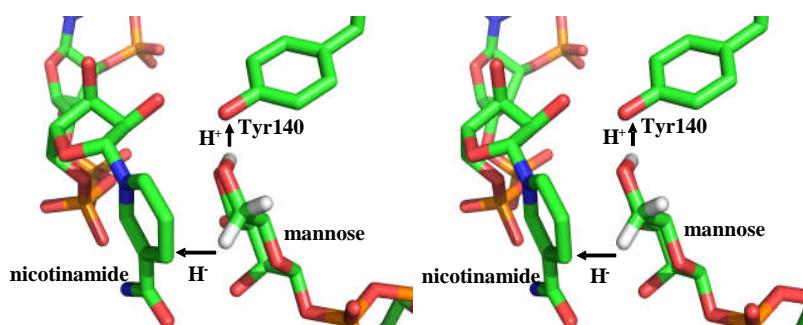


Figure 97. Stereo view of the contacts between C-6" of the sugar moiety of ADP-β-mannose (after a 120° rotation about the C-5"-C-6" bond) and C-4" of NADP⁺ and O6 of mannose and O of Tyr140 (2.1 Å). Hydrogens are shown in white, directions of proton and hydride transfers as arrows, and distances as dashed lines This figure has been produced using PyMol (DeLano, 2007).

Next, the C-5"-C-6" connection of the 6"-ketone is rotated and the opposite face of the C=O is shown to NADPH. At this point the Tyr would not be located in an appropriate way to add H⁺ to C=O (Morrison and Tanner, 2007) (Figure 98) and would therefore be required to change its position.

The catalytic region is large enough to allow this required movement of Tyr140 (Figure 99).

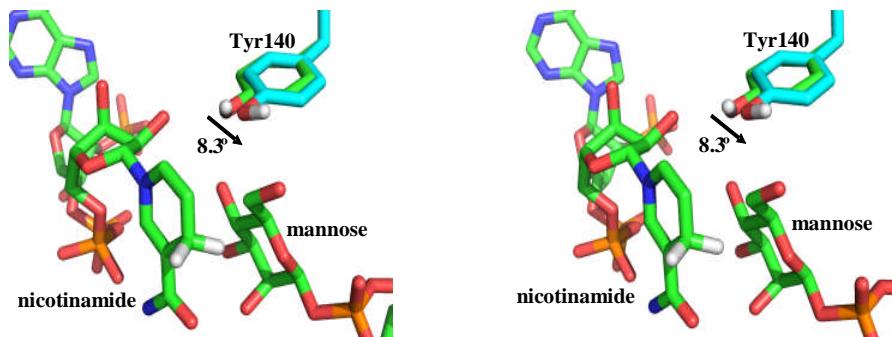


Figure 98. Stereo view of the same site as demonstrated in Figure 97. Hydrogens are shown in white. The required little movement of Tyr140 (only 8.3°) to accommodate a proper position for adding H⁺ to the C=O group of the intermediate (Morrison and Tanner, 2007) is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007).

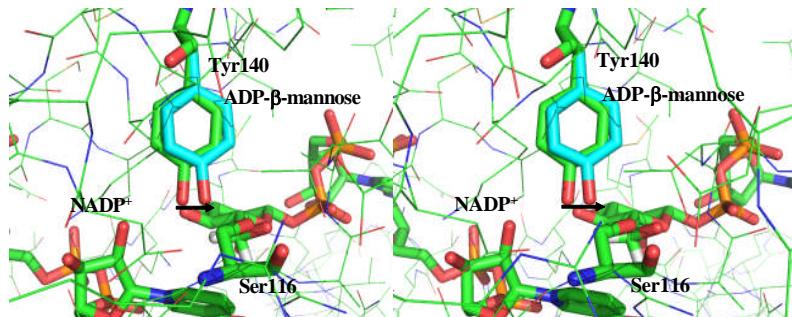


Figure 99. Stereo view of the active site of AGME. Tyr140 and atoms of the mannose moiety of ADP-β-mannose are shown as sticks. The required movement of the tyrosine residue into the other pocket of the catalytic region to be in an appropriate position to add H⁺ to the C=O group at position 6 of the rotated keto-intermediate (Morrison and Tanner, 2007) is shown as arrow. This movement would not cause any clashes of Tyr140 with any amino acids of the protein. Ser116, as most closely located residue, is now 2.4 Å from Tyr140. This figure has been produced using PyMol (DeLano, 2007).

Now, Tyr140 is properly located to protonate C=O which occurs together with H⁻ transfer from NADPH to C-6'' to generate NADP⁺ and ADP-L,D-Hep (Figure 100) (Morrison and Tanner, 2007).

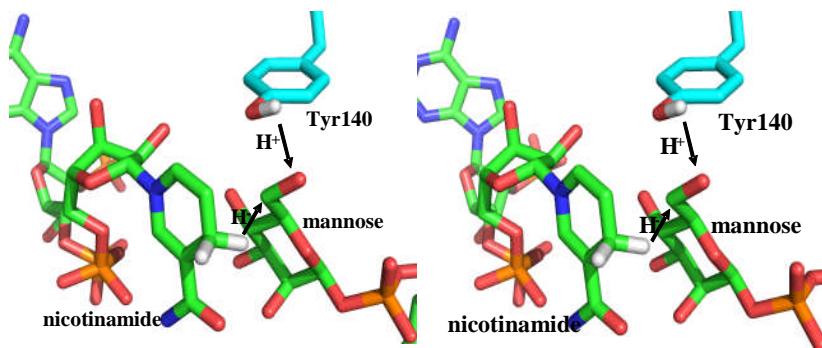


Figure 100. Stereo view of contacts between C-6'' of the sugar moiety of ADP- β -mannose and C-4'' of NADPH and the location of the rotated intermediate in respect to the moved Tyr140. Hydrogens are shown in white, directions of proton and hydride transfers as arrows. This figure has been produced using PyMol (DeLano, 2007)

The active site of AGME would also be large enough to accommodate the C-7''-OH of the substrate ADP-D,D-Hep (Figure 101A). C-7''-OH of ADP-D,D-Hep would not clash with any residues of the enzyme. The most closely located residue is Ser116 (Figure 101A) (2.6 Å between O7 of the substrate and OG of Ser116). However, in the case of ADP-L,D-Hep, O7 would be located too closely to Ser116 (2.2 Å), C-6'' of the nicotinamide ring (2.1 Å) and also Tyr140 (2 Å) (Figure 101B). For that reason small movements of either residues, niacinamide, or substrate would be required. A movement of the tyrosine of about 35° would be required to increase the distance to O7 of ADP-L,D-Hep (3.2 Å) (Figure 101B). However, Tyr140 would then clash with Ser116 and Ala118 (distances of 2 Å). For that reason a conformational change of the active site would be necessary.

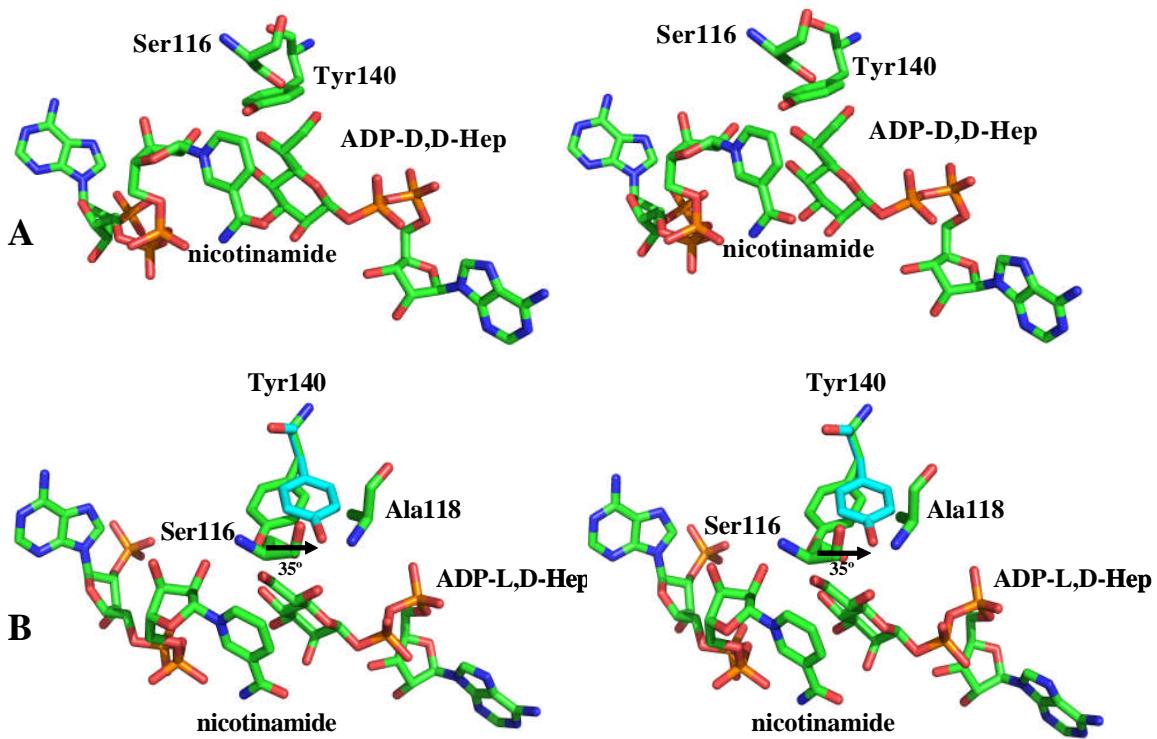


Figure 101. Stereo views of the interactions of ADP-D,D-Hep (A) and ADP-L,D-Hep (B) with residues of AGME. The nucleotide sugars, NADP and amino acids are shown as sticks. The 35° movement of Tyr140 is shown as an arrow. This figure has been produced using PyMol (DeLano, 2007)

An overview of the whole reaction is shown in Figure 102.

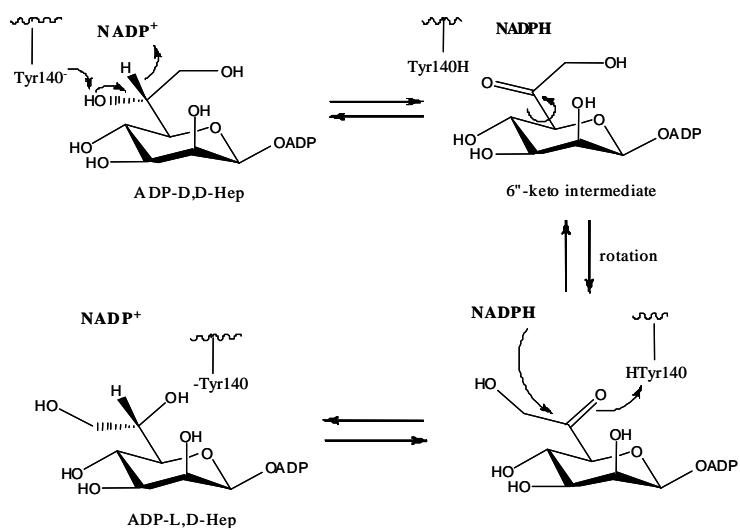


Figure 102. Proposed one-base mechanism for the reaction catalysed by AGME. Tyr140 is the catalytic site acid/base amino acid.

Future aims

The structure of AGME Y140F in complex with a substrate in a β -manno configuration, ADP- β -mannose, has been solved allowing us to propose a one-base mechanism for the epimerisation reaction catalysed by AGME. The role of Lys178 has not been elucidated yet. It could be involved in substrate binding. It could be attempted to crystallise the AGME K178M mutant in presence of ADP- β -mannose and/or ADP-D,D-Hep and/or the product ADP-L,D-Hep. A structure with one of those nucleotide sugars bound to the active site would indicate that Lys 178 is not important for sugar binding, whereas an empty active site would indicate that this lysine residue plays a crucial role in substrate binding. Further work could also include the design of inhibitors against AGME, to measure their inhibitory potencies and to try to get crystal structures of AGME/inhibitor complexes to study interactions of the inhibitors with this SDR enzyme.

CHAPTER 3

**STRUCTURES OF THE
CATALYTIC DOMAIN OF
STROMELYSIN-1 (SCD) IN
COMPLEX WITH NONPEPTIDE
INHIBITORS**

Introduction

3.1 Matrix metalloproteinases

Gross and Lapiere (Gross and Lapiere, 1962) discovered the first matrix metalloproteinase (MMP), collagenase, in the early 1960s (reviewed in Page-McCaw *et al.*, 2007). They discovered that there was a protein in tadpole tails during disambiguation which was able to degrade collagen in fibrils (Gross and Lapiere, 1962). Later, a collagenase of the interstitium, collagenase-1, was discovered (Birkedal-Hansen *et al.*, 1993). *In vitro*, collagenase-1 is involved in the degradation of collagens in fibrils, which are essential parts of the extracellular matrix (ECM) in vertebrates. This enzyme cuts amide bonds in collagens (Welgus *et al.*, 1983). After further studies had been performed the family of enzymes was found, named matrix metalloproteinases. The fact that high levels of matrix metalloproteinases were observed in various diseases in humans such as tumours and rheumatoid arthritis drove the interest in this protein family (see review Page-McCaw *et al.*, 2007). The fact that a higher level of matrix metalloproteinase activity can favour or block cancer development shows that the relationship between matrix metalloproteinases and disease is complex (reviewed in Page-McCaw *et al.*, 2007). Initially, it was assumed that matrix metalloproteinases are only involved in degradation of parts of the extracellular matrix but they also participate in other processes (reviewed in Page-McCaw *et al.*, 2007). They can play a role in cell movement, in the generation of “specific substrate-cleavage fragments” and in the control of tissue composition (Page-McCaw *et al.*, 2007). Moreover, they are also involved in the activation, deactivation or modification of signalling molecules (Figure 103) (Page-McCaw *et al.*, 2007).

Because of proteolysis of substrates of the extracellular matrix, carried out by matrix metalloproteinase, fragments which act distinctively compared to their precursor molecules are yielded. Examples for substrates of matrix metalloproteinases represent cell-adhesion molecules, tyrosine kinase receptors, chemokines, peptide growth factors, cytokines, other matrix metalloproteinases and proteolytic proteins (reviewed in Page-McCaw *et al.*, 2007).

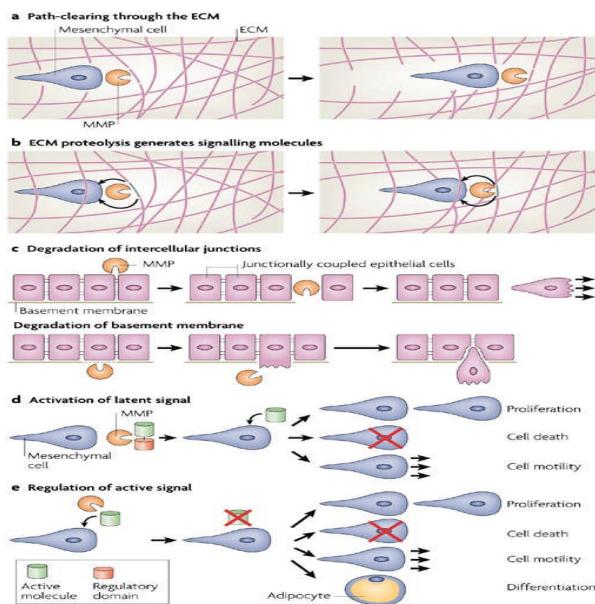


Figure 103. (taken from Page-McCaw *et al.*, 2007) Possible modes of matrix metalloproteinase action. Members of this enzyme family have the ability to cut parts of the extracellular matrix, which leads to more space for cells or tissues to migrate (a). Specific autocrine or paracrine signalling molecules can also be produced by cleavage carried out by matrix metalloproteinases (b). Matrix metalloproteinases are also able to directly control the composition of the tissue of the epithelium by cutting junctions between cells or the basement membrane (c). The action of cryptic signalling molecules can also be activated or altered by this family of proteinases which leads to various processes in cells (d). Deactivation or change of the function of signalling compounds which leads to modifications in cell death, proliferation, cell motility or differentiation can also be accomplished by matrix metalloproteinases (Page-McCaw *et al.*, 2007).

3.1.1 Structure of matrix metalloproteinases

Matrix metalloproteinases are referred to as “the matrixin subfamily of zinc metalloprotease family M10 in the MEROPS database (<http://www.merops.sanger.ac.uk/>)” (Nagase *et al.*, 2006). In general matrix metalloproteinases contain a prosegment of approximately 80 residues, a region of around 170 residues, responsible for catalysis, a linker or “hinge region” and a haemopexin (Hpx) region which is around 200 residues long. Highly conserved motifs in matrixins are the His-Glu-X-X-His-X-X-Gly-X-X-His cluster which interacts with Zn²⁺ and can be found in the catalytic region, and the “Cys-switch” fingerprint Pro-Arg-Cys-Gly-X-Pro-Asp which is part of the prosegment (Figure 104). The three His of the His-Glu-X-X-His-X-X-Gly-X-X-His cluster and the Cys of the “Cys-switch” sequence ligate the Zn²⁺ which is involved in catalysis. Because of the cysteine-zinc interaction pro-matrix metalloproteinases do not display any activity, as the H₂O is unable to interact with the catalytic Zn²⁺ (Nagase *et al.*, 2006). A “Met-turn” is present eight amino acids after the Zn²⁺ interacting region (Figure 104), which stabilises the structure surrounding this Zn²⁺ (Bode *et al.*, 1993; Nagase *et al.*, 2006).

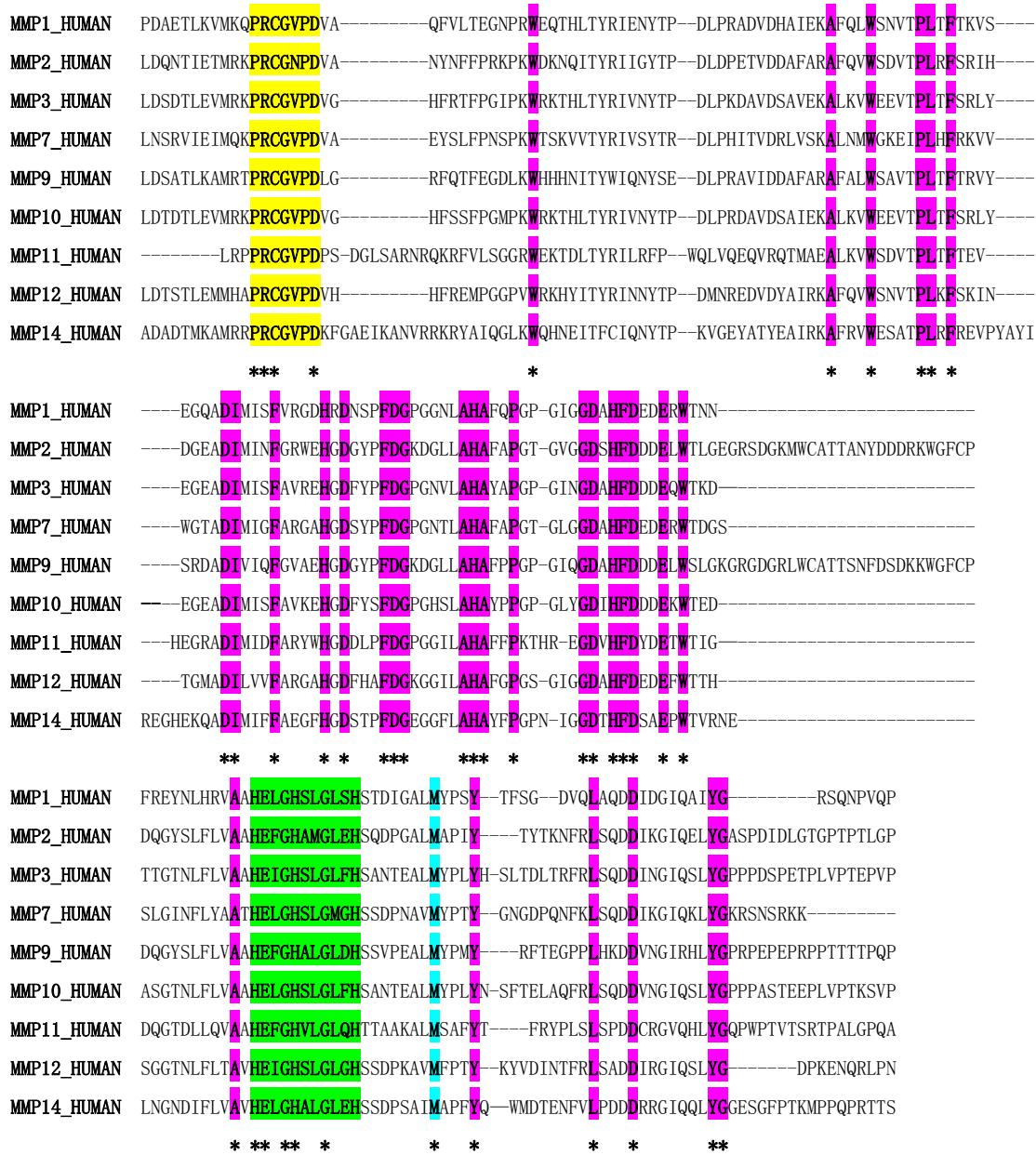


Figure 104. Structural sequence alignment of various MMPs (MMP-1, UniProtKB/TrEMBL entry P03956; MMP-2, UniProtKB/TrEMBL entry P08253; MMP-3, UniProtKB/TrEMBL entry P08254; MMP-7, UniProtKB/TrEMBL entry P09237; MMP-9, UniProtKB/TrEMBL entry P14780; MMP-10, UniProtKB/TrEMBL entry P09238; MMP-11, UniProtKB/TrEMBL entry P24347; MMP-12, UniProtKB/TrEMBL entry P39900; MMP-14, UniProtKB/TrEMBL entry P50281) from *Homo sapiens*. The conserved “Cys switch” fingerprint (Nagase *et al.*, 2006) sequence is highlighted in yellow, the Zn²⁺ associating motif in green, and the Met residue, 8 amino acids after the Zn²⁺ interacting cluster (Nagase *et al.*, 2006) in turquoise. Residues conserved in all nine matrix metalloproteinases are highlighted in magenta and marked with asterisks. This sequence alignment was carried out using the programme ClustalW (Chenna *et al.*, 2003).

The His-Glu-X-X-His-X-X-Gly-X-X-His cluster and the “Met-turn” also occur in other protein families such as ADAM (**a** disintegrin and metalloproteinase), ADAMTS (ADAM with thrombospondin fingerprint sequences), the astacins and serralysins (Nagase *et al.*, 2006). For that reason they are all named “metzincins” (Bode *et al.*, 1993).

The structures of the catalytic regions of matrixins are essentially identical. The secondary structural elements of this region are three α -helices and five β -strands which are linked by loops. One catalytic and one structural Zn^{2+} , and usually three Ca^{2+} which are responsible for the stabilisation of the protein are present in this domain. The prosegment of MMPs is composed of three α -helices and loops. The “Cys switch” is located in the region responsible for substrate interaction, however, the propeptide binds in a reverse way with respect to a peptide substrate (Nagase *et al.*, 2006). Cleavage of the prosegment allows the catalytic region to digest diverse substrates. The haemopexin region which occurs in the majority of matrixins consists of a “four-bladed β -propeller structure” that plays a role in associations between proteins and possibly substrate recognition (reviewed in Page-McCaw *et al.*, 2007). The haemopexin region also has an influence on protease activity and is involved in protease degradation as well (Overall, 2002).

3.1.2 Domains of matrix metalloproteinases

Due to different occurrence and arrangement of the domains (Figure 105) and the fact that matrix metalloproteinases favour distinct substrates, they can be divided into stromelysins, matrilysins, gelatinases, membrane-type (MT)-matrixins, collagenases and others (Visse and Nagase, 2003; Nagase *et al.*, 2006; Catania *et al.*, 2007). A detailed description of members of these families and their functions can be found in Nagase *et al.*, 2006.

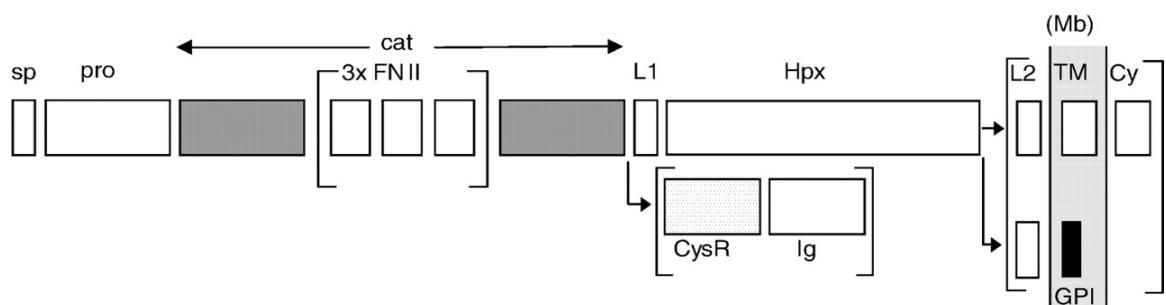


Figure 105. (taken from Nagase *et al.*, 2006) Organisation of the various domains of matrix metalloproteinases. sp: signal sequence; pro: prosegment; cat: region responsible for catalysis; FNII: fibronectin type II motif; L1: linker or “hinge-region”; Hpx: hemopexin region; CysR: cysteine rich; Ig: immunoglobulin part; L2: second connection; Mb: plasma membrane; TM: transmembrane domain; GPI: glycosylphosphatidylinositol anchor; Cy: tail located in the cytoplasm.

3.1.3 Stromelysin-1 (MMP-3)

Initially it was thought that MMP-3 participates in the synovitis cascade (Marcy *et al.*, 1991). The molecular weight of MMP-3 containing all domains is 52 kDa, and it needs zinc to be active (Chen *et al.*, 1999). Glycoproteins, several parts of the ECM such as fibronectin and laminin and also collagens and procollagens are targets of this matrixin (Welgus, 1991; Chen *et al.*, 1999). Furthermore MMP-3 can undergo autocatalysis and is also able to cleave a range of other members of the matrixin family (Nagase *et al.*, 1991; Chen *et al.*, 1999).

3.1.3.1 Structure of pro-stromelysin-1

Matrixins containing all their domains are flexible in their nature and for that reason it is not easy to obtain crystals of them (Dhanaraj *et al.*, 1996). Initially it has been reported that only entire MMP-1s from *Sus scrofa* (mature form) gave crystals of an appropriate stability to perform crystallographic studies. (Dhanaraj *et al.*, 1996; Li *et al.*, 1995). However, ten years later, the structure proMMP-1 from *Homo sapiens* was solved comprising the prosegment, catalytic region, hinge peptide and haemopexin region (Jozic *et al.*, 2005). However, most X-ray structures of MMPs such as stromelysin-1 (Kohno *et al.*, 2006; Pikul *et al.*, 2001; Dunten *et al.*, 2001; Cheng *et al.*, 2000; Steele *et al.*, 2000; Pavlovsky *et al.*, 1999; Chen *et al.*, 1999; Dhanaraj *et al.*, 1996; Becker *et al.*, 1995) are only of the catalytic domain. In the case of MMP-3 all published structures are those of the catalytic domain only except the one published by Becker *et al.*, 1995 which contains the propeptide.

A structure of stromelysin-1 including the Hpx domain has not been published yet. In prostromelysin-1, the structure of the segment, responsible for catalysis is almost identical to the catalytic region of the mature enzyme (Becker *et al.*, 1995). The prosegment of pro-stromelysin-1 is, as in other proMMPs (Nagase *et al.*, 2006), such as pro-MMP-1 (Jozic *et al.*, 2005) composed of three α -helices and a prosegment (Becker *et al.*, 1995) (Figures 107 and 108A and B).

In proMMP-3 this region contains amino acids 16-82. The first 15 amino acids and amino acids 31-39 are disordered (Becker *et al.*, 1995).

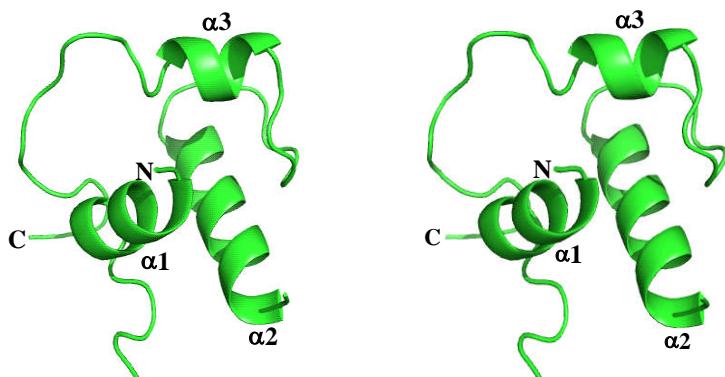


Figure 107. Stereo ribbon diagram of the prosegment of pro-MMP-3 (Pdb accession code 1slm). This figure has been produced using PyMol (DeLano, 2007).

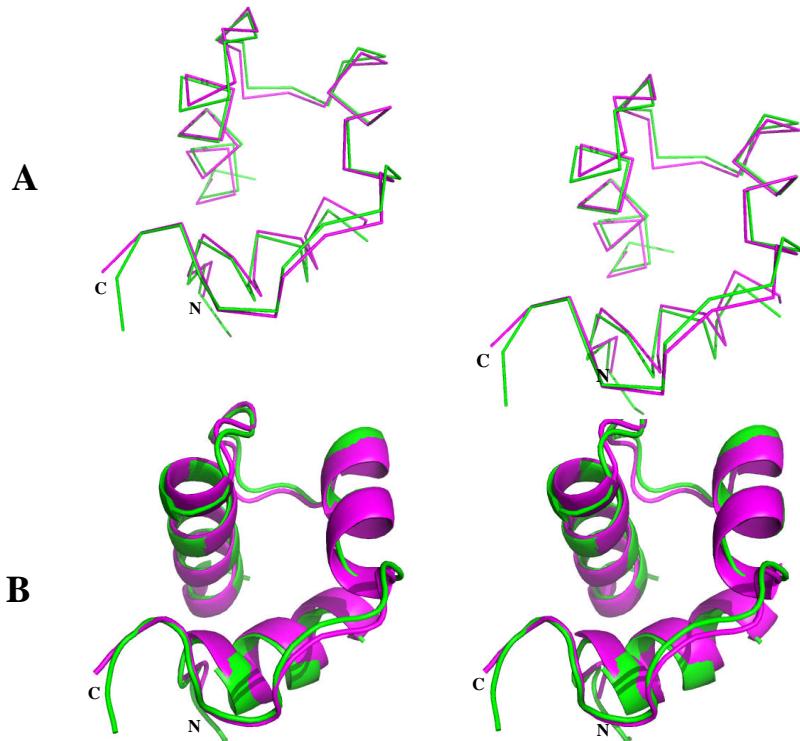


Figure 108. (A) Stereo view of $\text{C}\alpha$ tracing of the prodomains of pro-MMP-3 from *Homo sapiens* (in green, Pdb accession code 1slm) and human pro-MMP-1 (in magenta, Pdb accession code 1su3) after their superposition. (B) Stereo ribbon diagrams of both structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

The rms deviation between both prosegments is 0.65 Å for 52 $\text{C}\alpha$ positions. The prosegment occupies the catalytic region of pro-stromelysin-1, and there is an association between Cys75 and the Zn^{2+} crucial for catalysis (Figure 109) (Becker *et al.*, 1995).

Interestingly, the propeptide interacts in an opposite way to propeptide-derived blockers bound to stromelysin and collagenase (Becker *et al.*, 1995). The interactions of peptide-based blockers will be discussed in detail in section 3.1.3.3. The His-82-Phe83 connection is cut and processed stromelysin-1 generated (Cameron *et al.*, 1995; Becker *et al.*, 1995).

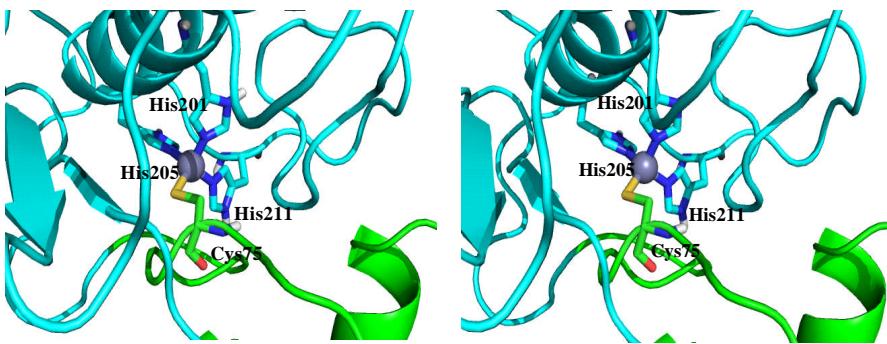


Figure 109. Stereo view of the region around the catalytic Zn^{2+} in pro-stromelysin-1 (Pdb accession code 1slm). Secondary structural elements of the prosegment and the catalytic domain are coloured in green and cyan, respectively. Residues from each of the two domains ligated with Zn^{2+} (grey sphere) are shown as sticks. This figure has been produced using PyMol (DeLano, 2007).

3.1.3.2 Structure of the catalytic domain of stromelysin-1

“Recombinant human truncated stromelysin (RHTS)” (Chen *et al.*, 1999) comprises the catalytic region of MMP-3. It comprises the region from Phe83 to Thr255 and its molecular weight is about 20 kDa (Chen *et al.*, 1999). RHTS displays similarity to stromelysin-1 containing all domains in regard to enzymatic activity, preference and susceptibility to blocking compounds (Marcy *et al.*, 1991; Ye *et al.*, 1992; Chen *et al.*, 1999).

The catalytic domain of stromelysin-1 displays “an α/β -type” (Chen *et al.*, 1999) fold common to all members of the matrixin family (Figure 110) (Chen *et al.*, 1999).

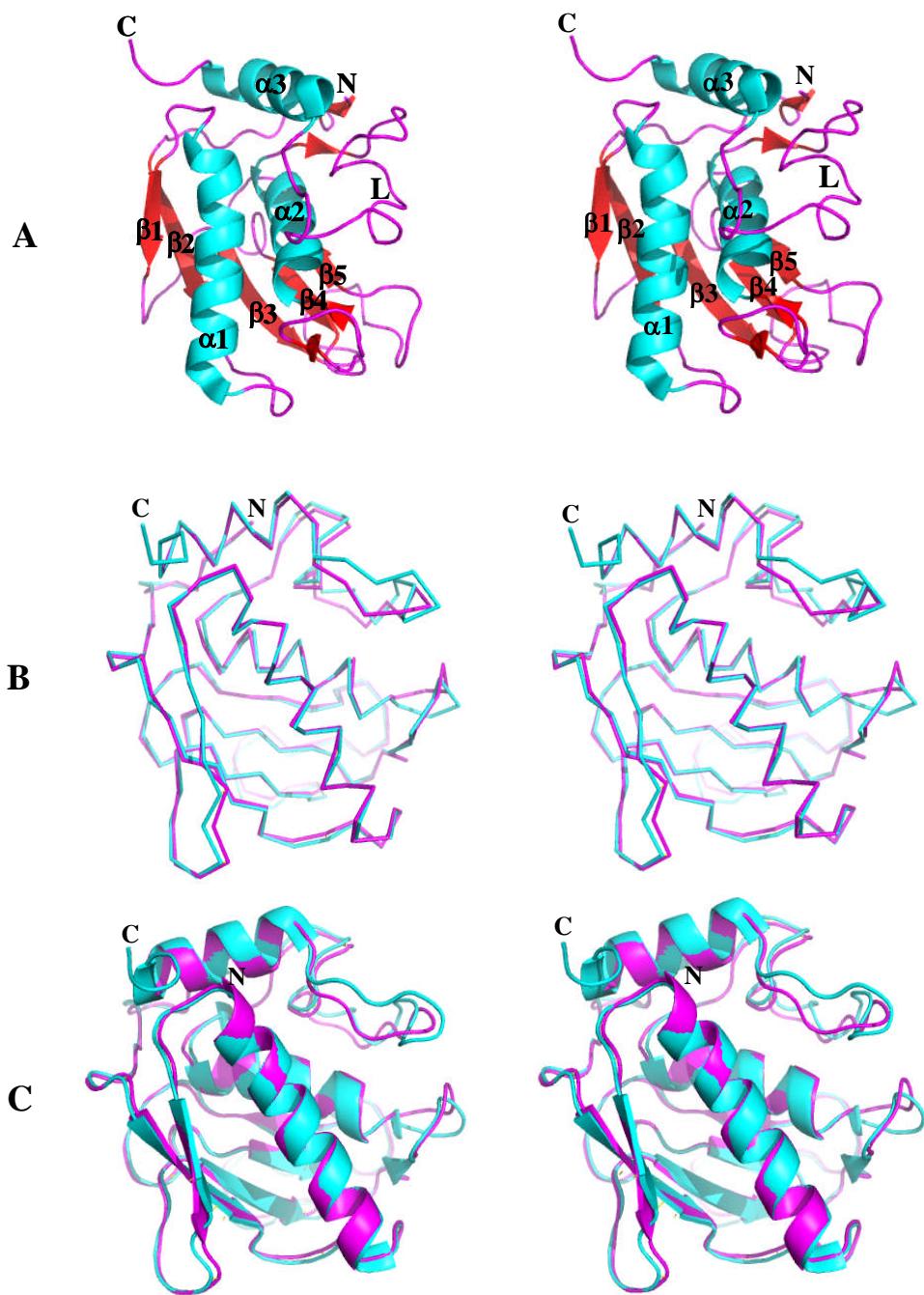


Figure 110. (A) Stereo view of the structure of RHTS (Pdb accession code 1cqr) (based on Chen *et al.*, 1999). α -helices are shown in cyan, β -sheets in red and loops in magenta (flexible loop from amino acids 210-234 (Chen *et al.*, 1999) labelled as L)). (B) C- α tracing of the catalytic domain of stromelysin-1 from *Homo sapiens* (in cyan, Pdb accession code 1cqr) and porcine MMP-1 (in magenta, Pdb accession code 1fbl) after their superposition. The rms deviation between both structures is 0.49 Å for 141 C α positions. (C) Ribbon stereo diagrams of both structures after their superposition. This figure has been produced using PyMol (DeLano, 2007).

3.1.3.2.1 Location of metal ion sites

The catalytic domain of stromelysin-1 contains two zinc ions. One is involved in catalysis and the second one is the structural Zn^{2+} (Dhanaraj *et al.*, 1996). The catalytic Zn^{2+} is ligated with NE2 of His201, His205 and His211 (Dhanaraj *et al.*, 1996). The distances between those nitrogens and Zn^{2+} vary between 2.1-2.2 Å (Figure 111). It is also possible that additionally to the interactions with the three histidines this zinc ion can also bind to one or two other atoms. These atoms can be part of either another protein or H_2O , or groups accomplished by a substrate or inhibitor (Chen *et al.*, 1999).

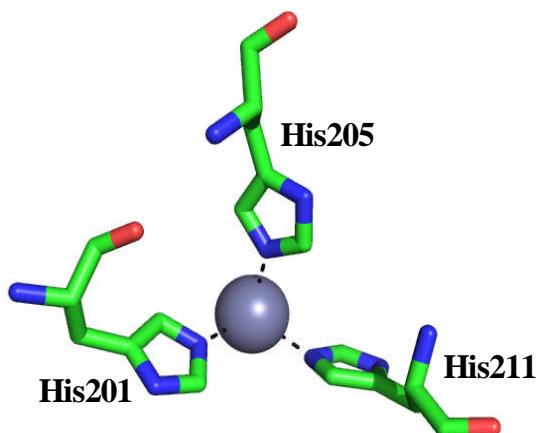


Figure 111. Close up view of the interactions between NE2 of His201, His205 and His211 and the catalytic Zn^{2+} in RHTS (Chen *et al.*, 1999) (Pdb accession code 1cqr). The three histidines are shown as sticks, and the catalytic zinc as grey sphere and distances as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

The structural Zn^{2+} is situated between the third and fourth β -helix and is ligated by NE2 of His 151 and His 166, ND1 of His179 and OD2 of Asp153 (Figure 112) (Dhanaraj *et al.*, 1996).

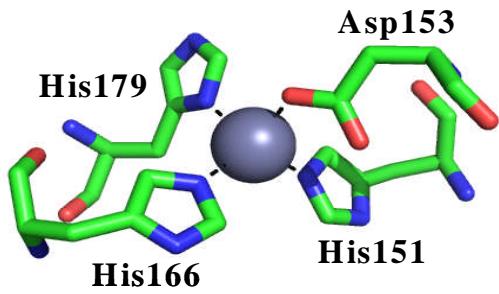


Figure 112. Close up view of the interactions of NE2 of His151 and His166, ND1 of His179 and OD2 of Asp153 with the structural Zn²⁺ in RHTS (Pdb accession code 1cqr) (Chen *et al.*, 1999). The three histidines and the aspartic acid are shown as sticks, the catalytic zinc as grey sphere and distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

Matrixins need Ca²⁺ to be active (Marcy *et al.*, 1991) and the majority contain three Ca²⁺ in their catalytic region (Nagase *et al.*, 2006; Chen *et al.*, 1999). In MMP-3, Ca1 is ligated by OD1 of Asp158, the oxygens of Gly159, Gly161 and Val163, OD2 of Asp181 and OE2 of Glu184 (Chen *et al.*, 1999; Dhanaraj *et al.*, 1996). Ca2 is ligated by OD1 of Asp107 and Asp182, O of Asp182 and Glu184 and two H₂O (Figure 113B) (Chen *et al.*, 1999). Ca3 is in close contact with O-atoms provided by Asp141, Gly173, Asn175 and Asp177 (Figure 113C) (Chen *et al.*, 1999).

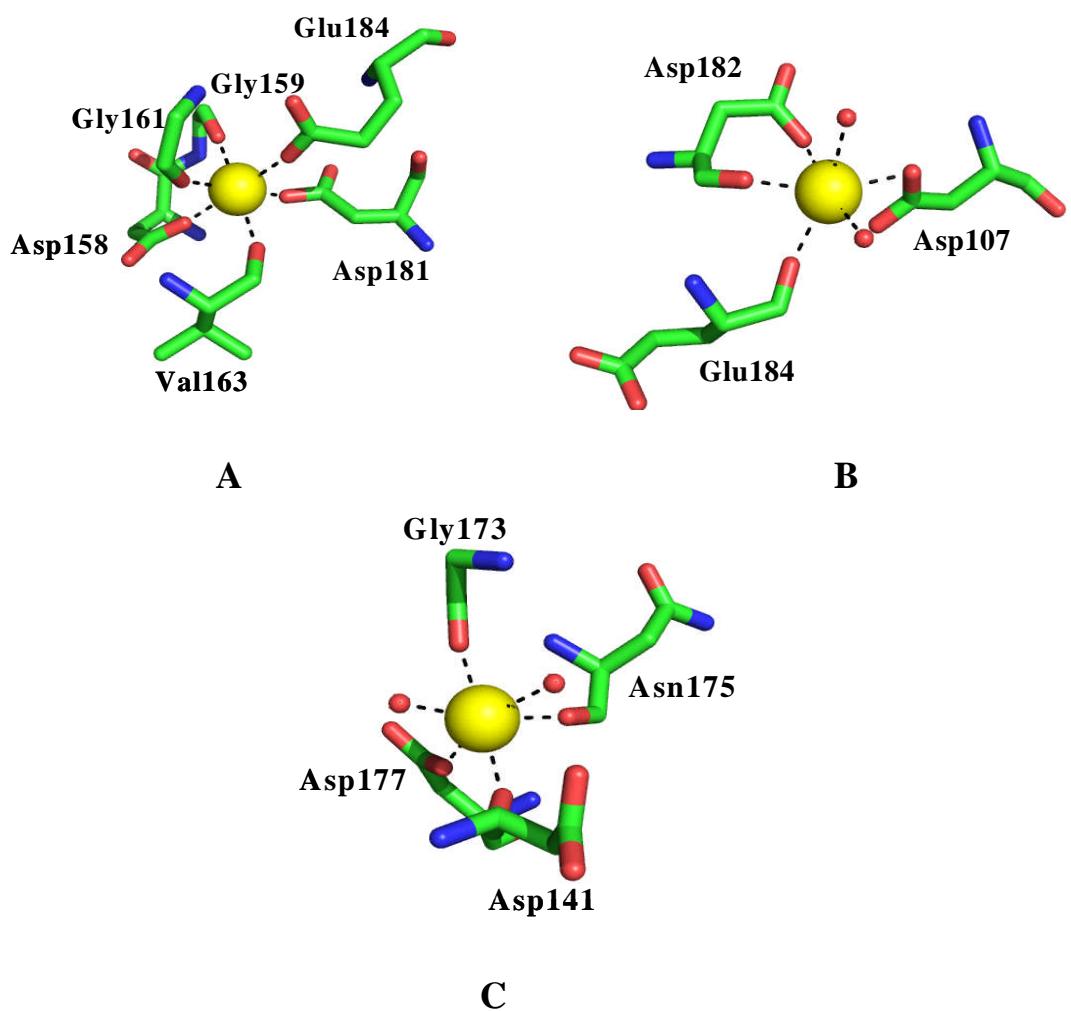


Figure 113. (A) Close up views of the interactions of OD1 of Asp158, the O-atoms of Gly159, Gly161 and Val163, OD2 of Asp181 and OE2 of Glu184 with Ca1 (Chen *et al.*, 1999; Dhanaraj *et al.*, 1996), (B) the interactions of OD1 of Asp107 and Asp182, the O-atoms of Asp182 and Glu184 and the two H₂O with Ca2 (Chen *et al.*, 1999), and (C) the interactions of OD1 of Asp177, O of Asp141, Gly173 and Asn175 and the two H₂O with Ca3 in RHTS (Chen *et al.*, 1999) (Pdb accession code 1cqr). Ligating residues are shown as sticks, the calcium ion as yellow sphere, waters as red spheres and distances as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

3.1.3.2.2 The active site and subsites of the catalytic domain of MMP-3

The catalytic region of MMP-3 including Zn²⁺ is situated in a big crevice (Figure 110A) surrounded by α 2, β 5 and a loop which shows flexibility (amino acids 210-234). His201 and His205 which coordinate the Zn²⁺ belong to α 2. β 5 can participate in the interaction with an inhibitor. The subsite S1' is formed by the loop. (Chen *et al.*, 1999). The S1' subsite is in general non-polar but its size is different depending on the particular matrixin. It has a great influence on the substrate preference. When a substrate interacts with the protein, the H₂O bound to Zn²⁺ moves. COO⁻ of Glu in the catalytic region enables the cleavage of the amide bond. It takes H⁺ from the shifted H₂O and consequently, the polarized H₂O can perform a nucleophilic attack on the C of C=O of the amide bond (Nagase *et al.*, 2006). Inhibitors interact with Zn²⁺ and have other regions denoted P1, P2, P1', P2' which match sites S1, S2, S1', S2' on the protein (Chen *et al.*, 1999).

In stromelysin-1 the S1' subsite is bulky and broad (Figure 114). It is built up by nonpolar amino acids such as Leu164, Leu197, Val198, Leu218, Met219, Pro221 and Leu222. Other residues which contribute to the formation of this pocket are His201 and Tyr223. Consequently, the P1' moiety of an inhibitor should not display any polarity and because of the large size of this cavity it can also be quite long. Crucial amino acids which bind to a P1' moiety are Leu164, Leu197, Val198, His201, Pro221, Leu222 and Tyr223. The S2' subsite of MMP-3 is also hydrophobic. Amino acids involved in the formation of this substrate binding site are Asn162, Val163 and Leu164. P2' and also P3' moieties can be comparatively hydrophilic because they point to the solvent. The P3' moiety is normally quite bulky in molecules which block matrixin 1, however, the same group does not need this feature in stromelysin-1 inhibitors (Chen *et al.*, 1999).

The S1 and S3 substrate binding sites (Figure 114) which belong to the “unprimed side” (Finzel *et al.*, 1998) (S1-S4) essentially create “one large open space” (Finzel *et al.*, 1998) which is made up by amino acids Tyr155, His166, Ala167 and Tyr168 (Finzel *et al.*, 1998). One wall of the S1 subsite is formed by the Phe residues at positions 86 and 210, His205, His211 and Zn²⁺. Tyr155 and Tyr168 form the other wall. The bottom of this substrate binding site is built up by amino acids Ala165-Tyr168 (Pavlovsky *et al.*, 1999). The S2 subsite (Figure 114), a “shallow depression” (Finzel *et al.*, 1998) is built up by His205, Phe210 and Phe86. Phe86, Pro87, Pro90 and Ala169 form the substrate binding site S4 (Finzel *et al.*, 1998).

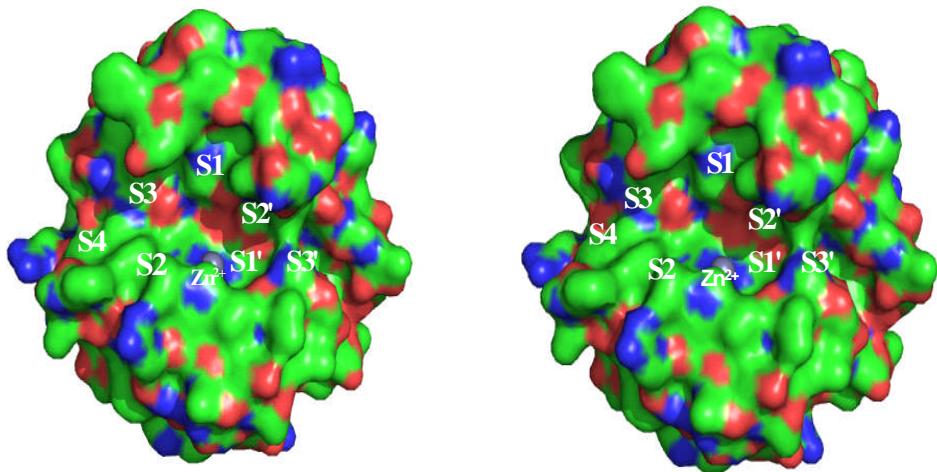


Figure 114. (based on Finzel *et al.*, 1998) A surface diagram to show the location of the subsites S1, S2, S3, S4, S1', S2' and S3' in the catalytic domain of stromelysin-1 (Pdb accession code 1usn) (shown as surface). Carbons are shown in green, oxygens in red and nitrogen in blue. The catalytic zinc ion is shown as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

3.1.3.3 Inhibitors of stromelysin-1

Insights into the blocking mechanism of endogenous tissue inhibitors of metalloproteinases (TIMPs) have been provided by several X-ray structures of matrixins in complex with them. These blockers are usually “wedge-shaped (Figure 115A) (Gomis-Rüth *et al.*, 1997; Fernandez-Catalan *et al.*, 1998). In the structure of the catalytic domain of MMP-3 complexed with TIMP-1 (Gomis-Rüth *et al.*, 1997), the first four amino acids of the protein (Cys1-Val4) and the amino acids Glu67-Cys70 are connected by a S-S bridge from an adjacent backbone which plugs into the catalytic region of the matrixin (Nagase *et al.*, 2006; Gomis-Rüth *et al.*, 1997). The α -NH₂- and the C=O group of Cys1 interact with the catalytic Zn²⁺ (Gomis-Rüth *et al.*, 1997). Additionally, H-bonds could also be generated between this α -NH₂-group and O of C=O of Ser 68 and O of COO⁻ of Glu202 of stromelysin-1 (Gomis-Rüth *et al.*, 1997) (Figure 115B) The side chain of Thr2 points into the S1' subsite (Gomis-Rüth *et al.*, 1997). If this amino acid is substituted, binding of the inhibitor to a matrixin will be altered, for instance, mutation to Gly at this position leads to an inactive inhibitor (Meng *et al.*, 1999).

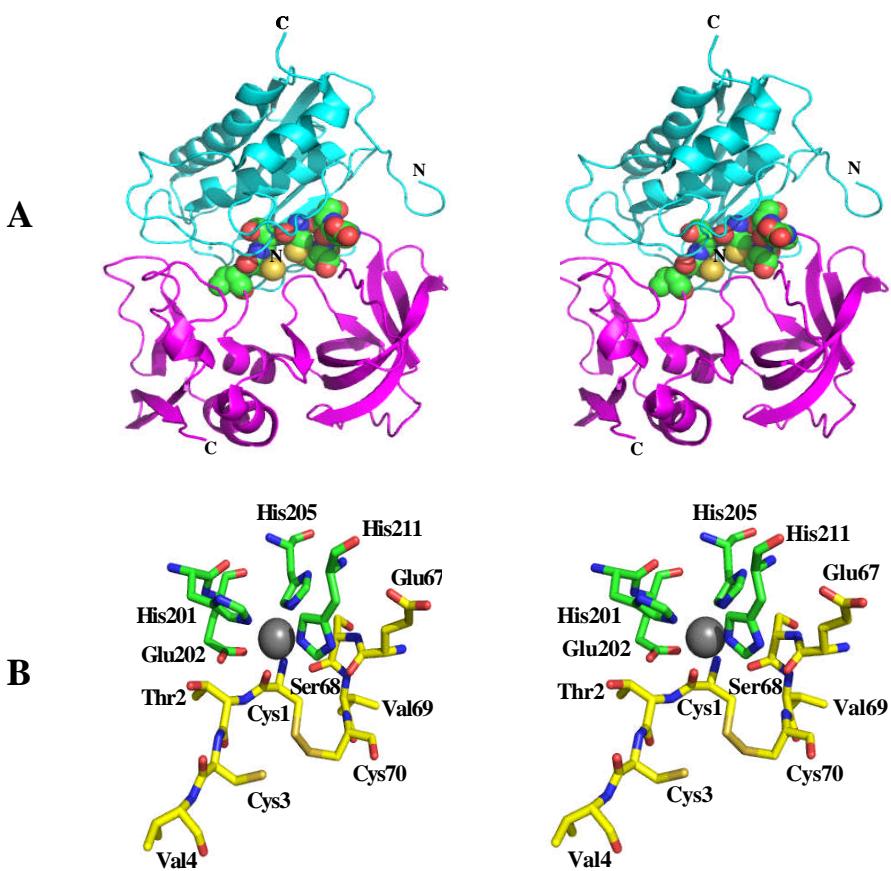


Figure 115. (A) (based on Nagase *et al.*, 2006; Gomis-Rüth *et al.*, 1997) Stereo ribbon diagram of the catalytic region of matrixin 3 (cyan) complexed with TIMP-1 (magenta) (Pdb accession code 1uea). Amino acids Cys1-Val4 and Glu67-Cys70 of the inhibitor are shown as spheres (based on Gomis-Rüth *et al.*, 1997 and Nagase *et al.*, 2006). (B) (based on Gomis-Rüth *et al.*, 1997) Stereo ribbon diagram of the residues in the vicinity of the catalytic Zn^{2+} . The three conserved histidines of MMP-3 and N and O of Cys1 of the inhibitor interact with the catalytic Zn^{2+} (grey sphere) (Gomis-Rüth *et al.*, 1997). Residues of the matrixin (carbons in green) and TIMP-1 (carbons in yellow) are shown as sticks. This figure has been produced using PyMol (DeLano, 2007).

TIMPs are native blockers of matrixins (Steele *et al.*, 2000) which bind their targets in a 1:1 ratio (Willenbrock and Murphy, 1994; Gomis-Rüth *et al.*, 1997; Steele *et al.*, 2000).

Because of their size (Steele *et al.*, 2000) (184-194 amino acids) (Nagase *et al.*, 2006) and broad inhibitory spectrum (Gomis-Rüth *et al.*, 1997) their use in therapy is restricted (Steele *et al.*, 2000). That is why the design of inhibitors of MMP-3 and other matrixins which are small and more specific is desirable (Steele *et al.*, 2000). The first synthetic inhibitors displayed a structure derived from the prodomain of matrixins (Fotouhi *et al.*, 1994; Steele *et al.*, 2000). It has been shown that this kind of inhibitors interact with the catalytic site of a particular MMP in an opposite direction with respect to the prosegment (Becker *et al.*, 1995; Dhanaraj *et al.*, 1996; Steele *et al.*, 2000). The inhibitors normally have a moiety such as OP(OR)R₂, OP(OR)₂R COO⁻ or R-CO-NH-O⁻, which coordinates the catalytic Zn²⁺ (Morphy *et al.*, 1995; Pavlovsky *et al.*, 1999).

An example for a complex structure of the catalytic domain of MMP-3 and a propeptide derived inhibitor was reported by Steele *et al.*, 2000. The bound inhibitor is named Ro-26-2812, contains a diphenyl group and the prosegment derived tripeptide Cys-Gly-Val, and has a half maximal inhibitory concentration of 1.6 μM. It interacts with the S1'- S3' subsites of the enzyme (Steele *et al.*, 2000) (Figure 116).

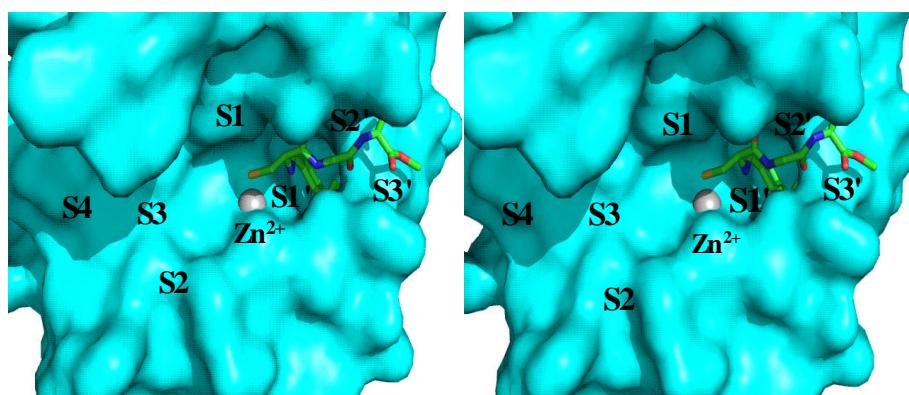


Figure 116. Surface representation of the interactions of Ro-26-2812 with subsites of the catalytic domain of stromelysin-1 (Pdb accession code 1c3i) The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Zn^{2+} ligates with the S-atom of the Cys-group of Ro-26-2812 (Figures 116 and 117). Its phenyl benzene moiety fills the S1' cavity (Figures 116 and 117). O of the COOH group of the Cys group of Ro-26-2812 generates a H-bond with N of Leu164 (2.6 Å) (Figure 117). O of the COOH group of the Gly of Ro-26-2812 forms a H-bond with N of Tyr223 and the amide of the same group is in close contact with O of Pro221 (Figure 117). The amide of the Val of Ro-26-2812 generates a H-bond with O of the COOH group of Asn162 (Steele *et al.*, 2000) (Figure 117).

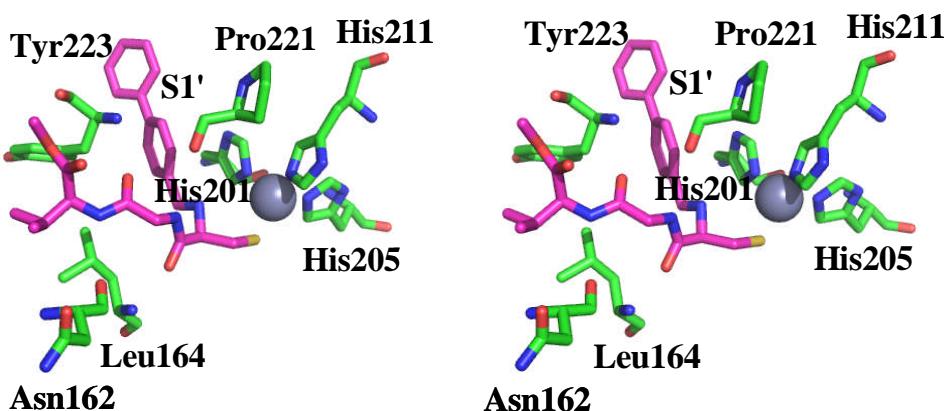


Figure 117. (based on Steele *et al.*, 2000) Stereo view of close interactions of Ro-26-2812 with residues of the catalytic domain of stromelysin-1 (Pdb accession code 1c3i). Amino acids (carbons in green) and the inhibitor (C-atoms in magenta) are shown as sticks, the catalytic Zn^{2+} as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Unlike Ro-26-2812, the Cys-Gly-Val tripeptide of the prosegment (Becker *et al.*, 1995) interacts with different subsites of MMP-3 (Steele *et al.*, 2000). Gly and Val bind to the S1 and S2 substrate binding sites (Steele *et al.*, 2000) whereas the S1' substrate binding site is not occupied (Becker *et al.*, 1995). Unlike in the Ro-26-2812/MMP-3 complex, there are no H-bonds between O-atoms of any COOH group in the tripeptide and the enzyme. The amide of Val does not generate any H-bonds either (Steele *et al.*, 2000) (Figure 118).

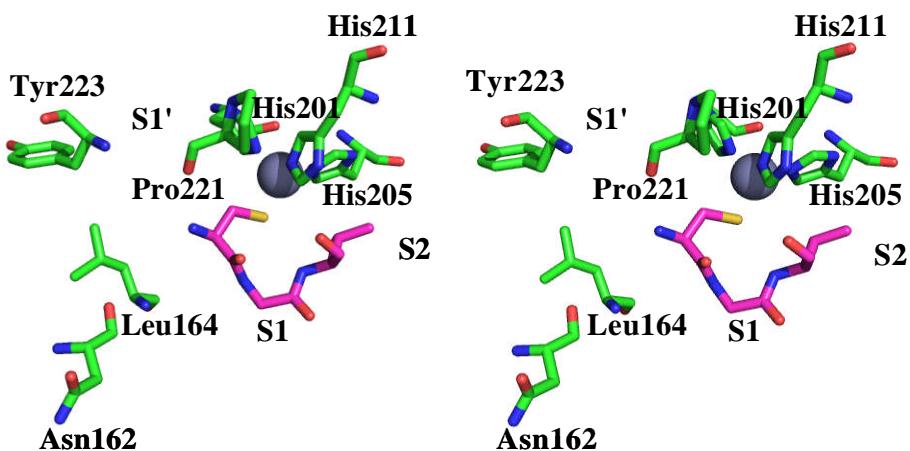


Figure 118. Stereo view showing the binding mode of the Cys-Gly-Val sequence of the propeptide to the catalytic domain of stromelysin-1 (Pdb accession code 1slm). Amino acids (carbons in green) and the tripeptide (C-atoms in magenta) are shown as sticks, the catalytic Zn^{2+} as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

If one compares the orientation and interactions of Ro-26-2812 to other peptide based inhibitors of the catalytic domain of stromelysin-1 (Becker *et al.*, 1995; Dhanaraj *et al.*, 1996; Esser *et al.*, 1997) they all bind in a similar fashion. In the structure of MMP-3 in complex with the hydroxamate inhibitor PD-140798, the subsites S1'-S3' are also occupied by the various moieties of the inhibitor. The substrate binding sites S1' and S2' are filled by leucine groups whereas a phenylalanine moiety points into the S3' subsite. Two O-atoms of the hydroxamic acid moiety of PD-140798 coordinate with the catalytic Zn^{2+} (Dhanaraj *et al.*, 1996). Becker and coworkers determined the crystal structure of the catalytic domain of stromelysin-1 complexed with the N-carboxyalkyl peptide inhibitor **I** (Figure 118) ($K_i = 0.23 \mu M$) (Becker *et al.*, 1995; Chapman *et al.*, 1993). The S1' substrate binding site is occupied by a homophenylalanine moiety (Becker *et al.*, 1995). The P2' group of this inhibitor is an Arg, the side chain of which is exposed to the solvent, the P3' moiety is an anilide (Gooley *et al.*, 1994). The COOH group of N-carboxyalkyl peptide inhibitor **I** interacts with the catalytic Zn^{2+} (Becker *et al.*, 1995). Based on this complex a series of more potent carboxyalkylpeptides was synthesized (Esser *et al.*, 1997).

An X-ray structure complexed with one of those inhibitors, L-764,004 (stromelysin-1-K_i = 2 nM) (Figure 119) was determined. It contains a 4-fluorodiphenylethyl moiety as P1', an iso-butyl moiety as P2' (Esser *et al.*, 1997) and the same anilide moiety as P3' group (Gooley *et al.*, 1994). Additionally a C_nH_{2n+1} chain with a benzolactam functions as P1 group (Esser *et al.*, 1997).

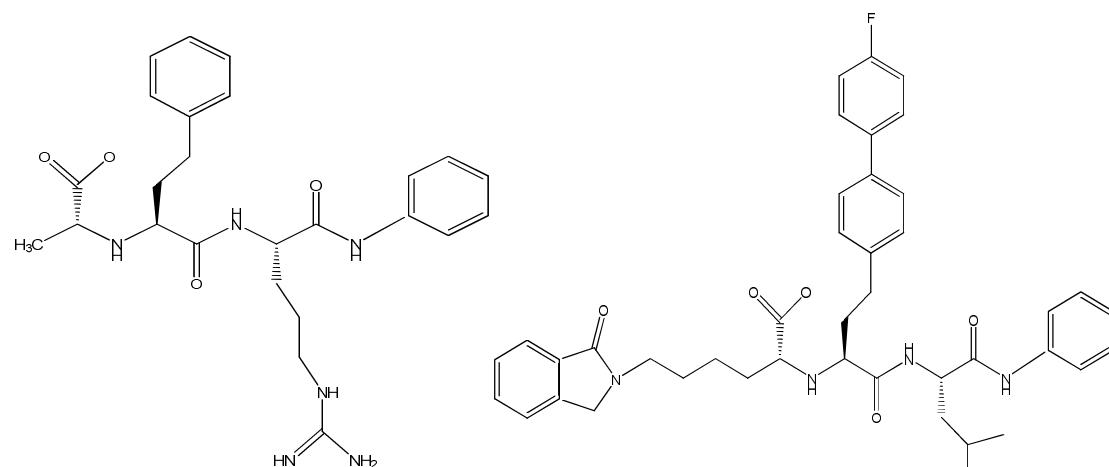


Figure 119. (after Becker *et al.*, 1995; Esser *et al.*, 1997) Chemical structures of the N-carboxyalkyl inhibitors I (Becker *et al.*, 1995; Chapman *et al.*, 1993) (left) and L-764,004 (Esser *et al.*, 1997) (right).

It has been shown that the fluorinated diphenylethyl moiety as P1' and C_nH_{2n+1} moieties as P1 groups raised the specific potency of this inhibitor for stromelysin-1 and also enhanced its activity in a mouse pleural cavity assay. After the secondary amine had been substituted with R₂C, and a (CH₃)₃C-group as P2' and a CH₃-group as P3' moiety had been introduced (Figure 120), the oral activity of the compound was increased (Esser *et al.*, 1997).

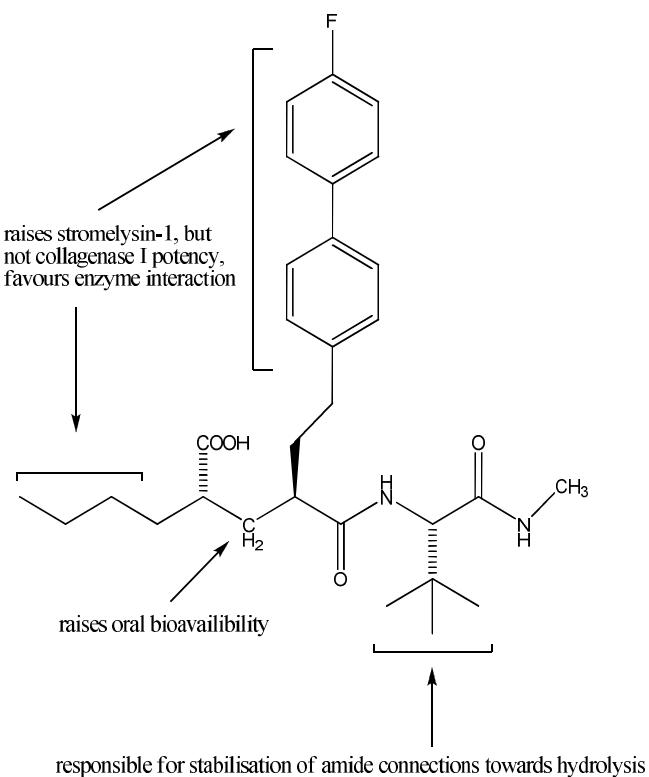


Figure 120. (after Esser *et al.*, 1997) Overview of the structure-activity relationship for P1'-lemonene-ethyl carboxylate blockers of stromelysin-1.

In general, many peptide based inhibitors are only very potent *in vitro*, block a range of matrixins (Pavlovsky *et al.*, 1999; Finzel *et al.*, 1998) and do not display high bioavailability (Pavlovsky *et al.*, 1999) or good solubility or long serum half life, thus are unsuitable for therapeutical applications (Finzel *et al.*, 1998). For that reason various so called “nonpeptide inhibitors” have been designed (Pavlovsky *et al.*, 1999; Finzel *et al.*, 1998). The design of most of these compounds is based on interactions with the S1' substrate binding site (Finzel *et al.*, 1998). Pavlovsky and coworkers determined four X-ray structures of stromelysin-1/nonpeptide inhibitor complexes (Pavlovsky *et al.*, 1999). All four inhibitors display specificity for MMP-3 (Ortwine *et al.*, 1997; White *et al.*, 1997; Pavlovsky *et al.*, 1999).

They all have a diphenyl piperidine moiety as P1' group. In the structures solved with the sulfonamide inhibitors I-III (Figure 121) bound, one O of the C=O group ligates with the catalytic Zn^{2+} . In the fourth structure with the hydroxamic acid based inhibitor IV (Figure 121) two O-atoms chelate with this zinc ion. Unlike inhibitors I and IV, inhibitors II and III contain an additional indole ring as P1 group. The interactions of a 1-benzazole ring with the S1 substrate binding site, however, do not appear to have a positive effect on K_i (Pavlovsky *et al.*, 1999) (Figure 121).

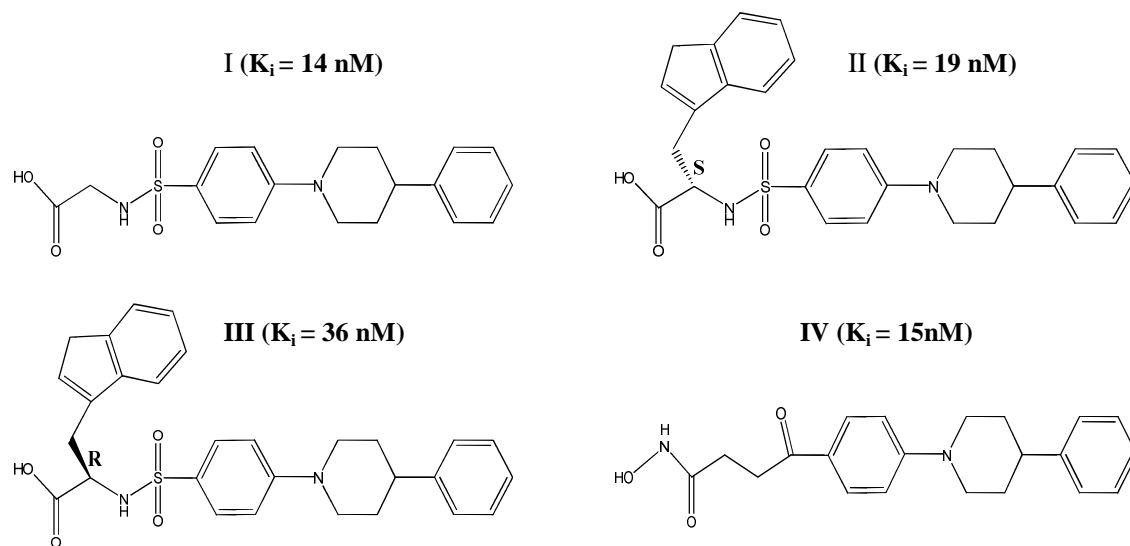


Figure 121. (after Pavlovsky *et al.*, 1999) Nonpeptide inhibitors I-IV co-crystallised with the catalytic domain of MMP-3.

K_i -values for stromelysin-1 are also shown (Pavlovsky *et al.*, 1999).

Compared to the binding mode of peptide based inhibitors those inhibitors do not have any groups interacting with the subsites S2' and S3' (Figure 122) (Pavlovsky *et al.*, 1999).

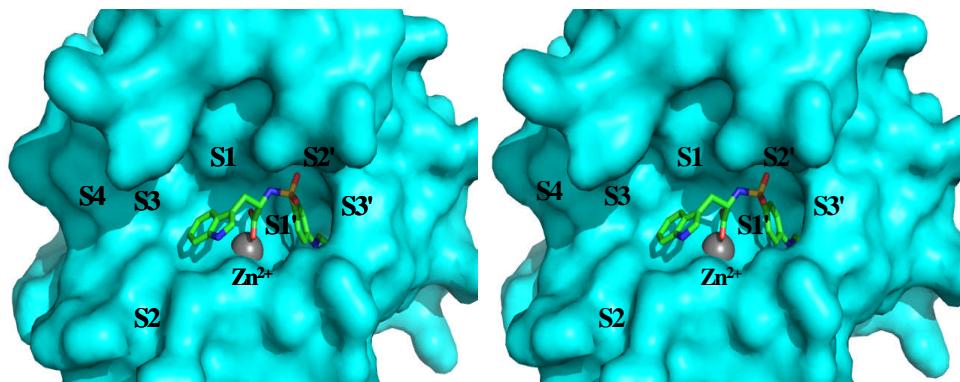


Figure 122. Surface representation of the interactions of inhibitor II with subsites of the catalytic domain of stromelysin-1 (Pdb accession code 1caq). The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

As already discussed in the previous examples of stromelysin-1 inhibitors almost all of these compounds contain hydrophobic P1' moieties (Kohno *et al.*, 2006). Often the group of choice is a methoxyphenyl (Figure 123A) (Pikul *et al.*, 2001; Almstead *et al.*, 1999; Cheng *et al.*, 1999; Cheng *et al.*, 2000;). Other inhibitors have a phenyl (Figure 123B) (Chen *et al.*, 1999) or a 4-phenoxyphenyl group (Figure 123C) (Dunten *et al.*, 2001).

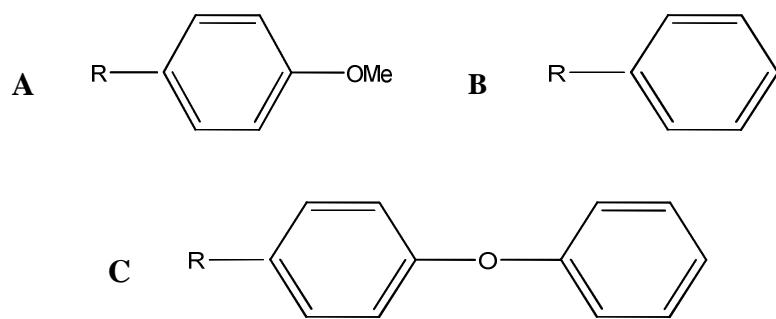


Figure 123. Chemical structures of methoxyphenyl (A), phenyl (B) and 4-phenoxyphenyl (C).

Particularly interesting is the structure of the catalytic domain of stromelysin-1 from *Homo sapiens* complexed with the inhibitor SM-25453 ($K_i = 5.6 \mu\text{M}$ for collagenase-1, $0.02 \mu\text{M}$ for stromelysin-1, $0.007 \mu\text{M}$ for collagenase-3) because the inhibitor's P1' moiety is a large aliphatic chain which has a polar guanidinomethyl group attached. The associations of this hydrophilic group with residues at the bottom of the S1' substrate site are very specific. The fact that there are four H-bonds between the guanidinomethyl group and amino acids in collagenase-3 and only three H-bonds of the same moiety with MMP-3 explains the higher K_i value of SM-25453 for MMP-13. Apart from this moiety SM-25453 also contains a hydroxamic acid as the Zn^{2+} chelating group and a D-lysine (Figure 124) (Kohno *et al.*, 2006).

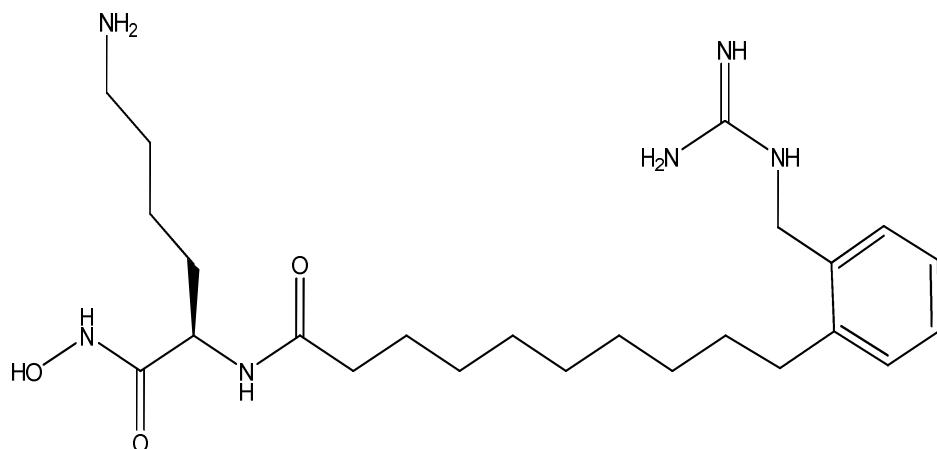


Figure 124. (after Kohno *et al.*, 2006) Chemical structure of SM-25453.

Other research groups focused their design on inhibitors to explore interactions with the S2' subsites of stromelysin-1 and other members of the MMP family (Cheng *et al.*, 1999; Cheng *et al.*, 2000; Pikul *et al.*, 2001). Pikul and coworkers solved the structure of the catalytic domain of stromelysin-1 in complex with an inhibitor which they called **4e**. This compound contains a hydroxamic acid moiety, a 6-oxohexahydropyrimidine ring connected with a sulfonamide group and a methoxyphenyl group.

Additionally a CH_2Ph group is attached to the hexahdropyrimidine ring (Figure 125) which points to the S2' substrate binding site. Compounds having $\text{C}_n\text{H}_{2n+1}$ or like **4e** a benzylic C_6H_5 -group at their hexahdropyrimidine rings proved to be more potent and specific for MMPs (Pikul *et al.*, 2001).

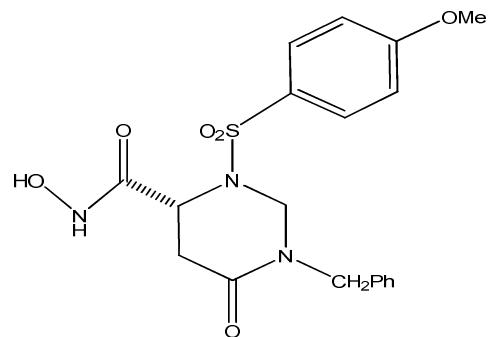


Figure 125. (after Pikul *et al.*, 2001) Chemical structure of **4e**.

Cheng and coworkers synthesized a series of inhibitors containing a hydroxamic acid, a sulfonamide, a methoxyphenyl and a remodelled proline ring moiety. Their studies revealed that an “sp² centre” (Cheng *et al.*, 1999) at carbon number 4 of this ring lead to a dramatic increase in the potency of those compounds. In particular, inhibitors with oxime or exomethylene groups at this position showed this feature. The X-ray structure of the catalytic domain of MMP-3 with one of those inhibitors, called **14** (Figure 126), was solved showing that the oxime group points to the S2' subsite (Cheng *et al.*, 1999).

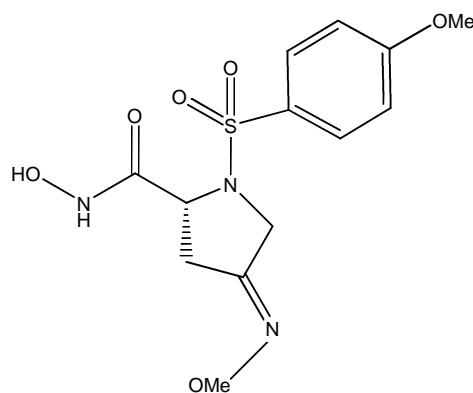


Figure 126. (after Cheng *et al.*, 1999) Chemical structure of compound **14**.

Further nonpeptide inhibitors against stromelysin-1 and other MMPs have been designed such as piperazine derived inhibitors (Figure 127A) (Cheng *et al.*, 2000), 2,4,6-pyrimidine triones (Figure 127B) (Dunten *et al.*, 2001) and thiadiazole inhibitors (Figure 127C and D) (Finzel *et al.*, 1998).

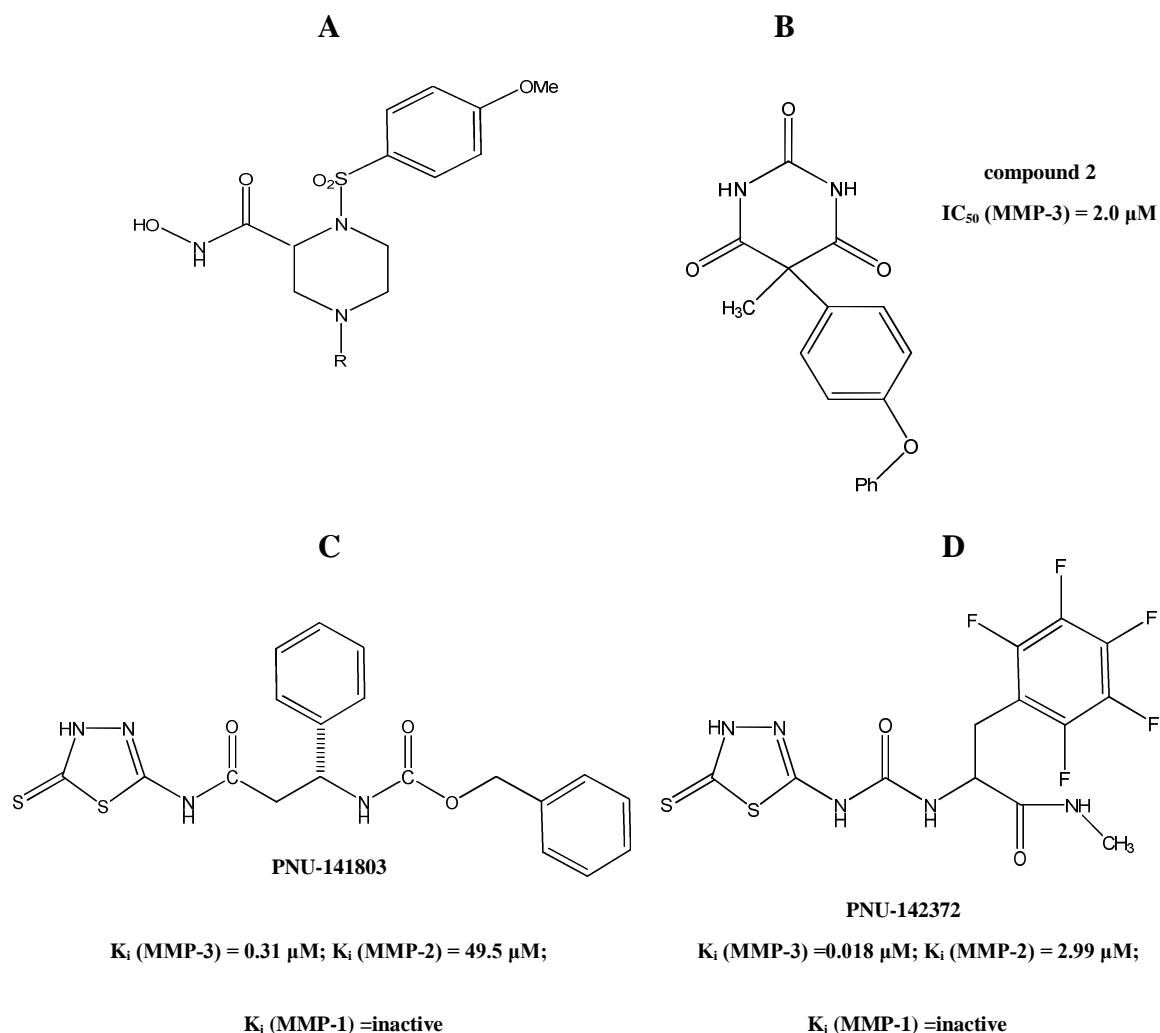


Figure 127. Chemical structures of piperazine based inhibitors (A) (after Cheng *et al.*, 2000), one 2,4,6-pyrimidine trione (B) (after Dunten *et al.*, 2001) and two thiadiazole inhibitors (C and D) (after Finzel *et al.*, 1998). An IC_{50} value for the 2,4,6-pyrimidine trione (Dunten *et al.*, 2001) and K_i values for both thiadiazole inhibitors for three different matrixins (Finzel *et al.*, 1998) are also presented.

In the structure of the catalytic domain of stromelysin-1 complexed to a pyrimidine trione inhibitor named “compound 2” (Dunten *et al.*, 2001), the interactions of this inhibitor with the catalytic Zn^{2+} are quite unique compared to the previously discussed inhibitors. Here, apart from the three His residues, three atoms of this 2,4,6-pyrimidine trione are in close distances to the catalytic Zn^{2+} , N3, O2 and O4 (Figure 128) (Dunten *et al.*, 2001).

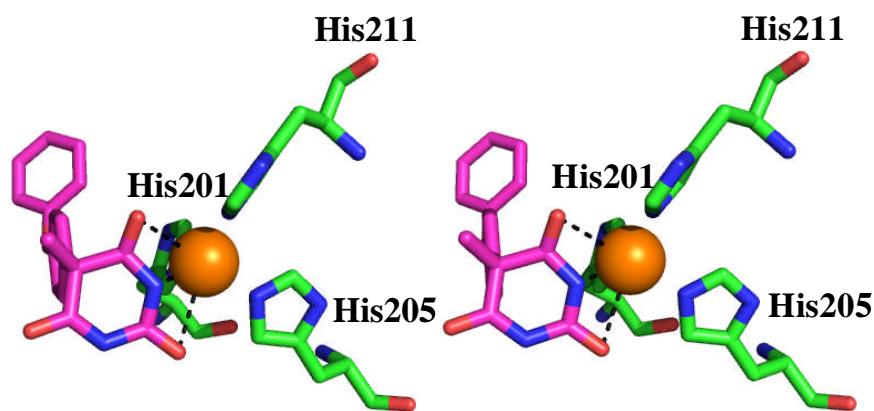


Figure 128. (based on Dunten *et al.*, 2001) Stereo view of the interactions of N3, O2 and O4 of compound 2 with the catalytic Zn^{2+} of the catalytic domain of stromelysin-1 (Dunten *et al.*, 2001) (Pdb accession code 1g4k). The three histidines ligating with Zn^{2+} (carbons in green) and the inhibitor (C-atoms in magenta) are shown as sticks, the catalytic Zn^{2+} as orange sphere and distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

The interesting feature of the structures with two thiadiazole inhibitors bound is that they exploit the unprimed subsites of stromelysin-1 (S1-S4) (Figure 129). Compound PN-142372 (Figure 127D) contains a carbamide and a penta-fluorophenylalanine moiety. The latter group is responsible for the high K_i value for MMP-3 compared to PNU-141803. PNU-141803 has amide, β -phenylalanine and O-benzylcarbamate groups (Figure 127C). The latter moiety can even extend to the S4 substrate binding site (Figure 129). In both inhibitors S2 of the thiadiazole ligates with the catalytic Zn^{2+} (Figure 129) (Finzel *et al.*, 1998).

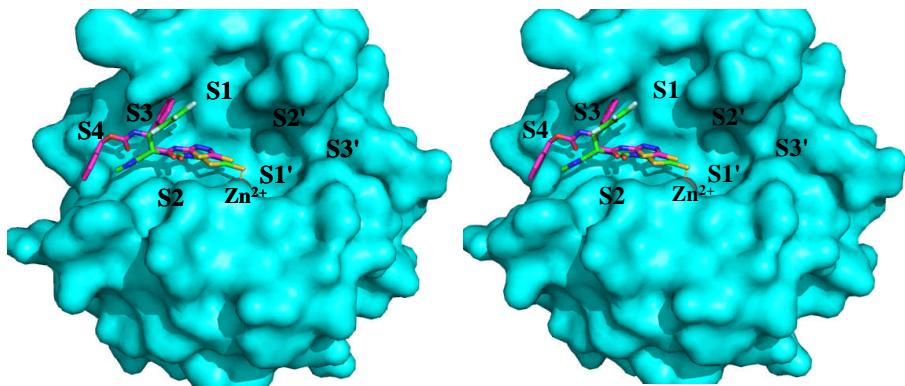


Figure 129. (after Finzel *et al.*, 1998) Surface representation of the interactions of inhibitors PNU-142372 (carbons in green) and PNU-141803 (C-atoms in magenta) with subsites of the catalytic domain of stromelysin-1 (Pdb accession codes 1usn and 2usn). The inhibitors are shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

Another series of nonpeptide inhibitors against MMP-3 characterised by hydroxamic acid and sulfonamide moieties has been synthesized (MacPherson *et al.*, 1997) (Figure 130).

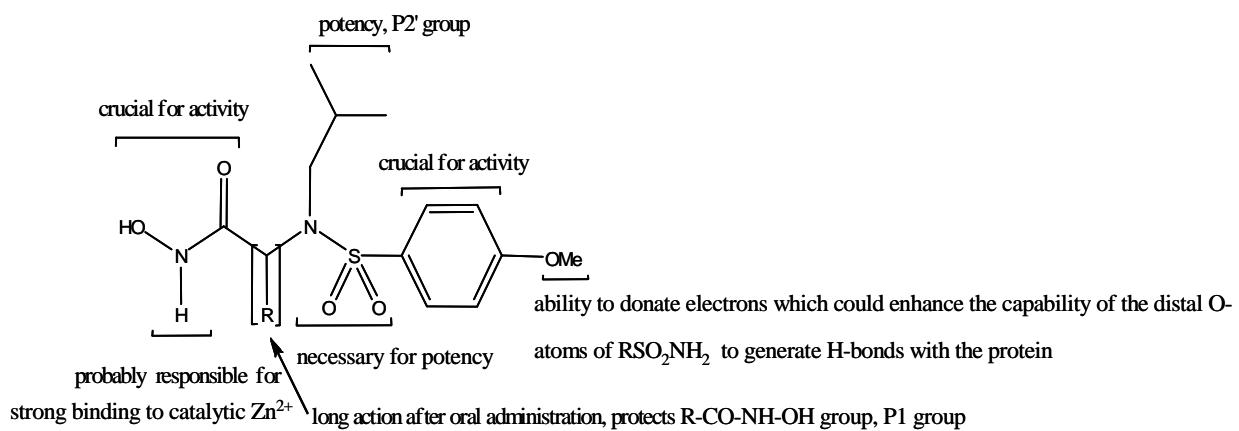


Figure 130. Chemical structure of inhibitor 3 and functions of its diverse moieties (after MacPherson *et al.*, 1997).

It was aimed to design compounds which could efficiently inhibit the decomposition of cartilage in rabbits.

Initial studies in optimising those compounds revealed that the hydroxamic acid moiety is crucial for their activities. Changing the benzol moiety of the para-methoxybenzenesulfonamide group to heterocycles (azabenzene, dimethylthiazole, thifuran) eliminated activity. The RSO_2NH_2 group was shown to be essential because its substitution by an amide connection resulted in an inactive compound. In general, addition of various different groups, especially non-polar ones, to N of RSO_2NH_2 lead to potent compounds. Reduced activity was only detected with $(\text{CH}_2)_5\text{NH}$, branching, or no groups attached. Most groups added to the position next to $\text{R}-\text{CO}-\text{NH}-\text{OH}$ were accepted. Additionally it was also observed that the preferred chirality centre at this position was R. The S-isomer or racemic mixture displayed less potency. Further studies were carried out to optimise “oral activity and duration of action” (MacPherson *et al.*, 1997) in an “*in vivo*” (MacPherson *et al.*, 1997) rabbit model (MacPherson *et al.*, 1997). The compound, named **70** or CGS 27023A was identified (Figure 131). The 1-methylethyl moiety attached at the position next to $\text{R}-\text{CO}-\text{NH}-\text{OH}$ proved to be crucial for long action of this inhibitor after oral administration. Its function is very probably to prevent $\text{R}-\text{CO}-\text{NH}-\text{OH}$ becoming metabolically inactivated (MacPherson *et al.*, 1997).

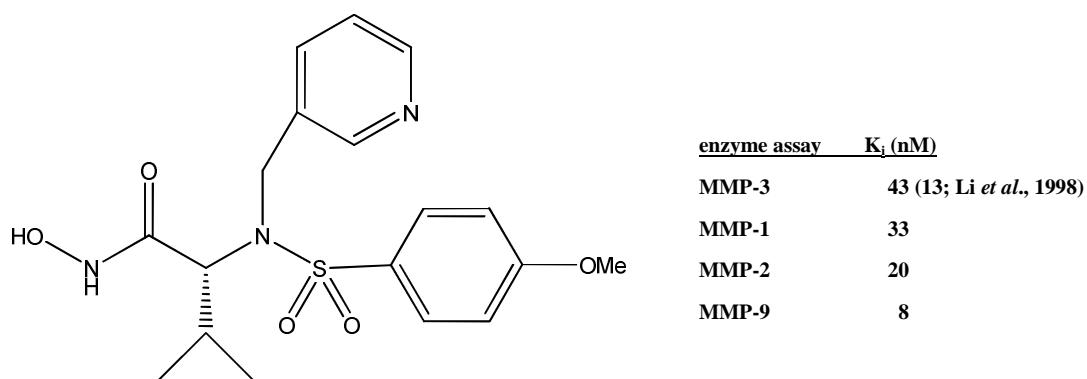


Figure 131. Chemical structure of inhibitor **70** (CGS 27023A) (MacPherson and Parker, 1995; MacPherson *et al.*, 1997).

K_i values for various matrixins are also shown (after MacPherson *et al.*, 1997).

3.3 Project aims

The objective of this project is the determination of X-ray structures of the catalytic domain of MMP-3 in complex with a series of inhibitors based on CGS 27023A (Figure 132), provided by our collaborators Erin Wilfong and Eric Toone from Duke University, Durham, USA.

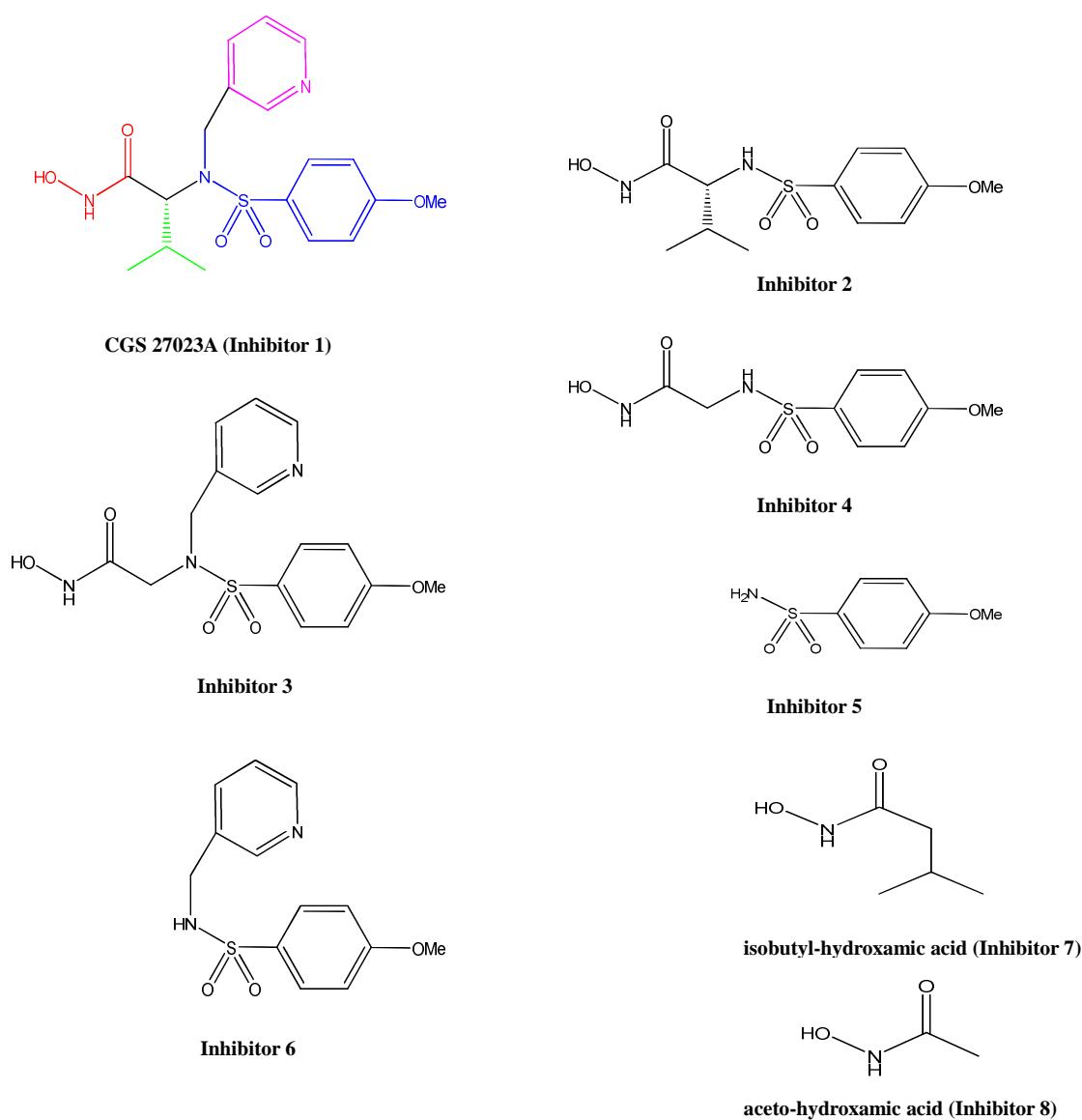


Figure 132. Chemical structures of inhibitors based on CGS 27023A. The various moieties of CGS 27023A are shown in different colours (hydroxamic acid in red, p-methoxybenzenesulfonamide in blue, pyridine in magenta, isopropyl in green).

An NMR structure of the catalytic domain of stromelysin-1 in complex with CGS 27023A has already been published (Gonnella *et al.*, 1997; Li *et al.*, 1998). The interactions of the diverse moieties of inhibitors **1-8** with subsites of stromelysin-1 and close contacts to residues of this enzyme are aimed to be studied with crystals structures of those complexes. Comparisons to the available NMR structure and other X-ray structures of stromelysin-1/inhibitor complexes will also be discussed.

Purification and crystallisation of the catalytic domain of human stromelysin-1 (SCD)

Summary

This section describes the purification and crystallisation of the catalytic domain of MMP-3 from *Homo sapiens*.

3.3 Purification

Purified enzyme of the catalytic domain of human stromelysin-1 (SCD) (Pavlovsky *et al.*, 1999; Li *et al.*, 1998) (amino acids 83-251) was provided by our collaborators Erin Wilfong and Eric Toone from Duke University, Durham, USA. After purification the protein was lyophilised in 2 mM Tris-HCl, pH 7.5, 10 mM CaCl₂ and 1 µM Zn (OAc)₂ and shipped to St. Andrews to perform crystallisation trials.

Initial crystallisation attempts failed, it was therefore decided to check the purity of the enzyme. About 5 mg of the lyophilised powder were dissolved in H₂O and loaded onto an SDS-PAGE gel (Figure 133)

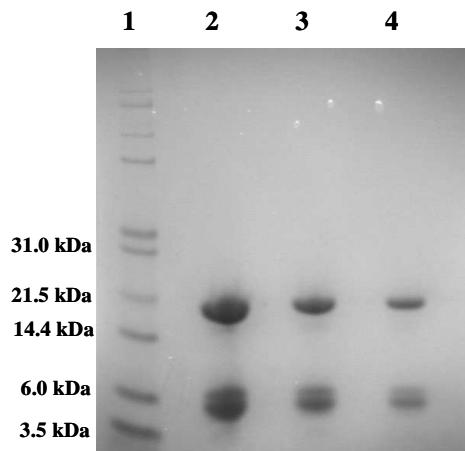


Figure 133. SDS-PAGE showing lyophilised SCD dissolved in H₂O. Lane 1: Mark12™ Protein Standard; lanes 2-4: SCD in H₂O in increasing dilutions.

One band at around 20 kDa corresponding to the catalytic domain of MMP-3 was observed. Additionally two other bands at around 6 kDa were present. Those two bands probably represent autolysis products of the full length enzyme. Consequently, it was decided to purify the protein further.

3.3.1 Size exclusion chromatography

About 14 mg of lyophilised enzyme powder were dissolved in 5 ml buffer containing 2 mM Tris-HCl, pH 7.5, 10 mM CaCl₂ and 1 μM Zn (OAc)₂.

Then the sample was filtered and loaded onto a HiPrepTM 16/60 SephadrylTM S-200 size exclusion column (Amersham Biosciences). This purification step was performed on the ÄKTAXpress (Amersham Biosciences). SCD fractions eluted in 2 mM Tris-HCl, pH 7.5, 10 mM CaCl₂ and 1 μM Zn (OAc)₂ were then checked for their purity by SDS-PAGE (Figure 134).

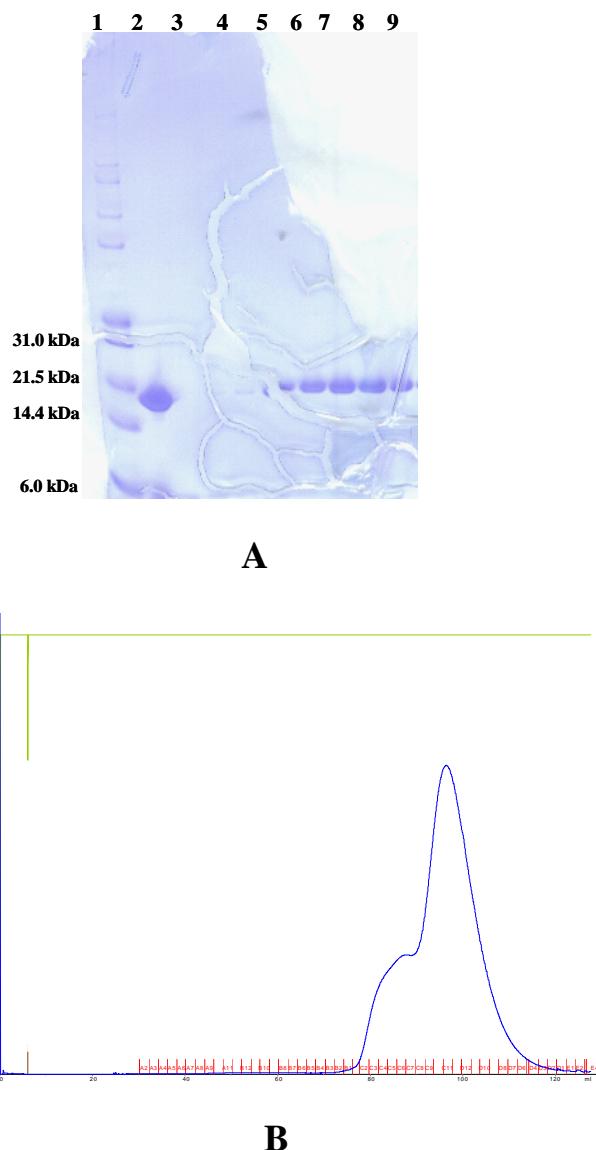


Figure 134. (A) SDS-PAGE showing the fractions of SCD eluted from the HiPrepTM 16/60 SephadrylTM S-200 size exclusion column. Lane 1: Mark12TM Protein Standard; lane 2: SCD before its application to the column: lanes 3-9: eluted SCD fractions; (B) Size exclusion chromatogram of SCD.

Fractions containing pure SCD (about 13 ml) were pooled and concentrated (VIVASPIN 20, 3,000 MWCO PES, Sartorius Stedim Biotech GmbH (SSB)) to 11.5 mg/ml (Figure 135).

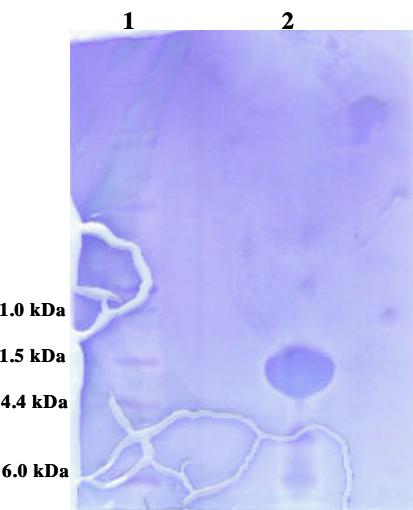


Figure 135. SDS-PAGE showing concentrated SCD. Lane 1: Mark12TM Protein Standard; lane 2: SCD at 11.5 mg/ml

3.4 Crystallisation

Prior to crystallisation SCD at a concentration of 11.5 mg/ml was spun for 5 min at 4°C. Then inhibitors **1-8**, dissolved in 100 % dimethyl sulfoxide (DMSO) (Sigma[®]) were added to final concentration of 4 mM. The eight samples were then incubated for 1-2 hours at room temperature. After that they were centrifuged at room temperature (Heraeus Biofuge Pico) and applied to crystallisation trials. Crystallisation experiments were carried out using the conditions 0.1 M ammonium sulfate, 0.1 M Na-cacodylate, pH 6.5, 20-30 % PEG 8000 in which well diffracting crystals complexed with four different nonpeptide inhibitors had been obtained (Pavlovsky *et al.*, 1999).

The sitting-drop vapour diffusion method was applied using Douglas Instruments 96-well crystallisation plates (Douglas Instruments) at room temperature. Protein and mother liquor were mixed in a ratio of 1:1 (2.5 µl + 2.5 µl, reservoir volume 100 µl). Sealed plates were then incubated at 20°C. Additionally the catalytic domain of stromelysin-1 was screened for other crystallisation conditions against commercially available 96-condition screens, JCSG⁺ and Classics (GE Healthcare), using the nanodrop crystallisation robot (Cartesian Dispensing Robot Honeybee 963, Genomic Solutions) on sitting drop plates (Innovaplate™ SD-2 Crystallography Plate, Innovadyne Technologies, Inc.). The 192 wells on those plates allowed screening of two different compounds (well # 1, compound **4**; well # 2, compound **8**) per crystallisation condition. Crystals large enough for diffraction experiments appeared after one to four weeks under the already published conditions (Pavlovsky *et al.*, 1999), but only when inhibitors **3** or **4** had been incubated with the protein solution (Figure 136). Crystals grew neither in presence of the other compounds, nor under any conditions of the two commercial screens.

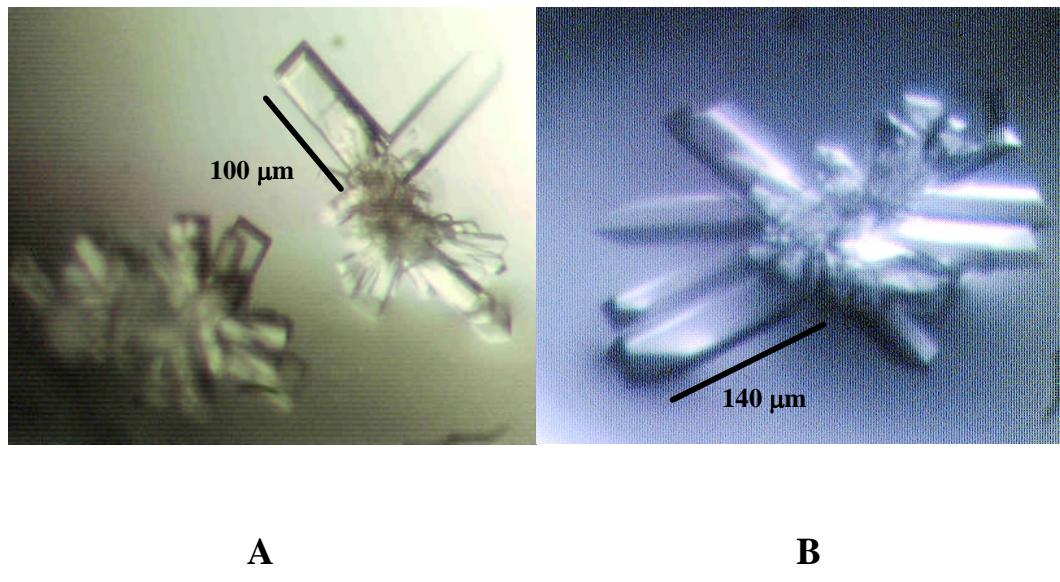


Figure 136. Crystals of SCD grown in presence of 2mM of compound **3** under the condition 0.1 M (NH₄)₂SO₄, 0.1 M Na-cacodylate, pH 6.5, 29% PEG 8000 (A) and in presence of 2mM of compound **4** under the condition 0.1 M (NH₄)₂SO₄, 0.1 M Na-cacodylate, pH 6.5, 26% PEG 8000 (B).

These crystals were cryoprotected in a solution containing 0.15 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M Na-cacodylate, pH 6.5, 30 % PEG 8000, 10 % PEG 400 and 2 or 4 mM of inhibitors **3** or **4** before being flash frozen in liquid nitrogen.

Discussion

The purified catalytic domain of stromelysin-1 was sent by our collaborators in a lyophilised form. Initial crystallisation and co-crystallisation trials with this sample using already published conditions (Chen *et al.*, 1999; Pavlovsky *et al.*, 1999) and also four different commercially available crystallisation screens (JCSG⁺, Classics, PEGs, pH clear) only yielded salt crystals. Consequently, the purity of the sample was checked by SDS-PAGE electrophoresis which revealed that additionally to full length RHTS, two other bands at a molecular weight of about 6 kDa corresponding to autolysis products were also present in the sample. The introduction of an additional purification step (size exclusion chromatography) was carried out to remove those contaminants. It was also very important to include CaCl_2 in the gel filtration buffer. If one withdraws calcium the protein becomes exceedingly unstable and degrades completely within hours. It can tolerate not having zinc for a couple of days but Ca^{2+} seems to be required for any degree of stability (personal communication with Erin Wilfong). Consequently, after the size exclusion step, the incubation time of SCD with the compounds was also kept short (1-2 hours at room temperature) and after that crystallisation trials were set up immediately. The fact that co-crystals could only be obtained in the presence of two of the compounds (**3** and **4**) is hard to explain as all of the inhibitors are very similar (see Figure 132). Perhaps, one could have tried more variations in the crystallisation conditions (also changing the concentration of ammonium sulfate and the pH of Na-cacodylate and/or using a wider concentration range of PEG 8000) and/or in protein and/or inhibitor concentrations).

Data collection, Processing, Molecular Replacement and Refinement of the catalytic domain of human MMP-3 complexed with inhibitors 3 and 4

Summary

Two complete X-ray datasets of SCD co-crystals, one in complex with inhibitor **3** and a second one complexed with compound **4**, were collected on the in-house RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku) to 2.4 and 2.5 Å, respectively. The diffraction data were processed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). An already solved structure of the catalytic domain of MMP-3 (Pdb accession code 1b8y) (Pavlovsky *et al.*, 1999) was decided to be used as a model to calculate a set of initial phase estimates using the technique of Molecular Replacement (MR). After the structures had been solved using PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004), incorrect amino acid rotamers were manually changed in COOT (Emsley and Cowtan, 2004). Isotropic refinement and TLS parameters were carried out in all steps of refinement using the programme REFMAC (Murshudov *et al.*, 1999). The final models were checked and validated using MOLPROBITY (Davis *et al.*, 2004).

3.6 Data collection, processing, structure solution and refinement of SCD in complex with inhibitor 4

3.5.1 Data collection and processing

Before data collection crystals were cryoprotected in a solution containing 0.15 M ammonium sulfate, 0.1 M Na-cacodylate, pH 6.5. 30 % PEG 8K, 4 mM compound **4** and 10 % PEG 400. Diffraction data were collected at -173.15 °C at a wavelength of 1.542 Å in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku). Data from one crystal to 2.5 Å resolution were collected in 0.5° oscillations (Figure 137) with 20s exposure. Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor et al., 2002). Initially it was not sure if the crystal belonged to the space group C222 or C222₁. However, PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004) could only find a correct solution in C222₁. Data quality assessment by phenix.xtriage (Zwart *et al.*, 2005) revealed pseudo translational symmetry (Table 34).

Table 34. Output from phenix.xtriage (Zwart *et al.*, 2005) showing pseudo translational symmetry and reindexing parameters.

The full list of Patterson peaks is:				
x	y	z	height	p-value(height)
(0.000, 0.000, 0.500) :			77.422	(8.535e-07)
(0.500, 0.000, 0.000) :			9.031	(2.276e-01)
If the observed pseudo translational are crystallographic the following spacegroups and unit cells are possible:				
space group	operator	unit cell of reference setting		
C 2 2 2 (a,b,2*c)	x, y, z+1/2	(56.48, 121.04, 46.87, 90.00, 90.00, 90.00)		

As a consequence reindexing of the data using the CCP4 programme Reindex was necessary. The data have an overall completeness of 99 % and 85.4 % in the highest resolution shell from 2.54-2.5 Å resolution. Data collection and processing statistics are shown in Table 35.

Table 35. Data collection statistics for the SCD complex crystal. Data for the highest resolution shell are shown in brackets.

Data collection	SCD
λ (Å)	1.542 Å
Resolution	50 – 2.5
Last shell (Å)	(2.54-2.5)
Spacegroup	C222 ₁
Cell (Å, °)	$a = 56.5, b = 121, c = 47$ $\alpha = \beta = \gamma = 90$
Unique reflections	11540
Average redundancy	3.6 (3.0)
I/σ	26.6 (4.8)
Completeness (%)	99 (85.4)
R_{merge}^*	0.065 (0.225)

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

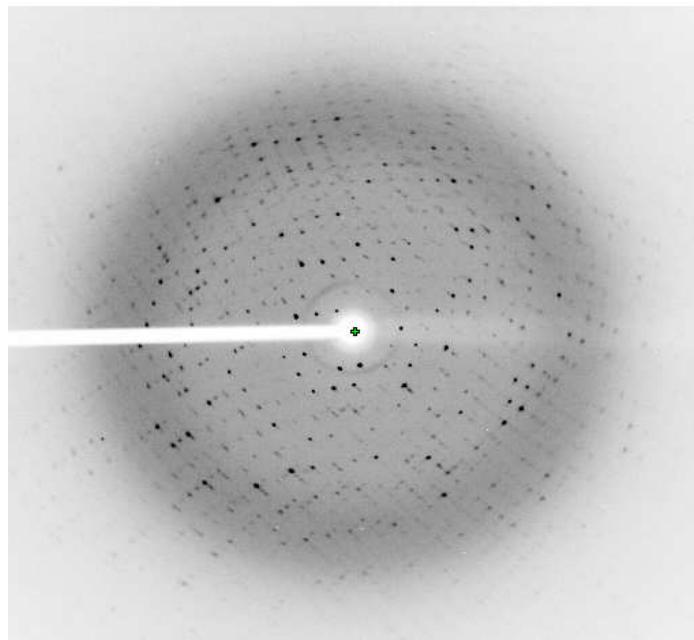


Figure 137. 0.5° oscillation X-ray image of the co-crystal of SCD collected in-house to 2.5 \AA resolution.

3.5.2 Structure solution

The structure of SCD in complex with inhibitor **4** was solved by molecular replacement using PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004). A monomer of the catalytic domain of stromelysin-1 (Pdb accession code 1b8y) was used as a search unit. A solution was found for one monomer in the asymmetric unit consistent with the Matthews coefficient calculation (Matthews, 1968) and a resolution range between 60.5 and 2.5 \AA . A clear rotation and translation solution (Z-scores of 15.1 and 21.4, respectively) for one monomer was found in the space group C222₁. Only one unfavourable molecular contact was observed for the final solution in crystal packing. The Molecular Replacement solution was then applied to the coordinates of the search model to work as the starting model in the refinement described in section 3.5.3

3.5.3 Refinement

REFMAC 5 (Murshudov *et al.*, 1999) was used to refine the structure of the catalytic domain of MMP-3 in complex with inhibitor **4**. Data in the range from 60.5-2.5 Å were used. Incorrect rotamers of the enzyme were modified manually in COOT (Emsley and Cowtan, 2004). TLS parameters (Table 36) and isotropic B-thermal factors were employed in all cycles of refinement.

Table 36. TLS groups employed throughout the refinement of SCD. “Residues” 1-5 are referred to the metal ions Zn²⁺ and Ca²⁺, “residue” 6 to compound **4**.

TLS
RANGE 'A 80.' 'A 260.' ALL
RANGE 'B 1.' 'B 6.' ALL

Waters and ligands (compound **4**, Zn²⁺, Ca²⁺, SO₄²⁻) were also added in COOT (Emsley and Cowtan, 2004). The R_{work} and R_{free} after the last round of refinement were 24 % and 25.5 %, respectively (Table 37). The quality of the structure was judged using MOLPROBITY (Table 38) (Davis *et al.*, 2004).

Table 37. Refinement statistics of SCD complexed with inhibitor 4. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Refinement	SCD
R_{work}^* %	24 (26.5)
R_{free}^* %	25.5 (21)
Number of atoms (average)	
B value (\AA^2)	
protein (mainchain)	636 (27.9)
protein (sidechain)	636 (28.3)
water	15 (44.4)
Zn ²⁺	2 (34.6)
Ca ²⁺	3 (32.1)
Inhibitor 4	17 (50.7)
SO ₄ ²⁻	5 (55.2)
rmsd bonds (\AA) / angles (°)	0.006 / 0.929
Ramachandran favoured (%)	94.8

$$* R_{work} = \sum_{hkl} | | F_{obs} - | F_{calc} | | / \sum_{hkl} | F_{obs} |$$

* For R_{free} the same calculation applies, but for only 5-10 % of reflections which are not included in the refinement.

Table 38. Output from MOLPROBITY (Davis *et al.*, 2007)

All-Atom Contacts	Clashscore, all atoms:	8.01	98th percentile* (N=271, 2.25Å - 2.75Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Rotamer outliers	0.00%	Goal: <1%
	Ramachandran outliers*	0.65%	Goal: <0.2%
	Ramachandran favored	94.77%	Goal: >98%
	Cβ deviations >0.25Å	1	Goal: 0
	MolProbity score	1.80	98th percentile* (N=6960, 2.25Å - 2.75Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	0.63%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

* Ramachandran outliers: Leu245

3.6 Data collection, processing, structure solution and refinement of SCD in complex with inhibitor 3

3.6.1 Data collection and processing

Before data collection crystals were cryoprotected in a solution containing 0.15 M ammonium sulfate, 0.1 M Na-cacodylate, pH 6.5. 30 % PEG 8K, 2 mM compound **3** and 10 % PEG 400. Diffraction data were collected at -173.15 °C at a wavelength of 1.542 Å in-house on the RA-Micro 7 HFM Table Top Rotating Anode X-Ray Generator (Rigaku). Data from one crystal to 2.4 Å resolution were collected in 0.5° oscillations (Figure 138) with 20s exposure. Indexing and merging of the data were performed using Denzo and Scalepack in the integrated package HKL2000 (Minor *et al.*, 2002). The crystal belonged to the space group C222₁. The data have an overall completeness of 99.3 % and 97.9 % in the highest resolution shell from 2.4-2.36 Å resolution. Data collection and processing statistics are shown in Table 39.

Table 39. Data collection statistics for the SCD complex crystal. Data for the highest resolution shell

are shown in brackets.

Data collection	SCD
λ (Å)	1.542 Å
Resolution	50-2.36
Last shell (Å)	(2.4-2.36)
Spacegroup	C222 ₁
Cell (Å, °)	a = 57, b = 120.5, c = 47 $\alpha = \beta = \gamma = 90$
Unique reflections	6909
Average redundancy	2.3 (1.9)
I/σ	21.3 (5.1)
Completeness (%)	99.3 (97.9)
R _{merge} *	0.062 (0.196.)

* $R_{\text{merge}} = \sum |I(k) - \langle I \rangle| / \sum I(k)$, where $I(k)$ is the value of the k_{th} measurement of the intensity of a reflection, $\langle I \rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

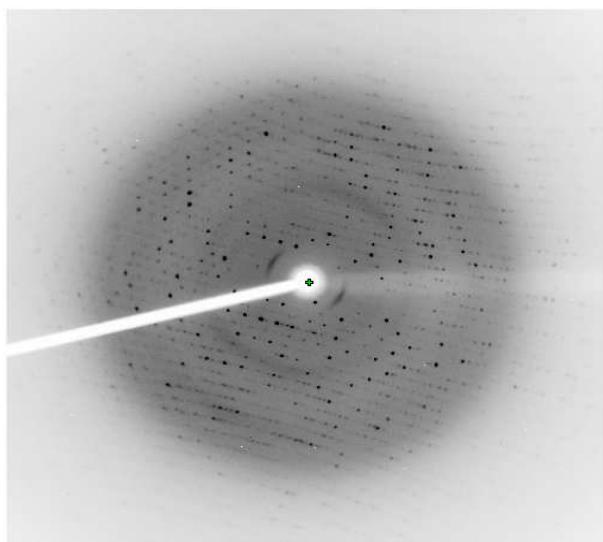


Figure 138. 0.5° oscillation X-ray image of the co-crystal of SCD collected in-house to 2.4 Å resolution.

3.6.2 Structure solution

The structure of SCD in complex with inhibitor **3** was solved by molecular replacement using PHASER (McCoy *et al.*, 2005; Storoni *et al.*, 2004). A monomer of the catalytic domain of stromelysin-1 (Pdb accession code 1b8y) was used as a search unit. A solution was found for one monomer in the asymmetric unit consistent with the Matthews coefficient calculation (Matthews, 1968) and a resolution range between 24.3 and 2.5 Å. A clear rotation and translation solution (Z-scores of 13.8 and 19.5, respectively) was found for one monomer. No unfavourable molecular contacts were observed for the final solution in crystal packing. The Molecular Replacement solution was then applied to the coordinates of the search model to work as the starting model in the refinement described in section 3.6.3

3.6.3 Refinement

REFMAC 5 (Murshudov *et al.*, 1999) was used to refine the structure SCD in complex with inhibitor **3**. Data in the range from 14-2.4 Å were used. Incorrect rotamers of the enzyme were modified manually in COOT (Emsley and Cowtan, 2004). TLS parameters (Table 40) and isotropic B-thermal factors were employed in all cycles of refinement. Waters and ligands (compound **3**, Zn²⁺, Ca²⁺, SO₄²⁻) were also added in COOT (Emsley and Cowtan, 2004). The R_{work} and R_{free} after the last round of refinement were 25.2 % and 28.2 %, respectively (Table 41). The quality of the structure was judged using MOLPROBITY (Table 42) (Davis *et al.*, 2004).

Table 40. TLS groups employed throughout the refinement of SCD. “Residues” 1-5 are referred to the metal ions Zn²⁺ and Ca²⁺, “residue” 6 to compound 3.

TLS GROUP 1
RANGE 'A 80.' 'A 104.' ALL
TLS GROUP 2
RANGE 'A 105.' 'A 212.' ALL
RANGE 'B 1.' 'B 6.' ALL
TLS GROUP 3
RANGE 'A 212.' 'A 237.' ALL

Table 41. Refinement statistics of SCD complexed with inhibitor 3. R_{work} and R_{free} in the highest resolution shell are shown in brackets.

Refinement	SCD
R _{work} * %	25.2 (27.1)
R _{free} * %	28.2 (35.6)
Number of atoms (average)	
B value (Å ²)	
protein (mainchain)	636 (54.9)
protein (sidechain)	636 (55.4)
water	16 (66.2)
Zn ²⁺	2 (55.1)
Ca ²⁺	3 (50.8)
Inhibitor 3	24 (60)
SO ₄ ²⁻	5 (50.6)
rmsd bonds (Å) / angles (°)	0.01 / 1.214
Ramachandran favoured (%)	89.5

$$* R_{\text{work}} = \sum_{\text{hkl}} |F_{\text{obs}} - |F_{\text{calc}}| | / \sum_{\text{hkl}} |F_{\text{obs}}|$$

* For R_{free} the same calculation applies, but for only 5-10 % of reflections which are not included in the refinement.

Table 42. Output from MOLPROBITY (Davis *et al.*, 2007)

All-Atom Contacts	Clashscore, all atoms:	23.15	51st percentile* (N=331, 2.15Å - 2.65Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Protein Geometry	Rotamer outliers	2.26%	Goal: <1%
	Ramachandran outliers*	1.31%	Goal: <0.2%
	Ramachandran favored	89.54%	Goal: >98%
	Cβ deviations >0.25Å	0	Goal: 0
	MolProbity score	2.69	54th percentile* (N=8058, 2.15Å - 2.65Å)
	Residues with bad bonds:	0.00%	Goal: <1%
	Residues with bad angles:	1.25%	Goal: <0.5%

* 100th percentile is the best among structures of comparable resolution; 0th percentile is the worst.

* Ramachandran outliers: Ser244

Discussion

Both complexes crystallised in the same space group and the cell dimensions are also identical (Tables 35 and 39). This is not surprising as the crystallisation condition is the same. Only the concentrations of the precipitant, PEG 8K, vary slightly. Furthermore, the two inhibitors are also very similar, compound 3 only has an additional pyridine moiety (see Figure 132). In one case data had to be reindexed because of pseudotranslational symmetry (see Table 34). Sometimes, problems occurred with other datasets from crystals grown under the same conditions during data processing. It was often quite difficult to mount a single crystal onto the goniometer because they usually grew in clusters (see Figure 136). As a consequence more than one lattice was frequently present in diffraction patterns. Structure solution and refinement were straightforward. A clashscore of all atoms of 8.01 (98th percentile) and a MOLPROBITY score of 1.8 (98th percentile) indicates the good quality of the SCD structure complexed with inhibitor 4 (Table 38). The corresponding scores for the SCD structure in complex with inhibitor 3 are much lower (Table 42) which means that still further refinement will be required before submission to the Protein Data Bank.

Structures of SCD in complex with inhibitors **3** and **4**

Summary

This section describes the two solved structures of SCD complexed with compounds **3** and **4**. First an analysis of both overall structures and bound ligands is performed. Then a more detailed description of the active sites in both structures is carried out, with focus on the interactions of the inhibitors with the enzyme. Differences and common features between both complexes, other solved SCD X-ray structures and the solution structure in complex with CGS 27023A (Inhibitor **1**) (Li *et al.*, 1998) are also discussed.

3.7 Overall structures

Unsurprisingly, because of their essentially identical sequence, both structures of SCD share the same fold as other structures of the catalytic domain of MMP-3 (Alcaraz *et al.*, 2007; Kohno *et al.*, 2006; Pikul *et al.*, 2001; Dunten *et al.*, 2001; Cheng *et al.*, 2000; Steele *et al.*, 2000; Pavlovsky *et al.*, 1999; Chen *et al.*, 1999; Dhanaraj *et al.*, 1996; Becker *et al.*, 1995). If one superimposes the 2.4 Å resolution structure with one of the already published SCD structures (Figure 139A) (Chen *et al.*, 1999), the rmsd is only 0.351 Å for 135 C α positions. Superposition of the 2.4 Å and 2.5 Å resolution structures gives an rmsd of 0.163 Å for 133 C α positions (Figure 139B). The asymmetric unit of both structures is composed of one monomer.

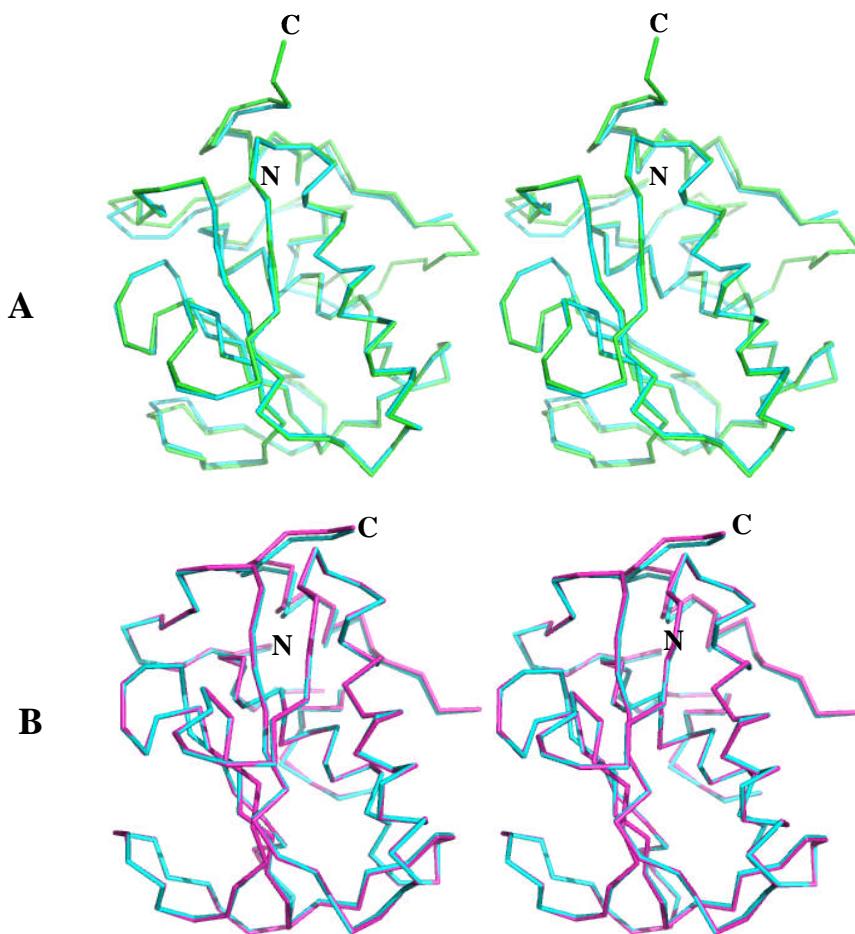


Figure 139. (A) Stereo view of the Cu tracing of SCD complexed with inhibitor 3 (cyan) and chain A of the native catalytic domain of stromelysin-1 (green) (Chen *et al.*, 1999) (Pdb accession code 1cqr). (B) Stereo view of the Cu tracing of SCDs complexed with inhibitors 3 (cyan) and 4 (magenta). This figure has been produced using PyMol (DeLano, 2007).

The monomers of both structures are essentially complete apart from some parts of the “long, flexible loop” (amino acids 210-234) (Chen *et al.*, 1999). In the SCD structure complexed with inhibitor 3, the electron density for amino acids 212-214 and 248 is weak. Amino acids 215, 216 and 225-230 are completely missing. Additionally the last two carboxyterminal residues (250-251) are also disordered.

In the lower resolution complex structure, the electron density for amino acids 213-216, 243 and 247 is not well-defined. As in the 2.4 Å resolution structure the region displaying residues 225-231 is also completely disordered and the two terminal amino acids are missing as well. Disordered amino acids in this flexible loop have also been reported in other X-ray structures of the catalytic domain of MMP-3 (Chen *et al.*, 1999; Dhanaraj *et al.*, 1996; Pavlovsky *et al.*, 1999). In the remainder of both structures electron density is only missing for a few side chains of amino acids such as Arg, Lys and Glu.

3.8 Ion binding sites and coordination

The positions and interactions of the catalytic and structural Zn^{2+} , as well as for the three Ca^{2+} , are essentially identical in both complexes and the same as in previously published SCD structures (Figures 140 and 141) (Kohno *et al.*, 2006; Pikul *et al.*, 2001; Dunten *et al.*, 2001; Cheng *et al.*, 2000; Steele *et al.*, 2000; Pavlovsky *et al.*, 1999; Chen *et al.*, 1999; Becker *et al.*, 1995).

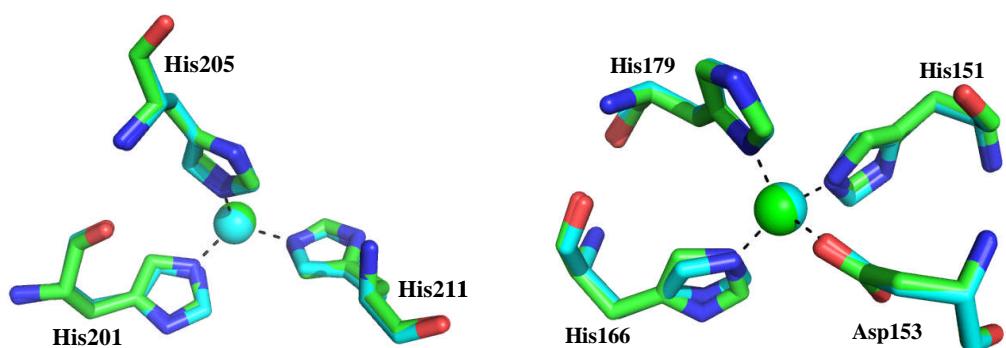


Figure 140. Interactions of both SCDs with the catalytic (left) and structural (right) zinc ions. Carbons of amino acids and Zn^{2+} are shown in green for the SCD structure complexed with inhibitor 3 and in cyan for the SCD structure in complex with inhibitor 4. This figure has been produced using PyMol (DeLano, 2007).

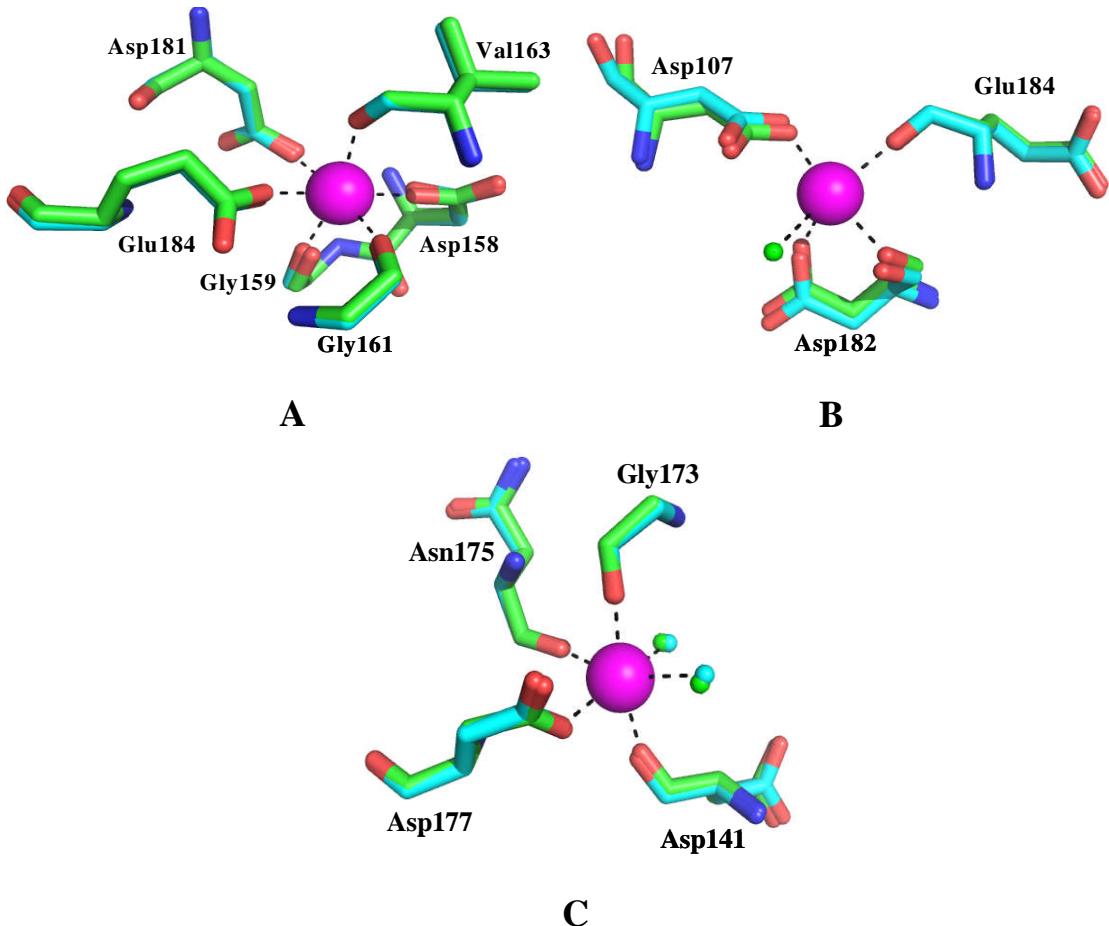


Figure 141. Interactions of both SCDs with the first (A), second (B) and third (C) calcium ion. Carbons of amino acids and waters are shown in green for the SCD structure complexed with inhibitor 3 and in cyan for the SCD structure in complex with inhibitor 4. Calcium ions are shown in magenta. This figure has been produced using PyMol (DeLano, 2007).

It has been reported that in some SCD structures the second Ca^{2+} was not fully occupied. As a consequence the B factor of this calcium ion was much higher than the two other Ca^{2+} (Table 43) (Becker *et al.*, 1995; Pavlovsky *et al.*, 1999). The same seems to apply to the structure of the catalytic domain of stromelysin-1 in complex with inhibitor 4.

There is no electron density for the two H₂O normally ligating with this Ca²⁺ (Chen *et al.*, 1999), and the B factor for this calcium ion is almost twice as high as the temperature factors for Ca1 and Ca3 (Table 43). In the SCD structure complexed with inhibitor **3** one water molecule is bound to Ca2 (Figure 141B). The B factors for each of the three Ca²⁺ do not differ significantly (Table 43).

Table 43. Temperature factors (Å²) of the first, second and third calcium ions (Ca1-Ca3) in both SCD complex structures compared to two other inhibited structures of the catalytic domain of MMP-3.

	Ca1	Ca2	Ca3
SCD + inhibitor 3	50	52	50
SCD + inhibitor 4	26	44	26
1b8y (Pavlovsky <i>et al.</i>, 1999)	14	40	14
1sln (Becker <i>et al.</i>, 1995)	8	30	13

One sulfate ion is also present in the SCD structures complexed with compounds **3** and **4** at exactly the same position as in the structures of the catalytic domain of MMP-3 in complex with other nonpeptide inhibitors (Pavlovsky *et al.*, 1999) (Figure 142). It is in close contact with Lys110, the first residue of α1 (amino acids 110-125) (Figure 142) (Pavlovsky *et al.*, 1999). Interestingly this SO₄²⁻ only occurs in the crystals with the space group P4₁2₁2 even though all complexes crystallised under the same conditions. There is no electron density for this anion in the other crystal forms (P6₅22, hexagonal and P2₁, monoclinic) (Pavlovsky *et al.*, 1999).

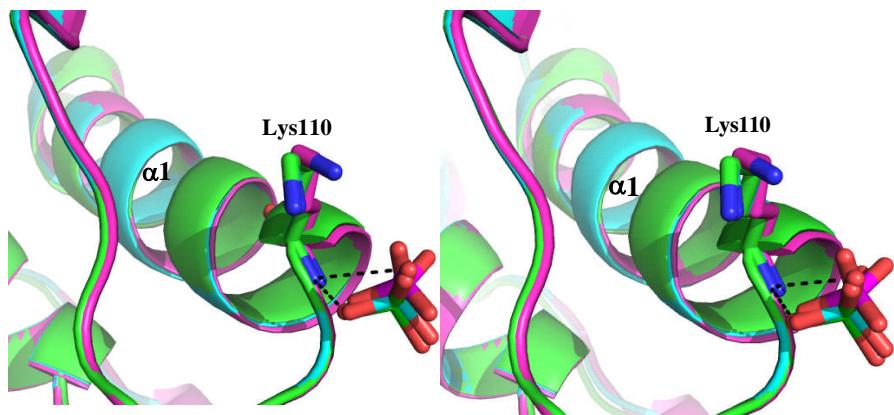


Figure 142. Stereo view of superimposed sulfate ion and α -helix 1 of SCD structures in complex with inhibitors 3 (green) 4 (cyan) and inhibitor I (magenta) (Pdb accession code 1b8y) (Pavlovsky et al., 1999). Lys110 is shown as sticks and close interactions ($< 3.5 \text{ \AA}$) between sulfate O-atoms of SCD complexed with inhibitor 3 and N of Lys110 as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

3.9 Interactions with inhibitors 3 and 4

There is clear electron density for one molecule of inhibitor **3** in the 2.4 \AA resolution (Figure 143A) and for one molecule of inhibitor **4** in the 2.5 \AA resolution structure (Figure 143B)

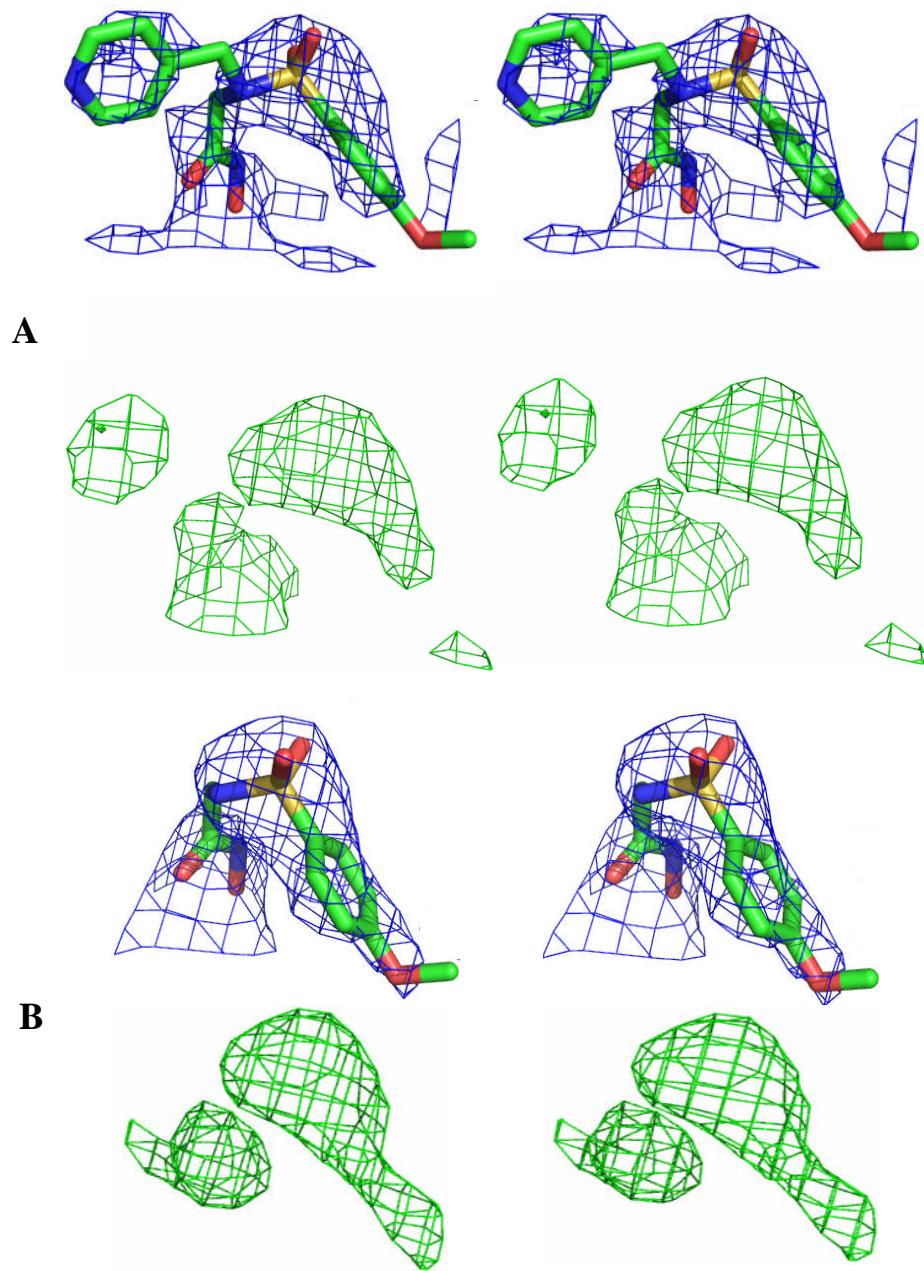


Figure 143. Unbiased $F_o - F_c$ at 3σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around compound 3 of the 2.4 Å structure of SCD (A) and unbiased $F_o - F_c$ at 3σ (green) and final $2F_o - F_c$ at 1σ (blue) maps around compound 4 of the 2.5 Å structure of SCD (B). Both inhibitor molecules are modelled in the $2F_o - F_c$ density. Both maps were calculated at final stages of refinement before addition of the compounds. This figure has been produced using PyMol (DeLano, 2007).

The methoxy group attached to the phenyl ring is flexible in both inhibitors as indicated by the missing electron density in both $2F_o - F_c$ and $F_o - F_c$ maps (Figure 143A and B). The average B factor of the methoxy group of inhibitor **4** in the 2.5 Å resolution structure is 53.6 \AA^2 , whereas the tightly bound hydroxamic acid moiety has an average temperature factor which is unsurprisingly lower (46.6 \AA^2). The average B factors of the corresponding groups of inhibitor **3** in the 2.4 Å resolution structure are 61.7 \AA^2 and 57.9 \AA^2 , respectively.

Because of the essentially same locations of the inhibitors, interacting residues and the catalytic zinc ions in both SCD structures (Figure 144), only the binding mode of inhibitor **3** will be discussed in detail due to its additional pyridine group.

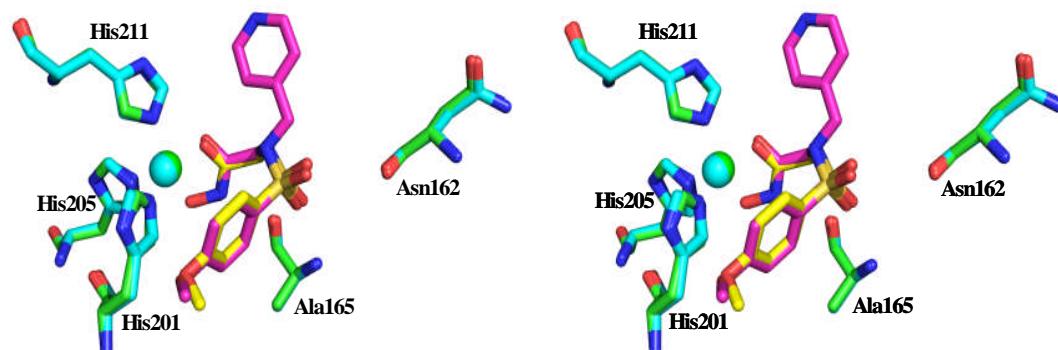


Figure 144. Stereo view of superimposed inhibitors **3** (carbons in magenta) and **4** (C-atoms in yellow) and adjacent residues (shown as green sticks in the 2.4 Å SCD structure and as cyan sticks in the 2.5 Å SCD structure). The zinc ions located at the exactly same position in both structures are shown as sphere(s) in green (higher resolution structure) and cyan (lower resolution structure). This figure has been produced using PyMol (DeLano, 2007).

The two O-atoms of the hydroxamic acid moiety of inhibitor **3**, O and OAE interact with the catalytic Zn^{2+} additionally to histidines 201, 205 and 211 (Figure 145) (Chen *et al.*, 1999; Pavlovsky *et al.*, 1999; Li *et al.*, 1998).

The C=O oxygen (O) is also in close distance to NE2 of His211 (2.6 Å) (Figure 145). Oxygen OAE shows interactions with more residues of SCD. It is in H-bonding distance to OE2 (Li *et al.*, 1998; Chen *et al.*, 1999) (2.5 Å) and OE1 of Glu202 (3.2 Å) (Pavlovsky *et al.*, 1999) and the NE2 nitrogen of both His 201 and 205 (3.1 Å and 2.9 Å, respectively) (Figure 145). The nitrogen NAQ which establishes a covalent connection with OAE (Chen *et al.*, 1999) could also generate H-bonds with O of Ala 165 (3 Å) (Li *et al.*, 1998; Chen *et al.*, 1999) and OE2 of Glu202 (2.9 Å) (Li *et al.*, 1998) (Figure 145).

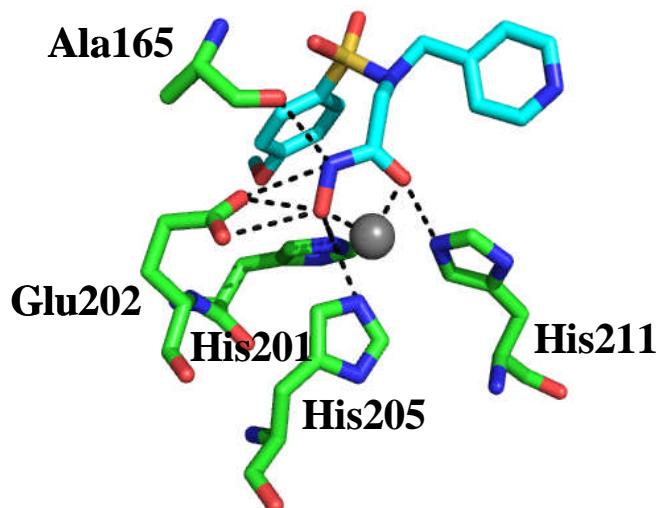


Figure 145. Close up view of the interactions of the hydroxamic acid moiety of inhibitor 3 (carbons in cyan) with the catalytic zinc (grey sphere) and residues of SCD (shown as sticks with C-atoms in green). Distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Oxygen OAC of the sulfonamide moiety of the inhibitor generates three H-bonds with the enzyme (Figure 146A), one with N of Leu164 (2.8 Å) (Li *et al.*, 1998; Pavlovsky *et al.*, 1999), one with N of Ala165 and one with O of Ala165 (3.5 Å). The second O-atom of the sulfonamide (OAD) is only in close contact with one amino acid of SCD, Asn162. The distance between this oxygen and O of Asn162 is 3.6 Å.

The methoxy phenyl moiety of inhibitor **3** (Figure 146) represents the P1' group which inserts into the S1' subsite of SCD (Li *et al.*, 1998) (Figure 147). There are only two residues of the enzyme adjacent to this moiety. The oxygen (OAR) of the methoxy group is located closely to ND1 of His201 (3.4 Å) and O of Leu197 (3.6 Å) (Figure 146B). The azine ring which fills the S2' subsite of the catalytic domain of MMP-3 (Figure 147) (Li *et al.*, 1998) does not associate with any amino acids at all and points to the solvent (Li *et al.*, 1998).

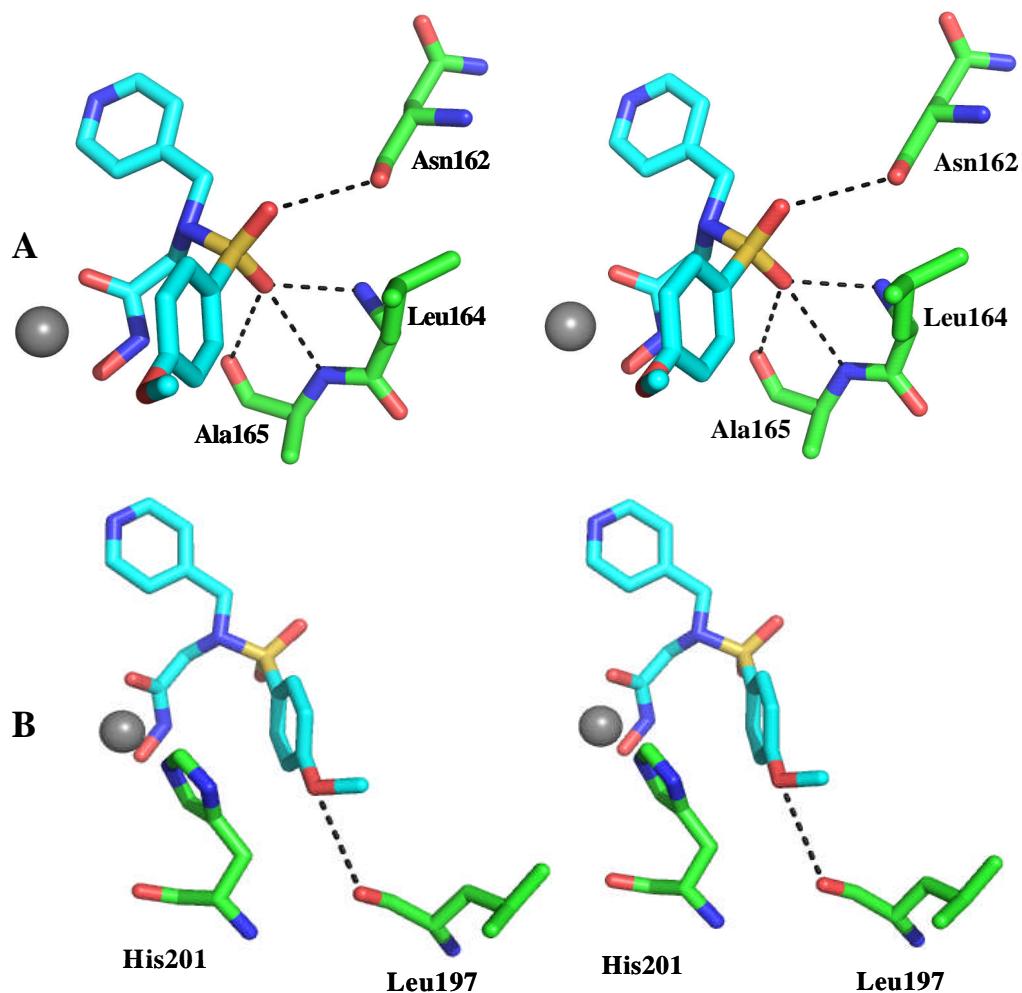


Figure 146. Stereo views of the interactions of the sulfonamide moiety (A) and the methoxy group (B) of inhibitor 3 with residues of SCD. Carbons of the inhibitor and residues of SCD are shown in cyan and green, respectively. The zinc is shown as grey sphere, distances as dashed lines. This figure has been produced using PyMol (DeLano, 2007).

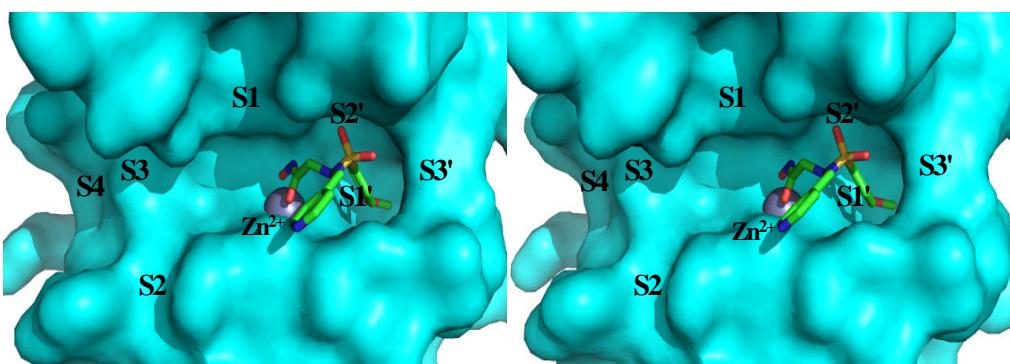
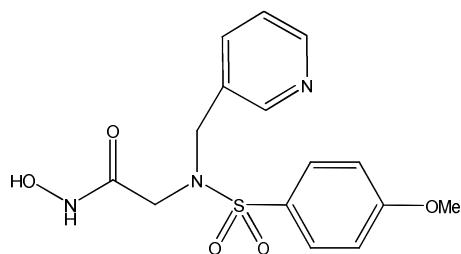


Figure 147. Surface representation of the interactions of inhibitor 3 with subsites of the catalytic domain of the 2.4 Å SCD structure. The inhibitor is shown as sticks and the catalytic zinc ion as grey sphere. This figure has been produced using PyMol (DeLano, 2007).

However, isothermal titration calorimetry (ITC) studies, performed by our collaborators (Wilfong, 2008), show that compound 3 binds thirteen times more tightly to stromelysin-1 than compound 4 (Figure 148). Also, the K_d value for CGS 27023A is more than six times lower than for compound 2 (Figure 148) which contains all the moieties of CGS 27023A except the pyridine ring. Consequently, it can be concluded that the azine ring contributes significantly to the potency of these inhibitors (MacPherson *et al.*, 1997). It is conjectured that the azine ring displaces a water molecule previously bound to the protein thereby increasing the entropy of the system compared to the non-azine molecule and given the similar enthalpies the binding stability increases.

Inhibitor 3

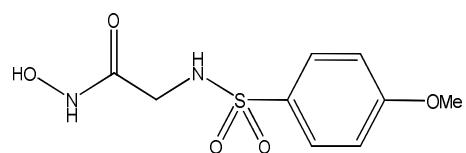


K_d = 22 nM

ΔG = -10.6 kcal/mol

ΔH = -10.7 kcal/mol

Inhibitor 4

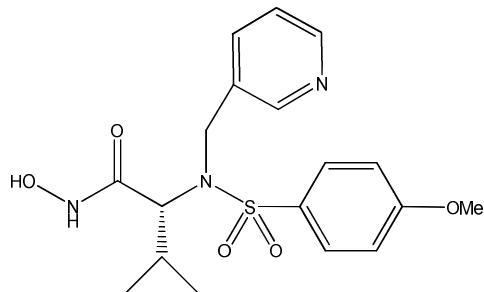


K_d = 296 nM

ΔG = -8.9 kcal/mol

ΔH = -13.6 kcal/mol

CGS 27023A

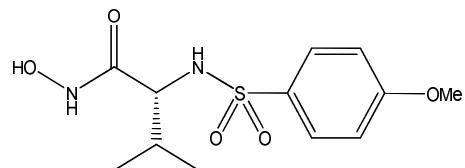


K_d = 7 nM

ΔG = -11.1 kcal/mol

ΔH = -9.0 kcal/mol

Inhibitor 2



K_d = 44 nM

ΔG = -10.2 kcal/mol

ΔH = -7.5 kcal/mol

Figure 148. (after Wilfong, 2008) Isothermal titration calorimetry (ITC) data for CGS 27023A and inhibitors 2-4.

3.10 Common features and differences in inhibitor binding between the crystal structure of SCD in complex with inhibitor 3 and the NMR structure of SCD complexed with CGS 27023A (Inhibitor 1)

The binding mode of both inhibitors is essentially the same in both structures. The two oxygens O and O1 of the hydroxamic acid moiety of CGS 27023A also function as ligands of the catalytic zinc apart from the three conserved histidines (Figure 149) (Chen *et al.*, 1999; Pavlovsky *et al.*, 1999; Li *et al.*, 1998). Like in the SCD structure complexed to inhibitor **3**, the nitrogen is also in H-bonding distance to O of Ala165 (Li *et al.*, 1998) (2.9 Å). Additionally the close distances of the hydroxylamine oxygen to NE2 of His201 and His205 (2.5 Å in both cases) are also common in both structures. However, there are also differences. Whereas only one atom of the hydroxamate of inhibitor **3** is in hydrogen bonding distance to NE2 of His201 (OAE) (Figure 145), both O-atoms and also the nitrogen of the same moiety are in very close contact with NE2 of this histidine (2.5, 2.4 and 3.1 Å, respectively) (Figure 149). His211 is positioned further away from the C=O oxygen of the hydroxamic acid in the CGS 27023A complex structure (Figure 149) which leads to an increase of the distance between this O-atom and NE of His211 (4.2 Å; 2.6 Å between the same atoms in the inhibitor **3** complex). The most striking differences between both structures lie in the interactions of the hydroxamate with Glu202. Li and coworkers reported that there could be an H-bond (about 2.3 Å) between the O-atom of the OH-group of the hydroxamate, which could carry H⁺, and OE2 of Glu202. Additionally this H-bond could be underpinned by the presence of the protective isopropyl moiety of this inhibitor which would obviate the fast replacement of H⁺ of the OH with H₂O (Li *et al.*, 1998). However, the side chain of Glu202 is positioned much too far away to make any interactions with the inhibitor at all (Figure 149).

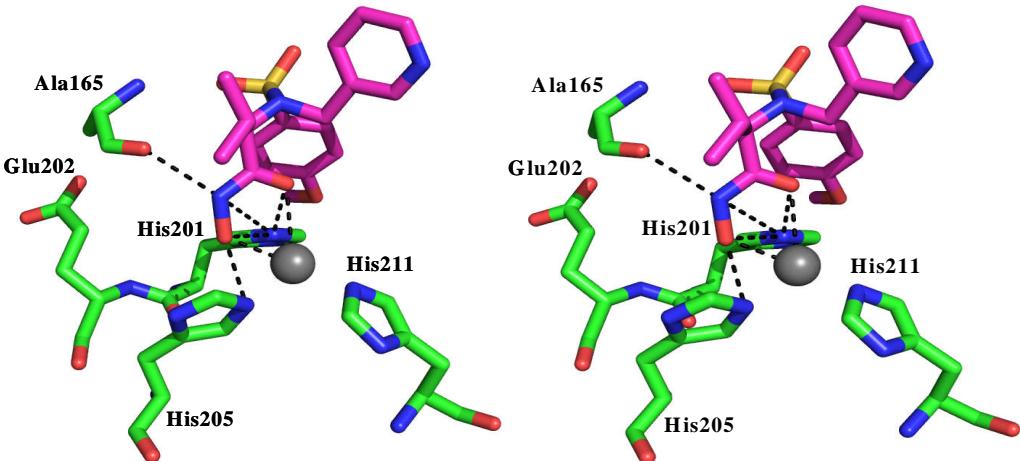


Figure 149. Stereo view of the interactions of the hydroxamic acid moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

Only one sulfonamide oxygen of inhibitor **3** (OAC) is in H-bonding distance with three atoms, provided by Leu164 and Ala165, (see Figure 146A) whereas both O-atoms of the sulfonamide can generate hydrogen bonds with residues of SCD (Li *et al.*, 1998). This is mainly possible due to the different positions of Asn 162 and Leu164 in the NMR structure complexed with CGS 27023A (Figure 150) compared to the 2.4 Å resolution crystal structure. OB1 (OAD in inhibitor **3**) is now located only 2.3 Å from O of Asn162 (3.6 Å in the SCD structure in complex with inhibitor **3**). Additionally it is in H-bonding distance to N of Leu164 (Li *et al.*, 1998). OB2 (OAC in inhibitor **3**) could also form an H-bond with O of Asn162 (3.4 Å). As in the 2.4 Å SCD complex structure, this O-atom of the sulfonamide is also in H-bonding distance with N of Leu164 (Li *et al.*, 1998) and Ala165 (2.4 and 2.7 Å, respectively). O of Ala165, however, is positioned slightly further away (3.7 Å).

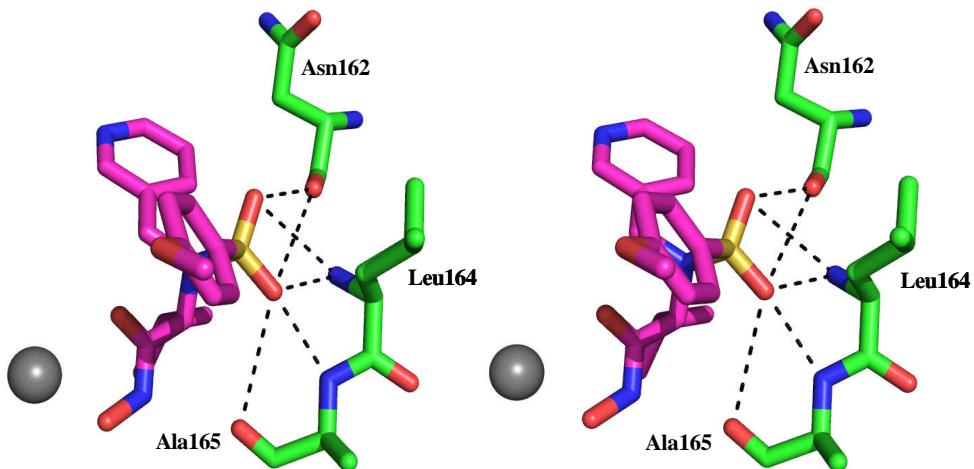


Figure 150. Stereo view of the interactions of the sulfonamide moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

There are also more interactions of the methoxy phenyl group of inhibitor **1** with residues of the enzyme than the same moiety of inhibitor **3** with SCD. Whereas the O-atom of the methoxy group of inhibitor **3** is only in H-bonding distance with ND1 of His201 (see Figure 146B), the corresponding O-atom of CGS 27023A could generate hydrogen bonds with three atoms, provided by His 201 (ND1, 2.6 Å), Leu218 (O, 2.6 Å) and Met219 (O, 3.2 Å) (Figure 151). The two latter amino acids are located more closely to the methoxy oxygen in the NMR structure (Figure 151) than the corresponding residues in the 2.4 Å X-ray structure. O of Leu197, however is positioned further away in Li and coworkers' structure due to the slightly different orientation of the flexible methoxy group (Figure 151).

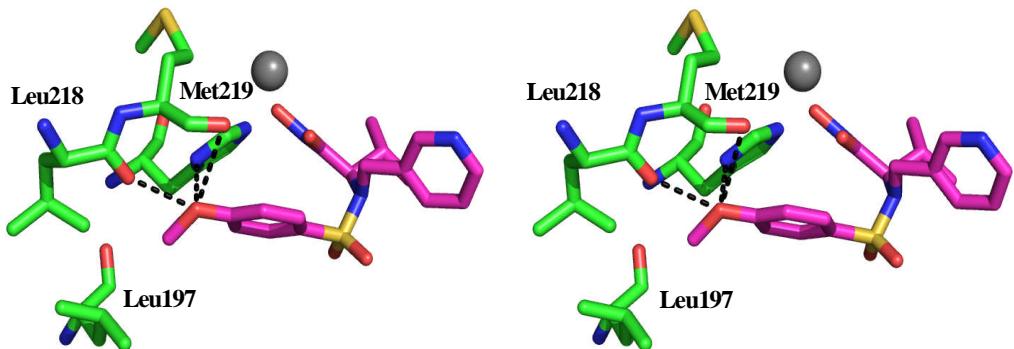


Figure 151. Stereo view of the interactions of the methoxy phenyl moiety of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in green) (Pdb accession code 1bm6). The catalytic zinc ion is shown as grey sphere and distances are shown as dotted lines. This figure has been produced using PyMol (DeLano, 2007).

The pyridine ring of CGS 27023A is rotated differently compared to the pyridine ring of inhibitor **3** (Figure 152). Whereas in the 2.4 Å crystal structure no contacts between this moiety of the latter compound and amino acids of SCD can be observed, two carbons of this ring of CGS 27023A are located closely to atoms of Asn162. CD1 is positioned 3.6 Å from O of Asn162 and CE1 is located 3.1 Å from CB of Asn162 (Figure 152).

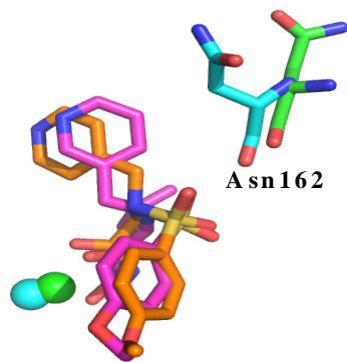


Figure 152. Close up view of the interactions of the pyridine ring of CGS 27023A (carbons in magenta) with residues of SCD (shown as sticks with C-atoms in cyan) (Pdb accession code 1bm6). The superimposed residues and inhibitor 3 of the 2.4 Å resolution SCD structure are also shown as sticks (C-atoms in green and orange, respectively). Zinc ions in the X-ray and NMR structure are shown as green and cyan spheres, respectively. This figure has been produced using PyMol (DeLano, 2007).

CGS 27023A also contains, in contrast to inhibitor **3**, an isopropyl moiety as P1 group which mainly points to the solvent (Figure 153) (Li *et al.*, 1998). This hydrophobic group is only in contact with one nonpolar residue, Val163.

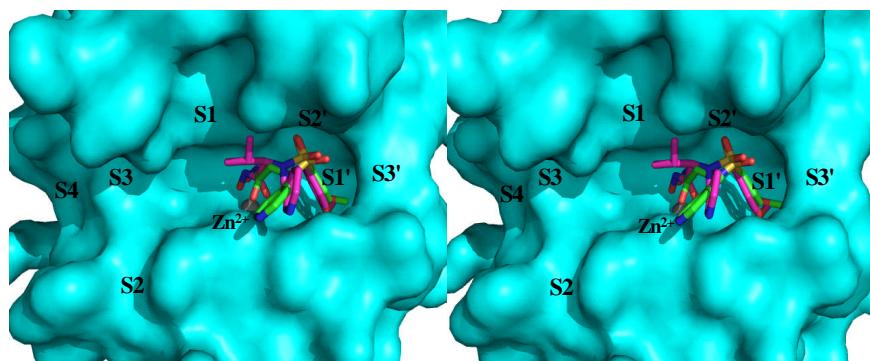


Figure 153. Surface representation of the interactions of inhibitors CGS 27023A (carbons in magenta) and **3** with subsites of the catalytic domain of the SCD structure. The surface of the 2.4 Å SCD crystal structure is shown. This figure has been produced using PyMol (DeLano, 2007).

It was reported that in the NMR structure (Li *et al.*, 1998) the part of the flexible S1' loop (Gooley *et al.*, 1996; Chen *et al.*, 1999; Li *et al.*, 1998) comprising amino acids 221-231 undergoes a remarkable modification in its conformation compared to other reported SCD structures (Li *et al.*, 1998). This distinct conformation (Figure 154) leading to a “compact enzyme-inhibitor complex” (Li *et al.*, 1998) could be responsible for the higher potency of this nonpeptide inhibitor compared to peptidic blockers which actually display more associations with the enzyme. Furthermore, this change in conformation limits access to the S1' subsite (Li *et al.*, 1998). In the SCD crystal structure complexed with inhibitor **3**, however, there is absolutely no sign of any similar conformation of the S1' loop (Figure 154). In addition, most of the residues of this loop region (amino acids 221-231) are disordered. (Figure 154).

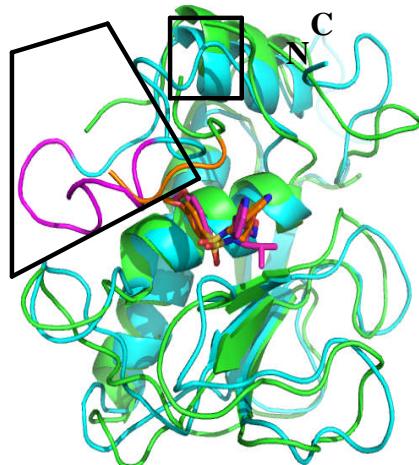


Figure 154. Ribbon diagram showing the 2.4 Å SCD structure complexed with inhibitor **3** (secondary structure elements in green, C-atoms of the inhibitor in orange) superimposed with the NMR structure of SCD in complex with inhibitor **1** (NMR structure with the lowest energy of the ensemble of the twenty structures deposited is used) (secondary structure elements in cyan, carbons of the inhibitor in magenta) (Pdb accession code 1bm6) (Li *et al.*, 1998) (Rmsd = 1.55 Å for 139 Ca positions). Both loop regions, amino acids 221-231 in the NMR structure, and residues 221-224 in the crystal structure are shown in magenta and orange, respectively. Additionally disordered parts of the entire flexible loop (amino acids 210-234) (Chen *et al.*, 1999) in the X-ray structure are boxed. This figure has been produced using PyMol (DeLano, 2007).

It has been shown that in the crystal structure of the catalytic domain of MMP-3 complexed with the inhibitor PGV-25727 the conformation of the S1' loop between amino acids 222 and 231 has been modified due to binding of this compound (Chen *et al.*, 1999). PGV-25727 (Figure 155) is similar to inhibitor **3** and CGS 27023A (Figure 132). In the uninhibited structure access to the S1' subsite is prevented by Tyr223. In the inhibited enzyme, however, Tyr223 was moved by Ph of PGV-25727 (Chen *et al.*, 1999). It has been reported that this highly conserved tyrosine residue might function as a “selective gatekeeper” (Dhanaraj *et al.*, 1996; Chen *et al.*, 1999) to the S1' pocket (Dhanaraj *et al.*, 1996; Chen *et al.*, 1999).

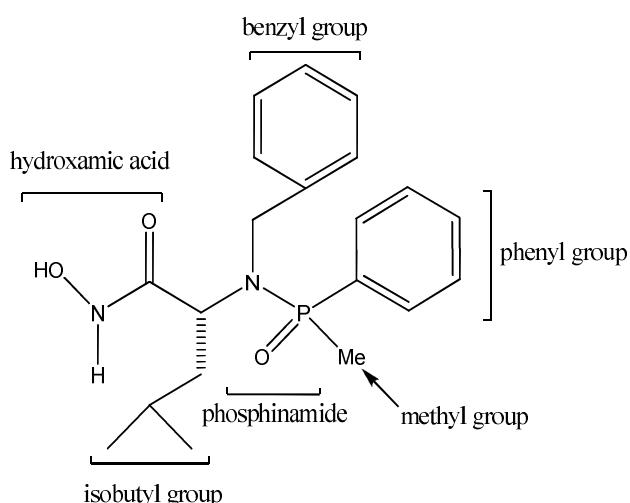


Figure 155. (after Chen *et al.*, 1999) Chemical structure of PGV-25727.

Comparison of the location of Tyr223 and the conformation of the flexible S1' loop (Gooley *et al.*, 1996; Chen *et al.*, 1999; Li *et al.*, 1998) between Chen and coworkers' complex structure and the SCD structure with inhibitor **3** bound reveals that the tyrosine is located at exactly the same position in both structures (Figure 156).

Additionally, parts of the S1' loop which are not disordered in the SCD structure complexed with inhibitor **3** are essentially at almost the same positions in both complexes (Figure 157B). Consequently, it can be proposed that the binding mode of inhibitors **3** and PGV-25727 and the resulting modification of the conformation of the flexible loop which leads to an increase in size of the S1' subsite (Chen *et al.*, 1999) are identical in both SCD structures.

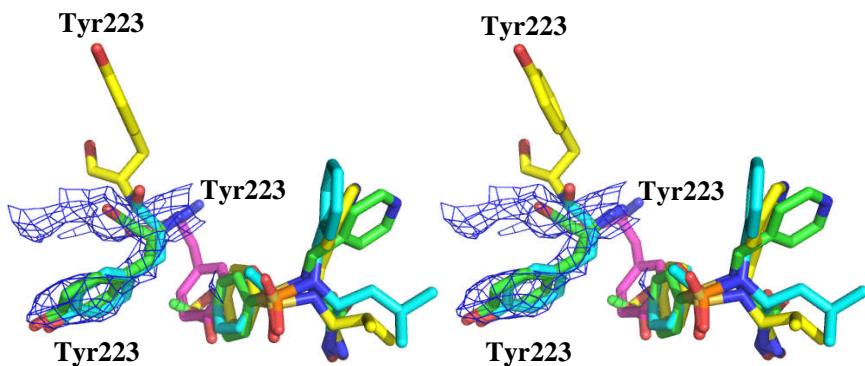


Figure 156. Stereo view of superimposed Tyr223 of SCD complexed with inhibitor 3 (carbons in green $2F_o - F_c$ density at 1 σ), SCD in complex with inhibitor PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999), native SCD (carbons in magenta) (Pdb accession code 1cqr) (Chen *et al.*, 1999) and the NMR structure of SCD complexed with CGS 27023A (C-atoms in yellow) (Pdb accession code 1bm6) (Li *et al.*, 1998). This figure has been produced using PyMol (DeLano, 2007).

Further comparisons between the 2.4 Å resolution SCD complex structure and other structures of the catalytic domain of stromelysin-1 complexed with other nonpeptide (Pavlovsky *et al.*, 1999) (see Section 3.1.3.3 and peptidic inhibitors containing P1' groups (Becker *et al.*, 1995) (see Section 3.1.3.3) reveals that in all those structures the location of Tyr223 is essentially identical and the S1' loop displays the same “open conformation” (Li *et al.*, 1998; Finzel *et al.*, 1998) (Figure 157A and B).

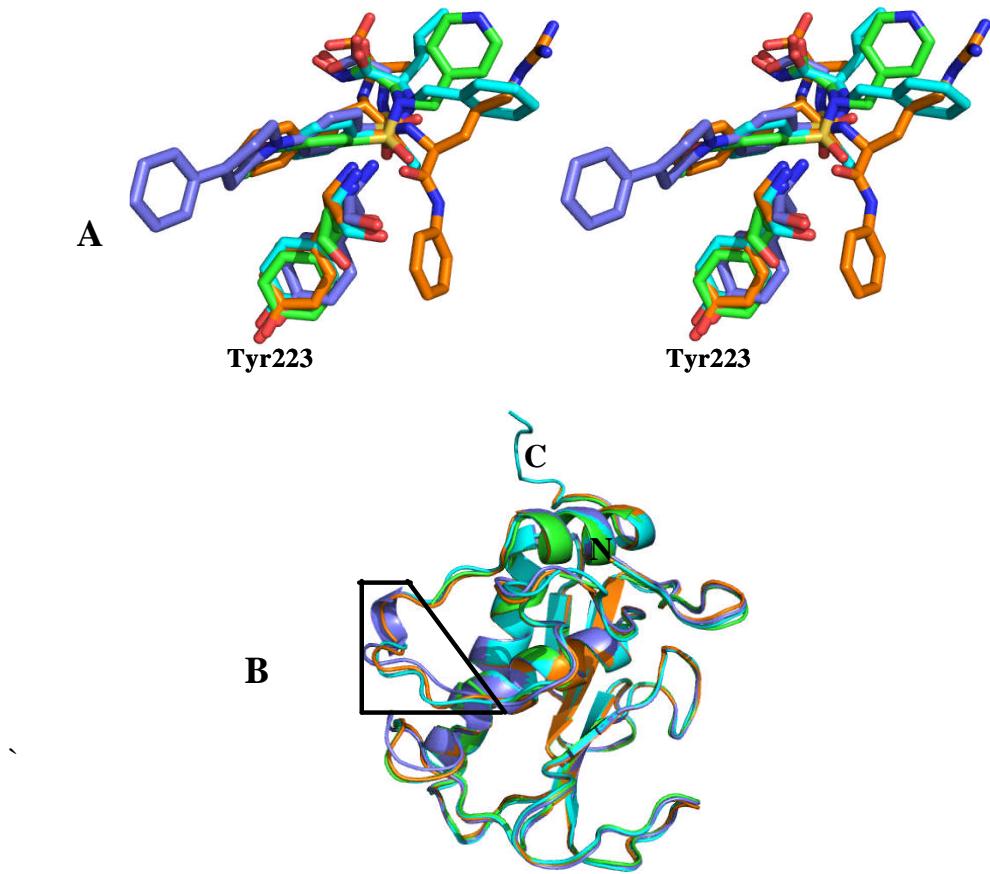


Figure 157. (A) Superposition of Tyr223 residues and inhibitors of SCD structures complexed with the nonpeptide inhibitors 3 (carbons in green), PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999) and I (carbons in blue) (Pdb accession code 1b8y) (Pavlovsky *et al.*, 1999) and the N-carboxyalkyl peptide inhibitor I (C-atoms in orange) (Pdb accession code 1sln) (Becker *et al.*, 1995). (B) Ribbon diagram of the same four superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed. This figure has been produced using PyMol (DeLano, 2007).

However, in SCD structures with inhibitors bound lacking a P1' group (see Section 3.1.3.3) (Finzel *et al.*, 1998), Tyr223 and also the loop from Leu222 to Arg231 adopt the same position and conformation (“S1'-closed conformation”) (Finzel *et al.*, 1998) such as observed in the native structure of the catalytic domain of MMP-3 (Finzel *et al.*, 1998) (Figure 158A and B).

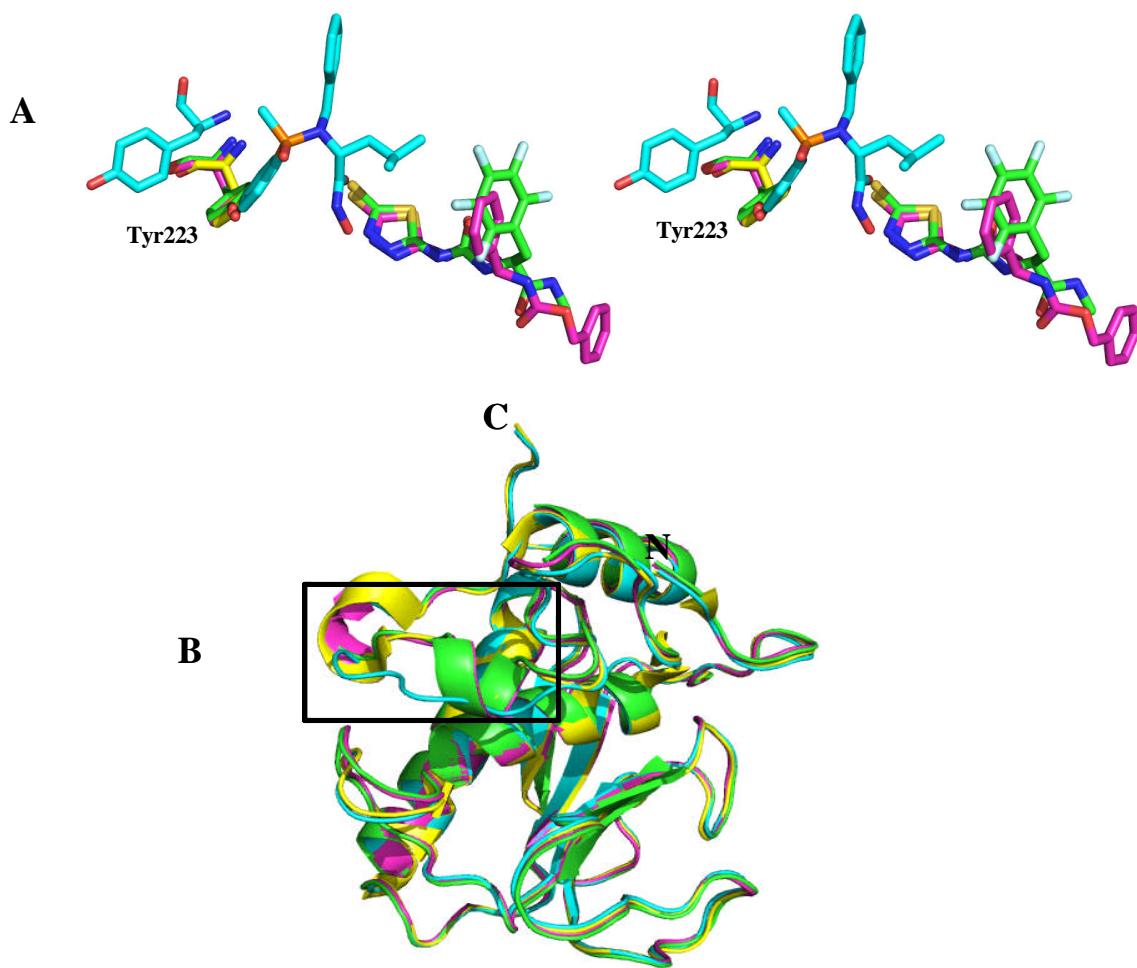


Figure 158. (A) Superposition of Tyr223 residues and inhibitors of SCD structures complexed with the nonpeptide inhibitors PGV-25727 (C-atoms in cyan) (Pdb accession code 1b3d) (Chen *et al.*, 1999), PNU-142372 (carbons in green) (Pdb accession code 1usn) (Finzel *et al.*, 1998), PNU-141803 (C-atoms in magenta) (Pdb accession code 2usn) (Finzel *et al.*, 1998) and the native SDR structure (carbons in yellow) (Pdb accession code 1cqr) (Chen *et al.*, 1999). (B) Ribbon diagram of the same four superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed. This figure has been produced using PyMol (DeLano, 2007).

The significant differences in the location of Tyr223 and the conformation of the S1' loop between the 2.4 Å resolution X-ray structure of SCD complexed with inhibitor **3** and the NMR structure in complex with the similar inhibitor, CGS 27023A (only contains an additional isopropyl group) (Li *et al.*, 1998), is hard to explain as the X-ray and NMR structures of the catalytic domain complexed with the same *N*-carboxyalkyl peptide inhibitor (Becker *et al.*, 1995; Gooley *et al.*, 1994; Gooley *et al.*, 1996) do not display such dramatic distinctions in this region (Figure 159 and B).

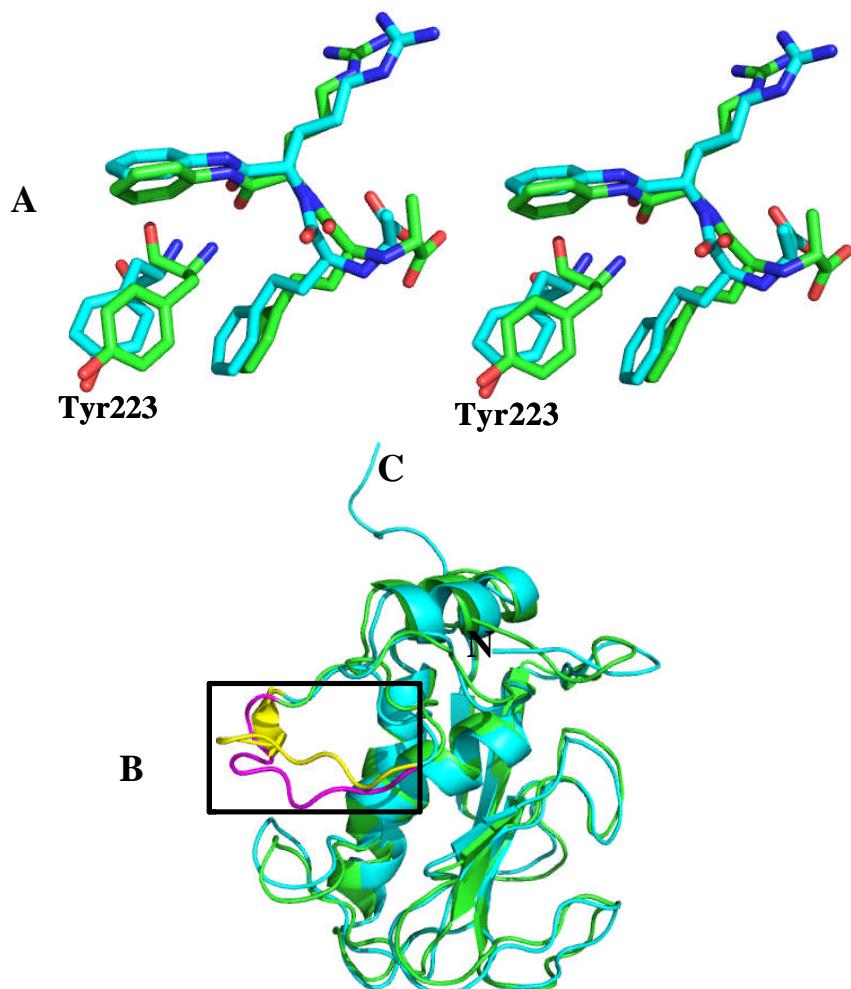


Figure 159. (A) Superposition of Tyr223 residues and the same *N*-carboxyalkyl peptide inhibitor of the crystal (C-atoms in green) (Pdb accession code 1sln) (Becker *et al.*, 1995) and NMR (C-atoms in cyan) (Pdb accession code 2srt) (Gooley *et al.*, 1994) structures of SCD. (B) Ribbon diagram of the same superimposed SCD structures (same colour scheme as in (A)). The part of the loop displaying amino acids Tyr223 to Arg231 (Chen *et al.*, 1999) is boxed and shown in different colours (magenta in the crystal and yellow in the NMR structure). This figure has been produced using PyMol (DeLano, 2007).

Future aims

The structures of the catalytic domain of stromelysin-1 in complex with inhibitors **3** and **4** have been solved to 2.4 and 2.5 Å, respectively. Comparisons to the already published NMR structure in complex with inhibitor **1** (Li *et al.*, 1998), which only contains an additional isopropyl moiety as P1 group (Li *et al.*, 1998) revealed significant differences in the conformation of the flexible S1'loop (Gooley *et al.*, 1996; Chen *et al.*, 1999; Li *et al.*, 1998). Further studies could include further attempts to get crystal structures of SCD in complex with the other inhibitors **1**, **2**, **5**, **6**, **7** and **8**. Particularly, structures with compounds **5** and **6** bound would be of great interest because of their missing hydroxamic acid moiety (see Figure 132). Consequently, the interactions with the catalytic zinc would have to be different.

References

- Adams, G. A.; Quadling, C.; Perry, M. B., D-glycero-D-manno-heptose as a component of lipopolysaccharides from Gram-negative bacteria. *Can J Microbiol* **1967**, 13, , 1605-13.
- Alcaraz, L. A.; Banci, L.; Bertini, I.; Cantini, F.; Donaire, A.; Gonnelli, L., Matrix metalloproteinase-inhibitor interaction: the solution structure of the catalytic domain of human matrix metalloproteinase-3 with different inhibitors. *J Biol Inorg Chem* **2007**, 12, (8), 1197-206.
- Allard, S. T.; Beis, K.; Giraud, M. F.; Hegeman, A. D.; Gross, J. W.; Wilmouth, R. C.; Whitfield, C.; Graninger, M.; Messner, P.; Allen, A. G.; Maskell, D. J.; Naismith, J. H., Toward a structural understanding of the dehydratase mechanism. *Structure* **2002**, 10, (1), 81-92.
- Allard, S. T.; Cleland, W. W.; Holden, H. M., High resolution X-ray structure of dTDP-glucose 4,6-dehydratase from Streptomyces venezuelae. *J Biol Chem* **2004**, 279, (3), 2211-20.
- Allard, S. T.; Giraud, M. F.; Naismith, J. H., Epimerases: structure, function and mechanism. *Cell Mol Life Sci* **2001**, 58, (11), 1650-65.
- Allard, S. T.; Giraud, M. F.; Whitfield, C.; Graninger, M.; Messner, P.; Naismith, J. H., The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from *Salmonella enterica* serovar Typhimurium, the second enzyme in the dTDP-l-rhamnose pathway. *J Mol Biol* **2001a**, 307, (1), 283-95.
- Almstead, N. G.; Bradley, R. S.; Pikul, S.; De, B.; Natchus, M. G.; Taiwo, Y. O.; Gu, F.; Williams, L. E.; Hynd, B. A.; Janusz, M. J.; Dunaway, C. M.; Mieling, G. E., Design, synthesis, and biological evaluation of potent thiazine- and thiazepine-based matrix metalloproteinase inhibitors. *J Med Chem* **1999**, 42, (22), 4547-62.
- Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J Mol Biol* **1990**, 215, (3), 403-10.
- Baker, E. N.; Drenth, J.; The thiol proteases: structure and mechanism. In Jurnak, F. A; McPherson, A.; Biological Macromolecules and Assemblies. *John Wiley & Sons* **1987**, 3, 313-68, New York.

Becker, J. W.; Marcy, A. I.; Rokosz, L. L.; Axel, M. G.; Burbaum, J. J.; Fitzgerald, P. M.; Cameron, P. M.; Esser, C. K.; Hagmann, W. K.; Hermes, J. D.; et al., Stromelysin-1: three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. *Protein Sci* **1995**, 4, (10), 1966-76.

Berger, E.; Arabshahi, A.; Wei, Y.; Schilling, J. F.; Frey, P. A., Acid-base catalysis by UDP-galactose 4-epimerase: correlations of kinetically measured acid dissociation constants with thermodynamic values for tyrosine 149. *Biochemistry* **2001**, 40, (22), 6699-705.

Berti, P. J.; Storer, A. C., Alignment/phylogeny of the papain superfamily of cysteine proteases. *J Mol Biol* **1995**, 246, (2), 273-83.

Birkedal-Hansen, H.; Moore, W. G.; Bodden, M. K.; Windsor, L. J.; Birkedal-Hansen, B.; DeCarlo, A.; Engler, J. A., Matrix metalloproteinases: a review. *Crit Rev Oral Biol Med* **1993**, 4, (2), 197-250.

Blankenfeldt, W.; Kerr, I. D.; Giraud, M. F.; McMiken, H. J.; Leonard, G.; Whitfield, C.; Messner, P.; Graninger, M.; Naismith, J. H., Variation on a theme of SDR. dTDP-6-deoxy-L-lyxo-4-hexulose reductase (RmlD) shows a new Mg²⁺-dependent dimerization mode. *Structure* **2002**, 10, (6), 773-86.

Bode, W.; Gomis-Rüth, F. X.; Stockler, W., Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXX GXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'. *FEBS Lett* **1993**, 331, (1-2), 134-40.

Boutselakis, H.; Dimitropoulos, D.; Fillon, J.; Golovin, A.; Henrick, K.; Hussain, A.; Ionides, J.; John, M.; Keller, P. A.; Krissinel, E.; McNeil, P.; Naim, A.; Newman, R.; Oldfield, T.; Pineda, J.; Rachedi, A.; Copeland, J.; Sitnov, A.; Sobhany, S.; Suarez-Uruena, A.; Swaminathan, J.; Tagari, M.; Tate, J.; Tromm, S.; Velankar, S.; Vranken, W., E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. *Nucleic Acids Res* **2003**, 31, (1), 458-62.

Brömmе, D.; Kirschke, H., N-peptidyl-O-carbamoyl amino acid hydroxamates: irreversible inhibitors for the study of the S2' specificity of cysteine proteinases. *FEBS Lett* **1993**, 322, (3), 211-4.

Bütschli, O., Einige Beobachtungen über Kiesel- und Kalknadeln von Spongiens. *Z. Wiss. Zool* **1901**, 69, 235-286.

Cameron, P. M.; Marcy, A. I.; Rokosz, L. L.; Hermes, J. D., Use of an active-site inhibitor of stromelysin to elucidate the mechanisms of prostromelysin activation. *Bioorg Chem* **1995**, 23, (4), 415-26.

Carmona, E.; Dufour, E.; Plouffe, C.; Takebe, S.; Mason, P.; Mort, J. S.; Menard, R., Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. *Biochemistry* **1996**, 35, (25), 8149-57.

Catania, J. M.; Chen, G.; Parrish, A. R., Role of matrix metalloproteinases in renal pathophysiologies. *Am J Physiol Renal Physiol* **2007**, 292, (3), F905-11.

Cha, J. N.; Shimizu, K.; Zhou, Y.; Christiansen, S. C.; Chmelka, B. F.; Stucky, G. D.; Morse, D. E., Silicatein filaments and subunits from a marine sponge direct the polymerization of silica and silicones in vitro. *Proc Natl Acad Sci U S A* **1999**, 96, (2), 361-5.

Chapman, K. T.; Kopka, I. E.; Durette, P. L.; Esser, C. K.; Lanza, T. J.; Izquierdo-Martin, M.; Niedzwiecki, L.; Chang, B.; Harrison, R. K.; Kuo, D. W.; et al., Inhibition of matrix metalloproteinases by N-carboxyalkyl peptides. *J Med Chem* **1993**, 36, (26), 4293-301.

Chen, L.; Rydel, T. J.; Gu, F.; Dunaway, C. M.; Pikul, S.; Dunham, K. M.; Barnett, B. L., Crystal structure of the stromelysin catalytic domain at 2.0 Å resolution: inhibitor-induced conformational changes. *J Mol Biol* **1999**, 293, (3), 545-57.

Cheng, M.; De, B.; Almstead, N. G.; Pikul, S.; Dowty, M. E.; Dietsch, C. R.; Dunaway, C. M.; Gu, F.; Hsieh, L. C.; Janusz, M. J.; Taiwo, Y. O.; Natchus, M. G.; Hudlicky, T.; Mandel, M., Design, synthesis, and biological evaluation of matrix metalloproteinase inhibitors derived from a modified proline scaffold. *J Med Chem* **1999**, 42, (26), 5426-36.

Cheng, M.; De, B.; Pikul, S.; Almstead, N. G.; Natchus, M. G.; Anastasio, M. V.; McPhail, S. J.; Snider, C. E.; Taiwo, Y. O.; Chen, L.; Dunaway, C. M.; Gu, F.; Dowty, M. E.; Mieling, G. E.; Janusz, M. J.; Wang-Weigand, S., Design and synthesis of piperazine-based matrix metalloproteinase inhibitors. *J Med Chem* **2000**, 43, (3), 369-80.

Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.; Thompson, J. D., Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **2003**, 31, (13), 3497-500.

Chowdhury, S. F.; Joseph, L.; Kumar, S.; Tulsidas, S. R.; Bhat, S.; Ziomek, E.; Menard, R.; Sivaraman, J.; Purisima, E. O., Exploring inhibitor binding at the S' subsites of cathepsin L. *J Med Chem* **2008**, 51, (5), 1361-8.

Chowdhury, S. F.; Sivaraman, J.; Wang, J.; Devanathan, G.; Lachance, P.; Qi, H.; Menard, R.; Lefebvre, J.; Konishi, Y.; Cygler, M.; Sulea, T.; Purisima, E. O., Design of noncovalent inhibitors of human cathepsin L. From the 96-residue proregion to optimized tripeptides. *J Med Chem* **2002**, 45, (24), 5321-9.

Cleland, W. W.; Frey, P. A.; Gerlt, J. A., The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* **1998**, 273, (40), 25529-32.

Coleman, W. G., Jr., The rfaD gene codes for ADP-L-glycero- D-mannoheptose- 6- epimerase. An enzyme required for lipopolysaccharide core biosynthesis. *J Biol Chem* **1983**, 258, (3), 1985-90.

Coleman, W. G., Jr.; Leive, L., Two mutations which affect the barrier function of the Escherichia coli K-12 outer membrane. *J Bacteriol* **1979**, 139, (3), 899-910.

Coulombe, R.; Grochulski, P.; Sivaraman, J.; Menard, R.; Mort, J. S.; Cygler, M., Structure of human procathepsin L reveals the molecular basis of inhibition by the prosegment. *EMBO J* **1996**, 15, (20), 5492-503.

Curnow, P.; Bessette, P. H.; Kisailus, D.; Murr, M. M.; Daugherty, P. S.; Morse, D. E., Enzymatic synthesis of layered titanium phosphates at low temperature and neutral pH by cell-surface display of silicatein-alpha. *J Am Chem Soc* **2005**, 127, (45), 15749-55

Curnow, P.; Kisailus, D.; Morse, D. E., Biocatalytic synthesis of poly(L-lactide) by native and recombinant forms of the silicatein enzymes. *Angew Chem Int Ed Engl* **2006**, 45, (4), 613-6.

Dahl, S. W.; Halkier, T.; Lauritzen, C.; Dolenc, I.; Pedersen, J.; Turk, V.; Turk, B., Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry* **2001**, 40, (6), 1671-8.

Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B., 3rd; Snoeyink, J.; Richardson, J. S.; Richardson, D. C., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **2007**, 35, (Web Server issue), W375-83.

Davis, I. W.; Murray, L. W.; Richardson, J. S.; Richardson, D. C., MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* **2004**, 32, (Web Server issue), W615-9.

Deacon, A. M.; Ni, Y. S.; Coleman, W. G., Jr.; Ealick, S. E., The crystal structure of ADP-L-glycero-D-mannoheptose 6-epimerase: catalysis with a twist. *Structure* **2000**, 8, (5), 453-62.

DeLano, W., The Pymol Molecular Graphics System. *DeLano Scientific* **2007**, Palo Alto, CA, USA.

Dhanaraj, V.; Ye, Q. Z.; Johnson, L. L.; Hupe, D. J.; Ortwine, D. F.; Dunbar, J. B., Jr.; Rubin, J. R.; Pavlovsky, A.; Humblet, C.; Blundell, T. L., X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure* **1996**, 4, (4), 375-86.

Ding, L.; Seto, B. L.; Ahmed, S. A.; Coleman, W. G., Jr., Purification and properties of the Escherichia coli K-12 NAD-dependent nucleotide diphosphosugar epimerase, ADP-L-glycero-D-mannoheptose 6-epimerase. *J Biol Chem* **1994**, 269, (39), 24384-90.

Ding, L.; Zhang, Y.; Deacon, A. M.; Ealick, S. E.; Ni, Y.; Sun, P.; Coleman, W. G., Jr., Crystallization and preliminary X-ray diffraction studies of the lipopolysaccharide core biosynthetic enzyme ADP-L-glycero-D-mannoheptose 6-epimerase from Escherichia coli K-12. *Acta Crystallogr D Biol Crystallogr* **1999**, 55, (Pt 3), 685-8.

Dong, C.; Major, L. L.; Allen, A.; Blankenfeldt, W.; Maskell, D.; Naismith, J. H., High-resolution structures of RmlC from Streptococcus suis in complex with substrate analogs locate the active site of this class of enzyme. *Structure* **2003**, 11, (6), 715-23.

Dunten, P.; Kammlott, U.; Crowther, R.; Levin, W.; Foley, L. H.; Wang, P.; Palermo, R., X-ray structure of a novel matrix metalloproteinase inhibitor complexed to stromelysin. *Protein Sci* **2001**, 10, (5), 923-6.

Eidels, L.; Osborn, M. J., Phosphoheptose isomerase, first enzyme in the biosynthesis of aldoheptose in *Salmonella typhimurium*. *J Biol Chem* **1974**, 249, (17), 5642-8.

Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **2004**, 60, (Pt 12 Pt 1), 2126-32.

Esser, R. E.; Angelo, R. A.; Murphey, M. D.; Watts, L. M.; Thornburg, L. P.; Palmer, J. T.; Talhouk, J. W.; Smith, R. E., Cysteine proteinase inhibitors decrease articular cartilage and bone destruction in chronic inflammatory arthritis. *Arthritis Rheum* **1994**, 37, (2), 236-47.

Esser, C. K.; Bugianesi, R. L.; Caldwell, C. G.; Chapman, K. T.; Durette, P. L.; Girotra, N. N.; Kopka, I. E.; Lanza, T. J.; Levorse, D. A.; MacCoss, M.; Owens, K. A.; Ponpipom, M. M.; Simeone, J. P.; Harrison, R. K.; Niedzwiecki, L.; Becker, J. W.; Marcy, A. I.; Axel, M. G.; Christen, A. J.; McDonnell, J.; Moore, V. L.; Olszewski, J. M.; Saphos, C.; Visco, D. M.; Hagmann, W. K.; et al., Inhibition of stromelysin-1 (MMP-3) by P1'-biphenylylethyl carboxyalkyl dipeptides. *J Med Chem* **1997**, 40, (6), 1026-40.

Evans, P., Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006**, 62, (Pt 1), 72-82.

Fairhead, M.; Johnson, K. A.; Kowatz, T.; McMahon, S. A.; Carter, L. G.; Oke, M.; Liu, H.; Naismith, J. H.; van der Walle, C. F., Crystal structure and silica condensing activities of silicatein alpha-cathepsin L chimeras. *Chem Commun (Camb)* **2008**, (15), 1765-7.

Fairhead, M.; van der Walle, C. F., The heavy-light chain loop of human cathepsin-L modulates its activity and stability. *Protein Pept Lett* **2008**, 15, (1), 47-53.

Fernandez-Catalan, C.; Bode, W.; Huber, R.; Turk, D.; Calvete, J. J.; Lichte, A.; Tschesche, H.; , K., Crystal structure of the complex formed by the membrane type 1-matrix metalloproteinase with the tissue inhibitor of metalloproteinases-2, the soluble progelatinase A receptor. *EMBO J* **1998**, 17, (17), 5238-48.

Filling, C.; Berndt, K. D.; Benach, J.; Knapp, S.; Prozorovski, T.; Nordling, E.; Ladenstein, R.; Jörnvall, H.; Oppermann, U., Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* **2002**, 277, (28), 25677-84.

Finzel, B. C.; Baldwin, E. T.; Bryant, G. L., Jr.; Hess, G. F.; Wilks, J. W.; Trepod, C. M.; Mott, J. E.; Marshall, V. P.; Petzold, G. L.; Poorman, R. A.; O'Sullivan, T. J.; Schostarez, H. J.; Mitchell, M. A., Structural characterizations of nonpeptidic thiadiazole inhibitors of matrix metalloproteinases reveal the basis for stromelysin selectivity. *Protein Sci* **1998**, 7, (10), 2118-26.

Fotouhi, N.; Lugo, A.; Visnick, M.; Lusch, L.; Walsky, R.; Coffey, J. W.; Hanglow, A. C., Potent peptide inhibitors of stromelysin based on the prodomain region of matrix metalloproteinases. *J Biol Chem* **1994**, 269, (48), 30227-31.

Frey, P. A., The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J* **1996**, 10, (4), 461-70.

Fujishima, A.; Imai, Y.; Nomura, T.; Fujisawa, Y.; Yamamoto, Y.; Sugawara, T., The crystal structure of human cathepsin L complexed with E-64. *FEBS Lett* **1997**, 407, (1), 47-50.

Gelb, B. D.; Shi, G. P.; Chapman, H. A.; Desnick, R. J., Pycnodynostosis, a lysosomal disease caused by cathepsin K deficiency. *Science* **1996**, 273, (5279), 1236-8.

Giri, S.; Trewyn, B. G.; Lin, V. S., Mesoporous silica nanomaterial-based biotechnological and biomedical delivery systems. *Nanomed* **2007**, 2, (1), 99-111.

Gomi, T.; Takata, Y.; Fujioka, M., Rat liver S-adenosylhomocysteine. Spectrophotometric study of coenzyme binding. *Biochim Biophys Acta* **1989**, 994, (2), 172-9.

Gomis-Rüth, F. X.; , K.; Betz, M.; Bergner, A.; Huber, R.; Suzuki, K.; Yoshida, N.; Nagase, H.; Brew, K.; Bourenkov, G. P.; Bartunik, H.; Bode, W., Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1. *Nature* **1997**, 389, (6646), 77-81.

Gonnella, N. C.; Li, Y. C.; Zhang, X.; Paris, C. G., Bioactive conformation of a potent stromelysin inhibitor determined by X-nucleus filtered and multidimensional NMR spectroscopy. *Bioorg Med Chem* **1997**, 5, (12), 2193-201.

Gooley, P. R.; O'Connell, J. F.; Marcy, A. I.; Cuca, G. C.; Axel, M. G.; Caldwell, C. G.; Hagmann, W. K.; Becker, J. W., Comparison of the structure of human recombinant short form stromelysin by multidimensional heteronuclear NMR and X-ray crystallography. *J Biomol NMR* **1996**, 7, (1), 8-28.

Gooley, P. R.; O'Connell, J. F.; Marcy, A. I.; Cuca, G. C.; Salowe, S. P.; Bush, B. L.; Hermes, J. D.; Esser, C. K.; Hagmann, W. K.; Springer, J. P.; et al., The NMR structure of the inhibited catalytic domain of human stromelysin-1. *Nat Struct Biol* **1994**, 1, (2), 111-8.

Gorlin, R. J.; Sedano, H.; Anderson, V. E., The Syndrome of Palmar-Plantar Hyperkeratosis and Premature Periodontal Destruction of the Teeth. A Clinical and Genetic Analysis of the Papillon-Lef'vre Syndrome. *J Pediatr* **1964**, 65, 895-908.

Graninger, M.; Nidetzky, B.; Heinrichs, D. E.; Whitfield, C.; Messner, P., Characterization of dTDP-4-dehydrorhamnose 3,5-epimerase and dTDP-4-dehydrorhamnose reductase, required for dTDP-L-rhamnose biosynthesis in *Salmonella enterica* serovar Typhimurium LT2. *J Biol Chem* **1999**, 274, (35), 25069-77.

Gross, J. W.; Hegeman, A. D.; Gerratana, B.; Frey, P. A., Dehydration is catalyzed by glutamate-136 and aspartic acid-135 active site residues in Escherichia coli dTDP-glucose 4,6-dehydratase. *Biochemistry* **2001**, *40*, (42), 12497-504.

Gross, J.; Lapierre, C. M., Collagenolytic activity in amphibian tissues: a tissue culture assay. *Proc Natl Acad Sci U S A* **1962**, *48*, 1014-22.

Gunčar, G.; Pungerčič, G.; Klemenčič, I.; Turk, V.; Turk, D., Crystal structure of MHC class II-associated p41 II fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. *EMBO J* **1999**, *18*, (4), 793-803.

Hart, T. C.; Shapira, L., Papillon-Lefevre syndrome. *Periodontol 2000* **1994**, *6*, 88-100.

Iler, R. K. in The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry, ed. Iler, R. K., *John Wiley & Sons* **1979**, 98-9, New York.

Jez, J. M.; Bennett, M. J.; Schlegel, B. P.; Lewis, M.; Penning, T. M., Comparative anatomy of the aldo-keto reductase superfamily. *Biochem J* **1997**, *326* (Pt 3), 625-36.

Jörnvall, H.; Persson, M.; Jeffery, J., Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. *Proc Natl Acad Sci U S A* **1981**, *78*, (7), 4226-30.

Jörnvall, H.; Persson, B.; Krook, M.; Atrian, S.; Gonzalez-Duarte, R.; Jeffery, J.; Ghosh, D., Short-chain dehydrogenases/reductases (SDR). *Biochemistry* **1995**, *34*, (18), 6003-13.

Jozic, D.; Bourenkov, G.; Lim, N. H.; Visse, R.; Nagase, H.; Bode, W.; , K., X-ray structure of human proMMP-1: new insights into procollagenase activation and collagen binding. *J Biol Chem* **2005**, *280*, (10), 9578-85.

Kallberg, Y.; Oppermann, U.; Jörnvall, H.; Persson, B., Short-chain dehydrogenases/reductases (SDR): coenzyme-based functional assignments in completed genomes. *Eur J Biochem* **2002**, *269*, 4409-17.

Kallberg, Y.; Oppermann, U.; Jörnvall, H.; Persson, B., Short-chain dehydrogenase/reductase (SDR) relationships: a large family with eight clusters common to human, animal, and plant genomes. *Protein Sci* **2002a**, *11*, (3), 636-41.

Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B.; Drenth, J., Structure of papain refined at 1.65 Å resolution. *J Mol Biol* **1984**, *179*, (2), 233-56.

Karrer, K. M.; Peiffer, S. L.; DiTomas, M. E., Two distinct gene subfamilies within the family of cysteine protease genes. *Proc Natl Acad Sci U S A* **1993**, 90, (7), 3063-7.

Katunuma, N.; Kominami, E., Abnormal expression of lysosomal cysteine proteinases in muscle wasting diseases. *Rev Physiol Biochem Pharmacol* **1987**, 108, 1-20.

Kawamura, T.; Ishimoto, N.; Ito, E., Enzymatic synthesis of uridine diphosphate N-acetyl-D-mannosaminuronic acid. *J Biol Chem* **1979**, 254, (17), 8457-65.

Kneidinger, B.; Marolda, C.; Graninger, M.; Zamyatina, A.; McArthur, F.; Kosma, P.; Valvano, M. A.; Messner, P., Biosynthesis pathway of ADP-L-glycero-beta-D-manno-heptose in Escherichia coli. *J Bacteriol* **2002**, 184, (2), 363-9.

Kohno, T.; Hochigai, H.; Yamashita, E.; Tsukihara, T.; Kanaoka, M., Crystal structures of the catalytic domain of human stromelysin-1 (MMP-3) and collagenase-3 (MMP-13) with a hydroxamic acid inhibitor SM-25453. *Biochem Biophys Res Commun* **2006**, 344, (1), 315-22.

Krissinel E.; Henrick, K., Detection of Protein Assemblies in Crystals. *Lecture Notes In Computer Science* **2005**, 3695, 163-174.

Laskowski R A.; MacArthur M. W.; Moss D.S.; Thornton J. M ., PROCHECK - a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **1993**, 26, 283-91.

Leytus, S. P.; Patterson, W. L.; Mangel, W. F., New class of sensitive and selective fluorogenic substrates for serine proteinases. Amino acid and dipeptide derivatives of rhodamine. *Biochem J* **1983**, 215, (2), 253-60.

Li, J.; Brick, P.; O'Hare, M. C.; Skarzynski, T.; Lloyd, L. F.; Curry, V. A.; Clark, I. M.; Bigg, H. F.; Hazleman, B. L.; Cawston, T. E.; et al., Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller. *Structure* **1995** 3, (6), 541-9.

Li, Y. C.; Zhang, X.; Melton, R.; Ganu, V.; Gonnella, N. C., Solution structure of the catalytic domain of human stromelysin-1 complexed to a potent, nonpeptidic inhibitor. *Biochemistry* **1998**, 37, (40), 14048-56.

Liu, H.; Naismith, J. H., A simple and efficient expression and purification system using two newly constructed vectors. *Protein Expr Purif* **2009**, 63, (2), 102-11.

Liu, Y.; Thoden, J. B.; Kim, J.; Berger, E.; Gulick, A. M.; Ruzicka, F. J.; Holden, H. M.; Frey, P. A., Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry* **1997**, 36, (35), 10675-84.

Ljunggren, A.; Redzynia, I.; Alvarez-Fernandez, M.; Abrahamson, M.; Mort, J. S.; Krupa, J. C.; Jaskolski, M.; Bujacz, G., Crystal structure of the parasite protease inhibitor chagasin in complex with a host target cysteine protease. *J Mol Biol* **2007**, 371, (1), 137-53.

MacPherson, L. J.; Bayburt, E. K.; Capparelli, M. P.; Carroll, B. J.; Goldstein, R.; Justice, M. R.; Zhu, L.; Hu, S.; Melton, R. A.; Fryer, L.; Goldberg, R. L.; Doughty, J. R.; Spirito, S.; Blancuzzi, V.; Wilson, D.; O'Byrne, E. M.; Ganu, V.; Parker, D. T., Discovery of CGS 27023A, a non-peptidic, potent, and orally active stromelysin inhibitor that blocks cartilage degradation in rabbits. *J Med Chem* **1997**, 40, (16), 2525-32.

MacPherson, L. J.; Parker, D. T., *U.S. Patent 5455258* **1995**.

Marcy, A. I.; Eiberger, L. L.; Harrison, R.; Chan, H. K.; Hutchinson, N. I.; Hagmann, W. K.; Cameron, P. M.; Boulton, D. A.; Hermes, J. D., Human fibroblast stromelysin catalytic domain: expression, purification, and characterization of a C-terminally truncated form. *Biochemistry* **1991**, 30, (26), 6476-83.

Maroteaux, P.; Lamy, M., La Pyknodysostose. *Presse Med* **1962**, 70, 999-1002.

Matthews, B. W., Solvent content of protein crystals. *J Mol Biol* **1968**, 33, (2), 491-7.

McCoy, A. J.; Grosse-Kunstleve, R. W.; Storoni, L. C.; Read, R. J., Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **2005**, 61, (Pt 4), 458-64.

McGrath, M. E., The lysosomal cysteine proteases. *Annu Rev Biophys Biomol Struct* **1999**, 28, 181-204.

Menard, R.; Carmona, E.; Plouffe, C.; Bromme, D.; Konishi, Y.; Lefebvre, J.; Storer, A. C., The specificity of the S1' subsite of cysteine proteases. *FEBS Lett* **1993**, 328, (1-2), 107-10.

Menard, R.; Carmona, E.; Takebe, S.; Dufour, E.; Plouffe, C.; Mason, P.; Mort, J. S., Autocatalytic processing of recombinant human procathepsin L. Contribution of both intermolecular and unimolecular events in the processing of procathepsin L in vitro. *J Biol Chem* **1998**, 273, (8), 4478-84.

Meng, Q.; Malinovskii, V.; Huang, W.; Hu, Y.; Chung, L.; Nagase, H.; Bode, W.; , K.; Brew, K., Residue 2 of TIMP-1 is a major determinant of affinity and specificity for matrix metalloproteinases but effects of substitutions do not correlate with those of the corresponding P1' residue of substrate. *J Biol Chem* **1999**, 274, (15), 10184-9.

Minor, W.; Cymborowski, M.; Otwinowski, Z., Automatic system for crystallographic data collection and analysis. *Acta Physica Polonica A* **2002**, 101, 613-9.

Morphy, J. R.; Millican, T. A.; Porter, J. R., Matrix metalloproteinase inhibitors: Current status. *Curr Med Chem* **1995**, 2, 743–62.

Morrison, J. P.; Read, J. A.; Coleman, W. G., Jr.; Tanner, M. E., Dismutase activity of ADP-L-glycero-D-manno-heptose 6-epimerase: evidence for a direct oxidation/reduction mechanism. *Biochemistry* **2005**, 44, (15), 5907-15.

Morrison, J. P.; Tanner, M. E., A two-base mechanism for Escherichia coli ADP-L-glycero-D-manno-heptose 6-epimerase. *Biochemistry* **2007**, 46, (12), 3916-24.

Mulichak, A. M.; Theisen, M. J.; Essigmann, B.; Benning, C.; Garavito, R. M., Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. *Proc Natl Acad Sci U S A* **1999**, 96, (23), 13097-102.

Müller, W. E.; Boreiko, A.; Wang, X.; Belikov, S. I.; Wiens, M.; Grebenjuk, V. A.; Schlossmacher, U.; Schröder, H. C., Silicateins, the major biosilica forming enzymes present in demosponges: protein analysis and phylogenetic relationship. *Gene* **2007**, 395, (1-2), 62-71.

Müller, W. E.; Krasko, A.; Le Pennec, G.; Schröder, H. C., Biochemistry and cell biology of silica formation in sponges. *Microsc Res Tech* **2003**, 62, (4), 368-77.

Murshudov, G. B.; Dodson, E. J., Simplified error estimation a la Cruickshank in macromolecular crystallography. *CCP4 Newsletter* **1997**, Chemistry Department, University of York, York, UK.

Murshudov, G. N.; Vagin, A. A.; Lebedev, A.; Wilson, K. S.; Dodson, E. J., Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* **1999**, 55, (Pt 1), 247-55.

Musil, D.; Zucic, D.; Turk, D.; Engh, R. A.; Mayr, I.; Huber, R.; Popovic, T.; Turk, V.; Towatari, T.; Katunuma, N.; et al., The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J* **1991**, 10, (9), 2321-30.

Nagase, H.; Suzuki, K.; Enghild, J. J.; Salvesen, G., Stepwise activation mechanisms of the precursors of matrix metalloproteinases 1 (tissue collagenase) and 3 (stromelysin). *Biomed Biochim Acta* **1991**, 50, (4-6), 749-54.

Nagase, H.; Visse, R.; Murphy, G., Structure and function of matrix metalloproteinases and TIMPs. *Cardiovasc Res* **2006**, 69, (3), 562-73.

Naismith, J. H., Inferring the chemical mechanism from structures of enzymes. *Chem Soc Rev* **2006**, 35, (9), 763-70.

Ni, Y.; McPhie, P.; Deacon, A.; Ealick, S.; Coleman, W. G., Jr., Evidence that NADP⁺ is the physiological cofactor of ADP-L-glycero-D-mannoheptose 6-epimerase. *J Biol Chem* **2001**, 276, (29), 27329-34.

Oppermann, U.; Filling, C.; Hult, M.; Shafqat, N.; Wu, X.; Lindh, M.; Shafqat, J.; Nordling, E.; Kallberg, Y.; Persson, B.; Jörnvall, H., Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact* **2003**, 143-144, 247-53.

Oppermann, U. C.; Persson, B.; Filling, C.; Jörnvall, H., Structure-function relationships of SDR hydroxysteroid dehydrogenases. *Adv Exp Med Biol* **1997**, 414, 403-15.

Ortwine, D. F.; Dhanaraj, V.; Dunbar, J. B. Jr.; Johnson, L. L.; Pavlovsky, A., Purchase, C. F II.; White, A. D.; Ye. Q-Z. Structure based design of matrix metalloproteinase inhibitors: What have the structure taught us? *Abstract of ACS National Meeting* **1997**, Las Vegas, Nevada.

Osborn, M. J.; Gander, J. E.; Parisi, E.; Carson, J., Mechanism of assembly of the outer membrane of *Salmonella typhimurium*. Isolation and characterization of cytoplasmic and outer membrane. *J Biol Chem* **1972**, 247, (12), 3962-72.

Overall, C. M., Molecular determinants of metalloproteinase substrate specificity: matrix metalloproteinase substrate binding domains, modules, and exosites. *Mol Biotechnol* **2002**, 22, (1), 51-86.

Page-McCaw, A.; Ewald, A. J.; Werb, Z., Matrix metalloproteinases and the regulation of tissue remodelling. *Nat Rev Mol Cell Biol* **2007**, 8, (3), 221-33.

Pavlovsky, A. G.; Williams, M. G.; Ye, Q. Z.; Ortwine, D. F.; Purchase, C. F., 2nd; White, A. D.; Dhanaraj, V.; Roth, B. D.; Johnson, L. L.; Hupe, D.; Humblet, C.; Blundell, T. L., X-ray structure of human stromelysin catalytic domain complexed with nonpeptide inhibitors: implications for inhibitor selectivity. *Protein Sci* **1999**, 8, (7), 1455-62.

Pegues, J. C.; Chen, L. S.; Gordon, A. W.; Ding, L.; Coleman, W. G., Jr., Cloning, expression, and characterization of the Escherichia coli K-12 rfaD gene. *J Bacteriol* **1990**, 172, (8), 4652-60.

Persson, B.; Kallberg, Y.; Oppermann, U.; Jörnvall, H., Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact* **2003**, 143-144, 271-8.

Pierce, M. M.; Raman, C. S.; Nall, B. T., Isothermal titration calorimetry of protein-protein interactions. *Methods* **1999**, 19, (2), 213-21.

Pikul, S.; Dunham, K. M.; Almstead, N. G.; De, B.; Natchus, M. G.; Taiwo, Y. O.; Williams, L. E.; Hynd, B. A.; Hsieh, L. C.; Janusz, M. J.; Gu, F.; Mieling, G. E., Heterocycle-based MMP inhibitors with P2' substituents. *Bioorg Med Chem Lett* **2001**, 11, (8), 1009-13.

Polgár, L., The catalytic triad of serine peptidases. *Cell Mol Life Sci* **2005**, 62, (19-20), 2161-72.

Polymeropoulos, M. H.; Ortiz De Luna, R. I.; Ide, S. E.; Torres, R.; Rubenstein, J.; Francomano, C. A., The gene for pycnodynatosostosis maps to human chromosome 1cen-q21. *Nat Genet* **1995**, 10, (2), 238-9.

Raetz, C. R., Biochemistry of endotoxins. *Annu Rev Biochem* **1990**, 59, 129-70.

Raetz, C. R.; Whitfield, C., Lipopolysaccharide endotoxins. *Annu Rev Biochem* **2002**, 71, 635-700.

Read, J. A.; Ahmed, R. A.; Morrison, J. P.; Coleman, W. G., Jr.; Tanner, M. E., The mechanism of the reaction catalyzed by ADP-beta-L-glycero-D-manno-heptose 6-epimerase. *J Am Chem Soc* **2004**, 126, (29), 8878-9.

Reeves, P. R.; in Bacterial Cell Wall (Ghuysen, J. M.; Hackenbeck, R., eds). *Elsevier Science Publisher* **1994**, 281-314, Amsterdam, The Netherlands.

Reeves, P. R.; Hobbs, M.; Valvano, M. A.; Skurnik, M.; Whitfield, C.; Coplin, D.; Kido, N.; Klena, J.; Maskell, D.; Raetz, C. R.; Rick, P. D., Bacterial polysaccharide synthesis and gene nomenclature. *Trends Microbiol* **1996**, 4, (12), 495-503.

Rhodes, G., Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models. *Academic Press* **2000**, Second edn, London.

Rizzi, M.; Tonetti, M.; Vigevani, P.; Sturla, L.; Bisso, A.; Flora, A. D.; Bordo, D.; Bolognesi, M., GDP-4-keto-6-deoxy-D-mannose epimerase/reductase from *Escherichia coli*, a key enzyme in the biosynthesis of GDP-L-fucose, displays the structural characteristics of the RED protein homology superfamily. *Structure* **1998**, 6, (11), 1453-65.

Rossmann, M. G.; Liljas, A.; Brändén, C. I.; Banaszak, L. J., Evolutionary and structural relationships among dehydrogenases. In: P.D. Boyer, Editor, (third ed.), *The Enzymes*. Academic Press **1975**, 11, 61–102, New York.

Rossmann, M. G.; Moras, D.; Olsen, K. W., Chemical and biological evolution of nucleotide-binding protein. *Nature* **1974**, 250, (463), 194-9.

Rowan, A. D.; Mason, P.; Mach, L.; Mort, J. S., Rat procathepsin B. Proteolytic processing to the mature form in vitro. *J Biol Chem* **1992**, 267, (22), 15993-9.

Rupp, B.; Segelke, B. W.; Krupka, H. I.; Lekin, T.; Schafer, J.; Zemla, A.; Toppani, D.; Snell, G.; Earnest, T., The TB structural genomics consortium crystallization facility: towards automation from protein to electron density. *Acta Crystallogr D Biol Crystallogr* **2002**, 58, (Pt 10 Pt 1), 1514-8.

Saftig, P.; Hunziker, E.; Wehmeyer, O.; Jones, S.; Boyde, A.; Rommerskirch, W.; Moritz, J. D.; Schu, P.; von Figura, K., Impaired osteoclastic bone resorption leads to osteopetrosis in cathepsin-K-deficient mice. *Proc Natl Acad Sci U S A* **1998**, 95, (23), 13453-8.

Sajid, M.; McKerrow, J. H., Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol* **2002**, 120, (1), 1-21.

Samuel, J.; Tanner, M. E., Mechanistic aspects of enzymatic carbohydrate epimerization. *Nat Prod Rep* **2002**, 19, (3), 261-77.

Schechter, I.; Berger, A., On the size of the active site in proteases. I. Papain. *Biochem Biophys Res Commun* **1967**, 27, (2), 157-62.

Schlieben, N. H.; Niefind, K.; Muller, J.; Riebel, B.; Hummel, W.; Schomburg, D., Atomic resolution structures of R-specific alcohol dehydrogenase from *Lactobacillus brevis* provide the structural bases of its substrate and cosubstrate specificity. *J Mol Biol* **2005**, 349, (4), 801-13.

Schnaitman, C. A.; Klena, J. D., Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* **1993**, 57, (3), 655-82.

Schröder, H. C.; Boreiko, A.; Korzhev, M.; Tahir, M. N.; Tremel, W.; Eckert, C.; Ushijima, H.; Müller, I. M.; Müller, W. E., Co-expression and functional interaction of silicatein with galectin: matrix-guided formation of siliceous spicules in the marine demosponge *Suberites domuncula*. *J Biol Chem* **2006**, 281, (17), 12001-9.

Schröder, H. C.; Perovic-Ottstadt, S.; Rothenberger, M.; Wiens, M.; Schwertner, H.; Batel, R.; Korzhev, M.; Müller, I. M.; Müller, W. E., Silica transport in the demosponge *Suberites domuncula*: fluorescence emission analysis using the PDMPO probe and cloning of a potential transporter. *Biochem J* **2004**, 381, (Pt 3), 665-73.

Schulze, F. E., Hexactinellida. Wissenschaftliche Ergebnisse der Deutschen Tiefsee-Expedition. *Fischer-Verlag* **1904**, 4, Jena.

Schuster, M.; Kasche, V.; Jakubke, H. D., Contributions to the S'-subsite specificity of papain. *Biochim Biophys Acta* **1992**, 1121, (1-2), 207-12.

Shimizu, K.; Cha, J.; Stucky, G. D.; Morse, D. E., Silicatein alpha: cathepsin L-like protein in sponge biosilica. *Proc Natl Acad Sci U S A* **1998**, 95, (11), 6234-8.

Simpson, T. L., The Cell Biology of Sponges. *Springer* **1984**, New York.

Smeland, T. E.; Cuevas, D.; Schulz, H., Epimerization of 3-hydroxy-4-trans-decenoyl coenzyme A by a dehydration/hydration mechanism catalyzed by the multienzyme complex of fatty acid oxidation from *Escherichia coli*. *J Biol Chem* **1991**, 266, (35), 23904-8.

Smith, S. M.; Gottesman, M. M., Activity and deletion analysis of recombinant human cathepsin L expressed in *Escherichia coli*. *J Biol Chem* **1989**, 264, (34), 20487-95.

Somers, W. S.; Stahl, M. L.; Sullivan, F. X., GDP-fucose synthetase from *Escherichia coli*: structure of a unique member of the short-chain dehydrogenase/reductase family that catalyzes two distinct reactions at the same active site. *Structure* **1998**, 6, (12), 1601-12.

Sottrup-Jensen, L.; Birkedal-Hansen, H., Human fibroblast collagenase- alpha-macroglobulin interactions. Localization of cleavage sites in the bait regions of five mammalian alpha-macroglobulins. *J Biol Chem* **1989**, 264, (1), 393-401.

Steele, D. L.; El-Kabbani, O.; Dunten, P.; Windsor, L. J.; Kammlott, R. U.; Crowther, R. L.; Michoud, C.; Engler, J. A.; Birktoft, J. J., Expression, characterization and structure determination of an active site mutant (Glu202-Gln) of mini-stromelysin-1. *Protein Eng* **2000**, 13, (6), 397-405.

Storoni, L. C.; McCoy, A. J.; Read, R. J., Likelihood-enhanced fast rotation functions. *Acta Crystallogr D Biol Crystallog* **2004**, 60, (Pt 3), 432-8.

Tahir, M. N.; Theato, P.; Müller, W. E.; Schröder, H. C.; Borejko, A.; Faiss, S.; Janshoff, A.; Huth, J.; Tremel, W., Formation of layered titania and zirconia catalysed by surface-bound silicatein. *Chem Commun (Camb)* **2005**, (44), 5533-5.

Tahir, M. N.; Theato, P.; Müller, W. E.; Schröder, H. C.; Janshoff, A.; Zhang, J.; Huth, J.; Tremel, W., Monitoring the formation of biosilica catalysed by histidine-tagged silicatein. *Chem Commun (Camb)* **2004**, (24), 2848-9.

Tanaka, N.; Nonaka, T.; Nakanishi, M.; Deyashiki, Y.; Hara, A.; Mitsui, Y., Crystal structure of the ternary complex of mouse lung carbonyl reductase at 1.8 Å resolution: the structural origin of coenzyme specificity in the short-chain dehydrogenase/reductase family. *Structure* **1996**, 4, (1), 33-45.

Tanner, M. E., NAD⁺-dependent Carbohydrate/Inositol Epimerases. In Comprehensive Biological Catalysis; Sinnott, M. (Ed.), *Academic Press* **1998**, Vol. III, Chapter 28ii, San Diego.

Tanner, M. E. Sugar Nucleotide-Modifying Enzymes. *Curr Org Chem* **2001**, 5, 169-192.

Tanner, M. E., Understanding nature's strategies for enzyme-catalyzed racemization and epimerization. *Acc Chem Res* **2002**, 35, (4), 237-46.

Tanner, M. E.; Kenyon, G. L. Inversion of Stereocenters. In Comprehensive Biological Catalysis; Sinnott, M. (Ed.), *Academic Press* **1998**, Vol. II, Chapter 21, San Diego.

Taylor, P. W., Bactericidal and bacteriolytic activity of serum against gram-negative bacteria. *Microbiol Rev* **1983**, 47, (1), 46-83.

Thoden, J. B.; Frey, P. A.; Holden, H. M., Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from Escherichia coli. *Biochemistry* **1996**, 35, (8), 2557-66.

Thoden, J. B.; Frey, P. A.; Holden, H. M., Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: implications for the catalytic mechanism. *Biochemistry* **1996a**, 35, (16), 5137-44.

Thoden, J. B.; Hegeman, A. D.; Frey, P. A.; Holden, H. M., Molecular Structure of DTDP-Glucose 4,6-Dehydratase from E. coli (unpublished).

Thoden, J. B.; Wohlers, T. M.; Fridovich-Keil, J. L.; Holden, H. M., Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. *Biochemistry* **2000**, 39, (19), 5691-701.

Toomes, C.; James, J.; Wood, A. J.; Wu, C. L.; McCormick, D.; Lench, N.; Hewitt, C.; Moynihan, L.; Roberts, E.; Woods, C. G.; Markham, A.; Wong, M.; Widmer, R.; Ghaffar, K. A.; Pemberton, M.; Hussein, I. R.; Temtamy, S. A.; Davies, R.; Read, A. P.; Sloan, P.; Dixon, M. J.; Thakker, N. S., Loss-of-function mutations in the cathepsin C gene result in periodontal disease and palmoplantar keratosis. *Nat Genet* **1999**, 23, (4), 421-4.

Turk, D.; Guncar, G.; Podobnik, M.; Turk, B., Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol Chem* **1998**, 379, (2), 137-47.

Turk, D.; Podobnik, M.; Popovic, T.; Katunuma, N.; Bode, W.; Huber, R.; Turk, V., Crystal structure of cathepsin B inhibited with CA030 at 2.0-A resolution: A basis for the design of specific epoxysuccinyl inhibitors. *Biochemistry* **1995**, 34, (14), 4791-7.

Turk, B.; Turk, D.; Turk, V., Lysosomal cysteine proteases: more than scavengers. *Biochim Biophys Acta* **2000**, 1477, (1-2), 98-111.

Turk, V.; Turk, B.; Turk, D., Lysosomal cysteine proteases: facts and opportunities. *EMBO J* **2001**, 20, (17), 4629-33.

Vagin, A.; Teplyakov, A., MOLREP: an Automated Program for Molecular Replacement. *J Appl Cryst* **1997**, 30, 1022-5.

Valvano, M. A.; Messner, P.; Kosma, P., Novel pathways for biosynthesis of nucleotide-activated glycero-manno-heptose precursors of bacterial glycoproteins and cell surface polysaccharides. *Micriobiology* **2002**, 148, (Pt 7), 1979-89.

van Hylckama Vlieg, J. E.; Tang, L.; Lutje Spelberg, J. H.; Smilda, T.; Poelarends, G. J.; Bosma, T.; van Merode, A. E.; Fraaije, M. W.; Janssen, D. B., Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *J Bacteriol* **2001**, 183, (17), 5058-66.

Visse, R.; Nagase, H., Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry. *Circ Res* **2003**, 92, (8), 827-39.

Wakil, S. J., Studies on the fatty acid oxidizing system of animal tissues. IX. Stereospecificity of unsaturated acyl CoA hydrase. *Biochim Biophys Acta* **1956**, 19, (3), 497-504.

Weaver, J. C.; Morse, D. E., Molecular biology of demosponge axial filaments and their roles in biosilicification. *Microsc Res Tech* **2003**, 62, (4), 356-67.

Welgus, H. G., Stromelysin: structure and function. *Agents Actions Suppl* **1991**, 35, 61-7.

Welgus, H. G.; Kobayashi, D. K.; Jeffrey, J. J., The collagen substrate specificity of rat uterus collagenase. *J Biol Chem* **1983**, 258, (23), 14162-5.

White, A. D.; Purchase, C. F II.; Baragi, V.; Finkel, M.; Hallak, H.; Hupe, D.; Johnson, L. L.; Kindt, E. K.; Laemont, K.; Ortwine, D. F.; Pavlovsky, A.; Renkiewicz, R. R.; Roth, B.D.; Schrier, D. J.; Ye, Q-Z., Selective non-peptide inhibitor of matrix metalloproteinases. *Abstract of ACS National Meeting* **1997**, Las Vegas, Nevada.

Wierenga, R. K.; Terpstra, P.; Hol, W. G., Prediction of the occurrence of the ADP-binding beta alpha beta-fold in proteins, using an amino acid sequence fingerprint. *J Mol Biol* **1986**, 187, (1), 101-7.

Wilfong, E., Thermodynamic Basis of Additivity in Ligand Binding: Matrix Metalloproteinase-3, Dissertation at Duke University, Durham, NC, **2008**.

Willenbrock, F; Murphy, G., Structure-function relationships in the tissue inhibitors of metalloproteinases. *Am J Respir Crit Care Med* **1994**, 150, (6 Pt 2), S165-70.

Yagel, S.; Warner, A. H.; Nellans, H. N.; Lala, P. K.; Waghorne, C.; Denhardt, D. T., Suppression by cathepsin L inhibitors of the invasion of amnion membranes by murine cancer cells. *Cancer Res* **1989**, 49, (13), 3553-7.

Yang, S. Y.; Elzinga, M., Association of both enoyl coenzyme A hydratase and 3-hydroxyacyl coenzyme A epimerase with an active site in the amino-terminal domain of the multifunctional fatty acid oxidation protein from Escherichia coli. *J Biol Chem* **1993**, 268, (9), 6588-92.

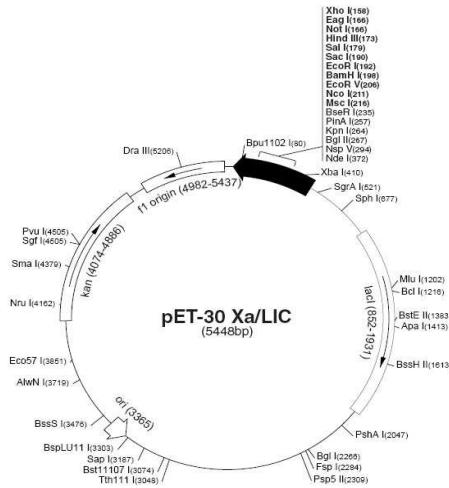
Yang, X. Y.; Schulz, H.; Elzinga, M.; Yang, S. Y., Nucleotide sequence of the promoter and fadB gene of the fadBA operon and primary structure of the multifunctional fatty acid oxidation protein from Escherichia coli. *Biochemistry* **1991**, 30, (27), 6788-95.

Ye, Q. Z.; Johnson, L. L.; Hupe, D. J.; Baragi, V., Purification and characterization of the human stromelysin catalytic domain expressed in Escherichia coli. *Biochemistry* **1992**, 31, (45), 11231-5.

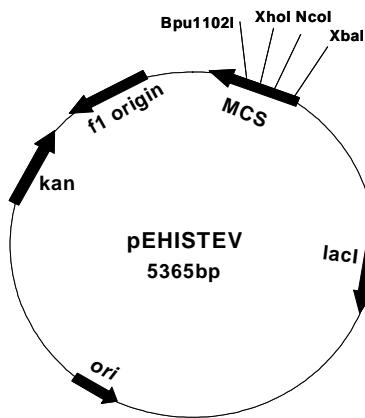
Zhou, Y.; Shimizu, K.; Cha, J. N.; Stucky ,G.D.; Morse D.E., Efficient catalysis of polysiloxane synthesis by silicatein α requires specific hydroxy and imidazole functionalities, *Angew Chem Int Ed Engl* **1999**, 38, 780–82.

Zwart, P. H.; Grosse-Kunstleve, R. W.; Adams, P. D., Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 newsletter* **2005**.

Appendix



(taken from http://ecoliwiki.net/collection/index.php/pET-30_Xa/LIC) The map of the Novagen pET-30 Xa/LIC vector, including cleavage sites of important restriction enzymes, the multiple cloning site (MCS, black arrow), the origin of replication (ori), a kanamycin resistance marker (kan), the f1 phage origin and the gene encoding the lac-repressor (*lacI*). Furthermore it encodes an aminoterminal hexa-histidine tag and contains a potentially cleavable Factor Xa site.



The map of the pEHISTEV vector including cleavage sites of important restriction enzymes, the multiple cloning site (MCS), the origin of replication (ori), a kanamycin resistance marker (kan), the f1 phage origin and the gene encoding the lac-repressor (*lacI*). Furthermore it encodes an N-terminal hexa-histidine tag and contains a tobacco etch virus (TEV) protease cleavage site (Liu and Naismith, 2009).

“Grid screens” around crystallisation conditions Classics # 74 and JCSG⁺ # 46

Starting condition: Classics, condition # 74

0.2 M CaCl₂

0.1 M Hepes-Na pH 7.5

28 % PEG 400

	0 mM Salt	100 mM Salt	200 mM Salt	300 mM Salt
CaCl ₂	1	2	3	4
MgCl ₂	1	5	6	7

All with 0.1 M Hepes-Na pH 7.5, 28 % PEG 400

pH	6.0	6.2	6.4	6.6	6.8	7.0	7.1	7.3	7.7	8.0
tube	8	9	10	11	12	13	14	15	16	17

All with 0.2 M CaCl₂, 28 % PEG 400. Na-Cacodylate for pH 6.0-7.0, Hepes-Na for pH 7.1-8

% PEG 400	25	26	27	29	30	31	32	33	34	35
tube	18	19	20	21	22	23	24	25	26	27

All with 0.1 M Hepes-Na pH 7.5, 0.2 M CaCl₂

Starting condition: JCSG⁺, condition # 46

0.2 M Calcium Acetate

0.1 M Na-Cacodylate pH 6.5

40 % PEG 300

	0 mM Salt	100 mM Salt	200 mM Salt	300 mM Salt
Ca-Acetate	28	29	30	31
Mg-Acetate		32	33	34

All with 0.1 M Na-Cacodylate pH 6.5, 40 % PEG 300

pH	6	6.2	6.4	6.6	6.8	7.0	7.1	7.3	7.7	8.0
tube	35	36	37	38	39	40	41	42	43	44

All with 0.2 M Calcium Acetate, 40% PEG 300. Na-Cacodylate for pH 6-7.0, Hepes-Na for pH 7.1-8

% PEG300	35	36	37	38	39	41	42	43	44	45
tube	45	46	47	48	49	50	51	52	53	54

All with 0.1 M Na-Cacodylate pH 6.5, 0.2 M Calcium Acetate

“Grid screens” around crystallisation conditions JCSG+ # 2 and # 78

Starting condition: JCSG⁺, condition # 2

0.1 M tri-Sodium citrate pH 5.5

20 % PEG 3000

	0.05 M salt	0.1 M salt	0.2 M salt
Tri-Sodium citrate pH 5.5	1	2	3

All with 20 % PEG 3000

pH	5.0	5.2	5.4	5.6	5.8	6.0
tube	4	5	6	7	8	9

All with 0.1 M tri-Sodium citrate, 20 % PEG 3000

% PEG 3000	17	18	19	21	22	23	24	25
tube	10	11	12	13	14	15	16	17

All with 0.1 M tri-Sodium citrate pH 5.5

Starting condition: JCSG⁺, condition # 78

0.2 M Na-malonate pH 7.0

20 % PEG 3350

	0.1 M salt	0.2 M salt	0.3 M salt
Na-malonate pH 7.0	18	19	20

All with 20 % PEG 3350

pH	6.6	6.8	7.2	7.4
tube	21	22	23	24

All with 0.2 M Na-malonate, 20 % PEG 3350

% PEG 3350	17	18	19	21	22	23	24	25
tube	25	26	27	28	29	30	31	32

All with 0.2 M Na-malonate, pH 7.0

“Grid screen” around crystallisation condition Classics # 5

Starting condition: Classics, condition # 5

0.1 M Hepes-Na pH 7.5

10 % 2-propanol

20 % PEG 4000

0.1 M HEPES-Na	pH 7.1	pH 7.3	pH 7.5	pH 7.7	pH 7.9
Tube	1	2	3	4	5

All with 10 % 2-propanol, 20 % PEG 4000

2-propanol	5 %	6 %	7 %	8 %	9 %	11 %	12 %	13 %	14 %	15 %
tube	6	7	8	9	10	11	12	13	14	15

All with Hepes-Na pH 7.5, 20 % PEG 4000

PEG 4000	15 %	16 %	17 %	18 %	19 %	21 %	22 %	23 %	24 %	25 %
tube	16	17	18	19	20	21	22	23	24	25

All with 0.1 M Hepes-Na pH 7.5, 10 % 2-propanol

“Grid screen” around crystallisation condition JSG⁺ # 46

Starting condition: JCSG⁺, condition # 46

0.2 M Calcium Acetate

0.1 M Na Cacodylate pH 6.5

40 % PEG 300

Ca-Acetate	100 mM Salt	200 mM Salt	300 mM Salt
tube	1	2	3

All with 0.1 M Na-Cacodylate pH 6.5, 40 % PEG 300

Na-cacodylate pH	6	6.2	6.4	6.6	6.8	7.0
tube	4	5	6	7	8	9

All with 0.2 M Calcium Acetate, 40% PEG 300

% PEG300	35	36	37	38	39	41	42	43
tube	10	11	12	13	14	15	16	17

All with 0.1 M Na-Cacodylate pH 6.5, 0.2 M Calcium Acetate

“Grid screens” around crystallisation conditions PEGs # 67, # 54 and # 76

starting condition: PEGs, condition # 67

0.2 M NH₄NO₃

20 % PEG 3350

[NH ₄ NO ₃]	0.1 M	0.15 M	0.2 M	0.25 M	0.3 M
Tube	1	2	3	4	5

All with 20 % PEG 3350

[PEG 3350]	15%	16%	17%	18%	19%	21%	22%	23%	24 %	25 %
Tube	6	7	8	9	10	11	12	13	14	15

All with 0.2 M NH₄NO₃

starting condition: PEGs, condition # 54

0.2 M NaCl

20 % PEG 3350

[NaCl]	0.1 M	0.2 M	0.3 M
Tube	1	2	3

All with 20 % PEG 3350

[PEG 3350]	15%	17%	19%	22%	25%
Tube	4	5	6	7	8

All with 0.2 M NaCl

starting condition: PEGs, condition #76

0.2 M Ca-Acetate

20 % PEG 3350

[Ca-Acetate]	0.1 M	0.15 M	0.2 M	0.25 M	0.3 M	0.2 M (Mg-acetate tetrahydrate)
Tube	1	2	3	4	5	6

All with 20 % PEG 3350

[PEG 3350]	15%	16%	17%	18%	19%	21%	22%	23%	24 %	25 %
Tube	7	8	9	10	11	12	13	14	15	16

All with 0.2 M Ca-Acetate

“Grid screens” around crystallisation conditions Classics # 82 and # 73

starting condition: Classics, condition # 82

0.01 M NiCl₂ hexahydrate

0.1 M Tris-HCl, pH 8.5

20 % PEG 2000 MME

[NiCl ₂]	0	0.01 M	0.03 M	0.05 M	0.07 M	0.1	0.2
Tube	1	2	3	4	5	6	7

All with 0.1 M Tris-HCl, pH 8.5, 20 % PEG 2000 MME

Tris-HCl	pH 7.7	pH 7.9	pH 8.1	pH 8.4	pH 8.6
Tube	8	9	10	11	12

All with 0.01 M NiCl₂, 20 % PEG 2000 MME

[PEG 2000 MME]	15%	16%	17%	18%	19%	21%	22%	23%	24 %	25 %
Tube	13	14	15	16	17	18	19	20	21	22

All with 0.01 M NiCl₂, 0.1 M Tris-HCl, pH 8.5

starting condition: Classics, condition # 73

0.1 M HEPES-Na pH 7.5

2 M (NH₄)₂SO₄

2 % PEG 400

[(NH ₄) ₂ SO ₄]	1.5 M	1.75 M	2 M	2.25 M	2.5 M
Tube	1	2	3	4	5

All with 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400

0.1 M HEPES-Na	pH 7.1	pH 7.3	pH 7.7
Tube	6	7	8

All with 2 M (NH₄)₂SO₄, 2 % PEG 400

[PEG 400]	1 %	3 %	4 %	5 %	6 %	7 %
Tube	9	10	11	12	13	14

All with 2 M (NH₄)₂SO₄, 0.1 M HEPES-Na, pH 7.5

[Spermidine]	10 mM	15 mM	20 mM	25 mM	30 mM
Tube	7	8	9	10	11

All with 2 M (NH₄)₂SO₄, 0.1 M HEPES-Na, pH 7.5, 2 % PEG 400