3 RESULTS & DISCUSSION

3.1 Crystallisation

Crystallisation conditions for AglA were first identified using a PEG-Ion screen. Once initial crystals were found, the conditions were refined by steps of 0.2 pH units and 0.5% precipitant concentration around the initial condition. The optimal general crystallisation conditions consisted of 5-9% PEG 6000, 1M LiCl and 100mM NaAc buffer, pH 4.4 – 4.6. Table 3-1 lists the detailed conditions of the crystals for which ultimately diffraction data were collected. Under these conditions, native AglA crystallized in rhombohedral plates which grew within 12 – 36 hours. The final size was usually 0.2 x 0.2 x 0.1 mm^3. Typical examples of crystal morphology are given in Figure 3.1.

In addition to a 1.8 Å native data set (called Nat-2) and a series of heavy atom derivative data sets to solve the phase problem, a series of co-crystallisation experiments were carried out in order to identify the active site and substrate binding sites for NAD^+, the Mn^{2+} ion and the sugar substrate. The details are listed in Table 3-1, in columns "Malmn1" (which contains Maltose, NAD^+ and Mn^{2+}), Ag12 (C25S, C174S double mutant) and Ag18 ($CoCl_2$ containing). Crystals of AglA turned out to be very robust, and addition of 25% glycerol to the mother liquor, plus increasing the PEG 6000 concentration to 20% (w/v), was sufficient to prevent ice formation. AglA crystals were usually incubated for 10-15 min in the cryo buffer and then cooled in liquid nitrogen or flash frozen in a 100K N_2 gas stream. The individual additives for the cryo buffer used on each specific crystal are outlined in Table 3-1.

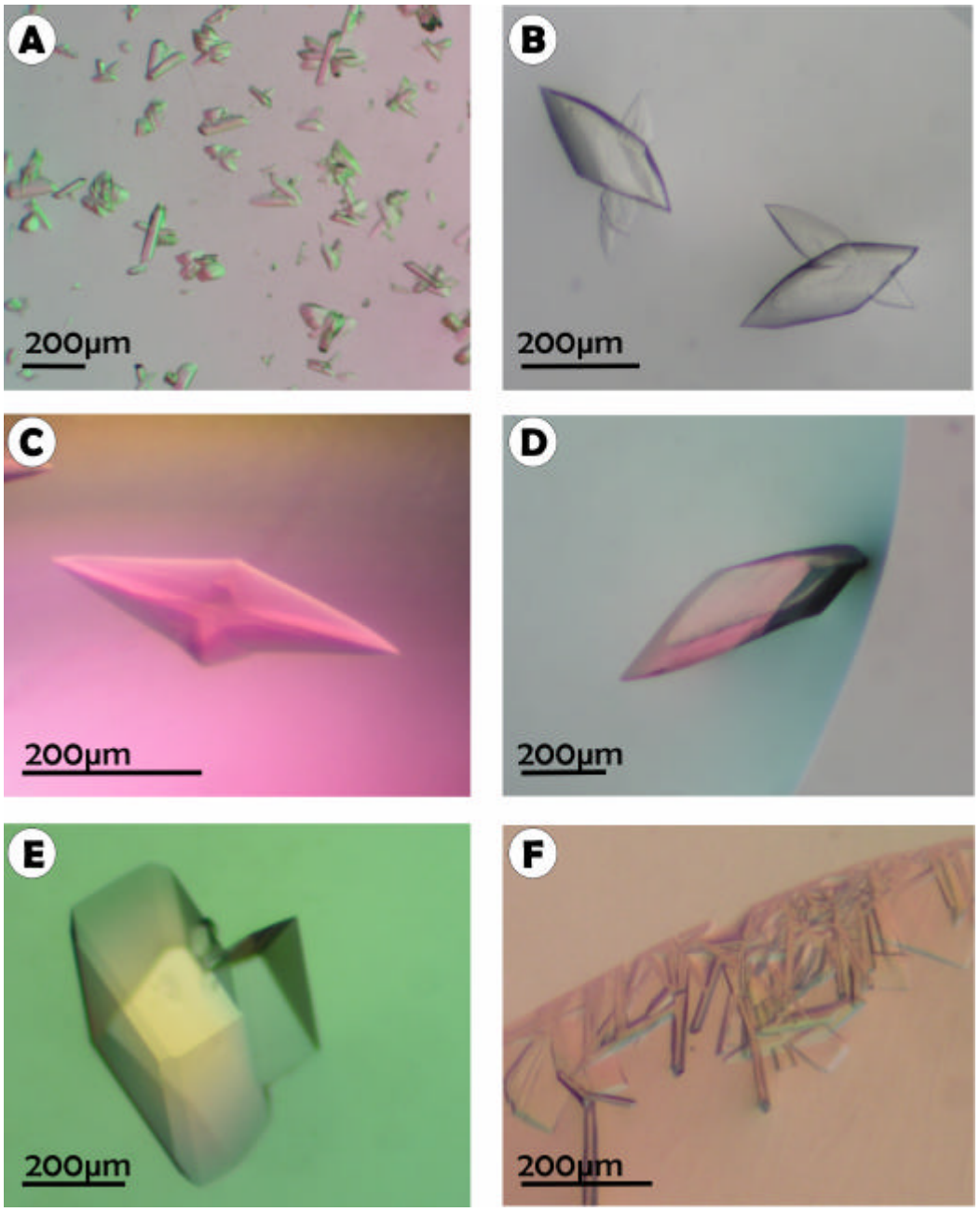


Figure 3.1 Examples of crystal morphology of AglA crystals

A. Examples of AglA crystals grown under non-optimised conditions. **B.** Typical AglA crystal morphology. The two smaller crystals could easily be separated from the larger prior to freezing. **C.** Morphology of AglA crystals with batch I after 6 months. Crystals still diffracted to 1.9 Å. **D.** and **E.** Alternative crystal morphologies which appeared occasionally with AglA batches II and III. They are isomorphous to AglA batch I crystals and diffracted to the same resolution. **F.** Crystals grown in the presence of high concentrations of DTT.

Table 3-1 Crystallisation and cryo conditions for each crystal measured

Crystal	Nat-2	MalMn1	Ag12	Ag18
Description	Apo form, protein only	Mn^{2+}, NAD^{+}, maltose	C25S, C174S mutant, Mn^{2+}, NAD^{+} maltose	$CoCl_2$, maltose
Protein Batch	AglA I	AglA II	AglA III	AglA I *
Protein Conc. (mg/ml)	12mg/ml	8mg/ml	16mg/ml	12mg/ml
Plate format	24- well	24- well	24- well	96- well
Reservoir				
Volume (µl)	500	500	500	100
Buffer (100mM)	NaAc pH 4.8	NaAc pH 4.4	NaAc pH 4.4	NaAc pH 4.6
Salt	1M LiCl	1M LiCl	1M LiCl	1M LiCl
PEG 6000 (w/v)	7%	8%	7%	9%
Additives		2mM $MnCl_2$ 2mM NAD^{+}	2mM $MnCl_2$ 2mM NAD^{+}	100mM $CoCl_2$
Drop size (µl)	2:2	2:2	2:2	0.5:0.5
Drop additives (0.4µl)		0.5M maltose	0.5M maltose	
Cryo buffer				
Buffer (0.1M)	NaAc pH 5.0	NaAc pH 5.0	Tris-Cl pH 7.5	NaAc pH 4.4
Salt	1M LiCl	1M LiCl	1M LiCl	1M LiCl
PEG 6000 (w/v)	20%	20%	20%	20%
Cryoprotectant	25% glycerol	25% glycerol	25% glycerol	25% glycerol
Additives		2mM $MnCl_2$ 2mM NAD^{+} 0.1M maltose	2mM $MnCl_2$ 0.1M maltose	100mM $CoCl_2$

* Protein was preincubated with 0.8M maltose prior to pipetting to drop

3.2 Data Collection, Processing and Refinement

3.2.1 Data Collection of Heavy Atom Derivatives

The crystals of AglA belong to the monoclinic space group $P2_1$ and diffract to between 1.7 Å and 2.1 Å on both home and synchrotron sources. They are usually isomorphous, with the best native data set exhibiting a unit cell of a= 74.6 Å, b= 85.0 Å, c= 83.6 Å (all other unit cells are given in Table 3-2 and Table 3-3). This results in a Matthews coefficient of V_m = 2.6 $Å^3$/ Da for 2 molecules per asymmetric unit, with a typical solvent content of 42%.

A series of heavy atom derivatives was prepared. The derivatised crystals were sufficiently isomorphous to allow phasing by the MIRAS method. X-ray data of four heavy atom derivatives were collected at the BAMline at BESSY II using a Mar345 imaging plate device as the detector. Data for further two derivatives and the native data set were collected on a MAR345 imaging plate detector mounted on an Enraf-Nonius FR571 rotating anode source and an Osmic MaxFlux mirror system.

Table 3-2 Heavy atom derivative data sets

Derivative	$HgCl_2$	$SmCl_3$	$Pb(Ac)_2$	PCMBS	PIP	Ta_6Br_{14}
Unit Cell	a=76.3Å, b=85.9 Å, c=83.3 Å, β=106.2°	a=75.3Å, b=85.6 Å, c=83.0 Å, β=105.7°	a=75.5 Å, b=85.9 Å, c=83.1 Å, β=106.2°	a=75.7 Å, b=85.7 Å, c=83.4 Å, β=106.0°	a=76.0 Å, b=86.0 Å, c=83.4 Å, β=106.1°	a=75.3 Å, b=85.0 Å, c=83.2 Å, β=105.5°
X-ray source	FR571	FR571	BAMline	BAMline	BAMline	BAMline
Resolution (Å)	25 - 2.45	15 - 1.85	15 - 3.0	15 - 2.4	15 - 2.0	10 - 2.4
Total Observations	71543	231821	212373	197454	299000	252530
Unique Observations	37253	77427	35354	40256	69629	39064
R_{sym} (final shell) (%)	6.8 (36.6)	6.1 (34.6)	6.1 (20.5)	5.6 (29.7)	7.4 (51.3)	5.0 (16.1)
I/σ (final shell)	10.0 (1.7)	20.0 (2.6)	25.7 (8.5)	26.3 (4.7)	16.2 (2.8)	33.9 (12.3)
Completeness (%)	97.6	89.6	100.0	99.4	99.4	100.0
Mosaicity (°)	0.27	0.70	0.84	0.67	0.58	0.50
Phasing to 1.85 Å						
Number of sites	3	2	4	2	5	2
Phasing power isomorphous	0.8	0.7	1.8	1.1	0.9	1.3
Phasing power anomalous	-	1.2	1.3	0.8	0.7	1.0

3.2.2 Data Collection from Cofactor- and Substrate-Treated Crystals

Several co-crystallisation experiments to define the sugar binding site in greater detail involved co-crystallisation with other sugar substrates and inhibitors. Experiments were performed to co-crystallise the known substrates galactose, raffinose and at concentrations of 50mM in conditions of 0.1M NaAc buffer pH 4.0 – 4.4, 1M LiCl and 6% – 9% PEG 6000. Data sets were collected, however the subsequent structure showed no difference density for the binding of a sugar molecule. Acarbose, a tetrasaccharide inhibitor, was both co-crystallised and soaked into AglA crystals, however it also showed no binding. Therefore, these structures are not discussed further in this work.

To mimic the conditions needed for activity, reducing agents were added to the crystallisation buffers. DTT produced very poor crystals (Figure 3.1, pg. 29, photo F) when co-crystallised, and crystals disintegrated quickly upon soaking. Only one crystal, Ag1, survived soaking for 5 minutes in cryo buffer 18, which contained both DTT and Tris-HCl buffer at pH 7.5.

The data sets presented in this work were collected on a variety of X-ray sources, as described in Table 3-3. All crystals diffracted well, to between 2.1 and 1.7 Å. The crystals all belonged to space group $P2_1$ and displayed isomorphous unit cells, varying by less than 1 Å in all.

3.2.3 Data Processing and Refinement

All data sets collected were processed with the HKL program suite, version 1.96, and refined using CNS, versions 1.0 and 1.1. The phase problem was solved by MIRAS and the initial structure of AglA was built using stat set Nat-2, as described in the Methods. All subsequent data sets from cofactor- and substrate-treated crystals were solved by rigid body refinement and the ligands localized in difference density maps. The refinement statistics of each data set are given in Table 3-4.

To demonstrate the quality of the model, the Ramachandran plots for chain A and B of the final apo-protein structure (Nat-2) are shown in Figure 3.2. All residues fall within allowed regions. Only two of the other structures show disallowed main chain conformations, and in both cases these are alternate side chain conformations.

Table 3-3 Data Collection Statistics

	Nat-2	Malmn1	Ag12	Ag18
X-ray source	DESY BW7A	FR507	DESY X13	BESSY Bl1
wavelength (Å)	0.80170	1.54179	0.80190	1.60514
Detector	MARCCD	MAR 345	MAR 345	MARCCD
Space Group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Unit cell parameters				
a (Å)	75.4	75.0	74.5	75.138
b (Å)	85.4	85.7	85.2	85.350
c (Å)	83.1	83.6	83.5	83.450
β (°)	106.1	106.0	105.8	105.746
Mosaicity (°)	0.56	0.51	0.41	0.72
Resolution range (Å)	1.80-20	1.90 - 25	1.70 – 20	2.1 - 20
Highest shell (Å)	1.80-1.86	1.90-1.97	1.70 – 1.75	2.10 – 2.14
Total reflections *	440448 (35516)	193023 (14722)	410367 (33844)	211041 (5365)
Unique reflections *	94611 (9428)	77496 (7375)	111680 (11108)	58174 (2304)
Multiplicity *	4.6 (3.8)	2.5 (2.4)	3.7 (3.0)	3.6 (2.3)
Completeness (%) *	100 (100)	96.6 (92.6)	99.9 (99.9)	97.6 (77.4)
R_{sym} *	0.073 (0.196)	0.038 (0.362)	0.051 (0.277)	0.061 (0.253)
I/σ *	20.4 (3.7)	18.9 (2.1)	21.7 (4.1)	17.7 (2.5)

* the data in parentheses refer to the highest resolution shell.

Table 3-4 Refinement statistics

	Nat-2	Malmn1	Ag12	Ag18
Resolution range (Å)	20 - 1.8	20 – 1.9	20 - 1.75	20 – 2.1
No. of Reflections				
work set	88106	75055	101665	53523
test set	2735	2347	3104	1651
R_{work} (%)	18.1	19.9	19.5	18.5
R_{free} (%)	22.6	25.6	21.7	23.0
No. of atoms:				
total atoms	8573	8523	8403	8378
protein atoms	7751	7721	7717	7721
substrate atoms	-	46	46	46
NAD^+ cofactor atoms	-	88	88	-
water molecules	822	646	552	591
metal ions	-	-	-	2
glycerol atoms	-	-	-	18
B-factor (Å^2)				
overall	34.6	32.5	30.04	42.8
protein*	33.4 / 34.1	31.2/32.0	29.5 / 29.5	41.2/43.0
substrate*	-	43.3/43.8	38.1 / 34.8	80.7/ 82.2
cofactor*	-	47.8/ 41.7	37.9 / 29.6	-
water molecules	42.9	40.5	36.5	45.1
metal ions	-	-	-	56.0/54.4
Wilson B-factor (Å^2)	29.7	28.1	25.8	32.6
rmsd bond lengths (Å)	0.014	0.012	0.005	0.005
rmsd bond angles (°)	1.59	1.53	1.22	1.16
Ramachandran				
most favoured (%)*	88.6 / 90	91.0 / 90.7	91.0 / 91.6	88.8 / 87.3
additional allowed (%)*	10.7 / 9.8	8.8 / 8.6	9.0 / 8.4	11.2 / 12.7
generously allowed (%)*	0.7 / 0.2	0.2 / 0.5	0.0 / 0.0	0.0 / 0.0
disallowed (%)*	0.0 / 0.0	0.0 / 0.2	0.0 / 0.0	0.0 / 0.0

*data given for each of monomer A / monomer B in the asymmetric unit

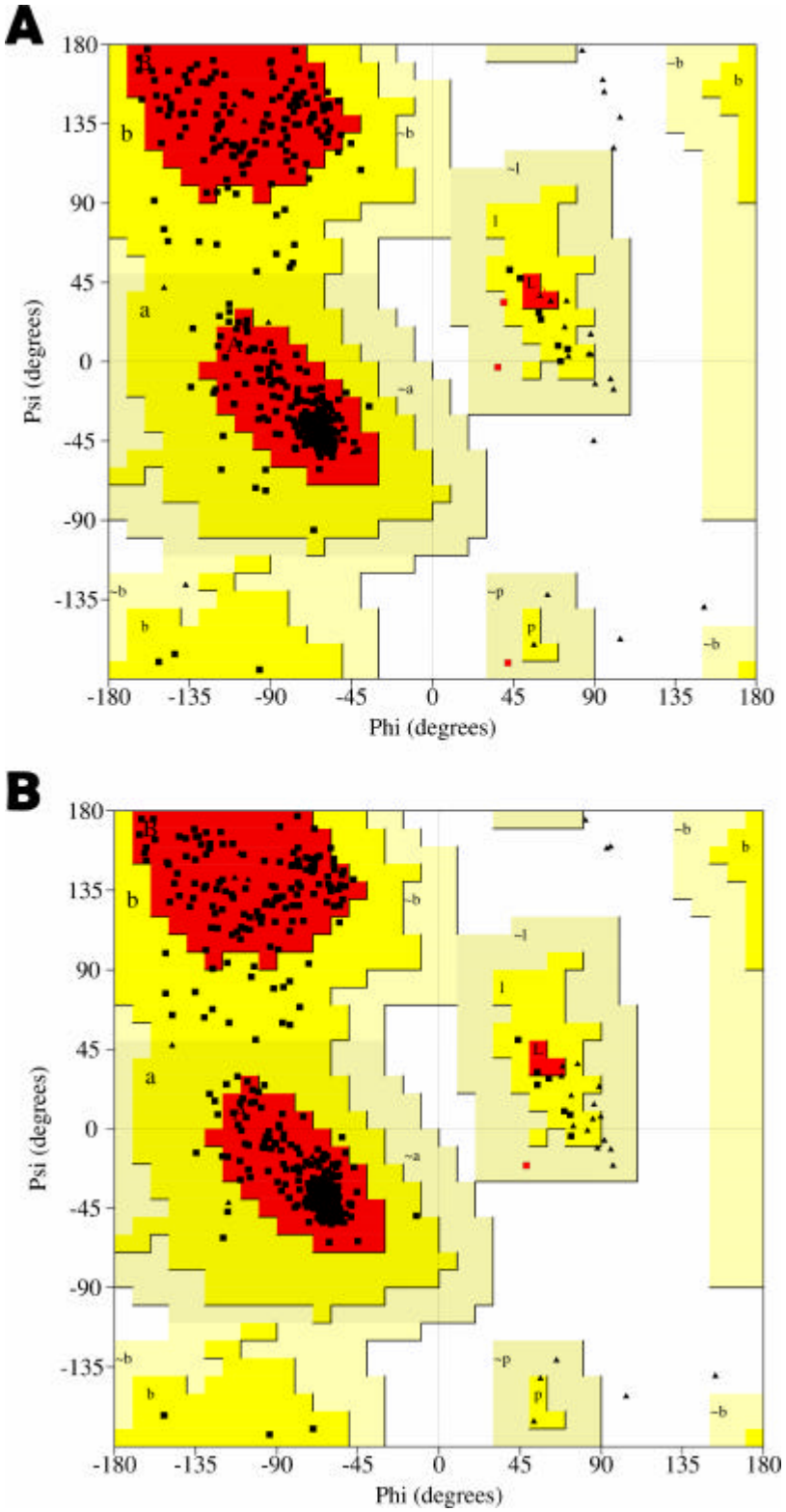


Figure 3.2 Ramachandran plots for the apo form of AglA (dataset Nat-2)

The plots are divided into energetically preferred regions (*red*), additionally allowed regions (*yellow*), generously allowed regions (*light yellow*) and disallowed regions (*white*). Glycines are represented by triangles. **A.** Plot for Nat-2 chain A **B.** Plot for Nat-2 chain B. Figure produced using PROCHECK (Laskowski et al., 1993)

3.3 Crystal Structure of AglA

The structure of the apo form of α-glucosidase A (AglA) from *Thermotoga maritima*, as solved by the MIRAS method, was initially refined to 1.8 Å resolution using diffraction data from crystal Nat-2 (Table 3-1). AglA crystallises with two molecules in the asymmetric unit, related by non-crystallographic symmetry. Due to the high resolution, the two independent molecules could be refined individually without NCS restraints. All of the 480 residues were built into the electron density, with the exception of residues 1-4 and 480. Several sections of the protein display high B-factor values (Figure 3.3). In particular two loops extending from residue 175-179 and 318-334 show notable disorder, with B-factors of 52.9 $Å^2$ and 65.2 $Å^2$, respectively, in monomer A, compared to a B value for all residues in monomer A of 30.7 $Å^2$.

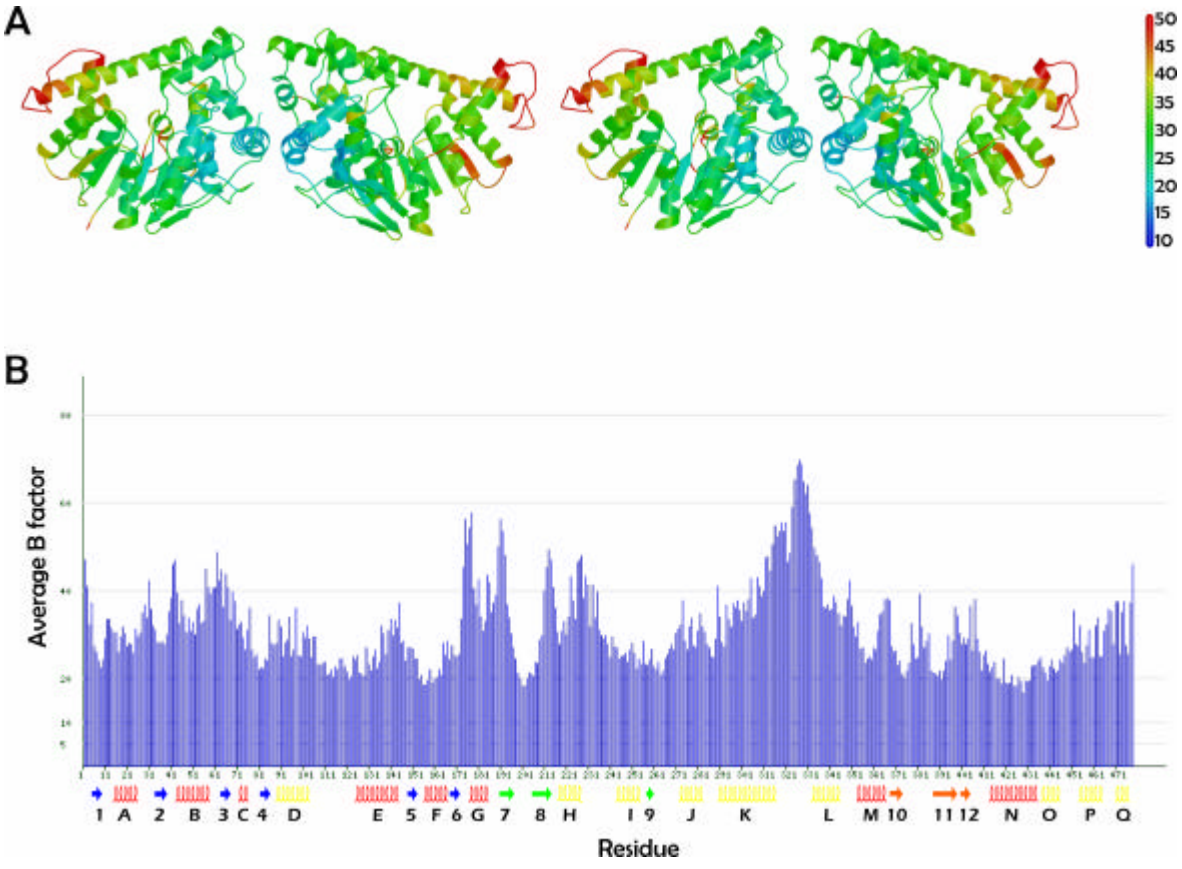


Figure 3.3 Stereo view of the asymmetric unit dimer, showing the B-factor distribution

A. The ribbon diagram is coloured by temperature factors, according to the bar shown at right. **B.** B-factor plot by residues, given for molecule A. Plot was created by the Uppsala Electron Density Server (Kleywegt et al., 2003).

Lowest B-factors are observed in the centre of the dimer, not in the centre of the monomers. The region most distant from the dimer centre shows the highest B-factors. This B-factor distribution may be partially caused by a rotational rigid body movement or static disorder of the whole dimer around the dimer centre as opposed to local flexibility.

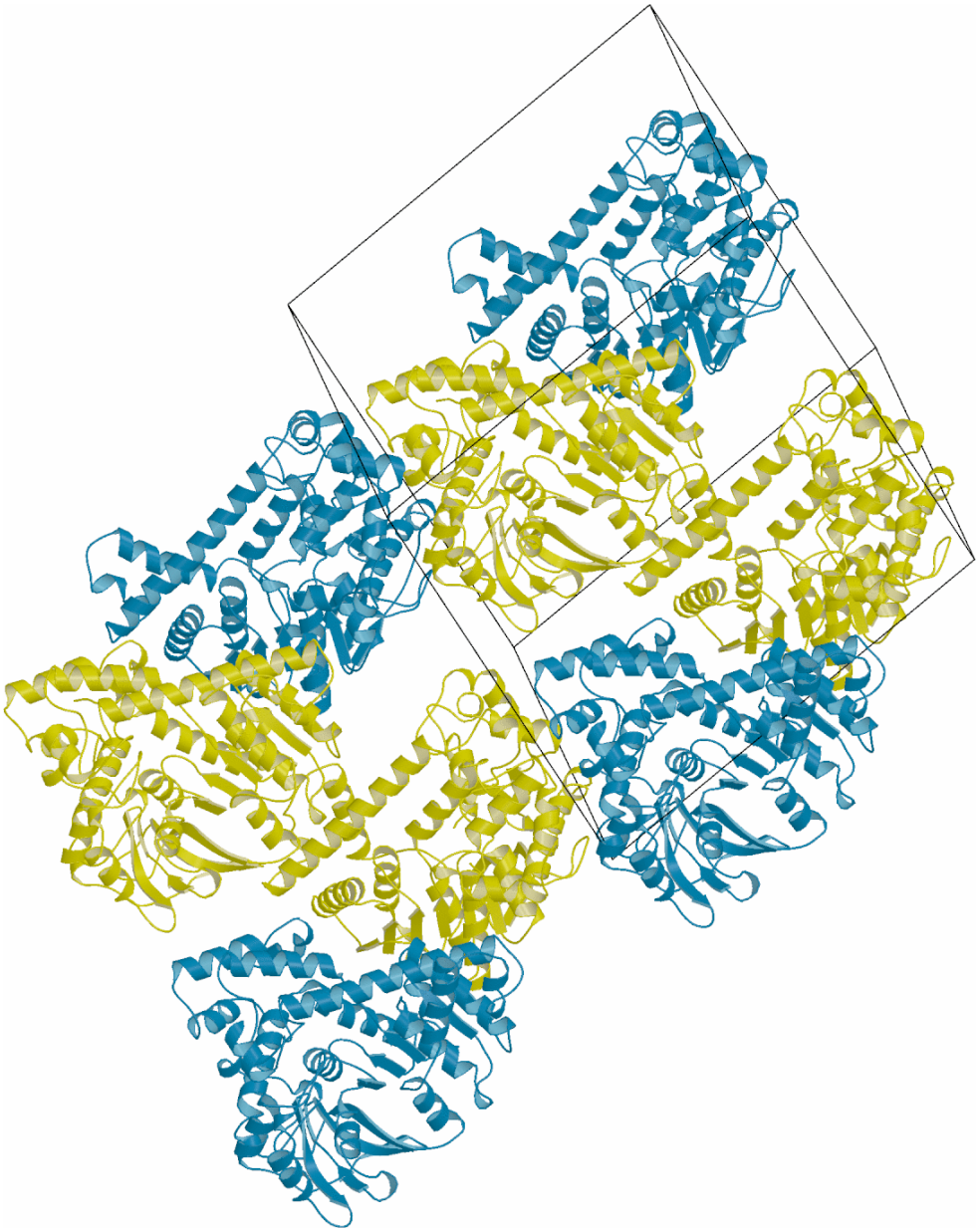


Figure 3.4 Crystal packing

The four molecules in the unit cell (two per asymmetric unit) are shown. Monomer A is coloured *yellow* while monomer B is *blue*. The unit cell is outlined in *black*.

3.3.1 Overall Protein Fold

The overall fold of AglA is shown in Figure 3.5. The monomer can be classified as a mixed α/β-fold type consisting of 17 α-helices (labeled A-Q) and 12 β-strands (labeled 1-12). The strands are organized into three β-sheets which are connected to each other by only two hydrogen bonds at each contact. Sheet I (blue in Figure 3.5) consists of β-strands 1to 6 and is a twisted parallel β-sheet, while the anti-parallel β-sheet II (β-strands 7 to 9, green in Figure 3.5) shows only a slight twist. Sheet III, also anti-parallel, consists of β-strands 10 to 12 (orange in Figure 3.5) and has a severe twist which causes the sheet to rotate by approximately 130° with reference to its starting point. Interestingly, this exceptional looking morphology is not found in any other glycosidase, but is in fact a well characterized fold observed in oxidoreductases. Sheets II and III, together with the surrounding α-helices form a typical oxidoreductase fold [CATH code 3.90.110.10 (Orengo et al., 1997)].

Sheet I is enclosed on both sides by α-helices A-G, which are antiparallel to the sheet, and these combine to form a classical NAD-binding Rossman fold consisting of six parallel β-strands interspersed by α-helices which fold to appear on each side of the six stranded β-sheet (Carugo and Argos, 1997). In AglA, this motif extends from residues 4 to 172 and includes the secondary structural elements 1-A-2-B-3-C-4-E-5-F-6. The first segment of this motif (β-α-β) has previously been predicted to extend from residues 4 to 38 in GH4 enzymes (Raasch et al., 2002; Thompson et al., 1998). There has been a modification to this typical topology in the form of helix D, which has been inserted into the fold between β-sheet 4 and helix E. Similar insertions have been observed in another group of NAD binding proteins, the D-glyceraldehyde phosphate dehydrogenases, where they are involved in substrate binding and catalysis. In AglA, this inserted helix does not partake in the NAD^+ binding domain, rather it extends above this fold, anchoring the long helix K and the C-terminal helices O, P and Q, via interactions with helix O.

Notably, helix K extends over a cleft in the structure which is located at the top of the Rossman fold and is the predicted binding site of the NAD^+ molecule. By extending over this cleft, helix K forms a tunnel through the middle of the protein which can be clearly seen in Figure 3.6. As mentioned in the Introduction, tunnel-shaped active sites have been observed in glucosidases of other families, although only in those with long substrates,

unlike AglA (Davies and Henrissat, 1995). The role of this tunnel will be further discussed in Section 3.6.

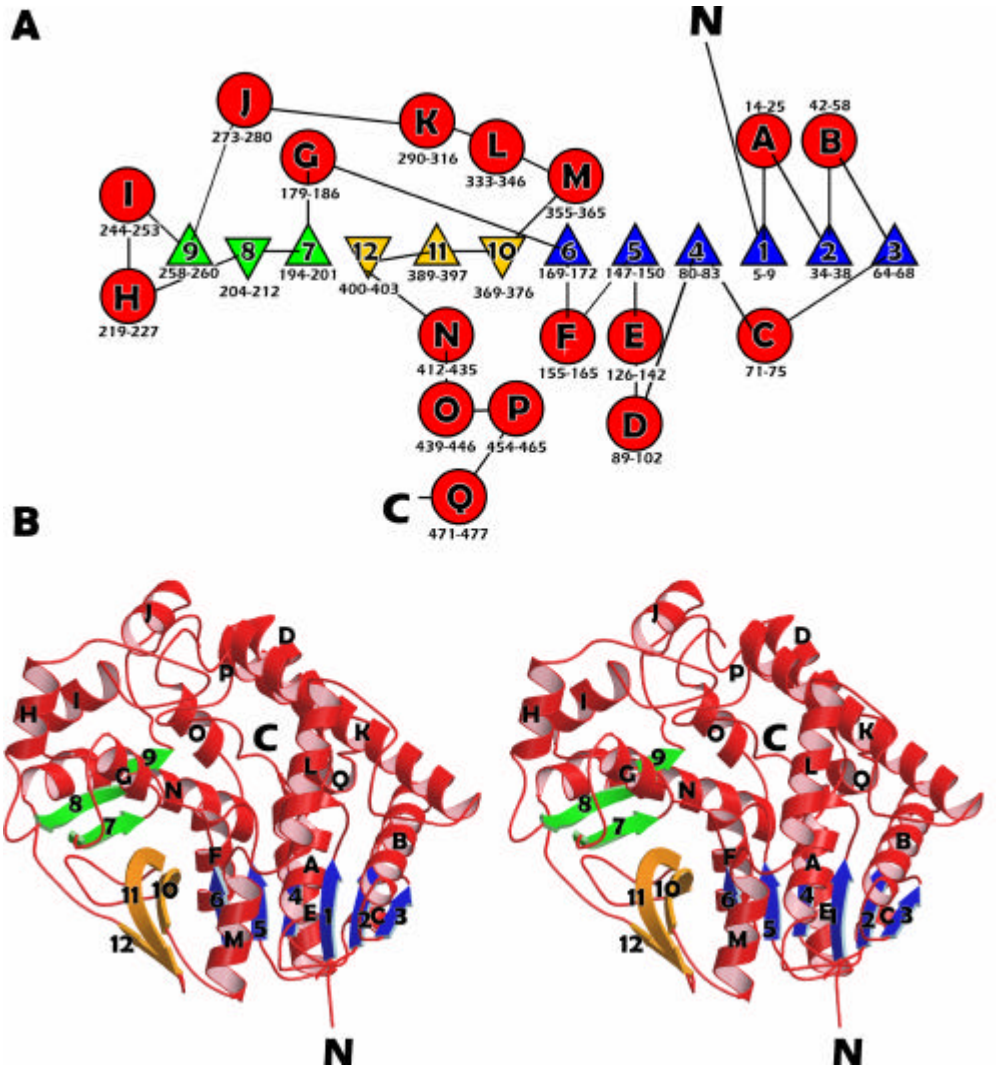


Figure 3.5 Topology and ribbon diagrams of the AglA monomer.

A. Topology diagram of AglA. Helices are represented by circles, β-strands by triangles. Colouring is as follows: Sheet I, *blue*; Sheet II, *green*; Sheet III, *orange*; helices, *red*. Secondary structure elements are labelled with the appropriate sequence numbering. **B.** Stereo view of monomer A, colouring and labelling as in A.

3.3.2 Differences Between the NCS-Related Molecules

All crystals of AglA used in this work belong to the same crystal form which contains two molecules in the asymmetric unit. These two independent protein molecules are related by a non-crystallographic (NCS) two-fold rotation axis. A least squares superposition of both molecules is depicted in Figure 3.6. The conformations of the two molecules are almost identical, but there is one notable exception. The superposition reveals a remarkable

difference in the orientation of a single helix, L between the two molecules. This helix L (residues 333 – 346) is rotated by 52° with respect to the same helix in the other molecule.

The overall root mean square deviation (RMSD), calculated for C_α atoms only, between the two monomers is 2.2 Å. This rather large value, however, can be dissected into an RMSD of 7.2 Å for residues 316-355, including helix L and its encompassing loops, and 0.7 Å for the rest of the structure. This indicates that the majority of the deviation between the two monomers is caused by the different orientation of helix L.

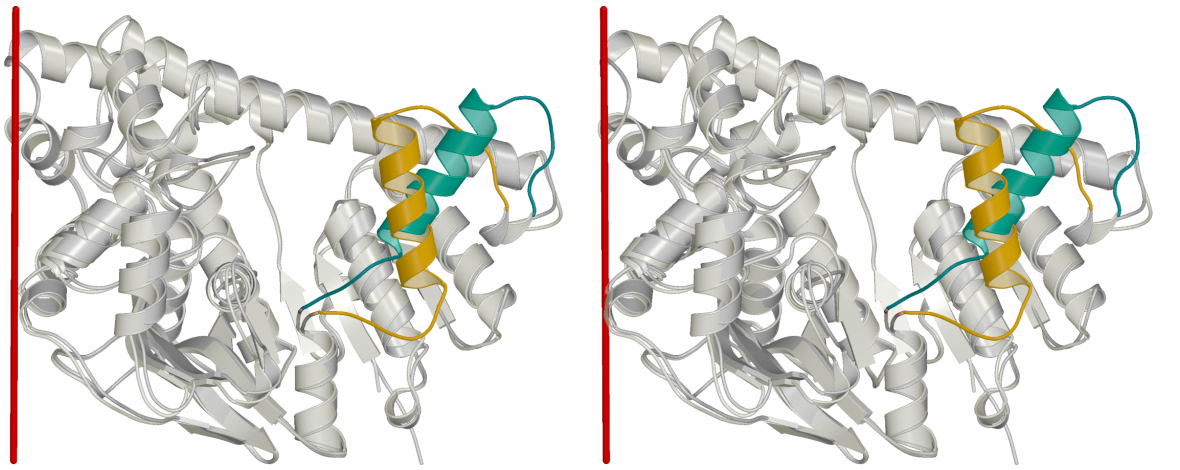


Figure 3.6 Least squares superposition demonstrating the asymmetry of helix L.

Monomer A is represented in *dark grey*, with helix L in *yellow*, while monomer B is a *light grey* and its helix L is *green*. The two-fold axis, around which monomer A was rotated to form the superposition, is indicated by the *red line*.

3.4 Dimer Organisation

Size exclusion chromatography has shown that AglA forms a homodimer in solution (Raasch et al., 2000). In the asymmetric unit there are indeed two molecules, however, it is not clear from the crystal packing if this represents the true biological dimer. A second possible dimer is present in the crystal, generated by a unit cell translation of the second monomer. Crystal packing and the two possible candidates for the dimeric form of AglA are illustrated in Figure 3.7.

The first dimer, dimer I, is formed by a rotation of 178.5° around the red non-crystallographic two-fold axis shown in Figure 3.7, and a minor translation of 0.23 Å along the axis. The interface is formed by 4 salt bridges and a number of hydrogen bond interactions, primarily between helices N and O, and by hydrophobic interactions between helices I and J (Figure 3.8). If dimer I would represent the physiological dimer in solution, the observed rotation of helix L (shown with red asterixes) between the two NCS-related molecules would not be part of the *in vivo* dimer interface and thus is formed as a result of crystal packing. The active sites, presumed at the position of the NAD^+ binding site, would then be located on opposing ends of the dimer and independent of each other.

The second possible dimer, dimer II, is formed by a unit cell translation of molecule B by 95.8 Å (Figure 3.7, marked with a dotted line). The interactions in this interface occur between the helices L and M of each subunit, thus involving the helix which displays the deviating conformations in both molecules. This interface could therefore suggest a physiological role in the dimer formation for the observed rotation of helix L between the NCS-related monomers. In this dimer, the two active sites are in close proximity, with the loop between helices L and M extending across the cleft of the opposing subunit, reducing the overall surface area of the cleft and the size of the opening to the active site. Although slightly smaller than the first interface and involving fewer hydrogen bonds and salt bridges, this second interface is comprised of a large number of hydrophobic interactions.

Analysis of the buried surface at the respective interfaces showed that for dimer I 1248 $Å^2$ of the monomeric surface was buried, compared to 1051 $Å^2$ in dimer II. These values are both in the expected range of a native interface (Janin and Chothia, 1990), and they are too similar to clearly identify the true dimer interface.

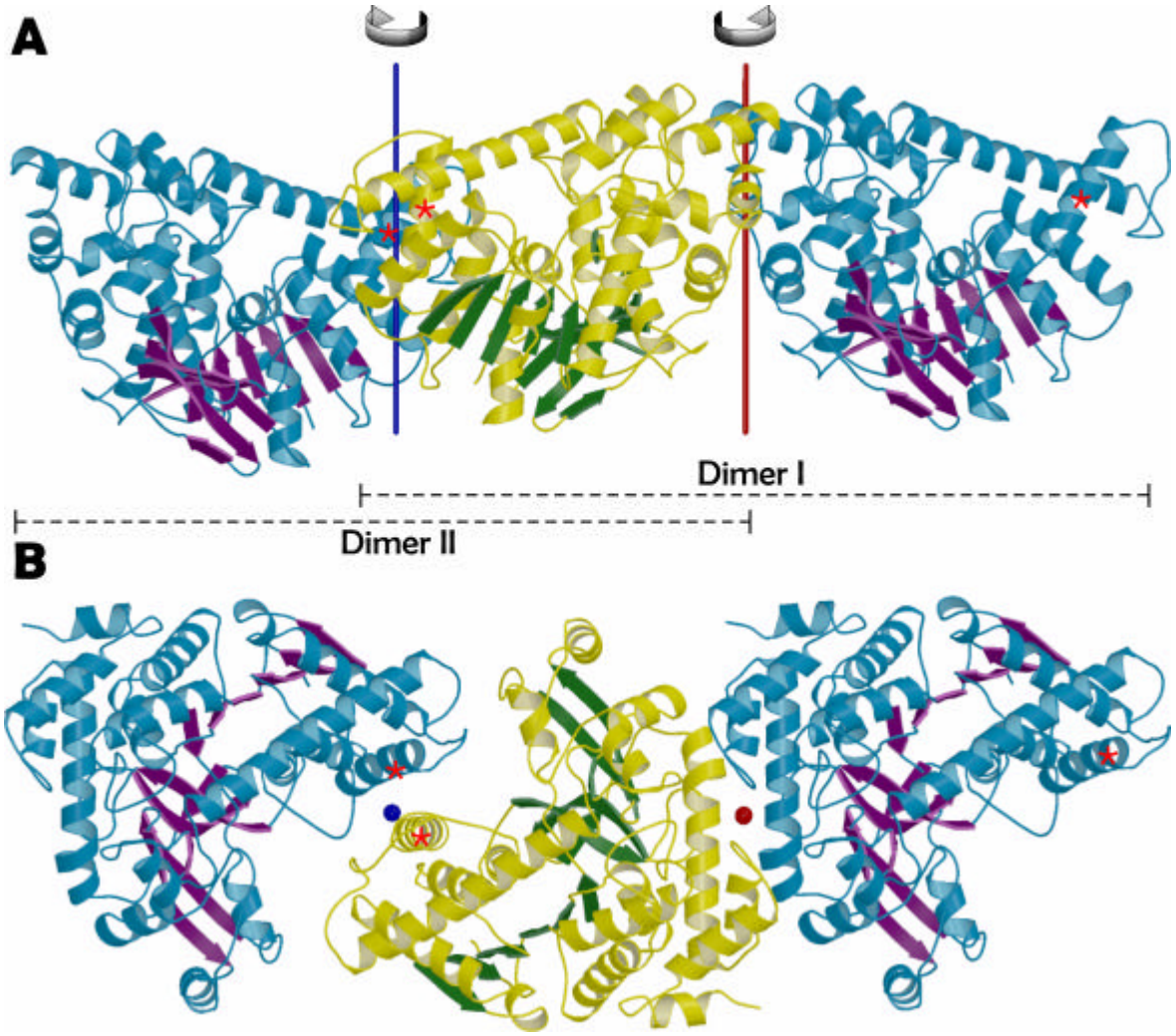


Figure 3.7 Crystal packing of AglA molecules

Packing of AglA molecules in the crystal. One monomer (monomer A) of the asymmetric unit is shown in *yellow* and *green*, the other (monomer B) in *blue* and *purple*. The two dimers are marked by the *dashed lines*. The B monomer in each dimer is related by a unit cell translation. The asymmetrically arranged helix L is indicated by the *red asterixes*. **A** Arrangement of AglA molecules in the crystal looking at the two-fold non crystallographic axis (represented by *blue* and *red lines*). **B** Crystal packing looking along the two-fold axes. Colouring as in A.

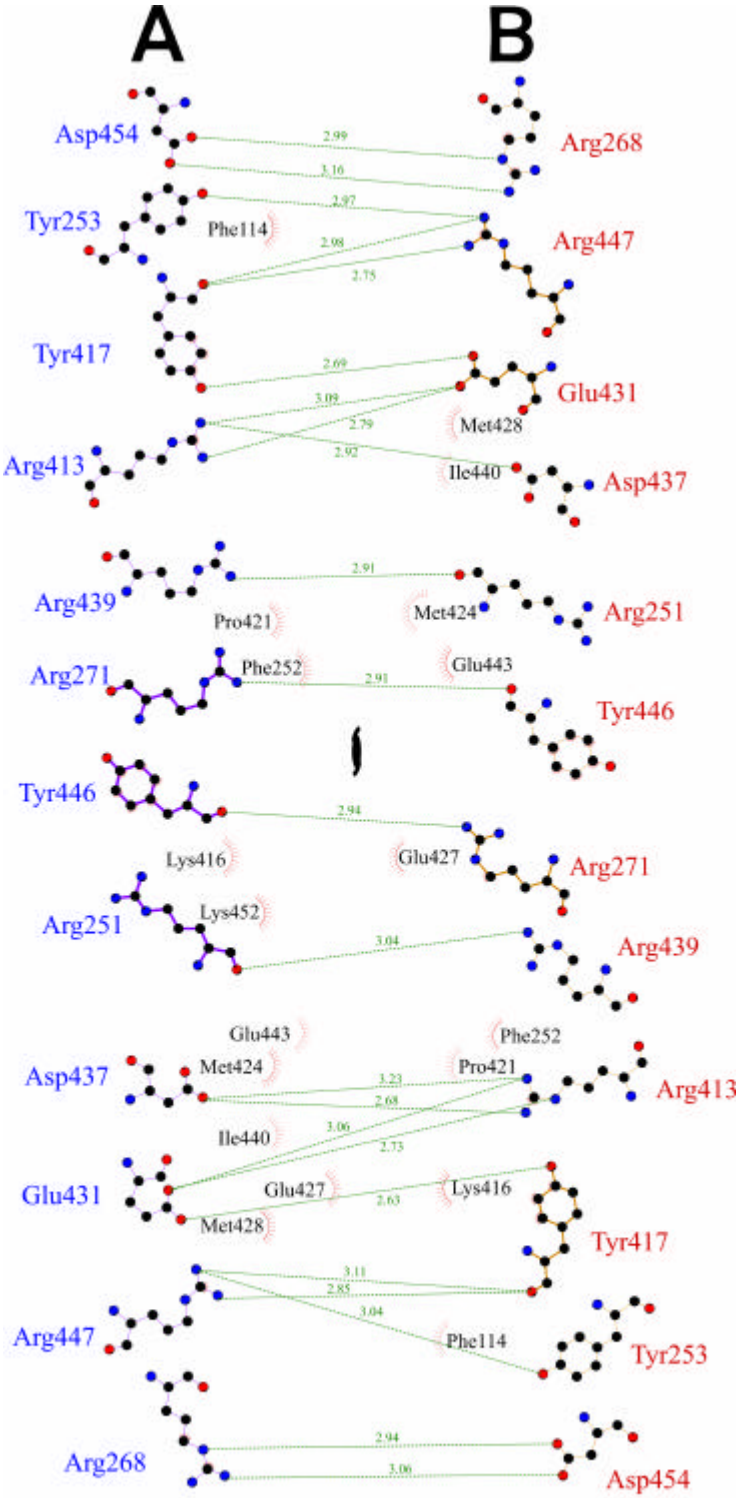


Figure 3.8 Dimer I interface

Schematic representation of the residues involved in the dimer I interface. Residues from monomer A are labelled in *blue*, those from monomer B in *red*. Hydrogen bonds are represented by *green lines* with distances as indicated. Hydrophobic interactions are represented by *red semi-circles*. The figure was prepared using Ligplot (Wallace et al., 1995).

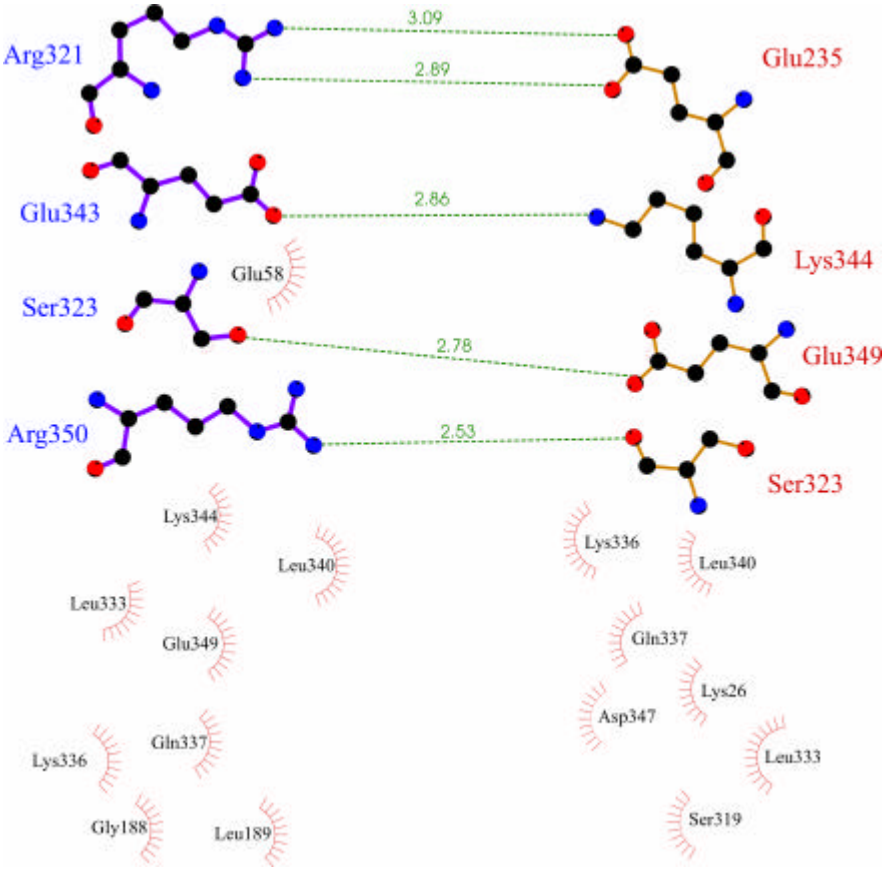


Figure 3.9 Dimer II interface

Schematic representation of the dimer II interface, which is proposed to be a crystal contact. Residues from monomer A are shown in *blue*, monomer B in *red*. Hydrogen bonds are indicated by *green lines* and the lengths of each indicated. Hydrophobic interactions are represented by *red semi-circles*. Figure was prepared using Ligplot.

To clarify which of the two possible dimers is present in solution, small-angle X-ray scattering (SAXS) was employed. SAXS data was collected on the Non Crystalline Systems (NCS) beamline at DESY, Hamburg. For each of the dimers, a scattering curve was predicted, to which the experimental results were fitted. The resulting fitted curves can be seen in Figure 3.10. Differences in the predicted scattering curves for the two possible dimers can be observed in the central region of the curve.

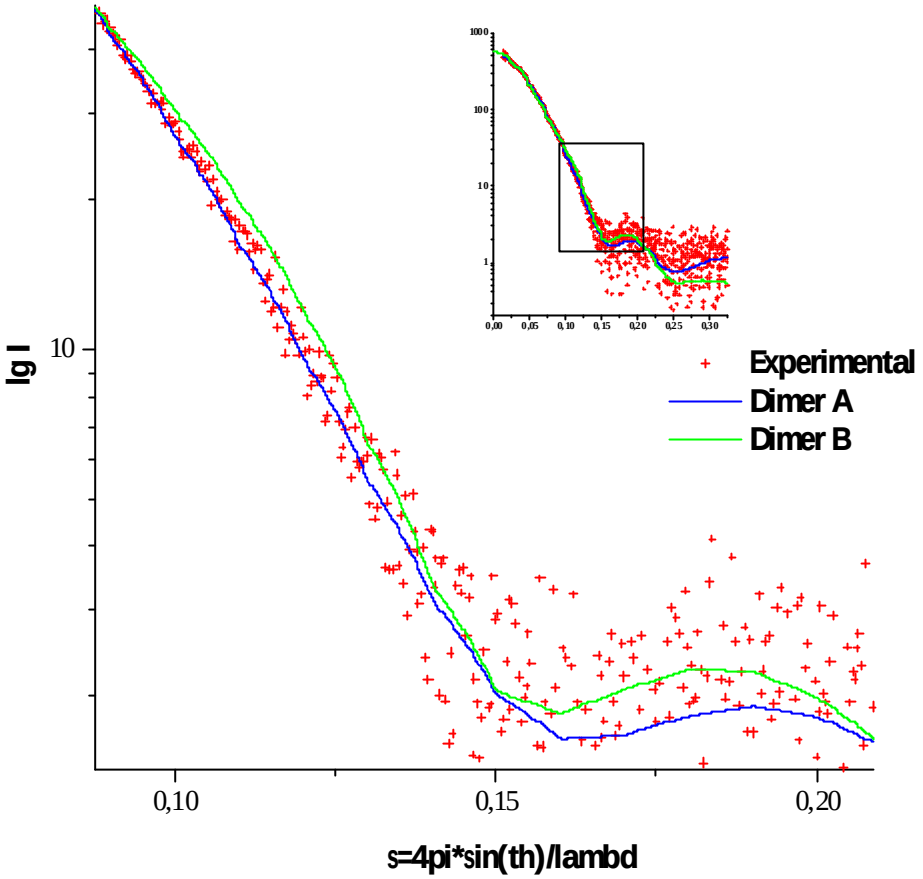


Figure 3.10 Scattering curve from small angle X-ray scattering (SAXS).

Inset graph displays the overall curve, while the boxed area is shown in the main figure. The curves calculated for dimers I (*blue*) and II (*green*) were fitted to the experimental data points (*red crosses*).

The fitted experimental data clearly support the presence of dimer I in solution. The χ^2 value, which reflects the agreement between the theoretical scattering of the predicted models and the experimental data, is χ^2 = 0.95 as opposed to χ^2 =1.47 for dimer II. A perfect fit results in a χ^2 value of 1.0. The presence of a 6 Å translation component for the screw operator relating the two monomers of dimer II, compared to 0.2 Å for dimer I, also favours the latter, as protein oligomer subunits are most often related by pure rotation, with little or no translational movement (Goodsell and Olson, 2000).

3.4.1 Importance of Dimerisation for Activity

Site directed mutagenesis experiments by Thompson and colleagues (1998) on the GH4 enzyme GlvA from *B. subtilis*, suggested that the residues Glu111 and Glu359 may be the potential general acid and general base in the active site. By sequence alignment (Figure 3.18), these are equivalent to AglA residues Glu113 and Glu391, respectively, neither of which is located at the catalytic site. In fact, Glu113 is located close to the dimer interface on a loop between helices D and E. This residue forms hydrogen bonds with only one other residue, Arg425, located on helix N of the interface in the same monomer (Figure 3.11). This indicates that the mutant, rather than directly interfering with binding at the interface, probably causes instability of helix N and thereby disrupts dimer formation which leads to inactivity. Other interactions stabilizing this helix are also shown in Figure 3.11. Although only one of several bonds present, the mutation of Glu113 to a glycine or an aspartate residue may create enough disruption to abolish enzyme activity, and suggests that dimerisation plays an important role in enzyme function. Further support for the hypothesis that oligomerisation is required comes from GlvA, which has been proven by analytical ultracentrifugation to be a homotetramer in solution, requiring Mn^{2+} to tetramerise from homodimers (Thompson et al., 1998). Although AglA is an active homodimer in solution regardless of the presence of a metal ion, this strengthens the proposal for a role of subunit assembly in the activation of GH4 enzymes.

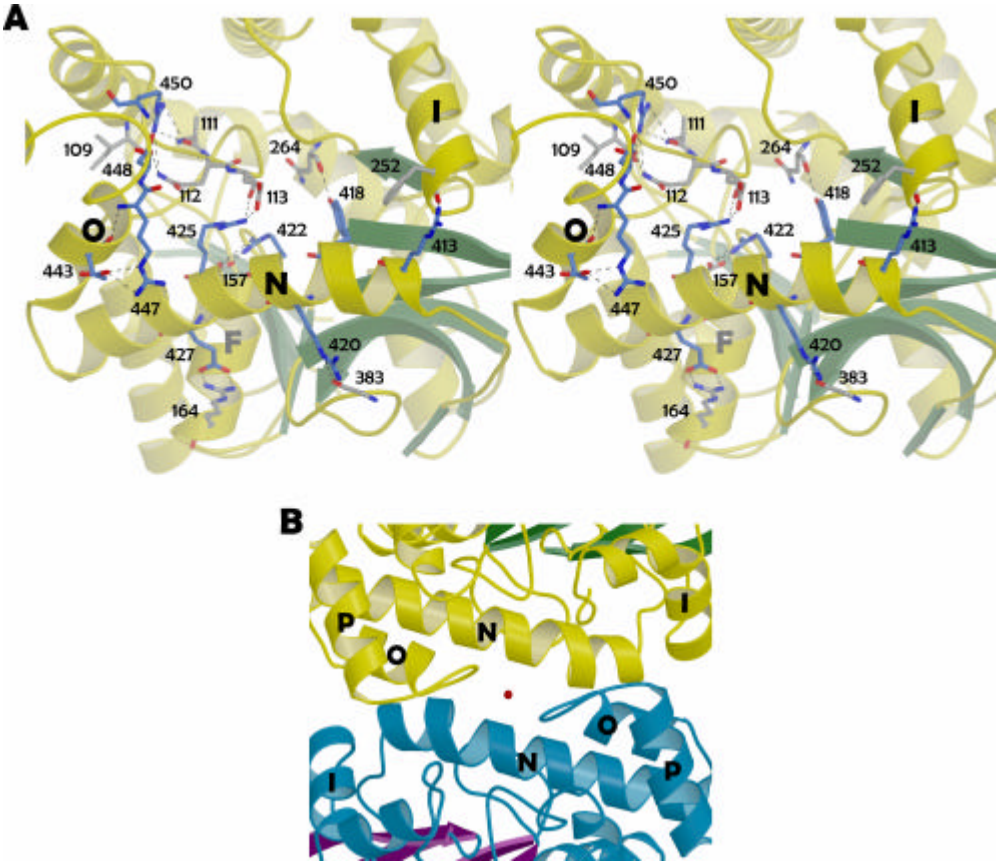


Figure 3.11 The interactions stabilising helix N of the dimer interface.

A. Stereo diagram of the view from the dimer interface on the interface-forming helix N of monomer A. The residues shown are those which stabilize the helix from monomer A. Residues from helix N and O are coloured *blue*, interaction partners in *grey*. Hydrogen bonds are shown by *dashed lines*. **B.** Top view of the dimer interface showing the locations of the helices in A. Molecule B is shown with *blue* helices and *purple* sheets

3.5 Similarity to Dehydrogenases

In order to identify structural relatives of AglA, a search in the Protein Data Bank (PDB) was performed using the DALI server (Holm and Sander, 1993). The closest relatives are listed in Table 3-5. It shows that none of AglA´s closest structural relatives are glucosidases, but instead all top ranking structural homologues belong to a subgroup of the NAD-dependent dehydrogenases, the 2-hydroxyacid dehydrogenases (Chapman et al., 1999). The lack of similarity to structurally known glucosidases reflects the sequence divergence from other families within the glycosyl hydrolases and can be observed by comparison with known glycosyl hydrolase clans (Figure 1.4, pg. 6). While the similarity between AglA and the dehydrogenases on a sequence level is only 17% or lower, the possible evolutionary relationship becomes obvious when comparing the 3D structures.

Table 3-5 Results of the comparison of AglA with other known structures using the DALI server

PDB	Description	Z	RMSD	LALI	%IDE
1CEQ-A	L-lactate dehydrogenase	21.6	3.2	274	17
1BMD	Malate dehydrogenase complexed with NADH	21.2	3.2	282	14
2CMD	Malate dehydrogenase	21.0	3.0	270	16
1HYH	1-2-Hydroxyisocaproate dehydrogenase	18.4	3.3	265	16
3LDH	Lactate dehydrogenase complexed with NADH	17.9	3.2	268	15
1DRW	Dihydrodipicolinate reductase	8.5	3.9	173	13
1GCA	Periplasmic galactose-binding protein for sugar uptake	8.0	2.8	112	9
2DRI	Periplasmic D-ribose binding protein for sugar uptake	7.9	2.6	102	9
1XEL	UDP-Galactose-4-epimerase	7.7	4.2	170	10
1FF9-A	Saccharopine reductase	7.7	5.7	153	8

Z = strength of structural similarity in standard deviations above expected
RMSD = positional root mean square deviation of superimposed C_α atoms in Angstroms
LALI = total number of equivalenced residues
%IDE = percentage of sequence identity over equivalenced positions

3.5 Similarity to Dehydrogenases

A superposition of AglA with lactate dehydrogenase (pdb code 3LDH) and malate dehydrogenase (pdb code 1BMD) is depicted in Figure 3.12. The core of the AglA protein, consisting of the NAD-binding oxidoreductase fold, is structurally homologous to the dehydrogenases. However, AglA contains extra secondary structure elements located on top of this dehydrogenase core, yet not organized into a distinct domain. The topology diagram (Figure 3.13) demonstrates that these segments (highlighted in yellow) are equivalent to a large insertion between residues 220 and 350 containing helices H to L and sheet 9, an inserted helix (helix D) between residues 89 and 104, and a C-terminal extension of around 50 amino acids encompassing helices O to Q. These regions will be referred to as non-DeHydrogenase, or non-DH regions.

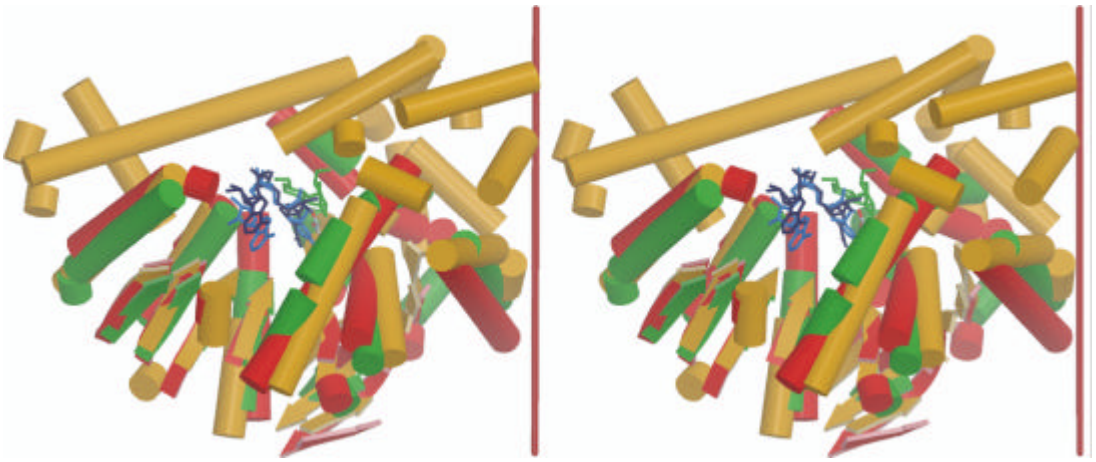


Figure 3.12 Least squares superposition of AglA with dehydrogenases

Stereo view of AglA (*orange*) superimposed with malate dehydrogenase, pdb code 1BMD (*green*), and lactate dehydrogenase, pdb code 3LDH (*red*). NAD^+ from AglA depicted in *dark blue*, from 1BMD in *light blue* and, from the 3LDH, pyruvate coloured by atom type. Maltose from AglA is shown in *green*. For clarity, loop regions are not shown.

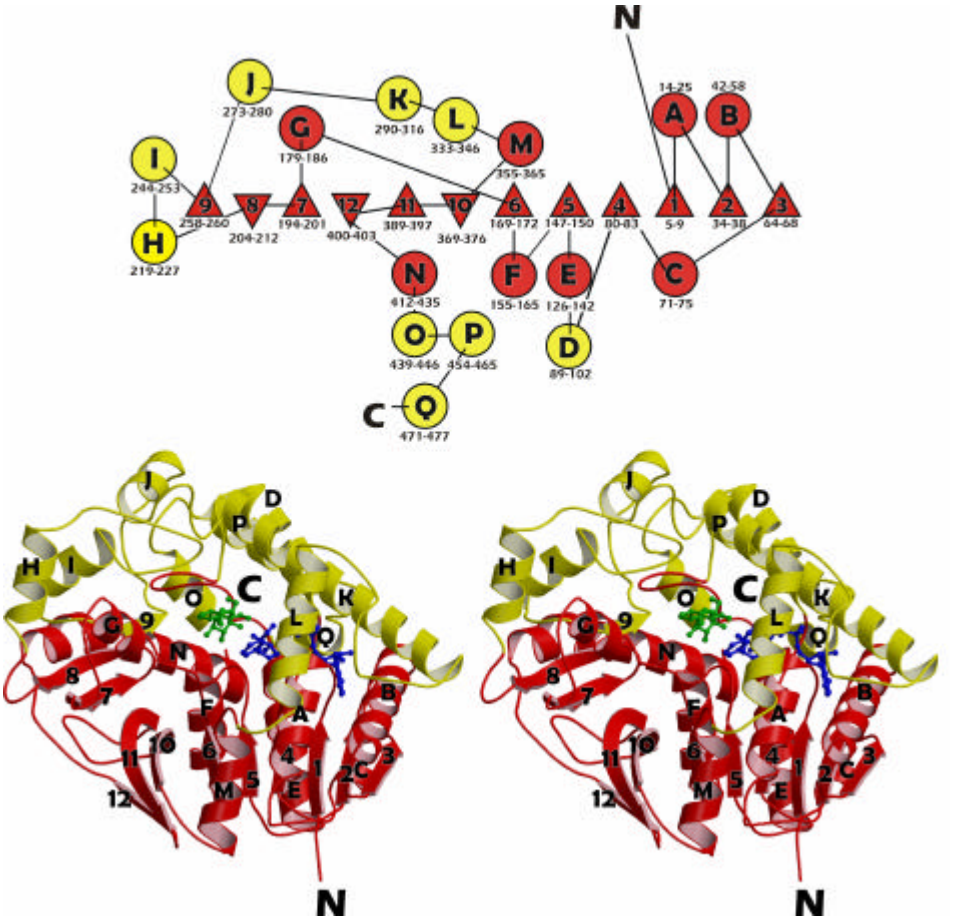


Figure 3.13 Homology to dehydrogenases.

A. Topology diagram, colour-coded according to homologogy with dehydrogenases (red), or non-homology (yellow). **B.** Secondary structure diagram shaded according to A. The NAD^+ and maltose molecules are shown in ball-and-stick representation in blue and green, respectively.

An interesting consequence of these added secondary structure elements is the covering of the very open active site present in dehydrogenases, by helix K. This has created a highly enclosed substrate binding site, as can be seen in Figure 3.12. The change to the exposed surface of the protein is shown by the space-fill models in Figure 3.14. The binding site of AglA greatly resembles that of malate dehydrogenase if the non-DH sections are removed [Figure 3.13, A(i), (ii) and B(i), (ii)], displaying the open cleft.

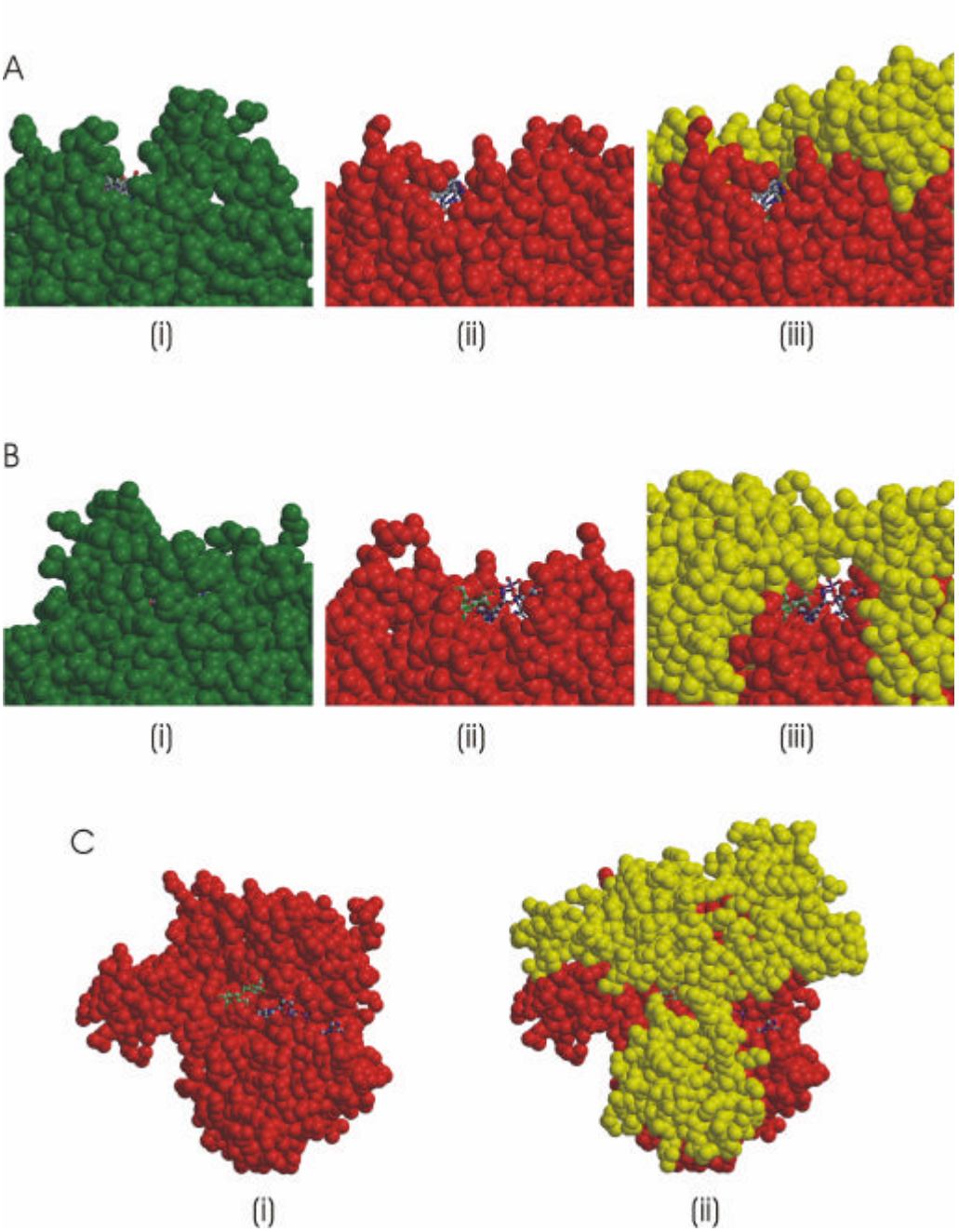


Figure 3.14 Space-filled models of AglA and malate dehydrogenase (MDH)

MDH is coloured *green*, the AglA homologous regions are coloured *red* and the non-DH regions *yellow*. Maltose (*green*) and NAD+ (*blue*) are shown in ball-and-stick representation. **A** (i) MDH looking at the cofactor binding end of the binding cleft (ii) AglA at the same view showing only the DH-homologous domains (iii) AglA with the non-DH sections included. **B** As for A, but showing the view from the other side of the protein at the substrate binding end. **C** View of AglA from above the protein looking at the active site, (i) only DH-homologous regions, (ii) entire protein.

The addition of these AglA-specific regions of the protein [yellow in Figure 3.14, A(iii), B(iii)], covering the binding site, may restrict the size of potential substrates. A view looking down at the top of the protein where the binding site is located [Figure 3.14, C(i) and (ii)], demonstrates how the binding cleft is covered, with much smaller access at both sides of the cleft, creating the tunnel-like structure which was discussed in Section 3.3.1, pg. 38.

3.5.1 The Dehydrogenase Active Site Arginines

Comparison of the binding of the substrate molecule by AglA to that of the dehydrogenases reveals an interesting relationship. The dehydrogenases contain three arginine residues located around the active site which are involved in substrate binding and which may exist in an open or closed state. Two of these arginines are found on a flexible loop which moves in to cover the binding site when the substrate is present, while the third arginine is found in the pocket of the active site. The open conformation is displayed in Figure 3.15 for porcine malate dehydrogenase (PDB code 5MDH) and the closed for the malate dehydrogenase from *Thermus flavus* (Chapman et al., 1999). The two arginines (Arg91 and Arg97) on the mobile loop are conserved in AglA as Arg97 and Arg107. These are located on helix D and the following loop, respectively, in positions which correspond to the open conformation of the dehydrogenase. This region is held "open" in the AglA structure by the insertion of helix D where there is only a loop in the dehydrogenases. The loop also interacts with the non-DH helix K over the NAD^+ binding site. However, another arginine, Arg263, and an aspartate, Asp260, occupy the same position as the arginines in the closed form - supplied from the non-homologous inserted loop between sheet 9 and helix G. The position occupied by the third arginine of the dehydrogenase binding pocket is, in AglA, the location of the active site disordered loop 6G (between strand 6 and helix G) consisting of residues 175 to 179.

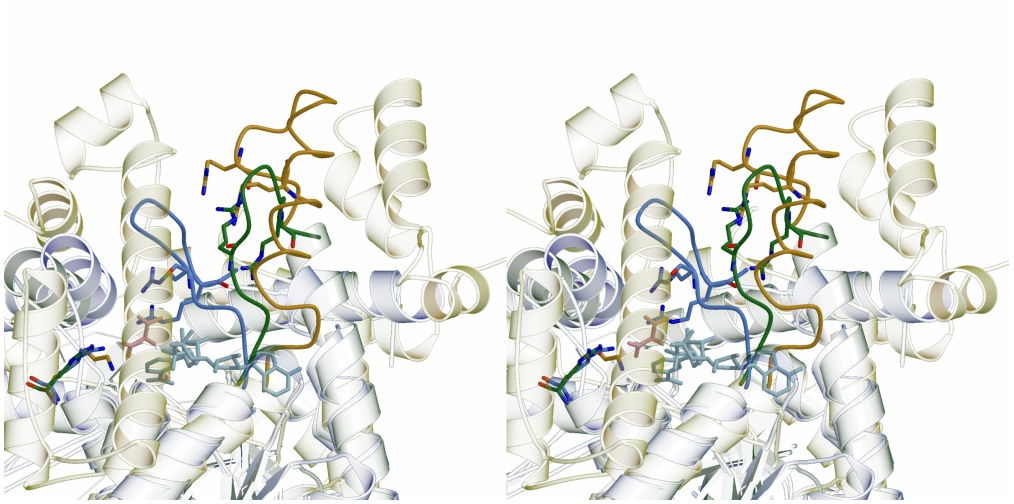


Figure 3.15 Stereo view of least squares superposition of the arginine loop

AglA (data set Malmn1) is coloured *yellow*, porcine malate dehydrogenase (pdb code 5MDH), *blue* and *Thermus flavus* malate dehydrogenase (1BMD), *green*. The loop containing two of the dehydrogenase active site arginines and the equivalent loop in the other structures are shown as traces of the C_α backbone, whereas the remainder of the structures are shown as a ribbon plot. The bound NAD^+ molecule of the dehydrogenase (*light blue*), and the pyruvate (*pink*) are shown in the active site.

3.6 Active Site

3.6.1 Overall Morphology of the Active Site

To identify the active site, AglA was co-crystallised with NAD^+ and maltose, as described in Table 3-1, pg. 30 (dataset Malmn1). The structure determined from this crystal clearly shows that the NAD^+ molecule does bind along the top of the previously identified Rossman-fold, whose β-sheets can be observed beneath the NAD^+ molecule in Figure 3.16 (Rossman, 1975) over which helix K extends, forming an enclosed tunnel through the protein. The presence of this tunnel in the apo-enzyme and its similarity to the tunnel structure of cellobiohydrolases (see Figure 1.5 pg. 8), could suggest the possibility of a sequential processing of substrates, as has been postulated for other tunnel-containing glucosidases (Davies and Henrissat, 1995). However, unlike these other enzymes, the co-crystal structure showed that the tunnel in AglA is completely filled by the NAD^+ cofactor, as can be seen in Figure 3.17.

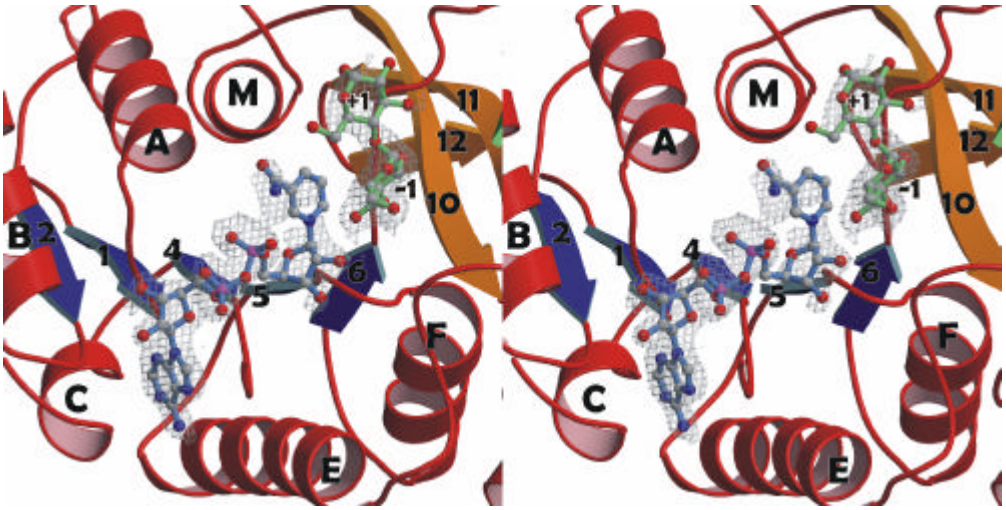


Figure 3.16 Stereo view of the active site with density for substrate and cofactor

The substrate maltose (*green*) and the cofactor NAD^+ (*blue*) are shown in ball-and-stick representation together with the $2F_o$-F_c electron density map surrounding the cofactor and substrate contoured at 1σ. For clarity, helix K, which covers the active site, is not shown.

This effectively blocks the tunnel and creates instead a deep cleft, lined at the bottom by the nicotinamide moiety of the NAD^+ molecule and surrounded by the loops between α-helices A to L and between helix I and β-sheet 9, and the helices G and M (Figure 3.16).

The maltose molecule could also be clearly identified in this structure, with the non-reducing end localised close to the nicotinamide moiety of the NAD^+ molecule in the cleft (Figure 3.16).

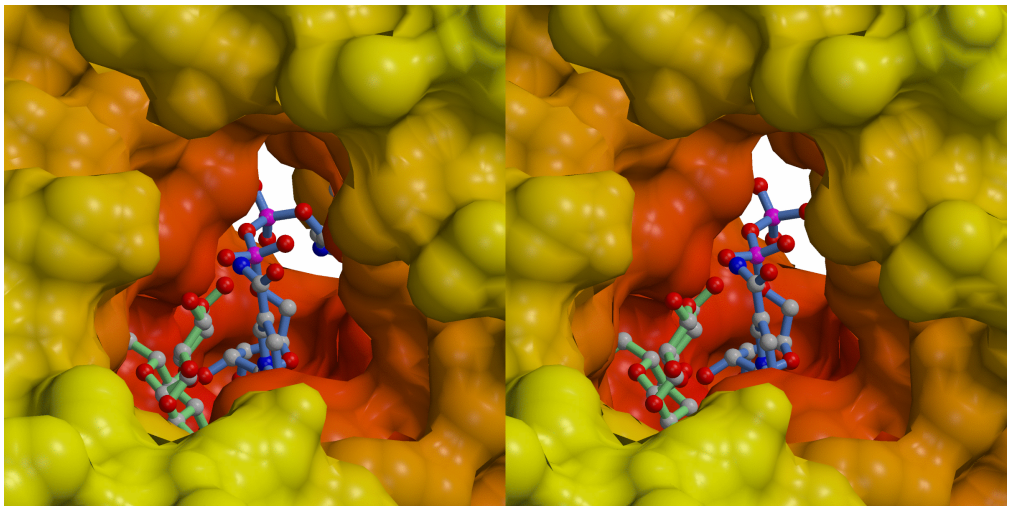


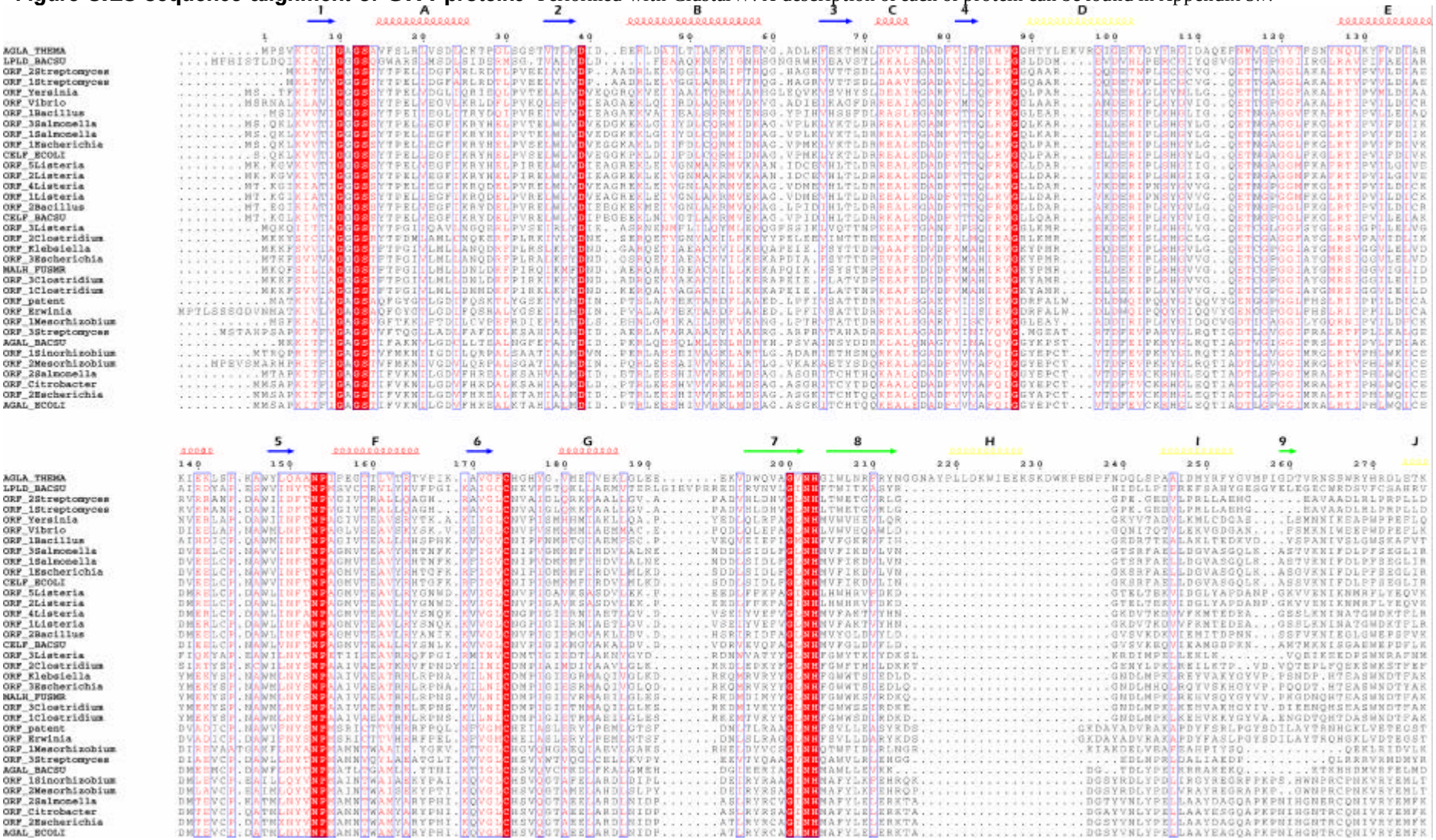
Figure 3.17 Stereo figure of the surface of the NAD^+ binding tunnel

The surface was calculated without substrate and cofactor and is coloured according to depth (*yellow* is closest to the viewer, *red* furthest away). Maltose (*green*) and the NAD^+ molecules (*blue*) are shown as ball-and-stick models.

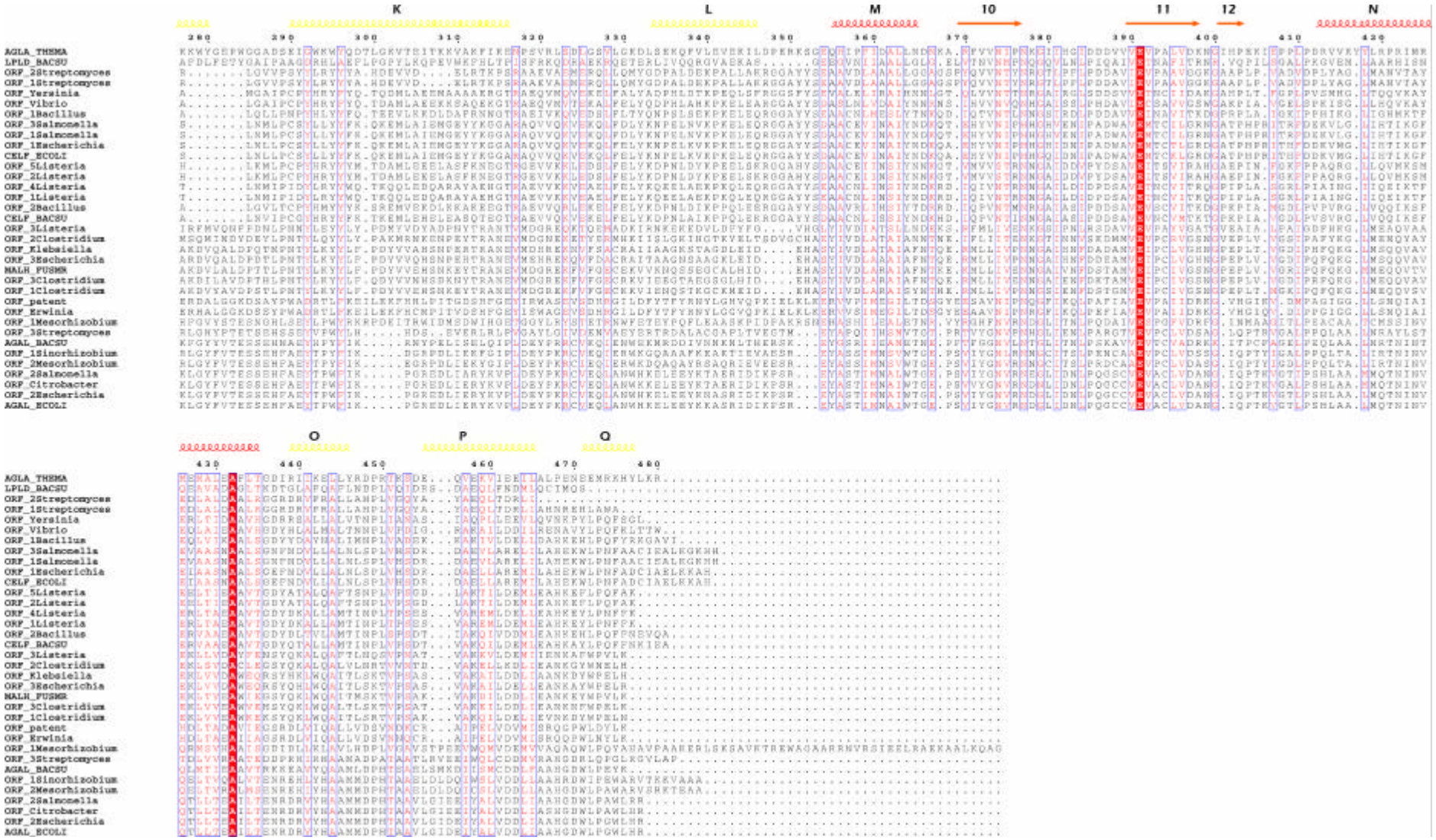
3.6.2 Binding of the NAD^+ Molecule

The NAD^+ molecule is bound to a well-described NAD binding domain termed a Rossman fold (Rossman, 1975). The presence of this binding domain in GH4 glucosidases had already been predicted for another GH4 protein, 6-phospho-α-glucosidase GlvA from *B. subtilis* (Thompson et al., 1998), and is the region of highest homology between all GH4 enzymes (see the alignment in Figure 3.18).

Figure 3.18 Sequence alignment of GH4 proteins Performed with ClustalW. A description of each of protein can be found in Appendix 5.7.



3.6 Active Site



The nicotinamide ring itself is the only part of the NAD^+ molecule which is not well defined in the electron density, suggesting that it is flexible in the active site pocket despite the presence of the substrate. The electron density is too badly defined to decide on which side of the nicotinamide ring the amide is located. For both orientations, refinement and inspection of the resulting $2F_o$-F_c and F_o-F_c maps did not resolve this ambiguity.

B-factor analysis of the NAD^+, the maltose substrate, and the surrounding residues (Figure 3.19) further underlines the flexibility of the NAD^+ nicotinamide moiety in the active site. To illustrate this, and to show the interactions of the nicotinamide ring with surrounding protein residues in both conformations, the orientation of the ring has been refined differently in each of the monomers. As can be seen in Figure 3.19, molecule A has considerably higher B-factors than molecule B. Interestingly, when the nicotinamide ring in monomer A was modelled as for molecule B, the B-factors remained high. This suggests that the flexibility of the active site is an intrinsic characteristic of this molecule in the crystal structure. Further description of the active site will concern the more defined molecule B, unless otherwise mentioned. Additionally, in agreement with PDB nomenclature, in this study all atoms belonging to the adenosine and adenosine phosphate will be prefixed with A, while all those belonging to the nicotinamide ring and adjacent sugar and phosphate will be prefixed with N.

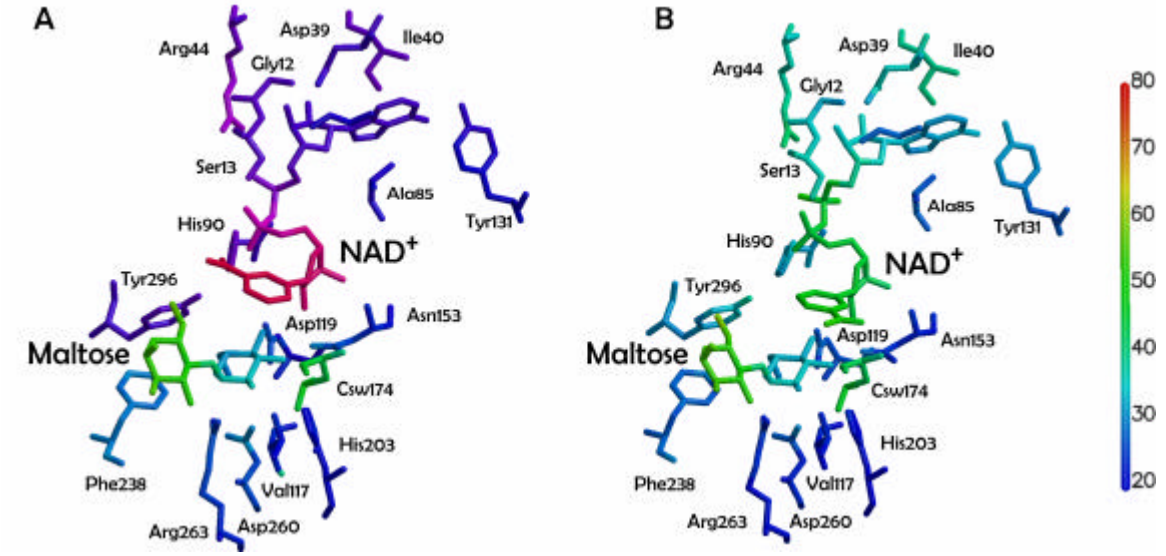


Figure 3.19 B-factor values of NAD+ and maltose in molecules A and B

NAD^+, maltose and surrounding residues have been coloured according to the B-factor of each atom (see scale bar, right). Residue Csw174 refers to the oxidized form of Cys174. **A.** Molecule A **B.** Molecule B.

The GXGS motif found at the N-terminal region of the protein is absolutely conserved throughout the GH4 family which can be seen in the sequence alignment in Figure 3.18. This motif in AglA (^{10}GAGS^{13}) has also been investigated by directed mutagenesis by Raasch and co-workers, who established that it is indeed required for NAD^+ binding by AglA (Raasch et al., 2002). A similar motif (e.g. GXGXXG) is found in all other dinucleotide-binding domains, and has been shown to be important for binding a structurally conserved water molecule which interacts with the dinucleotide pyrophosphate (Bottoms et al., 2002). In most structures analysed by Bottoms and co-workers, the water molecule is bound between the first or second conserved glycine residue, the last glycine residue, the C-terminal residue of β-sheet 4 and the dinucleotide pyrophosphate. Interestingly, this water molecule is absent in AglA.

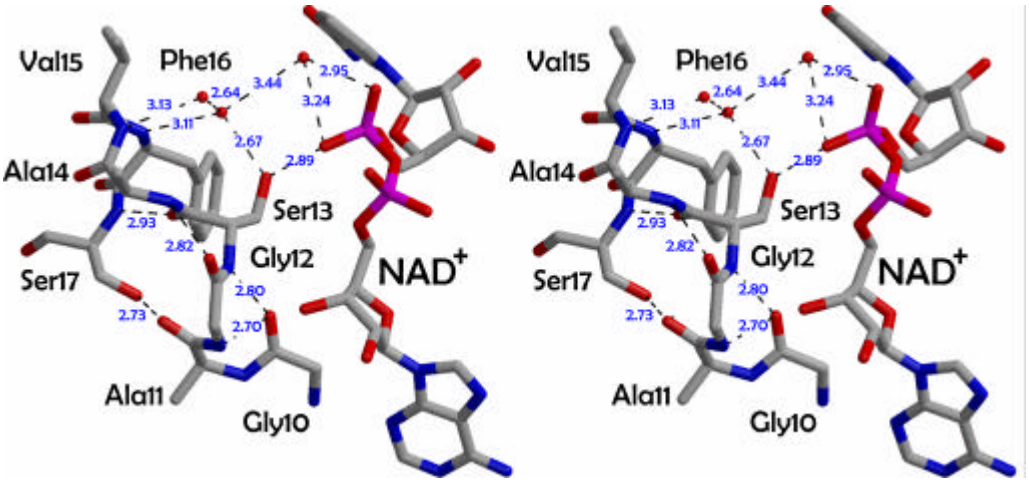


Figure 3.20 Stereo figure of the interactions of the conserved GXGS motif

The hydrogen bond network of the GXGS motif and additional residues in molecule B, showing the replacement of the structurally conserved wa ter by the side chain of Ser13.

The interactions of the GXGS motif in molecule B of AglA are shown in Figure 3.20. Gly10 and Gly12 are primarily involved in ensuring a tight turn at the end of β-sheet 1, however they do not bind a water molecule. It is Ser13 which is most interesting. Comparison with the malate dehydrogenase (pdb code 1BMD) which has the structural water (not shown), has indicated that the position of the water in AglA has been replaced by the side chain oxygen of Ser13. This binds to atom AO3', which also interacts with another water involved in a hydrogen bond network formed by the backbone nitrogens of Val15 and Phe16 and two additional waters. Ser13 is also stabilised by this network. The final coordinating residue of the structurally conserved water in other dinucleotide binding domains is Asp84, located at the end of β-sheet 4. This, however, is slightly further away in AglA, at a distance of 3.73 Å from the Ser13 side chain oxygen.

A number of other residues, not necessarily well conserved across GH4, are involved in NAD^+ binding. Stabilising the adenine ring via hydrogen bonds to atoms AN1 and AN6 is Tyr131 O^{η}, which is not found in this position in rest of the family. However, the adenosine sugar is well coordinated by the completely conserved Asp39 via hydrogen bonds between Asp39 $O^{\delta 1}$ and NAD AO2′, and between Asp39 $O^{\delta 2}$ and NAD AO3′. Arg44, whch coordinates one of the oxygens of the adenosine phosphate, is conserved in most of GH4 as an arginine or lysine. Also, the completely conserved Asn153 coordinates the nicotinamide sugar ring via both a backbone nitrogen and a backbone oxygen interaction with NAD NO3′.

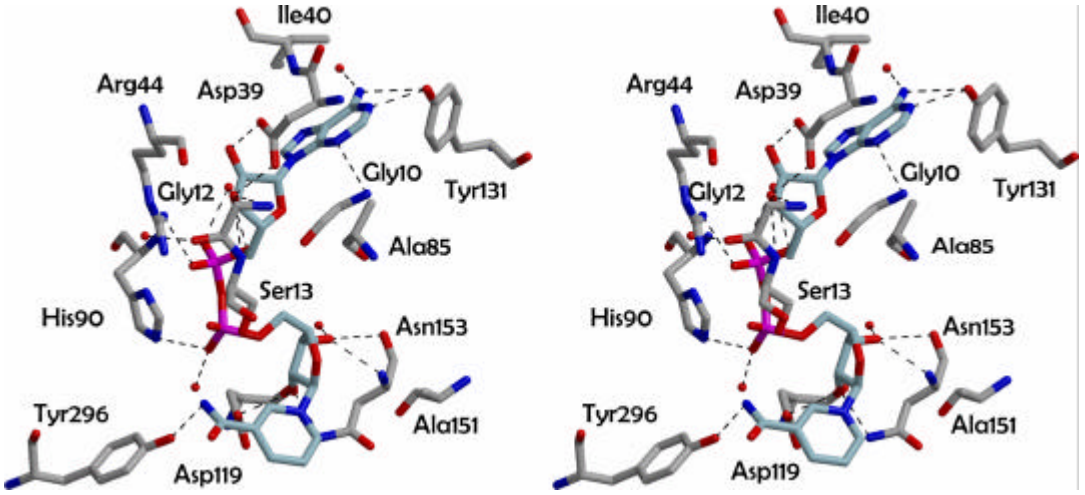


Figure 3.21 Stereo view of the NAD^+ binding site

Residues involved in the binding of NAD^+ (C atoms in *blue*) in monomer A are shown.

The position of the nicotinamide moiety of the NAD^+ molecule within the active site is in close proximity to the maltose molecule. This, together with the biochemical characterisation of its requirement for activity, suggests a participation of this group in the catalytic mechanism of AglA.

The exact role of the NAD^+ in the catalytic mechanism is yet to be determined. Activity assays on AglA by Raasch and colleagues, and on another GH4 enzyme, cellobiose-6-phosphate hydrolase (CelF) from *E. coli* by Thompson and co-workers, demonstrated that for these two enzymes, only NAD^+ can induce activity (Thompson et al., 1999) . $NADP^+$, NADH and NADPH were tested as well, and none of these demonstrated activation of the enzyme (Raasch et al., 2000).

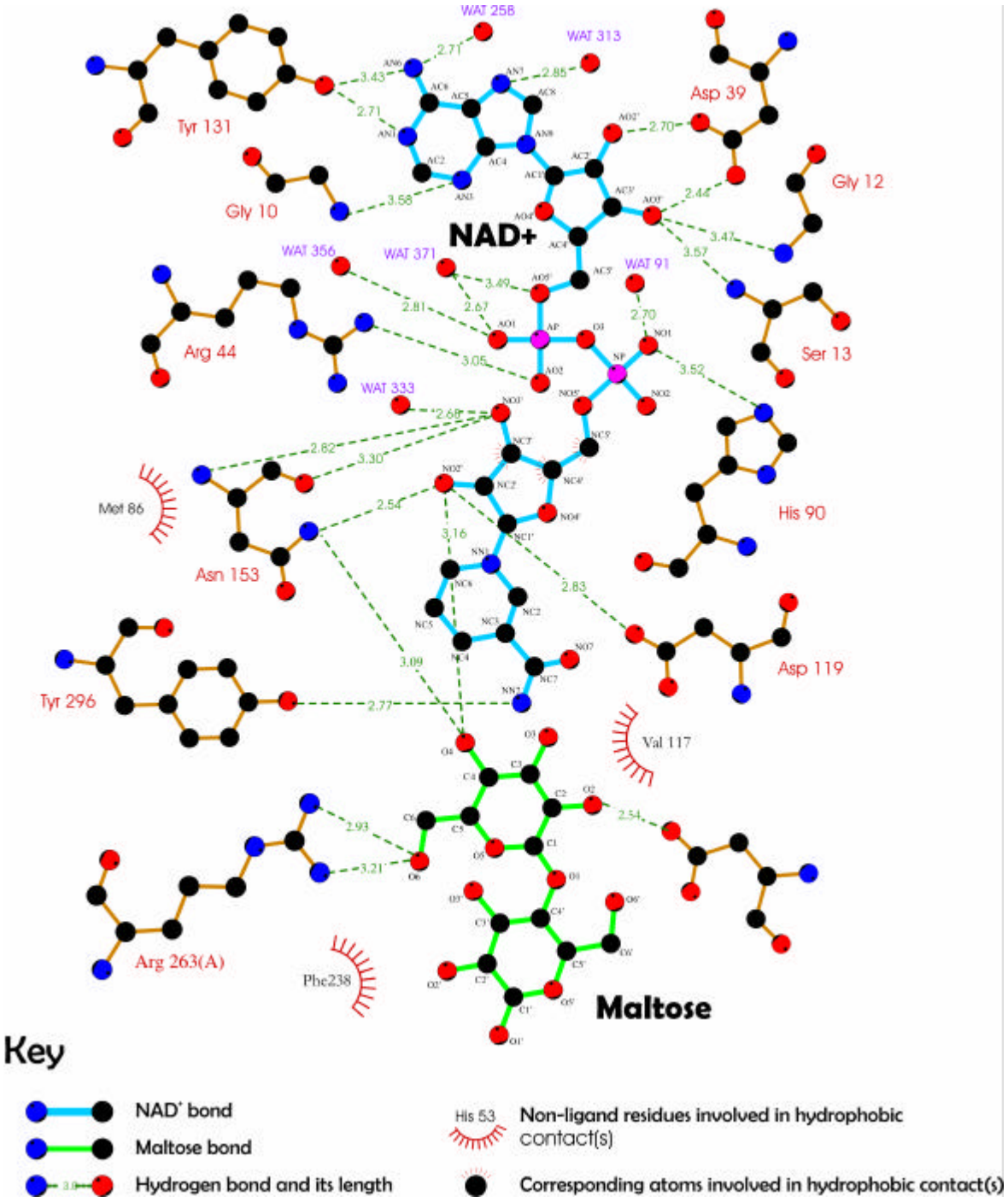


Figure 3.22 Schematic diagram of the active site showing NAD^+ binding residues

Shown is the monomer B binding site for crystal Malmn1. The atom numbering of NAD^+ and maltose corresponds to convention.

3.6.3 Binding Mode of the Maltose Molecule

The electron density maps of the cocrystal structure, malmn1, clearly indicated the presence of an intact maltose molecule (Figure 3.16). This demonstrates that the enzyme is inactive in the crystal, as maltose is the preferred substrate and should be hydrolysed to glucose. The non-reducing end of the maltose molecule, the -1 ring, is well defined in the electron density and bound by hydrogen bonds to Asn153, His203, Asp260, Asp119 and Arg263 as well as by hydrophobic interactions between Val117 (Figure 3.23) and the β-nicotinamide ring of the NAD^+ (Figure 3.22). In contrast, the electron density indicates that the reducing end (+1 ring) is more flexible, which corresponds with the fact that its only interactions with the protein is a single hydrophobic contact to Phe238.

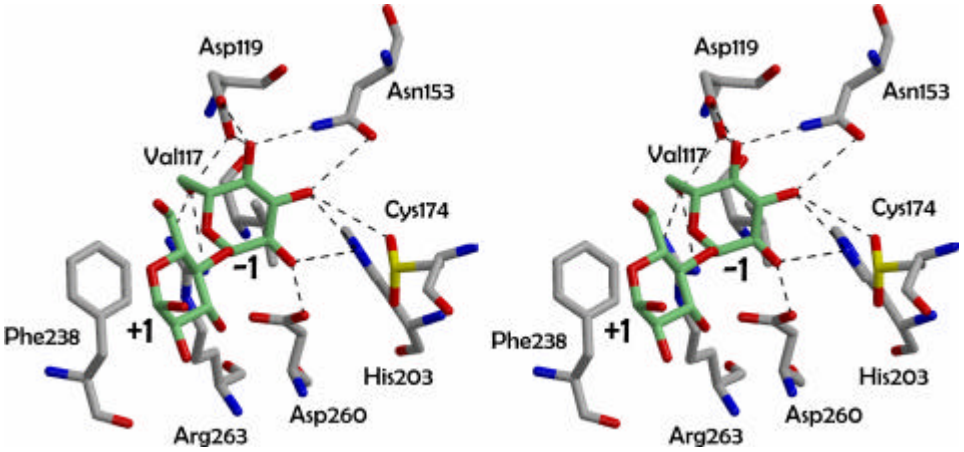


Figure 3.23 Maltose binding site in molecule A

Shown is the structure from crystal Malmn1, with Cys 174 modified to a sulfinic acid, see Section 3.6.5, pg. 64

This lack of tight binding to the second sugar ring correlates well with AglA´s wide substrate specificity and it´s exo-glucosidase activity which liberates one glucose residue at a time from the non-reducing end (Raasch et al., 2000). AglA cleaves 1-1, 1-3, 1-4, as well as 1-6 glycosidic linkages, and so there should not be tight binding to the sugar in the +1 site. The +1 ring should be able to "rotate" around its centre so that it may be connected via any of the O1, O3, O4 or O6 atoms to the first sugar. The hydrophobic stacking interaction of the sugar ring to Phe-238 supports this flexibility perfectly. Interestingly, such substrate recognition is not reflected by all GH4 glucosidases. GlvA has proven to be highly specific for glucoside substrates which are phosphorylated at the C6 atom of the glycone moiety,

while CelF hydrolyses only phospho-β-glucosides (Thompson et al., 1998; Thompson et al., 1999).

The maltose itself has quite a normal conformation in the active site. The dihedral angles around the glycosidic bond in AglA have been compared to previously studies onmaltose conformation. The torsion angles H1-C1-O1-C4′ and H4′-C4′-O1-C1 are -23.6° and -47.6°, respectively. These fall well into the region, determined by molecular mechanics for disaccharides, to represent the lowest energy for this bond (French and Dowd, 1993). Additionally, the dihedral angles O5-C1-O4′-C4′ and C1-O4′-C4′-C5′, with the respective angles of 71.0° and -144.7°, are similar to those determined for the glycosidic linkages in the oligosachharide inhibitor acarbose from multiple structures (Przylas et al., 2000a).

3.6.4 Potential Catalytic Residues of the Active Site

The active sites of a number of glycosidases have now been structurally characterised and the catalytic mechanisms well described (Davies and Henrissat, 1995; Zechel and Withers, 2000). The active site of AglA, however, bears no resemblance to these. The standard glucosidase active site contains an appropriately placed general base on one side of the glucose ring, with the corresponding general acid on the other side at the required distance to ensure inverting or retaining cleavage. The active site of AglA, however, contains the nicotinamide group of the NAD^+ molecule and an array of residues which do not correspond to those of other characterized glucosidases.

Site-directed mutagenesis on another GH4 enzyme, GlvA, from *B. subtilis*, suggested several acidic residues to be required for activity (Thompson et al., 1998). GlvA Asp41 can be seen by the sequence alignment in Figure 3.18 to be equivalent to AglA Asp39. This residue was predicted to be involved in NAD^+ binding, and the AglA structure confirms its position at the end of β-sheet 2, interacting with the O2′ and O3′ atoms of the adenine sugar ring. Mutation of this residue would thus disturb NAD^+ coordination, indirectly affecting the enzymatic activity. These experiments also predicted GlvA Glu111 as one of the catalytic residues, but, as discussed previously (Section 3.4.1, pg. 46), the homologous residue in AglA, Glu 113, is shown by this structure to be involved in stabilizing the dimer interface.

The second possible catalytic residue predicted for GlvA is Glu 359, whose equivalent by sequence alignment in AglA is Glu391. Glu391 is closer to the active site, situated on the

highly twisted β-strand 11 in sheet III. While not directly involved in the active site, it interacts with Cys174 via both a water molecule and Asn202, and through the latter, also interacts with His203. Cys174, Asn202 and His203 are completely conserved in GH4, and their roles will be discussed further. The strutcure suggests that an involvement of Glu391 in the positioning of these putative active site residues, rather than a role as a general acid or base, explains its requirement for activity.

A single putative general acid/base, Asp260, is located 3.4 Å from the C1 atom of the –1 ring of maltose (Figure 3.23). This residue is found at the end of strand 9 in β-sheet II, and it is this secondary structure which points the side chain into the active site. However Asp260 is not conserved in the family GH4 and it also lacks the required partner residue on the other side of the maltose molecule to partake in a typical glycoside hydrolysis reaction. The only totally conserved residues interacting directly with the substrate are His203 (interacting with O2 and O3 of the sugar in the -1 site) and Asn153 (interacting with O3 and O4 of the sugar in the -1 site), located at the bottom of the binding pocket (Figure 3.23), which are capable of interacting with a broad range of sugar substrates.

3.6.5 The Cysteine-Sulfinic Acid Cys174

Another completely conserved residue located in the active site is Cys174. Difference density around the sulphur of this cysteine displayed a “rabbit-ear” morphology which suggested it to be oxidized to a cysteine-sulfinic acid (Cys-SO_2H) (Figure 3.24 A). Upon inclusion of the two oxygen atoms, all unexplained difference density around the sulphur of the cysteine was gone (Figure 3.24 B) and the resulting B-factors for these oxygen atoms were the same as those of the rest of the residue. This finding might explain the observed requirement for reducing conditions for activity characterised for the protein *in vitro*, suggesting that this cysteine may be a catalytically important residue whose oxidation causes enzyme inactivation during purification, or even during heterologous expression in *E.coli*. The only other cysteine, Cys25, is found 24 Å away from the active site and does not show oxidation.

Cysteine-sulfinic acids (Cys-SOHs) have already been described in several protein structures, primarily as further oxidation states of the catalytically active but unstable cysteine-sulfenic acids (Hogg, 1990). Several different roles have been proscribed for these Cys-SOHs; in NADH peroxidases, NADH oxidases and peroxiredoxins it functions as

either a catalytically essential redox centre or a transient intermediate during peroxide reduction, and is thought to be stabilised by a nearby histidine residue and a potential Mg^{2+} ion (Choi et al., 1998). In nitrile hydratases, a Cys-SH, Cys-SOH and Cys-SO_2H form a "claw"-like structure which functions to coordinate a Fe(III) ion but plays no catalytic role (Nagashima et al., 1998). Two recent studies reported that the human protein phosphatase PTP-1B undergoes an unusual side chain modification where a cysteine, via formation of a sulfenic acid, forms a covalent bond with the backbone nitrogen of the following residue. This reversible reaction appears to be the way by which this enzyme is regulated via redox potential (Salmeen et al., 2003; Van Montfort et al., 2003).

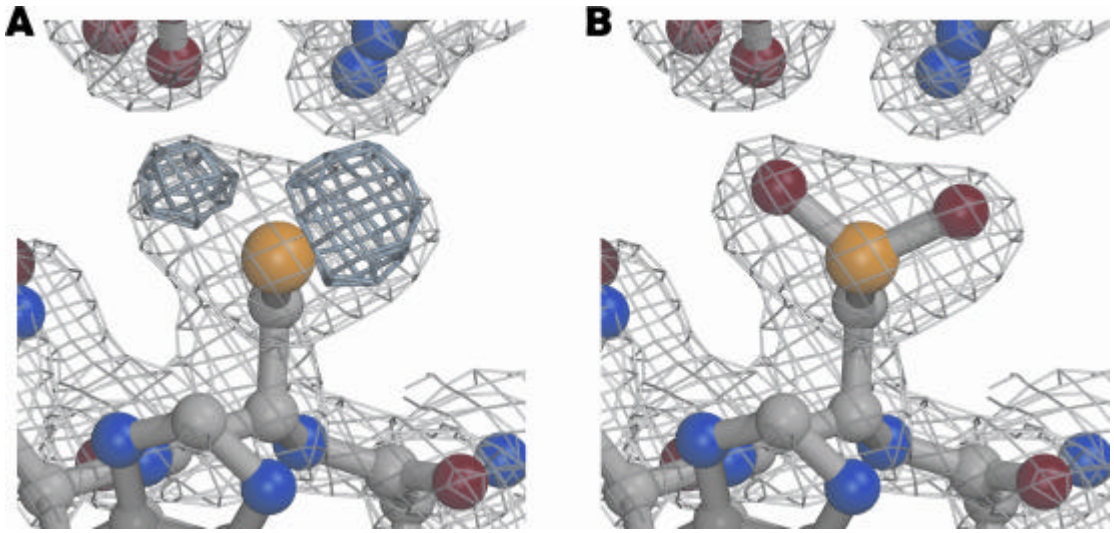


Figure 3.24 Identification of Cys-SO_2H

A. Cys174 with the $2F_o$-F_c map contoured at 1σ (*light grey*) and F_o-F_c map (*dark grey*) contoured at 3σ suggesting the presence of two oxygens. **B.** $2F_o$-F_c and F_o-F_c map after rebuilding to Cys-SO_2H, contoured at the same level. No unexplained difference density remains.

A functional Cys-SO_2H in a glucosidase has been described with the creation of a Glu→Cys mutant of the glucoamylase from *Aspergillus awamori* (Fierobe et al., 1998b). The accidental oxidation of the catalytic residue mutant led to a 1.6 fold increase in activity as compared to the wild type. This is explained as the more readily deprotonated Cys-SO_2H better stabilises the oxocarbenium-ion-like transition state than a glutamate does, hence increasing the k_{cat} (Fierobe et al., 1998a).

An opposing theory is the possibility of a role of the oxidized Cys-174 in the inactivation of the enzyme as an artifact of the purification procedure. Cys-SO_2H174 is hydrogen-bonded via one of its oxygens to the conserved His203 (Figure 3.25), a residue which could otherwise interact with the potential nucleophile Asp260. This could explain in two ways the inactivation of the enzyme: the CysSO_2H-bound histidine is a) unavailable to co-

ordinate the Mn^{2+} ion directly and/or, b) cannot orientate Asp260 in a position closer to the C1 of the maltose. Either scenario results in the reduced activity, recoverable by reduction, which is observed experimentally. Further experiments are required to determine whether it is this putative role in inactivation or a more direct role that Cys-SO_2H174 plays.

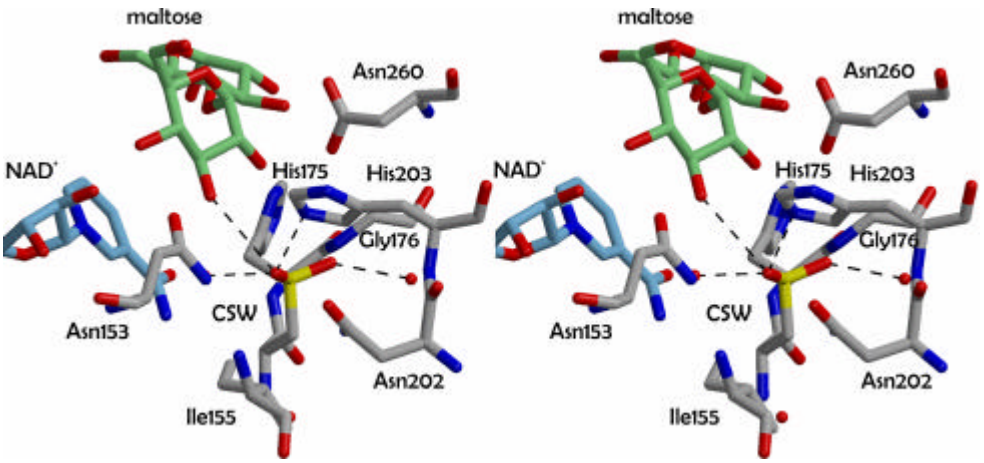


Figure 3.25 Residues interacting with Cys-SO_2H174

Stereo figure of the environment surrounding Cys-SO_2H174 (labelled CSW). Hydrogen bonds are drawn as *dashed lines*. The C atoms of NAD^+ are shown in *light blue*, of maltose in *light green* and water molecules are depicted as a *red dot*.

3.6.6 The Structure of a C174S Mutant

In an attempt to answer these questions, a C25S/C174S double mutant was prepared by our collaborators in the group of Prof. Liebl in Göttingen. This was crystallised and data collected (data set Ag12). The resulting structure showed very little difference to wild-type AglA, with the exception of the Cys174 loop near the active site. This was in fact more disordered than the wild-type and displayed alternative conformations between the two subunits. In monomer A the Ser174 side chain was located on the other side of the backbone compared to Cys-SO_2H174 and directed towards the centre of the protein (Figure 3.26 A). Like the wild-type apo-protein and substrate/cofactor-bound structures, this region in molecule A displayed higher B-factors than the corresponding loop in the other monomer – ranging from 30 $Å^2$ to 53 $Å^2$, compared to 25 $Å^2$ to 40 $Å^2$ in molecule B. In molecule B, the Ser174 O^d atom pointed towards the maltose in much the same manner as the Cys-SO_2H (Figure 3.26 B). An overlay of the two loops in relation to the maltose and nicotinamide ring of the NAD^+ is shown in Figure 3.26 C, which clearly demonstrates how the loops diverge from one another, affecting not only the positioning of the C174S side

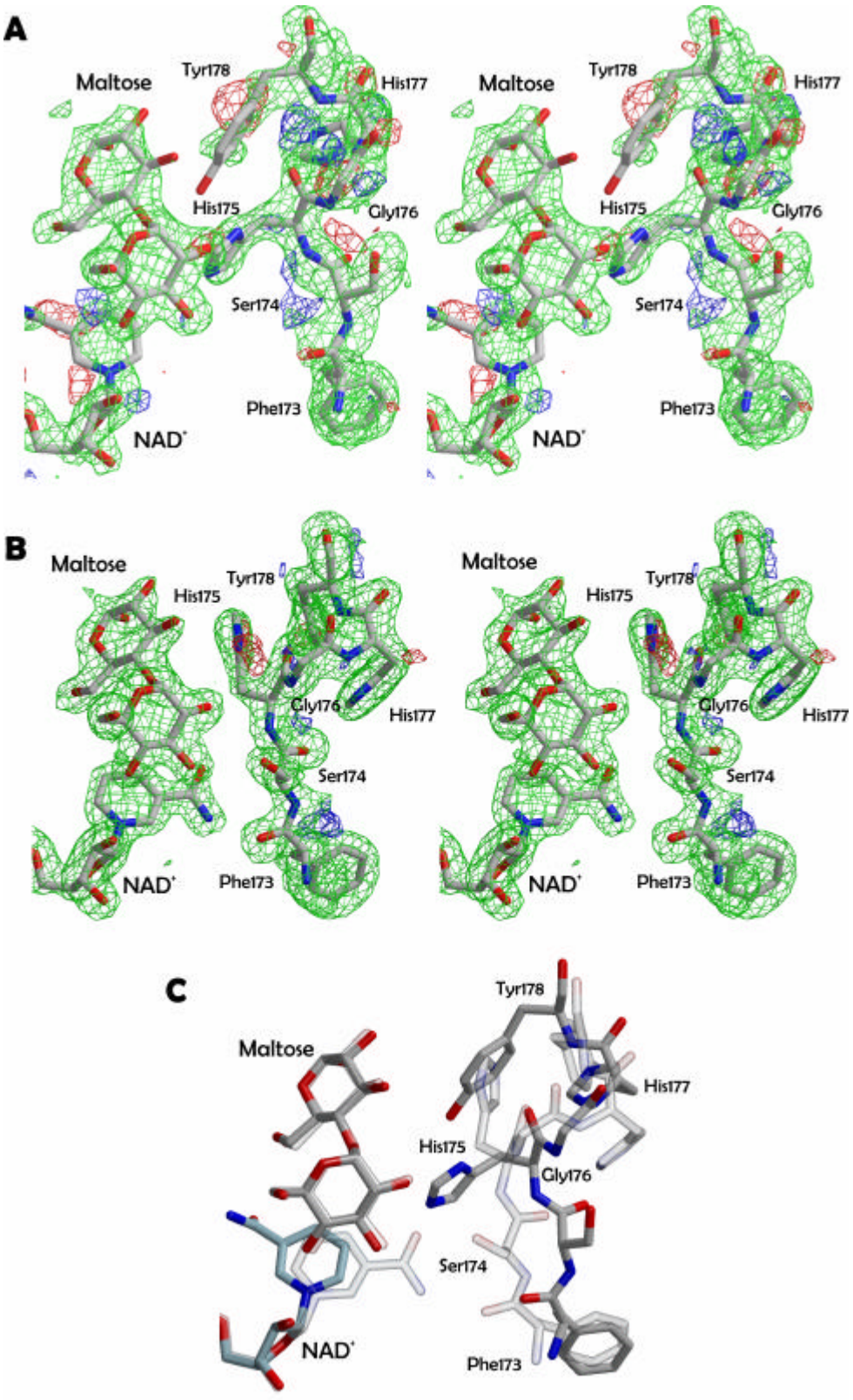


Figure 3.26 Flexibility of the C174S mutant at the active site

$2F_o$-F_c electron density maps, contoured to 1σ, are shown in *green* while F_o-F_c electron density maps, contoured to 3σ and -3σ, are shown in *blue* and *red* respectively. **A.** Stereo diagram of loop 6G with the nearby maltose and NAD^+ molecules. **B.** Stereo diagram of loop 6G with the nearby maltose and NAD^+ molecules. **C.** Overlay of loop6G, NAD+ and maltose from molecules A and B. Molecule B is transparent.

chain, but also the side chain positions of surrounding residues.

Biochemical characterization of this protein proved it to be inactive (W. Liebl, personal communication). Although we cannot determine whether this inactivation is due to removal of an essential part of the catalytic mechanism or an adverse affect on metal binding, we can conclude from the structure of this mutant that removal of the Cys-SO_2H174 has no affect on cofactor or substrate binding.

3.6.7 Localisation of the Metal Ion

Despite being present at concentrations of 10 mM in the crystallisation solution, the essential cofactor Mn^{2+} could not be located in the initial structure. According to the Metalloprotein Database (Castagnetto et al., 2002), one of the most common residues involved in co-ordinating Mn^{2+} ions in protein structures are histidines, and the binding pocket of AglA has several histidine residues located nearby the bound maltose. In particular His175 and His177 in the flexible loop 6G (average B-factor 52.9 $Å^2$) are situated adjacent to the maltose molecule. At the low pH of the crystallisation buffer (pH 4.6) the histidines are most probably protonated and therefore incapable of coordinating the Mn^{2+}. Another putative impediment to Mn^{2+} co-ordination is the oxidation of Cys174 as a coordination ligand, as discussed earlier.

To address this, co-crystallisation was performed with a series of metal ions at a concentration of 100mM. Crystals grew in the presence of $ZnCl_2$ and $CoCl_2$ and diffraction data was collected. The $CoCl_2$ crystal (Ag18, Table 3-1, pg. 30) did not contain NAD^+ and the refinement statistics (Table 3-4) show that the B-factors for the overall structure are slightly higher, but more importantly, the B-factor for the maltose residue is very high (~80 $Å^2$). This, together with the relatively poor resolution of this crystal suggests that either the absence of the NAD^+ cofactor, or the presence of cobalt ions, disturbed the binding of the substrate.

The F_o-F_c maps calculated for this structure displayed extra density in the active site, which correlated well to the presence of a bound cobalt ion. The coordination sphere of this cobalt ion consists of His203 $N^{\varepsilon 2}$, Cys-SO2H174 $O^{\delta 1}$ and the O3 atom of the -1 ring of the maltose molecule (Figure 3.27). These have distances of 2.18 Å, 1.87 Å and 2.14 Å, respectively, in molecule A with an identical coordination sphere in molecule B.

The presence of the cobalt ion in the active site and its interaction, not only with the putative catalytically-involved sulfinic acid, but also directly with the substrate suggests that this site is indeed the physiological metal-ion coordination site.

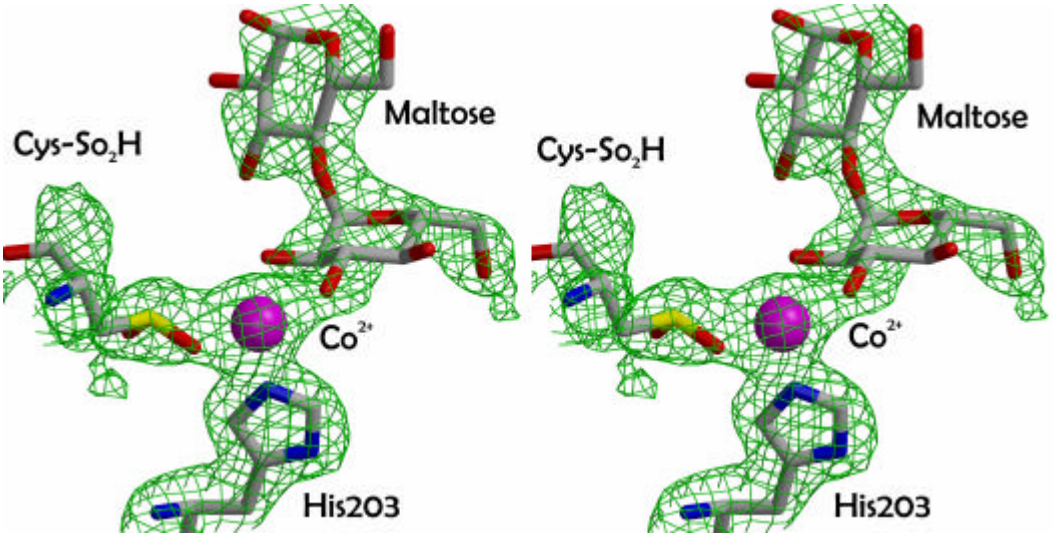


Figure 3.27 Stereo view of the cobalt binding site in molecule B of data set Ag18.

The cobalt ion (*purple*) is shown with its coordinating residues and the $2F_o$-F_c map of these residues, contoured to 1σ.

3.7 Similarity to Other GH4 Enzymes

3.7.1 Sequence Relationship of AglA to GH4

An alignment of representative members of GH4 is shown in Figure 3.18, pg. 56. As is highlighted in the alignment, the highest similarity occurs in the NAD-binding Rossman fold. At the position of helix D (residues 89 to 126), the insertion in the Rossman fold of AglA, AglA diverges from the rest of the family. This possibly indicates that helix D does not exist in other GH4 enzymes. Additionally, no sequence is aligned to that of helix H, and all align poorly after this insert. Attempts were made to align the last section of these sequences alone, with no conclusive results. This indicates that this section of the protein is not conserved across the family, but may have resulted from differing insertions in the sequence during evolution.

This idea is further reinforced by the observation that no or little sequence homology between AglA and other GH4 members is present in the region consisting of helices H through to L (residues 219 to 346). These include the helices which cover the active site and form the tunnel, and suggest that this morphology may not be present in all of GH4. As helix D is also required in AglA to stabilize these helices, the lack of similarity in both sections of the sequence may indicate a divergence within the family. This is easily observable in Figure 3.28, a dendogram calculated from the ClustalW sequence alignment. Here it is obvious that AglA (highlighted in red) is an outlier to the rest of the family.

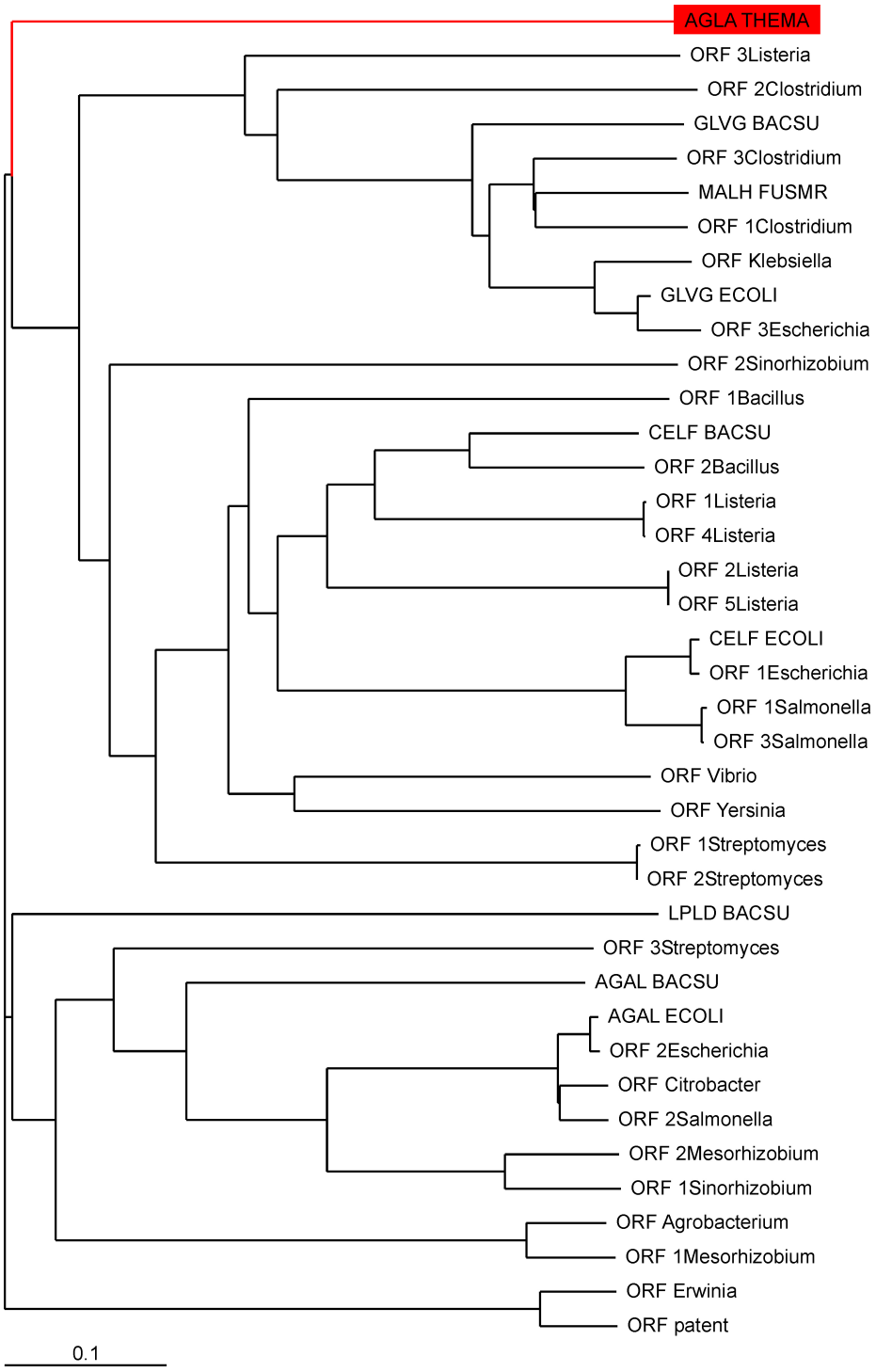


Figure 3.28 Dendogram of the ClustalW alignment of GH4 family members

The scale is 0.1 nucleotide substitutions per site. Dendogram was prepared using TREEVIEW (Thompson et al., 1994). Details of the abbreviations used in this figure can be found in Appendix 5.7, pg. 110.

3.7.2 GH4 Conserved Residues and Roles in Substrate Binding and Catalysis

The roles of those residues which are completely conserved across the family have been investigated in AglA, and the results discussed here are summarized in Table 3-6 and displayed in Figure 3.29.

Starting from the N-terminus, the first totally conserved residues across the GH4 family are found in the GXGS motif of the NAD^+ binding domain at the N-terminal end of the protein. These have already been discussed in detail in Section 3.6.2 on pg. 55.

Also located in the Rossman fold is Gly88. This is located near the adenine ring of the NAD^+, and coordinates a water molecule which binds to the AN7 atom of the adenine ring. However, it is also involved in a tight turn between β-sheet 4 and helix D, which may explain its absolute conservation.

The next two conserved residues are again in the Rossman fold. Asn153 binds the nicotinamide sugar ring via both the backbone oxygen and backbone nitrogen. Additionally, the side chain is directed towards the substrate and forms a hydrogen bond to O3 of the -1 ring. This dual job of coordinating both the substrate and the cofactor explains its conservation. Pro154 is involved in a sharp turn to start helix F and helps to maintain the integrity of the fold.

The conservation of Cys174, its modification to a cysteine-sulfinic acid, and the role it may play in the active site has been discussed in Section 3.6.5 on pg. 64. The residues Gly200, Asn202 and His203 are located in Sheet II, with Gly200 forming one of the hydrogen bonds to the next strand. Asn202 and His203 are located in the loop between these two strands and the side chains of both are directed towards the active site. Asn202 is located 3.2 Å below the Cys-SO_2H and may be important in positioning both the Cys-SO_2H and His203. As mentioned in the previous section, His203 is involved in metal ion coordination, thus explaining its presence in all GH4 glucosidases.

Several other sections of the sequence are also notable. Arg164 is present in over half of the family, or present as a glutamate, glutamine or lysine in the others. This residue forms two hydrogen bonds to Glu427 of the helix N located at the dimer interface and thus may be crucial for the integrity of the dimer. Also found functionally conserved across most of the family are Val390 and Glu391. These are located on β-strand 7 in the oxidoreductase fold. Val390 is present in a highly hydrophobic environment around which the β-sheet forms its

twist. Glu391 is only 2.8 Å from the conserved Asn202 suggested for the positioning of Cys174, and also hydrogen bonds to a water molecule, which forms a hydrogen bond to the Cys-SO_2H174.

Table 3-6 Role of the completely conserved residues in GH4

Residue	Location & Suggested Role
Gly10	NAD^+ binding
Gly12	NAD^+ binding
Ser13	NAD^+ binding
Asp39	NAD^+ binding
Gly88	Forms a tight turn between β-sheet 4 and helix D. Structural role.
Asn153	Located near active site. NAD^+ and maltose coordination.
Pro154	Creates sharp turn localizing Asn153 near active site. Structural role.
Cys174	Located in the active site, is oxidized to sulfinic acid and coordinates the metal ion.
Gly200	Located on β-strand 7 and forms H-bond to next strand in SheetII. Structural role.
Asn202	Located 3.2 Å beneath CysSO_2H174, and next to His203. Positioning of Cys174 / His203.
His203	Located near substrate. Metal ion coordination.

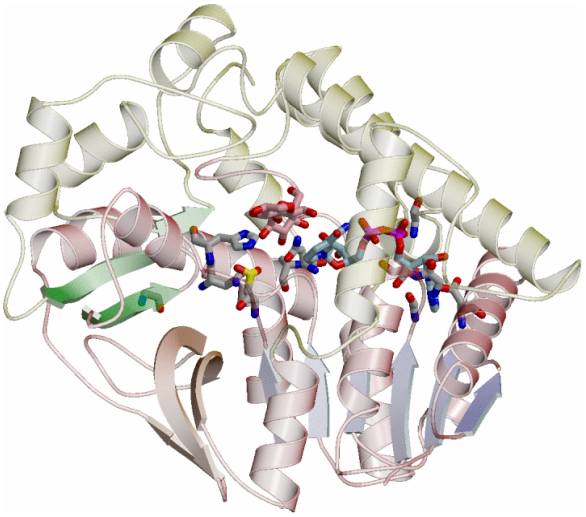


Figure 3.29 Location of GH4 conserved residues

The residues listed in Table 3-6 are shown in stick representation on the secondary structure of monomer A of AglA. The maltose molecule is highlight in *pink*, the NAD^+ molecule in *light blue*.

There is, however, one GH4 conserved sequence which is not conserved in AglA. A conserved GGX motif, where X is any hydrophobic residue, is present in the loop before the fourth helix (equivalent to AglA helix E) of the Rossman fold. These residues have been substituted in AglA by Tyr120, Tyr121 and Thr222.

Tyr120 OH is within hydrogen bonding distance to both Arg202 $O^{\delta 1}$ and Phe156 N, found at the end of sheet 7 and the start of helix F, respectively. It also has hydrophobic interactions to Pro154 and Ile155. Here, it effectively inserts itself into a pocket, keeping the loop on which it is found in position. In a similar manner, Tyr121 forms hydrophobic interactions with both Leu93 and Asp110 and forms a hydrogen bond to Arg97, thereby interacting with, and stabilizing, helix D, which may or may not be present in other GH4 enzymes. Both residues form strong backbone hydrogen bonds with other residues in the loop between helices D and E. Thr222, which has substituted a hydrophobic residue in the conserved GGX motif, is still involved in hydrophobic interactions with Glu157 of helix F. It is therefore consistent with the conservation of the GGX motif across GH4 that the backbone and hydrophobic interactions play an integral role in maintaining the conformation of this loop. Further, this suggests that the positioning and stability of this loop is important for all GH4 enzymes.

3.8 Similarity to Dehydratases

In addition to the similarity to dehydrogenases, the DALI analysis also revealed a similarity to another class of sugar-converting enzymes containing NAD-binding folds, namely the dehydratases. The relationship was highest (Z score = 7.1) for the dTDP-D-glucose-4,6-dehydratase (RmlB) from *Salmonella enterica* (pdb code 1KEU) (Allard et al., 2001). In this dehydratase the NAD^+ and the non-reducing end (-1 ring) of the sugar are bound in homologous positions to those found in AglA. Comparison of the sugar chain extension in AglA and the inhibitor in RmlB, demonstrated that they follow different paths. Similarities between the two active sites arise on closer inspection, where, in addition to the well conserved NAD^+ binding residues, the coordination of the O3 and O4 atoms in the AglA maltose -1 ring by Asn153 is performed by Tyr167 in RmlB. The proposed mechanism of RmlB involves an oxidation, dehydration and reduction step involving the NAD^+ as the redox cofactor (Figure 3.30). Also of interest is the conservation of AglA Cys174 to RmlB Cys194. Although participation of Cys194 in the proposed dehydratase catalytic

mechanism has not yet been suggested, transferring this reaction to AglA, with the characteristics of Cys174 discussed previously, could suggest a redox function as an integral part of the glycolysis reaction.

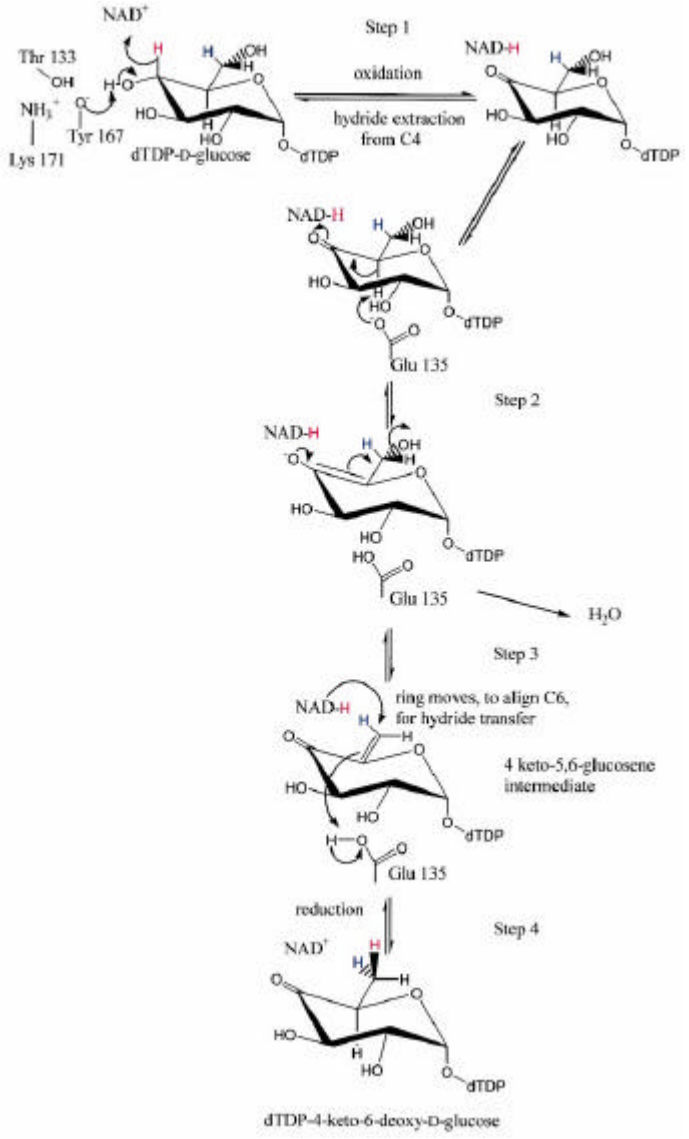


Figure 3.30 The proposed catalytic mechanism of dTDP-D-glucose-4,6-dehydratase (RmlB).

Step 1: NAD^+ extracts a hydride from the C4 position of the glucose ring. **Step 2:** Glu135 functioning as a general base removes a C5 proton. **Step 3:** the syn elimination of a water molecule between C5 and C6 to give the 4-keto-5,6-glucosene intermediate. **Step 4:** transfer of hydride from NADH to C6 of the glucose ring with inversion of configuration; Glu135, acting as a general acid, protonates C5. Taken from Allard et al. (2001).

3.9 Future Directions of Research

Although the crystal structure of AglA has offered many new insights into the GH4 famliy of glycosyl hydrolases, the information obtained from AglA produced under the conditions described here is limited. The inactivation of the protein, possibly due to oxidation during the purification process, must be prevented to elucidate clues to the mechanism of action. For this, anaerobic expression and purification must be employed, and crystallisation must also be performed under anaerobic conditions.

Another focus of future work should be the cloning, expression, purification and crystallisation of other members of GH4. The oxidation of the active site cysteine residue in AglA possibly is a result of expression in aerobic condtions. By focusing on a non-anaerobic organism, such as *E. coli*, whose endogenous GH4 enzyme is less likely to be oxidised so easily oxidized, protein may be obtained in an activate form. Investigation of a non-thermophilic could also provide insights into the thermostability of AglA, in comparison to other GH4 enzymes.