THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 278, No. 21, Issue of May 23, pp. 19151–19158, 2003
Printed in U.S.A.

Crystal Structure of *Thermotoga maritima* α-Glucosidase AglA Defines a New Clan of NAD^+-dependent Glycosidases*

Received for publication, November 14, 2002, and in revised form, February 13, 2003
Published, JBC Papers in Press, February 14, 2003, DOI 10.1074/jbc.M211626200

Jacinta A. Lodge‡, Timm Maier‡, Wolfgang Liebl§¶, Volker Hoffmann§, and Norbert Sträter‡‖**

From the ‡Institut für Chemie, Takustrasse 6, Freie Universität Berlin, D-14195 Berlin, the §Institut für Mikrobiologie und Genetik, Grisebachstrasse 8, Georg-August-Universität, D-37077 Göttingen, and the ‖Biotechnologisch-Biomedizinisches Zentrum, Am Deutschen Platz 5, Universität Leipzig, D-04103 Leipzig, Germany

Glycoside hydrolase family 4 represents an unusual group of glucosidases with a requirement for NAD^+, divalent metal cations, and reducing conditions. The family is also unique in its inclusion of both α- and β-specific enzymes. The α-glucosidase A, AglA, from *Thermotoga maritima* is a typical glycoside hydrolase family 4 enzyme, requiring NAD^+ and Mn^{2+} as well as strongly reducing conditions for activity. Here we present the crystal structure of the protein complexed with NAD^+ and maltose, refined at a resolution of 1.9 Å. The NAD^+ is bound to a typical Rossman fold NAD^+-binding site, and the nicotinamide moiety is localized close to the maltose substrate. Within the active site the conserved Cys-174 and surrounding histidines are positioned to play a role in the hydrolysis reaction. The electron density maps indicate that Cys-174 is oxidized to a sulfinic acid. Most likely, the strongly reducing conditions are necessary to reduce the oxidized cysteine side chain. Notably, the canonical set of catalytic acidic residues common to other glucosidases is not present in the active site. This, combined with a high structural homology to NAD-dependent dehydrogenases, suggests an unusual and possibly unique mechanism of action for a glycoside-hydrolyzing enzyme.

The glycosidic bond is one of the most stable bonds in nature, and with an estimated half-life of around 5 million years, this establishes glycoside hydrolases as among the most powerful enzyme catalysts known (1, 2). Glucosidases are a class of glycoside hydrolases that cleave the glycosidic bond to liberate a glucose molecule from an oligosaccharide. The α-glucosidases specifically hydrolyze the α-glucopyranosidic bond, thereby releasing an α-D-glucose from the nonreducing end of the sugar.

The general hydrolysis reaction of many glucosidases is well understood, involving either a single-step general acid/base catalysis or a double displacement mechanism, both involving a nucleophile-stabilized, oxocarbenium ion-like transition state and a pair of carboxylic acids in the active site (3). The first mechanism results in an inversion of the stereochemistry around the anomeric carbon, whereas in the second it is retained (4).

Glycoside hydrolase family 4 (GH4)[1] is a unique family of glucosidases that requires unusual cofactors and activation conditions, pointing toward a novel type of reaction mechanism (5–9). All except one of the GH4 enzymes investigated so far have shown a prerequisite for NAD^+ and a divalent cation, and some enzymes also require the addition of a reducing agent such as dithiothreitol or β-mercaptoethanol for activity (5–9). The requirement of NAD^+ for the hydrolysis reaction of a glycosidic bond has been described so far only for GH4 enzymes.

The role of NAD^+ for the activity of these enzymes is not yet understood. A sequence pattern at the N terminus of the GH4 enzymes is related to the "fingerprint" motif found in many of the classical NAD-binding enzymes (6). Mutations in this G*X*GS motif demonstrated that it is indeed important for NAD^+ binding and for the activity of *Thermotoga maritima* AglA (10). Spectroscopic measurements did not reveal the formation of NADH as a product or long lived intermediate of the catalytic reaction for several GH4 members (6, 10, 11), but a redox reaction during the hydrolysis reaction cannot be excluded.

Enzymes with various reaction specificities are found in this family, including α-glucosidases (*T. maritima* and *Thermotoga neapolitana* AglA (5)), α-galactosidases (*Escherichia coli* MelA (13)), 6-phospho-α-glucosidases (*Bacillus subtilis* GlvA (6), *Fusobacterium mortiferum* MalH (7), and *Klebsiella pneumoniae* AglB (14)), 6-phospho-β-glucosidases (*E. coli and B. subtilis* CelF (11)), and an α-D-glucuronidase (*T. maritima* Agu (9)). The GH4 family is unique in being the only family so far identified to contain enzymes that are specific for α-glycosidic bonds as well as enzymes that display β-specificity.

Currently no structures are available for a GH4 enzyme. In our ongoing effort to elucidate the catalytic mechanism of these enzymes, we have crystallized the 110-kDa homodimeric glucosidase AglA from *T. maritima*. Here we present the x-ray structure of the enzyme complexed with NAD^+ and the substrate maltose. The 1.9-Å resolution structure permits both a detailed description of the active site arrangement and implications regarding the catalytic mechanism. The structure further defines the evolutionary relationship of this unique enzyme family to other NAD-binding proteins.

EXPERIMENTAL PROCEDURES

Crystallization—T. maritima AglA was overexpressed in *E. coli* and purified as described in Ref. 5. The enzyme was crystallized using the hanging drop vapor diffusion method. Equal volumes (about 2 μl) of 10 mg/ml protein (in 50 mM Tris, pH 7.0) and reservoir solution (10% (w/v)

* This work was supported by the Deutsche Forschungsgemeinschaft. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 10BB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

¶ To whom correspondence may be addressed. Tel.: 49-551-393795; Fax: 49-551-394897; E-mail: wliebl@gwdg.de.

** To whom correspondence may be addressed. Tel.: 49-30-838-53456; Fax: 49-30-838-56702; E-mail: strater@chemie.fu-berlin.de.

[1] The abbreviations used are: GH4, glycoside hydrolase family 4; MIRAS, multiple isomorphous replacement and anomalous scattering; SAXS, small angle x-ray scattering.

TABLE I
Heavy atom derivatives

	$HgCl_2$	$SmCl_3$	$Pb(ac)_2$	PCMBS[a]	PIP[b]	Ta_6Br_{14}
Unit cell	a = 76.3 Å b = 85.9 Å c = 83.3 Å β = 106.2°	a = 75.3 Å 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	71,543	231,821	212,373	197,454	299,000	252,530
Unique observations	37,253	77,427	35,354	40,256	69,629	39,064
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

[a] PCMBS, *p*-chloromercuribenzene sulfonate.
[b] PIP, di-μ-iodobis(ethylenediamine)diplatinum(II) nitrate.

polyethylene glycol 6000, 1 M lithium chloride, 100 mM Tris/HCl, pH 4.6, and 2 mM $MnCl_2$) were mixed with 0.4 μl of 0.2 M maltose and equilibrated against the reservoir. Crystals of size 0.2 × 0.2 × 0.4 mm^3 appeared within 1–3 days.

Prior to data collection the crystals were transferred stepwise to a buffer containing 20% polyethylene glycol 6000, 1 M LiCl, 25% (v/v) glycerol, and 100 mM Tris/HCl, pH 5.0. After 10–30 min of incubation in the cryobuffer, the crystals were frozen in liquid nitrogen or in a nitrogen gas stream at 100 K. Heavy atom derivatives were dissolved in the reservoir buffer and added at a concentration of 5 mM to a 10-μl drop of the mother liquor containing a crystal. This was incubated overnight before being transferred to the cryobuffer and frozen as described.

Data Collection and Data Processing—X-ray data for four heavy atom derivatives were collected at the BAMline beamline at BESSY II (Berliner Elektronenspeicherring-Gesellschaft für Synchrotron Strahlung mbH II, Berlin, Germany) with a Mar345 imaging plate device as the detector (Table I). Data for two further 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. The software program Denzo/Scalepack (15) was used for data reduction (Table II).

Structure Determination and Refinement—Phases were determined by the multiple isomorphous replacement and anomalous scattering (MIRAS) method. All crystals belonged to the same crystal form of the space group $P2_1$ and were isomorphous with cell dimensions as given in Table II, having two monomers per asymmetric unit and a calculated Matthews parameter of V_m = 2.4. Six heavy atom derivatives have been used for phasing of the native data set to a 1.85-Å resolution. Heavy atom sites and initial phases were determined based on four derivatives collected at the BAMline using the software program SOLVE (16). Additional heavy atom sites of these four derivatives and of the $HgCl_2$ and $SmCl_3$ derivatives were located from difference Fourier maps. Phases were calculated with SHARP (17) using the isomorphous and anomalous contributions. After phase refinement and extension (software program DM (18)), the maps were readily interpretable. A model was built using the program O (19). Programs of the Collaborative Computational Project 4 software suite (18) and the MAPMAN (21), MOLEMAN2 (22), and LSQMAN (23) programs were used for many steps during structure solution and analysis. The structures were refined by simulated annealing and conjugate gradient energy minimization against maximum likelihood targets as implemented in the program CNS (24). Water molecules were picked automatically from $F_o - F_c$ electron density maps and verified manually.

Small Angle X-ray Scattering—SAXS measurements were performed on the X33-D1/2 beamline using multiwire proportional chambers with delay line readout at the EMBL outstation in Hamburg on the campus of the Deutsche Elektronen-Synchrotron research center (DESY). The scattering patterns were recorded at a sample detector distance of 1.7 m, equivalent to a range of momentum transfer of 0.02 $Å^{-1} < S < 0.35\ Å^{-1}$, with S being the scattering vector. The protein was purified as described (5) and diluted in 0.1 M Tris, pH 7.0, to concentrations between 3 and 15 mg/ml. Ten 1-min exposures at 15 °C were performed on each protein concentration with no protein degradation observed, and a buffer blank measurement was made before and after each protein sample. The programs SAPOKO and PRIMUS[2] were used to adjust for beam intensity and detector response, to average the frames, and to subtract the buffer background. Models of the two putative dimers and their fit to the experimental data were evaluated with the program CRYSOL (25).

TABLE II
Data collection and refinement statistics

Data collection	
X-ray source (wavelength in Å)	FR507 (1.54179Å)
Detector	Mar345
Space group	$P2_1$
Unit cell parameters	
a (Å)	75.0
b (Å)	85.7
c (Å)	83.6
β (°)	106.0
Mosaicity (°)	0.51
Resolution range (Å)	1.90–25
Highest resolution shell (Å)	1.90–1.97
Unique reflections	77,496
Multiplicity[a]	2.5 (2.4)
Completeness (%)[a]	96.6 (92.6)
R_{sym}[a]	0.038 (0.362)
I/σ[a]	18.9 (2.1)
Refinement	
Resolution (Å)	1.90
Highest resolution shell (Å)	1.90–1.97
Reflections (work set/test set)	75,935/2347
Non-hydrogen atoms	8504
R_{work}/R_{free}	0.196/0.234
Average B-factor ($Å^2$)	32.5
AglA (no. of atoms)	31.8 (7734)
NAD^+ (no. of atoms)	45.6 (88)
Maltose (no. of atoms)	44.6 (46)
Water (no. of molecules)	38.2 (646)
Root mean square deviation from ideal geometry	
Bond length (Å)	0.013
Bond angles (°)	1.53
Ramachandran plot	
Most favored (%)	90.4
Disallowed (%)	0

[a] The data in parentheses are for the highest resolution shell.

Structure Analysis and Preparation of Figures—Sequence alignment of GH4 enzymes was prepared using MULTALIN (26) and ESPRIPT (27). The topology diagram of AglA was prepared using TOPS (28). Structural homologues were identified using the DALI server (29). All figures of AglA were prepared using a graphical interface[3] for MolScript (30), CONSCRIPT (31), and Raster3D (32).

[2] P. V. Konarev, M. H. J. Koch, and D. I. Svergun, manuscript in preparation.
[3] N. Sträter, unpublished data.

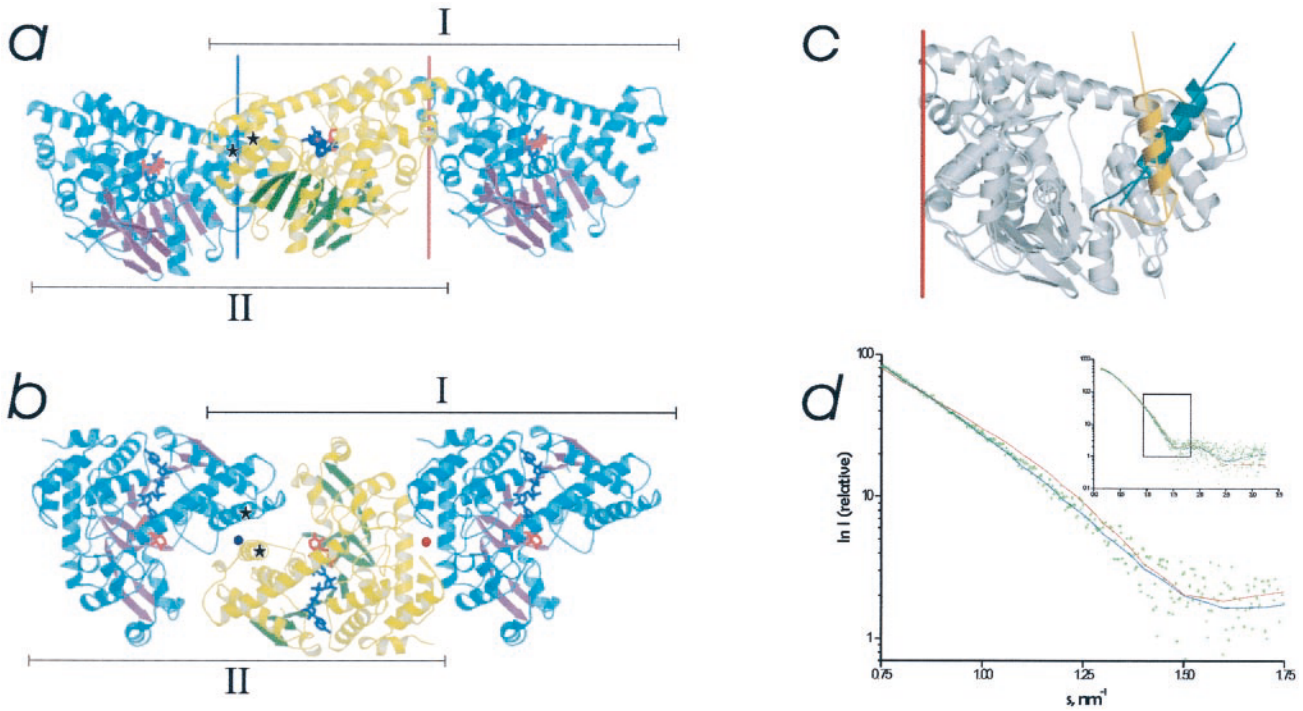

FIG. 1. **Crystal packing and dimer formation.** *a*, arrangement of molecules in the crystal looking at the 2-fold axis (represented by *blue* and *red lines*). One monomer (monomer A) of the asymmetric unit is shown in *yellow* and *green*, and the other (monomer B) is shown in *blue* and *purple*. The maltose is represented by *red* and the NAD^+ is represented by *blue rods*. Helix L in both monomers is indicated with a *star*. Possible dimers are indicated with *dotted lines*. *b*, crystal packing looking down the 2-fold axis. The coloring is as described for *panel a*. The described asymmetry between molecules A and B is observed around the axis and indicated by the *blue dot*. *c*, least squares superposition of monomer A (helix L is *orange*) and B (helix L is *green*), demonstrating the asymmetry of helix L. *d*, scattering curve from SAXS. The *inset graph* displays the overall curve, and the *highlighted area* is shown in the main figure. The curves calculated for dimers I (*blue*) and II (*red*) were fitted to the experimental data points (*green crosses*).

RESULTS

Overall Structure—α-Glucosidase A (AglA) from *T. maritima* was crystallized in complex with the substrate maltose and the cofactor NAD^+. The structure was solved by the MIRAS method using six heavy atom derivatives (Table I) and refined to a 1.9-Å resolution. The crystals were of the space group $P2_1$ with two monomers in the asymmetric unit. All of the 480 residues could be built into the electron density with the exception of residues 1–4 and 480. Several sections of the protein display high B-factor values; two loops in particular, residues 175–179 and 318–334, show notable disorder with B-factors of 52.9 and 65.2 $Å^2$, respectively, in monomer A compared with a B value for all residues in monomer A of 30.7 $Å^2$. Further details on structure determination and refinement are given in Table II.

The two AglA molecules in the asymmetric unit that are related by a noncrystallographic 2-fold rotation axis are shown in Fig. 1, *a* and *b*. The two molecules are almost identical with the remarkable exception of helix L, which is rotated by 51.7° with respect to the same helix in the other molecule (Fig. 1*c*). The overall root mean square deviation calculated for C-α atoms only between the two monomers is 2.18 Å. This, however, can be dissected into a root mean square deviation of 7.2 Å for residues 316–355, including helix L and its encompassing loops, and 0.71 Å for the rest of the structure. This clearly indicates that the majority of the deviation between the two monomers is caused by the different orientation of helix L. Unless noted otherwise, further descriptions of the structure of AglA in this paper concern monomer A.

Dimer Structure—The dimeric existence of AglA in solution has been confirmed by size exclusion chromatography (5); however, from the crystal structure it is not clear which of the protein contacts represents the true interface. The two possible candidates for the dimeric form of AglA are illustrated in Fig. 1, *a* and *b*. Dimer I is formed by hydrogen bond interactions primarily between helices N and O and by hydrophobic interactions between helices I and J. The presence of this dimer in solution would mean that the observed rotation of helix L between the two molecules related by non-crystallographic symmetry is a product of crystal packing, whereas the active sites, which are situated on opposing ends of the dimer, are independent of each other.

The second possible dimer, dimer II (Fig. 1, *a* and *b*), is formed by interactions between helices L and M of each subunit. This interface therefore could suggest a physiological role in the dimer formation for the observed rotation of helix L between the monomers related by non-crystallographic symmetry. 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, thereby reducing the overall surface area of the cleft and the size of the opening to the active site (Fig. 1*b*). Although smaller than the first interface and involving fewer hydrogen bonds, this second interface comprises 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 with 1051 $Å^2$ in dimer II. These values are both in the expected range of a native interface (33); however, they are too close to define clearly the true dimer interface. Further investigations were performed using SAXS experiments on the X33-D1/2 EMBL beamline at DESY (Fig. 1*d*). Differences in the predicted scattering curves for the two possible dimers can be observed in the central region of the curve. The fitted experimental data clearly supports the presence of dimer I in solution with a χ^2 value reflecting the agreement between the theoretical scattering of the predicted models

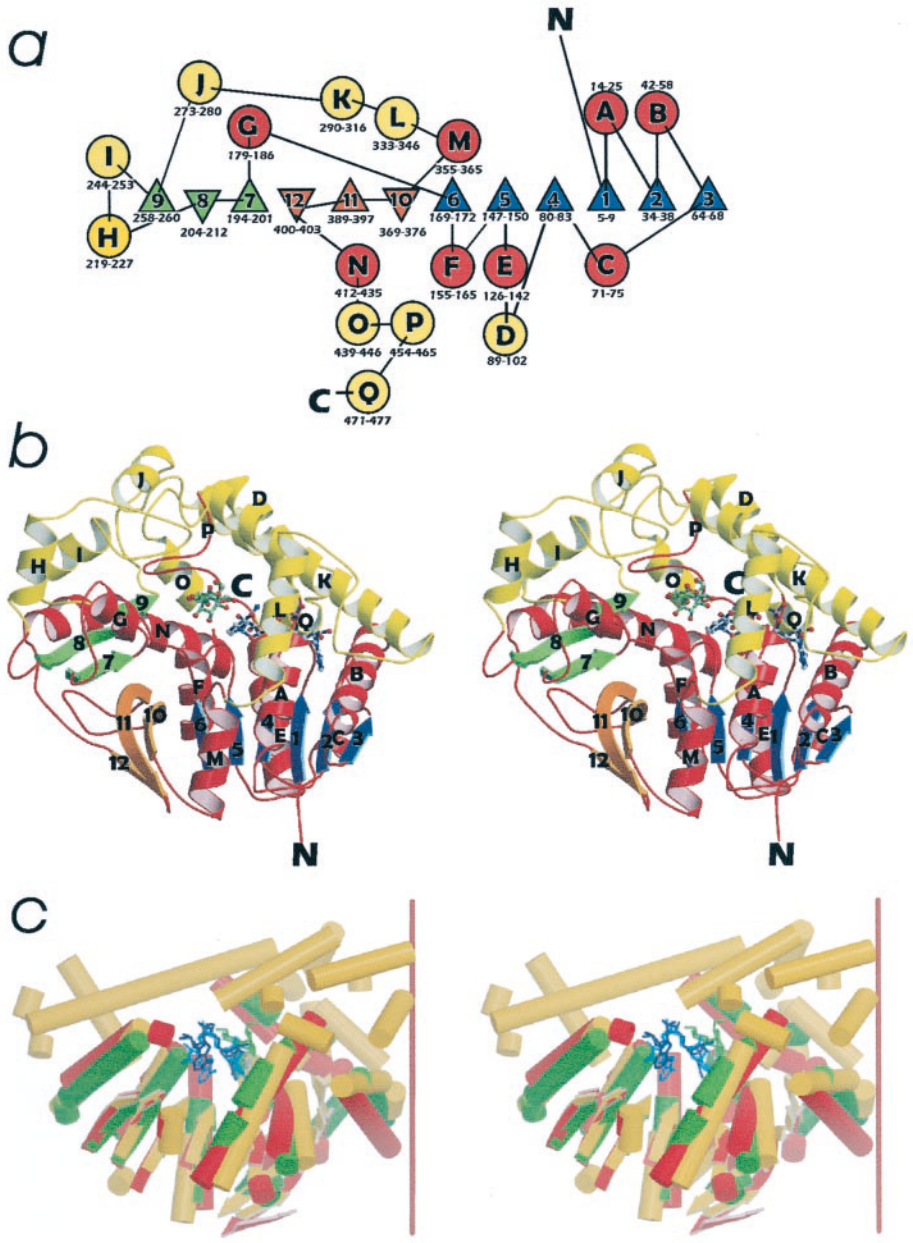

FIG. 2. **Structural homologues and substrate/cofactor density.** *a*, topology diagram of AglA monomer. Helices are represented by *circles*, and β-strands are represented by *triangles*. The coloring is as follows: sheet I, *blue*; sheet II, *green*; sheet III, *orange*; helices homologous to dehydrogenases, *red*; and helices unique to AglA, *yellow*. Secondary structure elements are labeled with the appropriate sequence numbering. *b*, stereo view of monomer A showing secondary structure. The coloring is as described for *panel a*. NAD^+ and maltose are shown with bonds drawn in *blue* and *green*, respectively. *c*, stereo view of a least squares superposition of AglA (*yellow*) with malate dehydrogenase (Protein Data Bank code 1BMD) (*green*) and lactate dehydrogenase (Protein Data Bank code 3LDH) (*red*). NAD^+ from AglA is depicted in *dark blue*, NAD^+ from malate dehydrogenase is shown in *light blue*, and the lactate dehydrogenase pyruvate is colored by atom type. Maltose from AglA is depicted in *green*. For clarity, loop regions are not shown.

($\chi^2 = 1$ for a perfect fit) and the experimental data of $\chi^2 = 0.95$ (as opposed to $\chi^2 = 1.47$ for dimer II). The presence of a 6-Å translation component for the screw operator relating the two monomers of dimer II compared with a 0.2-Å component for dimer I also favors the latter, as protein oligomer subunits are most often related by pure rotation with little or no translational movement (34).

Protein Fold—The overall fold of AglA is shown in Fig. 2, *a* and *b*. The monomer has 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 two hydrogen bonds at each contact. Sheet I consists of β-strands 1–6 and is a twisted parallel β-sheet, whereas the antiparallel β-sheet II (β-strands 7–9) shows only a slight twist. Sheet III, also antiparallel, consists of β-strands 10–12 and has a severe twist, which causes the sheet to rotate around 130° with reference to its starting point. Interestingly, this exceptional looking morphology is not found in any other glycosidases but is, in fact, a well characterized fold observed in oxidoreductases and sheets II and III, which together with the surrounding α-helices form a typical oxidoreductase fold (CATH code 3.90.110.10 (35)). Sheet I is enclosed on both sides by α-helices A–G, which are antiparallel to the sheet, and these combine to form a typical NAD-binding Rossman fold consisting of a β-α-β-α-β motif, which extends from residues 4 to 70 and includes the secondary structural elements 1-A-2-B-3. The first segment of this motif (β-α-β) has been predicted previously to extend from residues 4 to 38 in GH4 enzymes (10).

DISCUSSION

Similarity to Dehydrogenases and Dehydratases—A comparison of AglA with known structures using the DALI server (29) revealed that the closest structural relatives of AglA all belong to a subgroup of the NAD-dependent dehydrogenases, the 2-hydroxyacid dehydrogenases (*Z* scores, >15) (36). A superposition of AglA with lactate dehydrogenase (Protein Data Bank code 3LDH) and malate dehydrogenase (Protein Data Bank code 1BMD) is depicted in Fig. 2*c*. The core of the AglA protein consisting of the aforementioned NAD-binding and oxidoreductase folds is structurally homologous to the dehydrogenases. However, AglA contains extra secondary structure elements located on top of this dehydrogenase core, yet not organized

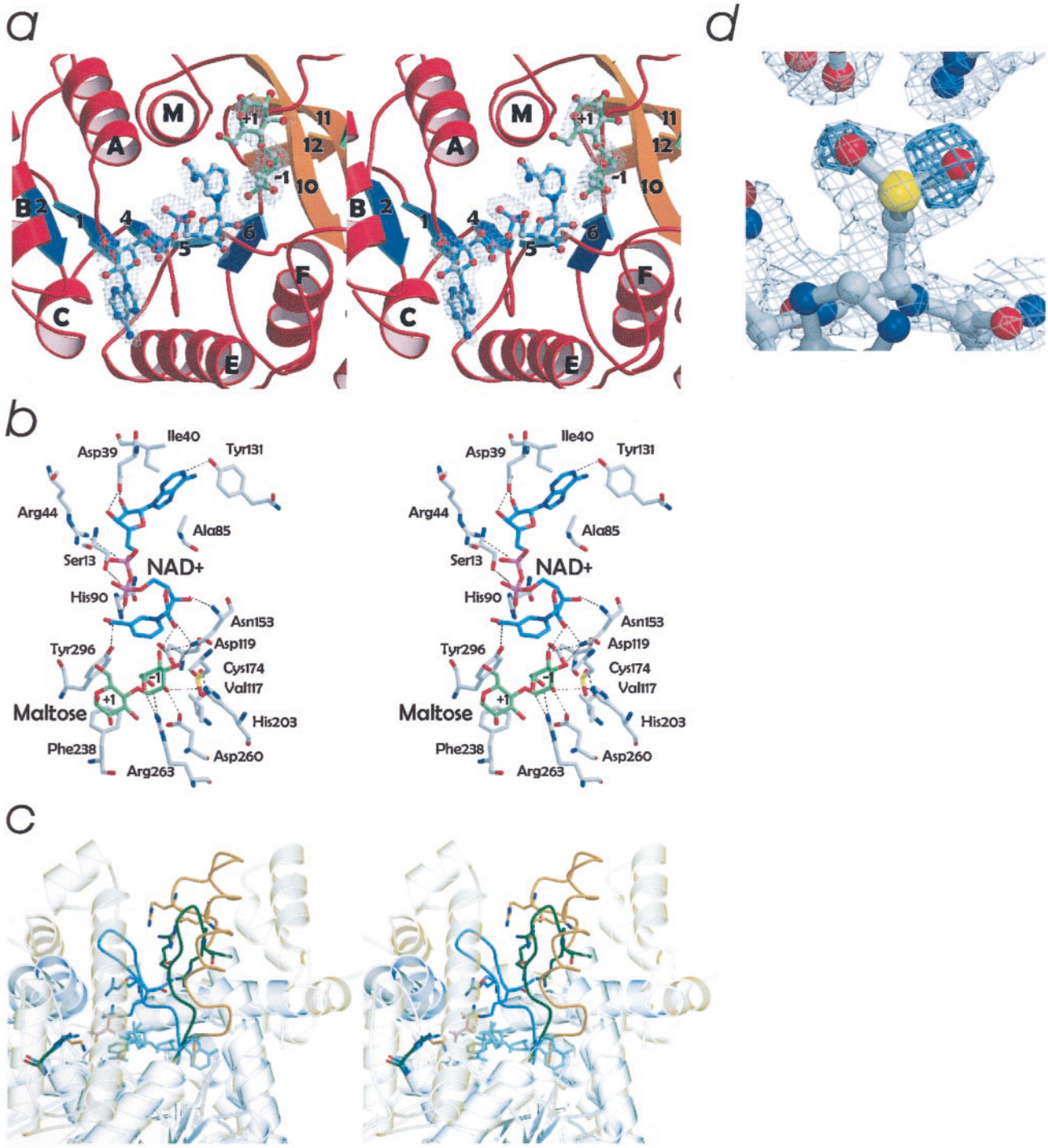

FIG. 3. **Active site of AglA.** *a*, stereo diagram of a $2F_o - F_c$ map displaying the density around the maltose and NAD^+ molecules contoured at 1.0 σ. NAD^+ and maltose are shown with bonds drawn in *blue* and *green*, respectively. For clarity helices K and L are not shown. *b*, schematic representation in stereo of the active site residues interacting with NAD^+ and maltose. Hydrogen bonds are indicated with *dotted lines*. *c*, stereo view of a least squares superposition of AglA (*gold*), porcine malate dehydrogenase (Protein Data Bank code 4MDH, *blue*), and dTDP-D-glucose 4,6-dehydratase (Protein Data Bank code 1KEU, *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 secondary structure elements. The bound NAD^+ of the dehydrogenase is shown in *light blue*, and the pyruvate is shown in *pink*. *d*, density surrounding Cys-174. The $F_o - F_c$ electron density omit map for the cysteine sulfinic acid is shown in *gray* with the positive density from the $F_o - F_c$ map calculated for cysteine indicated in *blue*.

into an obvious and distinct domain. The topology diagram (Fig. 2*a*) demonstrates that these segments (*highlighted* in *yellow*) are equivalent to a large insertion between residues 220 and 350 containing helices H–L and sheet 9, an inserted helix (helix D) between residues 89–104, and a C-terminal extension of around 50 amino acids encompassing helices O–Q. These regions are not found in any structurally known proteins.

A closer investigation of other sugar-converting enzymes containing NAD^+-binding folds revealed a similarity to a number of dehydratases, most specifically to the dTDP-D-glucose 4,6-dehydratase (RmlB; *Z* score, 7.1) from *Salmonella enterica* (Protein Data Bank code 1KEU) (37). In this dehydratase the NAD^+ and -1 sugar ring are bound in positions homologous to those found in AglA. However, the extension of the sugar chain,

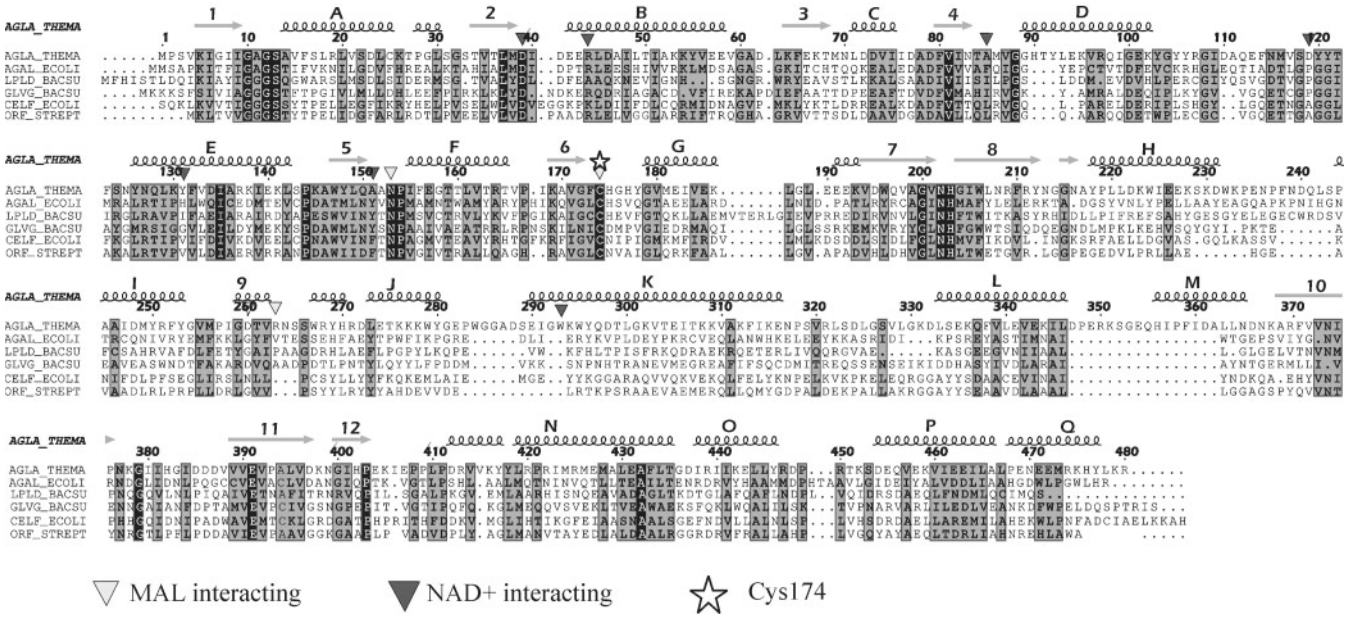

FIG. 4. **Alignment of glycoside hydrolase family 4.** Sequence alignments of several members of GH4, representing each phylogeny branch, are presented. Identity and homology levels shown are as determined for the entire family. Swiss-Prot accession numbers for each of the proteins shown are as follows: AGLA_THEMA, O33830; AGAL_ECOLI, P06720; LPLD_BACSU, P39130; GLVG_BACSU, P54716; CELF_ECOLI, P17411; and ORF_STREPT, CAB71199. Cys-174 is marked with a *star*, whereas those residues interacting with the maltose (*MAL interacting*) and NAD^+ molecules are indicated with *light* or *dark gray inverted triangles*, respectively.

or in the case of RmlB, the inhibitor, follows 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 O-3 and O-4 atoms in the AglA maltose −1 ring by Asn-153 is performed by Tyr-167 in RmlB. The proposed mechanism of RmlB involves an oxidation, dehydration, and reduction step involving NAD^+ as the redox cofactor. Also of particular interest is the conservation of AglA Cys-174 as Cys-194 in RmlB. Although participation of Cys-194 in the proposed dehydratase catalytic mechanism has not yet been suggested, the characteristics of AglA Cys-174 imply an importance and role in the active site that will be discussed below.

Active Site—Cocrystallization of AglA with NAD^+ and maltose clearly identified the location of the active site at the bottom of a deep cleft formed by NAD^+ and the loops between α-helices A and L, β-sheet 9 and helix I, and the helices G and M (Fig. 3*a*). The active site is covered by a single long helix (helix L, residues 290–316), a dominant feature unique to AglA and not found in related dehydrogenases. The protein without NAD^+ therefore contains a tunnel similar in appearance to those of the cellobiohydrolases and κ-carrageenases, the only glucosidases characterized previously with such a binding site topology (38, 39). However, although in the cellobiohydrolyases this tunnel has been suggested to aid in feeding long polysaccharides sequentially through the active site (39, 40), in AglA this tunnel is filled effectively by the NAD^+ molecule buried deep within it (41). The nicotinamide moiety remains extended into the cavity and thereby creates a deep pocket with the β-nicotinamide ring located at the bottom. The nicotinamide ring is the only part of the NAD^+ molecule that does not have a well defined electron density, suggesting that it is flexible in the active site pocket despite the presence of the substrate. The positioning of the nicotinamide moiety of the NAD^+ within the active site and in close proximity to the maltose molecule, as well as the biochemical characterization of its requirement for activity, suggest a participation of this group in the catalytic mechanism of AglA.

Coordination of Maltose—The electron density clearly indicated the presence of an intact maltose molecule. This demonstrates that the enzyme is inactive in the crystal as maltose is a good substrate and should be hydrolyzed to glucose (5). The nonreducing end of the maltose molecule (−1 ring) has a well defined electron density (see Fig. 3*a*) and is bound by hydrogen bonds to Asn-153, His-203, Asp-260, Asp-119, and Arg-263 and by hydrophobic stacking between Val-117 and the β-nicotinamide ring of NAD^+ (Fig. 3*b*). In contrast, the electron density maps indicate that the reducing end (+1 ring) is more flexible, which may be explained by the fact that its only interaction with the protein is a single hydrophobic stacking to Phe-238.

The binding of the substrate molecule by AglA in comparison with the dehydrogenases reveals an interesting relationship. The dehydrogenases contain around the active site three arginine residues, which are responsible for the substrate binding and may exist in an open (Protein Data Bank code 4MDH, porcine malate dehydrogenase) or closed state (Protein Data Bank code 1KEU, RmlB) (36, 37). Two of these arginines are found on a flexible loop that moves in to cover the binding site when the substrate is present, whereas the third arginine is found in the pocket of the active site (Fig. 3*c*) (36). The two arginines (Arg-91 and Arg-97) on the mobile loop are conserved in AglA as Arg-97 and Arg-107, which are located on helix D and the following loop, respectively, in positions that correspond to the open conformation of the dehydrogenase. This region is held in the open conformation in the AglA structure by the change from a loop to a helix structure (helix D) and its interaction with the large inserted helix K over the active site. However, another arginine, Arg-263, and an aspartate, Asp-260, occupy the same position as the arginines in the closed form and are supplied from the nonhomologous inserted loop between sheet 9 and helix G. In AglA the third arginine of the dehydrogenase-binding pocket is in the position of the active site disordered loop 6G (between strand 6 and helix G), which consists of residues 175–179. Interestingly, these conserved arginines are also present in the dehydratase RmlB, in which they are trapped, as in AglA, in the open conformation away from the active site.

Potential Catalytic Residues of the Active Site—The active sites of a number of glycosidases now have been characterized

structurally, and the catalytic mechanisms have been well described (2, 39). The active site of AglA, however, bears no resemblance to these. We have analyzed the residues surrounding the active site and their sequence conservation in GH4 (Figs. 3*b* and 4). Instead of an appropriately placed nucleophilic donor on one side of the maltose ring with the corresponding acceptor placed at the required distance to ensure inverting or retaining cleavage, the active site of AglA contains the nicotinamide group of the NAD^+ molecule and an array of residues that correspond more to the active sites of dehydrogenases than to those of glucosidases.

Site-directed mutagenesis on another GH4 enzyme, GlvA from *B. subtilis*, suggested several residues required for activity (6). GlvA Asp-41 is seen by sequence alignments (Fig. 4) to be equivalent to AglA Asp-39. This residue was predicted to be involved in NAD binding (6), and the AglA structure confirms its position at the end of β-sheet 2, interacting with the O-2′ and O-3′ atoms of the adenine sugar ring. The experiments on GlvA also suggested the residues Glu-111 and Glu-359 as the potential catalytic acid and nucleophile base in the active site. By sequence alignment, these are equivalent to AglA residues Glu-113 and Glu-391, respectively, neither of which is located at the catalytic site. Glu-113 is located at the dimer interface, suggesting that interference with dimerization played a role in the observed loss of activity of the mutant. This is of particular interest as GlvA is shown to be an active homotetramer that requires Mn^{2+} to tetramerize from homodimers (6). Glu-391 is closer to the active site, situated on the highly twisted β-strand 11 in sheet III. Although not directly involved in the active site, it interacts with Cys-174 via both a water molecule and Asn-202, and through the latter, it also interacts with His-203. Cys-174, Asn-202, and His-203 are conserved completely in GH4, and their roles will be discussed further. We suggest that an involvement of Glu-391 in the positioning of the active site residues explains its requirement for activity.

A single putative nucleophilic donor/acceptor, Asp-260, is located 3.4 Å from the C-1 atom of the maltose. This residue, however, is not conserved in GH4 and lacks the required partner residue on the other side of the maltose molecule. The only totally conserved residues interacting directly with the substrate are His-203 (interacting with O-2 and O-3) and Asn-153 (interacting with O-3 and O-4), located at the bottom of the binding pocket and capable of binding any sugar substrate.

The Cysteine Sulfinic Acid Cys-174—Another completely conserved residue located near the active site is Cys-174. Difference density around the sulfur of this cysteine displayed a "rabbit ear" morphology, which suggested that it was oxidized to a cysteine sulfinic acid (Cys-SO_2H) (Fig. 3*d*). This finding correlates well with the observed requirement for reducing conditions for activity characterized for the protein *in vitro*, suggesting that this cysteine may be a catalytically important residue in which oxidation causes enzyme inactivation during purification or even during heterologous expression in *E. coli*. The only other cysteine, Cys-25, is found 24 Å away from the active site and does not show oxidation.

Cysteine sulfinic acids have been described already in several protein structures, primarily as further oxidation states of the catalytically active but unstable cysteine sulfenic acids (Cys-SOHs). Several different roles have been proscribed for these Cys-SOHs. In NADH peroxidase, NADH oxidase, and peroxiredoxins, they function as either catalytically essential redox centers or transient intermediates during peroxide reduction and are thought to be stabilized by a nearby histidine and a potential Mg^{2+} ion (42). In nitrile hydratases a Cys-SH, Cys-SOH, and Cys-SO_2H form a "claw"-like structure that functions to coordinate a Fe(III) ion but plays no catalytic role (43). It is possible in AglA that a Cys-SOH is present in the active protein and is oxidized to Cys-SO_2H prior to crystallization.

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* (44). The accidental oxidation of the catalytic residue mutant led to a 1.6-fold increase in activity compared with the wild type. This is because the more strongly deprotonated Cys-SO_2H stabilizes the oxocarbenium ion-like transition state better than a glutamate, hence increasing the k_{cat} (12). Although an active Cys-SO_2H has yet to be identified in a wild-type glucosidase, its significance in light of the structural uniqueness of the active site of AglA cannot be ignored.

An opposing theory is the possibility of a role in the inactivation of the enzyme as an artifact of the purification procedure. Cys-SO_2H at position 174 is hydrogen-bonded via one of its oxygens to the conserved His-203, a residue which otherwise could interact with the potential nucleophile Asp-260. This could explain in two ways the inactivation of the enzyme: the Cys-SO_2H-bound histidine is (*a*) unavailable to coordinate the Mn^{2+} ion directly and/or (*b*) cannot pull Asp-260 into a position closer to the C-1 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_2H at position 174 plays.

Mn^{2+} Binding—Despite being present at concentrations of 10 mM in the crystallization solution, Mn^{2+} was not located in the structure. According to the metalloprotein data base (20), some of the most common residues involved in coordinating Mn^{2+} ions in protein structures are histidines, and the binding pocket of AglA has several located nearby the bound maltose. In particular His-175 and His-177 in the disordered loop 6G (B-factor of 52.9 $Å^2$) are situated adjacent to the maltose molecule. At the low pH of the crystallization buffer (pH 4.6), the histidines most probably are protonated and therefore incapable of coordinating the Mn^{2+}. Another putative impediment to Mn^{2+} coordination is the oxidation of Cys-174 as a coordination ligand, as discussed earlier.

Conclusion—The first crystal structure of a family 4 glucosidase shows that this family represents a new structural clan for the glycoside hydrolases with high structural homology to dehydrogenases. The structure defines the maltose binding mode and shows clearly the binding site of the NAD^+ cofactor as a typical Rossman fold. In agreement with previous evidence that NAD^+ is absolutely required for activity by these enzymes, the maltose molecule is located in close proximity to NAD^+. The finding that the maltose molecule is not turned over in the presence of NAD^+, Mn^{2+}, and reducing agents demonstrates that the enzyme as seen in the crystal structure is not in a catalytically active state. Also, a Mn^{2+} ion could not be located in the density maps. The oxidation of the potential catalytic residue Cys-174 to a sulfinic acid may be the cause of the inactivation; however, a catalytic role for the sulfinic acid is also possible.

Aside from NAD^+ and Cys-174, the structure shows additional residues that may play a role in the catalytic mechanism. However, in the absence of additional data on possible catalytic intermediates and the localization of the metal ion, the enzyme mechanism currently remains unclear. In particular, the role of NAD^+ has yet to be established, although currently a role in substrate binding, in transition state stabilization via its positive charge, as an acid/base catalyst, or even as a reduction/oxidation mechanism appears possible. These interesting features of the active site, combined with the uniqueness of the structure of AglA compared with other glucosidases and its

high homology to dehydrogenases, suggest an exciting new turn in classical views on glycosidase mechanisms.

Acknowledgments—We thank Uwe Müller and Timo Niedenzu of the Protein Structure Factory for assistance during crystallographic data collection at BESSY in Berlin and Dimitri Svergun for help with SAXS data collection at the EMBL/DESY in Hamburg. We also thank Carsten Raasch for protein purification.

REFERENCES

1. Wolfenden, R., Lu, X., and Young, G. (1998) *J. Am. Chem. Soc.* **120,** 6814–6815
2. Zechel, D. L., and Withers, S. G. (2000) *Acc. Chem. Res.* **33,** 11–18
3. Rye, C. S., and Withers, S. G. (2000) *Curr. Opin. Chem. Biol.* **4,** 573–580
4. Koshland, D. E. (1953) *Biol. Rev. Camb. Philos. Soc.* **28,** 416–436
5. Raasch, C., Streit, W., Schanzer, J., Bibel, M., Gosslar, U., and Liebl, W. (2000) *Extremophiles* **4,** 189–200
6. Thompson, J., Pikis, A., Ruvinov, S. B., Henrissat, B., Yamamoto, H., and Sekiguchi, J. (1998) *J. Biol. Chem.* **273,** 27347–27356
7. Thompson, J., Gentry-Weeks, C. R., Nguyen, N. Y., Folk, J. E., and Robrish, S. A. (1995) *J. Bacteriol.* **177,** 2505–2512
8. Robrish, S. A., Fales, H. M., Gentry-Weeks, C., and Thompson, J. (1994) *J. Bacteriol.* **176,** 3250–3256
9. Suresh, C., Rus'd, A. A., Kitaoka, M., and Hayashi, K. (2002) *FEBS Lett.* **517,** 159–162
10. Raasch, C., Armbrecht, M., Streit, W., Hocker, B., Strater, N., and Liebl, W. (2002) *FEBS Lett.* **517,** 267–271
11. Thompson, J., Ruvinov, S. B., Freedberg, D. I., and Hall, B. G. (1999) *J. Bacteriol.* **181,** 7339–7345
12. Fierobe, H. P., Clarke, A. J., Tull, D., and Svensson, B. (1998) *Biochemistry* **37,** 3753–3759
13. Liljestrom, P. L., and Liljestrom, P. (1987) *Nucleic Acids Res.* **15,** 2213–2220
14. Thompson, J., Robrish, S. A., Immel, S., Lichtenthaler, F. W., Hall, B. G., and Pikis, A. (2001) *J. Biol. Chem.* **276,** 37415–37425
15. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S., eds) pp. 56–62, SERC Daresbury Laboratory, Warrington, United Kingdom
16. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 849–861
17. de la Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276,** 472–494
18. Collaborative Computational Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
19. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
20. Castagnetto, J. M., Hennessy, S. W., Roberts, V. A., Getzoff, E. D., Tainer, J. A., and Pique, M. E. (2002) *Nucleic Acids Res.* **30,** 379–382
21. Kleywegt, G. J., and Jones, T. A. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52,** 826–828
22. Kleywegt, G. J. (1997) *J. Mol. Biol.* **273,** 371–376
23. Kleywegt, G. J. (1996) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **52,** 842–857
24. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
25. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) *J. Appl. Crystallogr.* **28,** 768–773
26. Corpet, F. (1988) *Nucleic Acids Res.* **16,** 10881–10890
27. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) *Bioinformatics* (*Oxf.*) **15,** 305–308
28. Westhead, D. R., Slidel, T. W. F., Flores, T. P. J., and Thornton, J. M. (1999) *Protein Sci.* **8,** 897–904
29. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233,** 123–138
30. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
31. Lawrence, M. C., and Bourke, P. (2000) *J. Appl. Crystallogr.* **33,** 990–999
32. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
33. Janin, J., and Chothia, C. (1990) *J. Biol. Chem.* **265,** 16027–16030
34. Goodsell, D. S., and Olson, A. J. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153
35. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) *Structure* **5,** 1093–1108
36. Chapman, A. D., Cortes, A., Dafforn, T. R., Clarke, A. R., and Brady, R. L. (1999) *J. Mol. Biol.* **285,** 703–712
37. Allard, S. T., Giraud, M. F., Whitfield, C., Graninger, M., Messner, P., and Naismith, J. H. (2001) *J. Mol. Biol.* **307,** 283–295
38. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) *Structure* **9,** 513–525
39. Davies, G., and Henrissat, B. (1995) *Structure* **3,** 853–859
40. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K., and Jones, T. A. (1990) *Science* **249,** 380–386
41. Rossman, M. G., Liljas, A., Brandon, C. I., and Banaszak, L. T. (1975) *Enzymes* **11,** 61–102
42. Choi, H. J., Kang, S. W., Yang, C. H., Rhee, S. G., and Ryu, S. E. (1998) *Nat. Struct. Biol.* **5,** 400–406
43. Nagashima, S., Nakasako, M., Dohmae, N., Tsujimura, M., Takio, K., Odaka, M., Yohda, M., Kamiya, N., and Endo, I. (1998) *Nat. Struct. Biol.* **5,** 347–351
44. Fierobe, H. P., Mirgorodskaya, E., McGuire, K. A., Roepstorff, P., Svensson, B., and Clarke, A. J. (1998) *Biochemistry* **37,** 3743–3752

Crystal Structure of *Thermotoga maritima* α-Glucosidase AglA Defines a New Clan of NAD $^{+}$-dependent Glycosidases

Jacinta A. Lodge, Timm Maier, Wolfgang Liebl, Volker Hoffmann and Norbert Sträter

J. Biol. Chem. 2003, 278:19151-19158.
doi: 10.1074/jbc.M211626200 originally published online February 14, 2003
