



**HAL**  
open science

# The Three-dimensional Structure of Invertase ( $\beta$ -Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases

Francois Alberto, Christophe Bignon, Gerlind Sulzenbacher, Bernard Henrissat, Mirjam Czjzek

## ► To cite this version:

Francois Alberto, Christophe Bignon, Gerlind Sulzenbacher, Bernard Henrissat, Mirjam Czjzek. The Three-dimensional Structure of Invertase ( $\beta$ -Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases. *Journal of Biological Chemistry*, 2004, 279 (18), pp.18903 - 18910. 10.1074/jbc.M313911200 . hal-01818938

**HAL Id: hal-01818938**

**<https://hal.science/hal-01818938>**

Submitted on 19 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Three-dimensional Structure of Invertase ( $\beta$ -Fructosidase) from *Thermotoga maritima* Reveals a Bimodular Arrangement and an Evolutionary Relationship between Retaining and Inverting Glycosidases\*

Received for publication, December 19, 2003, and in revised form, February 5, 2004  
Published, JBC Papers in Press, February 18, 2004, DOI 10.1074/jbc.M313911200

François Alberto, Christophe Bignon, Gerlind Sulzenbacher, Bernard Henrissat, and Mirjam Czjzek‡

From the Architecture et Fonction des Macromolécules Biologiques, CNRS and Université Aix-Marseille I & II, Institut de Biologie Structurale et Microbiologie, 31 Chemin Joseph Aiguier, 13402 Marseille cedex 20, France

*Thermotoga maritima* invertase ( $\beta$ -fructosidase) hydrolyzes sucrose to release fructose and glucose, which are major carbon and energy sources for both prokaryotes and eukaryotes. The name “invertase” was given to this enzyme over a century ago, because the 1:1 mixture of glucose and fructose that it produces was named “invert sugar.” Despite its name, the enzyme operates with a mechanism leading to the retention of the anomeric configuration at the site of cleavage. The enzyme belongs to family GH32 of the sequence-based classification of glycosidases. The crystal structure, determined at 2-Å resolution, reveals two modules, namely a five-bladed  $\beta$ -propeller with structural similarity to the  $\beta$ -propeller structures of glycosidase from families GH43 and GH68 connected to a  $\beta$ -sandwich module. Three carboxylates at the bottom of a deep, negatively charged funnel-shaped depression of the  $\beta$ -propeller are essential for catalysis and function as nucleophile, general acid, and transition state stabilizer, respectively. The catalytic machinery of invertase is perfectly superimposable to that of the enzymes of families GH43 and GH68. The variation in the position of the furanose ring at the site of cleavage explains the different mechanisms evident in families GH32 and GH68 (retaining) and GH43 (inverting) furanosidases.

invertase indicated that the enzyme operates with a molecular mechanism leading to overall retention of the anomeric configuration (2). The breakdown of sucrose is widely used as a carbon or energy source by bacteria, fungi, and plants. In plants, both glucose and fructose are implicated in the signaling pathways by which sucrose concentration functions as a key sensor of the nutritional status of plants, and, thus, invertase plays a fundamental role in controlling cell differentiation and development (3, 4). Commercially, invertase is mainly used in the confectionery industry, where fructose is preferred over sucrose because of a sweeter taste and a lower propensity to crystallize.

Although animals, including man, display a strong preference for sucrose-containing diets, their genomes do not encode invertases. Instead, they use a different and unrelated enzyme, sucrose  $\alpha$ -glucosidase (EC 3.2.1.48), to hydrolyze sucrose. The genomes of human gut microorganisms such as *Bacteriodes thetaiotamicron* (5) and *Bifidobacterium longum* (6) do possess invertase genes, demonstrating that these organisms benefit from the large intake of sucrose by humans.

Invertases are found in family GH32 of the sequence-based classification of glycoside hydrolases (afmb.cnrs-mrs.fr/CAZY) (7). This family, which includes over 370 members (as of January 2004) from plant, fungal, and bacterial origin, contains not only invertases but also other fructofuranosidases such as inulinase (EC 3.2.1.7), levanase (EC 3.2.1.65), and exo-inulinase (EC 3.2.1.80), and transfructosidases such as sucrose: sucrose 1-fructosyltransferase (EC 2.4.1.99) and fructan:fructan 1-fructosyltransferase (EC 2.4.1.100).

Glycoside hydrolases or glycosidases are a widespread group of enzymes displaying a great variety of protein folds and substrate specificities. They share a common defining feature in two critically located acidic residues, which make up the catalytic machinery responsible for the cleavage of glycosidic bonds. These two invariant residues have been identified experimentally in yeast invertase as an aspartate located close to the N terminus acting as the nucleophile (8) and a glutamate acting as the general acid/base (9). The enzymatic hydrolysis of glycosidic bonds has two possible stereochemical outcomes, inversion or retention of the anomeric configuration. Invertase is a retaining enzyme (2). With no known exception to date, the molecular mechanism appears conserved among the members of a given sequence-based family (10, 11). Sensitive sequence analyses coupled to structural comparisons have revealed significant similarities between representatives of different families, accompanied by a conservation of the catalytic machinery and of the stereochemical outcome of the reaction, reflecting

Invertase, the  $\beta$ -D-fructofuranosidase (EC 3.2.1.26) that cleaves sucrose into fructose and glucose is one of the earliest discovered enzymes. It was isolated in the second half of the 19th century, and its name was coined because the enzyme produces “invert” sugar, which is a 1:1 mixture of dextrorotatory D-glucose and levorotatory D-fructose (1). Because of its chemical structure, sucrose can be cleaved by either  $\alpha$ -glucosidase or  $\beta$ -fructofuranosidase activity. Koshland and Stein established that invertase is a  $\beta$ -fructofuranosidase by performing the reaction in  $^{18}\text{O}$ -labeled water and determining the  $^{18}\text{O}$  content of the products (2). The transfructosylation activity of

\* This work was partly funded by grant QLK5-CT-2001-00443 (EDEN) from the European Commission. 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 1UYP) 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 should be addressed. Tel.: 33-491-164-513; Fax: 33-491-164-536; E-mail: czjzek@afmb.cnrs-mrs.fr.

ancient divergence from a common ancestor to acquire novel substrate specificities (12). The evolutionarily, structurally, and mechanistically related families were grouped together in higher hierarchical level termed "clans" (10).

Threading analyses and homology modeling have led to the prediction that, as a member of glycosidase family GH32, invertase would display a six-bladed  $\beta$ -propeller fold related to that of influenza virus neuraminidase (13). However, the recent report on the three-dimensional structure of the family GH68 levanucrase from *Bacillus subtilis* (14) revealed that it had a novel five-bladed propeller fold, which has only been described previously for tachylectin (15) and for the family GH43  $\alpha$ -L-arabinanase from *Cellvibrio japonicus* (16). Recent detailed sequence analyses have revealed the existence of sequence motifs conserved in the glycosidase families GH32, GH43, GH62, and GH68, suggesting a possible structural relationship between these families (17) despite the opposite mechanisms in GH32 and GH68 (retaining) and GH43 (inverting). It should be noted that, because of the rapid mutarotation of furanoses, it is very difficult to experimentally determine the stereochemical course of the reaction catalyzed by furanosidases such as the family GH43  $\alpha$ -L-arabinofuranosidases. Three independent reports have, however, concluded that family GH43 enzymes operate by an inverting mechanism (18–20). The mechanism prevailing in family GH62 is not known.

After over a century of investigations and almost 40 years since the first crystal structure of a protein was solved, no three-dimensional structure of an invertase or of any member of glycosidase family GH32 has been reported. Here we report the three-dimensional crystal structure of *Thermotoga maritima* invertase. This thermostable enzyme has recently been biochemically characterized by Liebl *et al.* (21), who have determined that it liberates fructose from various substrates such as sucrose, raffinose, and inulin. The structure not only provides a template for all members of family GH32 (including invertases, inulinases, levanases, exo-inulinases, sucrose:sucrose 1-fructosyltransferases, and fructan:fructan 1-fructosyltransferases), but it also allows dissection of the exquisite details that distinguish retaining and inverting furanosidases with a perfectly superimposable catalytic machinery.

#### EXPERIMENTAL PROCEDURES

**Protein Cloning, Expression, and Purification**—Genomic DNA of *T. maritima* strain MSB8 (DSM 3109), kindly provided by Dr. Wolfgang Liebl (Georg-August-Universität, Göttingen, Germany), was used to amplify the invertase gene (GenBank™ accession number AAD36485). The *Escherichia coli* strains used were DH5 $\alpha$  for cloning experiments and BL21pLysS for expression. Vector pDONR is from Invitrogen, whereas vector pDEST170/I is the modified vector pDEST17 from Invitrogen by insertion of *lacO* and *lacI*, to prevent expression leakage. The invertase gene was amplified using INV-F (5'-TTCAAGC-CGAATTACTACTT-3') and INV-R (5'-TCACAACCATATGTCTCGA-3') primers containing recombination sequences for integration in Gateway™ vectors. PCR was performed using 500 ng of total genomic DNA of *T. maritima*, 300 nM each primer, 1.2 mM dNTP, 2.5 units of *Pfx* polymerase (Invitrogen), 1 $\times$  *Pfx* buffer (Invitrogen), and 1 mM MgSO<sub>4</sub>. The amplification program was 94 °C for 5 min followed by 30 cycles of 94 °C for 45 s, 55 °C for 30 s, and 68 °C for 2 min. The amplification was completed with a final extension at 68 °C for 10 min. The amplification product was purified by precipitation in 30% polyethylene glycol 8000 and 30 mM MgCl<sub>2</sub> and re-suspended in 50  $\mu$ l of TE buffer (10 mM Tris and 0.5 mM EDTA, pH 7.5). The PCR product was cloned in pDONR (Invitrogen) and then in pDEST170/I vectors as described in the manual supplied by Invitrogen (22) to obtain the plasmid pINV.

A single colony of BL21 pLysS containing the pINV plasmid was used to inoculate 40 ml of TBAC (Terrific broth supplemented with 100  $\mu$ g/ml ampicillin and 34  $\mu$ g/ml chloramphenicol). The culture was incubated overnight at 37 °C with constant shaking. This culture was used to inoculate 3 liters of TBAC. Incubation was done at 37 °C with vigorous shaking (240 rpm), and 0.5 mM of isopropyl-1-thio- $\beta$ -D-galactopyranoside was added when  $A_{600}$  reached 0.8. This induction was followed by

another incubation at 37 °C of 4 h. Cultures were pelleted and then re-suspended in 50 ml of lysis buffer (50 mM Tris pH 8, 150 mM NaCl, 10 mM imidazole, 1 mM EDTA, and 0.1% Triton X-100) containing 1 mM phenylmethylsulfonyl fluoride and 0.25 mg/ml lysozyme. This cell suspension was kept overnight at -80 °C. After thawing, the lysate was supplemented with 10  $\mu$ g/ml DNase I and 20 mM MgSO<sub>4</sub> and then incubated at 37 °C until it became fluid. The supernatant containing soluble proteins was separated from the pellet by centrifugation (20,000  $\times$  g) for 30 min at 4 °C.

The SeMet<sup>1</sup> protein was produced as follows. A single colony of BL21 pLysS containing the pINV plasmid was used to inoculate TBAC followed by overnight incubation at 37 °C. This culture was washed several times to remove the traces of TBAC medium and then used to inoculate 2 liters of M9 medium (medium from Difco supplemented with 2 mM MgSO<sub>4</sub>, 0.36% glucose, 100  $\mu$ M CaCl<sub>2</sub>, 100  $\mu$ g/ml ampicillin, and 34  $\mu$ g/ml chloramphenicol). Incubation was performed at 37 °C under vigorous shaking (240 rpm). When the  $A_{600}$  reached 0.5, 1.5 mM L-lysine, 1.5 mM L-phenylalanine, 1.5 mM L-threonine, 0.8 mM of L-leucine, 0.8 mM L-isoleucine, 0.8 mM L-valine, and 0.5 mM seleno-L-methionine (final concentrations) were added. After 30 min of incubation at 37 °C, 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside was added to the culture. After induction, expression was followed by measuring  $A_{600}$  until a value of 1.7 was reached. Culture lysis was done as described above.

In all cases, the supernatant of the 20,000  $\times$  g centrifugation was filtered (Amicon, 0.2- $\mu$ m pore-sized membrane), and the invertase was then purified in two steps. First, nickel affinity chromatography was performed using buffers containing 50 mM Tris, pH 8, 200 mM NaCl, and 50 and 500 mM imidazole for the washing and elution steps, respectively. Subsequently, the protein was submitted to gel filtration on a Sephadex column (Amersham Biosciences). The fractions containing the protein were pooled and concentrated to 11 mg/ml for the native protein and 8 mg/ml for the SeMet protein over ultrafiltration styrene acrylonitrile membranes (Millipore; cut-off was 30 kDa).

To verify that the N-terminal His-tag did not influence the enzymatic activity of invertase, the hydrolysis of sucrose by the purified protein was monitored. The method employed was adapted from Kidby and Davidson (23) and consisted of the measurement of reducing sugars by ferricyanide. Invertase (200  $\mu$ M) was incubated at 75 °C in 100 mM sodium acetate buffer (pH 5.5) and 120 mM sucrose, *i.e.* exactly the same conditions as those described by Liebl *et al.* (21). One hundred-microliter samples were taken at different times of incubation. The enzymatic reaction was revealed by mixing samples with 1 ml of reagent (1 mM K<sub>3</sub>Fe(CN)<sub>6</sub>, 130 mM Ca<sub>2</sub>O<sub>3</sub>, and 5 mM NaOH) and by heating the samples for 7 min at 95 °C. The activity was monitored by the decrease of  $A_{420}$  as a function of time and led to values (data not shown) very similar to those published by Liebl *et al.* (21).

**Crystallization of Native and SeMet-substituted Proteins**—Crystallization conditions were first investigated using two sparse matrix sampling kits (Molecular Dimensions and Stura Footprint). Optimized crystals of a suitable size were obtained by mixing 15% polyethylene glycol 1000, 150 mM Li<sub>2</sub>SO<sub>4</sub>, and 100 mM sodium citrate at pH 4.2 with 11 mg ml<sup>-1</sup> native protein. Crystals grew within 3 days at 20 °C by the vapor diffusion method. The conditions for the SeMet-substituted protein were 17% polyethylene glycol 1000, 50 mM Li<sub>2</sub>SO<sub>4</sub>, 1% isopropanol, and 100 mM sodium citrate buffer at pH 4.2. Here the drops were composed of 2  $\mu$ l of protein at a concentration of 8 mg ml<sup>-1</sup> with 1  $\mu$ l of reservoir solution. Both the crystals of native and SeMet-substituted protein belonged to space group P2<sub>1</sub> with unit cell parameters  $a = 94.2$  Å,  $b = 113.2$  Å,  $c = 129.6$  Å, and  $\beta = 98.96^\circ$ . The asymmetric unit contains six monomers giving a  $V_M$  value of 2.2 Å Da<sup>-1</sup> and 43% solvent content.

**Data Collection and Phasing**—Crystals were soaked in mother liquor supplemented by 15% glycerol (w/v) before flash freezing in a cryogenic nitrogen stream at 100 K. Diffraction data of native and SeMet-substituted protein crystals, both in space group P2<sub>1</sub>, were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) on beam lines ID14-EH2 and ID29 respectively (Table I). The data on SeMet-substituted crystals were collected at the absorption peak ( $\lambda = 0.97904$  Å) and phased using the SAD method. Forty of the 48 SeMet positions were determined by anomalous Patterson maps using the subroutine XPREP of the program package SHELX5.0 (24). The 40 sites were refined with SHARP (25), and the missing eight positions were found in the residual maps. The 48 selenium positions appeared to be arranged in a manner that suggested the presence of three dimers. Symmetry averaged initial phases (DMMULTI; Ref. 26), using two of

<sup>1</sup> The abbreviations used are: SeMet, selenomethionine; SAD, single wavelength anomalous dispersion.

TABLE I  
 Summary of data collection, phasing, and refinement statistics

Data sets	Native	$\lambda_1$ (Peak)
Wavelength (Å)	0.979	0.97904
High resolution (Å)	2.0	2.2
(Anomalous) completeness (%)	99.9 (99.9) <sup>a</sup>	99.3 (98.6)
Redundancy	3.1 (3.1)	4.2 (4.2)
$I/\sigma(I)$	8.0 (2.1)	6.8 (2.2)
$R_{\text{sym}}^b$	0.077 (0.402)	0.073 (0.211)
Phasing statistics		
Anomalous difference (%)		6.4
Figure of merit (overall)	0.429 (0.853) <sup>c</sup>	
Refinement statistics		
$R_{\text{cryst}}^d$ (%)	17.6	
$R_{\text{free}}^e$ (%)	22.0	
Overall $B$ -factor (Å <sup>2</sup> )	24.25	
Average $B$ -factors of MolA/MolB (Å <sup>2</sup> )	25.94/26.43	
MolD/MolE	26.59/26.54	
MolC/MolF	26.31/26.63	
R.m.s. deviation bond lengths <sup>f</sup> (Å)	0.027	
R.m.s. deviation bond angles <sup>f</sup> (°)	2.24	

<sup>a</sup> Numbers in parentheses indicate values for the highest resolution bin.

<sup>b</sup>  $R_{\text{sym}} = \sum [I_i - \langle I \rangle] / \sum \langle I \rangle$ , where  $i$  is the  $i$ th measurement and  $\langle I \rangle$  is the weighted mean of  $I$ .

<sup>c</sup> Figure of merit value in parentheses is calculated after density modification with the DM program.

<sup>d</sup>  $R_{\text{cryst}} = \sum |F_{\text{obs}}| - |F_{\text{calc}}| / \sum |F_{\text{obs}}|$ .

<sup>e</sup>  $R_{\text{free}}$  is the same as  $R_{\text{cryst}}$  for 5% of the data omitted from refinement totaling 10,694 reflections.

<sup>f</sup> R.m.s. is root mean square.

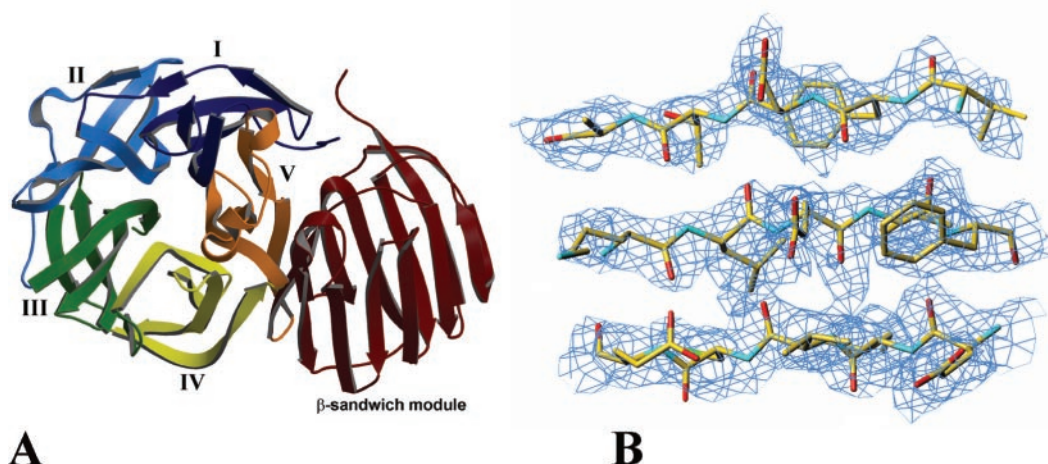


FIG. 1. Overall fold and experimental electron density map of invertase. A, ribbon representation of the monomeric unit of *T. maritima* invertase, highlighting the N-terminal  $\beta$ -propeller module, the five blades (numbered I–V), and the C-terminal  $\beta$ -sandwich module (dark red). B, section of the experimental map after phasing with SHARP (25), solvent-flattening with DMMULTI (26), and non-crystallographic symmetry averaging with RESOLVE (27). The experimental electron density map, contoured at a  $1\sigma$  level, shows three antiparallel  $\beta$ -strands in the  $\beta$ -sandwich module at the C-terminal region of the protein.

the three dimer positions, were subsequently used as input for RESOLVE (27), which automatically constructed initial  $C_\alpha$  tracing of all six monomers present in the asymmetric unit. The density modification step with RESOLVE also produced a  $\sigma_A$ -weighted  $2F_o - F_c$  map of excellent quality into which side chains were built manually with TURBO (28) (Fig. 1B) for one monomer. The relative positioning of all molecules within the asymmetric unit was then performed by molecular replacement (AmoRe; Ref. 29) using the first constructed monomer as search model.

**Structure Refinement**—The structural model of invertase was refined with REFMAC5 (30) with intermittent manual rebuilding and refining of individual B-factors after applying a TLS correction. Water molecules were added with ARP/wARP (30). The final model comprises  $6 \times$  protein residues 1–432 (2,592 amino acids), 12  $\text{SO}_4^{2-}$  ions, one buffer molecule (sodium citrate), and a total of 1,754 water molecules, which led to  $R$  and  $R_{\text{free}}$  values of 17.6 and 22%, respectively. A few residues lacked electron density and were therefore refined with occupations of 0.5. One short surface loop (residues 96–100) was highly disordered and displayed only clear density in one of the six invertase molecules. Ramachandran statistics (PROCHECK) indicated that, for the overall structure of the six molecules present in the asymmetric unit, 87.1% of the atoms are in the most favored region, and 12.6% are in additionally allowed regions. Details of refinement statistics are summarized in Table I.

## RESULTS AND DISCUSSION

**Overall Fold**—The crystal structure of the *T. maritima* invertase (residues 1–432) has been solved by SAD phasing of the SeMet-substituted protein at a maximal resolution of 2 Å. The SeMet-substituted as well as the native crystals belong to space group  $P2_1$  with unit cell parameters  $a = 94.2$  Å,  $b = 113.2$  Å,  $c = 129.6$  Å, and  $\beta = 98.96^\circ$ . The coordinates describing six copies of the invertase polypeptide chain and 1754 water molecules per asymmetric unit were refined to final  $R$ - and  $R_{\text{free}}$ -factors of 17.6 and 22%. One molecule of invertase is composed of two individual modules, namely a five-bladed  $\beta$ -propeller (residues 1–295) catalytic module linked to a C-terminal  $\beta$ -sandwich module (residues 306 to 432) by a 10-residue linker (Fig. 1). The ensemble of six bi-modular molecules arrange into three individual dimers, displaying 2-fold symmetry each. The three dimers are not related by any point group symmetry but by non-symmetrical rotations and translations. The dimer arranges around a pseudo 2-fold axis, bringing the  $\beta$ -sandwich domain of monomer A in contact with the  $\beta$ -propeller domain of

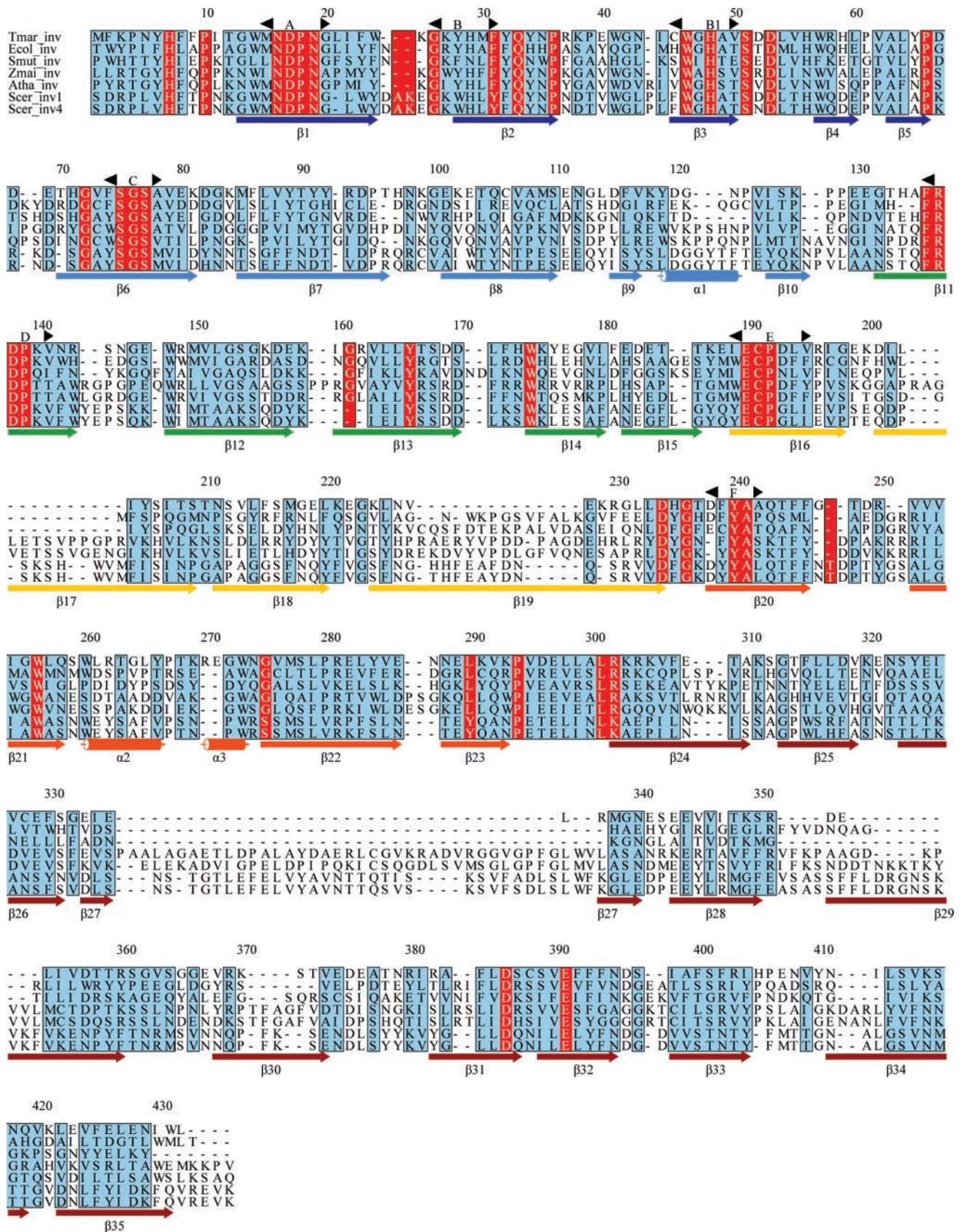


FIG. 2. Sequence alignment of a selection of family GH32 invertases. The sequences are identified as follows: *Tmar\_inv*, *T. maritima* invertase (Swiss-Prot O33833); *Ecol\_inv*, *E. coli* K12 invertase (Swiss-Prot P16553); *Smut\_inv*, *Streptococcus mutans* GS-5 invertase (Swiss-Prot P13522); *Zmai\_inv*, *Zea mays* invertase (Swiss-Prot O81189); *Atha\_inv*, *Arabidopsis thaliana* Landsberg *erecta* (GenBank™ BAA89048.1); *Scer\_inv1*, *Saccharomyces cerevisiae* invertase 1 (Swiss-Prot P10594); and *Scer\_inv4*, *S. cerevisiae* invertase 4 (Swiss-Prot P10596). The boxes shaded in red are strictly conserved residues, whereas the boxes shaded in light blue concern highly similar sequence regions. The sequence numbering and secondary structure elements (the color codes of the secondary structure elements are the same as in Fig. 1) correspond to the sequence of *T. maritima* invertase. The highly conserved motifs A through F, as defined by Pons *et al.* (13), are highlighted by left and right arrows above the sequences. The alignment was produced with ClustalX (46), and the figure was produced with ALSCRIPT (47).

FIG. 3. Close-up view of the catalytic site of *T. maritima* invertase. *A*, the residues surrounding the modeled sucrose molecule are most likely involved in binding and recognition. *B*, a glycerol molecule occupies the sugar binding site in a manner that mimics the presence of hydroxyl groups O4' and O6' of the fructose moiety of sucrose. Single letter amino acid abbreviations are used with position numbers.

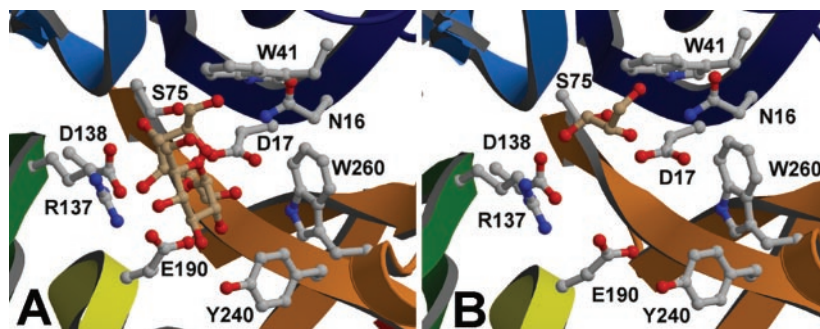


TABLE II  
Hydrogen bonding and close contacts between modeled sucrose and invertase active site residues

Sucrose atom	Invertase residue	Distance
		Å
Fructose O1'	Asp-17-O <sup>δ1</sup>	2.9
	Asp-17-O <sup>δ2</sup>	3.4
	Glu-190-O <sup>ε1</sup>	3.4
	Trp-260-N <sup>ε1</sup>	3.4
Fructose O2'	Asp-17-O <sup>δ2</sup>	3.4
	Asn-16-N <sup>δ2</sup>	3.5
Fructose O3'	Glu-190-O <sup>ε2</sup>	2.9
	Glu-190-O <sup>ε1</sup>	3.6
	Arg-137-N <sup>ε</sup>	2.9
Fructose O4'	Asp-138-O <sup>δ1</sup>	3.3
	Asp-138-O <sup>δ2</sup>	2.5
	Asp-138-O <sup>δ1</sup>	2.6
	Ser-75-N	3.0
	Ser-75-O <sup>γ</sup>	3.4
Fructose C6'	Phe-74-C <sup>γ</sup>	3.9
Fructose O6'	Asn-16-N <sup>ε2</sup>	3.3
	Gln-33-O <sup>ε1</sup>	3.3
Fructose C2'	Trp-41-N <sup>ε1</sup>	2.6
	Asp-17-O <sup>δ2</sup>	3.6
Glucose O1	Glu-190-O <sup>ε1</sup>	3.1
	Glu-190-O <sup>ε2</sup>	3.0
Glucose O2	Glu-190-O <sup>ε1</sup>	2.7
	Tyr-240-O <sup>η</sup>	4.0
Glucose O4	Arg-137-N <sup>η1</sup>	3.2
	Arg-137-N <sup>η2</sup>	3.8

monomer B and *vice versa*. Upon purification, the enzyme had a profile corresponding to a size of ~30 kDa. Nonspecific interaction of the enzyme with the Sephadex column may explain this elution behavior. The same behavior has already been observed previously (21), and, therefore, the oligomeric state cannot be determined by this method. Preliminary investigations by dynamic light scattering indicated that the *T. maritima* invertase is a monomer in solution (data not shown). Several oligomeric states have been reported for invertases of various sources (31, 32). Yeast invertase displays a dimeric substructure that may form even larger oligomers upon mannose binding (33). The overall monomer structure of *T. maritima* invertase has an elliptical shape with approximate dimensions of 63 × 43 × 45 Å with a negatively charged surface depression at the center of the β-propeller.

The clearly defined electron density revealed two amino acids in conflict with the GenBank™ sequence (A108 → V108 and V179 → A179). Therefore, the nucleotide sequence was checked twice (amplification from genomic DNA and the expression clone), and the two single base differences (C323 → T332 and T536 → C536) were only detected for the expression clone. As a consequence, these mismatches are attributed to misincorporation by the polymerase *Pfx*. Nevertheless, activity tests (see "Experimental Procedures") indicated that these mutations do not affect the enzymatic activity.

A five-bladed β-propeller structure has first been reported for tachylectin (15) and was found more recently for the en-

zymes α-L-arabinanase (16) and levansucrase (14) of the glycoside hydrolase families GH43 and GH68, respectively. Highly similar to the families GH43 and GH68 structures, the five β-sheets of invertase, labeled I–V (Fig. 1), adopt the classical "W" topology of four antiparallel β-strands. The N-terminal second strand lines the central cavity, and the C-terminal last strand is at the periphery, to which the β-sandwich module is connected by a short linker. Interestingly, and in contrast to levansucrase and α-L-arabinanase, the five bladed β-propeller of invertase does possess the short "molecular Velcro" that is typical of six- and seven-bladed β-propellers (15, 34, 35). The N-terminal first strand forms the outermost β-strand of the C-terminal blade V; however, only one hydrogen bond is formed across the sheet (Phe-8 O–Met-277 N, 2.88 Å). A similar short Velcro has also been observed in the six bladed β-propeller of *Vibrio cholerae* sialidase (36). As in all β-propeller structures, the β-strands forming the blades are strongly twisted, giving an angle of ~90° between the first and last β-strand of a blade. Insertions are common in this type of β-propellers, and, likewise, short stretches of <sub>3</sub><sub>10</sub>-helices are found inserted between several individual β-strands of the structure described here. They are, however, less extended than in the GH68 levansucrase, and from this perspective the β-propeller of invertase resembles more that of GH43 α-L-arabinanase.

**The Catalytic Active Site**—The catalytic active site is positioned at one end of the cavity at the center of the β-propeller with a funnel-like opening toward the molecular surface. It clearly has a pocket topology, which is fully consistent with the strict exo mode of action of the enzyme on the fructose polymer inulin (21). The three carboxylate groups of two aspartate (Asp-17 and Asp-138) residues and one glutamate (Glu-190) residue point to the center of the depression and generate a high negative charge at the active site. Reddy and Maley have shown that Asp-23 in yeast invertase (Asp-17 in *T. maritima* invertase) is the catalytic nucleophile (8), whereas Glu-204 (here Glu-190) is the general acid/base (9). In addition to the two regions containing the catalytic machinery, multiple sequence alignments of the GH32 family (Fig. 2) have revealed a number of other highly conserved amino acid stretches (13, 37). The inspection of the three-dimensional structure allows us to define possible roles for these highly conserved residues. For the family GH68 levansucrase, the sucrose complex of an inactive mutant has also been reported (14). Because the catalytic modules of invertase and levansucrase are structurally related, the superimposition of invertase with the sucrose-containing complex of levansucrase (PDB identification code 1PT2) allows us, by similarity, to infer the position of a sucrose molecule and model it in the active site of invertase (Fig. 3A). The crystal structure of invertase revealed a glycerol molecule, present in the substrate binding site, that mimicked the O4' and O6' hydroxyl groups of the substrate fructose moiety (Fig. 3B), and this helped define the precise position of the modeled sucrose molecule. This model shows that the second strictly conserved aspartate residue in motif D (for motif definitions see Ref. 13

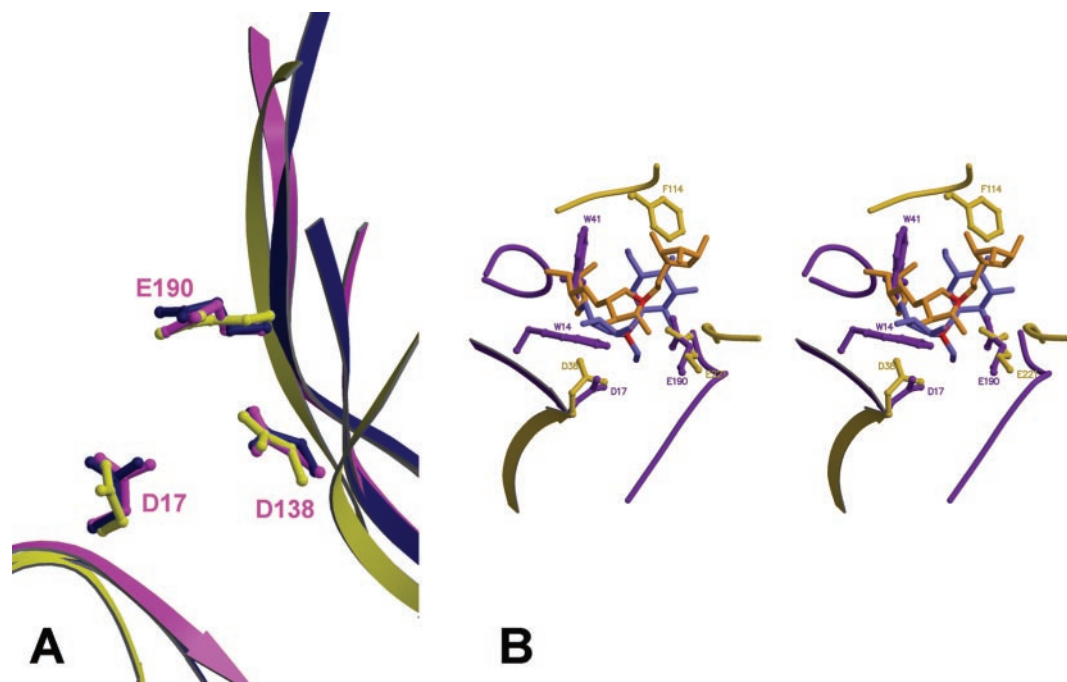
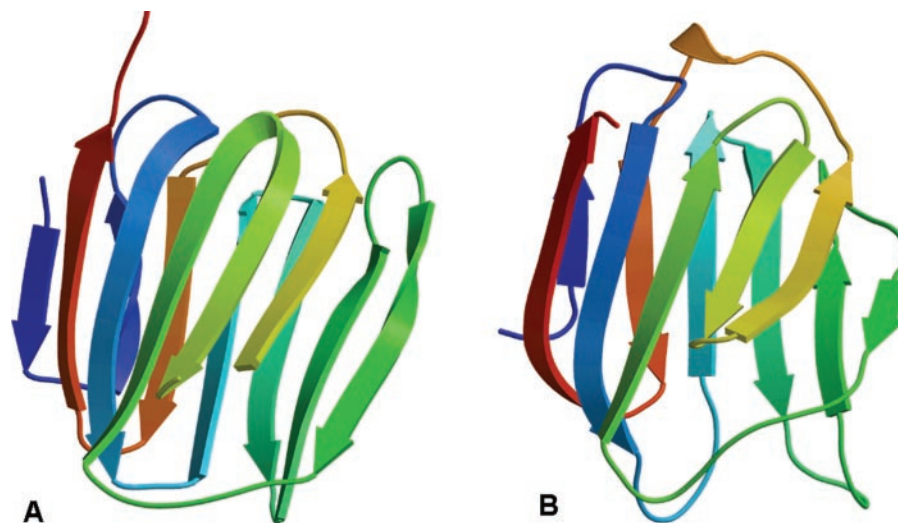


FIG. 4. **Structural comparison of families GH32, GH68, and GH43.** *A*, structural superimposition of the three strictly conserved residues in the catalytic sites of *T. maritima* invertase (magenta), *Bacillus subtilis* levansucrase (dark blue) and *Cellvibrio japonicus*  $\alpha$ -L-arabinanase (yellow). *B*, stereographic view of the superimposed catalytic active sites of  $\alpha$ -L-arabinanase (yellow) in complex with arabinotriose (orange; Protein Data Bank identification 1GYE), and invertase (dark purple) in complex with the modeled sucrose molecule (blue). The different binding modes of the two enzymes lead to a different position of the glycosidic bond with respect to the catalytic machinery. The anomeric carbons at the point of cleavage of both substrate molecules are colored red. The loops, including residues Trp-41 and Trp-14, which define the  $-1$  subsite in invertase, are either not present or are displaced in  $\alpha$ -L-arabinanase. In contrast, the loop containing Phe-114, which encloses the substrate in the binding cleft in  $\alpha$ -L-arabinanase, is absent in invertase. Single letter amino acid abbreviations are used with position numbers.

FIG. 5. **The C-terminal  $\beta$ -sandwich module.** *A*, ribbon representation of residues 306–432 of *T. maritima* invertase displaying the  $\beta$ -sandwich fold with colors ranging from blue at the N-terminal end to red at the C-terminal end. *B*, comparison of *A* to the structure of human galectin-3 (Protein Data Bank identification 1A3K) in approximately the same orientation, highlighting the similarity of the two structures.



and Fig. 2), Asp-138 in *T. maritima* invertase, forms hydrogen bonds to O3 and O4 of the fructose unit, whereas the neighboring Arg-139 is hydrogen-bonded to the glucose O4. Apparently, the pair of strictly conserved residues, “RD,” binds to characteristic hydroxyl groups of the substrate and, therefore, most likely plays a crucial role in substrate binding and recognition. Interestingly, the enzymes of family GH68, which hydrolyze the same substrates, also have the highly conserved motif “RDP,” whereas GH43 and GH62, which have a structurally related fold but hydrolyze different substrates, do not possess this motif and only have the aspartate residue in the same position. The highly conserved motif designated A by Pons *et al.* (13) contains the nucleophile Asp-17 and the preceding Asn-16, which forms a hydrogen bond to the O6 group of fructose, whereas the sequence regions designed B and B1 appear to be

structurally important, because the conserved aromatic residues are involved in hydrophobic interactions in the face-to-face packing of blades I and V and are not in the catalytic site. However the side chain of Trp-41, located between motifs B and B1, points into the active site and is most probably part of the aglycone binding pocket. Motif C contains residues involved in substrate binding such as Phe-74, which borders the fructose binding pocket, and Ser-75, which forms hydrogen bonds to the O4 hydroxyl of fructose (3.5 Å) and to catalytic nucleophile (2.9 Å). The sequence region E contains the general acid/base Glu-190 (3.1 Å from the glycosidic oxygen) and Cys-191, both located in the heart of the active site. This conserved cysteine is most probably important for transition state stabilization and/or the catalytic residue microenvironment, because it forms hydrogen bonds to Asp-17 (3.5 Å) and Asp-138 (3.6 Å). It

is interesting to note that enzymes of family GH-68 have an arginine replacing this cysteine, although they cleave highly similar substrates. The importance of these differences for binding, recognition, and catalysis will be investigated in the future by a study of inactivated invertase mutants in complex with oligosaccharides. See Table II for a comparison of hydrogen bonding and close contacts between modeled sucrose and invertase active site residues.

**Structural Relationship to Families GH68 and GH43 Five-bladed  $\beta$ -Propellers**—Based on detailed sequence analyses, a structural relationship between families GH32, GH43, GH62, and GH68 has been predicted (13, 17). The common five-bladed  $\beta$ -propeller fold, recently revealed by the structure determinations of members of family GH68 (14) and GH43 (16), confirmed this structural relationship. The crystal structure of invertase now proves that the catalytic modules of family GH32 enzymes also display the same five-bladed  $\beta$ -propeller fold. The superimposition of the catalytic module of invertase onto the two other enzymes leads to an overall root mean square deviation of 3.24 Å for 306 C $\alpha$  atoms in the case of the family GH43  $\alpha$ -L-arabinanase and 3 Å for 359 C $\alpha$  atoms in the case of the family GH68 levansucrase. Whereas levansucrase and invertase both retain the anomeric configuration at the site of cleavage,  $\alpha$ -L-arabinanase is an inverting enzyme (18–20). The most widely accepted (and documented) view of the difference between the catalytic machineries of retaining and inverting glycosidases is that, in the former, the two catalytic amino acids are  $\sim$ 5.5 Å apart, and in the latter this distance is generally  $\sim$ 9 Å, with the exception of  $\beta$ -helical enzymes such as polygalacturonase or  $\iota$ -carrageenase (38, 39). Remarkably, the three invariant amino acids Asp-17, Asp-138, and Glu-190 in GH32, defined as the catalytic residues in each of the families GH32, GH68, and GH43, superimpose rather well in all three enzyme structures (Fig. 4A), showing that the relatedness is not solely with the fold but also with the catalytic machinery. The structural superposition shows that there is no difference in the distances of the catalytic residues relative to each other, as has generally been observed in inverting *versus* retaining glycoside hydrolases (10, 12, 40, 41). Instead, it is the difference in the binding position of the sugar in the  $-1$  subsite (subsite nomenclature of Davies *et al.*; Ref. 48) that makes the difference in the catalytic mechanism of invertase and levanase on the one hand and  $\alpha$ -L-arabinanase on the other. The arrangement of the loops and residues surrounding the catalytic machinery in  $\alpha$ -L-arabinanase is such that the arabinosyl moiety in the  $-1$  subsite is bound in a position almost perpendicular to the fructofuranosyl moiety in invertase and levanase. Consequent to this different binding, the nucleophilic residues are only  $\sim$ 3.6 Å from the sugar C1 atom in invertase and levanase (14), whereas the distance C1–Asp-38 in  $\alpha$ -L-arabinanase is 6 Å, leaving room for a water molecule (16) (Fig. 4B). This different binding mode of the “glycone” part of the substrate fully explains the opposite stereochemical outcome of the reaction, despite a perfectly superimposable catalytic machinery.

**The  $\beta$ -Sandwich Module**—The C-terminal residues (from 306 to 432) of *T. maritima* invertase compose an individually folded  $\beta$ -sandwich consisting of two sheets of six  $\beta$ -strands. This module is connected to the catalytic module via a short, 10-residue-long linker region that is wrapped around the  $\beta$ -sandwich. Contrary to the catalytic module, which can be readily aligned with all other members of glycosidase family GH32, BLAST searches conducted with the C-terminal module of *T. maritima* invertase did not reveal a statistically significant sequence similarity with the equivalent regions in other family GH32 proteins. To detect possible relatedness beyond the detection level of BLAST, we have removed the easily

identifiable catalytic domain region in all complete family GH32 members and constructed a sequence library with the remaining C-terminal regions. PSI-BLAST searches conducted starting with the C-terminal region of plant or fungal or bacterial family GH32 members picked the *T. maritima* C-terminal domain after a few iterations, indicating that all GH32 family members will also be appended to a  $\beta$ -sandwich domain, such as that of *T. maritima* invertase.

The alignment of this module with the programs DALI (42) and 3D-PSSM (43) onto other  $\beta$ -sandwich structures revealed structural similarities with the  $\beta$ -sandwich in galectins, the Charcot-Leyden crystal protein, carbohydrate binding modules (CBMs), and other more distant proteins like lectins and exotoxin A. The highest similarity is observed with the human galectin-3 (Protein Data Bank identification, 1A3K; DALI Z-score, 10.9; root mean square deviation for 127 C $\alpha$  is 2.4 Å) (Fig. 5) and with the Charcot-Leyden crystal protein (Protein Data Bank identification, 1CLC; DALI Z-score, 10.7; root mean square deviation for 132 C $\alpha$  is 2.6 Å), which has recently been found to be a maltose binding galectin (44). It is interesting to note that six-bladed  $\beta$ -propeller glycosidases such as *Micromonospora viridifaciens* and *V. cholerae* sialidases have also been found appended with lectin-like domains (36, 45).

It has been observed that extracellular yeast invertase, a functionally active homodimer in solution, acquires maltose to self-assemble into higher oligomers upon transport and secretion (33). It is therefore tempting to postulate that the supplementary  $\beta$ -sandwich module of yeast invertase plays the role of a carbohydrate recognition domain involved in the higher oligomer formation. The distant similarity of the C-terminal module of *T. maritima* invertase, compared with the other members of the GH32 family, suggests that this module has perhaps lost this function in *T. maritima* invertase. Alternatively, this module might have evolved in *T. maritima* invertase to preserve stability at high temperature, even if the ancestral function of it has been lost. Proteins from hyperthermophilic organisms frequently adopt a modular as well as a multimeric structure. These two complementary features are thought to increase stability at high temperature by masking weak regions at the surface of the protein.

**Acknowledgments**—We thank Dr. Wolfgang Liebl (Georg-August-Universität, Göttingen, Germany) for his generous gift of *T. maritima* genomic DNA. We also thank the staff of the European Synchrotron Radiation sources for the provision of beam time and for technical assistance at the beamlines ID29 and ID-14 EH2. We thank Dr. J. Allouch and Dr. A. Gruez for helpful discussions.

## REFERENCES

- O'Sullivan, C., and Tompson, F. W. (1890) *J. Chem. Soc.* **57**, 854–870
- Koshland, D. E., Jr., and Stein, S. S. (1954) *J. Biol. Chem.* **208**, 139–148
- Sturm, A. (1999) *Plant Physiol.* **121**, 1–8
- Sturm, A., and Tang, G. Q. (1999) *Trends Plant Sci.* **4**, 401–407
- Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) *Science* **299**, 2074–2076
- Schell, M. A., Karmirantzou, M., Snel, B., Vilanova, D., Berger, B., Pessi, G., Zwahlen, M. C., Desiere, F., Bork, P., Delley, M., Pridmore, R. D., and Arigoni, F. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14422–14427
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316
- Reddy, V. A., and Maley, F. (1990) *J. Biol. Chem.* **265**, 10817–10820
- Reddy, A., and Maley, F. (1996) *J. Biol. Chem.* **271**, 13953–13957
- Davies, G., and Henrissat, B. (1995) *Structure* **3**, 853–859
- Gebler, J., Gilkes, N. R., Claessens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A., and Withers, S. G. (1992) *J. Biol. Chem.* **267**, 12559–12561
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7090–7094
- Pons, T., Olmea, O., Chinea, G., Beldarrain, A., Marquez, G., Acosta, N., Rodriguez, L., and Valencia, A. (1998) *Proteins* **33**, 383–395
- Meng, G., and Fütterer, K. (2003) *Nat. Struct. Biol.* **10**, 935–941
- Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R., and Bode, W. (1999) *EMBO J.* **18**, 2313–2322
- Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J., and Davies, G. J. (2002) *Nat. Struct. Biol.* **9**, 665–668
- Naumoff, D. G. (2001) *Proteins* **42**, 66–76



18. Pitson, S. M., Voragen, A. G., and Beldman, G. (1996) *FEBS Lett.* **398**, 7–11
19. Braun, C., Meinke, A., Ziser, L., and Withers, S. G. (1993) *Anal. Biochem.* **212**, 259–262
20. Kersters-Hilderson, H., Claeysens, M., van Doorslaer, E., and de Bruyne, C. K. (1976) *Carbohydr. Res.* **47**, 269–273
21. Liebl, W., Brem, D., and Gotschlich, A. (1998) *Appl. Microbiol. Biotechnol.* **50**, 55–64
22. Invitrogen (2002) *Gateway™ Technology, a Universal Technology to Clone DNA Sequences for Functional Analysis and Expression in Multiple Systems, Version C*, Invitrogen, Carlsbad, CA
23. Kidby, D. K., and Davidson, D. J. (1973) *Anal. Biochem.* **55**, 312–325
24. Sheldrick, G. M. (1990) *Acta Crystallogr. Sect. A* **46**, 467–473
25. de La Fortelle, E., and Bricogne, G. (1997) *Methods Enzymol.* **276**, 472–494
26. Cowtan, K., and Main, P. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54**, 487–493
27. Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1937–1940
28. Roussel, A., and Cambillau, C. (1991) *Silicon Graphics Geometry Partners Directory*, Vol. 88, Silicon Graphics, Mountain View, CA
29. Navaza, J. (1994) *Acta Crystallogr. Sect. A* **50**, 157–163
30. Computational Collaborative Project 4 (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50**, 760–763
31. Lee, H. S., and Sturm, A. (1996) *Plant Physiol.* **112**, 1513–1522
32. Li, Y., and Ferenci, T. (1996) *Microbiology* **142**, 1651–1657
33. Reddy, A. V., MacColl, R., and Maley, F. (1990) *Biochemistry* **29**, 2482–2487
34. Burmeister, W. P., Ruigrok, R. W., and Cusack, S. (1992) *EMBO J.* **11**, 49–56
35. Jawad, Z., and Paoli, M. (2002) *Structure* **10**, 447–454
36. Crennell, S., Garman, E., Laver, G., Vimr, E., and Taylor, G. (1994) *Structure* **2**, 535–544
37. Gunasekaran, P., Karunakaran, T., Cami, B., Mukundan, A. G., Preziosi, L., and Baratti, J. (1990) *J. Bacteriol.* **172**, 6727–6735
38. Cho, S. W., Lee, S., and Shin, W. (2001) *J. Mol. Biol.* **311**, 863–878
39. Michel, G., Chantalat, L., Fanchon, E., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) *J. Biol. Chem.* **276**, 40202–40209
40. Rye, C. S., and Withers, S. G. (2000) *Curr. Opin. Chem. Biol.* **4**, 573–580
41. Vasella, A., Davies, G. J., and Böhm, M. (2002) *Curr. Opin. Struct. Biol.* **6**, 619–629
42. Holm, L., and Sander, C. (1998) *Nucleic Acids Res.* **26**, 316–319
43. Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2000) *J. Mol. Biol.* **299**, 499–520
44. Ackerman, S. J., Liu, L., Kwatia, M. A., Savage, M. P., Leonidas, D. D., Swaminathan, G. J., and Acharya, K. R. (2002) *J. Biol. Chem.* **277**, 14859–14868
45. Gaskell, A., Crennell, S., and Taylor, G. (1995) *Structure* **3**, 1197–1205
46. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882
47. Barton, G. J. (1993) *Protein Eng.* **6**, 37–40
48. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem. J.* **321**, 557–559