Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose

François ALBERTO¹, Emmanuelle JORDI, Bernard HENRISSAT and Mirjam CZJZEK²–³

Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Universités Aix-Marseille I & II, Case 932, 163 Avenue de Luminy, 13288 Marseille Cedex 9, France

_Thermotoga maritima_ invertase (β-fructosidase), a member of the glycoside hydrolase family GH-32, readily releases β-D-fructose from sucrose, raffinose and fructan polymers such as inulin. These carbohydrates represent major carbon and energy sources for prokaryotes and eukaryotes. The invertase cleaves β-fructofuranosidic linkages by a double-displacement mechanism, which involves a nucleophilic aspartate and a catalytic glutamic acid acting as a general acid/base. The three-dimensional structure of invertase shows a bimodular enzyme with a five-bladed β-propeller catalytic domain linked to a β-sandwich of unknown function. In the present study we report the crystal structure of the inactivated invertase in interaction with the natural substrate molecule α-D-galactopyranosyl-(1,6)-α-D-glucopyranosyl-β-D-fructofuranoside (raffinose) at 1.87 Å (1 Å = 0.1 nm) resolution. The structural analysis of the complex reveals the presence of three binding-sites, which explains why *T. maritima* invertase exhibits a higher affinity for raffinose than sucrose, but a lower catalytic efficiency with raffinose as substrate than with sucrose.

Key words: crystal structure, exo-glycosidase, β-fructosidase, glycosidase GH32 family, invertase, retaining mechanism.

INTRODUCTION

Many bacteria and fungi, and approx. 15 % of flowering plants synthesize fructans, which are fructose oligosaccharides and polysaccharides synthesized from sucrose, the end product of photosynthesis [1]. The breakdown products of these storage carbohydrate reserves are used by many organisms as a source of carbon and energy [2,3]. Over 50 years ago Koshland and Stein [5] showed that yeast invertase-mediated hydrolysis of sucrose proceeds with overall retention of the β configuration of fructose. Subsequently, the catalytic residues of yeast invertase have been identified as aspartic acid, D23, acting as the nucleophile [6], and glutamic acid, E204, which serves as the general catalytic acid/base [7] in the double-displacement mechanism. Recently, the crystal structures of three GH32 enzymes, *Thermotoga maritima* (Tm) invertase [8], exo-inulinate from *Aspergillus awamori* [9] and a plant fructan 1-exohydrolase from *Cichorium intybus* [10] have all been reported. The three-dimensional structures of these three enzymes revealed a similar bimodular arrangement with a catalytic-domain folding in a rather unusual five-bladed β-propeller linked to a C-terminal β-sandwich domain of unknown function. So far, the five-bladed β-propeller fold has only been reported for tachylectin [11] and for two other glycoside hydrolase families, GH43 [12] and GH68 [13]. This structural resemblance has reinforced the establishment of an evolutionary relationship between retaining (GH32 and GH68) [13] and inverting (GH43) enzymes [8,8a].

The exo-inulinate from *A. awamori* (Aa exo-inulinate) was crystallized in the presence of fructose and thus permitted the identification of residues that constitute the fructose-binding pocket of the active site [9]. To complete the structural analysis of the active sites of GH32 family enzymes, we needed to obtain and characterize an enzyme–substrate complex that incorporated and extended beyond the cleavage site. For this purpose, the enzyme was inactivated by site-directed mutagenesis [14] and was co-crystallized or saturated with the substrate. Here we describe the crystal structure of a *Tm* invertase mutant E190D (inv-E190D) in complex with the trisaccharide raffinose [α-D-galactopyranosyl-(1,6)-α-D-glucopyranosyl-β-D-fructofuranoside; Figure 1A]. This substrate molecule is readily cleaved by the *Tm* invertase that can also use sucrose and fructofuranose polymers such as inulin as substrates [15]. In addition we discuss the structure of the enzyme–substrate complex in view of the enzyme’s specificity and we describe the detailed structure of the three binding-sites that comprise the active-site pocket of *Tm* invertase.

MATERIAL AND METHODS

Mutagenesis, protein expression and purification of the inactivated mutant inv-E190A

The plasmid pINV that was obtained by cloning the native invertase gene (*bfrA*) into the N-terminal His-tag encoding expression plasmid pDET17 [8] (Invitrogen) was subsequently used as a template for mutagenesis using the QuickChange Mutagenesis Kit (Stratogene). Amplifications were carried out using the primers E190AF (5′-CCACAAAAGAAATAGATT-3′) and E190AR (5′-CTTACAGATCGGG-ACACGCATTATTCTTTGTG-3′) to mutate Glu⁸⁰ into alanine (substitutions are underlined), and the primers E190DF (5′-CCACAAAGAAATAGATTCTTTGAGG-3′) and E190DR (5′-CTCAACAAGATCGGGACAATCTTTCTTTGTG-3′) to mutate Glu⁸⁰ into aspartate. Protein expression and purification were carried out essentially as described previously.

Abbreviations used: Aa, *Aspergillus awamori*; NCS, non-crystallographic symmetry; R.M.S.D., root mean square deviation; TBAC, Terrific Broth with ampicillin and chloramphenicol; *Tm*, *Thermotoga maritima*.

¹ Present address: INRA-UMR A408, Centre de Recherches d’Avignon Domaine Saint Paul, Site Agroparc, 84914 Avignon cedex 9, France.
² Present address: Station Biologique de Roscoff, Végétaux Marins et Biomolécules, UMR7139-CNRS-UPMC, Place George Teissier, BP74, 29682 Roscoff, France.
³ To whom correspondence should be addressed (email czjzek@sb-roscoff.fr).
for the native invertase [8]. Briefly, a single colony of Escherichia coli, BL21pLysS, containing the pINV-E190A plasmid was used to inoculate TBAC (Terrific Broth supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol). The culture was incubated at 37°C and was then used to inoculate 3 litres of TBAC. Induction was performed by adding 0.5 mM isopropyl-1-thio-D-galactopyranoside when D$_{max}$ reached 0.8 nm. After incubation at 37°C for another 4 h the cultures were pelleted and then submitted to the lysis procedure. After lysis the supernatant containing the soluble proteins was separated from the pellet by centrifugation (20000 g for 30 min at 4°C) and then filtered before two-step purification by nickel affinity chromatography and gel filtration on a Sephadex column (Amersham Biosciences). The fractions containing the protein were pooled and concentrated to 11 mg/ml on ultrafiltration styrene acrylonitrile membranes (Millipore).

### Activity assays

Invertase activity was determined following the method of Kidby and Davidson [16] by measuring the release of reducing sugars by the ferricyanide method. Invertase was incubated at 75°C in 100 mM sodium acetate buffer (pH 5.5) and 120 mM sucrose for 30 min at 4°C. The enzymatic reaction was revealed by mixing 1 ml of reaction buffer [1 mM K$_3$Fe(CN)$_6$, 130 mM sodium acetate and 5 mM NaOH] and by heating the samples for 7 min at 95°C. The activity was monitored by the decrease in A$_{420}$ as a function of time.

### Table 1  Summary of data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>invE190–raffinose complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray source</td>
<td>ESRF (ID14-EH2)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.934</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>32.1–1.87 (1.94–1.87)</td>
</tr>
<tr>
<td>Space group</td>
<td>P 2$_1$</td>
</tr>
<tr>
<td>R$_{int}$ (%)</td>
<td>0.040 (0.18)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>97.7 (91.7)</td>
</tr>
<tr>
<td>I/σ$_1$</td>
<td>16.2 (5.5)</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>220 454 (20270)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Refinement statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>32.1–1.87 (1.94–1.87)</td>
</tr>
<tr>
<td>R$_{int}$ (%)</td>
<td>19.8 (21.7)</td>
</tr>
<tr>
<td>R$_{free}$ (%)</td>
<td>22.9 (26.1)</td>
</tr>
<tr>
<td>Overall B-factor (Å$^2$)</td>
<td>25.5</td>
</tr>
<tr>
<td>R.M.S.D. in bond lengths (Å)</td>
<td>0.006</td>
</tr>
<tr>
<td>R.M.S.D. in bond angles (°)</td>
<td>1.103</td>
</tr>
<tr>
<td>Ramachandran plot</td>
<td></td>
</tr>
<tr>
<td>Most favoured regions (%)</td>
<td>88.3</td>
</tr>
<tr>
<td>Additionally allowed regions (%)</td>
<td>10.7</td>
</tr>
</tbody>
</table>

### Protein crystallization and X-ray diffraction data collection

The crystallization conditions for inv-E190D and inv-E190A were similar to those of the native protein [8]. Crystals of suitable size for X-ray diffraction experiments were grown in 15–17% PEG [poly(ethylene glycol) 1000], 150 mM Li$_2$SO$_4$, and 100 mM sodium citrate at pH 4.2, by the vapour-diffusion method using hanging drops. Co-crystallization experiments were set up under the same crystallization conditions except that the protein samples (100 µl of 11 mg/ml of protein) were mixed with 5 mM substrate solutions. All crystals that were tested failed to display interpretable electron density. Crystals of inv-E190D and inv-E190A were then soaked for 2 min in mother liquor saturated with d-(+)-raffinose (Sigma–Aldrich) and supplemented with 15% glycerol and then immediately flash frozen in a cryogenic nitrogen stream at 100 K. Trials with other substrate molecules, such as sucrose, kestose or nystose led to diffracting crystals but at lower resolution and none displayed interpretable electron density in the active-site pocket (results not shown).

Data collection was carried out at the ESRF [(European Synchrotron Radiation Facility), Grenoble, France] on beamline ID14-EH2. The only successful data set that gave a substrate–enzyme complex structure was collected from an inv-E190D crystal that had been soaked in raffinose as indicated above. The data collection statistics are reported in Table 1. The crystals of inv-E190D bound with raffinose belonged to space group P2$_1$, with unit cell parameters a = 94.5 Å (1 Å = 0.1 nm), b = 114.7 Å, c = 130.0 Å and β = 98.90°. The asymmetric unit contains six monomers giving a V$_{syn}$ value of 2.2 Å$^3$ Da$^{-1}$ and a 43% solvent content [17].

### Structural refinement

The structural model of inv-E190D was refined with REFMACS [18], using NCS (non-crystallographic symmetry) restraints and starting with the co-ordinates of the native invertase structure. The refinement cycles were alternated with intermittent manual rebuilding with TURBO-FRODO [19]. Individual B-factors were refined after the application of a TLS (translation, libration and screw-rotation) correction [20]. Water molecules were added by ARP/wARP (automated refinement procedure/weighted ARP) [21]. The final model is composed of six copies of the protein residues 2–432 (2592 amino acids), four SO$_4^{2–}$ ions, one buffer
molecule (sodium citrate), six raffinose molecules and a total of 1890 water molecules, which led to R and R\text{free} [22] values of 19.8 and 22.9% respectively, Ramachandran statistics (PROCHECK [23]) indicated that for the overall structure of the six molecules present in the asymmetric unit, 88.3% of the atoms are in the most favoured region and 10.7% are in additionally allowed regions. Details of refinement statistics are summarized in Table 1.

PDB accession numbers

The co-ordinates and structural factors have been deposited in the PDB under accession numbers 1W2T and 1W2Tsf for the inv-E190D mutant protein in complex with raffinose.

RESULTS AND DISCUSSION

Quality of the model and overall structure of the inactivated invertase

The structural analysis of glycoside hydrolases in complex with substrate molecules requires their inactivation before crystallization. For this purpose, two active site mutants of Tm invertase, inv-E190D and inv-E190A, were produced and tested for their activity towards sucrose (Figure 1B). Inv-E190D retains 14% of native invertase activity, whereas no activity was detectable for inv-E190A. Both Tm mutants crystallized with the same space group, P2₁, as the native protein, and in each case the asymmetric unit contained six independent invertase molecules. Both inv-E190D and inv-E190A mutants were used for soaking and cocrystallization experiments but only inv-E190D, in complex with raffinose, gave suitable high-resolution crystallographic data and is described below. Apparently, the glutamic acid to alanine substitution of the Tm invertase lacks a loop of 11 residues, which form the catalytic machinery, Asp17 and Glu190 in the case of Aa exo-inulinase, but the Tm invertase contains only one loop-insertion at the surface close to the active-site cleft, namely from residues Gly155 to Gly162 (Ala208 to His211 in Aa exo-inulinase), but none of these residues bind to the substrate molecule trapped in the active site of Tm invertase.

Three binding subsites in the active-site pocket of Tm invertase

Tm invertase is a typical exo-enzyme [24] and displays a 30 Å deep and 18 Å wide active-site pocket. The identity of the two residues that form the catalytic machinery, Asp17 and Glu190 in the case of Tm invertase, is now well established [6,7,9] and the putative roles for the different conserved stretches [25] in the sequences of GH32 family enzymes have recently been described [8,9]. The present structural analysis of the inactivated invertase–raffinose complex allows a much improved view of the substrate-binding site, in particular across the cleavage site. A single raffinose molecule is bound between the blades of each β-propeller module, defining three substrate binding subsites — I, +1 and +2 (Figure 3; subsite nomenclature adapted from Davies et al. [26] with cleavage taking place between —1 and +1, and the β-fructofuranosyl unit undergoing catalysis placed at subsite —1). For all six independent molecules within the asymmetric unit, the active sites and binding patterns are essentially the same and only one is described below.

The overall structure of the bound raffinose molecule is highly similar to that of raffinose in the crystalline state [27,28]. The only difference (Figure 3c) is a change of the torsion angle around the glycosidic linkage that undergoes catalysis (C2′–O1 in Figure 3c) by approx. 6°. This change is certainly induced by the replacement of the hydrogen-bonding pattern in the raffinose pentahydrate crystal structure by hydrogen bonds formed with the enzyme. At this stage the importance of this torsion for the catalytic mechanism is unclear. The superimposition of each of the three monoosaccharide units of the raffinose crystal structure on the enzyme-bound raffinose molecule did not reveal any notable ring distortion. This is not surprising, since the fructofuranosyl moiety does not require extensive distortion to place the glycosidic linkage in axial orientation for direct in-line nucleophilic attack by the enzyme [29].

The present complex of Tm inv-E190D with raffinose also supports the model that places sucrose in the active site of Tm invertase [8] that we had inferred from the presence of a glycerol molecule and from superimposition of the levansucrase–sucrose complex [13]. The disaccharide substrate modelled for Tm invertase perfectly superimposes on to the β-d-fructofuranosyl-α-L-galactopyranosyl-(1 → 6) moiety of raffinose, demonstrating that active-site residues such as Arg137, Glu188 or Thr260 (as described below) are responsible for the torsion angle geometry of the trapped substrate molecules.

The fructose moiety is located in subsite —1 at the deepest end of the active-site pocket, and is held in place by numerous amino acid residues, which form hydrogen bonds to its hydroxyl groups (Table 2). Of these, Trp290, Trp31, Gln33, Asn16 and Asp158 are clearly responsible for substrate specificity and recognition of the fructose moiety that is to be cleaved: Trp290 forms a hydrogen bond to O1′ of the fructose moiety, Trp31, Gln16 and Asn16 are within hydrogen bond distance of O5′, and Asp158 forms two bridging hydrogen bonds with O3′ and O4′ respectively (Figure 4). The same hydrogen bonding pattern has also recently been described.
Figure 2  Ribbon representation of the superimposition of the three three-dimensional structures of GH32 family enzymes

*Tm* invertase is coloured in red, *Ci* fructan 1-exohydrolase in blue and *Aa* inolinase in yellow. The arrows highlight specific loop insertions that most probably are responsible for modifying the substrate specificities of the three enzymes. Figures 2, 3a and 4 were produced with Molscript [30,31] and rendered with Raster3D.

Figure 3  Localization and structural arrangement of raffinose bound in the active-site pocket in *Tm* invertase

(a) Ribbon representation of the active-site pocket in *Tm* invertase. The raffinose molecule trapped in the active site is shown in coloured atoms as a stick-representation. (Oxygen atoms are coloured in red). Catalytic residues are shown in gold, aromatic residues in dark blue and other residues in green. Blue balls are water molecules. The substrate-binding subsites are annotated from $-1$ to $+2$. (b) Electron density calculated after the final refinement step for the *Tm* invertase inv-E190D in complex. The Fourier map (2Fo–Fc) is contoured at 1σ showing the trapped substrate molecule present in the active site of inv-E190D. (c) Stick representation of raffinose in complex with *Tm inv-E190D* (standard atom type colours) superimposed on to the crystal structure of raffinose pentahydrate (in blue). Figures 3(b) and 3(c) were prepared with TURBO-FRODO [19].

for fructose bound with the *Aa* exo-inulinase [9] and most of these residues are conserved throughout GH32 family enzymes.

The central α-linked glucose unit of raffinose in subsite $+1$ is surrounded by three conserved aromatic residues present throughout the GH32 family, these are Trp$^{65}$, Phe$^{74}$ and Trp$^{260}$. Nevertheless, none of these aromatic residues are involved in typical hydrophobic stacking interactions with the glucose unit. This may be important to allow adaptation to binding of other sugar units, such as fructose units of poly-fructofuranose and inulin for example, which are also substrates cleaved by *Tm*
invertase. Interestingly, the structure of the plant fructan 1-exo-hydrolase [10] revealed that Trp260 has no equivalent. Furthermore, this residue is part of a two-residue loop-insertion (Figure 2) that is not present in most plant invertases and that obstructs the access to the protein’s surface. This suggests that for longer-chain substrates, any additional sugar unit would not be bound to the protein, but would dangle out into the solvent. Nevertheless, the enzyme is also capable of cleaving shorter substrates such as sucrose, or longer polysaccharide chains of fructose, such as nystose, kestose or derivative polymers of inulin [15]. Unfortunately, several attempts to crystallize Tm invertase with nystose, kestose or longer fructofuranose chains were unsuccessful. The crystallographic data for inactivated Tm invertase in complex with sucrose (results not shown) did reveal electron density in the active site, but this was too diffuse to reliably construct a bound substrate molecule. This being the case, the diffuse density might be due to either high disorder of the bound molecules or to the fact that the affinity of the crystal structure for sucrose molecules is too low for them to become trapped. Indeed, the $K_m$ value reported for Tm invertase is best when raffinose is the substrate ($K_m = 15$ mM), followed by inulin ($K_m = 19$ mM) whereas the $K_m$ with sucrose as substrate is 4-fold greater than that for raffinose with a value of 64 mM [15]. However, when it comes to catalytic efficiency, the situation is reversed with $k_{cat}/K_m$ values of $4.1 \times 10^4$, $3.1 \times 10^4$ and $9.3 \times 10^3$ M$^{-1}$·s$^{-1}$ for sucrose, inulin and raffinose as substrates respectively. Thus filling the +2 binding subsite with the $\alpha$-galactose moiety of raffinose perhaps increases the affinity of Tm invertase for the substrate, but probably also reduces the catalytic efficiency of Tm invertase by making the disaccharide product more ‘sticky’.

The hydrogen-bonding pattern around subsite +2 involves only one direct protein–substrate contact between Glu$^{103}$ O$^{-1}$ and the galactose O6 hydroxyl group. However, several strong hydrogen bonds are mediated by water molecules: Tyr$^{22}$ binds via a water molecule (Tyr O$^{-1}$–Ow 1355; 2.64 Å) to the O5 ring oxygen of the galactose unit and Glu$^{103}$ O$^{-1}$ via Ow 1356 (a distance of 2.64 Å) again to the O6 hydroxyl group. Neither Tyr$^{22}$ or Glu$^{103}$ are conserved in GH32 family enzymes and are most probably responsible for substrate specificity in this particular enzyme.

![Figure 4](image-url) Stereographic view of the catalytic site of invertase in complex with raffinose

Catalytic residues are shown in gold, aromatic residues in dark blue and other residues in green. Hydrogen bonds are in cyan. Ow designates water molecules.
Determination of the structures of more enzyme–substrate complexes covering the large diversity of enzymes is still required to depict in detail the fine-tuning of substrate specificities that occur in the GH32 family. The present structural analysis of the inactivated \( Tm \) invertase in complex with raffinose certainly represents a first achievement that has helped to identify several loops and specific residues that appear to be important for determining substrate specificity. Subsequent investigation by mutagenesis experiments should shed further light on the structural determinants responsible for substrate specificity within the GH32 enzyme family.

The authors thank Dr Wolfgang Liebl (Georg-August-Universität, Göttingen, Germany) for generously providing \( T. \) maritima genomic DNA. We also thank the staff of the European Synchrotron Radiation source for provision of beam time and for technical assistance at the beamline ID14-EH2. We thank Dr G. Sulzenbacher, Dr C. Bignon and Dr A. Gruetz for helpful discussions. This work was partly funded by grant QLK5-CT-2001-00443 (EDEN) from the European Commission.

REFERENCES

8a Naumoff, D. G. (2001) \( \beta \)-Fructosidase superfAMILY: homology with some \( \alpha \)-arabinases and \( \alpha \)-xylanolases. Proteins 42, 62–76