Independent folding of the A and B domains in the α-amylase family

Gerard Pujadas* & Jaume Palau

Unitat de Biotecnologia Computacional, Departament de Bioquímica i Biotecnologia, Universitat Rovira i Virgili, Plaça de la Imperial Tàrraco, 1. Tarragona 43005, Catalonia (Spain); tel.: ++ 34 977 55 95 65, fax: ++ 34 977 55 95 97, e-mail: pujadas@quimica.urv.es

Although the activities of the α-amylase family of enzymes are different, they have a common ancestor. Their polypeptide chain always has a multi-domain arrangement (though the number of domains usually depends on the enzyme activity). The domains that we call A and B are always found in these structures. The A domain, which is the catalytic domain, is always a TIM-barrel fold. The B domain varies in sequence length and fold, and lies between the third β-strand and the third α-helix of the A domain. Its function has not yet been fully established. To determine whether the variability of the B domain affects the folding of the A domain, we studied the geometrical characteristics of the eight-stranded β-sheet at the core of the A domain. Our results show that the geometry of the TIM barrel does not depend on the length or fold of the B domain and supports the idea of an independent folding pathway for the A and B domains in α-amylase enzymes. Unwanted mutations that produce a different barrel geometry may be recognised by molecular chaperones and discarded as functional molecules. Our results show that family 77 enzymes have the same barrel geometry as family 13 enzymes. This supports the hypothesis that they have a common origin.

Key words: α/β barrel, TIM barrel, α-amylase family, family 13 glycoside hydrolases, family 77 glycoside hydrolases, clan GH-H, chaperones.

Abbreviations: AAMY, α-amylase; CA, carbon atoms in the α position; CGTase, cyclodextrin glucanotransferase; GH, glycoside hydrolases; rmsd, root mean square deviation; SI, similarity index; TIM, triosephosphate isomerase.

Introduction

α-Amylase (AAMY) and related enzymes are grouped in families 13 (Henrissat, 1991), 70 and 77 of glycoside hydrolases [GH; http://afmb.cnrs-mrs.fr/CAZY/index.html; (Coutinho & Henrissat, 1999)] and constitute the so-called GH-H clan or AAMY superfamily. At present, 27 different enzyme activities covering transferases (EC 2.x.x.x), hydrolases (EC 3.x.x.x) and isomerases (EC 5.x.x.x) have been identified as members of the clan GH-H (Coutinho & Henrissat, 1999; Janecek, 2000a,b; MacGregor et al., 2001). In all cases, these enzymes show: (i) a re-
taining mechanism of the substrate stereochemistry; (ii) an AAMY-type \((\beta/\alpha)\)_8-barrel fold for the catalytic domain; and (iii) the same catalytic residues in equivalent locations [Asp197, Glu233 and Asp300 in \(\beta_4, \beta_5\) and \(L_7\) (the loop after \(\beta_7\)), respectively, where the numbers correspond to the location of these residues in the pig pancreatic AAMY sequence (Qian et al., 1993)]. The sequences in the segments located around the \(\beta\)-strands \(\beta_2, \beta_3, \beta_4, \beta_5, \beta_7\) and \(\beta_8\) in the TIM-barrel domain are well-conserved (Friedberg, 1983; Nakajima et al., 1986; Vihinen & Mantsalä, 1989; Janecek, 1994a,b) and at least four of these segments are used as signatures for recognizing new enzymes as members of the superfamily (Takata et al., 1992). Moreover, two types of mammalian proteins without catalytic function have been identified as having a sequence that is clearly similar to those of the members of this clan (Janecek, 2000a). One of these is made up of proteins involved in transporting dibasic and neutral amino acids across the cell membranes (Bertran et al., 1992; Wells & Hediger, 1992). The other one acts as the

![Fig. 1. Schematic diagram of the structure of the A+B-domains in the Bacillus licheniformis (Figs 1A,B) and in the Bacillus stearothermophilus (Figs 1C,D) AAMYs. \(\alpha\)-Helices are shown as spiral ribbons and \(\beta\)-strands are drawn as arrows from the amino end to the carboxy end of the \(\beta\)-strand. The TIM-barrel fold corresponding to the A-domain is shown in light gray and the B-domain is shown in dark gray. Figures 1A,C are end views whose C-terminal side of the \(\beta\)-sheet is towards the reader. Figures 1B,D are side views whose C-terminal side of the \(\beta\)-sheet is towards the top of the page. The figure was produced with MOLSCRIPT v2.1.2 (Kraulis, 1991) and the PDB files 1VJS and 1HVX.](image-url)
4F2 heavy-chain cell surface antigen, which is a type II membrane glycoprotein involved in cell growth (QUACKENBUSH et al., 1987). In comparison with the closely-related enzymes in the AAMY family (oligo-1,6-glucosidases and α-glucosidases), these proteins lack at least some of the amino-acid residues essential for activity, and therefore do not contain all of the four conserved regions (NAKAJIMA et al., 1986) that are necessary for enzymatic members of clan GH-H (JANEČEK, 2000a,b).

Enzymes from the clan GH-H have a multi-domain arrangement (although the number of domains usually depends on the enzyme activity). As mentioned above, all the members of the AAMY superfamily have the TIM-barrel fold of the catalytic domain (the so-called A domain) in common (PUJADAS & PALAU, 1999). Also, all the enzymes from the GH-H clan have a domain (the so-called B domain) intercalated between the third β-strand and the third α-helix of the catalytic domain (see Figure 1). Most members of the AAMY superfamily also show a well-conserved sequence pattern near the C-terminus of the B domain.

![Different folds for the B domain in the clan GH-H structures](image)

Fig. 2. Different folds for the B domain in the clan GH-H structures. The figure was produced with MOLSCRIPT v2.1.2 (KRAULIS, 1991) and the PDB files indicated at the top of each B domain. Some B domains are representative of a group of clan GH-H structures: 1CGT for CGTases, 1HX0 for mammalian AAMys, 7TAA for fungal AAMys and 1HVX for AAMys from cluster BV (see Figure 3).
domain (JANECEK, 1992, 1995a). This conserved motif has not been identified in the glycogen branching and debranching enzymes (JANECEK et al., 1997), maltosyltransferase and maltoligosyltrehalose synthase (JANECEK, 2000b). Sequence comparison between enzymes and non-enzymatic members of the AAMY superfamily shows that the non-enzymatic members probably have a TIM-barrel domain. The same comparison shows that the B domain is also found in the transport proteins, although part of it seems to be absent in the 4F2 heavy-chain antigens (JANECEK et al., 1997; JANECEK, 2000a). The putative B domain of the transport proteins seems to be closely related in sequence and fold to the one from oligo-1,6-glucosidase (JANECEK et al., 1997; JANECEK, 2000a).

The sequence length and fold of the B domains of enzymes with different activities can be very different (JESPERSEN et al., 1993; JANECEK et al., 1997). This is also true for enzymes with the same activities (PUJADAS & PALAU, 2001) (Fig. 2). Nevertheless, different clan GH-H structures have the same kind of fold for the B domain (JANECEK et al., 1997; JANECEK, 2000b). Some types of B domain contain several β-strands and one or two α-helices (e.g. fungi AAMYs). Others (e.g. barley AAMY) have an irregular fold with no well-defined secondary structure elements, and isoamylase has a long region in place of domain B that forms a cluster with the β4 → α4 loop of the TIM barrel domain (KATSUYA et al., 1998). Despite this variability, a common origin for all B domains has not been discarded (JANECEK et al., 1997). The B domain plays an important role in:

(i) controlling the isoenzyme specificity properties in barley AAMY [i.e. substrate affinity/binding, catalysis, sensitivity to inhibitors and stability; (RODENBURG et al., 1994)];
(ii) binding the Ca²⁺ ion in AAMYs and related CGTases (BOEL et al., 1990; MACHIUS et al., 1995, 1998);
(iii) the thermostability of the full protein in some Bacillus species (HWANG et al., 1997; DECLERCK et al., 2000); and
(iv) controlling the product specificity of several Bacillus CGTases (NAMAMURA et al., 1994; SIN et al., 1994; PENNINGA et al., 1995).

The fact that the majority of the stabilizing mutations cluster in the B domain (DECLERCK et al., 2000) has led to speculation that Bacillus AAMYs are initially inactive by the partial unfolding of this domain (TOMAZIC & KLIBANOV, 1988; NIELSEN & BORCHERT, 2000).

So far, 24 sequences from the GH-H clan, which cover 9 different enzyme activities and 2 GH families, have been crystallized (Fig. 3). We therefore have enough structural information to investigate how the variability of the B domain affects the folding of the A domain. To do this, we studied the conservation of the geometrical characteristics of the β-sheet that forms the interior of the TIM barrel. This is coherent with the fact that the packing of side chains within the closed β-sheet has been described as one of the most important factors for maintaining a TIM barrel structure (LESK et al., 1989; WODAK et al., 1990). Our results suggest that the folding of the A domain does not depend on the characteristics of the B domain.

### Material and methods

The PDB codes for the crystallized sequences of the GH-H clan were obtained from the CAZy database from 26 July 2001 (COUTINHO & HENRISSAT, 1999). The corresponding protein structures were imported from the last on-line release of PDB [http://www.rcsb.org/pdb/](http://www.rcsb.org/pdb/) (BERMAN et al., 2000). When there was more than one structure for the same sequence, the one with best global quality (i.e. best resolution and lowest R-value) was selected for further calculations. In such cases, we did a structural superposition with the “Best fit” algorithm of Swiss-PdbViewer v3.7 to ensure that the β-sheet geometry of the TIM barrel in the selected PDB file was also representative of the other structures of lower quality (GUEX & PEITZCH, 1997). We used the same program and algorithm for the other superpositions needed in this paper. Only backbone atoms were used in all the superpositions between structures with different sequences. The root mean square deviation (rmsd) was used in all cases to quantify the degree of structural similarity. Figure 3 shows the PDB code for the structures we selected for the study.

In TIM-barrel domains, the internal β-sheet forms a closed structure that, in the case of the clan GH-H enzymes, resembles a hyperboloid. The shape of this closed β-sheet may be described by the geometry of the sections obtained when the hyperboloid is cut by equidistant planes -or layers- that are perpendicular to the axis of the barrel. Each layer is built of four residues belonging to four alternate β-strands. Only the residues whose side chains face towards the interior of the barrel are used to build the layers. The carbon atoms in the α position (CA) of the four residues that form a layer are roughly located in the same plane. Figure 4 shows a schematic of the four-layered scaffold in the interior of the TIM barrel of the human salivary AAMY (PDB code: 1SMD). The geometry of a layer may be described by the following parameters:

(i) the distances between the two pairs of opposite CA in a layer (e.g. Gly193/Arg252 and Gly193/Phe335 for the first layer in 1SMD);
(ii) the angle between the two segments that define the two pairs of opposite CA; and
(iii) the area of the layer. The latter parameter is calculated under the assumption that:

(i) the layers are...
elliptical and (ii) the axes of the ellipse are defined by the segments that join the two pairs of opposite CA in a layer. These parameters may be used to compare topologically identical layers from different structures. A parametrization of this kind was very useful for studying the evolution of the packing of the interior of the TIM barrel in β-amylase (Pujadas et al., 1996) and in identifying and characterizing the conformational change of β-strand 6 upon ligand binding on the same enzyme (Pujadas & Palau, 1997). We made all the calculations in an Excel v98 worksheet after identifying the layers.

The layer distribution in the β-sheet core of crystallized AAMYs from family 13 has recently been described (Pujadas & Palau, 2001). All crystallized AAMYs have a four-layered scaffold in the interior of their TIM barrel (see Figure 4). In this scaffold, the odd β-strands contribute to the second and fourth layers, whereas the even β-strands contribute to the first and third layers. In general, the layer distribution for the other clan GH-H structures has been obtained by structural superposition with the ISMD structure and further visual examination of the correctness of the superposition and layer assignment. This was a valid method for identifying the layer structure of cyclodextrin glucanotransferase (CGTase; 1CXL), maltotetraose-forming amylase (1GCY), isoamylase (1BF2), maltogenic α-amylase (1QHO and 1SMA) and neopullulanase (1BVZ). Once we had characterized the layer scaffold of 1CXL, we superimposed the other CGTases in the study (1CGT, 1PAM, 1SMD and 1UOK) with 1CXL and identified their placement. This was a valid method for identifying the layer structure of cyclodextrin glucanotransferase (CGTase; 1CXL), maltotetraose-forming amylase (1GCY), isoamylase (1BF2), maltogenic α-amylase (1QHO and 1SMA) and neopullulanase (1BVZ). Once we had characterized the layer scaffold of 1CXL, we superimposed the other CGTases in the study (1CGT, 1PAM, 1CYG and 1CIU) with 1CXL and identified their layers. The “Best fit” algorithm of Swiss-PdbViewer v3.7 (GueX & Peitsch, 1997) failed in the structural superposition of 1ESW (4-α-glucanotransferase), 1EH9 (α-amylase/glycosyltrehalose trehalohydrolase) and 1UOK (oligo-1,6-glucosidase) both with ISMD or 1CXL. For these “problematic” structures, we used the DALI web-server (http://www.ebi.ac.uk/dali/) to identify the best structural counterparts. These were 1BVZ, 7TAA and 1BF2 for 1ESW, 1EH9 and 1UOK, respectively. Once the structural counterparts were known, the layers for 1ESW, 1EH9 and 1UOK were found as shown above.

We studied the similarity of the sequences corresponding to the PDB files from Figure 3 in the following way. First, all sequences were shortened to produce informationally similar segments that correspond to the strict (β/α)8-barrel plus B domain (A+B segments) structure (see Figure 1). Therefore, neither the N-terminal tail nor domain C were considered. The amino-acid segments of reference used to define the A+B sequences were obtained after analyzing the secondary structure information included in the PDB files. More specifically, we took the PDB sequences from the first residue in βA1 (first β-strand in the A domain) to the last residue in βAS as our reference. Second, we used all the A+B segments to produce two sets of sequences, one corresponding to the strict domain A and the other to domain B. The boundary between domains has recently been obtained for the AAMYs (Pujadas & Palau, 2001). The limits for the other AAMY superfamily members were obtained after visually inspecting the structural superpositions used to define the layer distribution (see above). Domain A sequences were obtained by joining the two subsegments on the left and right of the corresponding domain B. Pairwise comparisons of all the sequences from the same domain were done with the Clustal V algorithm (Higgins & Sharp, 1989) and the commercial program MEGALIGN v4.03 from the Laser gene software package (1999, DNASTAR, Inc., London, UK) running in a Macintosh iBook. The similarity index (SI) between two sequences was calculated with the method of Wilbur & Lipman (1983), with a gap penalty of 3, a K-tuple of 1, 5 top diagonals and a window size of 5. The SI was calculated as the number of exactly-matching residues in this alignment minus a “gap penalty” for every gap introduced. The result was then expressed as a percentage of the length of the shorter sequence. The protein weight matrix was PAM 250.

**Results and discussion**

The characteristics of the B domain are extremely diverse

Figure 3 shows that the length of the B domain in crystallized sequences from the GH-H clan varies considerably [from 30 residues in the glycosyltrehalose trehalohydrolase from Sulfolobus Solfataricus Km1 (PDB code: 1EIH) to 105 residues in the AAMY from Bacillus licheniformis (PDB code: 1VJS)]. All possible pairwise comparisons between the sequences of the B domains from crystallized clan GH-H structures are shown below the diagonal line in Figure 5.

We can see that the sequences (SI ranges from 93.8 to 71.9%) and structures (rmsd ranges from 0.65 to 0.29 Å; results not shown) of the B domains in the CGTases are very similar when the B domain of 1CGT is used as a reference for superposition with the B domains from the other CGTases (see Figure 2). These results within CGTases are consistent with the fact that these structures: (i) were obtained from closely-related bacterial species (i.e. from the Bacillus/Clostridium taxonomic group); (ii) have highly similar A-domain sequences (SI ranges from 91.1 to 62.1%); and (iii) have a good level of structural superposition for the A domain (rmsd ranges from 0.95 to 0.72 Å; results not shown).

We have recently demonstrated (Pujadas & Palau, 2001) that if the sequences from the A domain are used, AAMYs may be classified into two clusters for archaen (AI and AII), eight for bacteria (from BI to BVIII) and three for eukaryota (EI, EII and EIII). Moreover, if we com-
Fig. 3. Clan GH-H structures used in this study. The different structures are sorted according to the following priority arrangement: (1) the GH family; (2) the EC number; and (3) the species from which the enzymes have been obtained. The following information is provided for the structures: the species, the PDB identification.
pare the AAMYs from different kingdoms, some of the above-mentioned clusters group to form inter-kingdom clusters (BVIII/EIII, AII/BV/EI and EI/BIV). Figure 5 shows that the SI values of other comparisons between the A domains from AAMY are below the 20% threshold. In the B domain, the similarity is high if we compare sequences from the same cluster (i.e. 2AAA and 7TAA; 1VJS and 1HXV; 1HNY, 1SMD, 1HX0 and 1JAE). The B domains of clusters EIII and BVIII are similar in length (71/61 and 60; see Figure 3), sequence (SI from 33.3 to 23.3%) and structure [rmsd from 1.12 to 0.87 Å (results not shown); see also Figure 2] and reflect the relationship between the A domain of these clusters (Pujadas & Palau, 2001). We have also demonstrated the close relationship between the A domains from clusters BV and EI (Pujadas & Palau, 2001). This similarity is not found in B domains, the SI values for the sequence comparison are 14.8 and 13.1%; Fig. 5) because they are very different in length (105/102 residues for 1VJX/1HVX and 61 for 1AVA; see Figure 3) and fold (Fig. 2). It is also remarkable that the B domain from 7TAA (Aspergillus oryzae) has a relatively high sequence similarity with the one from the inter-kingdom cluster formed by BVIII (1AQX) and EIII (1HNX, 1SMD, 1HX0 and 1JAE) (see Figure 5). The similarity also extends to 2AAA (the other structure in the EI cluster) when we compare the structures of the B domain, not their sequences (rmsd values range from 1.24 to 0.92 Å; results not shown). This is highly remarkable because AAMYs from fungi (EI) and animals (EII) diverged a long time ago (Janecek, 1994b; Pujadas & Palau, 2001), as is clearly shown by the low SI of the B domain of these AAMYs (15.1–12.0%; see Figure 5).

Two maltogenic AAMYs – one from Bacillus stearothermophilus C599 (PDB code: 1QHO) and the other from Thermus sp. IM6501 (PDB code: 1SMA) – have also been crystallized. When we compared the sequences for their A domains, their SI was 21.1% (Fig. 5) and their rmsd was 1.37 Å. The SI and rmsd in the comparison of their B domains were 12.7% and 1.38 Å respectively. In fact, it has recently been suggested that the enzymes classed as maltogenic AAMYs under the EC 3.2.1.133 should be re-classified into two different classes, one that seems to be specific for the α-1,4-linkage (i.e. 1QHO) and another that is active on α-1,4 and α-1,6 bonds (i.e. 1SMA) (MacGregor et al., 2001). The Bacillus sequence was very similar to CGTases (the SI ranged from 39.9 to 36.0% for the A domain and from 35.9 to 26.6% for the B domain). From the structural point of view, the B domain of 1QHO and CG-Tases have the same fold (see Figure 2). Also, the rmsd of the structural superposition of the B domains from the CGTases onto 1QHO ranged from 0.96 to 0.86 Å, and the maltogenic AAMY from Bacillus had a five-domain structure that is similar to the one often associated with CGTases (Dauter et al., 1999). From the sequence similarity between this maltogenic AAMY and CGTases, Janecek suggested that the role of the former was that of an "intermediary" enzyme among "true" AAMYs and "true" CGTases (Janecek, 1995b). On the other hand, the maltogenic AAMY from Thermus sp. (PDB code: 1SMA) was very similar to the neopullulanase from Thermoactinomyces vulgaris (PDB code: 1BVZ) in terms of sequence (54.7 and 50.9% for SI when we compared their A or B domains) and structure (rmsd was 1.17 or 0.91 Å when we compared either the A or the B domains; see also Figure 2). The close relationship between some maltogenic AAMYs and the neopullulanase from Thermoactinomyces vulgaris has recently been noted (Matzke et al., 2000). It has also been suggested that these sequences – along with other sequences from cyclo maltodextrinases – constitute a subfamily within the GH-H clan (Matzke et al., 2000). At present, it is suggested that these sequences belong to the superfamily GH-H.
that the two maltogenic AAMYs should be classified using the label “AM-type” (i.e. amylase-like) for the one from *Bacillus* and the label “N-type” (i.e. neopullulanase-like) for the one from *Thermus* sp. (MacGregor et al., 2001).

The results in this section therefore support previous findings that, in general, the evolution of the B domain matches that of the A domain, which therefore indicates that the insertion of the B domain in the TIM barrel is not a recent event (Janecek et al., 1997). Some “deviations” we have detected to this general rule are: (i) the lack of similarity and the differences in the sequence length of the B domains in AAMYs from clusters B and E (Puigadós & Palau, 2001) do not match the similarities in their A domains; and (ii) the similarities between the structure and, to a lesser extent, the sequence of the B domains from AAMYs in cluster EII and those from the interkingdom cluster formed by BVIII and EIII.

**Conservation of the barrel geometry**

The packing of side chains within the closed β-sheet is one of the most important factors for maintaining a TIM barrel structure (Lesk et al., 1989; Wodak et al., 1990). Any possible “external” influence on the folding of a TIM-barrel domain can therefore be analyzed by studying how the geometrical characteristics of the closed β-sheet are affected. The method used to parametrize this β-sheet is described in the Material and methods section (see above). The fact that the different clan GH-II structures share the same geometry for the TIM-barrel β-sheet (i.e. a hyperboloid shape and four layers) makes it easier to compare the data from the structures in Figure 3.

Figure 4 shows a schematic diagram of the four-layered scaffold in the interior of the TIM barrel of the human salivary AAMY (PDB code: 1SMD). The four-layered scaffold for the other structures in Figure 3 can be inferred by combining the information in Figures 4 and 6. Figure 6 shows that only a few positions have preserved the same residue throughout evolution in all reported structures [i.e. Asp96, Arg195 and Glu233 in β3, β4 and β5 respectively: the numbers correspond to the location of these residues in 1HX0 (the pig pancreatic AAMY sequence); note that Gln257 in 1CXL and Gln208 in 1BAG in β5 correspond to non-natural mutations]. Another position is also well-, though not strictly, preserved (i.e. Leu289 in the β4 of 1ESW is different from the Gly residue found in the other clan GH-II structures).
Fig. 5. Sequence comparison of all A-domain sequences in clan GH-H structures (above the diagonal line) and all B-domain sequences from the same structures (below the diagonal line). SI corresponding to the sequence comparison of all A-domain sequences in the clan GH-H structures of Figure 3 (above the diagonal line) and of all B domain-sequences from the same structures (below the diagonal line). Sequences from the polypeptide chains in the PDB files correspond to those in Figure 3 and are also sorted according to this figure arrangement. SI values that correspond to sequences from the same EC number are highlighted. A special highlight is also used for the SI values from AAMYs that belong to the same cluster or to inter-kingdom related clusters (i.e. BVIII/EIII and EI/BV).
Fig. 6. β-Strand segments involved in the packing of the TIM-barrel β-sheet in clan GH-H structures. The first and third residues of each of the eight segments (highlighted by inversion) face the interior of the barrel and are members of two different layers (see Figure 4). The second residue of the segments faces the external helices or coils. Residues that are in the same layer but different β-strand are in the same row and highlighted by the same type of inversion. The combined information of this figure and Figure 4 (built with the 1SMD structure) may easily be used to infer the four-layered scaffold in all clan GH-H structures. The sequence number in the corresponding PDB file for the residue that starts each segment is indicated before each one. PDB files are listed in the same order as in Figure 3. PDB files with the same EC number are grouped.

From the results in Figure 7 we can compare the degrees of conservation of the β-sheet in the structures in Figure 3. The level of variation is generally very low if we take into account that: (i) the residues that form the layers have—in general— not been conserved throughout evolution (see above); (ii) the similarity between the domain A sequences is generally low (see Figure 5 and the previous section); and (iii) the sequence of the TIM barrel is interrupted by the presence of the highly variable B domain. Figure 7 also shows that the geometry of the structure from the family GH-77 (PDB code: 1ESW) strongly agrees with that from the family GH-13. This is consistent with a common origin for the two families, which shows that both are part of the same GH clan (i.e. clan GH-H), and the idea that structural characteristics are better preserved throughout evolution than the characteristics of the sequence (Schulz & Schirmer, 1979).

Conclusions

Our results show that the length, fold and sequence of the B domain only slightly affect the geometry of the closed β-sheet that forms the core of the TIM-barrel domain. It should be noted that the B domain is a long loop, not an independent
Fig. 7. Distances (in Å) between pairs of layer-forming CA-atoms from opposite β-strands (first and second column for each layer), angles (in degrees) formed between the corresponding distance segments (third column for each layer) and area (in Å²) if an elliptical shape is assumed (fourth column for each layer). Note: β', x indicates the position of the β-strand in the TIM-barrel structure and y indicates the position of the residue in the corresponding three-residues-long β-strand segment (see Figure 6). PDB files are listed in the same order as in Figure 3. PDB files with the same EC number are grouped.

Another indirect way of showing that the β3 → α3 “loop” is a “true” domain involves finding clan GH-II members whose “loop” has a very long sequence. Such sequences exist; for instance, the β3 → α3 “loop” reaches 250 residues in the case of the human glycogen debranching enzyme [Swiss-Prot accession number: P35573 (Yang et al., 1992; Jespersen et al., 1993; Janecek et al., 1997)]. Although such a long “loop” is expected to constitute a “true” domain in the structure of
A)

B)

Fig. 8. Schematic of the structure of the A+B-domains in the maltosyltransferase from *Thermotoga maritima*. See Figure 1 for the criteria for representing the α-helices, the β-strands, the A- and B-domains and the different views. The space-filling model of the maltose ligand indicates the position of the active site (at the C-terminal side of the β-sheet in the core of the TIM-barrel fold). The figure was produced with MOLSCRIPT v2.1.2 (Kraulis, 1991) and the PDB file 1GJW.

this enzyme, we cannot prove this due to the lack of structural information. Fortunately, Liebl and coworkers have very recently described the structure of the maltosyltransferase from *Thermotoga maritima* [PDB code: 1GJW; (Roujénikova et al., 2001)], which has a β3 → α3 “loop” that is 155 residues long. This is the longest α3 → α3 “loop” so far crystallized (see Figure 3) and it therefore provides an excellent base for investigating whether it can or cannot be considered a “true” domain. Visual inspection of the PDB file clearly shows that the α3 → α3 “loop” constitutes a folded unit that is independent of the TIM-barrel domain (Fig. 8). A similar result is found with two other crystallized GH-H structures with a long α3 → α3 “loop” (i.e. 1VJS and 1HVX; see Figure I and Figure 3). We can therefore conclude that there is enough “structural” evidence to show that, at least in some GH-H enzymes, the α3 → α3 “loop” is a “true” domain (i.e. the concept of the B domain is structurally supported).

Maintaining the TIM-barrel scaffold of clan GH-H proteins or achieving the correct fold seems to be a matter of preserving (i) a sequence that can adopt the characteristic secondary-structure pattern of domain A and (ii) a set of residues that belong to β-strands at positions that are far apart in the sequence but which come together because of medium- and long-range interactions. Unwanted mutations leading to noncanonical or misfolded structures may be recognized by molecular chaperones and discarded as functional molecules (Ellis & Hartl, 1996; Welch & Brown, 1996). At this point it would be interesting to ascertain why the B domain – with different folds and lengths and inserted in the middle of the domain A sequence – has an almost unappreciable effect on the folding of the TIM barrel (see Figure 7). Our findings are difficult to understand if the molecular chaperones do not help to fold each domain correctly and independently of each other using a step-by-step mechanism of sequential domain folding. This conclusion is also supported by the geometrical analysis of the TIM-barrel β-sheet in the recently-described maltosyltransferase from *Thermotoga maritima*, where the B domain is the longest in crystallized clan GH-H structures (results not shown). It would obviously be very interesting to know whether the geometry of the β-sheet barrel core also remains unchanged in the circularly permuted members of the GH-H clan (i.e. GHs from family 70). In these GHs, the B domain is divided to precede and succeed the TIM barrel and, as a result of the permutation, there is an insertion of 140–150 residues in the α6 → β6 loop, [equivalent respectively to α8 and β1 in non-permuted clan GH-H sequences; (MacGregor et al., 1996)]. Unfortunately, no structural information is yet available for this family. We therefore think that chaperones preserve the geometrical characteristics of the TIM barrel which are not affected by the variability of domain B. Our hy-
pothesis about the role of chaperones in clan GH-H evolution is consistent with Csermely’s hypothesis (Csermely, 1997) that holds that chaperones are “mandatory for the evolution of our present-day catalysts”.

Acknowledgements

I thank Kevin Costello of the Language Service of our University for his help in writing the manuscript; Dr. Richard Haser and Dr. Bernard Henrissat for their comments, opinions and suggestions, which were invaluable for writing the conclusions; and Dr. Stefan Janecek for inviting me to attend the ALAMY Symposium. Finally, this work is dedicated to the memory of my friend, Prof. Jaume Palau, who with these results made his last contribution to science.

References


PRZYLUS, I., TERADA, Y., FUJII, K., TAKAHAMA, T., SAENGGER, W. & STRATER, N. 2000. X-ray structure of arabaco bond to amyloallactase from


Received October 29, 2001
Accepted February 12, 2002