

β -Helical catalytic domains in glycoside hydrolase families 49, 55 and 87: domain architecture, modelling and assignment of catalytic residues

Daniel J. Rigden^{a,*}, Octávio L. Franco^{a,b}

^aEmbrapa Genetic Resources and Biotechnology, Cenargen/Embrapa, S.A.I.N. Parque Rural, Final W5, Asa Norte, 70770-900 Brasília DF, Brazil

^bUniversidade Católica de Brasília, Brasília DF, Brazil

Received 24 July 2002; revised 19 September 2002; accepted 19 September 2002

First published online 3 October 2002

Edited by Robert B. Russell

Abstract X-ray crystallography and bioinformatics studies reveal a tendency for the right-handed β -helix domain architecture to be associated with carbohydrate binding proteins. Here we demonstrate the presence of catalytic β -helix domains in glycoside hydrolase (GH) families 49, 55 and 87 and provide evidence for their sharing a common evolutionary ancestor with two structurally characterized GH families, numbers 28 and 82. This domain assignment helps assign catalytic residues to each family. Further analysis of domain architecture reveals the association of carbohydrate binding modules with catalytic GH β -helices, as well as an unexpected pair of β -helix domains in GH family 55.

© 2002 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Glycosidase; Iterative database search; Fold recognition; Evolutionary relationship

1. Introduction

The determination of the structure of *Erwinia chrysanthemi* pectate lyase PelC in 1993 [1] revealed a new and unexpected class of protein fold – the right-handed β -helix. Subsequent structural determinations of other β -helical proteins have enabled key features of the fold to be identified [2,3]. Each rung of the β -helix contains three β -strands connected by turns of variable length. The rungs stack to produce three parallel β -sheets, named B1–3. This repetitive nature leads to extensive stacks of similar or identical residues, both internal, largely hydrophobic stacks and external hydrophilic stacks [3], and also enables the prediction of β -helical folds based on scanning for the corresponding supersecondary structural motif [4]. The main chain atoms of the β -sheets superimpose extremely well between different β -helical structures, even in the absence of significant shared sequence similarity. In contrast, the loops that contain catalytic residues are very variable in length and structure.

Although an increasing number of functions are now being linked with the β -helical fold [5], this domain is particularly associated with proteins that bind to carbohydrates [2]. This association, resulting from successive crystallographic structure determinations, has been given new strength by a recent bioinformatics investigation [6]. Among carbohydrate modifying enzymes, a particular diversity is evident for glycoside

hydrolases (GHs; EC 3.2.1.-). These have been classified according to similarity in sequence (and structure, where available) into 88 families in the CAZY database [7,8] now accessible online [9]. Grouping in this way enables the determination of important characteristics of all members of a family based on experimental evidence obtained for one. Typically interest focuses on the identity of the two acidic catalytic residues, nearly universally involved in cleavage of the glycoside bond [10], and whether the mechanism is of the inverting or retaining type.

The simple sequence analysis employed for grouping GH families can not be expected to detect more distant relationships, in which structural and functional features may be maintained during evolution while sequence similarity drops to undetectable levels. Such cases of distant homology become evident through structural determination or, increasingly, through the use of modern fold recognition methods [11–15]. Currently, GH families 28 and 82 are known to contain β -helical folds [16,17]. Here, we demonstrate the presence of catalytic β -helical domains in three further GH families, 49, 55 and 87, and provide evidence for a common evolutionary origin for all five GH families. In contrast to known β -helical GH structures, where the catalytic domain is generally found alone [16] or in combination with small domains of unknown function [17], GH families 55 and 87 contain multiple carbohydrate binding modules.

2. Materials and methods

Members of GH families 49, 55 and 87 were located in the CAZY database and retrieved from GenBank to give sets of nine, eight and four sequences, respectively. Sequence alignments were produced with T-Coffee [18], and manipulated and hand-edited with Jalview (<http://www.ebi.ac.uk/~michele/jalview/>). Jalview was also used to determine the four maximally diverse representatives of each family. Sequence motifs were sought using MEME [19] and secondary structure predicted with PSI-PRED [20]. Searches were made in the PFAM domain alignment database [21] (<http://www.sanger.ac.uk/Pfam>). Fold recognition experiments made use of the Structure Prediction META server [22] (<http://bioinfo.pl/meta>). Particular attention was paid to the results of the two consensus fold recognition analyses, Pcons2 [23] and the ‘Shotgun on 3’ consensus prediction (D. Fischer, unpublished) which produces a score based on the combined results of three independent methods, FFAS [24], 3D-PSSM [25] and Inbgu [26]. Pcons2 and ‘Shotgun on 3’ are currently the methods best able to distinguish between true and false positives (see <http://bioinfo.pl/Live-Bench/> [27]). Iterated sequence database searches were carried out using PSI-BLAST [28] at the NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>) employing either 0.01 or 0.001 as the *E*-value significance threshold. When searches with a certain GH family member produced members of another GH family among the significant results, a possible common evolutionary origin for the two families was suggested.

*Corresponding author. Fax: (55)-61-340 3658.

E-mail address: daniel@cenargen.embrapa.br (D.J. Rigden).

The family representatives detailed in Table 1 were used as input for the fold recognition and iterated database searches. GH family members were also submitted for BETAWRAP analysis [4] (<http://cuckoo.lcs.mit.edu:8080/BetaWrap/betawrap.html>) in order to provide independent evidence of the presence of β -helices, and to provide the predictions of rung locations that helped in the alignment of GH families 49, 55 and 87 to known structures. Model building was carried out with MODELLER 6 [29], employing an iterative modelling scheme in which model regions with unusual packing or solvent exposure characteristics in a set of 20 models, highlighted by PROSA II analysis [30], were interpreted as resulting from possible local alignment errors. Target–template alignment changes were then made and new sets of models produced and analyzed. Stereochemical analysis of the final model was done with PROCHECK [31]. The final model was deposited with the PDB under the code 1h3k. Structural superpositions were carried out with LSQMAN [32] and the CE methods [33] (<http://cl.sdsc.edu/ce.html>) and structural relationships of known structures explored in the SCOP database [34] (<http://scop.mrc-lmb.cam.ac.uk/scop/index.html>).

3. Results and discussion

Searches in the PFAM database [21] with members of GH families 49, 55 and 87 showed the presence of carbohydrate binding and other domains (see later) but gave no hint of a catalytic domain. This, along with their separate family status in the CAZY database [9], is an indication that simple sequence comparisons were unable to demonstrate evolutionary or structural relationships between these proteins and other better characterized enzymes. In such cases, fold recognition is often useful in search for more distant relationships as they combine sequence information with other derived characteristics, thereby enhancing sensitivity. This improved sensitivity comes with the price of the possibility of false positive results but consensus fold recognition methods help minimize these and detailed analysis usually enables remaining misleading results to be discarded.

Fold recognition studies with representative members of GH families 49, 55 and 87 (see Table 1) immediately revealed the presence of β -helices in these enzymes (Table 1). In each case, the scores (for the matching regions, stripped of extraneous sequence) are well in excess of the current best-scoring false positive scores for the Pcons2 and ‘Shotgun on 3’ methods of 1.31 and 40.1, respectively [27]. Initial results, in agreement with previous observation of an internal sequence repeat [35], suggested the presence of two domains in GH family 55 members and these were subsequently analyzed separately (Table 1). Individual fold recognition methods also strongly favored β -helical folds. For GH 49, GH 55 N-terminal half, GH 55 C-terminal half and GH 87 representatives, respectively, 3D-PSSM scores were 0.164, 0.0019, 0.101 and 0.0029, FFAS scores were 7.5, 50.0, 65.7, 10.7 and inbgu scores were 33.6, 79.7, 32.7 and 42.2. Two other GH families, numbers 28 and 82, have previously been shown crystallographically to contain catalytic β -helical domains [16,17]. It was therefore interesting to note, among the many significantly scoring β -helical proteins, that GH families 49 and 87 produced much better scores for GH family 28 structures than for GH family 82 structures. For GH family 55 there was a slight preference for GH family 82 structures. Independent confirmation of the presence of β -helices was obtained with BETAWRAP which gave significant P values in the ranges 1.7×10^{-4} – 2.2×10^{-3} , 3.8×10^{-4} – 7.1×10^{-3} and 5.8×10^{-3} – 2.8×10^{-3} , respectively, for GH families 49, 55 and 87.

As reviewed [2], the overall structural similarity between the

determined β -helical structures argues for their sharing a common evolutionary origin, despite a general lack of sequence similarity and detailed resemblance in core packing interactions. Low sequence similarity between GH families 49, 55 and 87 and their structural matches was also evident here – around 10% and 12% β -helical domain sequence identity, respectively, between family 49 members or family 87 members with *Aspergillus niger* polygalacturonase II (PDB code 1czf). The corresponding comparison of GH family 55 member N- or C-terminal halves with *Alteromonas* sp. α -carrageenase gives pairwise sequence identities of 11–15% and 10–18%. As fold recognition can produce significant results for structural analogues as well as distant structural homologues, we sought further evidence regarding evolutionary relationships with sensitive sequence comparisons carried out using PSI-BLAST. As shown in Fig. 1, the results are consistent with all these families sharing a common evolutionary origin since a clear network of connections can be achieved even at the conservative E -value threshold of 0.001. Remarkably, a single iteration at the more relaxed threshold of 0.01 is sufficient to demonstrate a relationship between GH families 82 and 28 despite their sharing insignificant overall sequence similarity (13% of 197 matched residues, for example, in a structural alignment of 1czf from family 28 with 1h80 of family 82) and totally different catalytic sites [17]. These and other results [6] counter the notion that PSI-BLAST is not well suited for work with β -helical proteins [4]. None of the results in Fig. 1 were previously reported [6]. The relationships between the previously identified β -helical ‘CASH’ domain [6], of around 150 residues, and the β -helical domains discussed here, typically with 300–350 residues, is unclear.

In the cases of GH families 49 and 87, the fold recognition alignments (Fig. 2) with 1czf and other GH family 28 structures immediately revealed the conservation of three aspartate residues, numbers 180, 201 and 202 in 1czf. These residues in 1czf have been experimentally shown by site-directed mutagenesis to participate in catalysis [36]. This enables the identification of catalytic residues in GH families 49 and 87 (Table 1) and is evidence for the existence of a closer evolutionary relationship between these families and family 28 than with

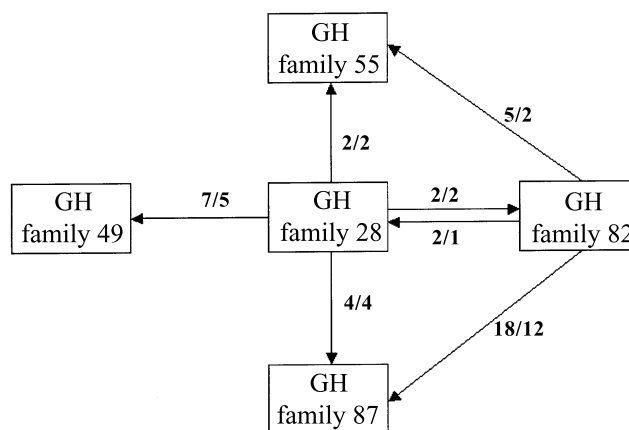


Fig. 1. Schematic representation of the GH family relationships revealed by PSI-BLAST [28]. The presence of family B in the significant results of searches made using family A is represented by an arrow from family A to family B. Each arrow is associated with the number of iterations required to demonstrate the relationship (x/y) at E -value thresholds of 0.001 (x) or 0.01 (y).

Table 1
CAZY [9] glycosidase families containing β -helical domains

Family	Activities	Number of non-redundant sequences in GenBank	Mechanism	Representative ^a	Pcons2 ^b	Shotgun on 3 ^c	Catalytic residues ^d
28	Polygalacturonase (EC 3.2.1.15) Exo-polygalacturonase (EC 3.2.1.67) Exo-polygalacturonase (EC 3.2.1.82) Rhamnogalacturonase (EC not defined)	405	Inverting	129934; <i>Aspergillus niger</i> ; polygalacturonase II; PDB 1czf [36]	–	–	D180, D201, D202, H223
82	ι -carrageenase (EC 3.2.1.-)	3	Inverting	10039456; <i>Alteromonas</i> sp.; ι -carrageenase; PDB 1h80 [17]	–	–	E245, D247, E310
49	Dextranase (EC 3.2.1.11) Isopullulanase (EC 3.2.1.57) Dextran-1,6 α -isomaltotriosidase (EC 3.2.1.95)	9	Inverting	1352244; <i>Penicillium minioluteum</i> ; dextranase [48]	4.93 1czf	99.3 1czf	D389, D412, D413
55	Exo-1,3-glucanase (EC 3.2.1.58) Exo-1,3-glucanase (EC 3.2.1.39)	8	Unknown	751495; <i>Ampelomyces quisqualis</i> ; exo-1,3-glucanase; Rotem et al. (direct database submission)	N-terminal half 3.96 1h80 C-terminal half 3.22 1h80	100 1rmg 100 1rmg	Probably among D233, D599 D304, E657, D679
87	Mycodextranase (EC 3.2.1.61)	4	Inverting	15128557; <i>Streptomyces</i> sp. mycodextranase [49]	2.77 1hg8	100 1hg8	D410, D429, D430

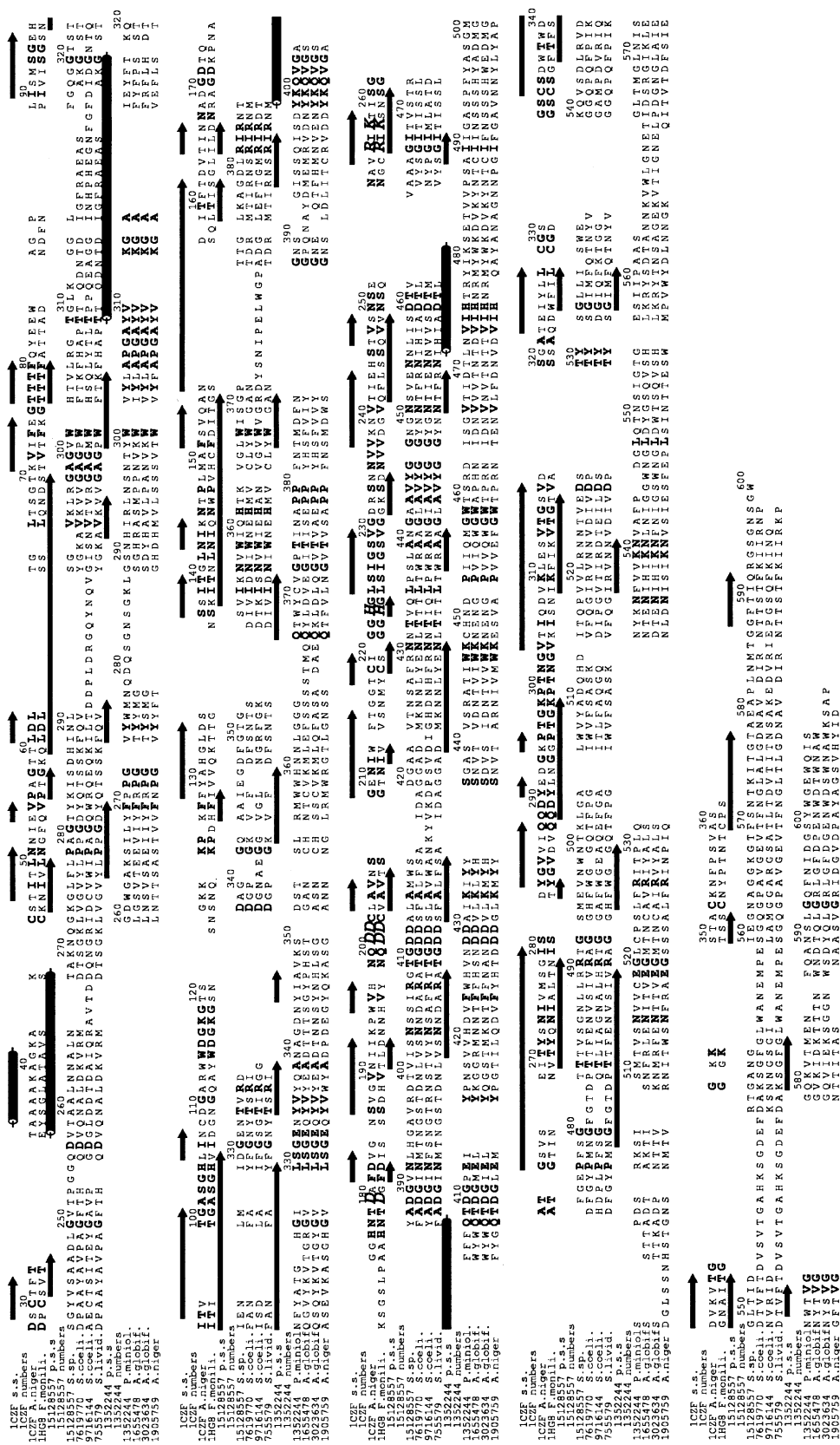
Experimentally determined structures are available for families 28 and 82. Structural predictions are presented for families 49, 55 and 87.

^aIn each case a GenBank ID is followed by species name, enzyme name, a PDB code in the cases of known structures, and reference.

^bScore by the Pcons2 consensus fold recognition method [23].

^cScore by the 'Shotgun on 3' consensus fold recognition method (D. Fischer, unpublished).

^dCatalytic residues, known or confidently predicted in the cases of experimentally determined structures, inferred from the results presented here in the remaining families.



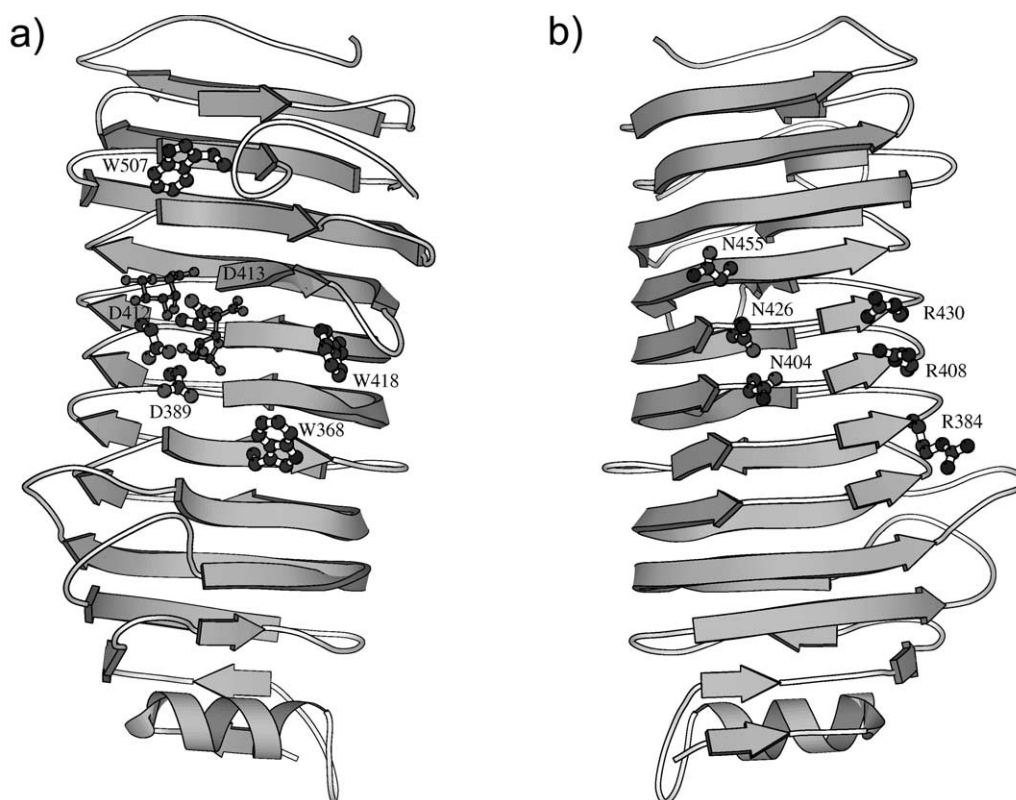


Fig. 3. MOLSCRIPT [47] diagram of the final model of family 87 representative *Streptomyces* sp. mycodextranase with secondary structure as defined by STRIDE [45]. a: View of the substrate binding cleft with catalytic acidic residues and highly conserved aromatic residues, possibly involved in carbohydrate binding [41], drawn as ball and stick. The bound galacturonate molecules recently observed in complex with GH family 28 *Stereum purpureum* endopolygalacturonase [40] are superimposed and drawn as thinner dark sticks. b: The reverse of the model showing the asparagine and arginine residues, suitably positioned for stacking interactions.

family 82. This is consistent with the PSI-BLAST results (Fig. 1) and the predominance of family 28 in the threading results obtained with families 49 and 87. Other residues, shown to be involved in substrate binding to *A. niger* polygalacturonase II [36], are not conserved in GH families 49 and 87, presumably reflecting the different substrates of these three families. The family 49 member *Penicillium minioluteum* dextranase has been the subject of a previous fold recognition and modelling study [37]. This reported a proposed structural relationship with sialidase and galactose oxidase [37] which is strongly disfavored by our results.

In the case of GH family 87, the alignment with GH family 28 member *A. niger* polygalacturonase II (PDB code 1czf) was of sufficient quality to permit model building of the catalytic β -helical domain, with the exception of a single 11 residue insertion. Modelling of β -helical proteins is particularly challenging since large alignment errors can readily result from missing β -strands [38]. Nevertheless, the alignment of catalytic site residues provided a confident alignment region in the center of the initial fold recognition-derived alignment.

From this, with the aid of BETAWRAP analysis, the alignment was carefully checked in both directions. During the rounds of iterative model building several buried charged residues were observed and alignment modifications made in order to reposition them. The objective model quality, measured by PROSA II [30] improved significantly during the process. The final model of *Streptomyces* sp. mycodextranase (Fig. 3) has a PROSA II score of -5.23 , corresponding to a pG value of 0.8 , well in excess of the 0.5 cutoff indicative of correct fold and largely correct alignment [39]. The overall stereochemical G -factor [31] of -0.36 is indicative of good stereochemistry and the model has 79% of residues in most favored regions of the Ramachandran plot with just a single disallowed residue. Several interesting features of the final model are shown in Fig. 3. In the trough, containing the catalytic site and along which the polysaccharide substrate lies [40], several aromatic residues, frequently involved in protein–carbohydrate interaction [41], are present (Fig. 3a). On the opposite side of the molecule (Fig. 3b), asparagine and arginine residues are suitably positioned for stacking interactions.

Fig. 2. β -Helical domain alignment of four maximally diverse members of GH families 87 (center) and 49 (lower), each labeled with GenBank ID and abbreviated species name, with *A. niger* polygalacturonase II (PDB code 1czf) and *Fusarium moniliforme* endopolygalacturonase (PDB code 1hg8) representing family 28. Representative sequences (Table 1) are numbered and their predicted secondary structures shown. The actual secondary structure of 1czf, as determined with STRIDE [45], and its numbering are shown above the alignment. Identities within the three groups are in bold and catalytic and substrate binding residues in GH family 28 [36] shown as large bold italic. The figure was made using ALSCRIPT [46].

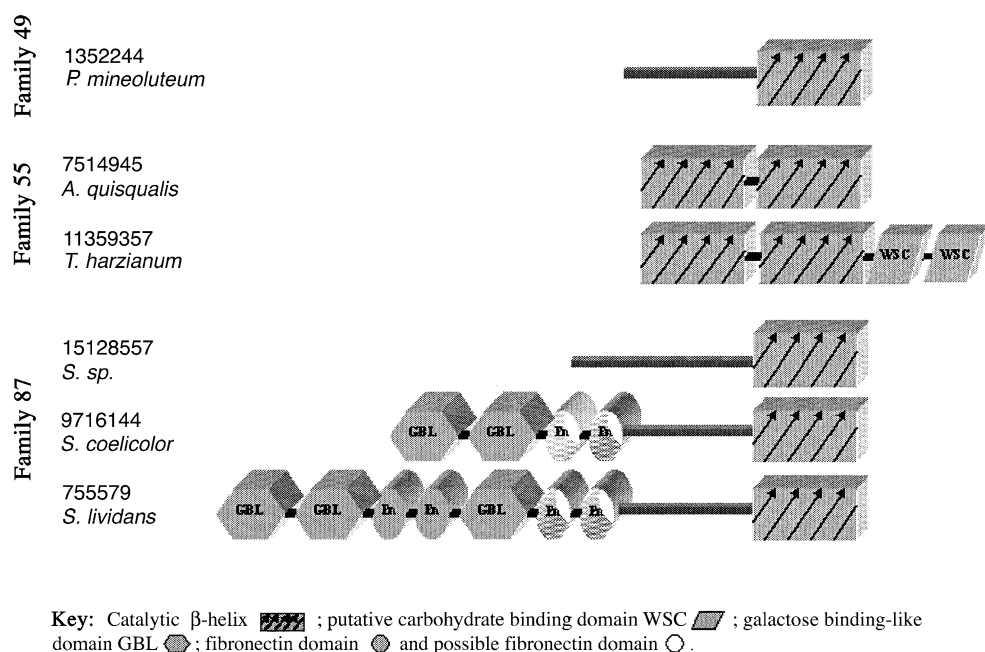


Fig. 4. Schematic representation of domain architectures for selected members of GH families 49, 55 and 87.

tions [3,17]. These residues – arginines 384, 408, 430 and asparagines 404, 426, 455 – are highly conserved (Fig. 2) within GH family 87 but are not shared with the template structure so that their appearance is suggestive of an accurate model. Electrostatic potential analysis (not shown) reveals a positive potential associated with the substrate binding cleft of the polygalacturonase template, contrasting with a more neutral substrate binding cleft in the mycodextranase. This difference is in agreement with the former enzyme acting on a negatively charged substrate and the latter an electrostatically neutral substrate.

The presence of two complete β -helical domains in the GH family 55 members (Table 1) seems to be unprecedented in the published literature. The alternative hypothesis that a single, long β -helix is present can be ruled out for several reasons. First, the amphiphilic α -helix that nearly always caps the N-terminal end of the β -helical fold is clearly predicted in both halves of GH family 55. Furthermore, alignment of N-terminal halves with C-terminal halves and MEME motif analysis (not shown) both identify two regions of significant sequence identity, at the N-terminal end of the fold, between the two halves of family 55 members. The alignment with α -carrageenase obtained from fold recognition shows that the first region corresponds to a calcium binding loop [17]. A DxxxDD motif (starting at positions 75 and 425 in the N- and C-terminal halves, respectively, of *Ampelomyces quisqualis* exo-1,3-glucanase) is well conserved in GH family 55 here with the exception of three members in which the entire loop is deleted. The second highly conserved region in the alignment of both halves comprises the first complete rung of the β -helix and contains two entirely conserved residues (Gly113 with Tyr115 and Gly452 with Tyr454, in the N- and C-terminal halves, respectively, of *A. quisqualis* exo-1,3-glucanase) which contact residues of the same calcium binding loop. Although these sequence similarities provide good evidence for an origin of family 55 in a gene duplication, as previously suggested

[35], it is striking that in the remainder of the sequences no similarity between N- and C-terminal halves is evident, perhaps further evidence of the rapid evolution associated with the β -helical fold [2].

In the case of GH family 55, the catalytic residues putatively identified in the family 82 α -carrageenase structure are not present in either β -helical domain. Alternative alignments with family 28 also show a lack of conservation of catalytic site residues. These results suggest the presence of a third kind of glycoside hydrolase catalytic site associated with the β -helical fold. Catalytic proton donating and accepting residues presumably lie among the set of conserved aspartates and glutamates [10], of which nine are present in GH family 55, four in the N-terminal domain and five in the C-terminal domain. Overall domain conservation does not help to identify a single catalytic domain with pairwise sequence identities between GH family 55 members in the N- and C-terminal halves of 39 and 40%, respectively. Also, the quality of the alignment of GH family 55 with α -carrageenase does not permit the identification of catalytic residues from predicted structural proximity. However, the trend towards location of catalytic site in the second half of the β -helical fold [16,17] helps to narrow down the possible catalytic residues to those shown in Table 1. GH family 55 represents a particularly interesting case for structural determination since three possible scenarios may be imagined, each of which would involve novelties. Both β -helical domains might contribute to the catalytic site, as has not yet been observed. Alternatively, each domain might carry its own, structurally different, catalytic site. We can also imagine the novel recruitment of one of the β -helical domains as a carbohydrate binding module to enhance the catalytic efficiency of the other domain.

Searches in the PFAM database [21] reveal that, with only five exceptions, each containing a single fibronectin domain [42], the 245 β -helical folds in GH family 28 are not coupled to other domains producing proteins around 400–500 residues

in length. However, as Fig. 4 shows, the β -helical catalytic domains considered here are coupled to other identifiable domains. The *Streptococcus lividans* GH family 87 member is annotated by PFAM as containing one F5/8 domain (accession PF00754), as found in blood coagulation factors, from residue 178 onwards, followed by two fibronectin domains (PF00060). Fold recognition experiments (not shown) gave similar, strongly, significant scores to F5/8 domains and to the structurally similar galactose binding like (GBL) domains, as seen, for example, at the N-terminus of galactose oxidase. A carbohydrate binding role makes excellent functional sense so this domain was reassigned as a carbohydrate binding module. Other fold recognition studies with the *S. lividans* GH family 87 member conclusively demonstrated the presence of a second GBL domain at the N-terminus and a third such domain following the fibronectin domains, which is in turn followed by a predicted all- β region giving borderline significant matches to several structures. The domain organization for the *S. lividans* GH family 87 member is suggestive of a duplication event involving a GBL domain in conjunction with two fibronectin domains so that this all- β region was tentatively annotated as containing two fibronectin domains (Fig. 4). One GH family 55 member from *Trichoderma harzianum*, GenBank ID 11359357 [43], contains two WSC domains (PF01822), putatively assigned a role in carbohydrate recognition, at its C-terminal end. These are of unknown structure, but fold recognition suggests a defensin-like structure (not shown). Interestingly, at the Fold level of the SCOP database [34], defensins are grouped, as knottins, with a variety of lectins and cellulose binding domains, supporting a carbohydrate binding function.

In summary, we have demonstrated that GH families 49 and 87 share an evolutionary origin with GH family 28, thereby enabling the identification of their catalytic residues. GH family 55 unexpectedly contains two β -helical domains, deriving from a gene duplication event, that probably also share a common evolutionary origin with other GH families, but which contain a third class of β -helix-associated GH catalytic site. Fold recognition experiments also assigned and reassigned probable carbohydrate binding modules in GH families 55 and 87 thus demonstrating a modular GH architecture, long apparent for other GH families [44], but not hitherto obvious for β -helical GHs.

Acknowledgements: The authors are grateful to Dr. Leszek Rychlewski for assistance with the META server.

References

- [1] Yoder, M.D., Keen, N.T. and Jurnak, F. (1993) *Science* 260, 1503–1507.
- [2] Jenkins, J., Mayans, O. and Pickersgill, R. (1998) *J. Struct. Biol.* 122, 236–246.
- [3] Jenkins, J. and Pickersgill, R. (2001) *Prog. Biophys. Mol. Biol.* 77, 111–175.
- [4] Bradley, P., Cowen, L., Menke, M., King, J. and Berger, B. (2001) *Proc. Natl. Acad. Sci. USA* 98, 14819–14824.
- [5] Leinala, E.K., Davies, P.L. and Jia, Z. (2002) *Structure* 10, 619–627.
- [6] Ciccarelli, F.D., Copely, R.R., Doerks, T., Russell, R.B. and Bork, P. (2002) *Trends Biochem. Sci.* 27, 59–62.
- [7] Henrissat, B. (1991) *Biochem. J.* 280, 309–316.
- [8] Henrissat, B. and Bairoch, A. (1993) *Biochem. J.* 293, 781–788.
- [9] Coutinho, P.M. and Henrissat, B. (1999) Carbohydrate-Active Enzymes server at URL: <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>.
- [10] Davies, G. and Henrissat, B. (1995) *Structure* 3, 853–859.
- [11] Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. and Davies, G. (1995) *Proc. Natl. Acad. Sci. USA* 92, 7090–7094.
- [12] Naumoff, D.G. (1999) *FEBS Lett.* 448, 177–179.
- [13] Naumoff, D.G. (2001) *Proteins* 42, 66–76.
- [14] Pons, T., Hernandez, L., Batista, F.R. and Chinea, G. (2000) *Protein Sci.* 9, 2285–2291.
- [15] Rigden, D.J. (2002) *FEBS Lett.* 523, 17–22.
- [16] Petersen, T.N., Kauppinen, S. and Larsen, S. (1997) *Structure* 5, 533–544.
- [17] Michel, G., Chantalat, L., Fanchon, E., Henrissat, B., Kloareg, B. and Dideberg, O. (2001) *J. Biol. Chem.* 276, 40202–40209.
- [18] Notredame, C., Higgins, D.G. and Heringa, J. (2000) *J. Mol. Biol.* 302, 205–217.
- [19] Bailey, T.L. and Elkan, C. (1994) in: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28–36, AAAI Press, Menlo Park, CA.
- [20] Jones, D.T. (1999) *J. Mol. Biol.* 292, 195–202.
- [21] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) *Nucleic Acids Res.* 30, 276–280.
- [22] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) *Bioinformatics* 17, 750–751.
- [23] Lundström, J., Rychlewski, L., Bujnicki, J. and Elofsson, A. (2001) *Protein Sci.* 10, 2354–2362.
- [24] Rychlewski, L., Jaroszewski, L., Li, W. and Godzik, A. (2000) *Protein Sci.* 9, 232–241.
- [25] Kelley, L.A., MacCallum, R.M. and Sternberg, M.J.E. (2000) *J. Mol. Biol.* 299, 499–520.
- [26] Fischer, D. (2000) in: *Pacific Symposium on Biocomputing*, Hawaii (Altman, R.B., Dunker, A.K., Hunter, L., Lauderdale, K. and Klein, T.E., Eds.), pp. 119–130, World Scientific, Singapore.
- [27] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) *Proteins* 45 (Suppl. 5), 184–191.
- [28] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) *Nucleic Acids Res.* 25, 3389–3402.
- [29] Sali, A. and Blundell, T.L. (1993) *J. Mol. Biol.* 234, 779–815.
- [30] Sippl, M.J. (1993) *Proteins* 17, 335–362.
- [31] Laskowski, R., MacArthur, M., Moss, D. and Thornton, J. (1993) *J. Appl. Crystallogr.* 26, 283–290.
- [32] Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M. and Jones, T.A. (2001) in: *International Tables for Crystallography*, Vol. F. Crystallography of Biological Macromolecules (Rossmann, M.G. and Arnold, E., Eds.), pp. 353–356, 366–367, Kluwer Academic, Dordrecht.
- [33] Shindyalov, I.N. and Bourne, P.E. (1998) *Protein Eng.* 11, 739–747.
- [34] Lo Conte, L., Brenner, S.E., Hubbard, T.J., Chothia, C. and Murzin, A.G. (2002) *Nucleic Acids Res.* 30, 264–267.
- [35] Donzelli, B.G., Lorito, M., Scala, F. and Harman, G.E. (2000) *Gene* 277, 199–208.
- [36] Van Santen, Y., Benen, J.A., Schroter, K.H., Kalk, K.H., Armand, S., Visser, J. and Dijkstra, B.W. (1999) *J. Biol. Chem.* 274, 30474–30480.
- [37] Pons, T., Chinea, G., Olmea, O., Bledarain, A., Roca, H., Padrón, G. and Valencia, A. (1998) *Proteins* 31, 345–354.
- [38] Williams, M.H., Shirai, H., Shi, J., Nagendra, H.G., Mueller, J., Mizuguchi, K., Miguel, R.N., Lovell, S.C., Innis, C.A., Deane, C.M., Chen, L., Campillo, N., Burke, D.F., Blundell, T.L. and de Bakker, P.I. (2001) *Proteins* 45 (Suppl. 5), 92–97.
- [39] Sanchez, R. and Sali, A. (1998) *Proc. Natl. Acad. Sci. USA* 95, 13597–13602.
- [40] Shimizu, T., Nakatsu, T., Miyairi, K., Okuno, T. and Kato, H. (2002) *Biochemistry* 41, 6651–6659.
- [41] Quicho, F.A. and Vyas, N.K. (1999) in: *Bioinorganic Chemistry: Carbohydrates*, pp. 441–457, Oxford University Press, New York.
- [42] Bork, P. and Doolittle, R.F. (1992) *Proc. Natl. Acad. Sci. USA* 89, 8990–8994.
- [43] Cohen-Kupiec, R., Broglie, K.E., Friesem, D., Broglie, R.M. and Chet, I. (1999) *Gene* 226, 147–154.

- [44] Henrissat, B. and Davies, G.J. (2000) *Plant Physiol.* 125, 527–531.
- [45] Frishman, D. and Argos, P. (1995) 23, 566–579.
- [46] Barton, G.J. (1993) *Protein Eng.* 6, 37–40.
- [47] Kraulis, J. (1991) *J. Appl. Crystallogr.* 24, 946–950.
- [48] Roca, H., Garcia, B., Rodriguez, E., Mateu, D., Coroas, L., Cremata, J., Garcia, R., Pons, T. and Delgado, J. (1996) *Yeast* 12, 1187–1200.
- [49] Okazaki, K., Amano, T., Morimoto, T., Iemoto, T., Kawabata, T., Hayakawa, S. and Akimitsu, K. (2001) *Biosci. Biotechnol. Biochem.* 65, 1684–1687.