β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven

John Jenkins, Leila Lo Leggio, Gillian Harris, Richard Pickersgill*

Protein Crystallography, Department of Protein Engineering, Institute of Food Research, Reading Laboratory, Earley Gate, Whiteknights Road, Reading, RG6 2EF, UK

Received 27 February 1995

Abstract Comparison of the recently determined crystal structures *Pseudomonas fluorescens* subsp. *cellulosa* family F xylanase, (1–3)-β-glucanase and (1–3,1–4)-β-glucanase and the catalytic domain of *E. coli* β-galactosidase reveals that they belong to a superfamily of 8-fold β/α-barrels with similar amino acid residues at their active sites. In the three families that these enzymes represent, the nucleophile is a glutamate, which is located close to the carboxy-terminus of β-strand seven. In addition all three enzymes have the sequence asparagine-glutamate close to the carboxy-terminus of β-strand four. This glutamate has been identified as the acid/base in the family F xylanases and is essential for catalysis in β-galactosidase. We suggest that the equivalent residue in the barley glucanases is the acid/base. Analysis of the sequences of family 1 β-glucosidases and family 5 cellulases shows that these enzymes also belong to this superfamily which we call the 4/7 superfamily.

Key words: Superfamily; Glycosyl hydrolase; Family F xylanase; 8-Fold β/α-barrel; β(1–3) and β(1–3,1–4) glucanase

1. Introduction

Catalysis by many glycohydrolases involves two essential carboxylates, one acting as proton donor and the other as the nucleophile [1]. Cleavage of the glycosidic bond results either in retention or inversion of configuration at the anomeric carbon. While the typical distance between the acid/base and the nucleophilic carboxylate in retaining enzymes is about 5 Å, in inverting enzymes they tend to be further apart, presumably because a water molecule must be accommodated between the nucleophile and substrate.

Henrissat [2,3] has compared all the available sequences of glycohydrolases (E.C. 3.2.1.x) using hydrophobic cluster analysis and classified these enzymes into families (currently 48, Henrissat, personal communication). Each family comprises enzymes which have clearly diverged from a common ancestor. As three dimensional structures have become available, further relationships between these families have been observed. The best example is the family of enzymes with an overall fold first seen in hen egg white lysozyme [4], where structural similarities were detected with the lysozymes from phage T4 [5], goose [6] and more recently chitinase [7,8] and a bacterial muramidase [9]. Thus families 19, 22, 23 and 24 in Henrissat's classification form a superfamily with a conserved fold.

In this paper we show how structural and sequence comparison of xylA from *P. fluorescens* subsp. *cellulosa* [10] with other glycohydrolases suggest that families 1 (β-glucosidases), 2 (β-galactosidases), 5 (family A cellulases), 10 (family F xylanases), and 17 (β-glucanases), also form a superfamily (Table 1). All of these enzymes act by a retaining mechanism except family 17 for which the stereochemistry of cleavage is unknown (Henrissat, personal communication).

2. Methods and databases

Native xylA from *Pseudomonas fluorescens* subsp. *cellulosa* has been recently solved [10]. Other atomic coordinates were obtained from the PDB (October 1994) [11]. Structures were displayed using QUANTA 4.0 [12] run on a Silicon Graphics Indigo^2 Elan. Structural alignments were made using MNYFIT 5.0 [13] running on a VAX 4400. Protein sequences were retrieved from the Swiss-Prot protein sequence database (release 30) or the PIR database (release 42). Fig. 2 was prepared using the program MOLSCRIPT [14].

3. Results and discussion

The architecture of the catalytic domain of *Pseudomonas fluorescens* subsp. *cellulosa* xylA, a family F xylanase, is the 8-fold β/α-barrel [10]. The substrate binding cleft has been identified by binding studies in the crystal and is formed by long loops at the carboxy-terminal end of β-strands 4 and 7 and short loops at the carboxy-terminal end of β-strands 5 and 6 (Fig. 2a). The active site acid/base and nucleophile of xylA are glutamate-127, located close to the carboxy-terminal end of β4, and glutamate-246, at the carboxy-terminal end of β7.

The barley glucanases GHS and GHR (family 17) are also 8-fold β/α-barrels [15], which can be superimposed on the xylA structure. RMS deviations for GHS and GHR are 1.33 Å (64 equivalent alpha-carbon atoms) and 1.25 Å (61 equivalent alpha-carbon atoms) respectively. In general, the β-strands superimpose much better than the α-helices.

Superimposition of the three structures brings the catalytic nucleophiles Glu^{231} (GHS) and Glu^{232} (GHR) close to Glu^{246} of xylA (Fig. 1). There is no carboxylate in xylA in a position corresponding to the proposed acid/base in the glucanases

*Corresponding author. Fax: (44) 734 267917.

Abbreviations: PDB, Brookhaven Protein Databank; xylA, xylanase A; GHS, (1–3)-β-glucanase; GHR, (1–3,1–4)-β-glucanase; *p*NPC, *p*-nitrophenyl cellobioside; β-gal, β-galactosidase; s.d.m., site directed mutagenesis.

SSDI 0014-5793(95)00252-9

(Glu^{288}). However Glu^{94} (GHS) and Glu^{93} (GHR), at the carboxy-end of β4, are equivalent to the catalytic acid/base Glu^{127} of xylA (Fig. 1) and are within the sequence IAVGNEV, highly conserved in family 17. The VGNE sequence breaks the polar/non-polar pattern of the beta strand, so that the two non-polar side chains Val and Gly both point towards the hydrophobic core, while the Asn and Glu side chains point towards the solvent. In xylA the corresponding Glu^{127} is found in the sequence WDVVNE (Table 1), again containing two non-polar groups (Val-Val) followed by the two polar Asn-Glu, which are in the same conformation as in the barley glucanases, i.e. with ϕ, φ angles typical of β-strand and left-handed α-helix respectively.

The distances between Oε atoms of Gluz^{4} (GHS) or Glu^{93} (GHR) and the nucleophilic Glu^{231}/Glu^{232} are 4.6 Å and 4.5 Å respectively, therefore consistent with Glu$^{94/93}$ being the acid/base catalysts in a retention mechanism. On the other hand, the distances between Oε atoms of Glu^{231} (GHR) or Glu^{232} (GHS) and Glu^{288}, which has been previously proposed as the acid/base by chemical labelling [16], are 8.2 Å and 8.4 Å respectively, which would be more consistent with an inversion mechanism. In absence of further evidence, the acid-base catalyst in barley glucanases has still to be convincingly assigned as either Glu^{94} (GHS) in a retention mechanism, or Glu^{288} in an inversion one.

Table 1
Alignment of the equivalent regions of families forming the 4/7 superfamily with the regions surrounding the acid/base and nucleophile in xylA

Family	Activities	Acid/base	Nucleophile
1	BGAL, BGLU PBGLU, LPH	WLTF**NE**P YF H I I I T M M	YIT**E**NGA V S V M I L
2	BGAL, BGLR	IWSLG**NE**S M CA P I V	ILC**E**YAH IS GA QT MV SV L
5	EG, EX	IYEIA**NE**P AF LL V V GM TC	FVT**E**WGT YCG F V IAS Y A VI S L
10	XYN, EG, CBH	WDVV**NE**A C I	KIT**E**LDV QV M H I
17	LAM, GH, XLAM, LIC	YIAAG**NE**V SV I I	VVS**E**SGW I

Abbreviations: BGAL, β-galactosidase; BGLU, β-glucosidase; PBGLU, 6-phospho-β-galactosidase; LPH, lactase/phlorzin hydrolase; BGLR, β-glucuronidase; EG, endoglucanase; EX, β(1–3) exoglucanase; XYN, xylanase; CBH, cellobiohydrolase; LAM, laminarinase; GH, β(1–3) and β(1–3),(1–4) glucanase; XLAM, exo-laminarinase; LIC, lichenase. β-glucosidase A from *Bacillus polymyxa*, β-galactosidase from *Escherichia coli*, endoglucanase I from *Bacillus subtilis*, xylA from *Pseudomonas fluorescens* subsp. *cellulosa* and GHS are used as representative sequences for each of the families with common alternative amino acids also given. The residues conserved across the families are in bold. The conserved glutamates are 165 and 352, 461 and 537, 169 and 257, 127 and 246 and 94 and 231.

Among the residues which have been directly shown to form the binding site for xylopentaose in xylA, Gln^{213}, Trp^{305} and Trp^{313}, which form the important subsite E adjacent to the cleaved bond, are equivalent to similar residues in GHS and GHR (Asn^{166}/Asn^{168}, Phe^{274}/Phe^{275} and Phe^{291}/Trp^{291}) and are conserved in family 17 (Fig. 1). Asn^{166} (GHS numbering) forms a hydrogen bond with Glu^{94}, equivalent to the one formed in xylA between Gln^{213} and Glu^{127}. Phe 291 is substituted by tryptophan in two sequences. *Streptomyces lividans* xylanase A can hydrolyse β(1–3) linked xylosidic bonds at a slow rate [17] and family F xylanases in general are capable of accepting glycosyl units at their substrate-binding site when hydrolysing *p*NPC, which also suggests similarity between families 17 and 10.

The structure of *E. coli* β-gal (family 1), also an 8-fold β/α-barrel, has been reported [18]. The barrel of β-gal is very distorted with α-helix 5 missing and β-strand 6 distorted. Catalysis involves Glu^{461} at the carboxy-terminus of β4 and Glu^{537} at the carboxy-terminus of β7. Glu^{537} is probably the nucleophile, while Glu^{461} is likely to bind an essential magnesium ion which may act as an electrophile in place of the proton in xylA (Sinnott, personal communication). Glu^{461} is found within the conserved sequence WSLGNES, which is similar to the corresponding sequence WDVVNEA in xylA (Table 1).

No structures are available for the β-glucosidases of families 1 and the family 5 cellulases. However, it has been suggested that these families are related [19] and that both are β/α-barrels [20]. A β/α-barrel structure for the family 5 cellulases is consistent with the far uv circular dichroism spectrum of a single domain family A cellulase from *Ruminoccus flavofaciens* (Mandy Scott, personal communication). Carboxylates, which have been identified as essential in family 1 and family 5 by s.d.m. [21], or have been proposed as catalytic because they are absolutely conserved [19], are in positions consistent with structurally and functionally equivalent roles to the acid/base and nucleophilic carboxylates in xylA. The similarity of the sequences around the nucleophilic glutamates of β-glucosidases and family F xylanases has already been noted [22]. In Table 1, we have aligned the conserved regions of representative sequences of the families which we propose form the 4/7 superfamily. This superfamily includes glycohydrolases with catalytic carboxylates at the ends of β-strands 4 and 7 of a 8-fold β/α-barrel.

Conserved glutamates close to the ends of β4 and β7 are not a common feature of non-glycosidase 8-fold β/α-barrel enzymes, for example they do not occur in xylose isomerase, ribulose 1,5-bisphosphate carboxylase or triose phosphate isomerase. Although the sequence asparagine-glutamate does occur close to the carboxy-end of β-strands in some non-glycosidase 8-fold β/α-barrel enzymes, such as the tryptophan synthase α-subunit, the bi-functional phospho-ribosyl anthranilate isomerase/indole-3-glycerol phosphate synthase and triose phosphate isomerase it is not conserved in the same enzyme from different species. Thus the conserved asparagine-glutamate close to the end of β4 and glutamate close to the end of β7 in the 4/7 superfamily is distintive and suggestive of divergent evolution from a common ancestor. It is also interesting to note that the family 18 glycosyl hydrolases recently solved, chitinase and hevamine [33,34], do not belong to the 4/7 superfamily as their active site carboxylates are close to the ends of β-strands 4 and 6 and the major axis of the β-barrel runs from β3 not β1.

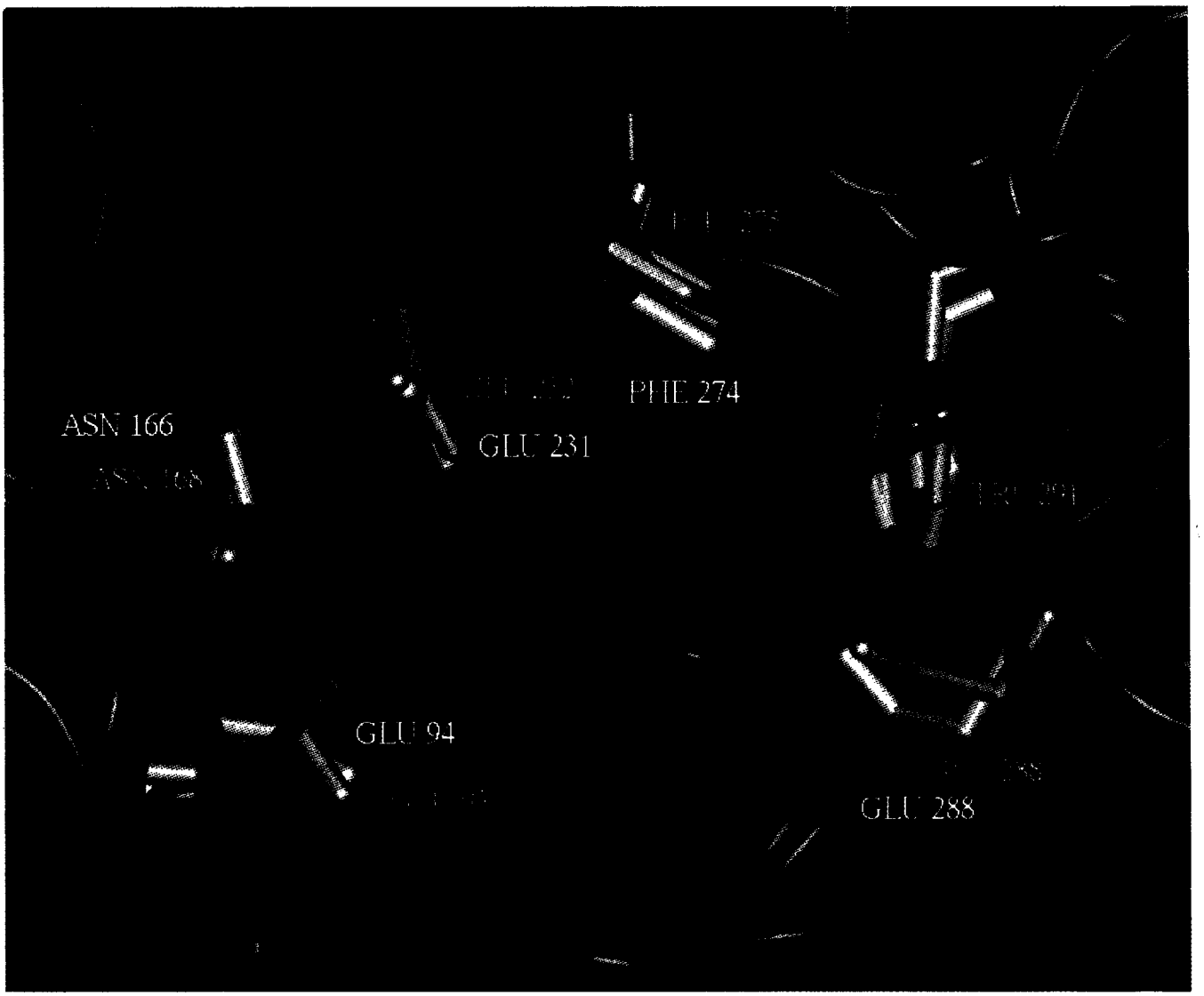

Fig. 1. Superimposition of the active sites of *Pseudomonas* xylA [10] (in red), the $\beta(1-3)$ glucanase (in white) and $\beta(1-3)(1-4)$ glucanase (in yellow) [15]. Distances between the glutamates and important substrate binding residues after superimposition of the structures are given below:

Xylanase	Glucanases (GHS/GHR)	Distance (Å)
E127 (Cδ)	E94/E93 (Cδ)	2.6/2.1
E246 (Cδ)	E231/E232 (Cδ)	1.3/1.2
Q213 (Cδ)	N166/N168 (Cγ)	2.0/1.6
W305 (Cγ)	F274/F275 (Cγ)	1.2/0.6
W313 (Cγ)	F291/W291 (Cγ)	3.6/2.5

The conserved E288 in the barley glucanases is also shown.

The sequence CINDYN is found in the family 5 endoglucanase I of *Butyrivibrio fibrisolvens* and in a number of family 10 sequences. This sequence occurs in a well conserved region of the xylanases at the carboxy-terminus of β-strand 5. In the endoglucanase it is found in approximately the correct position relative to the conserved active site glutamates to allow it to have the same structural and functional role and seems unlikely to have arisen by chance alone. The possibilities are that it is an evolutionary remnant, that the requirements of the structure have caused the sequences to converge or more likely that loops have been swapped during evolution. One can speculate that several of Henrissat's other families may fall into the 4/7 superfamily including families 39 [20] and 42 [3,23].

A comparison of xylA with other β/α-barrels further suggests that our proposed superfamily may be more distantly related to other enzymes. The shape of β/α-barrels has been the basis for their classification in previous papers [24]. The shape of the β/α-barrel in xylA is elliptical in cross section with the major axis running between β-strands one and five (Fig. 2a) and is very similar to that of the barley glucanases (Fig. 2b), but

Fig. 2. Comparison of β/α barrels. All structures were superimposed on the *P. fluorescens* xylA structure and are shown in the same orientation. The central β-barrels are shown in dark grey. The conserved carboxylates are drawn in ball and stick representation and the residue numbers and the preceding β-strand are listed below. The carboxylates are identified clockwise from the top left of each β-barrel. (a) XylA [10] is shown with bound xylopentaose substrate in ball and stick representation, and E246/β7 (nucleophile), E180/β5 (probably contributing to protonation of E127) and E127/β4 (acid/base); (b) GHS [15] is shown with the conserved residues E288/β8, E231/β7 (nucleophile) and E94/β4; (c) *Aspergillus niger* TAKA amylase [25] is shown with conserved residues D297/β7, E230/β5 and D206/β4, which are all demonstrated to be essential by s.d.m. [26]; (d) soybean β-amylase [27] is shown with conserved residues E380/β7, suggested as catalytic from crystallographic evidence, E186/β4 and D101/β3, found essential by s.d.m. [28]; (e) *Bacillus circulans* cyclodextrin glycosyltransferase [29], which has already been identified as similar to α-amylases, is shown with conserved residues D328/β7, E257/β5 and D229/β4, found essential by s.d.m. [26]; (f) xylose isomerase [30] is shown with the metal-binding D245/β7; (g) ribulose 1,5-bisphosphate carboxylate/oxygenase from *Rhodospirillum rubrum* is shown with conserved active site carboxylates D193/β2 and D194/β2, both probably involved in metal-binding [31]; (h) triose phosphate isomerase [32] is shown with the catalytic E165/β6.

clearly different from the circular barrel of ribulose 1,5-bisphosphate carboxylase (Fig. 2g) or the elliptical barrel of triose phosphate isomerase which has its major axis between strands 3 and 7 (Fig. 2h). More similar to xylA in shape are β-amylase and xylose isomerase, which also have their major axis running between strands 1 and 5, but are less elliptical (Fig. 2d and 2f), and to a lesser extent α-amylases and cyclodextrin glycosyltransferases (Fig. 2c,e), which are only slightly elliptical, with their major axis running between strands 2 and 6. The similarity between xylA and β-amylase was initially observed by Liisa Holm and Chris Sander (personal communication). As well as having a somewhat similar shape, both amylases and cyclodextrin glycosyltransferase share with the 4/7 superfamily the general position of catalytic carboxylates at the carboxy-terminus of β4 and β7, and xylose isomerase has a conserved metal-binding Asp at the end of β7 (as detailed in Fig. 2f).

In this paper we have shown structural and functional similarities between the glycohydrolases of family 1, 2, 5, 10 and 17. We suggest they form a superfamily with acid/base (or electrophile) at the end of β4 and nucleophile at the end of β7. We call this family the 4/7 superfamily. We further suggest that other glycohydrolases with the sequence asparagine-glutamate some 100 residues before a glutamate identified as the nucleophile may also fall into the 4/7 superfamily. Alternatively where a glutamate in the sequence asparagine-glutamate is essential for catalysis then a second glutamate some 100 residues later may be the nucleophile. This may help identify the active site residues in newly determined sequences.

Acknowledgements: We thank the EU for supporting Leila Lo Leggio under project AIR2-CT93-1272.

References

[1] Sinnott, M.L. (1990) Chem. Rev. 90, 1171–1202.
[2] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[3] Henrissat, B. (1993) Biochem. J. 293, 781–788.
[4] Blake, C.C.F., Koenig, D.F., Mair, G.A., North, A.C.T., Phillips, D.C. and Sarma, V.R. (1965) Nature 206, 757–761.
[5] Matthews, B.W. and Remington, S.J (1974) Proc. Natl. Acad. Sci. USA 71, 4178–4182.
[6] Weaver, L.H, Grütter, M.G., Remington, S.J., Gray, T.M., Isaacs, N.W. and Matthews, B.W. (1985) J. Mol. Evol. 21, 97–111.
[7] Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. and Robertus, J.D. (1993) J. Mol. Biol. 229, 189–193.
[8] Holm, L. and Sander, C. (1994) FEBS Lett. 340, 129–132.
[9] Thunnissen, A.M.W.H., Dijkstra, A.K., Rozeboom, H.J., Engel, H., Keck, W. and Dijkstra, B.W. (1994) Nature 367, 750–754.
[10] Harris, G.W., Jenkins, J.A., Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G.P., Laurie, J.I., Gilbert, H.J. and Pickersgill, R.W. (1994) Structure 2, 1107–1116.
[11] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–542.
[12] Quanta 4.0, Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA 02154, USA (1994).
[13] Sutcliffe, M.J., Haneef, I., Carney, D. and Blundell, T.L. (1987) Prot. Eng. 1, 377–384.
[14] Kraulis, P. (1991) J. Appl. Cryst. 24, 946–950.
[15] Varghese, J.N., Garrett, T.P.J., Colman, P.M., Chen, L., Høj, P.B. and Fincher, G.B. (1994) Proc. Natl. Acad. Sci. USA 91, 2785–2789.
[16] Chen, L., Fincher, G.B., and Høj, P.B. (1993) J. Biol. Chem., 268, 13318–13326.
[17] Biely, P., Kluepfel, D., Morosoli, R. and Shareck, F. (1993) Biochem. Biophys. Acta. 1162, 246–254.
[18] Jacobson, R.H., Zhang, X.-J., Dubose, R.F., and Matthews, B.W. (1994) Nature 369, 761–766.
[19] Gräbnitz, F., Seiss, M., Rücknagel, K.P. and Staudenbauer, W.L. (1991) Eur. J. Biochem. 200, 301–309.
[20] Belaich, A., Fierobe, H.-P., Baty, D., Busetta, B., Bagnara-Tardif, C., Gaudin, C. and Belaich, J.-P. (1992) J. Bacteriol. 174, 4677–4682.
[21] Chambers, R.S., Walden, A.R., Brooke, G.S., Cutfield, J.F. and Sullivan, P.A. (1993) FEBS Lett. 327, 366–369.
[22] Tull, D., Withers, S.G., Gilkes, N.R., Kilburn, D.G., Warren, R.A.J. and Aebersold, R. (1991) J. Biol. Chem. 266, 15621–15625.
[23] Hirata, H., Fukazawa, T., Negoro, S. and Okada, H. (1986). J. Bacteriol. 166, 722–727.
[24] Farber, G.K. and Petsko, G.A. (1990) Trends Biochem. Sci. 15 228–234.
[25] Swift, H.J., Brady, L., Derewenda, Z.S., Dodson, E.J., Dodson, G.G., Turkenburg, J.P. and Wilkinson, A.J. (1991) Acta Cryst. B 47, 535–544.
[26] Svensson, B. and Søgaard, M. (1993-review) J. Biotec. 29, 1–37.
[27] Mikami, B., Degano, M., Hehre, E.J. and Sacchettini, J.C. (1994) Biochemistry 33, 7779–7787.
[28] Totsuka, A., Nong, V.H., Kadowaka, H., Kim, C.-S., Itoh, Y. and Fukazawa, C. (1993) Eur. J. Biochem. 221, 649–654.
[29] Klein, C. and Schulz, G.E. (1991) J. Mol. Biol. 217, 737–750.
[30] Jenkins, J., Janin, J., Rey, F., Chiadmi, M., van Tilbeurgh, H., Lasters, I., De Maeyer, M., Van Belle, D., Wodak, S., Lauwereys, M., Stansenss, P., Mrabet, N.T., Snaauwaert, J., Matthyssens, G. and Lambeir, A.-M. (1992) Biochemistry 31, 5449–5458.
[31] Schneider, G., Lindqvist, Y. and Lundqvist, T. (1990) J. Mol. Biol. 211, 989–1008.
[32] Banner, D.W., Bloomer, A.C., Petsko, G.A., Phillips, D.C., Pogson, C.I., Wilson, I.A., Corran, P.H., Furth, A.J., Milman, J.D., Offord, R.E., Priddle, J.D. and Waley, S.G. (1975) Nature 255, 609–614.
[33] Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S. and Vorgias, C.E. (1995) Structure 2, 1169–1180.
[34] Terwisscha van Scheltinga, A.C., Kalk, K.H., Beintema, J.J. and Dijkstra, B.W. (1995) Structure 2, 1181–1189.