Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Anisotropic refinement of the structure of *Thermoascus aurantiacus* xylanase I

Susana Teixeira,[a]* Leila Lo Leggio,[b] Richard Pickersgill[c] and Christine Cardin[a]

[a]Chemistry Department, University of Reading, Whiteknights, Reading RG6 6AD, England, [b]Centre for Crystallographic Studies, Chemical Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark, and [c]Molecular and Cellular Biology, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS, England

Correspondence e-mail: s.c.m.teixeira@reading.ac.uk

Received 24 July 2000
Accepted 29 November 2000

PDB Reference: xylanase I, 1fxm.

The isotropic crystallographic model of the structure of xylanase I from *Thermoascus aurantiacus* (TAXI) has now been refined anisotropically at 1.14 Å resolution to a standard residual of $R = 11.1\%$ for all data. TAXI is amongst the five largest proteins deposited in the Protein Data Bank to have been refined with anisotropic displacement parameters (ADPs) at this level of resolution. The anisotropy analysis revealed a more isotropic distribution of anisotropy than usually observed previously. Adding ADPs resulted in high-quality electron-density maps which revealed discrepancies from the previously suggested primary sequences for this enzyme. Side-chain conformational disorder was modelled for 16 residues, including Trp275, a bulky residue at the active site. An unrestrained refinement was consistent with the protonation of the catalytic acid/base glutamate and the deprotonation of the nucleophile glutamate, as required for catalysis. The thermal stability of TAXI is reinterpreted in the light of the new refined model.

1. Introduction

T. aurantiacus xylanase I has a $(\beta\alpha)_8$ TIM-barrel fold and belongs to the family 10 of glycosyl hydrolases (Banner *et al.*, 1975; Jenkins *et al.*, 1995; Pickersgill *et al.*, 1998; Coutinho & Henrissat, 1999). Interest in such enzymes is because of their potential and actual applications in, to name but a few, paper and pulp technology, the recovery of oil from subterranean mines, clarification of juices and wine and the conversion of renewable biomass into liquid fuels (Kulkarni *et al.*, 1999). Understanding the structure of xylanases and how it correlates with their function is important to support studies aiming towards improving and using the properties of these enzymes in practical applications.

It has been observed before that the function and specificity of glycosyl hydrolases cannot be predicted from their overall fold (Davies & Henrissat, 1995). Instead, it is governed by subtle details of their three-dimensional structures. The same can be said in general about their thermostability. Previous studies have shown that the same basic fold, characteristic of a particular protein family, was present in thermophilic proteins and their mesophilic counterparts (Szilágyi & Závodszky, 2000). The higher precision of the model obtained through anisotropic refinement may be important to provide further insight into some more of these subtle details that can determine protein function, specificity and thermostability.

TAXI catalyses the hydrolysis of xylan, which represents the major group of hemicelluloses. It is an interesting xylanase as it has been shown to have a high degree of thermal stability, high activity and high specificity towards xylan. These char-

acteristics may be very useful in future applications in the pulp and paper industries, where reduction in the use of chlorine as a bleaching agent is being imposed by environmental regulations. Furthermore, TAXI does not possess the contaminating cellulase activity that is a characteristic of most fungal xylanases (Natesh *et al.*, 1999; Subramaniyan & Prema, 2000) and makes them less suitable for practical applications in biobleaching.

Previous crystallographic studies determined the structure of TAXI at 1.14 Å resolution (Lo Leggio *et al.*, 1999). The structure was then refined isotropically using *X-PLOR*. Although the near-atomic resolution data had reasonable quality, sequence ambiguities remained and there was still a lack of map quality to model some disordered residues. Anisotropic refinement, together with the availability of the DNA sequence (Bousson & Parriche, unpublished work), have resolved the ambiguities so that a detailed model for family 10 xylanases is now available.

2. 'Refinement' of the primary sequence

The amino-acid sequence of TAXI has been a subject of discussion since it was first published nine years ago (Srinivasa *et al.*, 1991). It was first obtained by chemical sequencing of proteolytic peptides but proved to be incorrect, as shown later by X-ray crystallography aided by multiple sequence alignment with other homologous xylanases. Through these methods two different sequences became available in 1999, with the structures deposited in the Protein Data Bank (Berman *et al.*, 2000) by Lo Leggio *et al.* (1999) (accession codes 1tax and 1tix; strain IMI 216529; sequence available from the SWISS-PROT database, accession code P23360) and by Natesh *et al.* (1999) (accession code 1tux; strain isolated from Indian soil). More recently, work by Bousson & Parriche (unpublished work; SPTREMBL accession code Q9UQZ4) on DNA sequencing provided further corrections to the primary sequence which were used to build the initial model for the anisotropic refinement.

It seems plausible to suggest that there are two discrepancies from the sequence deposited by Bousson & Parriche, which might be a consequence of genuine strain differences. The first relevant difference concerns Pro245. In the available sequence, this residue corresponds to a serine. However, the proline residue fits perfectly in this very clear region of the electron-density maps (Fig. 1). Neither the difference electron density nor the omit maps show any evidence for a serine residue at this position. Furthermore, it has been suggested before that the presence of prolines at three of the α-helices' N-termini in TAXI may be important to stabilize the protein's folded state (Lo Leggio *et al.*, 1999). This will be discussed further in §2.4. The second discrepancy was found for residue 217, which would be a serine according to the available sequence. However, a glycine fitted very well, whereas a serine generated negative $F_o - F_c$ electron-density maps and would not explain the shape of the omit maps calculated. The presence of a glycine at this position is supported by the presence of glycine in the sequence of the homologous family 10 fungal xylanases, with which TAXI has been shown to have a high sequence identity (see Fig. 2). Finally, the sequence deposited by Bousson & Parriche also includes an additional glutamine at the C-terminus. This region of the electron-density map is not clear. If the glutamine is present it is either statically disordered or has considerable thermal motion.

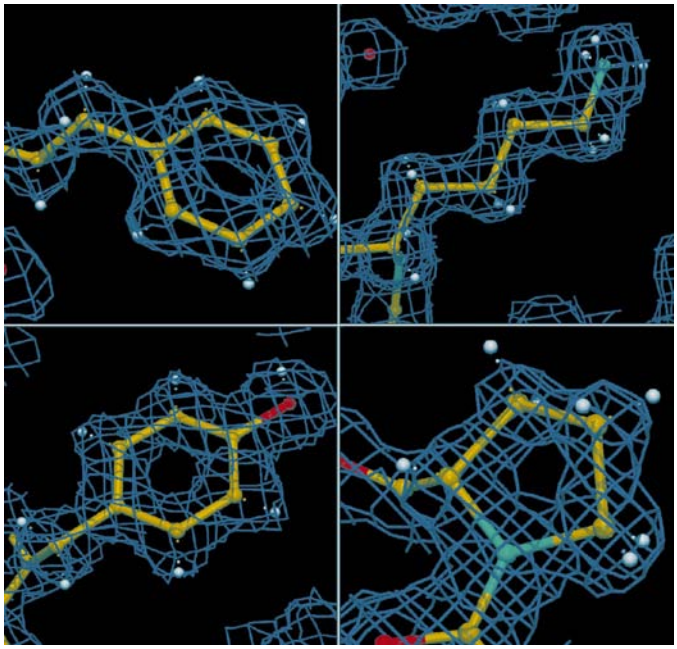

Figure 1
$2F_o - F_c$ electron-density maps contoured at the 1σ level for the side chains of residues Phe40 (top left), Lys50 (top right), Tyr170 (bottom left) and Pro245 (bottom right). H atoms are coloured in white.

2.1. Protonation state of the glutamates at the active site

A distance of 5.5 Å was observed between OE2 of the acid/base Glu131 and OE2 of the nucleophile Glu237. This is consistent with the retention of anomeric configuration during xylan hydrolysis. In the active enzyme and in the absence of xylan, the acid/base glutamate should be protonated while the nucleophile glutamate should be deprotonated (Sinnot, 1990; Davies & Henrissat, 1995).

The protonation states of the active-site glutamates were studied in an unrestrained refinement. In Glu131, the distances CD—OE1 and CD—OE2 are 1.325 ± 0.019 Å and 1.190 ± 0.018 Å, respectively. The target value in a restrained refinement (in *SHELX*) for equal anionic glutamate bond lengths is 1.249 Å. According to the values in the literature for bond lengths in crystalline organic compounds, the mean C—OH and C═O bond lengths in glutamic acid are 1.308 ± 0.019 and 1.214 ± 0.019 Å, respectively (Lide, 1991). This shows the protonation of OE1 in Glu131. For Glu237, bond lengths closer to the target value were found: 1.265 ± 0.016 Å for the CD—OE2 bond and 1.298 ± 0.017 Å for the CD—OE1 bond. These two bond lengths are more similar to each other, as expected for a delocalized 'double' bond character. Although the quality of the sharpened electron-density maps was not good enough to reveal H-atom positions

on these two glutamates, the C—O bond-length differences are appropriate. This evidence is only suggestive, as the H atoms cannot be directly located and the hydrogen-bonding pattern is ambiguous.

Given the positional standard uncertainties (s.u.s), the bond lengths measured seem to be accurate enough to provide experimental evidence of the protonation states of Glu131 and Glu237, supporting the proposed mechanism of catalysis, which predicts retention of configuration. There seems to be no reason to assume that the s.u.s were seriously underestimated, despite the fact that the block-diagonal calculation is just an approximation to the full-matrix least-squares refinement. Previous studies have shown that there is only a difference of 1% between calculations of the s.u.s performed with and without thermal parameters (McRee, 1999). Furthermore, the percentage of underestimation of the s.u.s owing to the use of the blocks should be minimum, as only one block containing all atoms was used.

It is reasonable to expect that with the increasing data quality and computing power, the s.u.s will be systematically determined for each new structure determination. Together with high-quality electron-density maps, this should provide enough evidence for X-ray crystallography to become a standard tool in studies of protonation states of atoms in macromolecules, as already happens with small molecules.

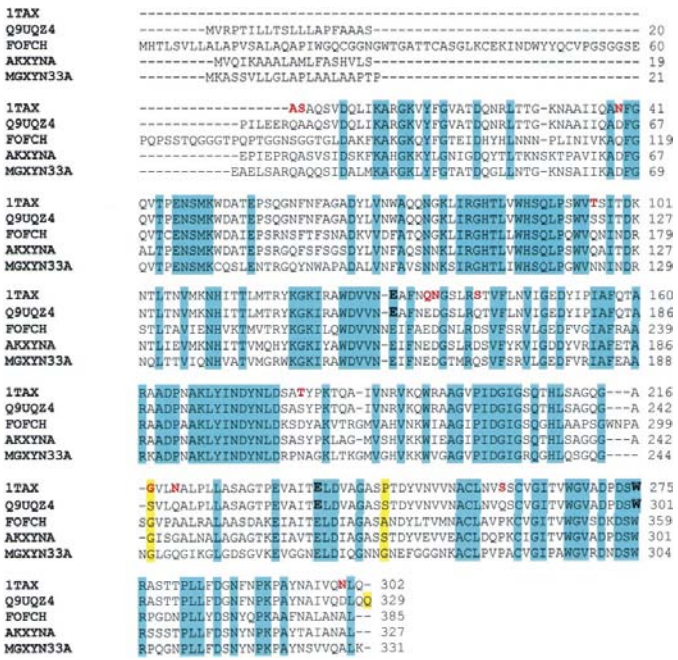

Figure 2
Multiple alignment of the sequences of 1tax (*T. aurantiacus* xylanase I obtained by Lo Leggio *et al.*, 1999; PDB), Q9UQZ4 (*T. aurantiacus* xylanase deposited by Bousson & Parriche; SPTREMBL), FOFCH (*Fusarium oxysporum* putative cellulase/xylanase; Sheppard *et al.*, 1994; EMBL), AKXYNA (*Aspergillus kawachii* xylanase A; Ito *et al.*, 1992; EMBL) and MGXYN33A (*M. grisea* xylanase B; Wu *et al.*, 1995; EMBL). The sequences are identified by their accession codes in the databases. The residues where mutations had to be made to the initial model are marked in red and the sequence discrepancies are highlighted in yellow. The active-site glutamates and tryptophan are in bold.

2.2. Side-chain disorder

The clear improvement of the quality of the maps after ADPs were introduced made possible a more precise fitting of side-chain disorder and solvent modelling. Alternate conformations previously not modelled were introduced for residues 8, 29, 46, 102, 109, 115, 180, 185, 259 and 275. Double conformations for residues 223 and 235 modelled by Lo Leggio *et al.* (1999) were removed, as the electron-density maps calculated clearly show that only one conformation could be fitted. The anisotropic displacement of the side-chain atoms in these two residues did not cover the space previously used to model the multiple conformations isotropically.

The most striking case is Trp275, where alternate conformations were clearly visible even though the electron density was not perfectly continuous. It was also possible to see its anisotropic movement. The occupancies of the two conformers were refined, showing just a slightly higher occupancy factor for one of them (0.52). This residue is thought to be very important in substrate binding to TAXI (see Fig. 3) by closing in on xylan (Lo Leggio *et al.*, in preparation). The alternate conformations and anisotropic movement suggest how and where the displacement of this residue takes place. It has been remarked before that the different modes of action of xylanases depend on subtle differences near the active site (Natesh *et al.*, 1999; Kulkarni *et al.*, 1999). The anisotropic behaviour and side-chain disorder of this residue might just be one such subtle difference from the previous models for TAXI.

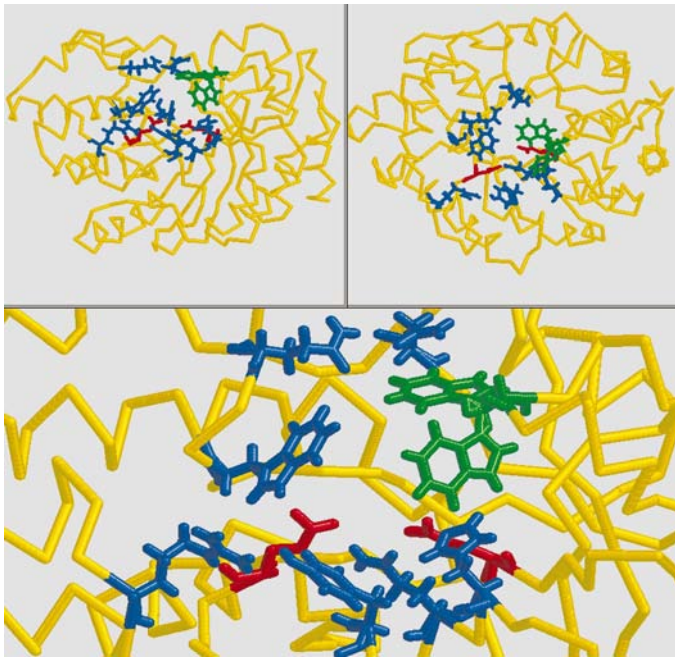

Figure 3
In the top-left corner a side view of the $(\beta\alpha)_8$ TIM-barrel fold of TAXI can be seen. A zoom on the active site shows the side chains of some selected residues (blue), illustrating the environment around the active-site glutamates (red). The alternate conformations of Trp275 are coloured in green. The top right corner shows a top view along the barrel.

Table 1
Final refinement statistics.

Refinement statistics	Isotropic	Anisotropic
Atoms		
Protein	165	171
Solvent	165	171
H atoms	—	2326
Resolution range (Å)	39.5–1.14	10–1.14
R factor (%, for all data)	18	11.1
Free *R* factor (%, for all data)	20.1	13.7
Goodness of fit (*SHELX*)	4.63	1.38
Restrained goodness of fit	4.50	1.18
R.m.s. deviations		
Bond length (Å)	0.009	0.015
Bond angles (Å)	1.493	2.205
Isotropic B_{eq} values (Å^2)		
Main chain	—	1.044
Side chain	—	3.080
Average *B* values (Å^2)	*B* value	Isotropic B_{eq}
Main chain	11.08	12.51
Side chain	13.21	17.65
Overall average *G* factor†	0.36	0.03
Double conformations modelled	8	16

† Obtained from *PROCHECK*.

2.3. Quality of the model

Table 1 lists the final refinement statistics. Judging by the overall improvement of the final statistics the anisotropic model seems to fit the data better, as expected considering the mean anisotropy of TAXI. Isotropic B_{eq} values from this refinement are higher than the *B* values obtained for the isotropic model (also obtained from data measured at 293 K). This indicates that the protein atoms undergo significant displacements, which might have been too restricted to an isotropic behaviour in the isotropic refinement. More contributions for the higher value of the isotropic B_{eq} can come from the extra conformational disorder modelled, as well as from the different restraints and/or the bulk-solvent modelling.

For residues at the active site and some residues directly connected to these, some strained geometry was observed. It is noteworthy that the nucleophile Glu237 was one such residue but Glu131 was not. It has been pointed out previously (Lawson *et al.*, 1997) that for the xylanase activity to be maintained there are positional requirements for the nucleophile but not for the acid/base glutamate.

2.3.1. Anisotropy analysis. Despite the relatively low completeness, the data quality in the 10–1.14 Å resolution range was high and proved to be sufficient to obtain very clear electron-density maps (see Fig. 1). These showed an overall agreement with the initial model, but some uncertainties remained until ADPs were added to the refinement, causing a drop of 4.8 and 4.5% in *R* and R_{free}, respectively. The mean anisotropy (ratio of the smallest and largest eigenvalues of the ADP matrix) is 0.54 for the protein and 0.52 for the solvent, with standard deviations of 0.16 and 0.12, respectively. Despite this considerable deviation from isotropy it is still a slightly higher value than expected, since previous studies showed a tendency for proteins with more than 50 residues to exhibit a mean anisotropy of 0.45 with a standard deviation of 0.15 (Merritt, 1999).

Table 2
Restraints tested: their effect on the residual *R* factors and on the mean anisotropy of the structure.

Restraints (*SHELXL*)		*R* factor (%, for all data)	R_{free} (%, for all data)	Mean anisotropy
All non-H atoms	Waters			
DELU 0.01/SIMU 0.1	ISOR 0.05	11.1	13.7	0.53
DELU 0.01/SIMU 0.08	ISOR 0.05	11.2	13.9	0.54
DELU 0.01/SIMU 0.04	ISOR 0.05	11.3	13.9	0.54
DELU 0.01/SIMU 0.02	ISOR 0.05	11.6	13.8	0.56
DELU 0.01/SIMU 0.02/ ISOR 0.1	ISOR 0.05	11.6	13.8	0.56
DELU 0.01/SIMU 0.01	ISOR 0.05	12.2	14.2	0.59
DELU 0.02/SIMU 0.02	ISOR 0.05	11.5	13.8	0.56
DELU 0.02/SIMU 0.1	ISOR 0.05	11.2	13.9	0.54
DELU 0.02/SIMU 0.1/ ISOR 0.08	ISOR 0.05	11.6	13.9	0.57
DELU 0.02/SIMU 0.015	ISOR 0.05	11.7	13.9	0.57

The distribution of anisotropy among the protein atoms of TAXI shows a deviation from the typical more anisotropic distribution curve. This same behaviour has been observed previously (Merritt, 1999) for the atomic resolution structures of *Bacillus lentus* subtilisin and *Penicillium janthinellum* penicillopepsin. Both these cases and TAXI could be genuine outliers for which the distribution of anisotropy serves a structural or/and functional purpose. Another possible explanation is that the deviation might result from the refinement procedure (more precisely, from the restraints used). To test this hypothesis, several refinements were performed (see Table 2) applying different strengths to the DELU, SIMU and ISOR restraints in *SHELXL* (see Sheldrick & Schneider, 1997). The runs were all made from the same starting point. The R_{free} and *R* residuals showed no significant falls and the mean anisotropy did not decrease, neither did the distribution of anisotropy deviate significantly towards a more anisotropic behaviour of the protein atoms. The initial DELU, SIMU and ISOR restraints thus seem appropriate, but the tests were not exhaustive nor were they repeated at other stages of the refinement.

It becomes apparent that with the information available so far, validation tools based on the analysis of protein anisotropy must still be approached with some caution. Further studies on this subject are certainly required and the effects of the refinement procedure on the accuracy of the results remain to be fully understood and assessed. The number of available structures currently deposited in the Protein Data Bank with more than 250 residues and refined with ADPs at high to atomic resolution is still quite small (see Table 3). To our knowledge, there are only four other structures in the Protein Data Bank larger than TAXI and refined anisotropically with atomic resolution.[1] However, the other three high-resolution structures also listed in Table 3 are certainly

[1] While this paper was being processed, Ducros *et al.* (2000) have independently reported the structure of *Streptomyces lividans* xylanase 10A, refined anisotropically to comparable resolution.

Table 3
List of structures deposited at the Protein Data Bank (at the time of writing) refined with ADPs and with size/resolution comparable to TAXI.

Protein	Residues in the a.u.	Resolution (Å)	R_1 (%)	PBD code; reference
Streptavidin (protein binding)	508	1.14	12.6	1swu; Freitag *et al.* (1999).
Salmonella typhimurium sialidase	379	1.05	11.6	3sil; Garman *et al.*, unpublished work.
P. janthinellum penicillopepsin	323	0.95	10.0	1bxo; Khan *et al.* (1998).
Phosphate-binding protein	321	0.98	11.7	1ixh; Wang *et al.* (1997).
T. aurantiacus xylanase	302	1.14	11.1	1fxm; this work.
Endonuclease IV (hydrolase)	285	1.02	12.4	1qtw; Hosfield *et al.* (1999).
Serine protease	269	0.78	10.1	1gci; Kuhn *et al.* (1998).
Cholera toxin B-pentamer	515	1.25	13.0	3chb; Merritt *et al.* (1998).

comparable owing to the relatively low completeness of the data for TAXI.

It is a fact that the number of structures refined with ADPs is increasing. The growing number of high-quality structures will surely bring important information to the parameterization used in refinement programs and to future studies on anisotropy and refinement protocols.

2.4. A structural basis for the thermostability of xylanases

In the previous paper on the structure determination of TAXI, it was suggested that the main structural determinants of thermostability among family 10 xylanases were the efficient packing of the hydrophobic core and the stabilization of the barrel α-helices, while the number of salt bridges was relatively unimportant (Lo Leggio *et al.*, 1999). This view remains unaltered by the correction of the TAXI sequence in this paper. In contrast, Natesh *et al.* (1999) focus on the general importance of salt bridges in the thermal stabilization of TAXI. Among others, they suggest that a particular salt bridge between Arg24 and Glu232 'cross-linking' β-strands 4 and 7 of the barrel (the strands bearing the catalytic residues) is very important. However, it should be noted that the salt bridge is also likely to be conserved in the highly homologous (around 60% sequence identity, including the required arginine and glutamate residues) *Magnaporthe grisea* xylanase (gi:AAC41684; Wu *et al.*, 1995) and *Hypocrea jecorina* xylanase III (Xu *et al.*, 1998), neither of which, to our knowledge, are thermostable.

The comparison of salt bridges in five xylanases (Lo Leggio *et al.*, 1999) was repeated to take into account the corrected sequence. The study was also extended to include three different distance cutoffs (3, 4 and 5 Å), either considering histidine residues as positively charged or not (details not shown). The results confirm the observation that the two thermostable family 10 xylanases TAXI and *Clostridium thermocellum* xylanase Z (Dominguez *et al.*, 1995) do not have more salt bridges than the mesophilic xylanases. While studies on large subsets of structures (Vogt *et al.*, 1997; Szilágyi & Závodszky, 2000) do indeed show that the number of salt bridges plays a prominent role as effector of thermostability, they also show that the mechanisms of thermal stabilization vary considerably from protein to protein.

There is a large body of literature highlighting the role of helix stabilization in thermostability, looking at different features of stabilizing factors (for example, Facchiano *et al.*, 1998). The stabilization of the helix dipole by charged side chains (Nicholson *et al.*, 1988) as well as the preferential presence of prolines at the N-termini of α-helices (Richardson & Richardson, 1988; Kim & Kang, 1999) have been identified previously as two helix-stabilizing factors in family 10 thermostable xylanases (Lo Leggio *et al.*, 1999). The data have now been re-evaluated according to the corrected sequence of TAXI and looking at the combined effect of these two factors in the helices forming the TIM barrel. Fig. 4 shows that the thermostable members of family 10 have a higher number of stabilized barrel helices and a lower number of destabilized barrel helices than their mesophilic counterparts (only taking these two factors into account). Helix stabilization alone may not be sufficient to increase thermostability, but it stands to reason that a thermostable protein needs thermostable helices. The importance of short 'thermo-helices' for the mechanisms of thermostability suggested by Natesh *et al.* (1999) does not seem convincing, since most of these helices are found in the *Streptomyces lividans* xylanase (Derewenda *et al.*, 1994) and all of them are found in the *Penicillum simplicissimum* structure (Schmidt *et al.*, 1998).

According to our analysis, the combination of better hydrophobic packing and higher stability of the barrel helices remains the most likely effector of thermostability in TAXI, although the salt bridge between β-strands 4 and 7 might also play a role.

3. Materials and methods

3.1. Data

TAXI crystallizes in space group $P2_1$, with unit-cell parameters a = 51.039, b = 68.302, c = 41.440 Å, β = 113.87°. The data used for the refinement consisted of the structure-factor file downloaded from the Protein Data Bank (accession code 1tax). The data were collected at 293 K using synchrotron radiation at beamline 9.6 at the SRS, Daresbury, England. It was obtained by Lo Leggio *et al.* (1999) by merging two data sets collected from two crystals of TAXI (over the resolution range 9.5–1.14 Å) using the program *SCALEPACK* (Otwinowski & Minor, 1997). This was performed in order to increase the overall completeness. The data corresponds to 80 628 unique reflections, with a mean $I/\sigma(I)$ of 9.8. The overall multiplicity of the data is 2.9, the $R_{\rm sym}$ is 7.8% and the completeness is 85%. In the resolution shell between 1.15 and 1.14 Å the multiplicity is 2.1, the $R_{\rm sym}$ is 23.8 and 11.5% of the

Table 4
Reflection statistics for the data used in the refinement.

Systematic absences are taken into account in calculating the percentages. The percentage with $I > 2\sigma$ is expressed relative to the theoretical number of unique reflections.

Resolution (Å)	N (unique)	Mean (I/s)†	$I > 2s$ (%)	Completeness (%)	Scale‡	R_1§ (all)	R_1 ($I > 2s$)
10.000–2.350	10217	17.39	93.83	94.43	1.003	0.135	0.134
2.350–1.870	10230	15.08	93.86	95.58	1.022	0.103	0.102
1.870–1.630	10020	11.59	88.29	91.62	1.029	0.094	0.092
1.630–1.466	10022	10.44	78.12	82.47	1.021	0.078	0.076
1.466–1.354	9767	8.24	74.02	81.68	1.019	0.083	0.079
1.354–1.266	10181	6.39	68.89	80.74	1.014	0.090	0.082
1.266–1.198	9954	4.73	65.08	80.33	1.015	0.098	0.086
1.198–1.140	10117	3.99	58.99	77.73	0.999	0.113	0.095
All data	80508	9.76	76.66	84.98	1.013	0.111	0.108

† $s = \sigma(I)$. ‡ $R_1 = \sum(|F_o - F_c|)/\sum(F_o)$. § Scale $= \sum(F_o)/\sum(F_c)$.

reflections have $I/\sigma(I) > 5.0$. The completeness in this shell is 78.8%.

Table 4 shows the reflection statistics for the data used in the refinement. Since the effective completeness is relatively low the data is referred to as having near-atomic resolution, despite the fact that the generally accepted threshold of 1.2 Å has been overtaken in the last resolution shell.

3.2. Model and refinement

The model for the refinement was based on the structure of TAXI obtained through isotropic refinement as described previously by Lo Leggio *et al.* (1999). Anisotropic refinement was carried out using *SHELXL* and the *SHELXPRO* interface (97-2 version; Sheldrick & Schneider, 1997), with a data-to-parameter ratio of 3.5 in the final refinement. Manual fitting of the model into $2F_o - F_c$ (1σ contour) and $F_o - F_c$ (3σ contour) maps was performed with the program *XTALVIEW*, release 3.2.1 (McRee, 1992).

5% of the reflections were used to calculate the $R_{\rm free}$ residual and were not included in the refinement at any stage. The initial model contained no waters and no side-chain disorder. It included the corrections to the amino-acid sequence according to the latest data deposited by Bousson & Parriche (SPTREMBL accession code Q9UQZ4). In the third run of the refinement, some of the side-chain conformational disorder was already quite clear and partly modelled. ADPs were added in the fourth run. In the following five runs, the side-chain disorder was modelled and minor fittings were made. In run 7, the water molecules were allowed to have half occupancies. Sharpened difference maps were calculated using the sharpening option in *SHELXPRO*, which multiplies the coefficients of the difference electron-density synthesis by $\langle F^2 \rangle^{1/4}$, where F^2 is the mean reflection intensity in the appropriate resolution shell (Sheldrick & Schneider, 1997). The sharpened maps did not show very clear hydrogen peaks in run 10. However, riding H atoms were added in the next run and caused a fall of approximately 1.3% in R and 1.5% in $R_{\rm free}$. The default weighting scheme of *SHELXL* was retained until run 12. From run 13 onwards, the recommended scheme output by the program was adopted (WGHT 0.2; see Sheldrick & Schneider, 1997). The refinement was completed after 18 runs, the last of which consisted mainly of minor fittings and solvent modelling. Although the resolution range used until run 16 was 39.5–1.14 Å, a closer look at the reflection data statistics clearly showed that the few data (120 from a total of 80 628 unique reflections) beyond 10 Å were not reliable, according to the corresponding scale and R factor. This might have been because of backstop-shadow problems and/or saturation. The range was therefore changed to 10–1.14 Å

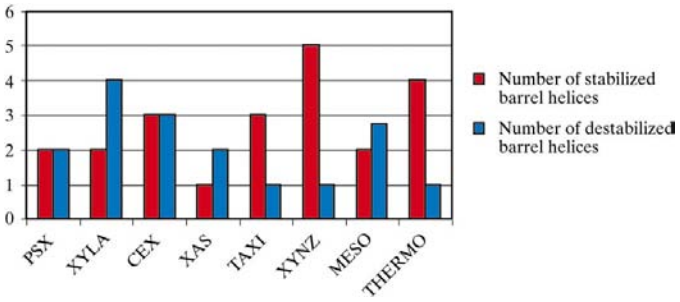

Figure 4
Barrel helices stabilized and destabilized in various family 10 xylanases. See §3 for the definition of stabilization. PSX is *P. simplicissimum* xylanase (Schmidt *et al.*, 1998), XYLA is *Pseudomonas fluorescens* xylanase A (Harris *et al.*, 1994), CEX is *Cellulomonas fimi* xylanase Cex (White *et al.*, 1994), XAS is *S. lividans* xylanase A (Derewenda *et al.*, 1994), TAXI is *T. aurantiacus* xylanase I and XYNZ is *Clostridium thermocellum* xylanase Z (Dominguez *et al.*, 1995). THERMO is the average of the two thermophilic enzymes (TAXI and XYNZ); MESO is the average of the remaining mesophilic enzymes.

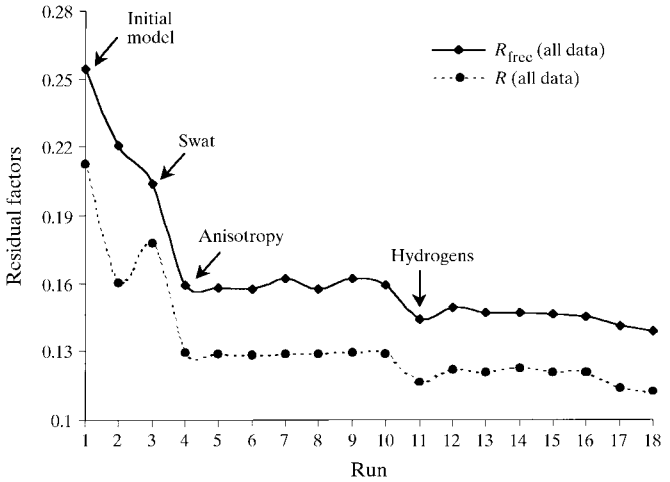

Figure 5
Progress of the refinement. The residual R factor is defined as $R = \sum ||F_o| - |F_c||/\sum |F_o|$.

(despite the possible implications in the solvent-information content of the data), without any major effects on the residuals R_{free} and R. DELU and SIMU restraints were applied to all protein non-H atoms within an effective standard deviation of 0.01 and 0.1 Å, respectively. ISOR restraints were also applied to solvent atoms, within an effective standard deviation of 0.05 Å. The SWAT option in *SHELXL* was used in order to model diffuse solvent (see Sheldrick & Schneider, 1997). Fig. 5 shows the progress of the refinement.

A block-diagonal unrestrained refinement was performed with the program *SHELXH*, using one block for all atoms and retaining the positional parameters only. The block approximation greatly reduced the required computer memory, making it possible to perform the calculation and obtain the s.u.s.

The sequences used in Fig. 2 were obtained through the Sequence Retrieval System Network Browser (SRS Version 6.1) for Databanks in Molecular Biology. The multiple sequence alignment was performed with *ClustalW* v.1.8 (Thompson *et al.*, 1994) through the WWW service at the European Bioinformatics Institute. The final TAXI model was deposited in the Protein Data Bank.

3.3. Validation

The program *PARVATI* (Merritt, 1999) was used on-line to analyse the final distribution of anisotropy among the protein and the solvent.[2] The analysis of *B* factors and some of the PDB file manipulations were performed using *BAVERAGE* and *PDBSET* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994), respectively. *PROCHECK* v.3.5 was used to perform the geometric analysis of the final model (Laskowski *et al.*, 1993).

The visual inspection of the structure was performed with *XTALVIEW*, release 3.2.1 (McRee, 1992). Fig. 1 was made with *Raster3D* (Merritt & Bacon, 1997).

3.4. Helix analysis

Only helices belonging to the TIM-barrel fold and consisting of at least six residues were considered. Helix definitions were taken from *PROMOTIF* (Hutchinson & Thornton, 1996). Proline at the N-terminal position, aspartate or glutamate as one of the first three residues or lysine, arginine or histidine as one of the last three residues were defined as favourable interactions; aspartate or glutamate as one of the last three residues or lysine, arginine or histidine as one of the first three residues were defined as unfavourable interactions. A helix was considered stabilized if the number of favourable interactions was more than the number of unfavourable interactions and destabilized if the reverse was true.

[2] Supplementary material has been deposited in the IUCr electronic archive (Reference: gr2088). Services for accessing these data are described at the back of the journal.

ST is grateful to the Chemistry Department of the University of Reading for financial support and to Dr E. Merritt and Dr G. Sheldrick for their invaluable advice. LL thanks the Danish National Research Foundation for financial support and Dr Anne Mølgaard and Henning Osholm Sørensen for helpful discussions.

References

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. & Waley, S. G. (1975). *Nature (London)*, **255**, 609–614.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coutinho, P. M. & Henrissat, B. (1999). *Carbohydrate-Active Enzymes Server*, http://afmb.cnrs-mrs.Fr/~pedro/CAZY/db.html.

Davies, G. & Henrissat, B. (1995). *Structure*, **3**, 853–859.

Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Shareck, F., Kluepfel, D. & Derewenda, Z. S. (1994). *J. Biol. Chem.* **33**, 20811–20814.

Dominguez, R., Souchon, H., Spinelli, S., Dauter, Z., Wilson, K. S., Chauvaux, S., Beguin, P. & Alzari, P. M. (1995). *Nature Struct. Biol.* **2**, 569–576.

Ducros, V., Charnock, S., Derewenda, U., Derewenda, Z., Dauter, Z., Dupont, C., Shareck, F., Morosoli, R., Kluepfel, D. & Davies, G. (2000). *J. Biol. Chem.* **275**(30), 23020–23026.

EU 3-D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.

Facchiano, A. M., Colonna, G. & Ragone, R. (1998). *Protein Eng.* **11**, 753–760.

Freitag, S., Le Trong, I., Klumb, L. A., Stayton, P. S. & Stenkamp, R. E. (1999). *Acta Cryst.* D**55**, 1118–1126.

Harris, G. W., Jenkins, J. A., Connerton, I., Cummings, N., Lo Leggio, L., Scott, M., Hazlewood, G. P., Laurie, J. I., Gilbert, H. J. & Pickersgill, R. W. (1994). *Structure*, **15**, 1107–1116.

Hosfield, D. J., Guan, Y., Haas, B. J., Cunningham, R. P. & Tainer, J. A. (1999). *Cell*, **98**(3), 397–408.

Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.

Ito, K., Ikemasu, T. & Ishikawa, T. (1992). *Biosci. Biotechnol. Biochem.* **56**, 906–912.

Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995). *FEBS Lett.* **362**, 281–285.

Khan, A. R., Parrish, J. C., Fraser, M. E., Smith, W. W., Bartlett, P. A. & James, M. N. (1998). *Biochemistry*, **37**, 16839–16845.

Kim, M. K. & Kang, Y. K. (1999). *Protein Sci.* **8**, 1492–1499.

Kuhn, P., Knapp, M., Soltis, S. M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.

Kulkarni, N., Shendye, A. & Rao, M. (1999). *FEMS Microbiol. Rev.* **23**, 411–456.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Lawson, S. L., Wakarchuck, W. W. & Withers, S. G. (1997). *Biochemistry*, **36**, 2257–2265.

Lide, D. R. (1991). *CRC Handbook of Chemistry and Physics*, 72nd ed., pp. 9-2–9-15. Boca Raton/Ann Arbor/Boston: CRC.

Lo Leggio, L., Kalogiannis, S., Bhat, M. K. & Pickersgill, R. W. (1999). *Proteins Struct. Funct. Genet.* **36**, 295–306.

McRee, D. E. (1992). *J. Mol. Graph.* **10**, 44–46.

McRee, D. E. (1999). *Practical Protein Crystallography*, 2nd ed., pp. 271–325. San Diego: Academic Press.

Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.

Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.

Merritt, E. A., Kuhn, P., Sarfaty, S., Erbe, J. L., Holmes, R. K. & Hol, W. G. (1998). *J. Mol. Biol.* **282**, 1043–1059.

Natesh, R., Bhanumoorthy, P., Vithayathil, P. J., Sekar, K., Ramakumar, S. & Viswamitra, M. A. (1999). *J. Mol. Biol.* **288**, 999–1012.
Nicholson, H., Becktel, W. J. & Matthews, B. W. (1988). *Nature (London)*, **336**, 651–656.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pickersgill, R., Harris, G., Lo Leggio, L., Mayans, O. & Jenkins, J. (1998). *Biochem. Soc. Trans.* **26**, 190–198.
Richardson, J. S. & Richardson, D. C. (1988). *Science*, **240**, 1648–1652.
Schmidt, A., Schlacher, A., Steiner, W., Schwab, H. & Kratky, C. (1998). *Protein Sci.* **7**, 2081–2088.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Sheppard, P. O., Grant, F. J., Oort, P. J., Sprecher, C. A., Foster, D. C., Hagen, F. S., Upshall, A., McKnight, G. L. & O'Hara, P. J. (1994). *Gene*, **150**(1), 163–167.
Sinnot, M. L. (1990). *Chem. Rev.* **90**, 1171–1202.
Srinivasa, B. R., Swaminathan, K. R., Ganapathy, C., Roy, R. P., Murthy, S. K. & Vithayathil, P. J. (1991). *Protein Seq. Data Anal.* **4**, 15–20.
Subramaniyan, S. & Prema, P. (2000). *FEBS Microbiol. Lett.* **183**(1), 1–7.
Szilágyi, A. & Závodszky, P. (2000). *Structure*, **8**, 493–504.
Thompson, J. D., Higgins, D. J. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vogt, G., Woell, S. & Argos, P. (1997). *J. Mol. Biol.* **269**, 265–271.
Wang, Z., Luecke, H., Yao, N. & Quiocho, F. A. (1997). *Nature Struct. Biol.* **4**, 519–522.
White, A., Withers, S. G., Gilkes, N. R. & Rose, D. R. (1994). *Biochemistry*, **33**, 12546–12552.
Wu, S. C., Kauffmann, S., Darvill, A. G. & Albersheim, P. (1995). *Mol. Plant Microbe Interact.* **8**, 506–514.
Xu, J., Takakuwa, N., Nogawa, M., Okada, H. & Morikawa, Y. (1998). *Appl. Microbiol. Biotechnol.* **49**, 718–724.