Prediction and Fourier-transform infrared-spectroscopy estimation of the secondary structure of a recombinant β -glucosidase from *Streptomyces* sp. (ATCC 11238)

Josep A. PEREZ-PONS,* Esteve PADROS† and Enrique QUEROL*‡

*Institut de Biologia Fonamental and Departament de Bioquímica i Biologia Molecular and †Unitat de Biofísica, Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

The secondary structure of a recombinant β -glucosidase (EC 3.2.1.21) from *Streptomyces* sp. (ATCC 11238) has been predicted by computer algorithms and also estimated by Fourier-transform IR spectroscopy. From curve fitting of the deconvoluted IR spectra, the most probable distribution of the

secondary-structural classes appears to be about 34% α -helix, 30% β -sheet, 25% reverse turns and 11% non-ordered structures. These data showed a good agreement with data from computer prediction (35% α -helix, 23% β -sheet, 31% reverse turns and 11% non-ordered structures).

INTRODUCTION

β -Glucosidases (β -glucoside glucohydrolase, EC 3.2.1.21) are enzymes that catalyse the hydrolysis of a β -D-glucosidic bond by transfer of glucosyl groups between oxygen nucleophiles. β -Glucosidases are involved in different metabolic roles, such as utilizing primary carbon sources, defence against phytopathogens and prevention of the accumulation of certain compounds (e.g. glucocerebrosides) in tissues. However, β -glucosidases, more specifically cellobiases, have been traditionally referred to as one of the three enzyme classes comprising the 'cellulase' system, catalysing the hydrolysis of cellobiose and cello-oligosaccharides to glucose [1,2]. The interest of studying β -glucosidases in the context of cellulolysis lies in the following factors: (i) the step they catalyse is rate-limiting, as cellobiose is an inhibitor of the other cellulolytic enzymes, i.e. endo- and exo-glucanases [3,4]; (ii) most microbial β -glucosidases are, in turn, sensitive to glucose inhibition; (iii) the amounts of β -glucosidase are generally small in crude cellulase preparations and also when produced intracellularly; (iv) β -glucosidase is usually the least stable of the cellulolytic enzymes and undergoes rapid inactivation during reaction [5]. A way of overcoming some of the above problems is to apply protein engineering to the analysis of the structure/function relationship and to the redesign of biotechnologically interesting properties. Although more than 15 β -glucosidase sequences are available [6,7], there is a lack of structural data; only recently have crystallization details of the enzymes from *Trifolium repens* and *Bacillus polymyxa* been reported in the literature [8,9]. In the absence of X-ray-crystallographic data, secondary-structure prediction used together with IR spectroscopy can generate important structural data, sufficient to suggest residues for site-directed mutagenesis replacement [10–12]. IR spectroscopy is one of the most comprehensive methods of studying protein secondary structure. The application of resolution-enhancement techniques has improved substantially the information obtainable from IR spectra, leading to the separation of otherwise overlapping bands [13–16]. In particular, the bands appearing in the amide-I region (1615–1700 cm^{-1}) are due principally to the absorption of the peptide in plane C=O

stretching vibration, and are indicative of the different classes of secondary structure [13–17].

In the present work we describe the analysis of protein structural traits by means of computer-predictive algorithms and the estimation of secondary-structure content by curve fitting of the deconvoluted IR spectra of a previously cloned and sequenced β -glucosidase from *Streptomyces* sp. ATCC 11238 [18].

EXPERIMENTAL

Computer analysis

Protein secondary structures were predicted as described by Chou and Fasman [19], and the hydropathy profile was plotted by the method of Kyte and Doolittle [20], by means of a previously reported computer program [21]. This program has two main advantages over other similar methods: (a) it minimizes the overprediction of β -turns and (b) uses Boolean decisions in the resolution of secondary-structure overlapping. The latest version of this program [22] can be used to make additional predictions, such as loops [23], Richardson rules for N-cap and C-cap residues [24], structural class of the protein [25], domain boundaries [26] and the reduced/disulphide-bridge state of cysteines [27]. Alignments were performed by the Clustal method [28] from LaserGene package (DNASTAR, Ltd.).

Fourier-transform IR (FTIR) spectroscopy

The recombinant enzyme was obtained and purified as reported elsewhere [18]. For FTIR analyses, the samples are dissolved in 25 mM sodium phosphate buffer, pH 7.0, prepared with either H_2O or $^2\text{H}_2\text{O}$, and kept overnight at room temperature. The solutions were prepared at concentrations of about 20 mg/ml in H_2O or 10 mg/ml in $^2\text{H}_2\text{O}$. An appropriate volume of the solution was applied between two CaF_2 windows using 6 μm spacers (in H_2O) or 25 μm spacers (in $^2\text{H}_2\text{O}$). IR spectra were acquired on a Mattson Polaris FTIR spectrometer equipped with a LgMCT detector, working at an instrument resolution of

Abbreviations used: FTIR spectroscopy, Fourier-transform IR spectroscopy.

‡ To whom correspondence should be addressed.

The sequence data are deposited in the EMBL/GenBank/DBJ Nucleotide Sequence Databases under the accession number Z29625.

2 cm⁻¹. Some 1000 scans were averaged using a sample shuttle, apodized with a triangle function and Fourier-transformed. The spectrometer was continuously purged with dry air (dew point better than -60 °C). The sample temperature (20 °C) was set using a home-made cell jacket of circulating water. The pure spectra of β -glucosidase were obtained by subtracting solvent spectra recorded under identical conditions. The criteria for a good subtraction was to obtain a flat line between 1800 and 2000 cm⁻¹. The same region was used to check the absence of residual water-vapour peaks.

Resolution enhancement of the spectra was performed using programs developed by Moffatt et al. [29]. Fourier self-deconvolution was carried out using a Lorentzian lineshape, a bandwidth of 14 cm⁻¹ and a resolution-enhancement factor of 2.5. The choice of these parameters was dictated by two considerations: (1) the necessity of avoiding the appearance of both negative side lobes and high frequency noise after the deconvolution treatment; (2) we tried to recompose the original spectrum by multiplying the curve-fitted bands by the enhancement factor, and adding all the bands [30]. We found that the above parameters gave the best fit to the original spectrum. Curve fitting was performed on the deconvoluted spectra by the procedures of Byler and Susi [13]. Fourier self-deconvolution and derivation gave the number and position of the component bands in the amide-I region. The bandwidths and intensities of the individual bands were also estimated from the deconvoluted and derivative curves. These were the input parameters for a least-squares iterative curve fitting that was performed to fit the bands of initial Gaussian lineshapes [31] to the deconvoluted spectrum. The peak positions, heights and bandwidths were allowed to vary simultaneously until a good fit was achieved. Peaks above 1700 cm⁻¹ and below 1715 cm⁻¹ are not amide-I bands, but were also introduced in the curve-fitting procedure to avoid the inclusion of a baseline fitting. The proportion of a particular structure was computed to be the area of the corresponding band, divided by the sum of the areas of all the bands having their maximum between 1700 and 1615 cm⁻¹.

RESULTS AND DISCUSSION

According to the algorithm of Chou and Zhang [25] for protein-class prediction, this β -glucosidase belongs to the $\alpha + \beta$ class, which is also corroborated by the distribution of the predicted secondary-structure stretches. Figure 1(a) shows the putative distribution of regular secondary structures along the sequence. Loops, predicted by the rules of Leszczynski and Rose [23], are also indicated. According to our predictions by the algorithm of Muskal et al. [27], the two cysteines (Cys-181 and Cys-230) would be involved in a disulphide bridge. The algorithm of Vonderviszt and Simon [26] indicates the following putative structural domains with boundaries that match with loop or coil stretches: domain I from residue 1 to 57; domain II from residue 58 to 282; domain III from residue 283 to 402; domain IV from residue 403 to the C-terminus (residue 479). It is remarkable that domain boundaries in the Bgl3 β -glucosidase match variable regions of β -glucosidases belonging to the family-1 glycosyl hydrolases [7] on multiple sequence alignment (not shown). Thus such a domain distribution might be conserved among these proteins provided that all of them are similar in size (between 450 and 500 amino acid residues). The protein sequence does not contain any proline/hydroxylamine acid-rich linker or inter-domain region separating functional domains, as do cellulases [32].

Figure 2 shows the deconvoluted spectra of β -glucosidase in the amide-I and amide-II regions, in H₂O and ²H₂O buffers. In

Figure 1 Computer analysis of the *Streptomyces* sp. β -glucosidase

(a) Amino acid sequence and secondary-structure prediction. α , stretches of α -helix; β , β -strand; L, loop. Turns and random coils are not indicated. Vertical arrows indicate the predicted domain (D I–D IV) boundaries. (b) Alignment of the putative functional region, including domain III and the first portion of domain IV from Bgl3. The glutamic acid residue indicated by an asterisk corresponds to the active-site nucleophile of the β -glucosidase from *A. taecalis* as identified by site-direct mutagenesis. The vertical arrow indicates the aspartic acid residue postulated as the acid-base catalyst in *A. taecalis*. Consensus sequences and conserved residues discussed in the text are indicated (*, hydrophobic residue). AGR, *A. taecalis* β -glucosidase; SSP, *Streptomyces* sp. β -glucosidase; TRI, *Trifolium repens* β -glucosidase. See the text for references.

H₂O, the spectrum shows several bands at 1680, 1657, 1651, 1642 and 1632 cm⁻¹, with shoulders at 1667 and 1624 cm⁻¹, and minor peaks at 1688 and 1616 cm⁻¹. This spectrum, together with that obtained using the ²H₂O solvent, is indicative of a complex secondary structure without a clear preponderance of any type of secondary-structural feature. Consideration of the spectrum in ²H₂O buffer in the amide-II region indicates that the majority of the amide protons have been exchanged, although an appreciable proportion still remains unexchanged. In the amide-region, the maximum has shifted from 1657 cm⁻¹ in the H₂O sample to about 1645 cm⁻¹, while similar features to those of the H₂O sample are seen.

Table 1 lists the peak positions, the relative areas obtained by band curve fitting of the deconvoluted spectra, and the assignments of each band. These are based on a number of recent FTIR studies [13,14,17,31,33–36]. In H₂O, the four bands at 1690, 1680, 1673 and 1667 cm⁻¹ are assigned to reverse turns, giving a total amount of about 25%. However, the small band found at 1690 cm⁻¹ may have some contribution from antiparallel β -sheet structures [15]. In general, it is found that this band has an intensity of about 10% of the 1632 cm⁻¹ band (see ref. [37] for a discussion). Therefore the antiparallel β -sheet

Figure 2 Deconvoluted spectra of the *Streptomyces* sp. β -glucosidase in the amide region

(a) In H_2O ; (b) in $^2\text{H}_2\text{O}$; both were in 20 mM sodium phosphate buffer, pH 7.0. Deconvolution was obtained using a bandwidth of 14 cm^{-1} and a resolution-enhancement factor of 2.5.

Table 1 Position, fractional areas and assignments of the amide-I bands of the *Streptomyces* sp. β -glucosidase

Frequency positions are rounded off to the nearest integer.

H_2O buffer			$^2\text{H}_2\text{O}$ buffer		
Frequency (cm^{-1})	Area (%)	Assignment	Frequency (cm^{-1})	Area (%)	Assignment
1690	3	Reverse turns	1683	2	Reverse turns
1680	5	Reverse turns	1676	6	Reverse turns
1673	4	Reverse turns	1667	6	Reverse turns
1667	13	Reverse turns	1657	13	Reverse turns
1657	14	α -Helix	1653	19	α -Helix
1651	22	Non-ordered/ α -helix	1644	24	Non-ordered/ α -helix/ β -sheet
1642	18	β -Sheet/ 3_{10} turns	1633	18	3_{10} Turns/ β -sheet
1632	19	β -Sheet	1625	8	β -Sheet
1624	2	β -Sheet	1617	3	β -Sheet

structures would contribute to the 1690 cm^{-1} band for only about 2% of total amide-I area. The band at 1657 cm^{-1} is assigned to α -helices, whereas the 1651 cm^{-1} band probably arises from both α -helices and non-ordered structures. In the absence of other evidence, we assign half of the total amount to each structure. The 1642 cm^{-1} band is also probably a composite. Although bands at these wavenumbers have been traditionally assigned to β -sheets, isolated 3_{10} turns have also recently been shown to absorb in this region [35]. We thus suggest that both β -sheets and isolated 3_{10} turns contribute to the 1642 cm^{-1} band. The bands at 1632 and 1624 cm^{-1} are attributed to β -sheets, together accounting for about 21% of the total.

As indicated above, more than half of the amide protons have exchanged for deuterons (Figure 2). Thus exposed structures, such as non-ordered regions and reverse turns and the ends of α -helices and β -sheets, would most probably exchange. According to Table 1 we could assume that the 1683 , 1676 , 1667 and 1657 cm^{-1} bands in $^2\text{H}_2\text{O}$ are the result of the shift in the corresponding bands in H_2O (bands at 1690 , 1680 , 1673 and 1667 cm^{-1}). These bands are assigned mainly to reverse turns, and their total area (about 27% in $^2\text{H}_2\text{O}$ compared with about 25% in H_2O) is consistent with this interpretation. As for the sample in H_2O buffer, there may be some small contribution of antiparallel β -structures to the 1676 cm^{-1} band [15]. The band at

1653 cm^{-1} can be assigned totally to α -helices, resulting from the shift of the 1657 cm^{-1} band in H_2O plus some unshifted α -helical band which appears at 1651 cm^{-1} in H_2O . The bands at 1644 and 1633 cm^{-1} in $^2\text{H}_2\text{O}$ are probably mixed bands, resulting from total shifts in the bands corresponding to non-ordered structures and 3_{10} turns respectively, plus partial shifts in bands corresponding to α -helices and β -sheets. Thus Bgl3 glucosidase would consist of about 34% α -helices (including 3_{10} turns), 30% β -sheets, 25% reverse turns and 11% non-ordered structures. In $^2\text{H}_2\text{O}$ these values are explained by the band shifts indicated above. For example, assuming 11% non-ordered structures, the bands between 1653 and 1617 cm^{-1} could be distributed into 34.5% α -helices and 26.5% β -sheets (see Table 1). These results match quite well with the secondary-structure content obtained by computer prediction i.e. 35% α -helix, 23% β -sheet, 31% reverse turns and 11% non-ordered structure.

From analysis of similarity to other glucosidases [6,7,38], we attempted to identify functional domains. A stretch extending from about half of domain III to the first portion of domain IV includes most of the residues that form the putative catalytic region (Figure 1a). A glutamic acid residue fully conserved among members of the family-1 glycosyl hydrolases is the active-site nucleophile (E358) in the β -glucosidase of *A. faecalis*, as recently demonstrated by region-directed mutagenesis [38]. The

region surrounding the active-site nucleophile from *A. faecalis* β -glucosidase and the alignment with the homologous region from the *Streptomyces* sp. enzyme is shown in Figure 1(b). According to the algorithm of Karplus and Schulz [39], the putative catalytic residues are included in a flexible region as would be expected for active sites. On the basis of their acid-base catalytic mechanism [40], most glycohydrolases use Glu/Asp residues as nucleophile/proton donor. On extensive point mutation of several residues surrounding the active-site region of *A. faecalis* β -glucosidase, Trimbur et al. [38] could not identify with certainty the residue (suggested to be D374) acting as the acid-base catalyst. So other residues might be postulated to play important roles in the enzyme function. In this sense, the sequence comparison with the shorter β -glucosidase from *T. repens* [41] could afford an alternative hypothesis (Figure 1b). Such a glucosidase lacks the C-terminal sequence containing the above-mentioned aspartic acid residue but still contains the consensus sequence ITENG which includes the nucleophilic glutamic acid residue. An alternative acid-base catalyst could be *Streptomyces* sp. D302 located in domain III at the end of a predicted α -helix and fully conserved in all of the family-I glycosyl hydrolases. In carbohydrate-protein interaction, van der Waals contacts of aromatic side chains on the saccharide ring have generally been observed [42,43]. In this sense, the motif GXNYY and a tryptophan residue (W356 in *Streptomyces* sp.) are strongly conserved, suggesting that putative aromatic residues are involved in substrate binding. These motifs are located in the predicted domain III; a GVNYY motif lies on the adjacent C-terminal side of the potential acid-base catalyst D302. It is remarkable that both D302 and the sequence GVNYY are found in a structural block (DFLGVNYYSP) related to glycosyl hydrolases as discovered by means of the algorithms Prosite [44] and Building blocks [45] implemented in a program reported elsewhere [46]. Furthermore, multiple alignment of the sequence forming domain III reveals that this domain is comprised of two mostly conserved sequence blocks separated by a variable region corresponding to predicted loops; D302 and the GVNYY motif are included in the N-terminal block of the domain, and MGW356 lies upstream of the nucleophilic active site. All these features can be observed in Figure 1(b). Also conserved in all family-I members is *Agrobacterium* R377, which, according to Trimbur et al. [38], is probably involved in binding the second sugar moiety of cellobiose.

The above results should provide useful information about the structure of this class of enzyme, taking into account the fact that, although we have obtained Bgl3 crystals suitable for X-ray-crystallographic studies (A. Guasch, J. A. Perez-Pons and E. Querol, unpublished work), their resolution is time-consuming work. In such cases, the combination of predictive methods and site-directed mutagenesis provides a method of analysing structure/function relationships further [10–12] and even facilitating crystallographic analysis.

This research was supported by grants BIO91-0477, BIO94-0912 to E.Q. and PB92-0622 to E.P. from the CICYT (Ministerio de Educación y Ciencia, Spain). The skilful technical assistance of Ms. E. Serrano is gratefully acknowledged.

REFERENCES

- Berghem, L. E. R. and Pettersson, L. G. (1974) *Eur. J. Biochem.* **46**, 295–305
- Béguin, P. and Aubert, J. P. (1994) *FEMS Microbiol. Rev.* **13**, 25–58
- Wood, T. M. and McCrae, S. I. (1982) *J. Gen. Microbiol.* **128**, 2973–2982
- Klyosov, A. A. (1990) *Biochemistry* **29**, 10577–10585
- Woodward, J. and Wiseman, J. (1982) *Enzyme Microb. Technol.* **4**, 52–53
- Henrissat, B. (1991) *Biochem. J.* **280**, 309–316
- Henrissat, H. and Bairoch, A. (1993) *Biochem. J.* **293**, 781–788
- Tolley, S. P., Barrett, T. E., Suresh, C. G. and Hughes, M. A. (1993) *J. Mol. Biol.* **229**, 791–793
- Sanz-Aparicio, J., Romero, A., Martínez-Ripoll, M., Madarro, A., Flors, A. and Polaina, J. (1994) *J. Mol. Biol.* **240**, 267–270
- Querol, E., Padrós, E., Juncosa, M., Planas, A. and Lloberas, J. (1992) *Biochem. Biophys. Res. Commun.* **184**, 612–617
- Planas, A., Juncosa, M., Lloberas, J. and Querol, E. (1992) *FEBS Lett.* **308**, 141–145
- Juncosa, M., Pons, J., Dot, T., Querol, E. and Planas, A. (1994) *J. Biol. Chem.* **269**, 14530–14535
- Byler, D. M. and Susi, H. (1986) *Biopolymers* **25**, 469–487
- Surewicz, W. J. and Mantsch, H. H. (1988) *Biochim. Biophys. Acta* **952**, 115–130
- Arrondo, J. L. R., Muga, A., Castresana, J. and Gofí, F. (1993) *Prog. Biophys. Mol. Biol.* **59**, 23–56
- Surewicz, W. K., Mantsch, H. H. and Chapman, D. (1993) *Biochemistry* **32**, 389–394
- Harris, P. I. and Chapman, D. (1992) *Trends Biochem. Sci.* **17**, 328–333
- Pérez-Pons, J. A., Cayetano, A., Rebordosa, X., Lloberas, J., Guasch, A. and Querol, E. (1994) *Eur. J. Biochem.* **223**, 557–565
- Chou, P. Y. and Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **47**, 45–148
- Kyte, J. and Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Parrilla, A., Domenech, A. and Querol, E. (1986) *Comput. Appl. Biosci.* **2**, 211–215
- Clotet, J., Cedano, J. and Querol, E. (1994) *Comput. Appl. Biosci.* **10**, 495–500
- Leszczynski, F. J. and Rose, G. D. (1986) *Science* **234**, 849–855
- Richardson, J. S. and Richardson, D. (1988) *Science* **240**, 1648–1652
- Chou, K. and Zhang, C. (1992) *Eur. J. Biochem.* **207**, 429–433
- Vonderviszt, F. and Simon, I. (1986) *Biochem. Biophys. Res. Commun.* **139**, 11–17
- Muskal, S. M., Holbrook, S. R. and Kim, S. H. (1990) *Protein Eng.* **3**, 667–672
- Higgins, D. G. and Sharp, P. M. (1989) *Comput. Appl. Biosci.* **5**, 151–153
- Moffatt, D. J., Kauppinen, J. K., Cameron, D. G., Mantsch, H. H. and Jones, R. N. (1986) *Computer Programs for Infrared Spectrophotometry*, NRCC Bulletin 18, National Research Council of Canada, Ottawa
- Cladera, J., Sabés, M. and Padrós, E. (1992) *Biochemistry* **31**, 12363–12368
- Prestrelski, S. J., Byler, D. M. and Liebman, M. N. (1991) *Biochemistry* **30**, 133–143
- Ong, E., Greenwood, J. M., Gilkes, N. R., Kilburn, D. G., Miller, R. C. and Warren, R. A. J. (1986) *Trends Biotechnol.* **7**, 239–243
- Fabian, H., Naumann, D., Misselwitz, R., Ristau, O., Gerlach, D. and Welfle, H. (1992) *Biochemistry* **31**, 6532–6538
- Holloway, P. W. and Mantsch, H. H. (1989) *Biochemistry* **28**, 931–935
- Kennedy, D. F., Crisma, M., Toniolo, C. and Chapman, D. (1991) *Biochemistry* **30**, 6541–6548
- Goormaghtigh, E., Cabiaux, V. and Ruyschaert, J. M. (1990) *Eur. J. Biochem.* **193**, 409–420
- Caughey, B. W., Dong, A., Bhat, K. S., Ernst, D., Hayes, S. F. and Caughey, W. S. (1991) *Biochemistry* **30**, 7672–7680
- Trimbur, D. E., Warren, R. A. J. and Withers, S. G. (1992) *J. Biol. Chem.* **267**, 10248–10251
- Karplus, P. A. and Schulz, G. E. (1985) *Naturwissenschaften* **72**, 212–213
- Sinnott, M. L. (1990) *Chem. Rev.* **90**, 1171–1202
- Oxtoby, E., Dunn, M. A., Pancoro, A. and Hughes, M. A. (1991) *Plant Mol. Biol.* **17**, 209–219
- Vyas, N. K. (1991) *Curr. Opin. Struct. Biol.* **1**, 732–740
- Maenaka, K., Kawai, G., Watanabe, K., Sunada, F. and Kumagai, I. (1994) *J. Biol. Chem.* **269**, 7070–7075
- Bairoch, A. (1992) *Nucleic Acids Res.* **20**, 2013–2018
- Henikoff, S. and Henikoff, J. G. (1991) *Nucleic Acids Res.* **19**, 6565–6572
- Fuchs, R. (1994) *Comput. Appl. Biosci.* **10**, 171–178