

## Refined Structure of the Chitinase from Barley Seeds at 2.0 Å Resolution

HYUN KYU SONG AND SE WON SUH\*

Department of Chemistry and Center for Molecular Catalysis, Seoul National University, Seoul 151-742, Korea

(Received 18 April 1995; accepted 3 July 1995)

### Abstract

Chitinase from barley seeds is a monomeric enzyme with 243 amino-acid residues and it plays a role as a defense protein. Its structure, previously determined at 2.8 Å resolution by multiple isomorphous replacement method, is mainly  $\alpha$ -helical [Hart, Monzingo, Ready, Ernst & Robertus, (1993). *J. Mol. Biol.* **229**, 189–193]. The crystallization and preliminary X-ray data of the same enzyme in a different crystal form has been reported independently [Song, Hwang, Kim & Suh, (1993). *Proteins*, **17**, 107–109], the asymmetric unit of which contains two chitinase molecules. As a step toward understanding the general principles of catalysis, reported here is the structure of chitinase from barley seeds in this crystal form, as determined by molecular replacement and subsequently refined at 2.0 Å resolution, with incorporation of partial data to 1.9 Å (*R* factor of 18.9% for 31 038 unique reflections with  $F_o > 2\sigma_F$  in the range 8.0–1.9 Å). The r.m.s. deviations from ideal stereochemistry are 0.013 Å for bond lengths and 1.32° for bond angles. A superposition of the two independent molecules in the asymmetric unit gives an r.m.s. difference of 0.55 Å for all protein atoms (0.43 and 0.74 Å for main-chain and side-chain atoms, respectively). When the refined model of each chitinase molecule in the asymmetric unit is superposed with the starting model, the r.m.s. difference for all shared protein atoms is 0.99 Å for molecule 1 and 0.85 Å for molecule 2, respectively. Through a sequence comparison with homologous plant chitinases as well as a structural comparison with the active sites of other glycosidases, key catalytic residues have been identified and the active site has been located in the three-dimensional structure of the barley chitinase. The present structure, refined at an effective resolution of 2.0 Å with incorporation of partial data to 1.9 Å, represents a significant improvement in resolution compared to the previously reported model. The improved resolution has enabled the location of solvent atoms, including water molecules near the catalytic residues, in addition to the positioning of protein atoms with greater accuracy.

### 1. Introduction

Chitinases {poly-[1,4- $\beta$ -(2-acetamido-2-deoxy-D-glucoside)] glycanohydrolase, E.C. 3.2.1.14} are distributed

in diverse organisms including bacteria, fungi, insects, plants and animals (Jeuniaux, 1966). They catalyze the hydrolysis of chitin, a biopolymer composed of  $\beta$ -1,4-linked *N*-acetylglucosamine (GlcNAc) units, which is the major component of the cell wall of many fungi and of the exoskeleton of insects. Chitinases play diverse roles in different organisms. For example, chitinases present in the digestive track of fish and those secreted by soil bacteria degrade the chitin for food and energy (Roberts & Selitrennikoff, 1988). Insect chitinases play an essential role during ecdysis by catalyzing the hydrolysis of integumental chitin (Dickinson, Keer, Hitchcock & Adams, 1989). Plant chitinases, belonging to the pathogenesis-related proteins (Eyal & Fluhr, 1991), retard the fungal growth by interfering with the biosynthesis of the cell wall (Schlumbaum, Mauch, Vögeli & Boller, 1986).

A new classification of chitinases has been recently proposed by Beintema (Beintema, 1994). It designates two groups as either b- or h-type, depending on the amino-acid sequence similarity to either barley endochitinase or hevamine, a chitinase/lysozyme from the rubber tree *Hevea brasiliensis*. The b-type includes classes I, II, and IV, while the h-type includes classes III and V. Many plant chitinases, including those isolated from the leaves and seeds of barley, maize, tobacco, potato, bean and rice, show a significant sequence similarity (Verburg, Smith, Lisek & Huynh, 1992). They are in general endochitinases and produce chito-oligosaccharides as principal products (Molano, Polacheck, Duran & Cabib, 1979). One of the four basic proteins from barley seeds has been identified as a chitinase with *pI* of 9.7 (Hejgaard & Bjørn, 1985). In barley, chitinase accumulates to high levels late during seed development and it synergistically inhibits the growth of fungi with two other antifungal proteins, *i.e.*, ribosome-inactivating protein and (1–3)- $\beta$ -glucanase (Leah, Tommerup, Svendsen & Mundy, 1991). The amino-acid sequence of the barley chitinase (BCH) deduced from its cDNA sequence (Leah *et al.*, 1991) indicated that it belongs to the class II of b-type chitinases (Beintema, 1994). The mature polypeptide chain, consisting of 243 amino-acid residues, has a molecular mass of 25 933 Da (Leah *et al.*, 1991). Class II chitinases are distinguished from class I enzymes by the absence of an N-terminal cysteine-rich domain similar to the rubber protein called hevein (Shinshi,

Neuhaus, Ryals & Meins, 1990). Class III chitinases (e.g., hevamine) show no sequence homology with either class I or II chitinases (Shinshi, Neuhaus, Ryals & Meins, 1990).

So far only three crystal structures of chitinases have been determined by the multiple isomorphous replacement method. They are (i) a b-type chitinase, endo-chitinase from barley (*Hordeum vulgare*) seeds at 2.8 Å resolution (Hart *et al.*, 1993); and (ii) two h-type chitinases, hevamine, an enzyme with chitinase/lysozyme activity from rubber tree (*Hevea brasiliensis* L.) latex at 2.2 Å resolution (Van Scheltinga, Kalk, Beintema & Dijkstra, 1994) and a bacterial exochitinase from *Serratia marcescens* at 2.3 Å resolution (Perrakis *et al.*, 1994). The structures of b- and h-type chitinases are completely different from each other, which may be expected from dissimilar amino-acid sequences. Previously, we reported the crystallization and preliminary X-ray crystallographic data of BCH (Song, Hwang, Kim & Suh, 1993). During the course of an independent structure analysis on this enzyme, its structure became known (Hart *et al.*, 1993) and thus we have employed the method of molecular replacement to solve the structure of BCH in the new crystal form, the asymmetric unit of which contains two independent molecules. In this paper we present the structure refined at 2.0 Å resolution, with incorporation of partial data to 1.9 Å

## 2. Experimental

### 2.1. Purification, crystallization, and X-ray data collection

Purification, crystallization, and preliminary X-ray data, including a description of the collection of native X-ray data to approximately 2.2 Å, have been reported previously (Song *et al.*, 1993). For this study a new set of native X-ray data has been collected to higher resolution from a better-diffracting crystal on a FAST TV area detector (Enraf-Nonius, Delft), using the MADNES software (Messerschmidt & Pflugrath, 1987). The native data set consists of 81 204 measurements of 31 962 unique reflections with an  $R_{\text{merge}}$  (on intensity) of 4.9% (rejecting 3.6% outliers). The completeness of the data is 95.4% between 28.0 and 1.99 Å, with the completeness of the shell between 2.09 and 1.99 Å being 79.3% and that of the shell between 1.99 and 1.90 Å being 26.3%. The crystal belongs to the monoclinic space group  $P2_1$  with unit-cell dimensions of  $a = 69.38$ ,  $b = 44.47$ ,  $c = 81.49$  Å and  $\beta = 111.99^\circ$ . There are two chitinase molecules in the asymmetric unit ( $V_M = 2.25$  Å<sup>3</sup> Da<sup>-1</sup>; solvent content = 45%).

### 2.2. Molecular replacement

The structure was determined by the molecular-replacement method (Crowther, 1972; Crowther & Blow, 1967) using the program *X-PLOR* Version

3.1 (Brünger, 1992a). The starting model was the previously reported structure (Hart *et al.*, 1993) of BCH (Protein Data Bank ID code 1BAA;  $R$  factor 0.20 for 5.0–2.8 Å data; r.m.s. deviations from ideal geometry are 0.020 Å for bond distances and 4.1° for bond angles). For the calculations of rotation function, only non-H protein atoms were included in the starting model. The probe structure factors were calculated by placing the starting model in an orthogonal unit cell with dimensions of  $100 \times 100 \times 100$  Å. The orientations of the search model were sampled by using the pseudo-orthogonal Eulerian angles. The rotation function search was carried out using a maximum Patterson radius of 29 Å and a step size of 2.5°. The best solution was obtained when data between 15.0 and 4.0 Å were used. Reflections with  $F_o > 2\sigma_F$  were used throughout the molecular replacement and subsequent refinement calculations. A consistent solution was obtained with other resolution ranges. Rigid-body Patterson correlation refinement (PC refinement) was then performed on the orientations produced by the above cross-rotation search for the same resolution range. After the PC refinement, 12 peaks with the highest correlation coefficients among the 277 cross-rotation solutions converged to two orientations corresponding to the two independent molecules in the asymmetric unit ( $\theta_1 = 303.7$ ,  $\theta_2 = 63.1$ ,  $\theta_3 = 189.4^\circ$  and  $\theta_1 = 149.8$ ,  $\theta_2 = 76.0$ ,  $\theta_3 = 119.3^\circ$ ). These two orientations are related by the non-crystallographic symmetry observed in the self-rotation function ( $\psi = 35.0$ ,  $\varphi = 135.0$ ,  $\kappa = 180.0^\circ$ ). Next the translation function was calculated using the X-ray data between 15.0 and 4.0 Å. The highest peak found in the translation search for each orientation yielded the translation vectors  $x_1 = 0.067$ ,  $z_1 = 0.311$  and  $x_2 = 0.222$ ,  $z_2 = 0.078$  in fractional coordinates for molecules 1 and 2, respectively. To determine the relative  $y$  position of molecule 2 with respect to molecule 1 in the asymmetric unit, a combined translation function was calculated. A significant peak with a correlation coefficient of 0.47 was found at a relative translation vector  $y = 0.289$ . The two independent molecules in the asymmetric unit, after being oriented and positioned according to the molecular replacement solutions, gave an  $R$  factor of 0.430 for data between 8.0 and 3.5 Å.

### 2.3. Refinement

The model of the two independent chitinase molecules in the asymmetric unit was then subject to refinement by the program *X-PLOR* Version 3.1 (Brünger, 1992a). Initially a rigid-body refinement was carried out with 8.0–6.0 Å data to further improve the positional and orientational parameters. The high-resolution limit of the diffraction data was increased step-wise from 6.0 to 3.0 Å, as the number of rigid bodies was increased from two to 20. The  $R$  factor at this stage was 0.438 for 8.0–3.0 Å data. Next, atomic positions were refined by

Table 1. Refinement statistics for barley chitinase

<i>R</i> factor	0.189
Free <i>R</i> factor	0.244
Resolution range (Å)	8.0–1.9
No. of unique reflections ( $F_o > 2\sigma_F$ )	31038
No. of non-H atoms	2 × 1831
No. of residues	2 × 243
No. of solvent atoms	279
Overall <i>G</i> factor	0.44

the conventional conjugate-gradient minimization, with higher resolution data being added in steps. The starting model had one discrepancy in the amino-acid sequence compared to the SWISS-PROT entry (chi2\_horvu): an arginine was modeled at position 17 instead of histidine. During the positional refinement, the electron density indicated that this residue is consistent with histidine and thus Arg17 in the starting model was replaced by histidine. The two independent molecules in the asymmetric unit were subject to the exact non-crystallographic symmetry until the positional refinement at 2.3 Å converged. At this stage the *R* factor reached 0.311 for 8.0–2.3 Å data.

In the next round of the positional refinement for the same resolution range, the atoms of the two non-crystallographically related molecules in the asymmetric unit were allowed to move independently. The *R* factor further dropped to 0.293. This model was then subject to the simulated-annealing refinement, employing the standard slow-cooling protocol (Brünger & Krukowski, 1990): from 4000 to 300 K (time step 0.5 fs; decrement of temperature 25 K; number of steps at each temperature 50). Solvent molecules, modeled as water, were located as high peaks in the  $F_o - F_c$  maps using *PEAKMAX/WATPEAK* in the *CCP4* program package (Collaborative Computational Project, Number 4, 1994) and were included in the subsequent rounds of refinement. When the *B* factor of a solvent molecule exceeded 60 Å<sup>2</sup>, it was removed. Higher resolution reflections up to 1.9 Å were added in steps. Further simulated-annealing refinement and the refinement of isotropic temperature factors (*B* factors) for individual atoms gave the final *R* factor of 0.189 for 31 038 unique reflections with  $F_o > 2\sigma_F$  in the range 8.0–1.9 Å. The final free *R* factor (Brünger, 1992*b*) is 0.244. A summary of the refinement statistics is given in Table 1.

#### 2.4. Model building and structure analysis

The  $2F_o - F_c$  and  $F_o - F_c$  electron-density maps were used as guides for manual rebuilding of the model, for which the interactive computer programs *FRODO* (Jones, 1985) running on an Evans & Sutherland PS390 graphics system and *CHAIN* Version 5.4 (Sack, 1988) running on a Silicon Graphics Indigo2 XZ workstation were used. At each step of the model rebuilding and refinement, the stereochemistry of the model

was assessed by the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). The presence of non-crystallographic symmetry was utilized wherever possible during manual rebuilding. Structural comparisons were made using the program *LSQKAB* in *CCP4* program package (Collaborative Computational Project, Number 4, 1994). The programs *CHAIN* Version 5.4 (Sack, 1988) and *SETOR* Version 4.13.0 (Evans, 1993) were used on a Silicon Graphics Indigo2 XZ workstation for model display. The comparison of amino-acid sequences was carried out by an electronic mail server (@ncbi.nlm.nih.gov for retrieval and blitz@embl-heidelberg.de for alignment).

### 3. Results and discussion

#### 3.1. Quality of the refined model

The refined model contains 3662 non-H protein atoms and 279 ordered water molecules, accounting for all 486 residues in the two independent molecules of chitinase in the asymmetric unit (Table 1). It gives an *R* factor of 18.9% for 31 038 unique reflections with  $F_o > 2\sigma_F$  in the range 8.0–1.9 Å. The r.m.s. deviations from ideal stereochemistry are 0.013 Å for bond lengths, 1.32° for bond angles, 24.4° for dihedral angles, and 2.3° for improper angles. The average positional error is estimated to be 0.20 Å from the Luzzati plot (Fig. 1) (Luzzati, 1952). Among 394 non-glycine and non-proline residues in the two molecules in the asymmetric unit, the numbers of residues lying in the most favoured, additionally allowed, generously allowed, and disallowed regions in the Ramachandran plot are 340 (86.3%), 52 (13.2%), 0 (0%), and two (0.5%), respectively (Fig. 2). Only a single residue in each molecule is in the disallowed region: His121 and His621 (In this paper, the residue numbering of molecule 1 starts from 1 and that of molecule 2 from 501.). The  $\psi$  angles of these residues

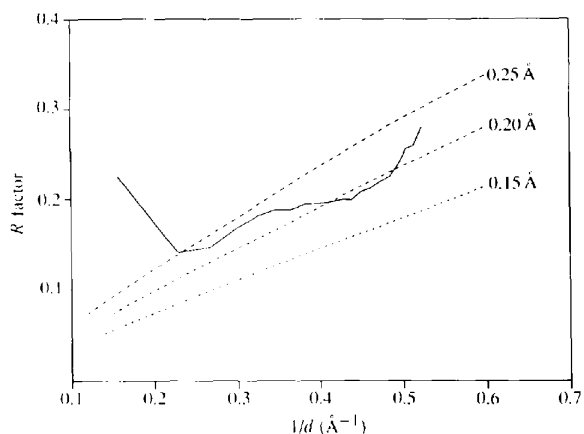


Fig. 1. Luzzati plot for barley chitinase. A total of 31 038 unique reflections with  $F_o > 2\sigma_F$  in the range 8.0–1.9 Å were used. The estimated mean coordinate error is approximately 0.20 Å.

lie slightly outside the allowed range but electron density around these residues is well defined. For molecule 1, the average  $B$  factors are  $14.1 \text{ \AA}^2$  for 972 main-chain atoms and  $17.4 \text{ \AA}^2$  for 859 side-chain atoms. For molecule 2, the values are slightly higher:  $15.4 \text{ \AA}^2$  for main-chain and  $18.8 \text{ \AA}^2$  for side-chain atoms. The  $B$  factors for 279 solvent molecules range from  $4.5$  to  $53.5 \text{ \AA}^2$ , with the average being  $32.9 \text{ \AA}^2$ . The average  $B$  factors for the main-chain and side-chain atoms of each residue in the two independent chitinase molecules are plotted as a function of residue number in Fig. 3. Besides the N- and C-termini of the chain, regions with  $B$  factors exceeding  $30 \text{ \AA}^2$  are residues 90–91, 100–101, and 162–163 in molecule 1 and residues 574 and 663–664 in molecule 2. The electron density for these residues is weak but the tracing of the peptide chain is not ambiguous. Except for the three regions (residues 73–74, 90–91, and 100–101), the  $B$ -factor plots for the two molecules show an excellent correlation to each other. Therefore, these plots can be useful as an indicator of the inherent flexibility of the segments. The overall  $G$  factor, which indicates the stereochemical quality of the main-chain, as defined in *PROCHECK* (Laskowski *et al.*, 1993), is better than the expected value at this resolution (Table 1).

### 3.2. Comparison of the two independent molecules in the asymmetric unit

The two independent molecules are related by the non-crystallographic symmetry: a rotation by an angle of  $179.1^\circ$  followed by a translation of  $27.8 \text{ \AA}$  along the

rotation axis. Molecule 2 is generated from molecule 1 by the following transformation,

$$\begin{aligned} X_2 &= -0.6431X_1 - 0.6782Y_1 + 0.3556Z_1 - 46.323 \\ Y_2 &= -0.6915X_1 + 0.3150Y_1 - 0.6500Z_1 + 51.836 \\ Z_2 &= +0.3289X_1 - 0.6639Y_1 - 0.6716Z_1 + 83.271. \end{aligned}$$

A superposition of the two independent molecules in the asymmetric unit gives an r.m.s. difference of  $0.55 \text{ \AA}$  for all 1831 non-H atoms in 243 residues ( $0.43 \text{ \AA}$  for 972 main-chain atoms and  $0.74 \text{ \AA}$  for 859 side-chain atoms). A superposition of the 243  $C\alpha$  pairs in the two molecules is shown in Fig. 4. The r.m.s. differences between the two molecules for the main-chain and side-chain atoms of each residue are plotted as a function of residue number in Fig. 5. The largest difference for main-chain atoms ( $> 1.3 \text{ \AA}$ ) is confined to the region between residues 90 and 98, which has one of the least conserved sequences. This large discrepancy can be a consequence of either the different crystalline environment or large coordinate errors in this region, or a combination of both. Among residues in this region, only two residues 90 and 91 in molecule 1 show high main-chain  $B$  factors ( $> 40 \text{ \AA}^2$ ), indicating the coordinate errors larger than average. Thus a large deviation in the region between residues 90 and 98 is partly due to the large coordinate errors of these two residues but it also reflects the inherent conformational variability of this region. Besides this region between residues 90 and 98, two residues 107 and

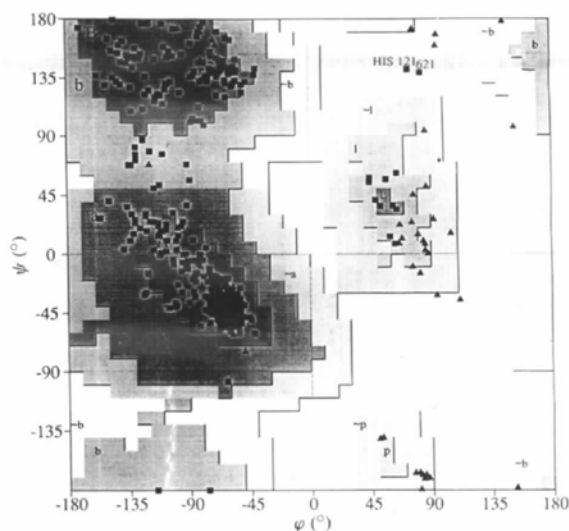


Fig. 2. Ramachandran plot for barley chitinase, created by *PROCHECK* (Laskowski *et al.*, 1993). Glycine and proline residues are denoted by triangles and all other residues by squares. The two labeled residues in the disallowed region are His121 in molecule 1 and His621 in molecule 2. The different regions defined by borderlines are labeled as: A, B, L, most favored; a, b, l, p, allowed; and ~a, ~b, ~l, ~p, generously allowed.

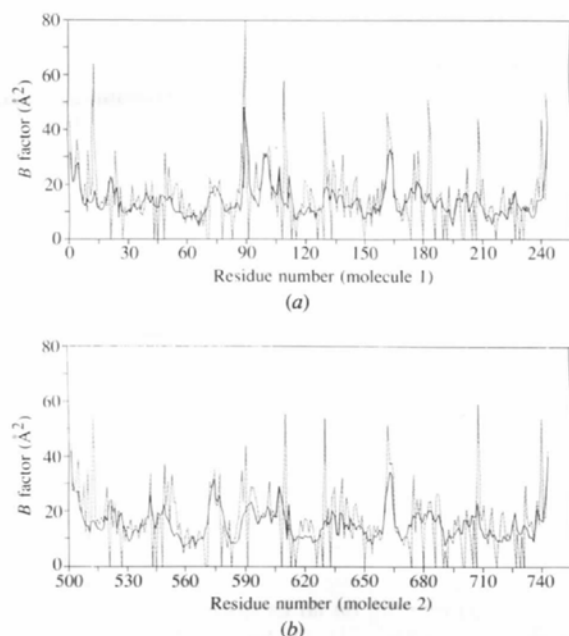


Fig. 3.  $B$ -factor plot. The average  $B$  factors for the main-chain (dashed lines) and side-chain atoms (dotted lines) of each residue are plotted as a function of residue number (a) for molecule 1 and (b) for molecule 2 in the asymmetric unit.

108 and the C-terminus show large deviations ( $> 1.1 \text{ \AA}$ ). In other parts of the chain, the magnitude of deviations is roughly what is expected from the estimated coordinate errors and the structural difference between the two independent molecules in the asymmetric unit should be regarded as only minor.

### 3.3. Comparison of refined and starting models

The previously reported model of BCH (Hart *et al.*, 1993), which was used as the starting model in our study, contained an amino-acid residue (Arg17 instead of His) different from the reported sequence deduced from the cDNA sequence (Leah *et al.*, 1991). However, our refinement indicated a clear electron density for histidine at this position. The starting model contained four *cis*-peptide bonds before proline residues at 140, 164, 178, and 241. In contrast, our refined model has only one *cis*-peptide before Pro163 in molecule 1 (and equivalently before Pro663 in molecule 2). When the refined model of molecule 1 is compared with the starting model, the r.m.s. deviations are 0.86 and 1.3  $\text{\AA}$  for main-chain and side-chain atoms, respectively, and 0.99  $\text{\AA}$  for all shared atoms, omitting the four non-identical atoms at sequence position 17 (Fig. 6a). For molecule 2, the values are 0.72 and 1.1  $\text{\AA}$  for main-chain and side-chain atoms, respectively, and 0.85  $\text{\AA}$  for all shared atoms (Fig. 6b). Large deviations are restricted mostly to the protruding

loop regions of the structure. This may be related to the inherent flexibility of these loop regions and the different crystalline environments. However, one has to be cautious in interpreting the results of this comparison, since the present structure and the starting model have been refined at different resolutions to different extents and they have been derived from different crystal forms.

### 3.4. Overall protein structure

Barley chitinase is an all  $\alpha$ -type structure, since nearly half (47%) of the 243 residues in the polypeptide chain are in  $\alpha$ - and  $3_{10}$ -helices. The overall folding of BCH in our refined model is essentially identical to the previously reported one (Hart *et al.*, 1993). BCH is composed of a single compact domain with approximate dimensions of  $45 \times 42 \times 40 \text{ \AA}$ . The structure can be divided into ten  $\alpha$ -helices (H1 through H10), two  $3_{10}$ -helices (G1 and G2), and loops connecting these secondary-structure elements (Fig. 7). Our assignment of the secondary-structure elements indicated in Fig. 7 is largely in agreement with the previous one by Hart *et al.* (Hart *et al.*, 1993). The discrepancies worth mentioning are: first, the helix H5 is only four residues long (residues 142–145) in our refined model instead of seven residues long (residues 139–145) in the previous model and, second, two  $3_{10}$ -helices has been assigned in our model instead of none in the previous model.

The polypeptide chain of BCH has three disulfide bridges (Cys23–Cys85, Cys97–Cys105, and Cys204–Cys236) and a free cysteine (Cys223). The three disulfide bonds are well separated from each other in the structure and are positioned on the long loop

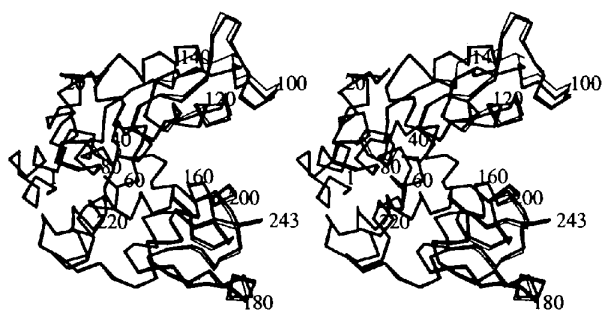


Fig. 4. Stereo diagram showing the superposition of  $C_{\alpha}$  atoms in the two independent molecules in the asymmetric unit. Thick and thin lines represent molecules 1 and 2 in the asymmetric unit, respectively. Every 20th residue is labeled.

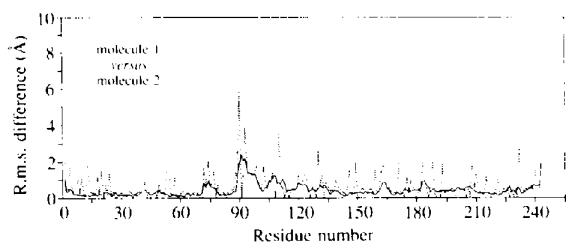


Fig. 5. Plot of the difference between the two independent molecules in the asymmetric unit. The r.m.s. differences for the main-chain (dashed lines) and side-chain atoms (dotted lines) of each residue are plotted as a function of residue number.

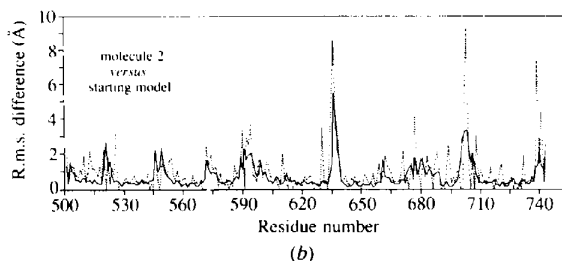
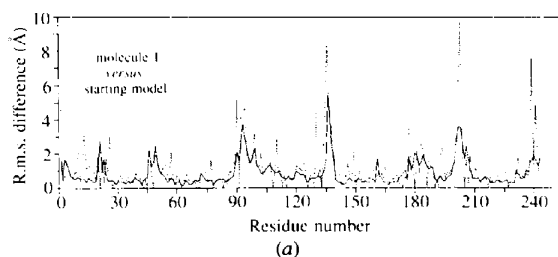


Fig. 6. Plot of the difference between the refined and starting models. (a) Molecule 1 versus the starting model and (b) molecule 2 versus the starting model. The r.m.s. differences for the main-chain (dashed lines) and side-chain atoms (dotted lines) of each residue are plotted as a function of residue number.

regions. The residues adjacent to the disulfide bonds have relatively low *B* factors in both molecules 1 and 2 in the asymmetric unit (Fig. 3). It appears that the disulfide bonds reduce the flexibility of the loop regions by cross-linking them.

The structure of BCH, a b-type enzyme, is completely different from the recently reported structures of h-type chitinases: hevamine, a plant enzyme with chitinase/lysozyme activity from the rubber tree (*Hevea brasiliensis*) latex (Van Scheltinga *et al.*, 1994) and a bacterial chitinase from *Serratia marcescens* (Perrakis *et al.*, 1994). The h-type chitinases have typical  $(\beta/\alpha)_8$ -barrel folds, which are present in many polysaccharide-hydrolyzing enzymes such as  $\alpha$ - and  $\beta$ -amylases (Larson, Greenwood, Casio, Day & McPherson, 1994; Mikami, Degano, Hehre & Sacchettini, 1994). In addition to the  $(\beta/\alpha)_8$ -barrel domain, the bacterial chitinase has two extra domains (Perrakis *et al.*, 1994).

### 3.5. Identification of the active-site and catalytic residues

Fig. 8 shows the alignment of amino-acid sequences of 11 plant chitinases: nine basic chitinases [from barley (Leah *et al.*, 1991), rice (Zhu & Lamb, 1991), mouse-ear cress (Samac, Hironaka, Yallaly & Shah, 1990), rape (Hamel & Bellemare, 1993), tomato (Danhash, Wage-makers, Van Kan & De Wit, 1993), tobacco (Shinshi, Mohnen & Meins, 1987), kidney bean (Broglie, Gaynor & Broglie, 1986), potato (Gaynor, 1988), garden bean (Vad, De Neergaard, Madriz-ordenana, Mikkelsen & Collinge, 1993)] and two acidic chitinases [from petunia (Linthorst, Vanloon, Memelink & Bol, 1991), western balsam poplar (Parsons, Bradshaw & Gordon, 1989)]. These chitinases are expected to have very similar three-dimensional structures, because their sequence identity is above 50%. From this alignment one can recognize three stretches of the most highly conserved residues at 63–72, 113–118 and 192–205. In addition to these regions, two shorter stretches of four strictly conserved residues are at 156–159 and 233–236. Fig. 9 shows the distribution of conserved residues in the three-dimensional structure. Since functionally and/or structurally important residues, particularly those residues involved in catalysis and in substrate binding, are expected to be highly conserved, the active site of the enzyme can be recognized as the elongated cleft in the structure (Figs. 7 and 9). The same active site has been suggested by Hart *et al.* (1993) and it is further supported by the identification of Glu67 as the key catalytic residue by Holm & Sander (Holm & Sander, 1994). This residue, located near the C-terminus of ( $\alpha$ -helix H3, lies right at the center of the presumed active site (Figs. 7 and 9).

Holm & Sander proposed the invariant Glu67 of BCH to be the equivalent of Glu35 in hen egg-white lysozyme (HEWL) (Holm & Sander, 1994). This is consistent with its location in the presumed active-site cleft (Figs. 7 and 9) and with the fact that the amino-acid sequence

around this residue is highly conserved (Fig. 8). It is worth mentioning that both Glu67 of BCH and Glu35 of HEWL are positioned near the C-terminus of  $\alpha$ -helices. The  $pK_a$  values of carboxylate groups in these residues can be raised by the helix dipole moment (Hol, 1985). In pig pancreatic  $\alpha$ -amylase, a chloride ion close to the general acid Glu233 raises the  $pK_a$  of the carboxylate group (Qian, Haser, Buisson, Duee & Payan, 1994).

Holm & Sander further proposed a structural equivalence for 11 pairs of segments between BCH and human lysozyme (Holm & Sander, 1994). They suggest that the invariant Gly115 in BCH is structurally equivalent to the invariant Gly55 in human lysozyme, which is part of a tight  $\beta$ -turn. If the proposed structural equivalence is accepted, the equivalent of Asp53 in human lysozyme (or HEWL Asp52) should be a residue near position 113 in BCH. However, neither Asp nor any other suitable residue can be found in the sequence, -Tyr-Tyr-Gly(113)-Arg-Gly-Pro-Ile-Gln-, since the side chains of Tyr111 and Tyr112 point away from the carboxylate group of Glu67 and the side-chain atoms of Gln118 make hydrogen bonds with the protein backbone. In view of this difficulty, one has to exercise caution in making predictions on the basis of the structural resemblance between BCH and lysozymes, as proposed by Holm & Sander (Holm & Sander, 1994).

Carboxylate C atoms of Glu35 and Asp52 in HEWL are separated by 7.1 Å, while, the carboxylate C atoms of Glu233 and Asp300 in pig pancreatic  $\alpha$ -amylase, two of three catalytically important residues, are separated by 7.0 Å (Larson *et al.*, 1994). A superposition of these residues with Glu35 and Asp300 of HEWL demonstrated a remarkable structural resemblance of the active sites of the two polysaccharide hydrolyzing enzymes (Larson *et al.*, 1994). In the case of a bacterial chitinase from *S. marcescens*, the closest carboxylate O atoms of Asp391 and Glu315 are 6.9 Å apart [the corresponding distance in HEWL is 6.2 Å (Perrakis *et al.*, 1994)]. To identify a residue in BCH, which is equivalent to HEWL Asp52, the negatively charged or polar residues around Glu67, except histidine, are illustrated in Fig. 10. They are Gln63 (7.2 Å), Thr64 (7.0 Å), Ser65 (8.9 Å), Thr68 (7.6 Å), Thr69 (8.7 Å), Gln118 (7.1 Å), Thr195 (9.4 Å), Asn199 (6.5 Å), Glu203 (7.3 Å) and Asp214 (9.5 Å) (the distance given in parentheses is between the carboxylate C atom of Glu67 and the structurally equivalent atom in the nearby residue, averaged over the two molecules in the asymmetric unit). Asp214 can be ruled out to be the second catalytic residue, because the imidazole ring of His66 lies between the carboxylate O atoms of Glu67 and Asp214. As pointed out by Holm & Sander (Holm & Sander, 1994), Thr68 and Gln118 make hydrogen bonds to the protein backbone, playing a similar role in maintaining the geometry of the active site as the highly conserved residues Ser36 and Gln58 in human lysozyme. Side-chain O atoms of Gln63, Thr64, Ser65, Thr69 and Thr195 point in unfavorable directions

so that these residues can also be ruled out. Therefore, it is concluded that one of the remaining two residues Asn199 and Glu203 is likely to be the equivalent of HEWL Asp52. The amino-acid sequence around these residues is indeed one of the most highly conserved regions in plant chitinases (Fig. 8). The above distance information, however, is ambiguous in distinguishing between the two. If the catalytic mechanism of BCH is similar to that of HEWL, the negatively charged Glu203 would be preferred. However, further experimental data regarding the reaction mechanism of BCH are required before an unambiguous choice can be made.

Besides Glu67, Asn199 and Glu203, most of the residues which form the surface of the presumed active-site cleft are well conserved. They include residues His66, Thr69, Phe86, Glu89, Tyr96, Ile117, Tyr123, Asn124, Trp156, Phe157, Gln162, Lys165, Ile198, Ile202, Asp214 and Arg215. They are likely to be involved in recognizing and binding the substrate. The invariant residue Asn124 was indeed proposed to hydrogen bond to the substrate by Holm & Sander (Holm & Sander, 1994). However, it is not likely to be the equivalent of HEWL Asp52, because it is located a little too far from Glu67 (Fig. 10). Chemical modification (Verburg, Smith, Lisek & Huynh, 1992) and mutagenesis experiments (Verburg, Rangwala, Samac, Luckow & Huynh, 1993) on a tyrosine residue of homologous chitinases equivalent to Tyr123 of BCH suggest that this residue does not participate directly in catalysis. The presence of aromatic and aliphatic residues at the active-site cleft are consistent with the significant non-polar protein-carbohydrate interactions found in pig pancreatic  $\alpha$ -amylase (Qian *et al.*, 1994).

### 3.6. Solvent structure and the crystal packing

Bound solvent molecules, mostly waters, are an integral part of the protein structure. Often some of the bound waters play structurally and/or functionally important roles. However, in order to locate the solvent molecules in the crystal with confidence, the protein structure has to be refined at high resolution, typically better than about 2.5 Å. The previous structure determination by Hart *et al.* (1993) was made at 2.8 Å and thus the solvent structure around the chitinase molecule was not reported. In this study, the structure has been refined at 2.0 Å resolution and thus it was possible to identify a total of 279 ordered solvent molecules, modeled as water, in association with the two independent chitinase molecules in the asymmetric unit. Most of the waters are bound to the surface of the protein. Exceptions are the eight buried water molecules, which occupy structurally equivalent positions inside the protein, in each chitinase molecule in the asymmetric unit.

As shown in Fig. 11, the molecules of BCH are tightly packed in the crystal. This explains the excellent diffraction capability of the present crystal form. The

Table 2. Direct intermolecular hydrogen-bonding contacts

Possible hydrogen bond		Distance (Å)	Symmetry
Glu89 (OE2)	Ser100 (OG)	3.45	2*
Gly133 (O)	Asp635 (N)	3.38	1
Asp135 (N)	Gly633 (O)	3.15	1
Gly174 (O)	Ser507 (OG)	2.65	N.c.s.†
Ser177 (OG)	Ile505 (O)	2.96	N.c.s.
His206 (O)	Gly543 (N)	3.48	2
Asp224 (OD2)	Tyr730 (N)	3.05	1
Tyr230 (OH)	Gly731 (O)	3.31	1
Tyr230 (N)	Asn732 (OD1)	2.91	1
Gly231 (O)	Asn732 (ND2)	3.00	1
Lys526 (O)	Ser593 (OG)	2.89	2
Asp595 (N)	His706 (O)	3.30	2
Gly608 (O)	Ser710 (N)	2.88	2
Arg610 (NE)	Asp709 (OD2)	3.15	2
Arg610 (NH2)	Asp709 (OD2)	2.65	2
Asp641 (OD1)	Arg711 (NH2)	3.33	2
Asp646 (OD2)	Arg740 (NH1)	3.22	1

\* The atoms in the second column belong to the molecule generated by applying the following symmetry operation to the molecule in the first column: (1)  $x+1, y+1, z+1$ ; (2)  $-x, y+\frac{1}{2}, -z$ . † N.c.s., non-crystallographic symmetry.

Table 3. Water-mediated intermolecular hydrogen-bonding contacts

Possible hydrogen bond			Distance (Å)		Symmetry
Asp32 (O)	Wat 438	Thr547 (OG1)	2.79	2.59	2*
Asp95 (OD2)	Wat 314	Gly71 (O)	3.09	2.82	2
Thr98 (O)	Wat 934	Arg90 (O)	3.46	2.69	2
Pro99 (O)	Wat 934	Arg90 (O)	3.19	2.69	2
Pro99 (O)	Wat 989	Gln118 (OE1)	3.44	2.97	2
Ser100 (OG)	Wat 924	Glu89 (OE2)	2.66	2.68	2
Tyr123 (O)	Wat 339	Gly91 (O)	2.67	2.65	2
Trp176 (O)	Wat 847	Ser504 (O)	3.01	3.24	N.c.s.†
Trp176 (O)	Wat 847	Ile505 (O)	3.01	3.47	N.c.s.
Ser177 (OG)	Wat 847	Ser504 (O)	3.48	3.24	N.c.s.
Ser177 (OG)	Wat 847	Ile505 (O)	3.48	3.47	N.c.s.
Tyr230 (N)	Wat 993	Asn732 (ND2)	3.48	2.98	1
Gly231 (N)	Wat 993	Asn732 (ND2)	2.79	2.98	1
Ser238 (OG)	Wat 961	Gly548 (N)	3.49	2.81	2
Arg592 (O)	Wat 453	Gly705 (O)	2.63	3.26	2
Ala601 (O)	Wat 992	Asp20 (OD2)	2.84	2.76	1
Gly608 (O)	Wat 960	Asp709 (OD2)	3.04	2.91	2
Arg610 (NE)	Wat 960	Asp709 (OD2)	2.91	2.91	2
Arg630 (NH1)	Wat 407	Arg130 (O)	2.56	2.98	1
Arg630 (NH2)	Wat 407	Arg130 (O)	2.93	2.98	1
Arg683 (O)	Wat 848	Asp532 (OD1)	3.14	2.99	1
Ala684 (O)	Wat 903	Thr530 (OG1)	3.25	2.61	1
His706 (O)	Wat 370	Asp595 (OD2)	3.25	2.94	2
Gln708 (O)	Wat 997	Pro607 (O)	2.80	2.91	2
Asn733 (ND2)	Wat 421	Asn232 (OD1)	3.40	2.61	1
Arg740 (NH2)	Wat 903	Thr530 (OG1)	2.97	2.61	1

\* The atoms in the third column belong to the molecule generated by applying the following symmetry operation to the molecule in the first column: (1)  $x+1, y+1, z+1$ ; (2)  $-x, y+\frac{1}{2}, -z$ . † N.c.s., non-crystallographic symmetry.

packing interactions involve mostly salt bridges or hydrogen bonds. Intermolecular hydrogen bonds are made either directly between the protein atoms or are mediated by water molecules (Tables 2 and 3). Despite



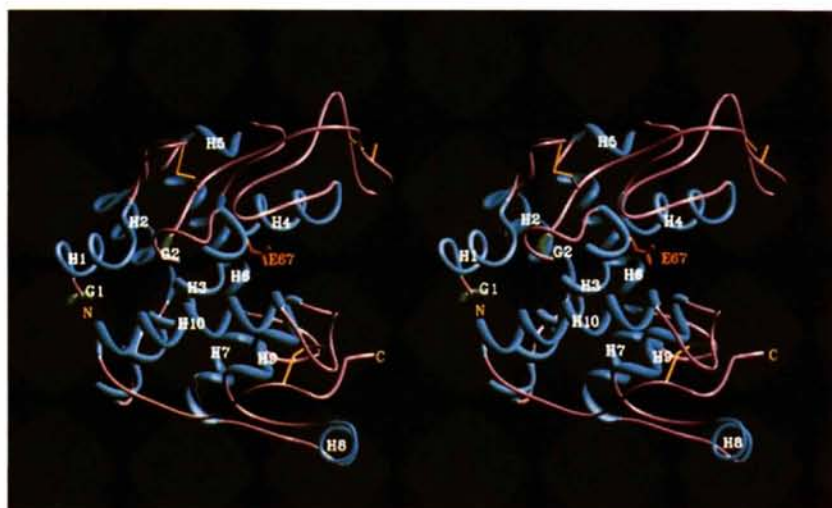


Fig. 7. Schematic drawing of BCH made with the program *SETOR* (Evans, 1993). Sky-blue ribbons, green ribbons, and pink ropes represent  $\alpha$ -helices,  $3_{10}$ -helices and the loops. The N- and C-termini of the chain are indicated. The side chain of the catalytic residue Glu67 at the active-site cleft is shown in orange. The three disulfide bonds are drawn in yellow. The secondary-structure elements are assigned as follows: (i) ten  $\alpha$ -helices, H1, 7–16; H2, 30–41; H3, 49–68; H4, 122–132; H5, 142–145; H6, 147–159; H7, 167–174; H8, 180–185; H9, 191–198; H10, 210–226 and (ii) two  $3_{10}$ -helices, G1, 2–4; G2, 79–81. The programs *PROCHECK* (Laskowski *et al.*, 1993), *YASSPA* in *O* (Jones, Zou, Cowan & Kjeldgaard, 1991), and *SETOR* (Evans, 1993) were used for the assignment of secondary structures.

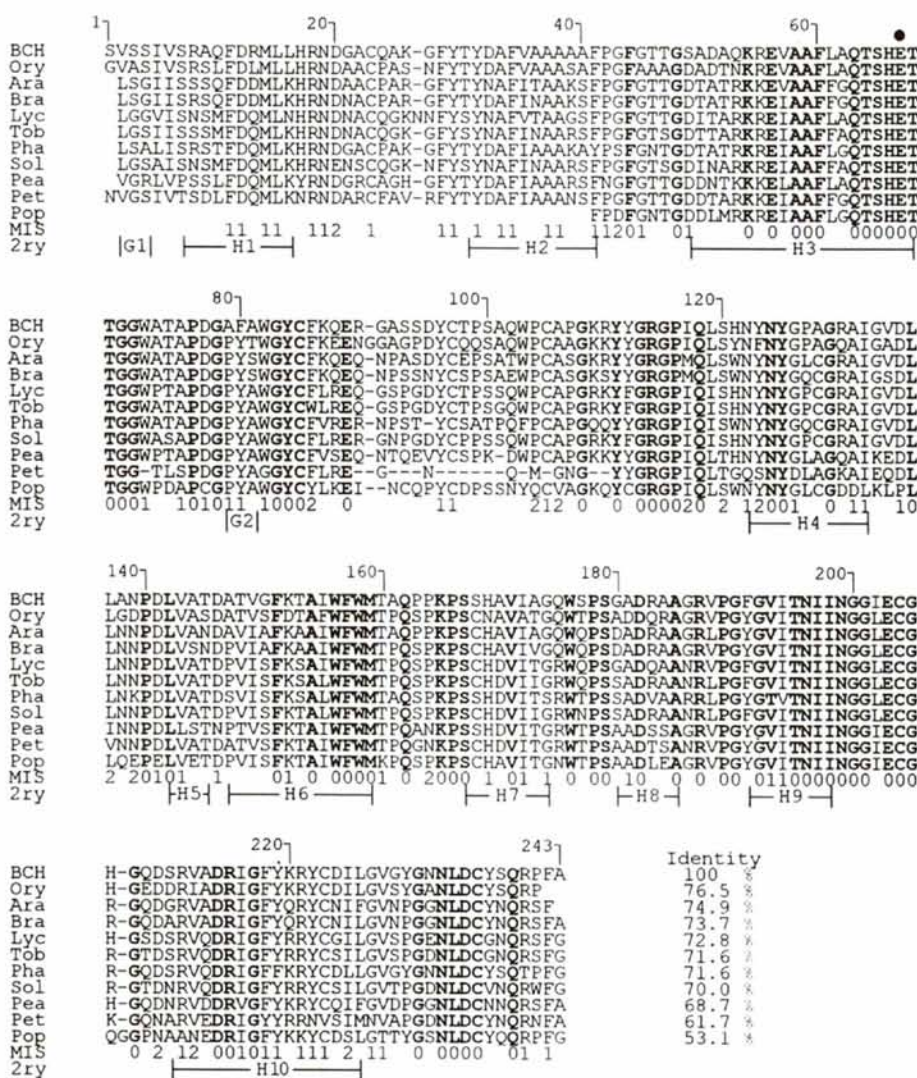


Fig. 8. Amino-acid sequence alignment of 11 plant chitinases. Nine basic chitinases [from barley (BCH), rice (Ory), mouse-ear cress (Ara), rape (Bra), tomato (Lyc), tobacco (Tob), kidney bean (Pha), potato (Sol), garden bean (Pea)] and two acidic chitinases [from petunia (Pet), western balsam poplar (Pop)] are aligned. Our assignment of the secondary structural elements of barley chitinase is also given [H and G represent  $\alpha$ - and  $3_{10}$ -helices, respectively]. The numbering system is for the amino-acid sequence of BCH deduced from the nucleotide sequence. Residues conserved in all 11 sequences are in bold. In addition, the number of mismatches up to two is given below the sequence for highly conserved residues.



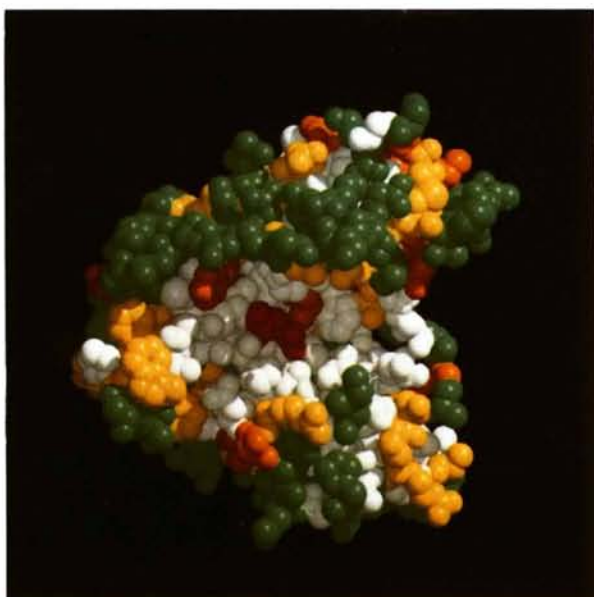


Fig. 9. CPK representation of BCH. Residues conserved in all 11 plant chitinases are colored white except the catalytic residue Glu67, which is drawn in red. Those residues with one or two mismatches are colored yellow or orange, respectively. The remaining residues are in green. This view is obtained by rotating the chitinase model in Fig. 7 around the vertical axis by about  $90^\circ$ .

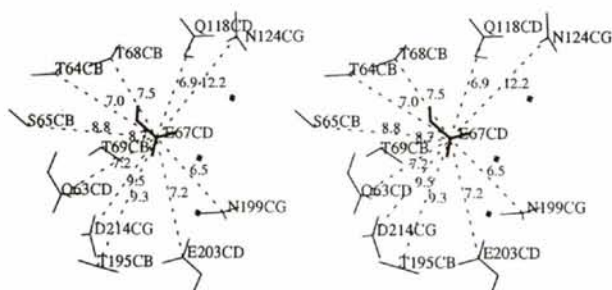


Fig. 10. The negatively charged or polar residues, except histidine, around the catalytic residue Glu67 in molecule 1. The distances from the carboxylate C atom of Glu67 are given. Water molecules within 6 Å from the carboxylate C atom of Glu67 are also shown as  $x$ .



Fig. 11. Stereo diagram illustrating the packing of BCH molecules in the monoclinic crystal. There are four protein molecules in the unit cell. Some neighboring molecules are included for clarity. Thick and thin lines represent the two independent molecules in the asymmetric unit and neighboring molecules, respectively. Molecule 1 is in the lower left corner of the unit cell and molecule 2 at the center. Crystallographic  $a$  and  $c$  axes are horizontal and vertical, respectively.

the fact that BCH is rich in  $\alpha$ -helices, residues in the loop regions make more intermolecular contacts than those in  $\alpha$ -helices. The active-site cleft of molecule 1 is filled with solvent only, whereas part of the active-site cleft of molecule 2 (residues 86–92) is involved in intermolecular contacts. (Fig. 11). There are numerous ordered water molecules in the active-site clefts of both molecules and a few water molecules are found close to the presumed catalytic residues. Water molecules within 6 Å distance from the carboxylate C atom of Glu67 are shown in Fig. 10. Among them, functionally significant water molecules seem to be the ones which make hydrogen bonds to both Glu203 and His66 (Glu203 OE1...O Wat821, 2.62 Å and His66 NE2...O Wat821, 2.90 Å; Glu703 OE1...O Wat442, 2.71 Å and His566 NE2...O Wat442, 2.82 Å) and to Glu67 in molecule 1 (Glu67 OE1...O Wat885, 2.76 Å and Glu67 OE2...O Wat885, 3.00 Å). No water is bound to the catalytic residue Glu567 in molecule 2. In pig pancreatic  $\alpha$ -amylase, two water molecules exist between the side-chain atoms of Glu233 and Asp300 in the free structure, while one of them is displaced by about 1 Å upon inhibitor binding (Qian *et al.*, 1994).

This work was supported by the Korea Science and Engineering Foundation (grant No. 92-24-00-04) and Center for Molecular Catalysis. We thank the Inter-University Center for Natural Science Research Facilities for providing the X-ray equipment and computer graphics. The atomic coordinates and structure factors have been deposited with the Protein Data Bank.\*

\* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1CNS, R1CNS5F). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: AS0689). At the request of the authors, the structure factors will remain privileged until 23 June, 1999.

## References

- Beintema, J. J. (1994). *FEBS Lett.* **350**, 159–163.
- Brogliè, K. E., Gaynor, J. J. & Brogliè, R. M. (1986). *Proc. Natl Acad. Sci. USA*, **83**, 6820–6828.
- Brünger, A. T. (1992a). *X-PLOR Manual, Version 3.1: A System for Crystallography and NMR*. New Haven: Yale University Press.
- Brünger, A. T. (1992b). *Nature (London)*, **355**, 472–474.
- Brünger, A. T. & Krukowski, A. (1990). *Acta Cryst.* **A46**, 585–593.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Crowther, R. A. (1972). *Fast Rotation Function*. In *The Molecular Replacement Method, a Collection of Papers on the Use of Non-Crystallographic Symmetry*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon and Breach.
- Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
- Danhash, N., Wagemakers, C. A. M., Van Kan, J. A. L. & De Wit, P. J. G. M. (1993). *Plant Mol. Biol.* **22**, 1017–1029.

- Dickinson, K., Keer, V., Hitchcock, C. A. & Adams, D. J. (1989). *J. Gen. Microbiol.* **135**, 1417–1421.
- Evans, S. V. (1993). *J. Mol. Graphics*, **11**, 134–138.
- Eyal, Y. & Fluhr, R. (1991). *Oxford Surv. Plant Mol. Cell Biol.* **7**, 223–254.
- Gaynor, J. J. (1988). *Nucleic Acids Res.* **16**, 5210.
- Hamel, F. & Bellemare, G. (1993). *Plant Physiol.* **101**, 1403–1403.
- Hart, P. J., Monzingo, A. F., Ready, M. P., Ernst, S. R. & Robertus, J. D. (1993). *J. Mol. Biol.* **229**, 189–193.
- Hejgaard, J. & Bjørn, S. E. (1985). *Physiol. Plant.* **64**, 301–307.
- Hol, W. G. J. (1985). *Prog. Biophys. Mol. Biol.* **45**, 149–195.
- Holm, L. & Sander, C. (1994). *FEBS Lett.* **340**, 129–132.
- Jeuniaux, C. (1966). *Methods Enzymol.* **8**, 644–650.
- Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Larson, S. B., Greenwood, A., Casio, D., Day, J. & McPherson, A. (1994). *J. Mol. Biol.* **235**, 1560–1584.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Leah, R., Tommerup, H., Svendsen, I. & Mundy, J. (1991). *J. Biol. Chem.* **266**, 1564–1573.
- Linthorst, H. J. M., Vanloon, L. C., Memelink, J. & Bol, J. F. (1991). *Plant Mol. Biol.* **15**, 521–523.
- Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
- Messerschmidt, A. & Pflugrath, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- Mikami, B., Degano, M., Hehre, E. J. & Sacchettini, J. C. (1994). *Biochemistry*, **33**, 7779–7787.
- Molano, J., Polacheck, I., Duran, A. & Cabib, E. (1979). *J. Biol. Chem.* **254**, 4901–4907.
- Parsons, T. J., Bradshaw, H. D. Jr & Gordon, M. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 7895–7899.
- Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S. & Vorgias, C. E. (1994). *Structure*, **2**, 1169–1180.
- Qian, M., Haser, R., Buisson, G., Duce, E. & Payan, F. (1994). *Biochemistry*, **33**, 6284–6294.
- Roberts, W. K. & Selitrennikoff, C. P. (1988). *J. Gen. Microbiol.* **134**, 169–176.
- Sack, J. S. (1988). *J. Mol. Graphics*, **6**, 244–245.
- Samac, D. A., Hironaka, C. M., Yallaly, P. E. & Shah, D. M. (1990). *Plant Physiol.* **93**, 907–914.
- Schlumbaum, A., Mauch, F., Vögeli, U. & Boller, T. (1986). *Nature (London)*, **324**, 365–367.
- Shinshi, H., Mohnen, D. & Meins, F. Jr (1987). *Proc. Natl Acad. Sci. USA*, **84**, 89–93.
- Shinshi, H., Neuhaus, J.-N., Ryals, J. & Meins, F. Jr (1990). *Plant Mol. Biol.* **14**, 357–368.
- Song, H. K., Hwang, K. Y., Kim, K. K. & Suh, S. W. (1993). *Proteins*, **17**, 107–109.
- Vad, K., De Neergaard, E., Madriz-ordenana, K., Mikkelsen, J. D. & Collinge, D. B. (1993). *Plant Sci.* **92**, 69–79.
- Van Scheltinga, A. C. T., Kalk, K. H., Beintema, J. J. & Dijkstra, B. W. (1994). *Structure*, **2**, 1181–1189.
- Verburg, J. G., Rangwala, S. H., Samac, D. A., Luckow, V. A. & Huynh, Q. K. (1993). *Arch. Biochem. Biophys.* **300**, 223–230.
- Verburg, J. G., Smith, C. E., Lisek, C. A. & Huynh, Q. K. (1992). *J. Biol. Chem.* **267**, 3886–3893.
- Zhu, Q. & Lamb, C. J. (1991). *Mol. Gen. Genet.* **226**, 289–296.