

# Structural Analysis Shows Five Glycohydrolase Families Diverged From a Common Ancestor

JON D. ROBERTUS,\* ARTHUR F. MONZINGO,  
EDWARD M. MARCOTTE, AND P. JOHN HART

*Department of Chemistry and Biochemistry, University of Texas, Austin,  
Texas 78712*

**ABSTRACT** We have solved the X-ray structure of barley chitinase and bacterial chitosanase. Structural constraints predicted these would work by an inverting mechanism, which has been confirmed biochemically. The two enzymes were compared with lysozymes from goose (GEWL), phage (T4L), and hen (HEWL). Although the proteins share no significant amino acid similarities, they are shown to have a structurally invariant core containing two helices and a three-stranded  $\beta$  sheet that form the substrate binding and catalytic cleft. These enzymes represent a superfamily of hydrolases arising from the divergent evolution of an ancient protein. The glycohydrolase superfamily can be structurally divided into a bacterial family (chitosanase and T4L), and a eucaryotic family represented by chitinase, GEWL, and HEWL. Both families contain the ancestral core but differ at the amino and carboxy termini. The eucaryotes have a small N terminal domain, while the prokaryotes have none. The C terminal domain of the eucaryotic family contains a single  $\alpha$ -helix, while the prokaryotic domain has three antiparallel helices. *J. Exp. Zool.* 282:127-132, 1998. © 1998 Wiley-Liss, Inc.

Because the metabolism of polysaccharides is so crucial to modern life forms, it is reasonable to assume that it is a very ancient activity. It is also reasonable to assume that the ancient ancestor enzymes may have diverged widely over this long time. Henrissat and Bairoch ('93) have classified over 50 families of glycohydrolases based on their amino acid sequences. Among these are chitinases, chitosanase, and lysozymes, all involved in the hydrolysis of  $\beta$ -1-4 linked polysaccharides. Chitinases cleave chitin, a linear polymer of N-acetylglucosamine (GlcNAc), chitosanases hydrolyze chitosan, a polymer of aminoglucose (GlcN) polymer, and lysozymes hydrolyze peptidoglycans found in bacterial cell walls. These generally contain alternating  $\beta$ -1-4 linked residues of GlcNAc and N-acetylmuramic acid.

We recently solved the X-ray structure of a 243 residue chitinase isolated from barley (Hart et al., '93, '95). This protein is homologous to a large family of plant chitinases and is the structural archetype for that family. That is, the structure observed for the barley enzyme is likely to be an excellent model for any of the enzymes in the family; a ribbon drawing of barley chitinase is shown in Fig. 1A. We also solved the structure for the 238 residue chitosanase from *Streptomyces* N174 (Marcotte et al., '96). Again, this protein is homologous to a number of other enzymes, so our

model is likely to represent an entire family of bacterial chitosanases. A ribbon drawing of chitosanase is shown in Fig. 1B.

A major goal of any structural analysis is to describe the way in which substrates are bound to the active site of an enzyme and how catalysis may occur. We have built models of substrate binding to both barley chitinase and to the bacterial chitosanase. In this way we could predict which residues along the cleft were likely to be involved in substrate binding. The putative interactions between barley chitinase and its substrate is shown in Figure 2. Similar interactions are made between chitosanase and its substrate. Because chitosan is a polycation, however, more of the interactions are specific ion pairings between the C2 amines and enzyme carboxylates. In this figure, water is shown being polarized by a base, Glu 89, and attacking the substrate sugar D. Bond cleavage occurs between the D and E sugars.

Mechanisms of action for glycohydrolases fall into two broad classes, retaining and inverting,

---

Grant sponsor: National Institutes of Health; Grant numbers: GM 30048 and GM35989; Grant sponsor: Foundation for Research and the Welch Foundation.

\*Correspondence to: Dr. Jon D. Robertus, Department of Chemistry and Biochemistry, Welch Hall 5.266, University of Texas, Austin, TX 78712. E-mail: Robertus@utbc01.cm.utexas.edu

Received 27 March 1998; Accepted 27 March 1998

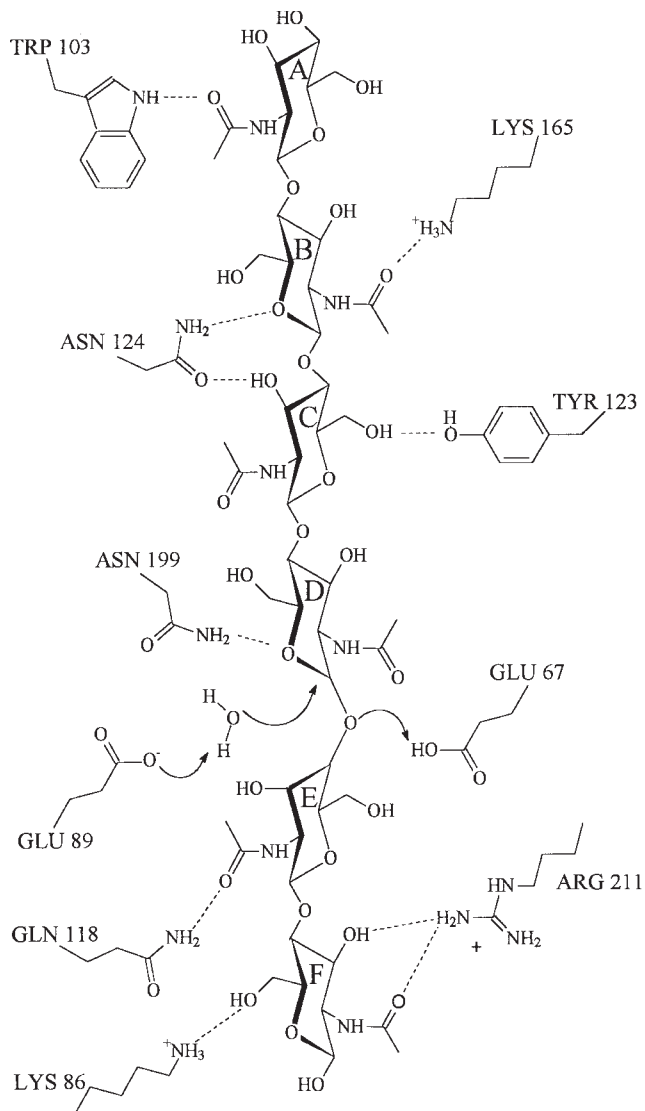
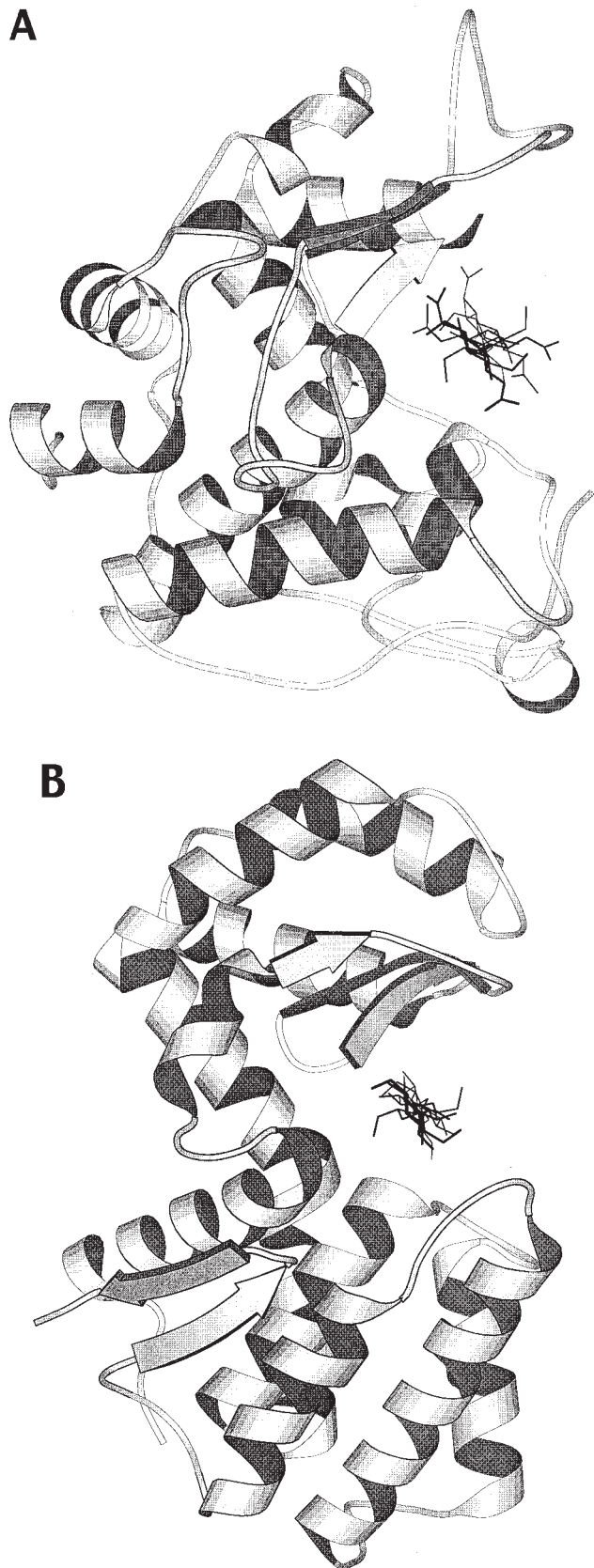


Fig. 2. Models of substrate binding to the active sites of barley chitinase. Sugars are labeled A through F from the nonreducing end. Hydrolysis by water breaks the substrate between sugars D and E.

as shown in Fig. 3. Our model building suggested that two carboxylates were responsible for the catalysis by both enzymes. These were Glu 67 and Glu 89 in barley chitinase, and Glu 22 and Asp 40 in chitosanase. In both cases we hypothesized that the mechanism would be an inverting one, because the space between the "second carboxy-

Fig. 1. The structures of two novel glycohydrolases. (A) A ribbon representation of barley chitinase based on the 1.8 Å structure (Hart et al., '95). (B) A ribbon drawing of chitosanase based on the 2.4 Å structure (Marcotte et al., '96). This figure was drawn using the program MOLSCRIPT (Kraulis, '91).

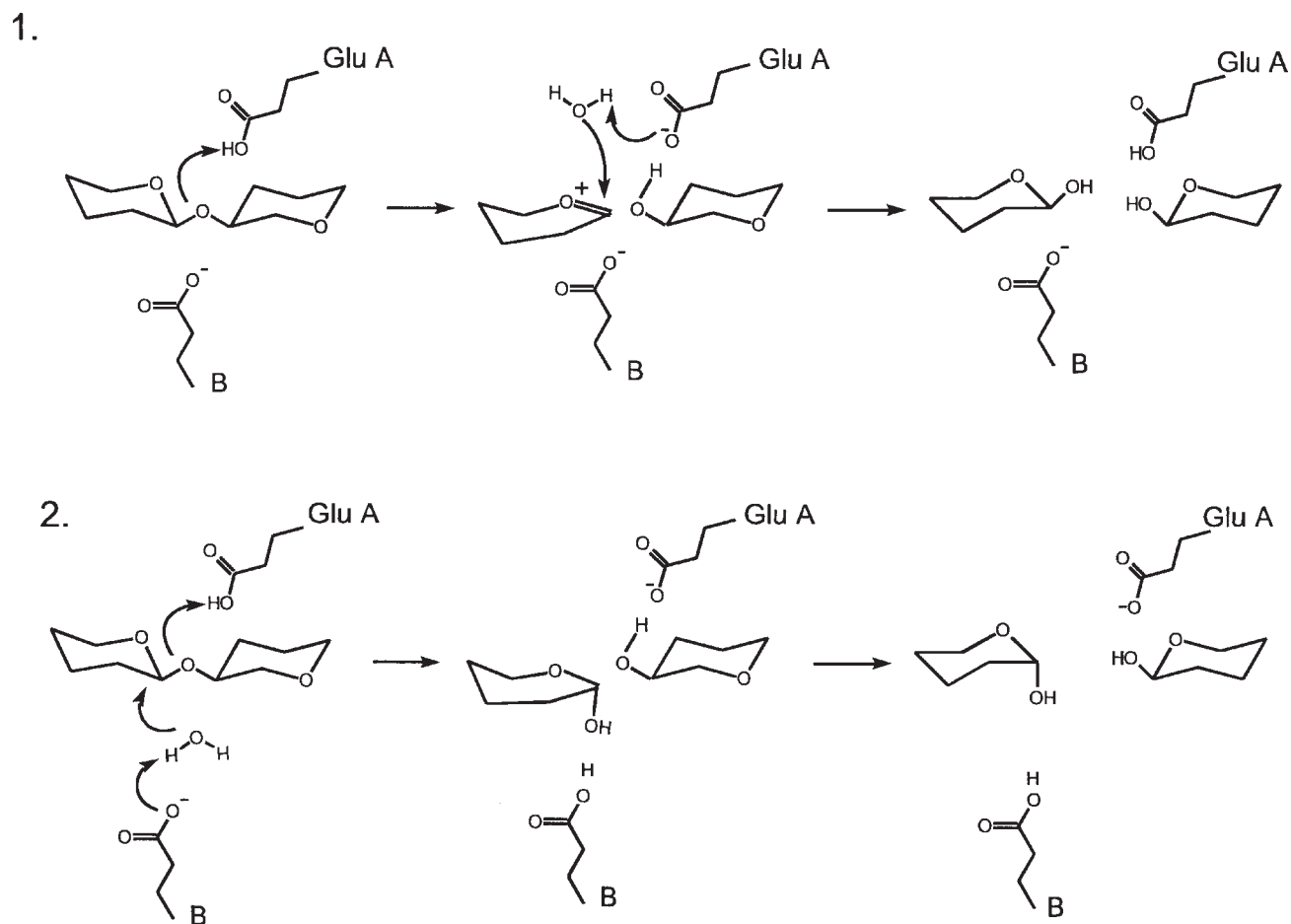


Fig. 3. Two common mechanisms of action for glycohydrolases. The top panel (1) shows three steps in a retaining mechanism. The leaving group is protonated by an acid labeled Glu A. The cationic transition state is electrostatically stabilized by an interaction with a second carboxylate, B. Water attacks from the same side as the leaving group,

retaining the  $\beta$  anomeric configuration of the sugar. The lower panel (2) shows an inverting mechanism. Here, the second carboxylate acts as a general base to activate water for a nucleophilic attack on the backside of the susceptible glycosidic bond, inverting the anomeric position to the  $\alpha$  form.

late" and the susceptible glycosidic bond demanded that an attacking water be interposed (Fig. 2). This mechanism was confirmed biochemically by Dr. Tamo Fukamizo, who showed that the hydrolysis products were in fact inverted from the  $\beta$  to the  $\alpha$  anomer (Fukamizo et al., '95; Hollis et al., '97).

Prior to our solution of the chitinase and chitosanase structures, the structures of three lysozymes, each representing a separate class of enzymes, had been solved. These were the lysozyme from hen egg white, HEWL, (Blake et al., '65); bacteriophage T4, T4L (Blake et al., '65); and GEWL, from goose (Grutter et al., '83). HEWL and T4L were compared and it was initially proposed that the two molecules were unrelated, consistent with their lack of obvious

amino acid sequence similarity (Matthews and Remington, '74). It was subsequently shown by structural comparisons that, despite the lack of sequence similarity, HEWL and T4L were almost certainly related by divergent evolution (Rossmann and Argos, '76; Matthews et al., '81). Although the two proteins differed in size and total folding pattern, they shared several major elements of secondary structure and their active sites appear to bind substrate in a similar fashion. Later the structure of GEWL was compared to these, and shown to have a similar core structure (Grutter et al., '83). As will be clear from this review, however, these earlier comparisons are complicated by a lack of information about larger structural patterns pertinent to the procaryotic or eucaryotic ori-

gin of the enzymes. Once several members of each group were available, these patterns could be discerned.

Although chitinase and chitosanase showed no amino acid sequence similarity to any of the lysozymes, the clearly bilobal and highly helical nature of the proteins suggested that they might be structurally related. We carried out a comparison in which secondary structural elements,  $\alpha$  helices and  $\beta$  sheets, were superimposed (Monzingo et al., '96). One example of the many pairwise comparisons made is shown in Fig. 4, a least squares superposition of chitosanase and T4L, a protein of 164 amino acids. In this particular pairing we found that 106 residues in various secondary structural elements occupied essentially the same relationships in space and they differed by a root

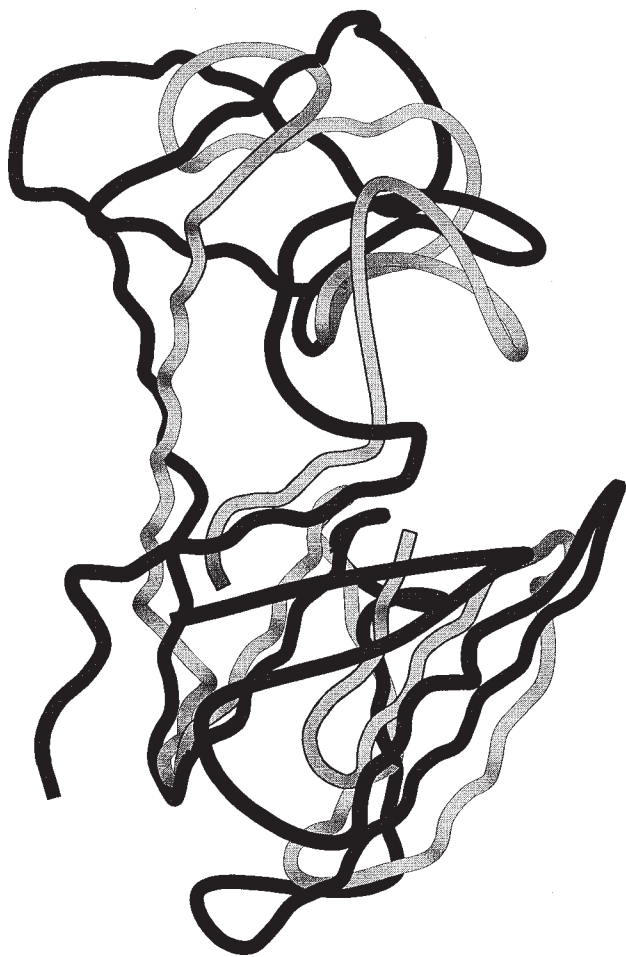


Fig. 4. The superposition of bacterial chitosanase and phage T4 lysozyme. The larger chitosanase protein is black and T4L is gray. The proteins are aligned in roughly the same way as Fig. 1, that is, looking down the prominent substrate binding cleft on the right-hand side of the figure.

mean squared distance of only 3.7 Å. In our study we compared all possible pairs of enzymes and these comparisons led to the conclusion that, despite lacking any significant sequence homology, the five proteins shared a common core structure. This core forms a globular unit with an elongated polysaccharide binding site. It is roughly 100–150 residues long and contains a number of helices and sheets that occupy the same position and orientation in space. The larger cores, like those of chitinase, have fairly large inserts in loop regions, but the folding pattern is always the same.

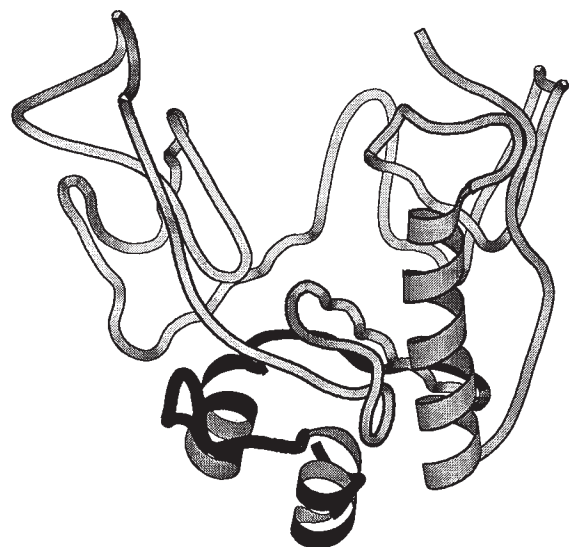
Of particular importance is the conservation of a three-stranded  $\beta$  sheet and two crossing helices within the core, which taken together form the polysaccharide binding and catalytic site of each enzyme. The geometry of this subsection of the core is very well conserved (Monzingo et al., '96). Table 1 shows the residues that compose the ancient core protein in all five modern glycohydrolases. HEWL has only 129 residues, compared to 243 for chitinase, and most of its structural elements appear to have been reduced to a minimalist level. For example, HEWL is the only protein of the group lacking helix I; this has been replaced by an extended chain along the same space.

It is also clear from our analysis that the ancient core protein has been added to during the evolution of the modern glycohydrolases. The procaryotic members of this superfamily, bacterial chitosanase and T4L, lack an amino (N) terminal domain; the three eucaryotic members possess one, of around 50 residues. The N terminal domain of HEWL, like most elements of that protein, has been reduced to a bare minimum. The procaryotic enzymes have a large carboxy, or C, terminal domain. In both observed cases it has three  $\alpha$ -helices and is about 80 residues in length. In contrast, the C terminal domain of the eucaryotes is smaller, about 40 residues, and contains a single helix. This is also shown in Table 1 but displayed more graphically in Fig. 5.

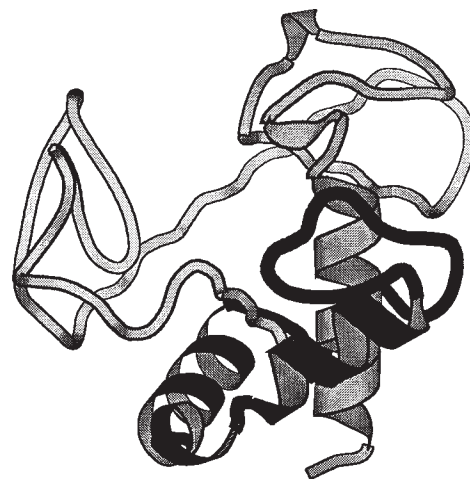
It is reasonable to conclude from the structural analysis that the 150 residue conserved core of the

TABLE 1. Domains of the glycohydrolase superfamily

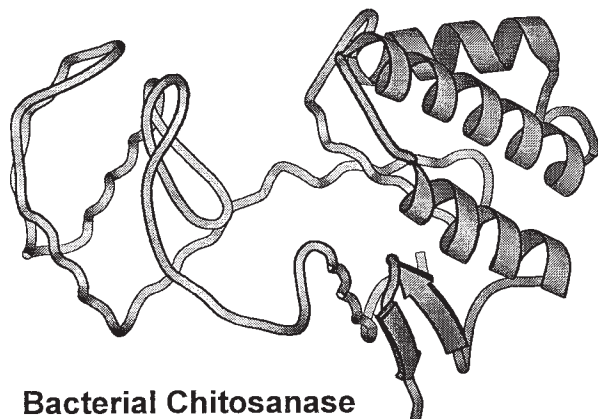
Protein	Core residues	N-terminal residues	C-terminal residues
T4 lysozyme	1–105	—	106–164
Chitosanase	1–151	—	152–236
Chitinase	49–199	1–48	200–243
GEWL	62–147	1–61	148–185
HEWL	25–109	1–24	110–129



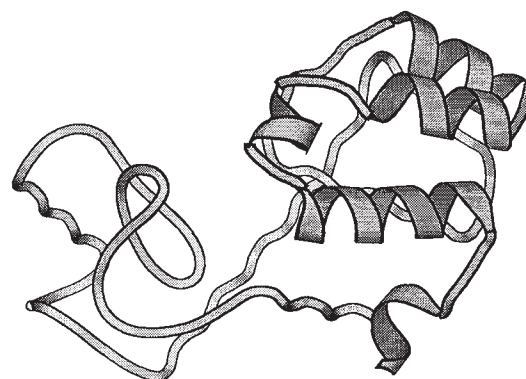
Barley Chitinase



Goose Lysozyme



Bacterial Chitosanase



Phage Lysozyme

Fig. 5. A gallery from the glycohydrolase superfamily. The top row shows two eucaryotic enzymes, barley chitinase and GEWL, while the lower row shows two procaryotes, chitosanase and T4L. In each case the conserved ancestral cen-

tral core is shown as a light gray coil, the carboxy terminal domains in a wide ribbon is shaded a darker gray, and the amino terminal domains are shown as wide black ribbon.

glycohydrolases represents the folding of an ancient  $\beta$  glycohydrolase. This protein has existed since before the procaryotic/eucaryotic split. After the split, procaryotes modified the core hydrolase by adding a relatively large C terminal domain. It is unclear how this improved the enzyme, but presumably it increased stability or allowed a wider range of substrate affinities to evolve. The eucaryotes modified the core protein by adding a small N terminal and a small C terminal domain. Since it is now known that chitinase, chitosanase and T4L are all cata-

lytic inverters, it is reasonable to suppose that the ancestral core enzyme also was an inverter. HEWL is probably the most thoroughly studied member of the superfamily and is known to have a retaining mechanism. This is unusual and probably arose as the enzyme evolved. HEWL is very small, only 129 residues, and has undergone substantial modification from the ancestral protein. All other members have the mechanistic second carboxylate (B in Fig. 2) on the first strand of the active site three-stranded  $\beta$ -sheet. This position allows room for the

interposing water that becomes the nucleophile in the hydrolytic reaction. HEWL has the second carboxylate on the second strand, bringing it too close to the bound polysaccharide substrate to accommodate a water. As a consequence, the mechanism of action changed to the retaining form.

As described above, we feel that the plant chitinase, bacterial chitosanase, and lysozymes from hen, goose, and phage all belong to a single superfamily—despite a lack of any significant sequence identity. It is the conserved protein fold that shows the relationship. We believe these proteins diverged from a common ancestor. Is it not possible that they have converged to a similar fold from widely different parents? This is possible but very unlikely. The best example of clearly convergent protein evolution is that between the mammalian serine proteases, like chymotrypsin, and the bacterial proteases like subtilisin. A comparison of these proteins shows that they have identical active site geometries but that the rest of the protein molecules exhibit no folding or topological similarities (Robertus et al., '72). In this case radically different polypeptide folding units evolved to form catalytic pockets which are essentially optimum for peptide hydrolysis, but this site is clearly embedded in widely differing matrices. The similar topology among the five glycohydrolases argues strongly that the proteins diverged from an ancient ancestor and have filled a number of related roles centered on the hydrolysis of  $\beta$  linked polysaccharides.

#### LITERATURE CITED

- Blake, C.C.F., D.F. Koenig, G.A. Mair, A.C.T. North, D.C. Phillips, and V.R. Sarma (1965) Structure of hen egg-white lysozyme. *Nature*, *206*:757–761.
- Fukamizo, T., Y. Honda, S. Goto, I. Boucher, and R. Brzezinski (1995) Reaction mechanism of chitosanase from *Streptomyces* sp. N174. *Biochem. J.*, *311*:377–383.
- Grutter, M.G., L.H. Weaver, and B.W. Matthews (1983) Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? *Nature*, *303*:828–831.
- Hart, P.J., A.F. Monzingo, M.P. Ready, S.R. Ernst, and J.D. Robertus (1993) Crystal structure of an endochitinase from *Hordeum vulgare* L. seeds. *J. Mol. Biol.*, *229*:189–193.
- Hart, P.J., H.D. Pflugger, A.F. Monzingo, T. Hollis, and J.D. Robertus (1995) The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. *J. Mol. Biol.*, *248*:402–413.
- Henrissat, B., and A. Bairoch (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.*, *293*:781–788.
- Hollis, T., Y. Honda, T. Fukamizo, E.M. Marcotte, and J.D. Robertus (1997) Kinetic analysis of barley chitinase. *Arch. Biochem. Biophys.*, *344*:335–342.
- Kraulis, J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.*, *24*:946–950.
- Marcotte, E., A.F. Monzingo, S.R. Ernst, R. Brzezinski, and J.D. Robertus (1996) X-ray structure of an anti-fungal chitosanase from *Streptomyces* N174. *Nature Struct. Biol.*, *3*:155–162.
- Matthews, B.W., and S.J. Remington (1974) The three dimensional structure of the lysozyme from bacteriophage T4. *Proc. Natl. Acad. Sci. USA*, *71*:4178–4182.
- Matthews, B.W., M.G. Grutter, W.F. Anderson, and S.J. Remington (1981) Common precursor of lysozymes of hen egg-white and bacteriophage T4. *Nature*, *290*:334–335.
- Monzingo, A.F., E.M. Marcotte, P.J. Hart, and J.D. Robertus (1996) Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. *Nat. Struct. Biol.*, *3*:133–140.
- Robertus, J.D., R.A. Alden, J.J. Birktoft, J. Kraut, J.C. Powers, and P.E. Wilcox (1972) An X-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry*, *11*:2439–2449.
- Rossmann, M.G., and P. Argos (1976) Exploring structural homology of proteins. *J. Mol. Biol.*, *105*:75–96.