npg © 1996 Nature Publishing Group http://www.nature.com/nsmb

Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core

Arthur F. Monzingo, Edward M. Marcotte, P. John Hart and Jon D. Robertus

Barley chitinase, bacterial chitosanase, and lysozymes from goose (GEWL), phage (T4L) and hen (HEWL) all hydrolyse related polysaccharides. The proteins share no significant amino-acid similarities, but have a structurally invariant core consisting of two helices and a three-stranded β-sheet which form the substrate-binding and catalytic cleft. These enzymes represent a superfamily of hydrolases which are likely to have arisen by divergent evolution. Based on structural criteria, we divide the hydrolase superfamily into a bacterial family (chitosanase and T4L) and a eucaryotic family represented by chitinase and GEWL. Both families contain the core but have differing N- and C-terminal domains. Inclusion of chitinase and chitosanase in the superfamily suggests the archetypal catalytic mechanism of the group is an inverting mechanism. The retaining mechanism of HEWL is unusual.

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

Correspondence should be addressed to J.D.R.

Metabolism of polysaccharides is crucial to most life forms and is presumably a very ancient activity. Among the enzymes involved in breakdown of sugar-based polymers are cellulases, chitinases, chitosanases and lysozymes. Endochitinases catalyse the hydrolysis of chitin, a linear polymer of β-1–4 linked N-acetylglucosamine (GlcNAc). Chitosanases hydrolyse chitosan, a β-1–4 linked aminoglucose (GlcN) polymer, which may contain some fraction of GlcNAc. Lysozymes hydrolyse peptidoglycans found in bacterial cell walls; these contain alternating β-1–4 linked residues of GlcNAc and N-acetylmuramic acid. These polysaccharides resemble one another chemically and structurally. Because of this, one might anticipate that chitinases, chitosanases and lysozymes would be structurally related enzymes. Indeed, some lysozymes are very good chitinases; HEWL hydrolyses chitin about one-half as efficiently as it attacks its natural substrate. However, amino-acid sequences show no obvious similarity among these three types of enzymes.

Endochitinases are generally monomeric proteins between 25,000–40,000 M_r. We have recently solved the X-ray structure of the 243-residue barley enzyme[1]. Amino-acid sequence alignment shows that this protein is representative of a large number of plant chitinases. A model for chitin binding and for the catalytic mechanism have been proposed[2]. In this issue we report the first X-ray structure of a chitosanase, from *Streptomyces* N174[3]. Binding of the polycationic chitosan substrate for this 238-residue enzyme has been modelled and a plausible mechanism of action proposed.

The X-ray structures of three lysozymes, each representing a class of enzymes, have been solved and the analysis has led to an interesting view of the evolution of that enzyme family. The first lysozyme X-ray structure was that of the 129-residue HEWL[4]. Comparison with the 164-residue bacteriophage T4 lysozyme (T4L)[5] reveals that the proteins had no apparent similarity in amino-acid sequence and the structures appeared to differ significantly. Although they were both folded into two distinct lobes, it was initially concluded that HEWL and T4L were probably evolutionarily unrelated. It was subsequently shown by structural comparisons that, despite the lack of sequence similarity, HEWL and T4L were almost certainly related by divergent evolution[6,7]. Although different in size and total folding pattern, the proteins share major elements of secondary structure and their active sites appear to bind substrate in a similar fashion. Structural conservation was shown to be much stronger than amino-acid conservation. This notion was extended to include lysozyme from goose egg-white (GEWL). Again, the 185-residue GEWL lacks any obvious sequence similarity to HEWL or T4L, and although the protein fold taken as a whole is novel, GEWL contains the core structure of the lysozyme fold[8]. This analysis showed that, after alignment of secondary structural elements, HEWL and T4L shared 74 equivalent residues with an r.m.s. deviation of 3.8 Å. GEWL shared about 90 residues compared to either HEWL or T4L, deviating by about 3.2 Å. Finally, the X-ray structure of the 618-residue periplasmic soluble lytic transglycosylase (SLT) revealed it to contain a 167-residue C terminal domain structurally related to HEWL[9]. Again there is no obvious sequence similarity between the proteins.

Holm and Sander[10] have suggested that barley chitinase may resemble the lysozyme family, based upon a computer match of selected coordinates of the models. A detailed analysis of the refined X-ray structure suggested similarities between chitinase and HEWL which

© 1996 Nature Publishing Group http://www.nature.com/nsmb

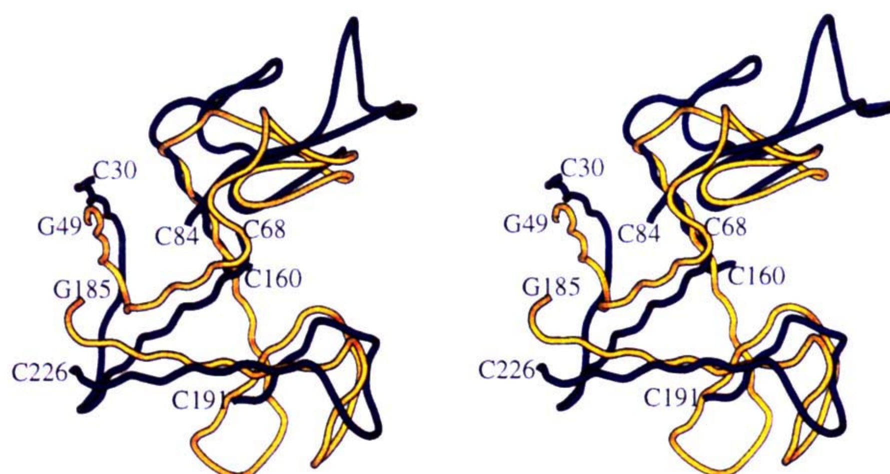

Fig. 1 Superposition of chitinase and GEWL. Chitinase (blue) and GEWL (yellow) share substantial structural overlap in this comparison. GEWL residues 49–185 are shown. Chitinase residues are shown between 30–226, although several elongated loops, such as residues 69–83, have been removed for clarity. This figure was drawn using the program MOLSCRIPT[25].

facilitated the modelling of substrate in the chitinase structure[2]. In this paper we show that the bacterial chitosanase has strong structural relations to chitinase and to the three lysozymes. We quantify the structural relationships, describe key features of the secondary structural elements conserved across this extremely broad superfamily, and suggest some evolutionary relationships among the proteins.

Comparison of five hydrolases

The five enzymes involved in this comparison vary considerably in size, from 243 residues for chitinase to 129 residues for HEWL. It is clear that even if the structures are related in an evolutionary sense, there must be substantial differences in portions of the polypeptide chains. All of the proteins have a high α-helix content, a bi-lobal structure, and a pronounced active-site cleft with a three-stranded β-sheet forming one wall. These features can be used as a rough guide to structural alignments.

Coordinates of HEWL, T4L and the trisaccharide NAM-NAG-NAM bound to HEWL were taken from the Brookhaven Protein Data Bank (accession numbers 6LYZ, 2LZM and 9LYZ, respectively). The coordinates for GEWL were obtained from Brian Matthews and Larry Weaver. Model observation, manipulation and least-squares fitting were done on an Evans and Sutherland PS390 graphics system using the program FRODO[11].

Fig. 1 shows a superposition of major segments of two proteins, barley chitinase and GEWL, aligned in a least squares sense. Similar alignments have been made for each pair of proteins (see Fig. 4 of accompanying paper[3] for a comparison of chitosanase with T4L). It is clear that many elements of secondary structure are shared between chitinase and GEWL.

Our analysis suggests that chitinase and GEWL have 81 Cα positions which are structurally conserved. This means they are in analogous places on corresponding secondary structural elements. The r.m.s. difference between these 81 atoms was 3.2 Å. Atoms occupying similar places in space, but contributed from different portions of the peptide chain were not counted as homologues. The Holm and Sander[12] elastic similarity score for these 81 atoms was 584, indicating significant structural homology. For comparison, the HOMOLOGY program of Rossmann and Argos[6] matched 101 Cα atoms between chitinase and GEWL, obtaining an r.m.s. deviation of 4.2 Å; and a similarity score of 568. Table 1 summarizes these scores for all 10 pairs of enzymes.

Fig. 2*a* summarizes the elements of secondary structure, arranged along the peptide chain, which can be aligned in space among the five enzymes. Chitinase helix C (residues 49–68) has a structural homologue in each of the other four proteins. The elements of a three-stranded sheet, downstream from this helix, are also present in all five proteins. The only other feature found in all five enzymes is the homologue of chitinase helix F. Other structural features are conserved among subsets of this five-member group. For example, helices structurally equivalent to chitinase helix B are found in GEWL and HEWL, while homologues of helix D are found in T4L and chitosanase. It is possible to expand this comparison (Fig. 2*b*), showing the amino acids within the various elements which were aligned by the least-squares analysis.

Two α-helices (C and F in chitinase) and three antiparallel β-strands which form a sheet are the only invariant elements of secondary structure among chitinase, chitosanase, and the three lysozymes (Fig. 2*a*). Fig. 3 shows a least-squares superposition of these elements from all five proteins. Binding of the polysaccharide substrate is included for HEWL[13]. It is evident that the invariant secondary structural elements form the heart of the substrate-binding and catalytic pocket in this group of enzymes. These elements form the core structure which is conserved in all five of these hydrolases.

Structurally important residues

All of these enzymes are polysaccharide hydrolases; they probably bind substrate in similar orientations, and may act through similar or related mechanisms. The best studied mechanism is that of HEWL. HEWL has a binding site which can accommodate six hexose sugars in sites called A–F from the nonreducing end. Hydrolysis occurs between sugars D and E in what is generally thought to be an S_N1 mechanism[14]. Glu 35 protonates the leaving group O4. Oxocarbonium ion character develops on sugar D and this may be stabilized by Asp 52, a second carboxylate. Water replaces the E sugar and attacks the oxocarbonium, creating a product sugar with the same anomeric configuration (β) as the substrate. That is, the reaction is said to proceed with reten-

© 1996 Nature Publishing Group http://www.nature.com/nsmb

tion of configuration. It has also been argued that retaining enzymes like HEWL involve a transient covalent intermediate between the D sugar and the second carboxylate, Asp 52. It has even been possible to trap covalent intermediates in retaining enzymes like β-glucanase, using 2-fluoro sugar derivatives[15] (reviewed in ref. 16).

In contrast to the retaining mechanism of HEWL, it has been shown that some glycohydrolase mechanisms proceed with inversion of the product anomeric hydroxyl. Such a mechanism is likely to involve coordinated protonation of the leaving group from the β side and attack on the α side by water; the water may be polarized by a nearby base. The chitinase from *Dioscorea opposita* (yam) has been shown by NMR to proceed with inversion of product[17]. Structural analysis suggested that the homologous barley enzyme may also be an inverting enzyme[2], and this has been confirmed chemically (T. Fukamizo, personal communication). The *Streptomyces* chitosanase is an inverting enzyme[3,18] and it has also been suggested that T4L may proceed with inversion of product anomer[19]. This is based on the formation of a covalent intermediate between a substrate and a mutant T4L with an added carboxylate near the position occupied in HEWL by Asp 52.

The glycohydrolases of this superfamily appear to require at least two catalytic residues, whether they are retainers, like HEWL, or the more common inverters. In either case the mechanism requires an acid to protonate the susceptible glycosidic linkage. The least-squares analysis shows that the homologue of Glu 35 from HEWL is structurally conserved in all five enzymes. These Glu residues are 67 in barley chitinase, 73 in GEWL, 22 in chitosanase, and 11 in T4L. All rest near the C-terminal end of the helix labelled C in chitinase (Fig. 2).

The situation is less clear for the second catalytic residue, a carboxylate. Retaining enzymes require a carboxylate very near to sugar D to stabilize the oxocarbonium ion which develops in the S_N1 mechanism. This is provided by Asp 52 in HEWL. In contrast, inverting enzymes are likely to proceed by an S_N2-like mechanism. They may require a second carboxylate which is farther removed from the D sugar; this relatively large separation allows room for an attacking water between the second carboxylate and the back, or α, side of the D sugar. Indeed, it has been observed that the distances between the acid and the second carboxylate are larger for inverting enzymes than for retaining enzymes[20]. The second carboxylate resides in the three-stranded β-sheet but the position is variable. Asp 52 of HEWL is on the second strand, close to the D sugar site. In chitinase, Glu 89 is in the first strand with sufficient room between it and the substrate to accommodate the attacking water. Asp 20 from T4L is also in the first strand of the β-sheet while Asp 40 of chitosanase is in a loop between the first and second strands.

It has been hypothesized that GEWL has no second carboxylate, largely because two candidates, Asp 86 and Asp 97, are too far from the homologous Asp 52 position of HEWL to carry out the same function[21]. If GEWL is an inverting enzyme, then the second carboxylate must be far enough away to accommodate the attacking water. Asp 86 is on the loop between the first and second strand of the β-sheet and appears to be well positioned to serve as the base in an inverting hydrolytic mechanism. Asp 97 is also a plausible candidate; it occupies a position similar to Ser 120 of chitinase, a residue implicated in positioning the attacking water in that enzyme[2].

Outside of the catalytic site, the analysis of amino-acid similarities among the five proteins becomes even more difficult. This is consistent with the observation that standard sequence comparison programs reveal no apparent relationships. Our analysis reveals at least five interactions which appear to help define and stabilize the invariant core structure of the five proteins. To facilitate discussion, Table 2 lists residues in the conserved core elements of each of the proteins which appear to be playing corresponding roles. These residues, along with the catalytic glutamic acids, are shown in bold in Fig. 2*b*.

The features of regions that are outside the catalytic site but which appear to be elements of a common core are as follows:

Table 1 Comparison of chitinase, chitosanase and the lysozymes

	Chitinase	GEWL	Chitosanase	T4L	HEWL
Chitinase		81[1] 3.2 Å[2] 584[3] 15%[4]	77 3.0 Å 481 17%	76 4.0 Å 399 16%	61 2.6Å 357 10%
GEWL	101 4.0 Å 568		63 32 Å 218 8%	58 2.4Å 361 5%	55 1.9Å 353 15%
Chitosanase	90 3.5 Å 560	87 4.2 Å 276		106 3.7 Å 820 13%	45 3.2 Å 75 11%
T4L	74 3.3 Å 381	68 3.0 Å 408	109 3.9Å 787		37 2.6Å 122 11%
HEWL	66 3.9 Å 257	86 3.1 Å 579	65 4.3 Å 77	66 4.1 Å 137	

[1]Number of residues in common.
[2]R.m.s. distance of paired α-carbons.
[3]Elastic similarity score calculated as described by Holm and Sander[12].
[4]Sequence identity of common residues.
Above and to the right of the grey boxes, the table shows information calculated from pair-wise superpositions based upon common secondary structure. These were done initially by aligning corresponding α-carbons of the β-sheet and helices C and F (chitinase nomenclature). For each pair of proteins, the superposition was examined on the graphics system, and additional helices which were topologically consistent and appeared to potentially superimpose were selected to be included in the least-squares superposition. Corresponding atoms were chosen such that the helices could be superimposed with a minimal rotation and translation. In general, helices which had been matched but which had an overall negative similarity score were removed from the superposition. The helix E scores for the HEWL:chitinase and chitinase:chitosanase pairs were positive; for consistency, the match of the E helices of HEWL and chitosanase, which had a negative score, was included in the superposition. For comparison, lbeow and left of the grey boxes are the results of calculations based on pair-wise superpositions performed by the HOMOLOGY program[6] using corresponding α-carbons of the β-sheet and helices C and F as the initial set of atoms. Sequence identity was not calculated for alignments created using this program.

© 1996 Nature Publishing Group http://www.nature.com/nsmb

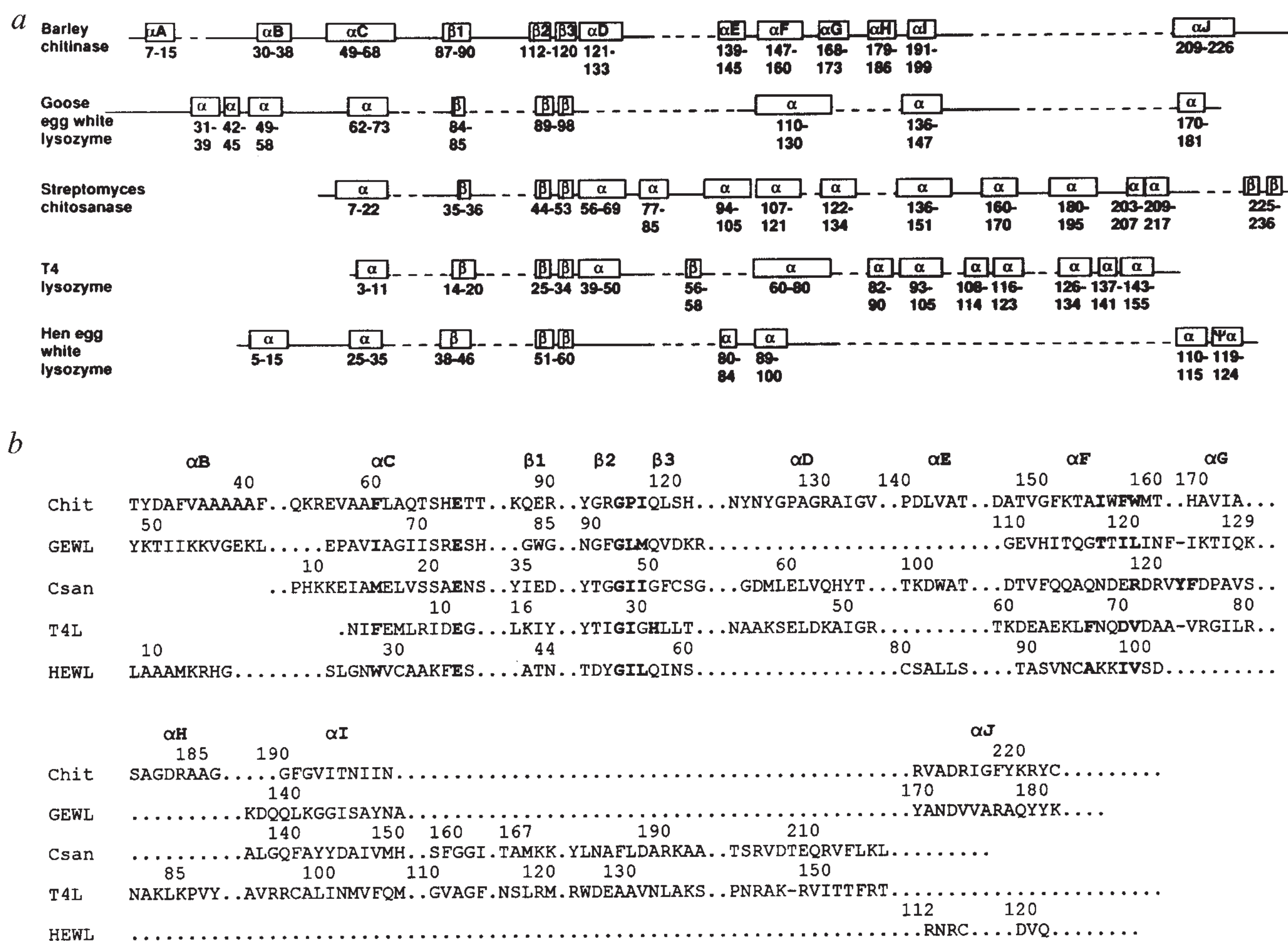

Fig. 2 *a*, Alignment of secondary structure elements between chitinase, chitosanase, and three lysozymes. Boxes are labelled as α-helices (α) or β-strands (β); those with parameters outside the normal values are labelled as pseudo (ψ) Chitinase helices are identified by their order as A–J for discussion in the text. Those elements which align in space are aligned in the same column in this figure. For example, the invariant helices (analogs of αC and αF) form a stack of five. *b*, The amino-acid sequences of the conserved core elements. The amino acids used in the least-square superpositions are shown for each of five enzymes. The catalytic glutamatic acid and the conserved hydrophobic core residues are shown in bold type. The secondary structural elements of chitinase are also indicated.

1. There is always a bulky hydrophobic amino acid seven residues upstream from the catalytic glutamic acid; it resides on the conserved C helix (Fig. 2*b*; Table 2, first column). This is Phe 60 in chitinase, Ile 67 in GEWL, Met 15 in chitosanase, Phe 4 in T4L and Trp 28 in HEWL. In each case the nonpolar side chain projects into the heart of the core structure.

2. The hydrophobic residue from the C helix fits into a hydrophobic socket formed by various residues on the second conserved helix of the core, which is labelled F in chitinase. These contacts are listed in the first column of Table 2. In chitinase, it would be formed by the side chains of Ile 155 and Trp 158, one helical turn downstream on the F helix. The other proteins have similar sockets, although the socket for chitosanase is displaced about a half turn in space. That socket also involves contact of Met 15 with a methylene carbon (Cδ) of Arg 118 whereas all other sockets are composed of purely nonpolar side chains.

3. The β2 strand contains an invariant Gly, the only residue besides the catalytic Glu which is strictly conserved. This is Gly 115 in chitinase and 92, 47, 28 and 92 in GEWL, chitosanase, T4L and HEWL respectively. The Gly appears to be conserved to maintain packing within the core. If any L-amino acid were substituted, its side chain would clash with some atoms from β3, the F helix or other features, depending on the various residues in any given protein. For example, an L-amino acid at 115 in chitinase would collide with side chains of Ile 117 on β3 and with Thr 153 on the F helix. An L-amino acid at position 47 of chitosanase would clash with Phe 51 of β3 and Gln 114 on helix F.

4. The β1 strand of the sheet contains a nonpolar residue that may make several contacts within the core. The most consistent contacts are with a conserved hydrophobic residue on β2 one residue down stream from the invariant Gly described above, although other contacts may also be made. For example, chitinase Tyr 84 from β1 contacts Pro 116 from β2, and also Phe 151 on the F helix.

© 1996 Nature Publishing Group http://www.nature.com/nsmb

Table 2 Conserved interactions of hydrophobic core residues

	αC-------------αF		β1--------β2		β3---------αF	
Chitinase	F60	I155 / W158	Y84	P116	I117	F157
GEWL	I66	T117 / L120	A76	L93	M94	I119
Chitosanase	M15	R118 / F123	Y32	I48	I49	Y122
T4L	F4	F67 / V71	L13	I29	H31	D70
HEWL	W28	A95 / V99	F38	I55	L56	I98

The conserved contacts between β1 and β2 strands are shown in the middle column of Table 2.

5. As shown in the last column of Table 2 there is generally a hydrophobic amino acid two residues downstream from the invariant Gly; this residue marks the beginning of β3. The hydrophobic residue interacts with several other nonpolar groups, but always contacts a nonpolar group on the F helix, which lies between the two socket residues described above. For example, Ile 117 on chitinase β3 contacts Phe 157 on helix F. T4L stands out in Table 2 for this interaction because it has charged residues His 31 on β3 and Asp 70 on the F helix. In fact these two groups form an ion pair or hydrogen bond pair and so conserve the positional interaction, even though it is a polar one in this case.

The conserved interactions among the core elements of chitinase are shown in Fig. 4. Similar interactions are made in the other four proteins although the details may vary. In describing and discussing these interactions it is important to remember that we have attempted to distill out from a myriad of contacts those which appear to be conserved and which may form the main underlying scaffolding for the protein cores. Each core consists of many other interactions which are difficult to relate between proteins but which may be very important for any given protein or even for subgroups of the five proteins.

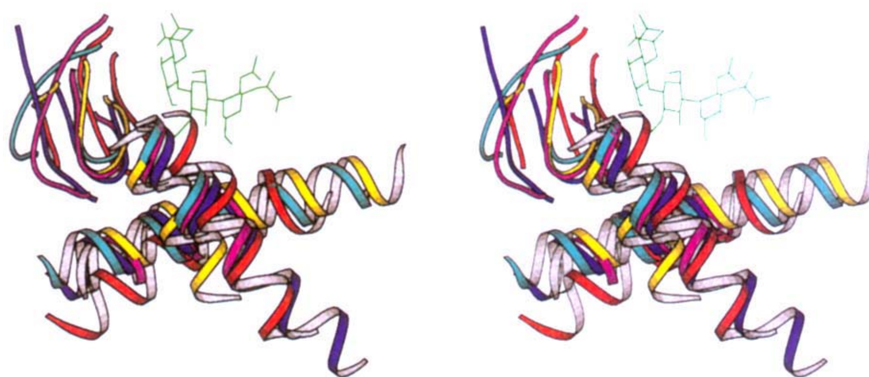

Fig. 3 Invariant structural elements in chitinase, chitosanase, and lysozyme. Three elements of secondary structure, the β-sheet and helices C and F, are superimposed from barley chitinase (blue), chitosanase (red) and three lysozymes: hen egg white (purple), T4 (cyan) and goose egg white (yellow). A trisaccharide observed in HEWL is shown in blue in its proper position with respect to that protein. This figure was drawn using the prgram MOLSCRIPT[25].

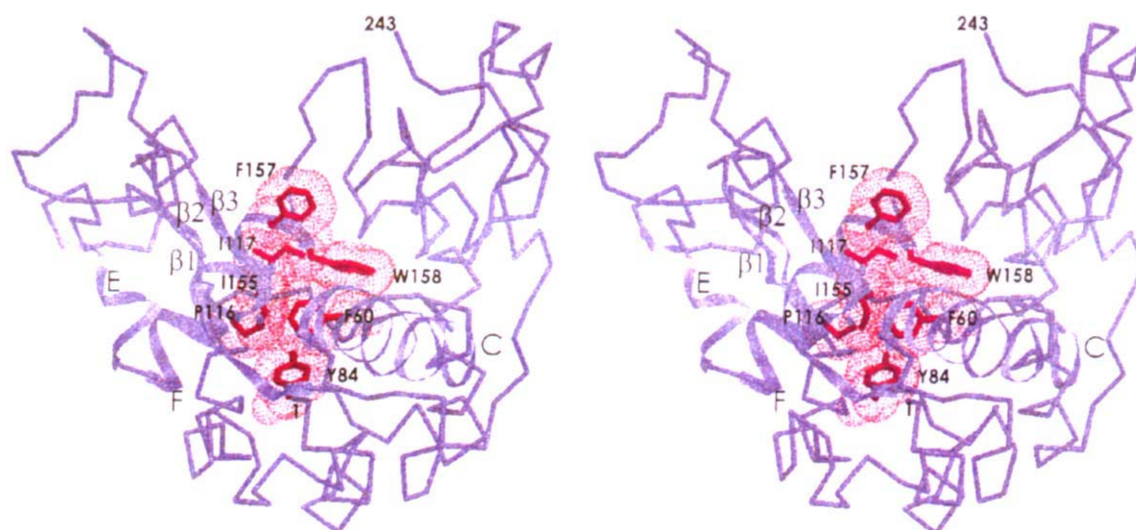

Fig. 4 The hydrophobic interactions securing the invariant core elements from chitinase. The molecular backbone is shown as an Cα trace except the core elements are shown as ribbons. The side chains of key hydrophobic residues are shown as red bonds with Van der Waals surfaces. The figure was produced using the program RASMOL[26].

© 1996 Nature Publishing Group http://www.nature.com/nsmb

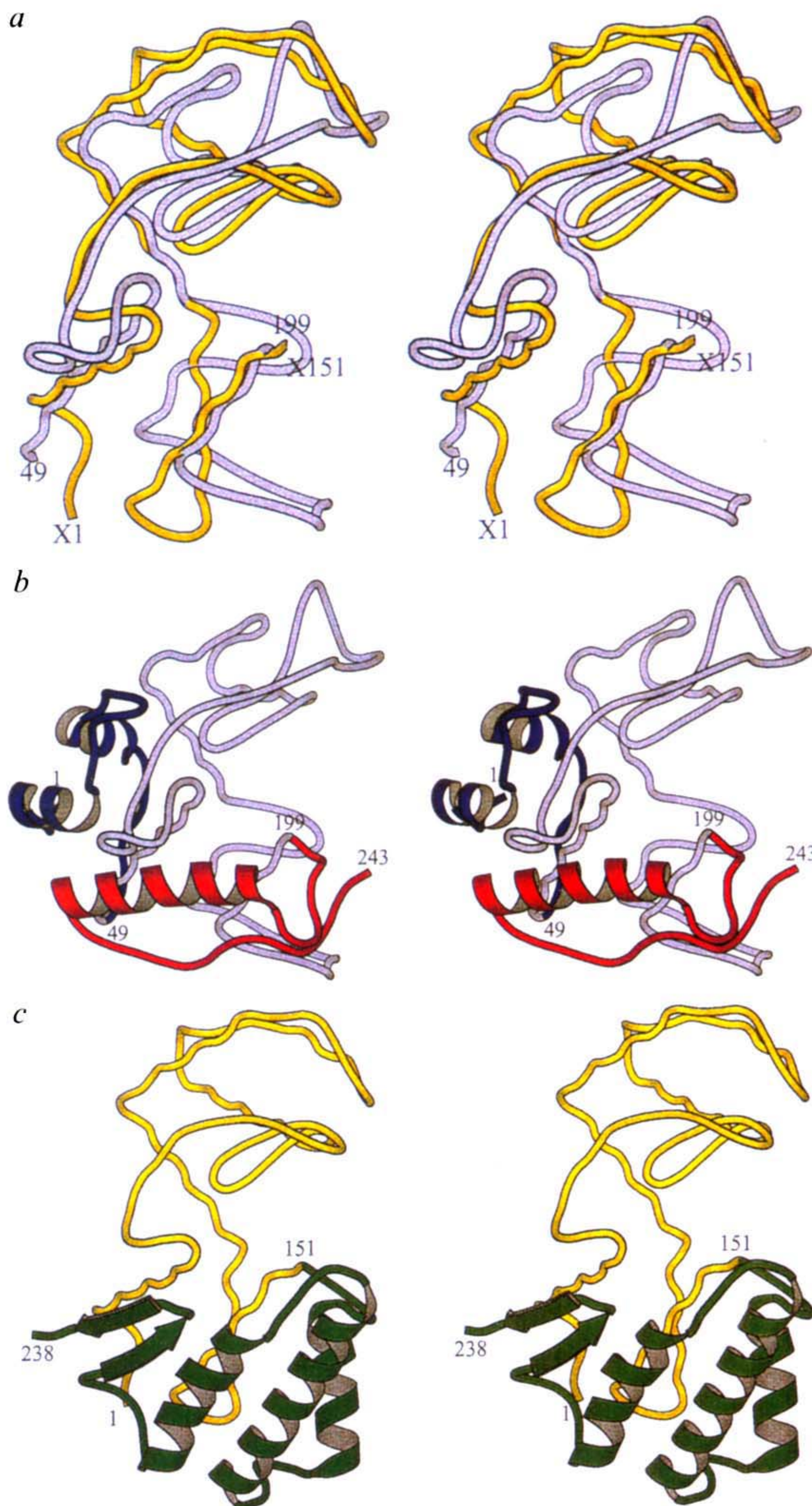

Fig. 5 Structural comparison of the eucaryotic and bacterial families of hydrolases. *a*, A least squares superposition of the extended core of barley chitinase and bacterial chitosanase. The chitinase core, a light blue coil, consists of residues 49–199. The chitosanase extended core, a yellow coil, contains residues X1–X151. *b*, The Cα trace of chitinase. The N-terminal domain, residues 1–48, is blue while the C-terminal domain, residues 200–243, is red. *c*, The Cα trace of chitosanase. The C-terminal domain, residues 152–236, is green. This figure was drawn using the program MOLSCRIPT[25].

Implications for evolutionary relationships

Matthews and co-workers, making pair-wise comparisons of lysozymes, observed that there appeared to be a common core, similar to the one described here. They showed that any one lysozyme might also have structural overlap with another outside the conserved core areas, and these contacts differed markedly between pairs[8]. This is also true when chitinase and chitosanase are added to the comparison.

An obvious question concerning these structural comparisons is whether they reflect divergence from a common ancestor or convergence to a stable folding pattern. Matthews and co-workers concluded that the lysozymes probably reflect divergence from a common ancestor[7] and we feel this is also true when the chitinase and chitosanase proteins are included in the comparison. All five enzymes appear to hold similar substrates in a similar manner with respect to conserved secondary structural core elements and the key catalytic Glu. This situation contrasts with that for the serine proteases chymotrypsin and subtilisin which are considered to be an example of convergent evolution. In that case, substrates are oriented in a very similar manner with respect to a catalytic triad at the active site[22] but the folding pattern and secondary structural elements are very different between the two proteins.

Another argument put forward by the Matthews group in favour of divergence is also strengthened by addition of chitinase and chitosanase to the comparison with lysozymes. That is the argument that the overlap between various nonessential elements of the structures favours divergence from a common ancestor (Fig. 2*a*). The core elements define the substrate-binding and catalytic sites of the enzymes but the overlap of other nonessential structural elements suggests slow divergence from a common ancestor. It seems unlikely for example that the analogue for chitinase helix I would arise by chance in both T4L and GEWL.

Two families

Fig. 2*a* and Table 1 can be analysed to deduce relations among the five hydrolases. The large number of corresponding atoms and high similarity score suggest that the two proteins which are the most similar in structure are bacterial chitosanase and phage T4 lysozyme. It is reasonable to postulate that T4L arose in phage by capture of a bacterial host protein; this pair shows evidence of a relatively recent split.

© 1996 Nature Publishing Group http://www.nature.com/nsmb

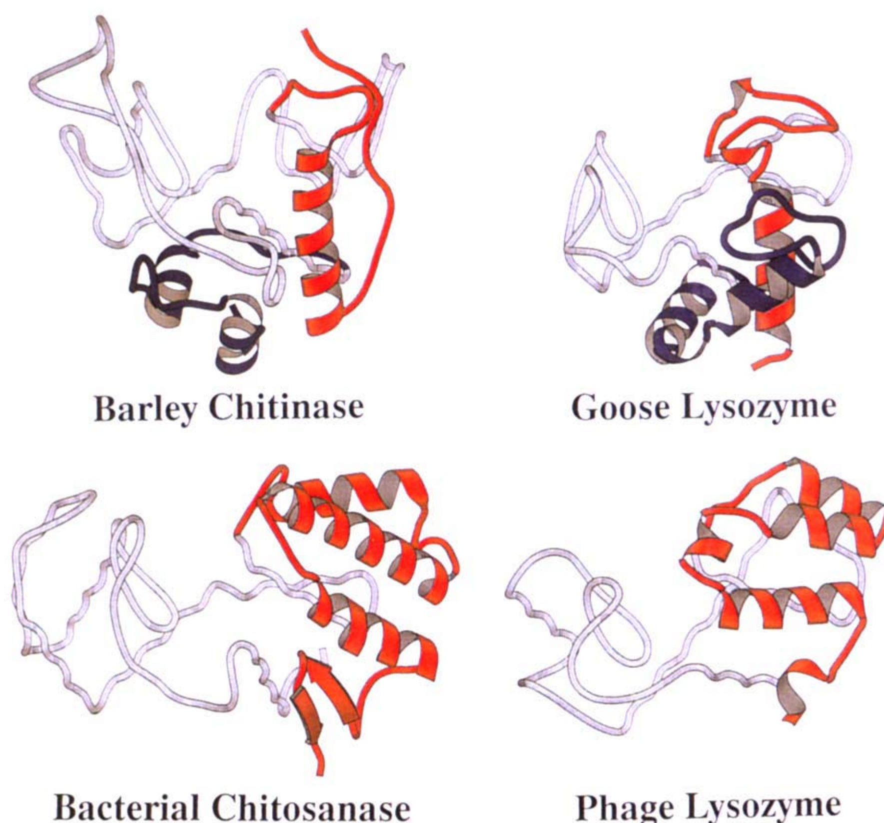

Fig. 6 A gallery from the glycohydrolase superfamily. The top row shows two eucaryotic enzymes, barley chitinase and GEWL, while the lower row shows two procaryotic enzymes, chitosanase and T4L. In each case the conserved central core is shown as a gray ribbon, N-terminal domains are shown in blue and C-terminal domains in red. This figure was drawn using the program MOLSCRIPT[25].

In a structural sense, the next closest pair is barley chitinase and GEWL. It is informative to consider these as representatives of a eucaryotic family of hydrolases, to be compared with the bacterial family, represented by chitosanase and T4L. It is clear from Fig. 2*a* that both families (chitinase/GEWL and chitosanase/T4L) have similar secondary structural elements running from α-helix C through α-helix I, using the chitinase nomenclature. Fig 5*a* shows a superposition of this region of barley chitinase, representing the eucaryotic family, and chitosanase, representing the bacterial family. The region contains the invariant core elements found in all five hydrolases, but the core can be extended to include structural elements through Helix I. The end of Helix I forms the centre of the substrate-binding wall opposite the three-stranded sheet. Each protein has a hydrophobic residue at the end of this helix which anchors it to the other core elements. In chitinase this I helix residue is Ile 198; it makes hydrophobic contacts with groups from helix C, β3, and helix F.

The two families can be seen to differ markedly in their modifications to the extended core. The eucaryotic proteins have several α-helices at the N terminus which are absent in the bacterial family. As shown for chitinase (Fig. 5*b*), these form a folding unit on the opposite side of the enzyme from the active site. The N-terminal domain for GEWL is even larger, containing three well defined helices and additional extended chain. The bacterial enzymes have no extensive structure added to the N terminus of the conserved core. It is unclear if the eucaryotic N-terminal domain has been added to the extended core by the eucaryotes, or deleted by the bacteria.

The two families also differ at the C terminus. As shown (Fig. 5*b*), the eucaryotes add a small domain to the C terminus of helix I to complete the protein. This domain consists of ~40 residues. It begins with a short chain which runs from the end of the I helix along the active site wall. It then forms a helix (J) which lies outside of the core C helix and they cross at a ~45° angle (Fig. 5*b*). The N-terminal half of the J helix forms part of the active site wall. In GEWL the C-terminal domain finishes with the helix, but in chitinase an extended chain runs back up towards, and terminating near, the domain origin.

By contrast, the procaryotic C-terminal domain is longer, roughly 80 residues; it contains three prominent helices, comprising around 60 residues, switching back and forth in a roughly antiparallel fashion. These helices have their axes approximately normal to the elongated active-site cleft, and the helix ends form the top wall of that cleft. Following the helices, the procaryotic C-terminal domain has a polypeptide tail. In chitosanase the tail forms a two-stranded antiparallel β-sheet (Fig. 5*c*) but in T4L it forms a short helix. It is evident that the C-terminal domains from procaryotes and eucaryotes have no amino-acid sequence or structural relationship and appear to represent independent additions to an ancient extended core structure. The structural archetypes of the eucaryotic and procaryotic families are shown by orienting two members of each family in a similar fashion (Fig. 6).

The structure of the catalytic domain (residues 449–618) of the 70,000 M_r bacterial soluble lytic transglycosylase (C-SLT) has been analysed recently[23]. The enzyme is not a proper hydrolase, but instead produces a 1,6-anhydro-muropeptide. The catalytic domain does have a lysozyme-like fold, however, and is thought[23] to have a particularly close resemblance to GEWL. Although atomic coordinates have not been deposited for the protein, visual inspection shows that the structure of C-SLT agrees very well with our new classification of the superfamily into procaryotic and eucaryotic families. Compared to other procaryotic enzymes, C-SLT appears to have a short N-terminal segment which links it to the main body of the protein. This linker does not resemble the extensive N-terminal domains of chitinase or GEWL, but serves to bridge the C-SLT domain

© 1996 Nature Publishing Group http://www.nature.com/nsmb

to the main body of SLT. The core of C-SLT closely resembles that of the glycohydrolases. Tyr 552 in C-SLT plays the role of Ile 198 in chitinase, anchoring the I helix into the conserved core as described above. Following the I helix analog, C-SLT has a three-helix C terminal domain which is characteristic of the other procaryotic enzymes. Compared with the other procaryotic enzymes, the third helix of C-SLT is tilted slightly and falls into a position very similar to that of the single helix of the eucaryotic C-terminal domains. This may account for the apparent structural similarity to GEWL, although the overall fold of C-SLT more closely resembles those of chitosanase and T4L.

Henrissat and Bairoch have proposed a classification scheme for glycosyl hydrolases, based on amino-acid sequences[24]. In this scheme barley chitinase is in class 19, *Streptomyces* chitosanase in 46, and lysozymes GEWL, T4L and HEWL in classes 23, 24 and 22 respectively. It is clear from this and other structural comparisons that many proteins which appear to be unrelated based on sequence are in fact distantly related as seen in their protein fold. Presumably when more structures are known from the many glycosyl hydrolase families it will be possible to reorganize the classification scheme into a few superfamilies, such as that described here. This may be useful in the description of enzyme evolution as well as in the characterization of enzyme mechanism and the rational design of proteins and inhibitors.

HEWL is unusual

HEWL is significantly smaller than the other proteins and may well represent a minimal hydrolase protein. Table 1 shows that HEWL is more closely related to the eucaryotic family than to the bacterial family. Fig. 2*a* confirms this, showing HEWL has the characteristic N-terminal domain of the eucaryotes, although this has been reduced to a single short helix. In addition, all the elements downstream of helix F have been minimized, although they are still recognizably related to the eucaryotic family. Helices G, H and I have been truncated to a short piece of chain. However, this chain continues to form the substrate wall opposite the three-stranded β-sheet. The chain segment is centered on Trp 108, whose bulky side chain anchors it into the invariant core, making hydrophobic contacts with the analogs of helices C and F, and with β3. The helices at the C terminus of HEWL have the position and orientation of the eucaryotic family. HEWL is also the only enzyme of the group so far proven to proceed by a mechanism which retains the β anomer in the product. The other four hydrolases in this superfamily have been proven to, or are likely to, proceed with anomeric inversion. The mechanistic uniqueness of HEWL may be correlated with its small size. For example, evolution of Asp 52 on the second strand of the β-sheet, near the substrate D site, may have allowed the switch from an inverting to a retaining mechanism. Since the first β-strand and long loops between the first and second strands were no longer required, they may have been able to tolerate selective shortening.

Note added in proof. *A paper reporting that T4 lysozyme is an inverting enzyme[27] was published while this manuscript was in review. The finding strengthens our prediction that this family of glycohydrolases evolved from an ancestor with an inverting mechanism of action.*

Received 2 June 1995; accepted 2 January 1996.

Acknowledgments
We wish to thank B. Matthews and L. Weaver for providing the coordinates of goose egg-white lysozyme. This work was supported by grantsfrom the National Institutes of Health, from the Foundation for Research and the Welch Foundation.

1. Hart, P.J., Monzingo, A.F., Ready, M.P., Ernst, S.R. & Robertus, J.D. Crystal structure of an endochitinase from *Hordeum vulgare* L. seeds. *J. Mol. Biol.* **229**, 189–193 (1993).
2. Hart, P.J., Pfluger, H.D., Monzingo, A.F., Hollis, T. & Robertus, J.D. The refined crystal structure of an endochitinase from *Hordeum vulgare* L. seeds at 1.8 Å resolution. *J. Mol. Biol.* **248**, 402–413 (1995).
3. Marcotte, E., Monzingo, A.F., Ernst, S.R., Brzezinski, R. & Robertus, J.D. X-ray structure of an anti-fungal chitosanase from *Streptomyces* N174. *Nature Struct. Biol.* **3**, 155–162 (1996).
4. Blake, C.C.F. *et al.* Structure of hen egg-white lysozyme. *Nature* **206**, 757–761(1965).
5. Matthews, B.W.& Remington, S.J. The three dimensional structure of the lysozyme from bacteriophage T4. *Proc. Natl. Acad. Sci. USA* **71**, 4178–4182 (1974).
6. Rossmann, M.G.& Argos, P. Exploring structural homology of proteins. *J. Mol. Biol.* **105**, 75–96 (1976).
7. Matthews, B.W., Grütter, Anderson, M.G., W.F. & Remington, S.J. Common precursor of lysozymes of hen egg-white and bacteriophage T4. *Nature* **290**, 334–335 (1981).
8. Grütter, M.G., Weaver, L.H. & Matthews, B.W. Goose lysozyme structure: an evolutionary link between hen and bacteriophage lysozymes? *Nature* **303**, 828–831 (1983).
9. Thunnissen, A.W.H., Dijkstra, A.J., Kalk, K.H., Rozeboom, H.J., Engel, H., Keck, W.& Dijkstra, B.W. Doughnut-shaped structure of a bacterial muramidase revealed by X-ray crystallography. *Nature* **367**, 750–753 (1994).
10. Holm, L. & Sander, C. Structural similarity of plant chitinase and lysozyme from animals and phage. *FEBS Lett.* **340**, 129–132 (1994).
11. Jones, T.A. in *Computational Crystallography* (ed. Sayre, D.) 303–317 (Oxford University Press, 1982).
12. Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
13. Kelly, J.A., Sielecki, A.R., Sykes, B.D., James, M.N. G. & Phillips, D.C. X-ray crystallography of the binding of the bacterial cell wall trisaccharide NAM-NAG-NAM to lysozyme. *Nature* **282**, 875–878 (1979).
14. Blake, C.C.F. *et al.* Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. Roy. Soc.* **B167**, 378–388 (1967).
15. Withers , S.G. *et al.* Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. *J. Am. Chem. Soc.* **112**, 5887–5889 (1990).
16. Sinnot, M.L. Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90**, 1171–1202 (1990).
17. Fukamizo, T., Koga, D. & Goto, S. Comparative biochemistry of chitinases-anomeric form of the reaction product. *Biosci. Biotech. Biochem.* **59**, 311–313 (1995).
18. Fukamizo, T., Honda, Y., Goto, S., Boucher, I. & Brzezinski, R. Reaction mechanism of chitosanase from *Streptomyces* sp. N174. *Biochem. J.* **311**, 377–383 (1995).
19. Kuroki, R., Weaver, L.H. & Matthews, B.W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262**, 2030–2033 (1993).
20. Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J. & Withers, S.G. Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* **116**, 11594–11595 (1994).
21. Weaver, L.H., Grütter, M.G. & Matthews, B.W. The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* **245**, 54–68 (1995).
22. Robertus, J.D. *et al.* An X-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry* **11**, 2439–2449 (1972).
23. Thunnissen, A.W.H., Isaacs, N.W. & Dijkstra, B.W. The catalytic domain of a bacterial lytic transglycosylase defines a novel class of lysozymes. *Proteins* **22**, 245–258 (1995).
24. Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J.* **293**, 781–788 (1993).
25. Kraulis, J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
26. Sayle, R. RASMOL V2.5 Molecular visualization program (Glaxo Research and Development, 1994).
27. Kuroki, R., Weaver, L.H. & Matthews, B.W. Structure-based design of a lysozyme with altered catalytic activity. *Nature Struct. Biology* **2**, 1007–1011 (1995).