Three-dimensional structure of the human 'protective protein': structure of the precursor form suggests a complex activation mechanism

Gabby Rudenko[1,2], Erik Bonten[2], Alessandra d'Azzo[2] and Wim GJ Hol[1,3*]

[1]Department of Biological Structure, Biomolecular Structure Program, Box 357742, School of Medicine, University of Washington, Seattle, WA 98195-7742, USA, [2]Department of Genetics, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, USA and [3]Howard Hughes Medical Institute, Box 357742, School of Medicine, University of Washington, Seattle, WA 98195-7742, USA

Background: The human 'protective protein' (HPP) forms a multi-enzyme complex with β-galactosidase and neuraminidase in the lysosomes, protecting these two glycosidases from degradation. In humans, deficiency of HPP leads to the lysosomal storage disease galactosialidosis. Proteolytic cleavage of the precursor form of HPP involves removal of a 2 kDa excision peptide and results in a carboxypeptidase activity. The physiological relevance of this activity is, as yet, unknown.

Results: The crystal structure of the 108 kDa dimer of the precursor HPP has been elucidated by making extensive use of twofold density averaging. The monomer consists of a 'core' domain and a 'cap' domain. Comparison with the distantly related wheat serine carboxypeptidase dimer shows that the two subunits in the HPP dimer differ by 15° in mutual orientation. Also, the helical subdomain forming part of the cap domains is very different. In addition, the HPP precursor cap domain contains a 'maturation' subdomain of 49 residues which fills the active-site cleft. Merely removing the 'excision' peptide located in the maturation subdomain does not render the catalytic triad solvent accessible.

Conclusions: The activation mechanism of HPP is unique among proteases with known structure. It differs from the serine proteases in that the active site is preformed in the zymogen, but is blocked by a maturation subdomain. In contrast to the zinc metalloproteases and aspartic proteases, the chain segment physically rendering the catalytic triad solvent inaccessible in HPP is not cleaved off to form the active enzyme. The activation must be a multi-step process involving removal of the excision peptide and major conformational changes of the maturation subdomain, whereas the conformation of the enzymatic machinery is probably almost, or completely, unaffected.

Structure 15 November 1995, **3**:1249–1259

Key words: human protective protein, lysosomal storage disease, protease precursor activation, serine carboxypeptidase

Introduction

A defect in the human 'protective protein' (HPP), which is thought to be identical to cathepsin A and thus is also known as protective protein/cathepsin A (PPCA), has been identified as the primary genetic lesion underlying galactosialidosis [1], a lysosomal storage disease inherited as an autosomal recessive trait. Patients with this disorder are biochemically diagnosed as having drastically reduced β-galactosidase and neuraminidase activities in the lysosomes [2–4]. Three clinical phenotypes of this syndrome are distinguished according to age of onset: an 'early infantile', a 'late infantile' and a 'juvenile/adult' form. Slightly different spectra of symptoms are observed accompanying each clinical form. These include skeletal dysplasia, dysmorphism, progressive neurological deterioration, reduced life expectancy, and in some cases mental retardation, as well as impairment of cardiac and kidney function (for reviews, see [5,6]).

The molecular basis of galactosialidosis has been shown to be more complicated than G_{M1}-gangliosidosis, Morquio B and sialidosis, syndromes in which either β-galactosidase or neuraminidase activities are deficient (reviewed in [7,8]). In cultured galactosialidosis fibroblasts, β-galactosidase appears to be produced at normal levels [1,9], but its half-life is reduced from ten days to less than one day [10]. Loss of the β-galactosidase activity is due to rapid degradation of the enzyme upon arrival in the lysosomal compartment [11] and it is likely that neuraminidase suffers the same fate. These findings have led researchers to propose, and subsequently demonstrate, a deficiency in galactosialidosis patients of a third lysosomal protein, HPP, which serves to stabilize β-galactosidase and neuraminidase in a multi-enzyme complex [1,9]. Upon uptake of HPP from the culture medium by galactosialidosis fibroblasts, both β-galactosidase and neuraminidase activities are indeed restored [1].

HPP is synthesized as a 542 amino acid precursor with a molecular weight of 54 kDa [12] and dimerizes soon after synthesis in the endoplasmic reticulum [13]. The sequence contains two glycosylation sites (Asn117 and Asn305) and nine cysteines. After transport to the acidic endosomal/lysosomal compartments, the precursor undergoes a protease-mediated maturation process. A polypeptide of approximately 2 kDa, called the 'excision' peptide is removed from the protein yielding a 32 kDa and a 20 kDa chain held together by disulfide bridges [14]. Amino acid sequencing of tryptic peptides from mature HPP isolated from placenta has shown that residue Arg284 is still present in the 32 kDa chain after maturation [12]; while Met299 is the first residue in the

*Corresponding author.

20 kDa chain [12,15]. The excision peptide thus comprises, at most, residues Met285–Arg298. The maturation event can be mimicked *in vitro* by limited proteolysis with trypsin, probably cleaving at sites Arg284, Arg298 as well as Arg292 [14].

Surprisingly, HPP has approximately 30% sequence identity to the wheat and yeast serine carboxypeptidases (abbreviated CPW and CPY respectively), both members of the hydrolase fold family [16–18]. The serine carboxypeptidases have peptidase activity at acidic pH (pH 4.5–5.5) as well as deamidase and esterase activities at pH 7 (reviewed in [19,20]). Mutagenesis studies and enzymatic assays have revealed that HPP, but only its mature form, possesses serine carboxypeptidase activity. Based on sequence alignments with serine carboxypeptidases, the catalytic triad in HPP has been proposed to be formed by the residues Ser150, His429 and Asp372 [12].

HPP appears to be a multi-functional lysosomal enzyme. It occurs in lysosomes in a high molecular weight multi-enzyme complex with β-galactosidase and neuraminidase [21–26], but a large amount of lysosomal HPP is found in an unassociated dimeric form as well [27]. Although the exact physiological substrate(s) of HPP are not known, there is evidence that HPP may be secreted to participate extracellularly in the deactivation of selected bio-active peptides such as endothelin I [15,28–31]. While loss of the protective capacity of HPP is directly linked to galactosialidosis, it is unclear to what extent loss of the enzymatic activity contributes to the clinical phenotype. Given its pleiotropic and distinct functions, as well as the capacity to be secreted and taken up by different cell types, HPP may have a function outside the multi-enzyme complex and possibly outside the lysosomes as well.

This paper describes the crystal structure determination of the precursor form of HPP by a combination of molecular replacement and twofold density averaging. The structure presented here is the first of an enzyme associated with a human lysosomal storage disease, and the third human lysosomal enzyme structure determined after cathepsin B [32] and cathepsin D [33,34].

Results and discussion

The HPP precursor monomer

The HPP precursor monomer has dimensions of 60 Å×50 Å×70 Å and is present as a dimer in our crystal structure. The monomer contains approximately 20% β structure and 30% helical structure. The protein fold can be divided into two domains: a 'core' domain, as commonly found in members of the hydrolase fold family [18] and a 'cap' domain (Fig. 1a). The core domain, comprising residues 1–182 as well as 303–452, contains a central ten-stranded β sheet. An additional ten α helices and two small β strands occur on both sides of the central β sheet. The cap domain can be divided into a 'helical' subdomain consisting of three α helices (residues 183–253) and a 'maturation' subdomain consisting of a three-stranded mixed β sheet (residues 254–302). A Cα trace of the monomer is shown in Figure 1b.

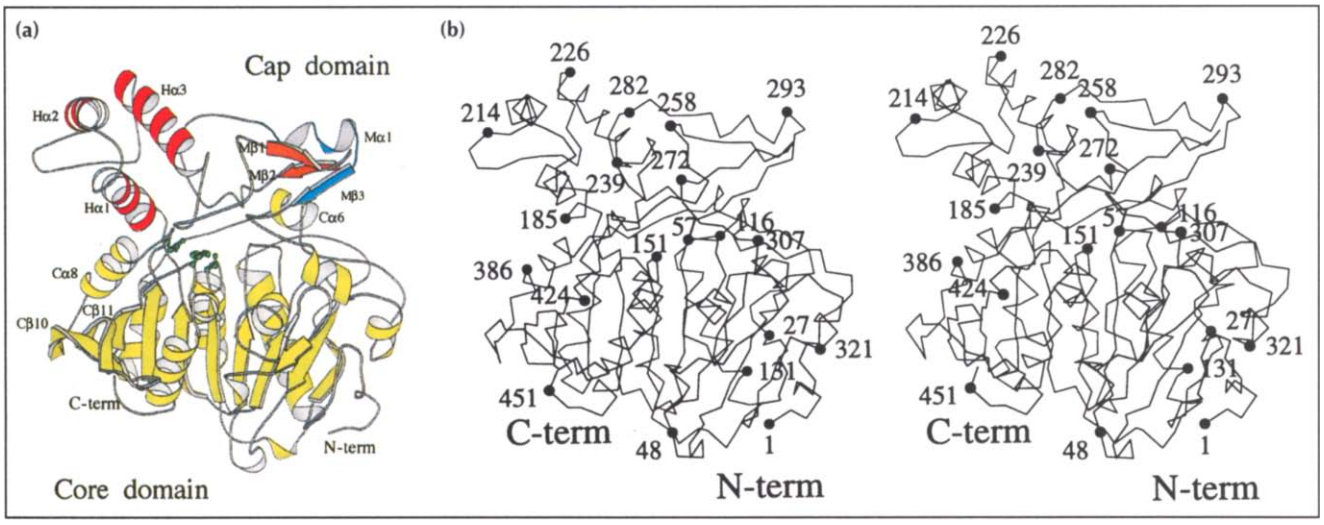

Fig. 1. (a) Schematic ribbon diagram of the HPP monomer (monomer 1). The 'core' domain is shown in yellow. The 'cap' domain consists of a 'helical' subdomain (red) and a 'maturation' subdomain (orange). The 'excision' peptide, located in the maturation subdomain, is shown in light blue. The side chains of the catalytic triad Ser150, His429 and Asp372 (from right to left) are in green. Some secondary structure elements (assigned according to DSSP [58]) are labeled. Core domain: Cβ1, 21–27; Cβ2, 32–39; Cβ3, 50–54; Cα1, 63–67; Cβ4, 73–75; Cβ5, 82–84; Cβ6, 94–98; Cα2, 118–135; Cβ7, 144–149; Cα3, 152–163; Cβ8, 171–177; Cα4, 307–313; Cα5, 316–321; Cα6, 336–341; Cα7, 350–359; Cβ9, 363–369; Cα8, 377–386; Cβ10, 391–401; Cβ11, 407–416; Cβ12, 419–424; Cα9, 431–434; Cα10, 436–447. Cap domain: Hα1, 183–196; Hα2, 202–212; Hα3, 226–240; Mβ1, 261–264; Mβ2, 267–270; Mα1, 290–293; Mβ3, 296–299. Note that for monomer 2 the secondary structure assignments in the cap domain are slightly different from those in monomer 1. Residues in Mβ1 are in a region of poor density and Mα1 is an extended coil. See also text. **(b)** Stereo diagram of the Cα trace of the HPP monomer 1 with numbering of selected residues. (Figure generated using MOLSCRIPT [59].)

There appear to be four disulfide bridges per monomer: Cys60–Cys334, Cys253–Cys303, Cys212–Cys228 and Cys213–Cys218. The first two disulfide bridges connect secondary structure elements forming the active-site cleft. The last two disulfide bridges play a critical role in keeping the helices Hα2 and Hα3 in the helical subdomain together (Fig. 2). A type II′ turn is observed between residues 214 and 217. The only free cysteine (Cys375) is located in the active-site cleft. There is one *cis*-proline (Pro101) observed per subunit. Electron density is clearly visible for two *N*-acetylglucosamine sugar moieties at position Asn117 and one *N*-acetylglucosamine moiety at position Asn305 per monomer. The sugars extend out into the solvent, making few if any interactions with the protein surface.

The HPP precursor dimer

The physiological dimer is V-shaped, with the two monomers related to each other by a rotation κ=179.6°. There are major differences (i.e. greater than 0.7 Å for more than three sequential Cαs) in the two crystallographically distinct monomers forming the dimer, in the regions 211–240, 255–272, 279–296 and 404–406, resulting in a root mean square deviation (rmsd) of 1.27 Å for 449 Cα atoms. The region 211–240, containing Hα3, differs slightly in the two monomers by an approximate rigid-body rotation of 2°. Residues 281–293 not only have a different spatial conformation in the two subunits, but also different secondary structure. In monomer 1, residues 287–289 form a 3_{10} helix immediately followed by an α helix composed of residues 290–293. In monomer 2, a 3_{10} helix is observed for residues 285–288, while residues 289–293 have no distinct secondary structure. If the four regions showing major deviations (67 residues) are excluded from the comparison, the two monomers are very similar and can be superimposed with an rmsd of 0.26 Å for 383 Cα atoms.

HPP precursor dimer interface and accessible surface

The cap domain (Hα1, Hα2 and two loops) and the core domain (Cα8, Cβ10 and Cβ11) contribute approximately equally to formation of the dimer interface, as seen in Figure 3. The helical subdomain from one monomer packs between the core domain and the helical subdomain of the other monomer, while the maturation domains are not involved in the dimer interface. A number of residues appear to play a crucial role at the dimer interface. Residues Tyr196, His197 and Tyr250 interact with their counterpart from the other subunit, while residues Phe377 and Phe412 fit in hydrophobic pockets formed by the symmetry-related monomer. The total solvent-accessible surface of the HPP precursor dimer is about 34200 Å^2 as calculated by GRASP [35]. The dimer interface involves 1600 Å^2 of buried accessible surface per monomer and is approximately 60% non-polar and 40% polar/charged [35]. Thus, the interface buries approximately 10% of the total solvent-accessible surface per monomer.

The surface of the HPP dimer appears to be that of a typical globular protein, tightly packed with no large protrusions except for part of the cap domain (Mβ1, Mβ2 and Mβ3), and loop 402–408. The solvent-accessible surface of the HPP precursor dimer, as calculated by GRASP [35], is 45% non-polar and 55% polar/charged. The surface properties of the protective protein are likely to be tailored precisely for its protective task. Unfortunately, little is known about the exact composition of the multi-enzyme complex, although stoichiometry and binding sites for β-galactosidase to HPP have been proposed [24,36,37].

Crystal packing

Each monomer in the crystal interacts with four non-crystallographically related monomers. The most extensive contact, by far, is that generating the physiological dimer. The other three contacts are crystal contacts between two monomers from different dimers, and average approximately 200–800 Å^2 per monomer. The largest of these crystal contacts involves the excision peptides from two monomers making intimate contact with each other (region 265–267, 281–295 from monomer 1 with residues 281–293 from monomer 2). Together these loops create an intermolecular buried surface of 1680 Å^2 for the two monomers, a surface about half the size of the dimer interface. This large contact possibly stabilizes an otherwise very flexible area, and may explain the good diffraction qualities of the $P2_12_12$ crystals (in sharp contrast to the five other HPP precursor crystals forms found diffracting to 3.5 Å or poorer resolution). Whether this crystal contact reflects any physiological significance remains to be established.

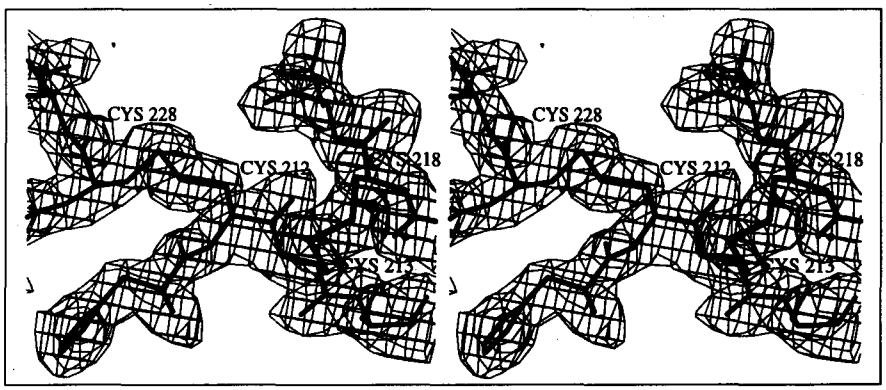

Fig. 2. Stereoview of the disulfide bridges Cys212–Cys228 and Cys213–Cys218 as revealed in the SIGMAA-weighted $2mF_o$–DF_c electron-density map [50]. The map was calculated from the model refined to 2.2 Å and contoured at 1σ. (Figure drawn with O [52].)

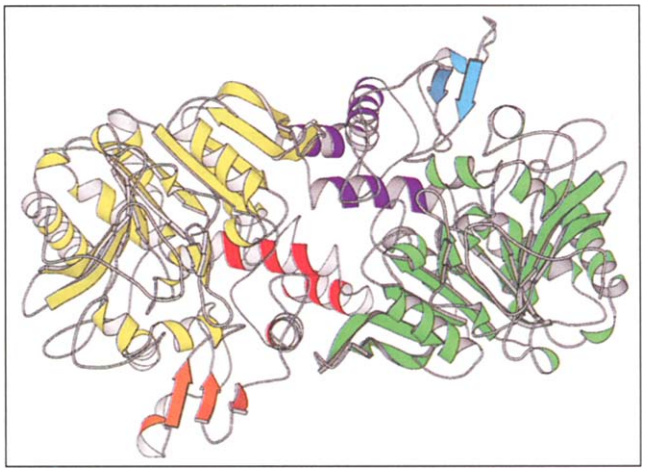

Fig. 3. Schematic ribbon diagram of the HPP dimer viewed approximately along the twofold axis. For monomer 1, the core domain is yellow while the cap domain consists of a helical subdomain in red and a maturation subdomain in orange. For monomer 2, the core domain is green, while the cap domain consists of a dark blue helical subdomain and a light blue maturation subdomain. (Figure generated using MOLSCRIPT [59].)

HPP and the hydrolase family

HPP belongs to the hydrolase fold family which includes enzymes with different catalytic functions such as the serine carboxypeptidases [16,17], dehalogenase, various lipases and acetylcholinesterase [18]. Although the central core is the same (a central β sheet flanked by α helices on both sides), the cap domains in this protein family are quite diverse, both with respect to their folds as well as their sizes (Fig. 4). HPP has one of the largest cap domains, with 121 residues forming the three-helix bundle of the helical subdomain and the three-stranded β sheet of the maturation subdomain.

Major differences, and comparison with the serine carboxypeptidases

The overall fold of the monomers from HPP, CPW and CPY is similar [16,17]. The complete core domains of HPP and CPW superimpose with an rmsd of 1.7 Å for 302 Cα atoms and have 38% sequence identity. Deleting major deviating loops to yield only the central part of these domains allows HPP to superimpose on CPW with an rmsd of 1.2 Å (293 equivalent Cαs with 40% sequence identity for CPW). The 271 equivalent Cαs in CPY (with 42.2% identity) superimpose on HPP with an rmsd of 1.2 Å.

The cap domain in HPP differs significantly from the CPW and CPY counterparts. The HPP precursor structure reveals a large maturation subdomain not present in the structures of CPW and CPY for which the structures of the enzymatically active forms are known, but not the precursors. All three enzymes do contain a three-helix bundle in the cap domain (see Fig. 4), albeit with low sequence identity in this region (~12%). Superimposing the helical bundle in the HPP and CPW cap domains reveals that helix Hα1 in HPP maintains the same general orientation with respect to the core domain (differing by only 7.4°), but helices Hα2 and Hα3 have undergone major rotations (by κ=28.5° and κ=93.4° respectively). Given the integral role of the cap domain in forming the dimer interface it is of interest to compare the dimers of HPP and CPW. (The structure of CPY was not considered because it is found as a monomer in the crystal structure [17].)

The orientation of the monomers in forming the dimer, differs between HPP and CPW. Superposition of the core domain of one monomer from each dimer shows that the second pair of monomers forming the respective

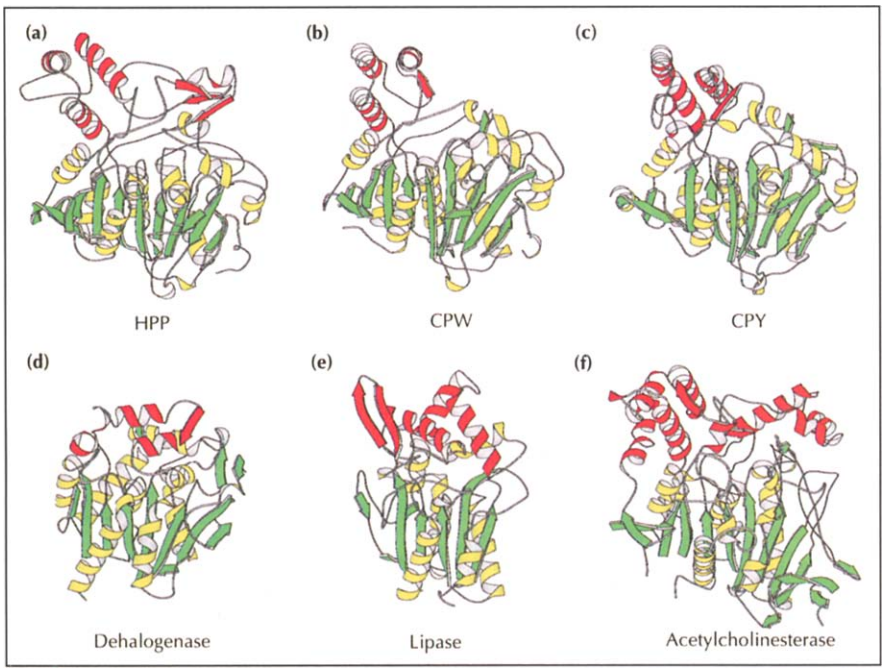

Fig. 4. Topological comparison of six members of the hydrolase fold family. The arrangement of structural elements in the central core domain (in green and yellow) of the different proteins is generally similar. The cap domains (in red) vary greatly. The secondary structure assignments were generated with O, using structures taken from the Brookhaven Protein Data Bank (PDB). The following structures are shown (references and PDB entry codes are given in parentheses): **(a)** the HPP precursor (this work), cap domain consists of two subdomains one α-helical and the other mainly β sheet; **(b)** CPW (3SC2; [16]), cap domain helical; **(c)** CPY (1YSC; [17]), cap domain helical; **(d)** dehalogenase (2HAD; [45]), cap domain helical but quite different from the serine carboxypeptidases; **(e)** lipase from *Pseudomonas glumae* (1TAH; [61]), cap domain mixed helical and β strands; **(f)** acetylcholinesterase (1ACE; [44]), cap domain large and predominantly helical. (Figure generated using MOLSCRIPT [59].)

dimers differ by a remarkable 15° in orientation (Fig. 5). Thus, it appears that the extensive differences in the cap domains, in particular the difference in orientation of the helical bundles with respect to the core domain, lead to a different arrangement of the subunits in the dimers of HPP and CPW.

Catalytic triad and enzymatic mechanism

Our structure shows that the precursor HPP has all the elements proposed for the enzymatic machinery of the serine carboxypeptidases [16,17], and is now the third member of this family to have its structure elucidated. The catalytic triad in the active site of HPP is formed by residues Ser150, His429 and Asp372 in agreement with the proposal by Galjart *et al.* [12]. The Oγ of Ser150 forms a good hydrogen bond with the Nε1 of His429 with an N...O distance of 2.8 Å. The Nδ1 of His429 is 2.7 Å removed from the Oδ2 and 3.3 Å from the Oδ1 of Asp372. The carboxylate of Asp372 and the imidazole of His429 in HPP are non-planar, making an angle of approximately 60° between the imidazole and the carboxylate. Two backbone amides appear to orient the carboxylate group of Asp372: the N of Ala374 is 3.0 Å from Oδ1 of Asp372 and the N of Cys375 is 2.9 Å from Oδ2 of Asp372. The oxyanion hole proposed to stabilize the negatively charged tetrahedral intermediate in serine carboxypeptidases [16] is formed in HPP by the backbone amides of Gly57 and Tyr151. The 32 atoms of the catalytic triad residues plus the oxyanion hole amides from HPP, CPY and CPW superimpose with an rmsd of 0.4 Å (see Table 1).

In HPP, a pair of glutamic acid residues (Glu69 and Glu149) is positioned near the catalytic triad, with their carboxylate groups interacting with each other. The carboxylate groups are located at approximately 8 Å from the Oγ of Ser150, and lie at the bottom of the active site. The two carboxylates interact with each other via hydrogen bonds. In addition, an asparagine (Asn55) is oriented such that it forms a hydrogen bond to each of the two carboxylate groups of the glutamic acid pair, at an Nδ2 (Asn) to Oε1/Oε2 (Glu) distance of 3.0 Å and 3.6 Å respectively. These three residues are conserved between HPP, CPW and CPY and have been implicated in regulating the low pH optimum for the carboxypeptidase activity found in the serine carboxypeptidases [16].

Biochemical data have suggested that a functional group with an apparent pK_a value of pH 5.5 binds the C-terminal carboxylate group of peptide substrates and is responsible for the observed pH optimum (reviewed in [19,20]). Liao *et al.* [16] have suggested that at pH 5.5 or below, one or both glutamic acid residues must be uncharged, while at higher pH deprotonation of one or both of the carboxylates may result in unfavorable electrostatic interactions. This would disturb the hydrogen-bonding pattern or result in structural perturbations causing the observed increase in K_m for peptide substrates. The

Table 1. Superposition of the proposed catalytic machinery of the serine carboxypeptidases with known three-dimensional structures.

	HPP		CPW	ΔHPP–CPW	CPY	ΔHPP–CPY
Catalytic triad	Ser150	N	Ser146	0.3	Ser146	0.4
		Cα		0.4		0.5
		C		0.3		0.4
		O		0.3		0.4
		Cβ		0.9		1.1
		Oγ		1.5		0.9
	His429	N	His397	0.2	His397	0.4
		Cα		0.3		0.4
		C		0.3		0.5
		O		0.5		0.6
		Cβ		0.3		0.6
		Cγ		0.3		0.5
		Cδ2		0.7		0.5
		Cε1		0.4		0.5
		Nδ1		0.3		0.4
		Nε2		0.7		0.5
	Asp372	N	Asp338	0.2	Asp338	0.2
		Cα		0.1		0.1
		C		0.1		0.1
		O		0.2		0.1
		Cβ		0.3		0.2
		Cγ		0.2		0.1
		Oδ1		0.2		0.3
		Oδ2		0.4		0.1
Oxyanion hole	Gly57	N	Gly53	0.1	Gly53	0.3
		Cα		0.2		0.4
		C		0.1		0.4
		O		0.3		0.8
	Tyr151	N	Tyr147	0.3	Tyr147	0.2
		Cα		0.2		0.1
		C		0.3		0.2
		O		0.5		0.2
pH regulators*	Asn55		Asn51	0.2	Asn51	0.2
	Glu69		Glu65	0.3	Glu65	0.7
	Glu149		Glu145	0.4	Glu145	0.4

The residues forming the proposed catalytic machinery are strictly conserved between the three serine carboxypeptidases. The deviation in distance between the atoms from HPP and the equivalent atoms in CPW or CPY after superposition is given in Å.
*Averaged over all atoms.

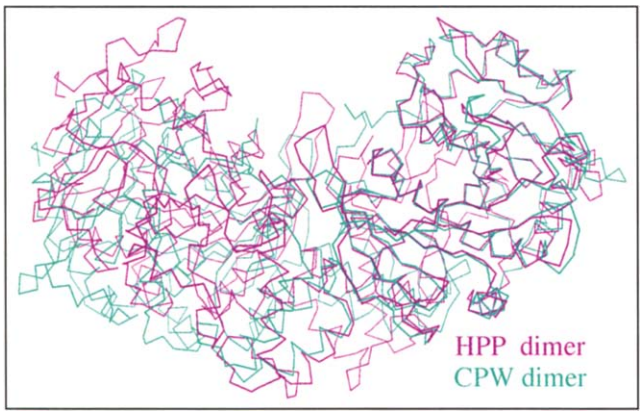

Fig. 5. Superposition of the Cα traces of HPP (mauve) and CPW (cyan) dimers based on alignment of the core domains from the subunits shown on the right-hand side. Notice the remarkable difference in mutual orientation (of 15°) of the two subunits on the left-hand side of the two dimers. (Figure drawn with O [52].)

CPW and CPY structures have been determined at pH 5.7 and at pH 6.5–7.0 respectively, while the structure of HPP has been determined at pH 8. As the arrangement of the Glu-Glu-Asn cluster in all three structures is essentially identical (see Table 1), our structure seems to rule out any significant conformational changes of these residues at up to 2.5 pH units above that optimal for carboxypeptidase activity. The high degree of conformational conservation, however, does indicate a role in some characteristic shared by all three enzymes.

From our comparison it is clear that the enzymatic machinery in the inactive precursor HPP has a very high degree of structural similarity to that found in the structures of the fully active CPW and CPY enzymes.

Active site and substrate specificity

HPP has a substrate preference for hydrophobic residues in the P1 and/or P1′ binding pockets [29]. In CPW, the P1′ pocket was identified as consisting of two tyrosine residues (Tyr60 and Tyr239) which form a long channel, capped by two acidic residues (Glu272 and Glu398) [16], explaining the high preference of this enzyme for arginine and lysine as the leaving group [38]. In HPP, the analogous residues are Tyr247 and Asp64, forming the sides of the P1′ pocket with Met430 and Thr304 at the far end. However, a more detailed description of the P1, P2 and P1′ binding pockets and a better understanding of their observed biochemical preferences requires structural information on the mature form of HPP.

Inactivation mechanism of the precursor form

The active-site cleft in the precursor is blocked by numerous residues from the maturation subdomain. The catalytic triad is rendered solvent inaccessible by residues Asn275, Ile276 and Phe277. These residues are part of the polypeptide Asp272–Phe277 which we call the 'blocking' peptide. This peptide is held down predominantly by hydrophobic contacts of Leu273, Ile276 and Phe277 to the core domain residues Gly57, Cys60, Leu180, Leu190, Val191, Leu232, Val235, Ala236, Ile246, Leu280, Leu282, Met299 and Ala374 (Fig 6). Residue Asn275 of the blocking peptide appears to fill what might be part of the P1 binding pocket in the mature form. The blocking peptide does not assume a conformation that a productive peptide substrate would adopt. It is carefully positioned to avoid being cleaved by the nearby catalytic residues. In particular, Gly274 (with Ramachandran angles ϕ=66° and ψ=28°) appears critical in allowing the polypeptide chain to adopt a conformation in which its main chain is at a safe distance from the catalytic triad. The P1′ binding pocket seems to be completely filled by Pro301 interacting with Thr304, Tyr247, Cys60 and Cys334. Thus, substrate binding is prevented in the precursor form by the inaccessibility of the substrate-binding pockets.

In our crystal structure, the excision peptide (residues 285–298) in both monomers extends out from the protein surface rendering it accessible to solvent and proteases. Of the three cleavage sites thought to be utilized upon trypsin digestion, Arg284 and Arg292 are particularly well exposed. The main-chain atoms of Arg298 are less accessible, being sandwiched between the strand Mβ2 and a loop N-terminal to helix Cα6, while the side-chain atoms of Arg298 are only partially solvent accessible due to a salt bridge with Glu264.

Proposed maturation event and extent of conformational rearrangement

Our structure reveals that the inactivation mechanism of HPP is based on blocking of the active site, not upon conformational changes of the residues involved in catalysis or transition-state stabilization. The excision peptide itself does not bind in the active-site cleft. Hence removing it is not in itself sufficient to allow solvent or substrate access to the active site and maturation must be accompanied by conformational changes.

Based on the HPP precursor structure and the comparison with CPW and CPY, it is proposed that upon maturation a region comprising approximately residues 254–284 rearranges to free the P1 and P2 binding sites, while the residues 299–302 rearrange to free the P1′ binding pocket (see Fig. 7). The linker connecting these two segments of polypeptide chain is the 14 amino acid

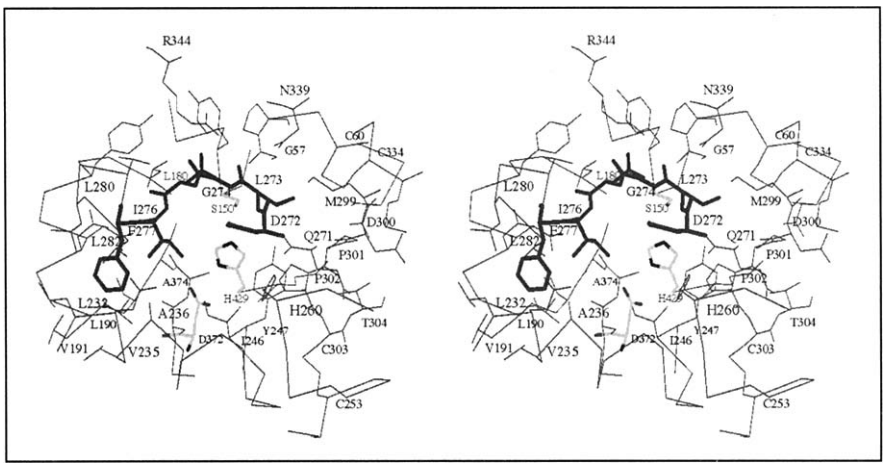

Fig. 6. Close-up of the 'blocking' peptide (residues 272–277), shown in thick lines bound in the active site, rendering the catalytic triad solvent inaccessible. Side chains are shown for residues making extensive contacts with the blocking peptide. The catalytic triad is shown in gray. (Figure drawn with O [52].)

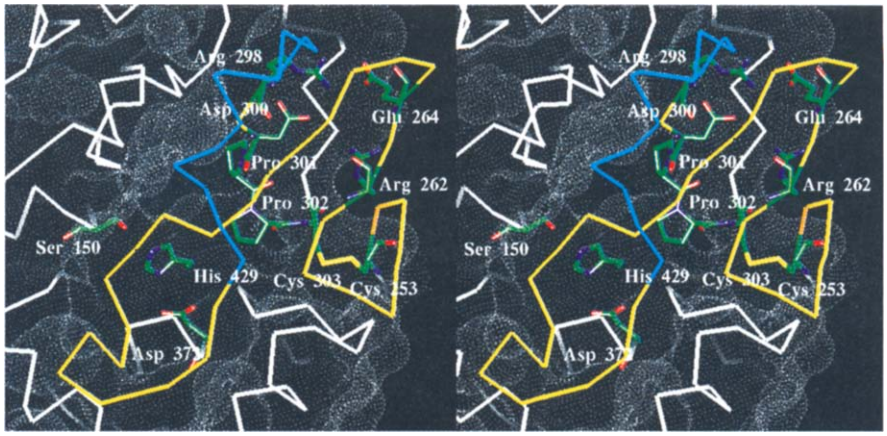

Fig. 7. Stereoview of the elements proposed to be involved in the activation mechanism of the precursor form of HPP. The Cα trace of the maturation subdomain is shown in yellow, except for the excision peptide shown in light blue. Relevant side chains are depicted and labeled. The solvent-accessible surface was calculated and visualized together with the atomic coordinates by Insight II (Biosym Technologies Inc., San Diego, CA).

excision peptide Met285–Arg298. The number of residues involved in these rearrangements is likely to be limited by the Cys253–Cys303 disulfide bridge which is conserved in the serine carboxypeptidase family. This critical disulfide serves to keep the secondary structure elements together at the far end of the P1′ pocket.

It is not clear what creates the driving force behind the conformational changes required to open up the active site. No biochemical studies have been performed to investigate the extent to which a pH shift and/or the cleavage event(s) trigger conformational rearrangement during the maturation event. The structure of the HPP precursor does yield some hints. A rectangular array of salt bridges is observed between Arg262, Asp300, Glu264 and Arg298, four residues located on strands Mβ1 and Mβ3 of the mixed β sheet found in the maturation subdomain. This cluster of residues is strategically positioned at the base of the excision peptide, close the core domain and 'shields' the mixed β sheet via side-chain interactions (see Fig. 7). These residues are strictly conserved among the human, mouse and chicken protective proteins [27]. A similar double salt bridge has been observed in pepsinogen, an aspartic protease zymogen, between the proenzyme segment (Arg8P) and the enzyme (Arg308, Glu13, Asp304) and has been proposed to function as a pH-dependent trigger for the removal of the proenzyme segment [39]. In the HPP precursor a similar scenario might be envisaged, although the lack of biochemical data makes this purely speculative. Upon arrival in the endosome/lysosome, the charge cluster would be affected by a shift from neutral to acidic pH. Protonation of either the asparagine or the glutamic acid residue or both, would result in unfavorable electrostatic interactions and destabilization of this charge cluster. This might in turn promote partial unfolding of the maturation subdomain, allowing easier access to additional potential cleavage sites, or it might stimulate removal of the blocking peptide which fills the active site in the precursor.

The maturation mechanism for HPP appears to be novel among those proteases for which the three-dimensional structure of the zymogen is known. To our knowledge three different mechanisms have been elucidated so far, for the conversion of a protease zymogen to a mature enzyme. First, in the serine protease family, activation of chymotrypsinogen and trypsinogen has been shown to entail multiple cleavage events of the polypeptide chain allowing small rearrangements in the enzymatic machinery to occur (involving the oxyanion hole, the S1 primary binding site and formation of an internal salt bridge in the active site, as reviewed in [40]). Thus, activation involves structural changes to reposition selected elements of the enzymatic machinery into a catalytically competent conformation.

In the second mechanism, found in the zinc-containing metalloprotease family, it has been shown for procarboxypeptidases A and B that an extensive N-terminal domain blocks the active site [41]. Removal of this domain by cleavage of a single peptide bond renders the active site accessible, and no further rearrangements are needed to obtain activity. Thus, the catalytic machinery in the zymogen is already in a conformation suitable for catalysis and inactivation results from physical obstruction by residues subsequently removed during the activation process.

A third activation mechanism has been revealed in the aspartic proteases by comparing the structures of pepsinogen and pepsin [39]. An N-terminal domain, held down by ionic interactions, blocks the active site. Upon a shift to low pH, this domain is thought to (partially) dislocate from the active site, undergoing conformational changes that allow it to be autocatalytically removed [39]. Hence a conformational change precedes (autocatalytic) cleavage and the residues involved in blocking the active site are removed to generate the mature enzyme.

In HPP a fourth mechanism for protease zymogen activation is revealed (see Fig. 8). In this case, the catalytic triad in the precursor form is in a catalytically competent conformation. Enzymatic activity is prevented by the blocking peptide. The blocking peptide is, however, not the same as the excision peptide. This leads to a distinct difference from other maturation mechanisms. After disappearance of the excision peptide, up to 35 residues filling the active-site cleft in the HPP precursor must rearrange to render the catalytic triad solvent accessible, but these residues do not get cleaved off. Removal of the excision peptide, and possibly a shift to lower pH in the

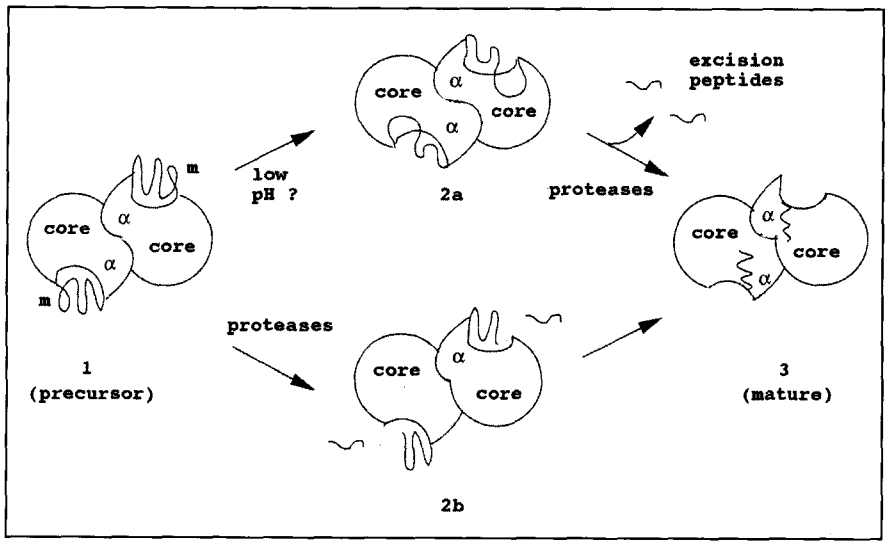

Fig. 8. Schematic representation of the proposed activation of HPP. The active-site cleft is formed by the core domain and the helical subdomain (indicated as α). The maturation subdomain (m) contains the residues which block the active-site cleft rendering the precursor enzymatically inactive (structure 1). In the acidic endosome/lysosome, the precursor undergoes activation. In activation pathway 2a, conformational rearrangements induced by low pH might render the excision peptide more accessible to proteases as a first step, followed by cleavage of the polypeptide chain removing the excision peptide. Alternatively, in pathway 2b, proteolytic cleavage of the excision peptide might trigger the total rearrangement, removing the blocking peptide from the active site and thus generating the fully active enzyme (structure 3).

endosome/lysosome, appears to be a trigger for this event. Surprisingly, the maturation mechanism of the serine carboxypeptidases HPP, CPW and CPY may differ from each other as well. This is clearest for CPY, in which a 91-residue polypeptide is cleaved off the N terminus to convert the zymogen to an active enzyme [42], as opposed to the excision of a peptide from within the zymogen generating a two-chain active form as is the case for HPP and CPW [14,19].

Considering the hydrolase fold family as a whole, the catalytic triad is housed in the core domain and the various cap domains attenuate the biological function by influencing entirely different properties such as: enzyme kinetics (exemplified by the interfacial activation of lipases [43]); substrate channeling (as proposed for acetylcholinesterase [44]); substrate recognition (as proposed by Franken *et al.* [45] for dehalogenase and by Endrizzi *et al.* [17] for CPY and CPW); and enzyme inactivation (as shown here for HPP). Nature seems to have employed a common theme for totally different purposes.

Biological implications

The lysosomal storage disease, galactosialidosis, has been linked to a deficiency in the human 'protective protein' (HPP), believed to be identical to cathepsin A. HPP is thought to form a multienzyme complex with β-galactosidase and neuraminidase in the lysosomes, protecting these two glycosidases from degradation, in their harsh acidic and proteases-rich environment. The structural basis of the mechanism by which this protection is achieved is unknown.

Enzymatically, HPP is classified as a serine carboxypeptidase. It shares 30% sequence identity with the wheat and yeast serine carboxypeptidases (CPW and CPY). In its precursor form, HPP is inactive, but upon maturation, entailing removal of a 2 kDa 'excision' peptide, carboxypeptidase activity is released [14]. We have solved the structure of the precursor form of HPP to 2.2 Å at pH 8, and found it to be a dimer. The precursor form of HPP is of interest as it is the only form of the enzyme which is secreted by cells and can be taken up by others via endocytosis. This latter property is of potential therapeutic significance for the treatment of galactosialidosis.

The HPP monomer can be divided into a 'core' domain housing the catalytic triad and a 'cap' domain consisting of a 'helical' subdomain and a 'maturation' subdomain. The core domains of HPP, CPW [16] and CPY [17] have very similar folds that classify them as members of the hydrolase fold family [18]. However, significant differences are observed in the arrangement of the structural elements in the HPP helical subdomain as compared with its counterpart in CPW and CPY. Although HPP and CPW both form dimers, comparing the configuration of the subunits that generate the dimer in each enzyme reveals a significant (15°) rotational difference, which would have been hard to predict by homology modelling.

The precursor structure enables us to propose an inactivation mechanism that has not been seen before in any of the other known zymogen structures of proteases. The catalytic triad of HPP seems to have an arrangement poised for catalysis. However, the triad and substrate-binding site are rendered inaccessible to solvent and substrate by residues from the maturation subdomain binding in the active-site cleft. Most of these residues are found in a strand called the 'blocking' peptide which does not overlap with the 2 kDa excision peptide. After removal of the excision peptide, up to 35 additional residues must rearrange in order to unblock the active-site cleft. It is not clear what the driving force is for the conformational changes. A strategically positioned pair of salt bridges,

comprising Arg262, Arg298, Glu264 and Asp300 at the base of the excision peptide, may become destabilized at low pH, unraveling this region of the structure, allowing easier access to cleavage sites and/or promoting the rearrangement event.

A number of research groups are currently involved in designing enzyme and gene therapy procedures for several lysosomal storage diseases. Insight into the three-dimensional structure, protein functioning and stability of HPP, the first enzyme associated with a lysosomal storage disease for which a structure has been determined, may prove useful in future designs of an adequate therapy procedure for galactosialidosis. Information from the three-dimensional structure of HPP, might also aid in designing an engineered form of HPP with increased stability and a longer half-life.

Material and methods

Purification and crystallization

The HPP precursor was over-expressed in a baculovirus system [14]. The precursor could be purified to better than 95% homogeneity using ion-exchange chromatography starting from a partially pure preparation containing both the precursor and the cleaved forms. Extensive screening yielded various crystal forms (Rudenko *et al.*, unpublished data). The crystals suitable for data collection and the structure determination were grown in polyethylene glycol 8000 at pH 8.0–8.3 in space group $P2_12_12$ with cell dimensions a=115.04 Å, b=148.11 Å, c=80.97 Å.

Data collection, data processing and reduction

Diffraction data were collected at the Stanford Synchrotron Radiation Laboratories to 2.0 Å at −178°C on a MAR imaging plate at a wavelength of 1.08 Å on beamline 7-1. The data were processed and reduced with the CCP4 suite [46] using the programs MOSFLM version 5.2, REFIX [47], ROTAVATA, AGROVATA and TRUNCATE. The data set contains 436709 reflections between 32.3 Å and 2.2 Å, reducing to 67740 unique reflections. The I/σ for the highest resolution shell of the data used in the structure determination (2.26–2.20 Å) is 5.7. The data is 95.7% complete to 2.2 Å with an R_{sym} of 5.1%. The V_m [48] is 3.2 $Å^3$ Da^{-1} for two monomers in the asymmetric unit, corresponding to a solvent content of 62%.

Molecular replacement

A multi-alanine search model for a monomer was constructed from the atomic coordinates for the wheat serine carboxypeptidase [16], aided by sequence alignment made by the GCG program PILEUP (Genetics Computer Group, Madison, Wisconsin). Residues identical between the two proteins were left in the model, residues with the same sized side chain were left unaltered and the rest of the residues were truncated to alanines. The search model contained 331 residues. Data between 4–8 Å were used for the rotation function, Patterson correlation refinement and translation function calculations carried out by X-PLOR [49]. A Patterson cutoff vector of 21 Å was used. Two orientations were found for the monomer and the individual translation functions gave a clear position for each orientation with peaks 7.7σ and 8.8σ above the mean. A combined translation function placed the two monomers relative to the same crystallographic origin, and yielded a packing in which the two monomers formed a physiological dimer as judged by their extensive interactions. This initial structure solution gave an R factor of 51.5% for data between 6 Å and 4.5 Å, which decreased to 49.2% after rigid-body refinement.

Density modification and map interpretation

Electron-density maps were calculated using the programs from the CCP4 computing package [46]. In the initial SIGMAA-weighted $2mF_o–DF_c$ map [50] reasonable density appeared to be present for 294 residues. Therefore, 158 residues, or 35%, were still missing per monomer. By using cycles of twofold averaging with the program RAVE [51] and gradual model expansion, 148 residues per monomer could be added. For map visualization and model building the program O [52] was used on a Silicon Graphics workstation.

Refinement and model validation and analysis

Rigid-body and positional refinement were carried out using the program X-PLOR [49] with the stereochemical parameters of Engh and Huber [53]. At this stage the model was still missing 20 residues (10 per subunit). Eighteen of these residues could gradually be incorporated in the model during additional cycles of positional refinement and model building. The refinement procedure started out using data from 8–3.0 Å and was gradually extended to include data from 8–2.2 Å. B-values for individual atoms were refined once data between 2.4 Å and 2.2 Å had been added. The program ARP [54] was used to add crystallographic waters and as a tool to check for bad areas in the model. The programs PROCHECK [55], WHATIF [56] and PROFILE 3D [57], as well as the geometrical analysis generated by X-PLOR have been used to check the structure for errors. For the final refinement cycle, 10% of the weakest data were rejected using an $|F_{obs}|/\sigma$ cutoff and anisotropic scaling was applied between F_{obs} and F_{calc} to compensate for a rapid fall-off in intensity along the crystallographic a axis. The R factor for the current model is 21.3% for data between 8–2.2 Å; the R_{free} is 26.8% (see Table 2). The estimated rms coordinate error according to SIGMAA [50] is 0.282 Å. The rmsds from ideality are 0.012 Å for bond lengths and 1.72° for bond angles. The current model consists of 902 residues out of a total of 904 for the dimer, numbered 1–452 for monomer 1 and monomer 2. In addition, 296 waters and six *N*-acetylglucosamine moieties have been incorporated. The electron density for residues 259 and 260 from monomer 2 is not well

Table 2. Data statistics for refinement.

Resolution (Å)	R_{cryst} (%)	Completeness (%)	Cumulative completeness (%)
8.00–4.21	22.0	93.5	93.5
4.21–3.42	19.3	92.7	93.1
3.42–3.02	20.6	88.6	91.6
3.02–2.75	21.3	84.2	89.8
2.75–2.56	22.0	78.6	87.6
2.56–2.42	22.2	73.1	85.2
2.42–2.30	22.7	64.6	82.3
2.30–2.20	24.0	56.9	79.2
8.00–2.20	21.3	82.9	

$R_{cryst}=\Sigma||F_{obs}(h)|-|F_{calc}(h)||/\Sigma|F_{obs}(H)|$. Figures for the completeness of data included in the refinement are given after rejecting 10% of the weakest reflections using an $|F_{obs}|/\sigma$ cutoff for data between 8 Å and 2.2 Å and partitioning 5% of the data in a test set for monitoring the R_{free}.

defined, and these residues have not been included in the model. In total, 14 side chains are not visible and an alanine side chain has been incorporated into the model. For monomer 1, these residues are: Lys15, Lys27, Glu46, Phe113, Glu326 and Val393. For monomer 2, these residues are: Glu46, Gln215, Lys217, Phe261, Tyr263, Lys265, Lys296 and Glu326. The average individual B-factor for main-chain atoms is 16.6 $Å^2$ and for side-chain atoms 18.3 $Å^2$. The program GRASP [35] was used for the molecular surface analyses and the program SUPPOS (Dijkstra *et al.*, unpublished program) was used for all the least-square superpositions of various models and parts of the models.

Coordinates have been forwarded to the Brookhaven Protein Data Bank.

Acknowledgements: Dr H Galjaard (Clinical Genetics Foundation, Rotterdam) is most gratefully acknowledged for support in the initial stages of the project. Dr J Remington kindly provided the atomic coordinates of the wheat and the yeast serine carboxypeptidases prior to publication; Drs G Kleywegt and A Jones made the RAVE averaging software and documentation available prior to publication. Drs R Read and B Hazes are thanked for the encouragement and helpful advice during the early stages of the structure determination. The Stanford Synchrotron Radiation Laboratories provided excellent facilities. GR is greatly indebted to the lecturers of the FEBS/EACMB Workshop, Århus 1992, the ESF Molecular Replacement Workshop in Paris (1992) (especially Dr I Tickle), and the EMBO course in Heidelberg (1993). The discussions with Dr B Dijkstra and the Groningen crystallography group are kindly acknowledged, as well as Dr C Verlinde for creating the computer graphics cluster in Seattle. These studies were supported in part by the NIH Cancer Center Support CORE Grant P30-CA21765 and the American Lebanese Syrian Associated Charities (ALSAC). WGJH acknowledges receipt of a major equipment grant from the Murdock Charitable Trust.

References

1. d'Azzo, A., Hoogeveen, A., Reuser, A.J.J., Robinson, D. & Galjaard, H. (1982). Molecular defect in combined β-galactosidase and neuraminidase deficiency in man. *Proc. Natl. Acad. Sci. USA* **79**, 4535–4539.
2. Wenger, D.A., Tarby, T.J. & Wharton, C. (1978). Macular cherry-red spots and myoclonus with dementia: co-existent neuraminidase and β-galactosidase deficiencies. *Biochem. Biophys. Res. Commun.* **82**, 589–595.
3. Andria G., Strisciuglio, P., Pontarelli, G., Sly, W.S. & Dodson, W.E (1981). Infantile neuraminidase and β-galactosidase deficiencies (galactosialidosis) with mild clinical courses. In *Sialidases and Sialidosis. Perspectives in Inherited Metabolic Diseases.* (Tettamanti, G., Durand, P. & Di Donato, S., eds), vol. **4**, pp. 379–395, Edi. Ermes, Milan.
4. Lowden, J.A., Cutz, E. & Skomorowski, M.A. (1981). Infantile type 2 sialidosis with β-galactosidase deficiency. In *Sialidases and Sialidosis. Perspectives in Inherited Metabolic Diseases.* (Tettamanti, G., Durand, P. & Di Donato, S., eds), vol. **4**, pp. 261–279, Edi. Ermes, Milan.
5. d'Azzo, A., Andria, G., Strisciuglio, P. & Galjaard, H. (1995). Galactosialidosis. In *The Metabolic and Molecular Bases of Inherited Disease.* 7th Edition. (Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D., eds), vol. **II**, pp. 2825–2837, McGraw Hill, Inc., New York.
6. Okamura-Oho, Y., Zhang, S. & Callahan, J.W. (1994). The biochemistry and clinical features of galactosialidosis. *Biochim. Biophys. Acta* **1225**, 244–254.
7. Thomas, G.H. & Beaudet, A.L. (1995). Disorders of glycoprotein degradation and structure: α-mannosidosis, β-mannosidosis, fucosidosis, sialidosis, aspartylglucosaminuria, and carbohydrate-deficient glycoprotein syndrome. In *The Metabolic and Molecular Bases of Inherited Disease.* 7th Edition. (Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D., eds), vol. **II**, pp. 2529–2561, McGraw Hill, Inc., New York.
8. Suzuki, Y., Sakuraba, H. & Oshima, A. (1995). β-Galactosidase deficiency (β-galactosidosis): G_{M1}-gangliosidosis and Morquio B disease. In *The Metabolic and Molecular Bases of Inherited Disease.* 7th Edition. (Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D, eds), Vol **II**, pp. 2785–2823, McGraw Hill, Inc., New York.
9. Hoogeveen, A.T., Verheijen, F.W. & Galjaard, H. (1983). The relation between human lysosomal β-galactosidase and its protective protein. *J. Biol. Chem.* **258**, 12143–12146.
10. van Diggelen, O.P., Schram, A.W., Sinnott, M.L., Smith, P.J., Robinson, D. & Galjaard, H. (1981). Turnover of β-galactosidase in fibroblasts from patients with genetically different types of β-galactosidase deficiency. *Biochem. J.* **200**, 143–151.
11. van Diggelen, O.P., Hoogeveen, A.T., Smith, P.J., Reuser, A.J.J. & Galjaard, H. (1982). Enhanced proteolytic degradation of normal β-galactosidase in the lysosomal storage disease with combined β-galactosidase and neuraminidase deficiency. *Biochim. Biophys. Acta* **703**, 69–76.
12. Galjart, N.J., *et al.*, & d'Azzo, A. (1988). Expression of cDNA encoding the human 'protective protein' associated with lysosomal β-galactosidase and neuraminidase: homology to yeast proteases. *Cell* **54**, 755–764.
13. Zhou, X.-Y., Galjart, N.J., Willemsen, R., Gillemans, N., Galjaard, H. & d'Azzo, A. (1991). A mutation in a mild form of galactosialidosis impairs dimerization of the protective protein and renders it unstable. *EMBO J.* **10**, 4041–4048.
14. Bonten, E.J., Galjart, N.J., Willemsen, R., Usmany, M., Vlak, J.M. & d'Azzo, A. (1995). Lysosomal protective protein/cathepsin A: role of the 'linker' domain in catalytic activation. *J. Biol. Chem.* **270**, in press.
15. Jackman, H.L., *et al.*, & Erdös, E.G. (1990). A peptidase in human platelets that deamidates tachykinins. Probable identity with the lysosomal 'protective protein'. *J. Biol. Chem.* **265**, 11265–11272.
16. Liao, D.I., Breddam, K., Sweet, B., Bullock, T. & Remington, S.J. (1992). Refined atomic model of wheat serine carboxypeptidase II at 2.2 Å resolution. *Biochemistry* **31**, 9796–9812.
17. Endrizzi, J.A., Breddam, K. & Remington, S.J. (1994). 2.8 Å-structure of yeast serine carboxypeptidase. *Biochemistry* **33**, 11106–11120.
18. Ollis, D.L., *et al.*, & Goldman, A. (1992). The α/β hydrolase fold. *Protein Eng.* **5**, 197–211.
19. Breddam, K. (1986). Serine carboxypeptidases. A review. *Carlsberg Res. Commun.* **51**, 83–128.
20. Remington, S.J. & Breddam, K. (1994). Carboxypeptidases C and D. *Methods Enzymol.* **244**, 231–248.
21. Verheijen, F., Brossmer, R. & Galjaard, H. (1982). Purification of acid β-galactosidase and acid neuraminidase from bovine testis: evidence for an enzyme complex. *Biochem. Biophys. Res. Commun.* **108**, 868–875.
22. Yamamoto, Y., Fujie, M. & Nishimura, K. (1982). The interrelation between high- and low-molecular-weight forms of G_{M1}-β-galactosidase purified from porcine spleen. *J. Biochem.* **92**, 13–21.
23. Yamamoto, Y. & Nishimura, K. (1987). Copurification and separation of β-galactosidase and sialidase from porcine testis. *Int. J. Biochem.* **19**, 435–442.
24. Potier, M., Michaud, L., Tranchemontagne, J. & Thauvette, L. (1990). Structure of the lysosomal neuraminidase–β-galactosidase–carboxypeptidase multienzymic complex. *Biochem. J.* **267**, 197–202.
25. Scheibe, R., Hein, K. & Wenzel, K.-W. (1990). Lysosomal β-galactosidase from rat liver: purification, molecular forms and association with neuraminidase. *Biomed. Biochim. Acta* **49**, 547–556.
26. Hubbes, M., d'Agrosa, R.M. & Callahan, J.W. (1992). Human placental β-galactosidase. Characterization of the dimer and complex forms of the enzyme. *Biochem. J.* **285**, 827–831.
27. Galjart, N.J., Morreau, H., Willemsen, R., Gillemans, N., Bonten, E.J. & d'Azzo, A. (1991). Human lysosomal protective protein has cathepsin A-like activity distinct from its protective function. *J. Biol. Chem.* **266**, 14754–14762.
28. Jackman, H.L., Morris, P.W., Deddish, P.A., Skidgel, R.A. & Erdös, E.G. (1992). Inactivation of endothelin I by deamidase (lysosomal protective protein). *J. Biol. Chem.* **267**, 2872–2875.
29. Jackman, H.L., Morris, P.W., Rabito, S.F., Johansson, G.B., Skidgel, R.A. & Erdös, E. (1993). Inactivation of endothelin-1 by an enzyme of the vascular endothelial cells. *Hypertension* **21**, 925–928.
30. Itoh, K., Kase, R., Shimmoto, M., Satake, A., Sakuraba, H. & Suzuki, Y. (1995). Protective protein as an endogenous endothelin degradation enzyme in human tissues. *J. Biol. Chem.* **270**, 515–518.
31. Hanna, W.L., Turbov, J.M., Jackman, H.L., Tan, F. & Froelich, C.J. (1994). Dominant chymotrypsin-like esterase activity in human lymphocyte granules is mediated by the serine carboxypeptidase called cathepsin A-like protective protein. *J. Immunol.* **153**, 4663–4672.
32. Musil, D., *et al.*, & Bode, W. (1991). The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J.* **10**, 2321–2330.
33. Metcalf, P. & Fusek, M. (1993). Two crystal structures for cathepsin D: the lysosomal targeting signal and active site. *EMBO J.* **12**, 1293–1302.

34. Baldwin, E.T., *et al.*, & Erickson, J.W. (1993). Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc. Natl. Acad. Sci. USA* **90**, 6796–6800.
35. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296.
36. Pshezhetsky, A.V., Levashov, A.V. & Wiederschain, G.Y. (1992). Regulation of the G_{M1}-galactosidase supramolecular structure and catalytic activity *in vitro*. *Biochim. Biophys. Acta* **1122**, 154–160.
37. Pshezhetsky, A.V., Elsliger, M.-A., Vinogradova, M.V. & Potier, M. (1995). Human lysosomal β-galactosidase–cathepsin A complex: definition of the β-galactosidase-binding interface on cathepsin A. *Biochemistry* **34**, 2431–2440.
38. Breddam, K., Sørensen, S.B. & Svendsen, I. (1987). Primary structure and enzymatic properties of carboxypeptidase II from wheat bran. *Carlsberg Res. Commun.* **52**, 297–311.
39. James, M.N.G. & Sielecki, A.R. (1986). Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* **319**, 33–38.
40. Creighton, T.E. (1984). Zymogen activation. In *Protein Structures and Molecular Principles*. pp. 439–443, W.H. Freeman & Co., New York.
41. Guasch, A., Coll, M., Avilés, F.X. & Huber, R. (1992). Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J. Mol. Biol.* **224**, 141–157.
42. Winther, J.R. & Sørensen, P. (1991). Propeptide of carboxypeptidase Y provides a chaperone-like function as well as inhibition of the enzymatic activity. *Proc. Natl. Acad. Sci. USA* **88**, 9330–9334.
43. Smith, L.C., Faustinella, F. & Chan, L. (1992). Lipases: three-dimensional structure and mechanism of action. *Curr. Opin. Struct. Biol.* **2**, 490–496.
44. Sussman, J.L., *et al.*, & Silman, I. (1991). Atomic structure of acetylcholinesterase from *Torpedo californica*: a prototypic acetylcholine-binding protein. *Science* **253**, 872–879.
45. Franken, S.M., Rozeboom, H.J., Kalk, K.H. & Dijkstra, B.W. (1991). Crystal structure of haloalkane dehalogenase: an enzyme to detoxify halogenated alkanes. *EMBO J.* **10**, 1297–1302.
46. Collaborative Computing Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
47. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **26**, 795–800.
48. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
49. Brünger, A.T. (1992). *X-PLOR: Version 3.1. A System for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT.
50. Read, R.J. (1986). Improved Fourier coefficient for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
51. Kleywegt, G. & Jones, T.A. (1994). Halloween... masks and bones. In *From First Map to Final Model*. (Bailey, S. Hubbard, R. & Waller, D., eds), pp. 59–66, SERC Daresbury Laboratory, Warrington, UK.
52. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
53. Engh, R. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392–400.
54. Lamzin, V.S. & Wilson, K.S. (1993). Automated refinement of protein models. *Acta Cryst. D* **49**, 129–147.
55. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291.
56. Vriend, G. (1990). WHAT IF: a molecular modelling and drug design program. *J. Mol. Graphics* **8**, 52–56.
57. Lüthy, R., Bowie, J.U. & Eisenberg, D. (1992). Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83–85.
58. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
59. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
60. Noble, M.E.M., Cleasby, A., Johnson, L.N., Egmond, M.R. & Frenken, L.G.J. (1993). The crystal structure of triacylglycerol lipase from *Pseudomonas glumae* reveals a partially redundant catalytic aspartate. *FEBS Lett.* **331**, 123–128.

Received: **14 Jul 1995**; revisions requested: **8 Aug 1995**;
revisions received: **12 Sep 1995**. Accepted: **15 Sep 1995**.