Materials and Methods

Production and Purification of Recombinant Human Granzyme B

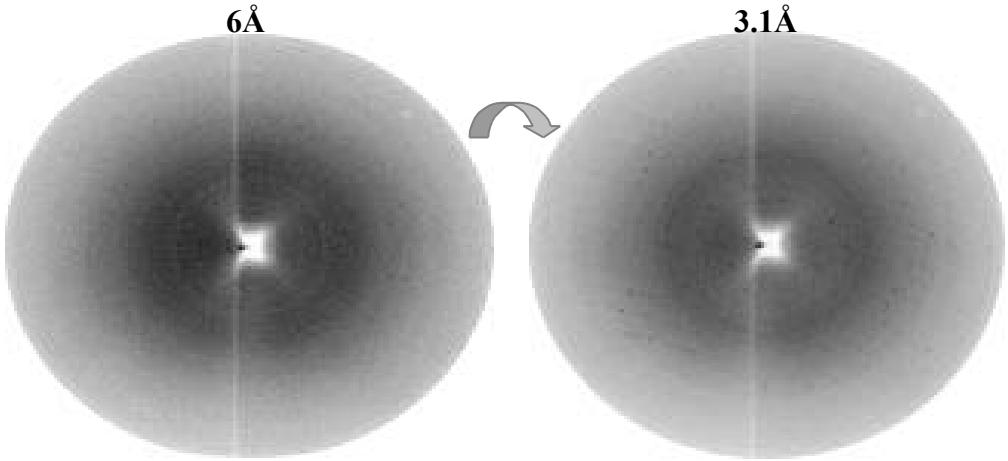
Mature human granzyme B was amplified from a human leukocyte cDNA library by PCR using the primers 5'-AGGATCCCGTGGATGATGACAAGATCATCGGGGGACAT-GAG-3' and 5'-GAATTCTTAGTAGCGTTTCATGGT-3'. These primers were designed to delete the signal sequence and insert an enterokinase cleavage site at the N-terminus. A *BamH1* site at the N-terminus and an *EcoRI* site at the C-terminus were introduced to facilitate subsequent cloning steps. The Ubiquitin-EkCS-GrB construct was cloned in a pAcGP67-B baculovirus transfer vector (Pharmingen, San Diego, CA). SF9 cells (Invitrogen, Carlsbad, CA) cotransfected with this vector and Baculogold DNA (Pharmingen) were used to produce the first, second and third generation of virus. High Five cells were then used to express the protein as a secreted product (Wang et al., 1998).

Medium was typically collected after 72 hours of incubation at 27°C and directly loaded on a SP-Sepharose column (Pharmacia, Sweden) equilibrated in 20 mM Hepes, 20 mM NaCl pH 7.0. The protein was eluted with a linear NaCl gradient (0.02 to 1 M) and protein fractions were identified by UV absorption and SDS-PAGE. The fractions containing rGzmB were pooled and the active enzyme was produced by incubation with enterokinase (125 Sigma Units/50 ml) for 24-36 h at 25 °C. rGzmB was then applied to a MonoS column (Pharmacia, Sweden) and purified with a NaCl gradient (eluted at about 500 mM NaCl). The enzyme was concentrated to 3-5 mg/ml, flash-frozen in liquid nitrogen and stored at –70°C. The concentration of rGzmB and of catalytically active protein was determined as previously described (Harris et al., 1998). Further rGzmB concentrations were determined assuming a specific activity of 1.8 μM of product per nMol of GzmB and min, measured under standard conditions (0.1 M Hepes, 0.15 M NaCl, 0.01 % dodecyl maltoside, 1 mM Ac-Ile-Glu-Thr-Asp-pNA and 2% dimethylformamide, pH 7.5). Prior to crystallization, the protein buffer was changed to 10 mM Hepes pH 7.4, 50 mM NaCl, using a Nap-10 column (Pharmacia, Sweden) and the sample was further concentrated to 7.7 mg/ml using a Centricon of 10 kDa cut-off (Amicon,USA).

Crystallization

Crystals suitable for diffraction analysis were obtained by sitting drop vapour-diffusion procedures. 1 µl of the protein solution was mixed with 2 µl of a solution consisting of 36% PEG-8000, 0.24 M ammonium sulphate and 4 mM Zwittergent® 3-12 (Hampton Research). Crystals grew within three months at 20 °C to maximal dimensions of 0.3 x 0.2 x 0.5 mm^3. These crystals belong to the space group $P2_12_12_1$ (cell constants a=41.74 Å, b=114.31 Å, c=135.52 Å**)** and contain two molecules per asymmetric unit. Perfluoropolyether oil PFO-X125/03 (Lancaster) was added to the crystallization drop before transferring the crystals to a capillary-free mounting system (Kiefersauer et al., 2000). Upon a change from 90% to 85% humidity at 16°C, the crystalline order of the crystals, diffracting initially to maximally 6 Å, dramatically improved showing reflections with Bragg spacings to beyond 3.0 Å. Most of the crystals tested proved to be twinned, however, and showed a strong decay during exposure rendering the collection of a complete data set from a single crystal impossible.

Diffraction pattern of GzmB crystals before and after humidity transformation:



No suitable cryo-conditions have been found so far, impairing all attempts to freeze the crystals. Thus, two incomplete data sets to 3.1 Å resolution obtained from two different crystals were collected at room temperature on a 300 mm MAR Research image plate detector using monochromatized CuKα radiation from a RIGAKU rotating anode X-ray generator. These two data sets were evaluated with MOSFLM (Leslie, 1991), merged, reduced and scaled (Table 1) using programs supported by the Collaborative Computational Project No. 4 (The CCP4 suite, 1994).

The structure was solved by molecular replacement with AMoRe (Navaza, 1994) using a modified CatG search model, in which all non-identical residues between the two proteinases were changed to alanine. Rigid body refinement in AMoRe resulted in an R-factor of 39.8% and a correlation coefficient of 52.3% for the range 15.0 to 3.5 Å. Model building was done on a SGI graphics workstation using MAIN (http://www-bmb.ijs.si/doc).

Calculation of the electron density maps and crystallographic refinement were performed with X-PLOR initially applying high non-crystallographic symmetry restraints for main and side-chain atoms, which were progressively released in the course of the refinement. Several cycles of model building, conjugate gradient minimization and simulated annealing using X-PLOR (Brünger, 1993) resulted in a structure with good stereochemistry. The target parameters of Engh and Huber (Engh and Huber, 1991) were used for the protein moiety. Individual, highly restrained atomic B-values were refined, and no sigma cut-off was applied to the whole procedure. During the improvement of the protein model, the Fourier maps also revealed good electron density for three oligossacharide units of the Asn65 *N*-linked carbohydrate. No attempts were made to identify and include ordered water molecules but density near Ser195 was interpreted as a sulfate ion. Both main chains, in agreement with the sequence of human GzmB deposited in SwissProt database (entry P10144), start with the N-terminal Ile16 and are fully defined up to Lys243. The last two residues are not defined by electron density probably due to enhanced flexibility. The main chain around segment Arg172-His173 is placed in relatively poor electron density, and a few exposed side chains projecting into the solvent are not defined by electron density so that their occupancy have been set to zero. The final crystallographic R-Factor is 22.7 % for all reflections from 15.0 to 3.2 Å.

A Ramachandran plot calculated using the program PROCHECK (Laskowski et al., 1993) shows that most of the residues fall into the most favoured or additionally favoured regions. This is above average for a structure determined at 3.1 Å.

Acknowledgements

We thank Milton T. Stubbs, Martin Renatus, Jens Kaiser and Jens Lüders for helpful discussions. W.B. acknowledges the financial support by the "Training and Mobility" and the "Biotech" programmes (ERBFMRXCT98 and BIO4-CT98-0418) of the European Union, by the HFSP programme (RG-203/98), by the SFB469 of the University of Munich, and by the Fonds der Chemischen Industrie, and H.R. the financial support by NIH grants AG10599, AR42931 and HL50523.

Table 1 . Statistics for Data Collection and Refinement.

Space group	$P2_12_12_1$
Cell constants (Å)	
a	41.74
b	114.31
c	135.52
Resolution (Å)	3.1
Reflections measured	40,709
R_{merge} (%)[a]	19.9
Unique reflections	11,275
Completeness (%), overall	83.2
Completeness (%), outermost (3.25-3.2 Å) shell	68.2
Reflections used for refinement	11,071
Resolution range (Å)	15-3.1
R-value (%)[b]	22.7
R_{free} (%)[c]	32.5
R.m.s deviations	
Bond lenghts (Å)	0.011
Bond angles (°)	1.511
Average B-value	27.8
Average B-value/SD[d] ($Å^2$)	2.27

[a] $R_{merge}=\sum_{hkl}| I - I|/\sum_{hkl}|I|$

[b] $R_{factor}=\sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$

[c] R_{free} is the R-value calculated with 500 reflections that were not used for the refinement.

Accession Numbers

Atomic coordinates have been deposited at the PDB under the ID code 1FQ3.

Paper 2. In preparation.

Entering the Cell in Pairs: Is Granzyme B Dimer the Form Internalized by Target Cells?

Introduction

Cytotoxic T lymphocytes (CTLs) and natural killer (NK) cells induce apoptosis in viral infected and cancer cells through two independent mechanisms, the Fas ligand/Fas-mediated and the granule exocytosis pathways (for a recent review see Krammer, 2000). A widely accepted model of the latter mechanism postulates synergy between neutral serine proteinases collectively termed granzymes, and the pore-forming perforin (Bleackley et al., 1988; Tschopp and Nabholz, 1990). Perforin released by CTLs and NK cells inserts into the membrane of the target cell and polymerizes in a calcium-dependent manner. The formed perforin channels are believed to facilitate access of the granzymes to the cytosol of the condemned cell, either passively or by reparative pinocytosis (for a recent review see Page et al., 1998). However, other experimental evidence supports receptor-mediated endocytosis instead (Pinkoski et al., 1998; Shi et al., 1997). Granzyme B (GzmB) induces apoptosis, a function shared with the caspases, and apparently related to its unique substrate specificity among serine proteinases. Identified GzmB targets include the executioner caspase 3 (Darmon et al., 1995), and death substrates such as poly(ADP-ribose) polymerase (Froelich et al., 1996).

Newly synthesized lysosomal enzymes acquire mannose 6-phosphate (M6P) residues during their biosynthesis in the endoplasmic reticulum and Golgi complex. Subsequently, these glycoproteins are sorted from the Golgi into a prelysosomal compartment by two different P-type lectins, the 46-kDa cation-dependent mannose 6-phosphate receptor (CD-MPR) (Dahms et al., 1987), and the 300-kDa cation-independent mannose 6-phosphate receptor/insulin-like growth factor II receptor (CI-MPR) (Lobel et al., 1987; Morgan et al., 1987) (for a recent review see Traub and Kornfeld, 1997). The same mechanism seems to play a major role in the trafficking of GzmB to the lytic granules of cytolytic lymphocytes (Griffiths and Isaaz, 1993). In keeping with their intracellular transport function, the two M6P-receptors are predominantly found in the

trans-Golgi network and in endosomes. However, after dissociation of the receptor-ligand complex in a prelysosomal compartment not all receptor molecules are recycled back to the Golgi, but a small portion (about 10-20%) moves to the plasma membrane. The subset of cell surface-exposed CI-MPR, but not CD-MPR molecules, perform important roles in the internalization of exogenous ligands, usually M6P-containing glycoproteins (Kornfeld, 1992; Stein et al., 1987). In addition, CI-MPR possesses distinct binding sites for non-glycosylated proteins, such as insulin-like growth factor II (IGF-II) (Morgan et al., 1987). Recently, Motyka and co-workers have shown that also human GzmB is internalized by target cells via CI-MPR (Motyka et al., 2000). Further, strong evidence indicates that GzmB binding and internalization proceeds via recognition of M6P residues in the carbohydrate moiety(ies) attached to the proteinase.

Is GzmB a Dimer Under Physiological Conditions?

We have recently solved the three-dimensional crystal structure of human granzyme B, using partially glycosylated recombinant protein produced in a baculovirus system (Estébanez-Perpiñá et al., 2000). One unexpected feature of the structure analysis was the observation of a GzmB dimer, stabilized by the interdigitation of oligosaccharide chains attached to Asn-65{71} in the two monomers (employing the chymotrypsinogen numbering system for GzmB, with numbers for the unprocessed zymogen given in curly brackets). The second potential *N*-glycosylation site at Asn-98{104} is not utilized in recombinant GzmB (unpublished observations). The crystal structure of the recombinant rat enzyme produced in *Escherichia coli* has recently been reported (Waugh et al., 2000). In this structure, monomeric, non-glycosylated GzmB forms a complex with an engineered variant of the *E. coli* inhibitor ecotin. Our structural finding, together with observations that binding and uptake of GzmB in target cells is effected by CI-MPR in a mannose 6-phosphate–dependent manner (Motyka et al., 2000), and that receptor dimerization is an essential element of the internalization mechanism (Byrd et al., 2000; York et al., 1999), strongly suggests that the glycosylated GzmB dimer is the form preferentially recognized by CI-MPR.

In this regard, human GzmB purified from granules of NK cells co-elutes with the disulfide-linked GzmA (apparent molecular mass 60-70 kDa) in gel filtration assays

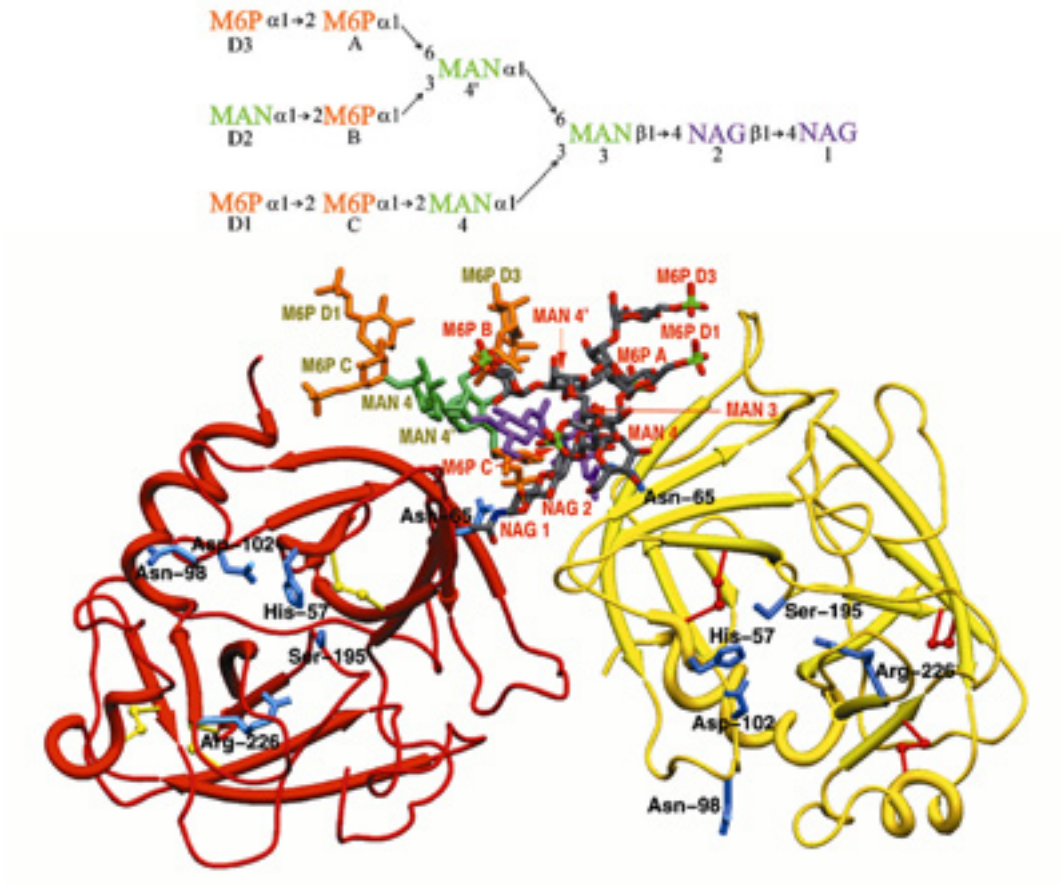
(Hameed et al., 1988). Human GzmB isolated from CTLs has been shown to exist in dimeric and monomeric forms (Poe et al., 1991). Further, GzmB is secreted from cytotoxic cells as a neutral high molecular weight form (apparent M_r > 100 kDa), which possesses pro-apoptotic activity (Galvin et al., 1999). This form was regarded as a complex of the proteinase with acidic glycosaminoglycans. In view of the structural evidence, GzmB storage in lytic granules and its secretion as an oligosaccharide-stabilized dimer appear to us as an alternative, simpler explanation. In solution, recombinant GzmB expressed in insect cells appears predominantly as a monomer (data not shown). This fact indicates that formation of a stable dimer depends on the presence of branched high-mannose sugar chains and/or their phosphorylation, perhaps because of the highly basic character of the protein-protein interface (Estébanez-Perpiñá et al., 2000).

Putative Location and Model of M6P Residues in Human Granzyme B

To our knowledge, the composition and structure of *N*-linked oligosaccharides attached to asparagine residues Asn-65{71} and Asn-98{104} in human GzmB has not been reported. The reduction in apparent molecular weight after EndoH treatment (~5-6 kDa as estimated from SDS-PAGE; see e.g. Griffiths and Isaaz, 1993; Sun et al., 1999) suggests that high mannose-type chains could be found at either or both positions. However, recombinant GzmB expressed in insect cells is not glycosylated at Asn-98{104}. Therefore, we limited ourselves to extend by model building the well-defined Asn-65{71} oligosaccharide moieties that cross-link the two GzmB monomers in our current crystal structure, to include M6P-moieties at the positions known to be phosphorylated in high mannose-type oligosaccharides (Varki and Kornfeld, 1983) (Fig. 1).

Fig. 1. Putative Location and Model of M6P Residues in Human Granzyme B.

Ribbon plot of the GzmB dimer with all possible phosphorylated mannose residues included (the inset shows an schematic diagram of a fully phosphorylated high mannose-type oligosaccharide). In analyzed glycoproteins, however, oligosaccharide species containing one or two phosphomonoester moieties predominate; see e.g. Dittmer et al., 1997). The crystal structure of human GzmB (Estébanez-Perpiñá et al., 2000) was used to generate the fully glycosylated model. Modeling was done interactively with MAIN (Turk, 1992), and the final model was minimized with XPLOR (Brünger, 1993). Notice the position of the second putative glycosylation site (Asn-98{104}).



The Cation-Independent Mannose-6-Phoshate Receptor

Comparative sequence analysis indicates that the GzmB receptor CI-MPR contains 15 tandem repeats of a domain (150 amino acids in length), which is homologous to the single M6P-binding extracytoplasmic domain of the cation-dependent

receptor (Dahms et al., 1987; Morgan et al., 1987). The structure of the CD-MPR ligand-binding domain has recently been solved by X-ray crystallography (Roberts et al., 1998), and the elements responsible for the recognition of M6P-containing oligosaccharides have further been explored by X-ray analysis of its complex with a yeast-derived pentamannosyl phosphate (Olson et al., 1999a) and site-directed mutagenesis (Olson et al., 1999b). This domain crystallizes as a dimer in which both subunits associate via a large (~1500 $Å^2$) apolar interface, locating the two independent ligand binding sites at a linear distance of approximately 40 Å (Roberts et al., 1998). These features are in agreement with the fact that CD-MPR dimers represent the predominant form of the receptor on cell membranes. Further, dimer formation does not depend on ligand binding and/or dissociation (Punnonen et al., 1996).

Two distinct high affinity binding sites for M6P (Tong et al., 1989) have been localized to domains 3 and 9 in the extracytoplasmic domain of CI-MPR (Dahms et al., 1993; Marron-Terada et al., 1998; Westlund et al., 1991). The strict conservation of several residues critical for binding of lysosomal enzymes (Olson et al., 1999b), together with the overall conservation of core residues (Roberts et al., 1998), indicates that the two P-type lectins exhibit a common fold as well as similar M6P-binding pockets.

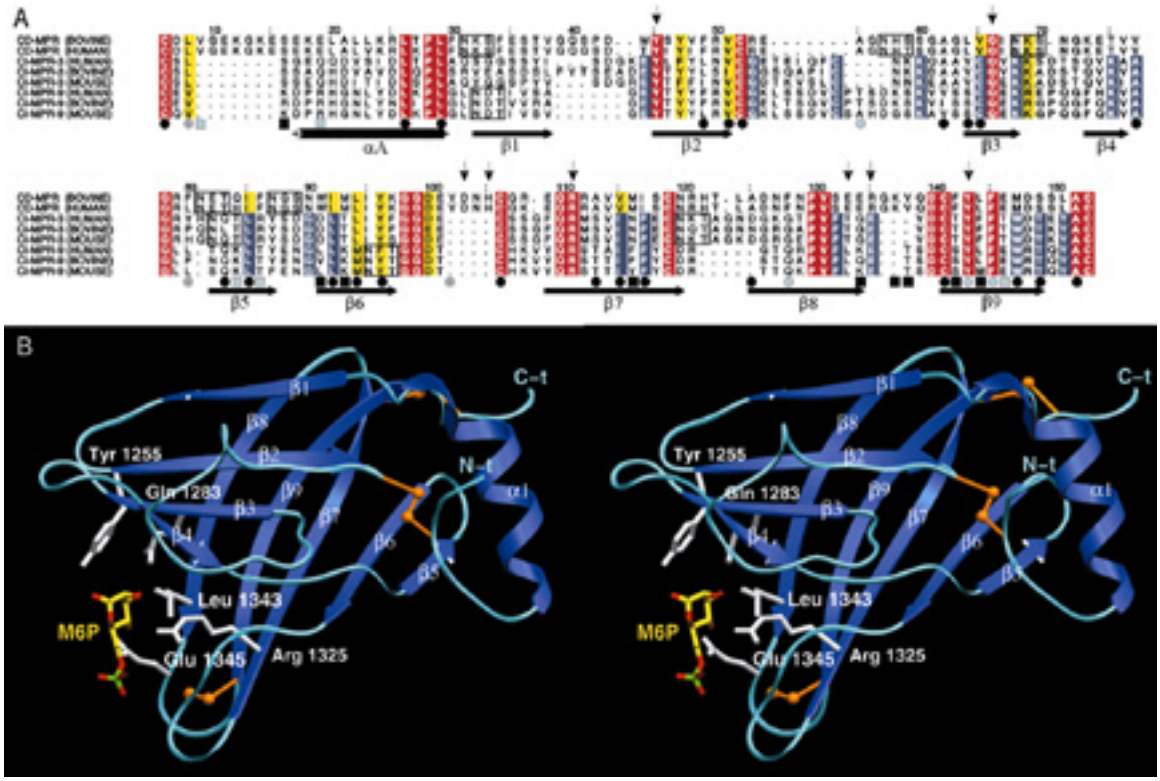
To investigate the probable binding mode of GzmB to its cell surface receptor, we generated three-dimensional models of domains 3 and 9 of human CI-MPR, taking advantage of their sequence homology with bovine CD-MPR (Fig. 2A). The three-dimensional model of human CI-MPR-9 bound to M6P is presented in Fig. 2B.

Fig. 2. Modeled Structures of M6P-Binding Domains in Human CI-MPR

(A) Sequence alignment of the ligand binding domains of the cation-dependent receptor (human and bovine proteins), and domains 3 and 9 in the human, bovine, and murine CI-MPRs. The alignment was produced with the program PileUp from the Wisconsin Package, manually modified at a few sites to conform with the secondary structure of CD-MPR (Roberts et al., 1998) (given below the alignment). Numbers refer to bovine CD-MPR, as used for structure description. Strictly conserved residues are shadowed red, other highly similar residues are shadowed yellow. Residues strictly conserved in the M6P-binding domains of CI-MPR from different species are

shadowed blue. Putative *N*-glycosylation sites are boxed. Residues buried in the hydrophobic core of CD-MPR are marked by black (relative accessible surface area below 1%) or gray circles (accessible surface area between 1 and 5%). Notice that residues involved in receptor dimerization are usually not conserved in CI-MPR domains; these residues are indicated by black (if over 80% of the relative side chain surface area becomes buried upon dimerization) or gray squares (between 50 and 80% buried). Arrows indicate the CD-MPR residues involved in ligand binding. Notice that the M6P-binding residues Gln-66, Arg-111, and Tyr-143 in CD-MPR are conserved throughout.

(B) Ribbon plot of human CI-MPR domain 9 (secondary structure elements dark blue, loops light blue). Side chains of residues predicted to contact the bound M6P ligand (color-coded) are shown. Although essentially conserved, minor differences in the substrate binding pockets of domains 3 and 9 (for instance, in the β1-β2 and β3-β4 hairpins that would face the mannose residue preceding the terminal M6P moiety) would account for their slightly different ligand binding properties (Marron-Terada et al., 1998; 2000). CI-MPR numbering refers to the unprocessed receptor.



Oligomerization State of CI-MPR

In contrast to the dimeric character of the cation-dependent receptor, the oligomerization state of CI-MPR is a matter of controversy, with multiple experimental evidence supporting both the presence of monomeric and dimeric forms (Byrd et al., 2000; Marron-Terada et al., 2000; Stein et al., 1987; York et al., 1999). Recombinant truncated forms containing extracytoplasmic domains 1-3, 7-9, and 7-11 exist as monomers (Marron-Terada, 2000). On the other hand, chimeric variants that only include domains 11-15 and 13-15 dimerize on the cell surface (Byrd et al., 2000), suggesting that ligand binding and receptor dimerization may be effected by distinct domains (see below). In any case, it is remarkable that the residues (mostly hydrophobic) involved in receptor dimerization in CD-MPR are missing in CI-MPR domains 3 and/or 9, or are replaced by hydrophilic residues. For instance, the central Phe-86 – Phe-142 pair of interacting aromatic side chains (Roberts et al., 1998) are replaced by arginine/threonine and threonine/serine, respectively (Fig. 2A). Thus, under physiological conditions it would be extremely unlikely that the ligand-binding CI-MPR domains form homodimers similar to the one constitutively formed by CD-MPR.

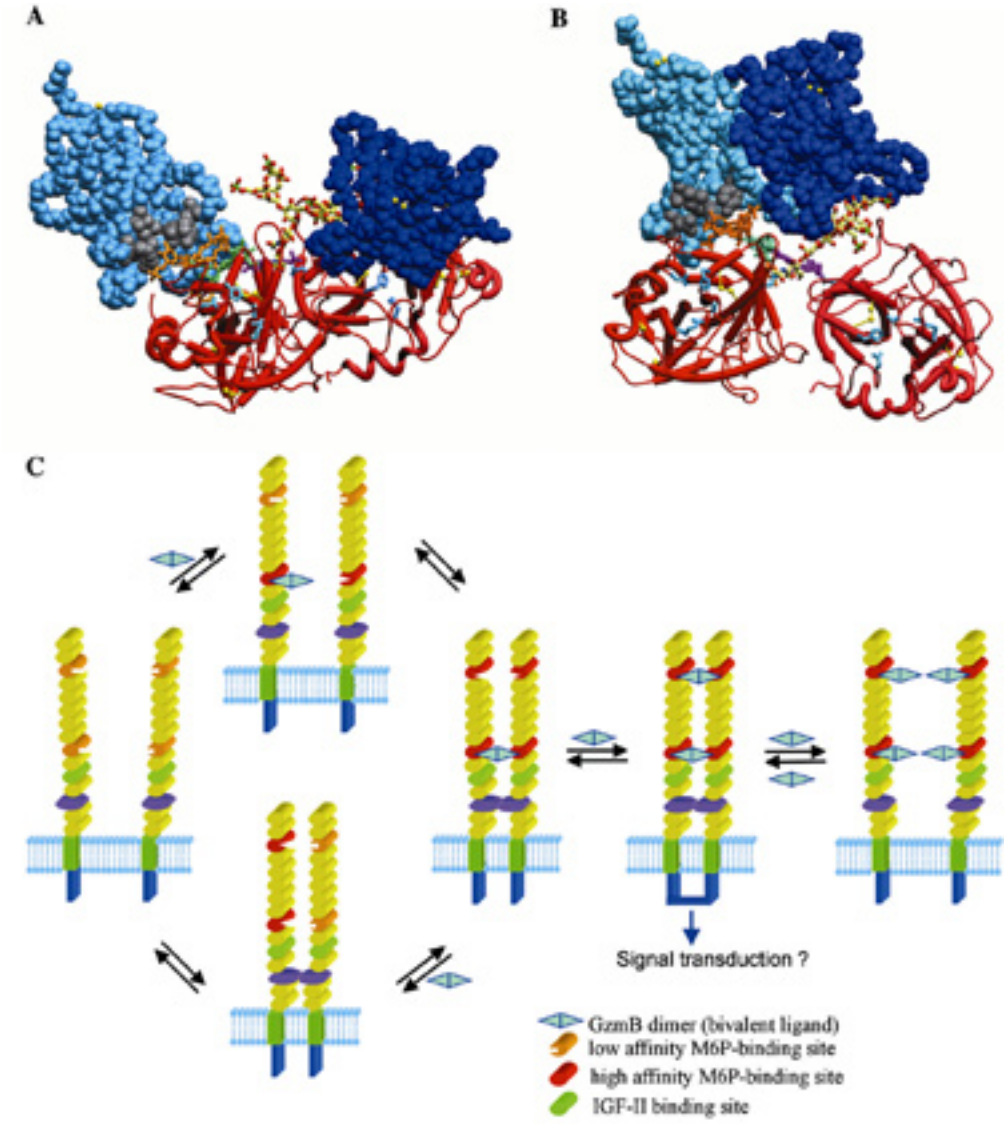
In accord with the structural and biochemical evidence discussed above, we assumed binding of separate CI-MPR domains to an 'uncovered' mannose 6-phosphate moiety in each Asn-65–linked oligosaccharide chain of GzmB. An inspection of the relative arrangement of M6P moieties in the glycosylated GzmB dimer (Fig. 1) suggests that the M6P-binding domains could dock in strikingly different ways. For instance, in an arrangement characterized by receptor binding to M6P residue D3 in both GzmB monomers, a GzmB dimer would be sandwiched between the ligand-binding regions of two proximal monomeric receptors (Fig. 3A). In a second arrangement, binding to M6P moieties D1 and D3 in GzmB monomers A and B, respectively, would locate both CI-MPR modules approximately on 'top' of the dimer (Fig. 3B). Thus, it would seem that CI-MPR possesses a greater flexibility for ligand selection than the dimeric CD-MPR, which might be essential for rapid interactions with heterogeneously glycosylated ligands. In this regard, the cation-independent receptor has a higher affinity for lysosomal enzymes, recognizes a broader spectrum of these glycoproteins than does the cation-dependent receptor (Kornfeld, 1992; Sleat and Lobel, 1997), and is therefore more

efficient than the latter in targeting lysosomal enzymes to the lysosome (Sohar et al., 1998; Stein et al., 1987).

Fig. 3. Putative Binding Modes of Granzyme B to CI-MPR.

(A) and (B) Two distinct CI-MPR-9 modules could dock to the GzmB dimer via different exposed mannose 6-phosphate units. Considering that CI-MPR strongly prefers M6P($\alpha 1 \rightarrow 2$)Man disaccharides (Tong et al., 1989), favored receptor binding to M6P residues C, D1, and D3 could be expected (compare Fig. 1).

(C) Schematic drawing of two putative mechanisms of GzmB uptake. Most of the 15 extracytoplasmic CI-MPR domains are represented as yellow motifs, while the M6P-binding domains 3 and 9 are shown as light orange or red motifs (low and high affinity forms, respectively). The IGF-II binding domain 11 and the 'irregular' domain 13 are given in green and pink color, respectively. At high receptor concentrations on the cell surface, dimerization occurs (Byrd et al., 2000). The ligand binding sites on each monomer, however, could be only partially accessible, or be oriented relative to each other in a non-productive manner (i.e., not suitably placed for simultaneous binding to a single bivalent ligand, schematically indicated with light orange motifs). Alternatively, and at a lower receptor density, monomeric CI-MPR would predominate. Binding of a bivalent ligand to one receptor molecule would facilitate interactions with a neighboring molecule and induce dimerization. Negative cooperativity (Byrd and MacDonald, 2000) occurs because excess ligand competes for the available binding sites, weakening or disrupting the 2:2-dimer.



The apparent discrepancy between recent results concerning the oligomeric state of CI-MPR suggests two complementary pathways of ligand binding and uptake (Fig. 3C): Either (1) pre-formed dimeric, but not fully accessible forms of CI-MPR exist on the cell surface, which are stabilized upon binding of bivalent M6P-bearing ligands as high-affinity forms (Byrd and MacDonald, 2000; Byrd et al., 2000), or (2) binding of bivalent ligands to a monomeric receptor favors ligand binding to the homologous domain of a neighboring receptor molecule, promoting receptor dimerization (York et al., 1999), perhaps accompanied by the creation/exposure of a second high-affinity binding site by the other pair of M6P-binding domains. Domain 13, which harbors a unique inserted fibronectin type II domain, would be a likely candidate for receptor dimerization.

This module possesses a highly conserved cluster of aromatic residues (e.g., Phe-1909, Tyr-1914, Trp-1926, and Trp-1933 in human CI-MPR), which form a solvent-exposed hydrophobic surface in homologous domains of known three-dimensional structure (see e.g. Pickford et al., 1997). We notice that the affinity of truncated receptor forms containing both domains 11 and 13 for IGF-II is comparable to that of the wild-type receptor (Devi et al., 1998), while domains 7-11 exhibit a 40-fold lower affinity for this ligand (Marron-Terada et al., 2000). The affinity-enhancing activity effected by domain 13 is easily explained in terms of its contribution to receptor dimerization.

Even though the Asn-98 site could also be glycosylated in native GzmB, carbohydrate chains attached to Asn-65 and Asn-98 of the same GzmB molecule would extend away in opposite directions (Fig. 1). Thus, the M6P-residues of these two oligosaccharide chains would be located up to 70 Å apart (phosphate-phosphate distance). It is conceivable that this arrangement of putative binding sites does not support receptor dimerization. More relevantly, the reduced conformational freedom of oligosaccharide chains attached to the GzmB monomers at Asn-65 would facilitate interactions with separate but neighboring CI-MPR molecules. Both features would strongly favor recognition of the GzmB dimer over the (hypothetical) double-phosphorylated monomers.

Paper 3. In Preparation.

Crystal Structure of Human pro-Granzyme K at 2.2 Å Reveals a Novel Mechanism for Zymogen Stabilization.

3.1) Introduction to Granzyme K

Human granzyme K (GzmK) is a proteinase of approximately 28 kDa that was initially isolated from the granules of interleukin-2 (Il-2)-activated peripheral blood mononuclear cells. It was the third esterase characterized from these granules, initially named granzyme 3 (tryptase-2), and was the second Z-Lys-SBzl-cleaving esterase found in lymphokine-activated killer cells (Hameed *et al*., 1988, Sayers, 1994; Przetak, 1995). GzmK shares a substantial degree of sequence similarity (approx. 40-50%) with other granzymes (Gzms) (see Fig. 19). GzmB and GzmA are the most abundant granzymes in the secretory granules of cytotoxic T-lymphocytes and natural killer cells, whereas GzmK, together with GzmH, are only present in minor amounts. These later Gzms remain still poorly studied, primarily due to a lack of sufficient quantities of those proteins. Recently, however, two groups have managed to obtain milligram amounts of recombinant pro-GzmK, in *Bacillus subtilis* (Babé. *et al*., 1998) as well as in *E. coli* (Wilharm *et al*., 1999).

3.1.1) GzmK Substrate Specificity

GzmK exhibits a substrate specificity similar to GzmA, since it preferentially hydrolyses peptide bonds C-terminal to both Arg or Lys (see Fig. 13). Measurements of *Kcat*/*Km* for single-residue thioester substrates show approximately a two-fold preference for a Lys versus Arg residue at P1. No activity is observed for ester substrates with other residues at the P1 position (Babé *et al*., 1998).

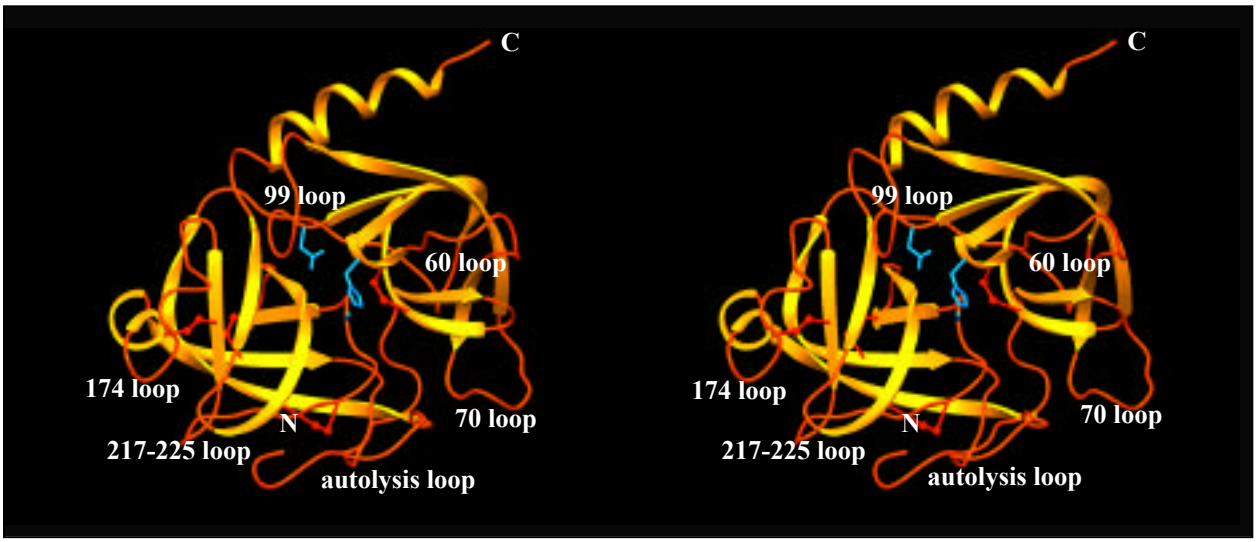
3.1.2) Inhibitors of GzmK

GzmK substrate hydrolysis is blocked most efficiently by inhibitors that contain Lys or Arg at position P1. GzmK is inhibited by the synthetic compounds D-Phe-Pro-Arg-chloromethyl ketone, phenyl-methyl-sulfonyl fluoride, benzamidine, and by the Kunitz-type inhibitor aprotinin. The plasma-derived inter-α-trypsin inhibitor complex, its

bikunin subunit, and the second carboxy-terminal Kunitz-type domain of bikunin were identified as GzmK genuine physiological inhibitors. Most probably, these inhibitors neutralize extracellular GzmK after T-cell degranulation, thus avoiding unspecific damage of bystander cells at sites of inflammatory reactions (Wilharm *et at*., 1999).

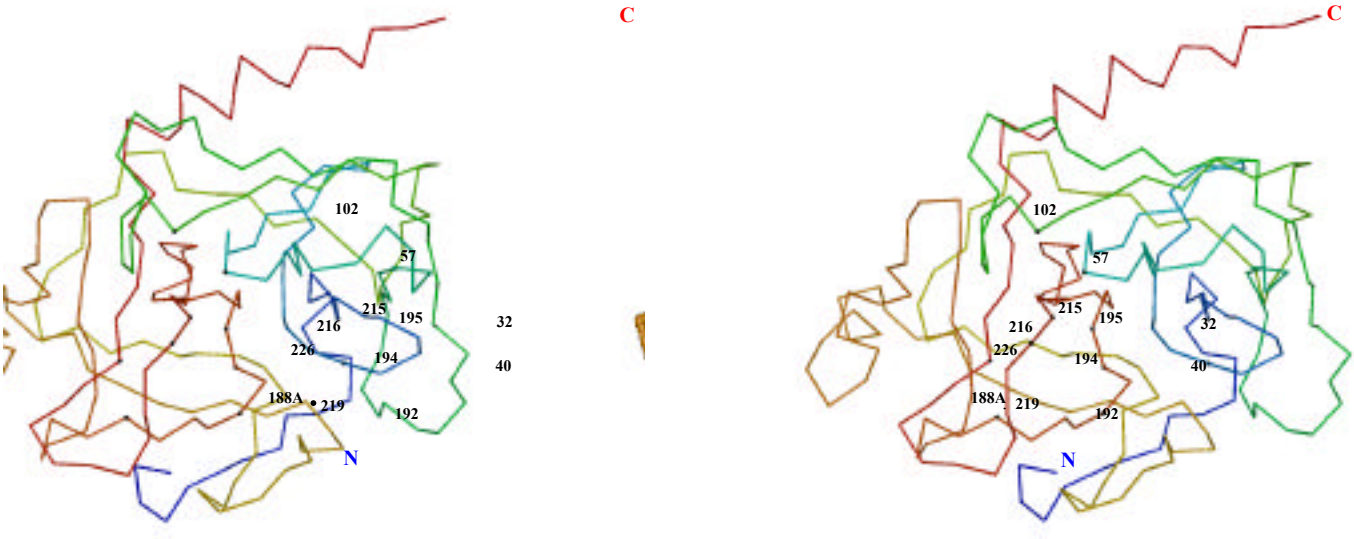
3.2) Overall Structure of Pro-granzyme K

Human pro-GzmK is an oblate ellipsoid with diameters of about 35 and 50 Å. As for other members of the trypsin-like serine proteinase family (Bode and Huber, 1986), the GzmK polypeptide chain essentially folds into two six-stranded β-barrel domains held together by three trans-domain straps (Fig. 17). The surface is made up of several turn structures, two single turn α-helices from Ala56 to Cys58 and from Arg165 to Asn169, and the C-terminal mixed 3_{10}-α-helix (residue Lys234 to Leu243). The catalytic residues Ser195 (Ala195 in the mutant crystallographically studied), His57 and Asp102 are located at the junction of the two β-barrels, while the active-site cleft runs along the surfaces of both barrels perpendicular to this junction (Fig. 1).



3.2.1) Fig. 1. Crystal Structure of Human Pro-Granzyme K.

Stereo ribbon representation of the pro-GzmK molecule. Pro-GzmK is shown in the standard orientation. The catalytic triad residues and the disulfide bridges are represented as blue and red stick models, respectively. The figure was done with Setor (Evans, 1990).



3.2.2) Fig. 2. Stereoview Cα trace of pro-Granzyme K in standard orientation.

The position of residues 32, 40, 57, 192, 194, 195, 215, 216, 219, 226 are indicated. This figure was drawn with BOBSCRIPT (Esnouf, 1997; Esnouf, 1999).

3.3) Electrostatic Surface Potential

A distinct feature of the pro-GzmK molecule is its uneven charge distribution: positively charged residues particularly cluster to the “east” and the “north-west” of the active site as seen in the standard orientation (Fig. 3). These patches of positively charged residues very much resemble the fibrinogen recognition exosite (anion binding exosite I) and the heparin binding site (exosite II) in thrombin (Bode et al., 1992).

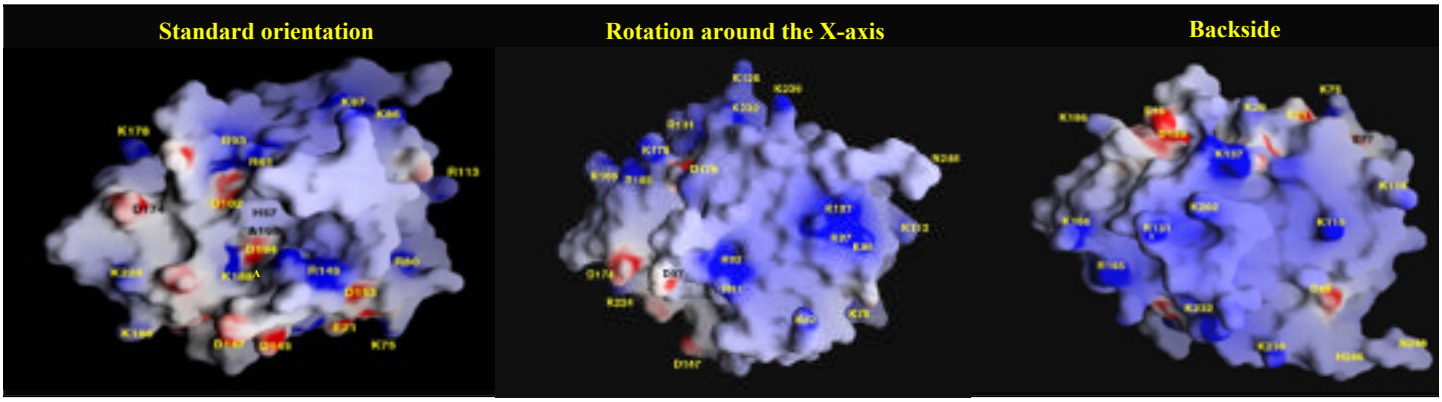
Compared to other serine proteinases, the percentage of polar residues is not particularly high in GzmK. However, the 8 arginine and 23 lysine residues are not charge-balanced by the 6 glutamate and 11 aspartate residues, rendering pro-GzmK very basic (isoelectric point 10.2). Most charged side chains are, at least partially, exposed to

the bulk solvent. Exceptions are the catalytic Asp102 and Asp194, which are located within the surface shell and the S1 pocket, respectively.

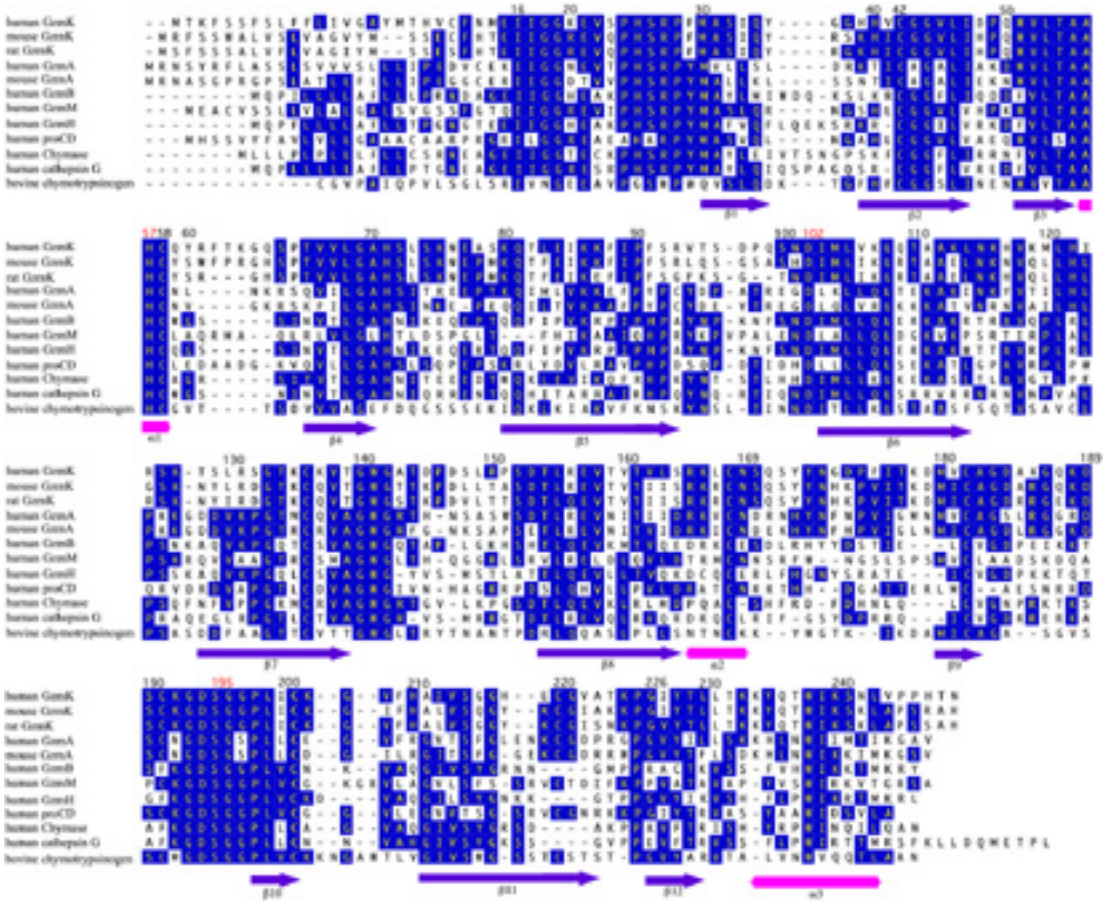
The juxtaposition of lysines (K188A and K192), arginines (R61, R93 and R159) and the catalytic H57 in and around the S1-pocket, charge-compensated only by Asp102 and Asp194, gives rise to an overall positive electrostatic potential around the negatively charged S1 pocket (Fig. 3). This contrasts with the strong negative potential of most other serine proteinases, in particular of CatG, at this surface site (Hof *et al*., 1996). In pro-GzmK, this basic region extends towards the east via Arg149 into a larger positively charged surface patch made up mainly of the side chains of Lys75, Arg80, Lys113, and Lys116 (Fig. 3).

3.3.1) Fig.3. Solid Surface Representation of Pro-GzmK

The protein is shown in three different orientations: in the standard orientation (left panel), after a 90° rotation around the X-axis (central panel), and a the backside of pro-GzmK (right panel). The colours indicate positive (blue) and negative (red) electrostatic potential at the molecular surface contoured from +10 kT/e to -10 kT/e. The figure was made with GRASP (Nicholls *et al*., 1993).



3.4) Fig. 4. Structure-based Amino Acid Sequence Alignment



The sequence of human GzmK (Shersta *et al*., 1997) is aligned with those of human chymase (Caughey *et al*., 1991), human cathepsin G (Salvesen *et al*., 1987), and bovine chymotrypsinogen A (Hartley and Shotton, 1971). The numbers shown refer to bovine chymotrypsinogen A. The β-sheets (β1 to β12) and α-helices (α1 to α3) of GzmK are indicated by arrows and cylinders, respectively. The catalytic triad residues are indicated with red numbers. The alignment was prepared with DNAStar Package. A proper alignment with bovine chymotrypsinogen requires residue insertions at positions 60, 170, 175, 185, 215, 233, and deletions at 36, 127, 203-204, 206-207, and 217-218.

3.5) Pro-GzmK Active-site Cleft

The residues of the active-site triad, Ser195 (mutated to Ala in the form studied here), His57, Asp102, and other catalytic elements such as the oxyanion hole are arranged at the centre of the cleft. The still non-formed specificity pocket (S1) opening to the west of Ala195 is bordered by segments Val213-Cys220 (the entrance frame), Ser190-Ala195 (the basement), Lys225-Tyr228 (the back of the pocket), and the Cys191-Cys220 disulphide bridge (closing the pocket towards the south). Unlike GzmB and GzmM, but in agreement with the granzymes A and H, CatG and human chymase, residue 228 has an almost strictly conserved aromatic side chain as in most other chymotrypsin-like proteinases (e.g. a phenylalanine in CatG or a tyrosine in the chymases). The side chain of His57 is slightly rotated outwards of the pocket (see Figs. 6 and 7), which could be a crystallographic artifact due to the S195A mutation, compared to chymotrypsinogen, trypsinogen, GzmB, CatG, chymase A, elastase and pro-CFD, where it is facing Gly216. Position 215, one of the substrate binding sites which is a glycine only in case of human and rat GzmK, points outwards, a conformation not favourable for possible substrate interactions (Figs. 5 and 6). Normally, this position is occupied by an aromatic residue, which is oriented in the opposite direction (Fig. 6). Another difference with other serine proteinases and their zymogens is the location of the side chain of Asp189, which is directed away from the pocket. The side chain has to turn around 180° to give rise to the active conformation of GzmK.

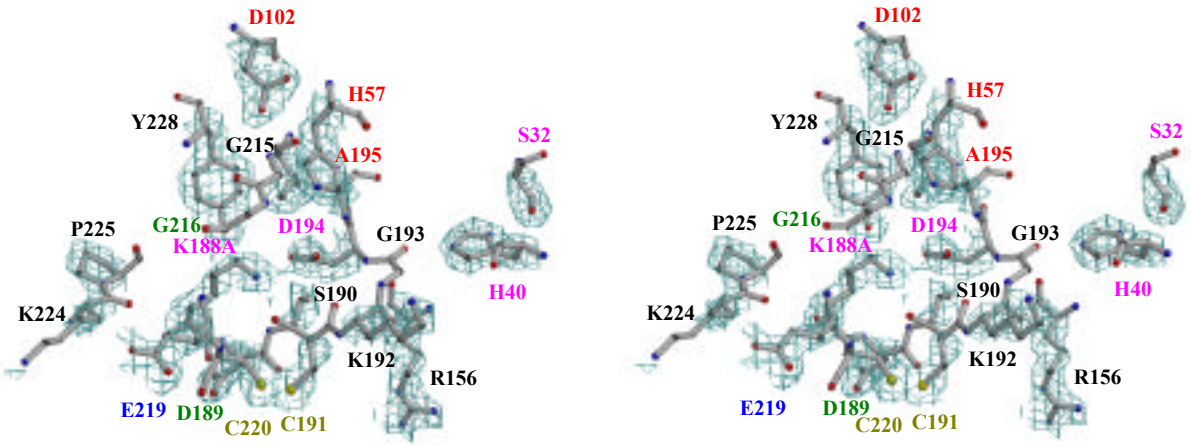


Fig.5 (previous page). Stereoview of pro-granzyme K active site pocket. The catalytic residues are shown in red (His57-Asp102-Ala(Ser)195; in pink are indicated the "zymogen triad" (missing in pro-GzmK) (His40-Asp194-Ser32), K188A is also shown in pink as it stablishes an hydrogen bond with Asp194; the substrate specificity elements G216 and Asp189 are in green; the 191-220 disulfide bridge is shown in dark yellow; residue 219 (shared with FIXa) is coloured blue. Other residues located in the specificity pocket are shown in black.

3.6) Comparison of Pro-GzmK with Several Serine Proteinases

Given the conserved mechanism of serine proteinase activation, we assumed that the zymogen structural characteristics found in previously solved zymogen structures will be also present in the structure of pro-GzmK. However, several novel features could be seen in the structure of this GzmK zymogen.

3.6.1) Comparison of Pro-GzmK with Other Serine Proteinase Zymogens

There are only few structures of serine proteinases zymogens solved so far, i.e., trypsinogen (Bode *et al*., 1976, Fehlhammer *et al*., 1977), chymotrypsinogen A (Wang *et al*., 1985), bovine ternary complex (Gomis-Rüth, 1997), pro-complement factor D (Jing *et al*., 1998) and prethrombin (Vijayalakschmi, 1994). An optimal superposition of pro-GzmK with the following four zymogens including all the α-carbon atom pairs within 1.5 Å results in the following topological values calculated with the program Top3d. **Table 1**:

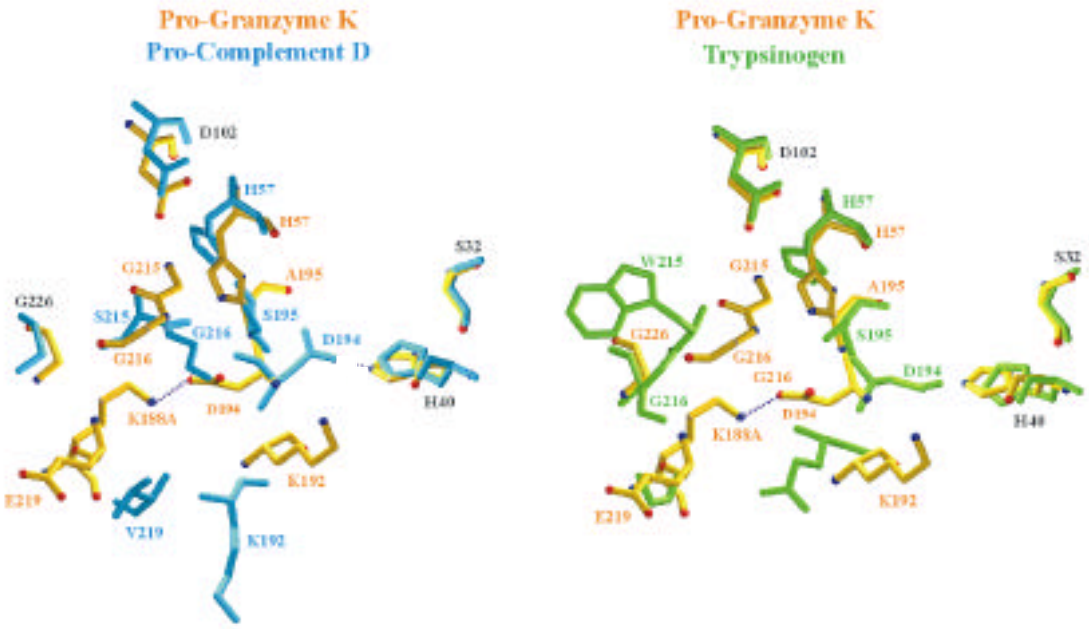
GzmK compared with	Matching residues	Identical residues	Identity (%)	α-carbon atoms r.m.s deviation
Pro-CD	180	81	46.0	1.09
ChymoA	180	74	41.1	1.30
Trypsinogen	189	74	39.2	1.24
Prethrombin-2	182	58	31.9	1.29

We have compared the crystal structure of pro-GzmK with that of pro-complement D and trypsinogen. Trypsinogen, the inactive precursor of trypsin, is very similar in structure to trypsin in about 85% of the molecule (mean deviation of internal

main chain atoms: 0.2 Å), the rest being different (Bode *et al*., 1976, Felhammer *et al*., 1977). The activation domain is highly flexible and displays no significant electron density (Bode and Huber, 1978). The chain segments, including residues 184 to 193 and 217 to 233 (which form the specificity pocket in trypsin), the residues preceding Tyr20, and the autolysis loop between residues 142 and 152, are partially and completely flexible, respectively. The structure of pro-GzmK revealed several unexpected features in comparison with trypsinogen. To begin with, although residues Ser32, His40, and Asp194 are conserved in pro-GzmK, they do not form any zymogen triad since Asp194 adopts a completely unusual conformation. This residue, which normally makes a hydrogen bond with His40, is rotated about 180°, adopting a conformation more closely related to the one observed in active serine proteinases. Besides, Asp194 makes a hydrogen bond with residue Lys188A. This newly described hydrogen bond between Asp194 and Lys188A represents a novel serine proteinase zymogen stabilization mechanism (Fig. 21). During activation, the side chain of Asp194 usually has to rotate 180°. Secondly, the activation domain in trypsinogen was not defined, whereas in pro-GzmK the above mentioned segments are fully defined by electron density, and thus show much less thermal disorder. On the other handside, the pro-GzmK polypeptide chain cannot be traced in the Fourier and difference Fourier maps from its N-terminus up to Gly18, similar to free trypsinogen (Bode *et al*., 1976; Feldhammer *et al*., 1977) and the complex trypsinogen-pancreatic trypsin inhibitor (PTI) (Bode *et al*., 1978).

We have also compared the GzmK structure with that of pro-Complement factor D (pro-FD). Complement factor D (FD) is a serine proteinase from the alternative pathway of the complement cascade. FD is unique among serine proteinases in that it requires neither enzymatic cleavage to be active nor inactivation by an inhibitor for its control. Regulation of FD activity in blood is apparently achieved by a novel induced-fit mechanism that depends on conformational changes effected upon substrate binding, which allow reversible expression of proteolytic activity (Kim *et al*., 1995). Its natural substrate, C3bB, seems to induce these conformational changes, resulting in reorientation of the catalytic centre and of the substrate binding site, both of which have atypical structures ((Narayana *et al*., 1994), (Fig. 6)). The structure of pro-FD is highly flexible and exhibits disordered conformation in five regions: 22, 71-76, 143-152, 187-193, and

215-223. These regions are involved in interactions with the N-terminus in the mature enzyme (see below).



3.6.1.1) Fig. 6. Superposition of pro-GzmK with pro-CD (A) and trypsinogen (B) around the entrance of the S1 pocket. Pro-GzmK is represented as a yellow stick model whereas pro-CD and trypsinogen are shown as blue and green stick models, respectively. The most outstanding differences between the granzyme zymogen and the other two serine proteinase zymogens are, firstly, the atypical conformation of Asp194, which instead of forming an hydrogen bond with His40 (that together with Ser32 form the zymogen triad) is hydrogen-bonded to Lys188A (this hydrogen bond is shown in a violet dashed line).

3.6.2) Comparison of Pro-GzmK with Active Serine Proteinases (i.e. Granzyme B and Active Complement D)

An optimal superposition of the α-carbon atom pairs within 1.5 Å from pro-GzmK structure with those of human GzmB (Estébanez-Perpiñá *et al*, 2000), cathepsin G (Hof *et al*., 1996), human chymase (McGrath *et al*., 1997, Pereira *et al*., 1999), leukocyte elastase (Bode *et al*., 1986, Sihna *et al*., 1987), and blood coagulation factor IXa

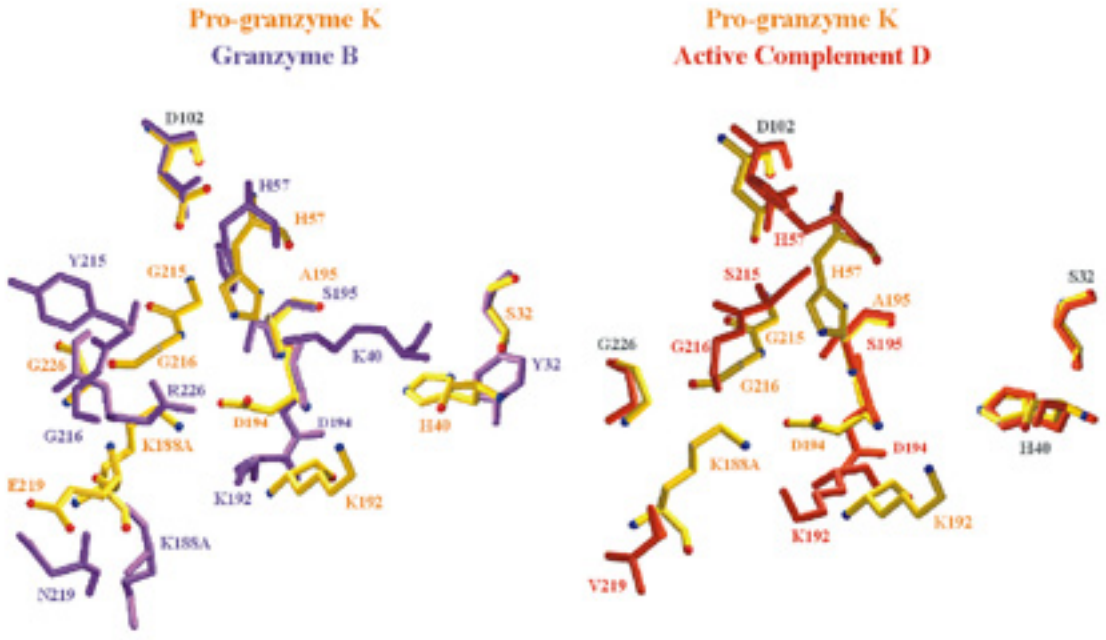
(Brandstetter *et al*., 1995) results in the following topological values calculated with the program Top3d. **Table 2**:

GzmK with	Matching residues	Identical residues	Identity (%)	α-carbon atoms r.m.s deviation
GzmB	189	84	44.4	1.14
Chymase	186	79	42.5	1.07
CatG	185	75	40.5	1.06
Elastase	182	66	36.3	1.17
FIXa	191	58	30.4	1.35

Thus topologically, pro-GzmK most closely resembles GzmB, pro-CD and human chymase (what fits well to the proposed co-evolution suggested by Smyth et al., 1996). The topological equivalence with chymotrypsinogen forms the basis for the sequence alignment of the eleven serine proteinases shown in Figure 4.

As already mentioned, factor D is a very efficient proteinase that apparently does not require activation cleavage by a specific activating enzyme or inhibition by a natural proteinase inhibitor. These unique functional characteristics were shown to correlate well with the atypical structural features present in the structure of active factor D (Narayana *et al*., 1994, Jing *et al*., 1998) where atypical conformations for the catalytic triad residues His57 and Asp102, and the loop 214-218 that forms the entrance frame to the S1 specificity pocket were observed. Jing *et al*. proposed that loop 214-218 acts as a self-inhibitory loop that dictates the resting state conformation of FD (Jing *et al*, 1998). The model of Volnanakis and Narayana (1996) proposes that the natural substrate of FD, factor B, upon binding induces a reversible conformational change in this proteinase switching it from the resting state to an active state, where the catalytic triad and the substrate binding site are functional.

In factor D, the imidazol side chain of His57 is in an atypical *trans* orientation whereas in pro-GzmK structure, His57 also adopts an atypical *gauche* conformation (see Fig.6). Both conformations of His57 deviate from the active tautomer conformation (the catalytically active *gauche* orientation).



3.6.2.1) Fig. 7. Superimposition of pro-GzmK with active complement D and granzyme B around the S1 pocket. Pro-GzmK is represented as a yellow stick model, whereas GzmB and active FD are shown in violet and dark orange stick models, respectively. It is noteworthy to point out in these figures the atypical non-catalytical conformation of His57 in both pro-GzmK and active factor D. In pro-Gzm K, His57 adopts a non-functional *gauche* orientation and in active factor D a *trans* orientarion (Kim *et al*., 1995).

3.6.3) Coagulation Factor IXa and Pro-GzmK: the Role of Residue 219

Coagulation factor IXa (FIXa) is a key serine proteinase in blood coagulation and its hereditary deficiency causes hemophilia B. FIXa and pro-GzmK have a glutamic acid residue at position 219, which normally is a Gly among other trypsin-like serine proteinases. Glu219 in FIXa (like the equivalent Gly-residues in all the other trypsin-like enzymes) exhibits a high energy main chain conformation (Brandstetter *et al*., 1995),

which in proteins is generally adopted only by Gly residues. In the pro-GzmK structure, E219 also adopts the energetically unfavourable conformation found in FIXa.

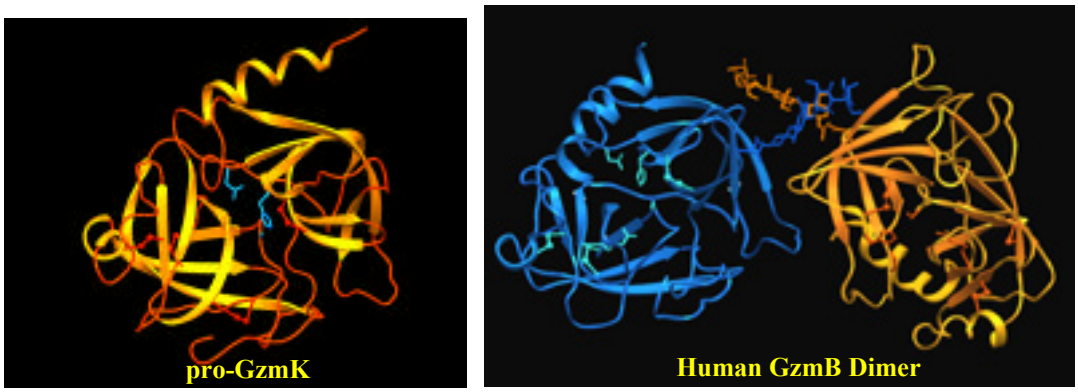
3.6.4) Overall Comparison of Pro-GzmK with Human GzmB

We have previously solved the crystal structure of recombinant human Granzyme B (GzmB) (Estébanez-Perpiñá *et al*., 2000), which is the prototypic member of the Gzms. GzmB has been proven to directly trigger apoptosis (programmed cell death) in target cells by activating several executioner pro-caspases (i.e. pro-caspase-3 and pro-caspase-7) as well as Bid, a pro-apoptotic Bcl-2 family member. GzmB exhibits an unusual substrate specificity as it cleaves peptide bonds after an aspartic residue. This unique substrate specificity is conferred by Arg226 located at the back of the specificity pocket. On the contrary, GzmK cleaves peptide bonds following an Arg or Lys residue. Instead of an Arg at position 226, this proteinase contains a Gly (see Fig. 4). Active GzmB and the zymogen of GzmK have a similar overall fold as can be seen in Fig. 8, however, major differences can be found in some of the loops that define the environment of the active site cleft and in the specificity pocket (which is still incompletely formed in GzmK zymogen) (for more details see Fig. 8).

Pro-Granzyme K
Granzyme B

3.6.4.1) Fig. 8. Stereo ribbon plot of pro-GzmK (in yellow) superimposed with human GzmB (in blue) (Estébanez *et al*., 2000). The view is in the standard orientation. The figure was done with Setor (Evans, 1990).

Recombinant human GzmB expressed in a baculovirus system crystallized as a dimer interconnected through the oligosaccharide chains *N*-linked to Asn65. The question whether a GzmB dimer would be physiologically relevant is still under investigation. Human GzmK, however, lacks consensus sequences for N-glycosylation. Furthermore, the crystallized protein was obtained from *E.coli*. The two structures are compared in the following figure.



3.6.4.2) Fig. 9. Side to side ribbon plots of pro-GzmK (left) and GzmB (right). The sugar chains N-linked to Asn-65 were two units of N-acetylglucosamine and 1 mannose units.

3.7) Discussion

Until recently no structural evidence has been available for this important class of serine proteinases involved in the protection of our body against viral infections and tumor cells (Waugh, 2000; Estébanez-Perpiñá *et al*., 2000; Rotonda, 2001). We have solved the crystal structure of recombinant uninhibited human granzyme B, structure that unexpectedly proved to be a dimer stabilised by the interdigitation of the *N*-linked sugar units (Estébanez-Perpiñá *et al*., 2000). Structural work on granzyme zymogens was missing, however, and thus a structure-based view of their activation mechanism. We have crystallised the recombinant human pro-GzmK mutant S195A and have solved the corresponding crystal structure at room temperature (293 K) and at cryo-conditions (125 K).

In contrast to the large number of active serine proteinases studied by X-ray crystallography the structures of over 20 different serine proteinases have been deposited in the Protein Data Bank, only few zymogens have been solved so far. The crystal structure of human pro-GzmK reveals an overall fold similar to that found in other zymogens of serine proteinases. However, the structure also revealed unexpected features when compared with the structures of other zymogens. A common characteristic of these zymogens is the presence of a zymogen triad, which appeared to be a general feature of the pancreatic serine proteinase zymogens (Madison *et al*., 1993). Since residues Ser32, His40, Asp194 are conserved in pro-GzmK, it was supposed that this zymogen also possesses the zymogen triad like trypsinogen and chymotrypsinogen. Most strikingly, however, the zymogen triad is missing in pro-GzmK as Asp194 is rotated about 180° with respect to all other known zymogens. Hence, Asp194 has a conformation more closely related to that found in the active enzymes, and the zymogen is stabilised via a previously unobserved mechanism: Asp 194 is forming a salt bridge with Lys188A. The conservation of this residue inserted in the 188-loop of mouse GzmA and human GzmB suggests that probably their respective zymogens will also use this novel mechanism of zymogen stabilization. Upon activation, this salt bridge between Asp194-Lys188A has to be broken, but Asp194 does not have to undergo any rotation because it is already near the active conformation. Besides, the salt bridge between Asp194-Lys188A might prevent Asp194 from forming a salt bridge with Arg156, which could trigger generation of the catalytic machinery. In tPA the topologically equivalent residue Lys156 forms a salt bridge with Asp194, thus enabling activity of the enzyme without proteolysis of the Arg15-Ile16 peptide bond (Renatus *et al*., 97).

Another characteristic of the pro-GzmK structure is Gly215, which is in other serine proteinases one of the important elements of substrate binding site. The segment 214-226 is slightly rotated clock-wise, with the rotation axes being fixed by the Cys220 bridged to Cys191. The whole segment loops out to the east, narrowing the entrance to the active site triad.

Another uncommon feature of the pro-GzmK structure is the presence of a glutamate at position 219 (contrary to the majority of serine proteinases, where a Gly residue can be found),which adopts an unusual conformation. This characteristic is only

shared with the blood coagulation factor IXa where this residue causes incomplete active site formation (cryptic S1 pocket), even after zymogen activation, and hence the low activity of this enzyme towards chromogenic substrates (Brandstetter, 1995). In FIXa, this partially occluded specificity pocket might explain the several orders of magnitude reduction in catalytic efficiency (*Kcat*/*Km*) of FIXa towards chromogenic amide substrates and its poor binding capability for synthetic inhibitors compared, e.g., with FXa (Brandstetter *et al*., 1995, Bode *et al*., 1996). This destabilized state of FIXa is reminiscent of the mobile and occluded S1 pocket of the trypsinogen-like zymogens, which generally precludes binding of substrates and inhibitors can be forced into the active state by very high-affinity active site binding (Huber and Bode, 1978, Vijayalakschmi *et al*., 1994) (see Fig. 22 for the position of E219). In FIXa, E219 conformation results in a partially collapsed pocket, which is incompatible with normal substrate binding. Residue E219 in pro-GzmK also adopts the energetically high conformation found in FIXa. Possibly, GzmK displays likewise lower activities and therefore no special substrates have been described so far.

It is tempting to speculate that a similar situation occurs in GzmK. It is plausible that upon activation the GzmK active site may not be complete and ready to bind and cleave its specific substrates. Like in FIXa, there is possibly a cofactor or the substrate itself that upon binding will induce the conformational changes necessary to make GzmK capable of cleaving. This fact would suggest then a possible induced-fit mechanism for GzmK towards its substrate(s) may occur. To date, no physiological substrate for GzmK has been described, however.

The first 4 N-terminal residues, and the last C-terminal residues (after Leu243) are not defined by electron density indicating high flexibility. The 99-loop (segment 94-99) of pro-GzmK, which in most other trypsin-like proteinases forms subsites S2 and S4, is totally disordered as indicated by the lack of electron density in both crystal structures. The mobile 99-loop is a characteristic of GzmK zymogen as no other zymogen structure has shown such a flexibility in this region. On the contrary, loops that have been described to be flexible in zymogens are well defined in both pro-GzmK crystals.

Altogether, there are several structural reasons why the binding of a substrate to pro-GzmK is hindered. An unproductive approach of the scissile peptide bond may be

due to the non-productive conformation adopted by residues 189 (at the back of the specificity pocket), 215, 216 and 219 (the latter as in FIXa) located in the frame to the entrance of the specificity pocket. Furthermore, the catalytic His57 also adopts a non-functional conformation, as it is slightly rotated outwards from the pocked.

It was interesting to see that in fact most of the loops of pro-GzmK resemble more the active enzymes than previously analyzed zymogens. Furthermore, the crystal structure of human pro-GzmK suggests that granzyme zymogens might be more similar to active serine proteinases than to the known zymogens. Whether this is a typical feature of all granzyme zymogens has to be confirmed in future structures.

Interaction Between GzmK and Cathepsin C

In vivo, pro-GzmK wild type is converted to active GzmK via cathepsin C–catalyzed removal of the N-terminal Met-Glu dipeptide allowing the formation of the salt bridge between the new N-terminal Ile16 and Asp194. Once the pro-GzmK structure was solved, we intended also to study the mechanism of recognition between pro-GzmK and CatC. In other words, we wanted to find out whether there was any interaction pattern that would help to target CatC to pro-GzmK. After docking both proteins (D. Turk, personal communication), it was not possible to find a distinct electrostatic interaction match of the zymogen with its activator protein. It is likely that the interaction of pro-GzmK (and the other granzymes) with CatC is merely restricted to their N-terminal four residues and does not involve specific interaction in the surrounding surface of both interacting partners. The N-terminus is highly flexible in pro-GzmK and the other zymogens and extrudes into the solvent being easily accessible to CatC. Thus, the binding between GzmK and CatC is not dependent on the matching of complementary surfaces (i.e., polar charged residues interact with other polar/oppositely charged residues with neutralising partners, and non-polar residues are engaged in Van der Waals interactions with other non-polar residues).

Physiological Implications

Although the role of all granzymes is still unclear, GzmB clearly promotes rapid induction of apoptosis in target cells (virus-infected or tumoral cells) (Zapata, 1998). Granzymes are thus one of the most important tools by which cytotoxic lymphocytes eliminate threatening cells to our body. The reason why the innate immune system needs several granzymes is not clear, but they most probably ensure an adequate response to the broad spectrum of viruses and other intracellular pathogens. Several viruses code for various caspases and GzmB inhibitors as an escaping mechanism to the infection. Thus, it is possible that this apparent redundancy of granzymes will allow induction of apoptosis despite inhibition of caspases/GzmB.

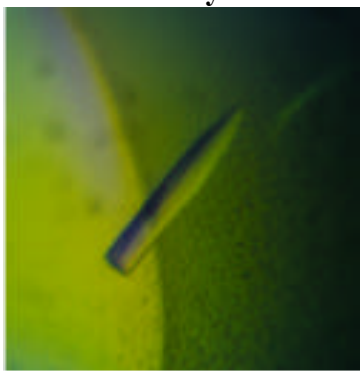
GzmK and GzmA share the same substrate specificity as they cleave peptide bonds after a lysine or arginine residue (P1 residue). However, it is possible that both Gzms might recognise distinct substrates *in vivo* as they do not share a common specificity for P2 and P3 sites. For instance, GzmA carries an Arg at position 99, while this residue is a glutamine in GzmK. The same explanation could be applied to mouse and rat GzmK, both differing from the human homologue at that position.

3.8) Material and Methods

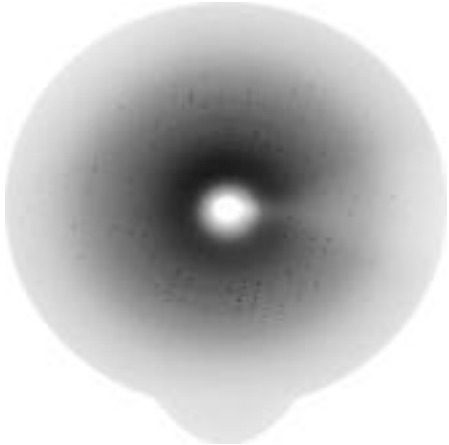
3.8.1) Crystallization of Human Pro-GzmK

Crystals suitable for diffraction analysis were obtained by sitting drop vapour-diffusion procedures. One µl of the protein solution (10 mg/ml) was mixed with 2 µl of a solution consisting of 3,9 M sodium format. Crystals grew within three months at 20°C to maximal dimensions of 0.3 x 0.2 x 0.5 mm^3 (see image below). **Fig.10**:

GzmK Crystal



Diffraction Pattern Cryo-data Set



3.8.2) Data Collection and Data Processing

A complete data set was recorded at room temperature on a 300 mm MAR Research image plate detector using monochromatized CuKα radiation from a RIGAKU rotating anode X-ray generator. The crystal was directly mounted from the drop in a thin-walled glass capillary, and diffracted to 2.9 Å. A second complete data set to 2.23 Å was collected under cryo-conditions. Crystals belong to the space group $P2_12_12_1$ and contain one molecule per symmetric unit. The two data sets were independently evaluated with MOSFLM (Leslie, 1991), merged, reduced and scaled (Table 1) using programs supported by the Collaborative Computational Project No. 4 (The CCP4 suite, 1994).

The structure was solved by molecular replacement with AMoRe (Navaza, 1994) using a modified Granzyme B search model, in which all non-identical residues between the two proteinases were changed to alanine. Rigid body refinement in AMoRe resulted in a R_{free}-factor of 49% for both crystals, for the range 15.0 to 2.9 Å and 15.0 to 2,2 Å. Model building was done on a SGI graphics workstation using MAIN (http://www-bmb.ijs.si/doc).

Calculation of the electron density maps and crystallographic refinement were performed with X-PLOR. Several cycles of model building, conjugate gradient minimization and simulated annealing using X-PLOR and CNS (Brünger, 1993) resulted in a structure with good stereochemistry. The target parameters of Engh and Huber (Engh and Huber, 1991) were used. Individual, highly restrained atomic B-values were refined, and no sigma cut-off was applied to the whole procedure. The main chain is in agreement with the sequence of human GzmK deposited in SwissProt database (entry P10144), starts with the N-terminal Met14 and is fully defined up to the residue 245. There is continuous electron density from Gly19 to Ser92 and Ser100 to Leu243.

The last four residues are not defined by electron density probably due to enhanced flexibility. The main chain around segment 94-99 in both crystals is placed in very poor electron density, and a few exposed side chains projecting into the solvent are not defined by electron density so that their occupancy have been set to zero. The final crystallographic R_{free}-Factor is 32% for all reflections from 15.0 to 2.9 Å and 29% for all reflections from 15.0 to 2.23 Å, respectively.

A Ramachandran plot calculated using the program PROCHECK (Laskowski et al., 1993) shows that most of the residues fall into the most favoured or additionally favoured regions. This is above average for both structures determined at 2.9 Å or 2.23Å.

3.9) Table 3. Statistics for Data Collection and Refinement

Zymogen GzmK S195A	**Cryo data set**	**RT data set**
Molecules/asymmetric unit	1	1
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell constants (_):		
a	31.46	31.75
b	78.16	78.86
c	83.62	85.28
Resolution (_)	2.23	2.90
Reflections measured	65669	39313
R_{merge} (%)	12.5	14.7
Unique reflections	23808	10537
Completeness (%), overall	97.1	94.5
Completeness (%), outermost shell	95.1	92.5
Reflections used for refinement	22617	9771
Resolution range (_)	22 - 2.23	22 - 2.90
R-value (%) initial	50.1	52.0
R_{free} (%) initial	49.2	49.4
R-value (%) final	23.24	23.48
R_{free} (%) final	29.73	31.39
r.m.s standard deviations: Bond length (_) Bond angles (°) Average B-value/SD (_)	 0.008 1.45 23.6	 0.007 1.38 26.8

[a] $R_{merge}=\sum_{hkl}| I - I|/\sum_{hkl}|I|$

[b] $R_{factor}=\sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$

[c] R_{free} is the R-value calculated with 500 reflections that were not used for the refinement.

Article 4.

Crystal Structure of a Novel Midgut Procarboxypeptidase from the Cotton Pest *Helicoverpa armigera*

4.1) Introduction

Helicoverpa armigera (Lepidoptera: *Noctuidae*), also known as bollworm or corn earworm, has a widespread distribution in tropical, subtropical and warm temperature regions. *H. armigera* is a polyphagous pest of 181 plant species, including cotton (*Gossypium* species) and maize (*Zea mays*), and is predicted to become an important pest in other crops such as tobacco (*Nicotiana tabacum*) (Harsulkar *et al*., 1999). It is one of the most serious insect pests in important cotton producing countries like Australia, India and China. The economic problems caused by this voracious insect are enormous, as exemplified by the crisis in the cotton production in China after the occurrence of mass outbreaks of *H. armigera* in 1992 (Shen *et al*., 1993).

Until recently, studies of the protein digestion in the insect gut have mainly concentrated on the initial phases of protein and peptide degradation through endopeptidase (trypsin and chymotrypsin-like proteases) activities (Terra *et al*, 1994). However, the products of such endopeptidase cleavage are large peptides, which must be further degraded to be taken up by the insect gut cells (Billingley *et al*., 1990). Consequently, exopeptidases such as carboxypeptidases are also significant factors in the overall process of protein digestion.

Recently, a novel procarboxypeptidase from *H. armigera* (PCPAHa), the first enzyme of this class from a lepidopteran insect, has been characterised by expressing its encoding cDNA in insect cells (Bown et al, 1998). The pre-proprotease sequence contains 426 amino acid residues and exhibits sequence homology with the metallo-carboxypeptidases from mammalian species, and with carboxypeptidases from other invertebrates. Removal of a predicted signal peptide and of the activation peptide results in a product of approximately 35.5 kDa, which is comparable in size to the pancreatic mammalian carboxypeptidases (Vendrell *et al.,* 2000). PCPAHa belongs to the A form of carboxypeptidases and preferentially cleaves aliphatic and aromatic residues. This

carboxypeptidase is active over a broad pH range (7.5 – 10), displaying maximal activity at pH 8.0. This is in agreement with the report of a maximal peptidase activity in the alkaline region determined in other phytophagous lepidopteran species (Lenz *et al*., 1991; Ortego *et al*., 1996; Ferriera *et al*., 1994).

We have determined the three-dimensional structure of this novel procarboxypeptidase from *H. armigera*. This is the second structure reported for an insect proteinase (Botos *et al*., 2000) and the first for an insect carboxypeptidase. The structure of PCPAHa resembles the mammalian pancreatic procarboxypeptidases, containing a long pro-segment that inhibits the active enzyme without having any direct interaction with the zinc and the active site region.

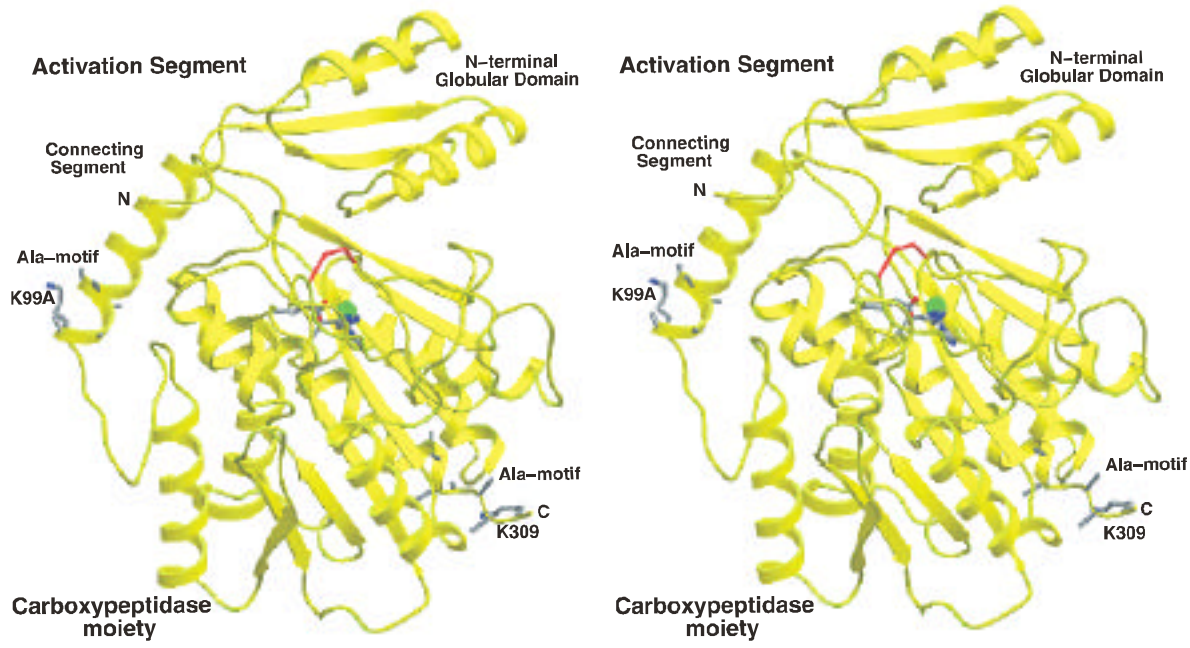
4.2) Results and Discussion

4.2.1) Overall Structure

Like other mammalian procarboxypeptidases (Vendrell e*t al*., 2000), the *H. armigera* procarboxypeptidase (PCPAHa) has a globular shape with two independent domains: the large pro-segment or activation segment, consisting of 91 residues, and the active enzyme moiety with 318 residues (Fig. 1). Upon proteolytic cleavage of the zymogen, the interaction of the pro-segment with the enzyme moiety is weakened allowing the release of the pro-segment. The possible presence of more than one activation site would be in agreement with the connecting loop (Fig. 1), being considerably long, lacking any regular structure and containing three basic residues (see also Fig. 2B). One particular feature of PCPAHa when compared to the related pancreatic PCPs (PCPA2, PCPB and PCPA1) (García-Sáez *et al*., 1997, Coll *et al*., 1991, Guasch *et al*., 1992) is the slight rotation shift of 2-3° of the whole pro-segment relative to the carboxypeptidase moiety (see Fig. 2A). This shift does not seem to compromise the inhibitory capacity of the pro-segment, as the proenzyme has no detectable intrinsic activity (Bayés *et al*., unpublished).

4.2.2) Fig. 1. Three-dimensional Structure of PCPAHa.

Stereo ribbon plot representation of PCPAHa. The pro-segment is folded into a globular N-terminal domain consisting of an antiparallel α/ antiparallel β open-sandwich, and a α-helical connecting segment. The catalytic moiety is formed by a central twisted eight stranded β-sheet flanked by eight α-helices, together forming a globular α/β protein. The zinc ion (green sphere) is coordinated in the active site by two histidines and one glutamic acid. The single disulfide bridge is shown in red colour. Lys99A, Lys309 and the Ala-motifs are shown as grey stick models.



4.2.3) Activation Segment of PCPAHa

The activation segment of PCPAHa can be described as composed of two different regions (Fig. 1): the large N-terminal globular domain which shields the pre-formed wide active-site depression of the enzyme, and the connecting segment consisting of a long α-helix (α3) that links this globular domain with the carboxypeptidase moiety (Coll *et al.*, 1991). The globular N-terminal domain (residues His4A to Val82A) exhibits an open-sandwich antiparallel-α /antiparallel-β fold, made up by two α-helices and four β-strands arranged with a β1α1β2β3α2β4 topology (Fig. 1). Despite the relatively low

sequence identity of the pro-segment of PCPAHa to those of mammalian carboxypeptidases (around 20%), the overall folding topology is conserved (Fig. 2). This structural similarity is higher to the A forms, which lack the 3_{10} helix insertion (43A-46A) that is typical in the B forms. The C-terminal α-helical connecting segment comprising residues Lys83A to Lys99A is followed by a loop that links the pro-peptide with the CP moiety. Noteworthy is the presence of five consecutive Ala residues followed by an exposed Lys at the end of the connecting segment (Fig. 1). This endoprotease cleavage motif that forms the last turn of the α3-helix has never been described before for procarboxypeptidases, and its role is still unclear. Due to its four turns, this α3-helix resembles the connecting helices of PCPA forms, where four and five-turn helices are found (PCPA1 and PCPA2) compared to the two-turn α3-helix observed in PCPB. Due to the sequence homology and from previous biochemical studies (Bayés *et al.,* unpublished), it is known that the *in vitro* activation by Lys-C endoproteinase occurs at Lys99A, located at the end of the activation segment (Figs. 1 and 2). Although the activation with Lys-C endoproteinase is more efficient, the proenzyme can also *in vitro* be activated by bovine trypsin at Arg4, located just five residues behind Lys99A. On the other hand, the sequence of the carboxypeptidase purified from larval gut extracts is LSFDKIHSYEEVDAYLQELAKEFPNVVTVV, which coincides with residues numbered 7-36 in figure 2B.

4.2.4) Carboxypeptidase Moiety of PCPAHa

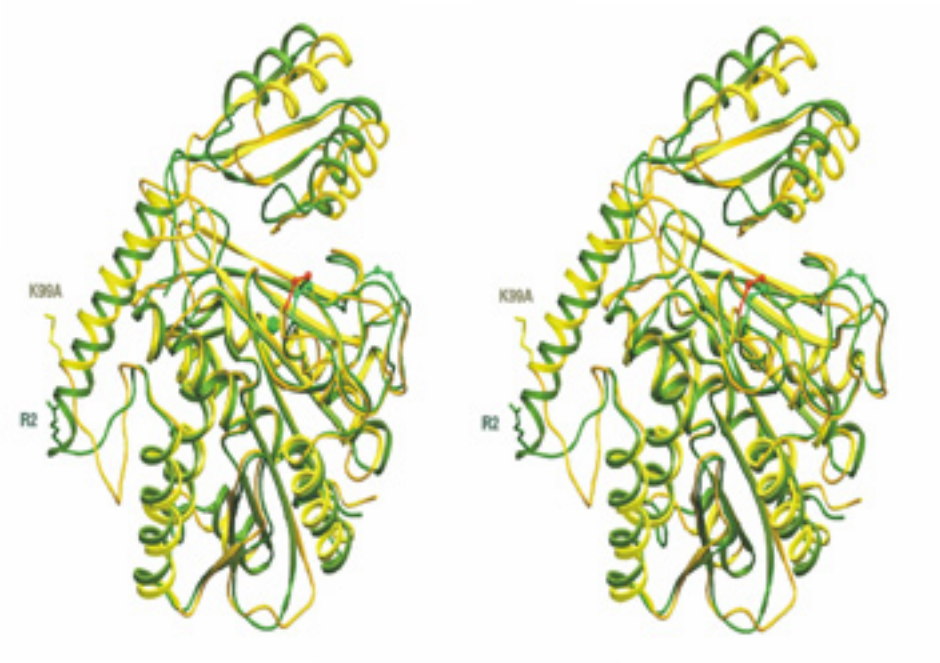
The PCPAHa catalytic moiety exhibits the same fold as the other known carboxypeptidase structures (Coll *et al*., 1991; Guasch *et al*., 1992; Gomis-Rüth *et al*., 1995; García-Sáez *et al*., 1997): a central twisted eight-stranded β-sheet (β5-β12) flanked by eight α-helices (α4-α11), which together form a globular α/β protein (Fig. 1). Compared to the structure of the search model (human PCPA2), several insertions and deletions located in the loops that connect the secondary structure elements are required to align both enzymes (Fig 2A), namely: a three residue insertion in the loop between β6 and β7, a four residue deletion between α5 and α6, a three residue insertion between α9-β11 and a final three residue insertion after β12. Another $(Ala)_6Lys$ motif, similar to that found in the connecting segment (see above), is located at the end of the last α-helix

(α11) which contains six consecutive Ala residues compared to the five described for the activation segment (Figs. 1 and 2B).

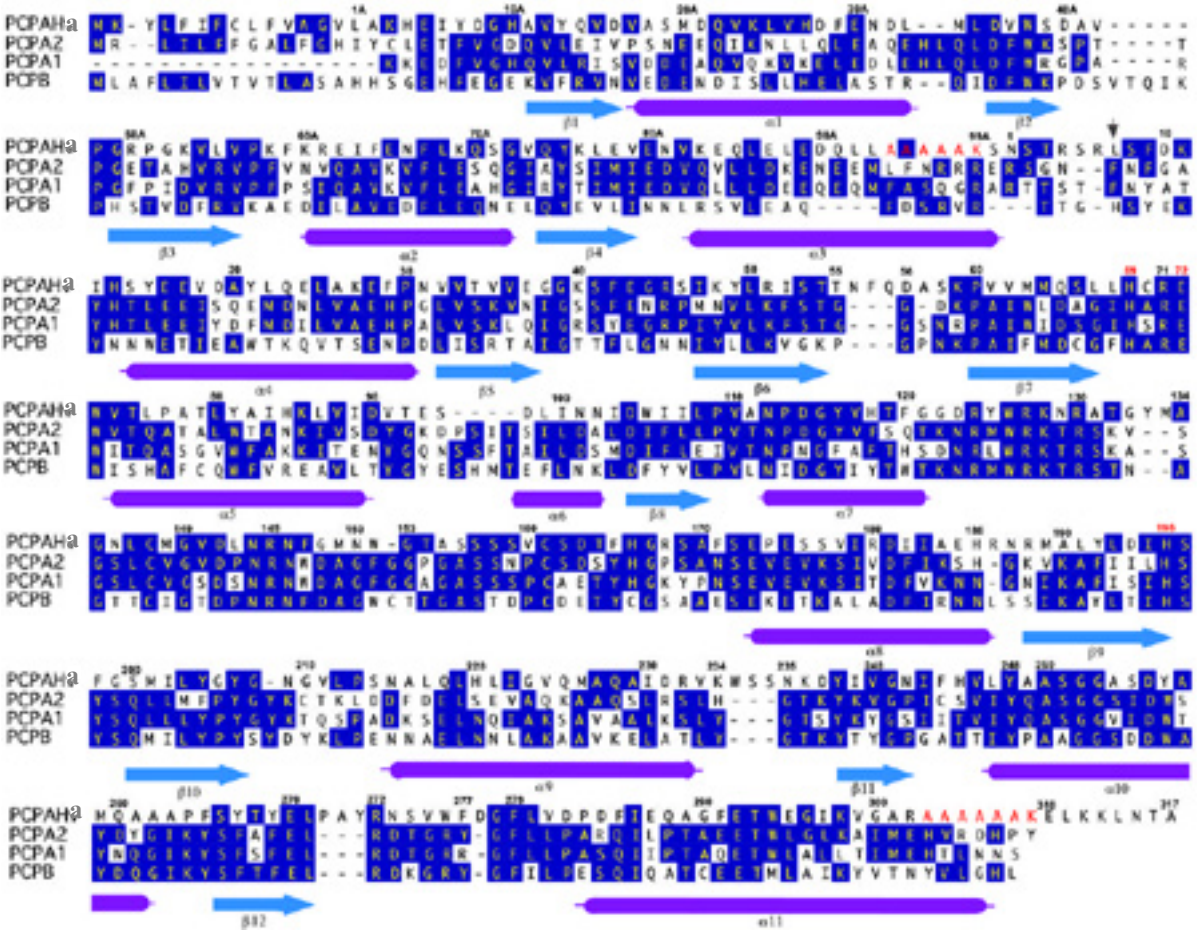
PCPAHa possesses only one disulfide bond (Cys138 – Cys161), being the only conserved in all known procarboxypeptidases. However, this protease has an additional buried free-cysteine (Cys70), following the His69 from the zinc binding motif, replacing the Ala70 currently found in the other enzymes (Figs. 2B and 3). No catalytic role has been ascribed to this cysteine so far, despite its proximity to the active site residues.

4.2.5) Fig. 2. Topological and Sequence Comparison of PCPAHa and Human PCPA2.

A) Stereo ribbon plot superimposition of PCPAHa (yellow) and human PCPA2 (green). The main differences between both enzymes are located in the loops between the secondary structure elements. A slight shift is observed in the activation segment of PCPAHa compared to human PCPA2. The figure was prepared with Setor (Evans, 1990).

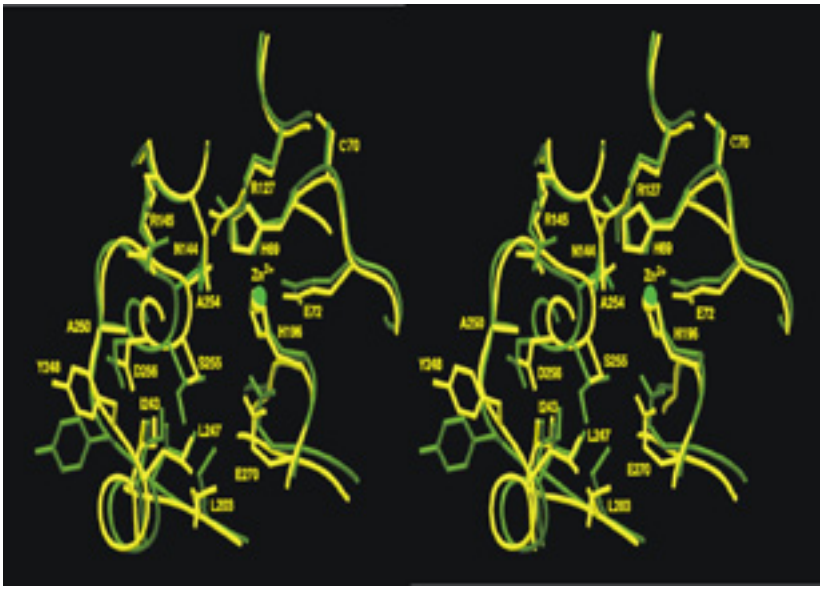


B) Structure-based Amino Acid Sequence Alignment Between *H. armigera* PCPAHa, Human PCPA2, Porcine PCPA1 and Porcine PCPB. The numbering of both moieties of the proenzyme is based on three-dimensional topological alignments with the other known procarboxypeptidase structures (Coll et al., 1991; Guasch et al., 1992, García-Sáez et al., 1997). The numbers of the amino acid positions in the activation segment are followed by an "A" to distinguish them from positions in the enzyme moiety. An arrow on Leu7 indicates the N-terminus of the mature enzyme isolated from larval extracts. The β-strands (β1 to β12) and α-helices (α1 to α11) are shown by arrows and cylinders, respectively. The alignment was prepared with the DNASTAR software package (Madison, USA).



The active site cleft of the PCPAHa is placed at the C-edge of the central part of the β-sheet of the catalytic moiety. The active site is composed of a zinc atom and the different active subsites (S1', S1, S2, S3 and S4, nomenclature according to Schechter & Berger, 1967). In PCPAHa, the typical Zn^{2+} coordination also described for the other CPs is found (Fig. 3): the zinc ion is pentahedrally coordinated by His69 Nδ1, His196 Nδ1, both Glu72 Oε1 and Oε2 oxygen atoms, and the "catalytic" water molecule, which in addition is linked to the general base Glu270 (distance 3.7 Å). As in the other CP structures, the correct arrangement of His196, one of the catalytic Zn^{2+} ligands, requires the formation of a characteristic *cis*-peptide bond between Ser197 and Phe198. In the absence of a substrate, the substrate binding site is apparently empty in the crystal structure of PCPAHa, but contains several disordered solvent molecules. The essential residues for catalysis have been identified on the basis of previous studies made with CPs and their zymogens (Christianson & Lipscomb, 1989; Kim & Limpscomb, 1990; Coll *et al*., 1991; García-Sáez *et al*., 1997; Avilés *et al*., 2000).

The S1' subsite, which has a dual positive-charged/hydrophobic nature to anchor and neutralize the scissile C-terminal residue of the substrate, is lined by residues Leu203-Gly207, Ile243-Asp256 and Tyr267-Thr268 (Figs. 3 and 4). The hydrophobic character of the pocket is mainly provided by Ile243, Ala250 and Leu247, these. residues are also conserved in the other CPA. However, the Ser255 at the bottom of the pocket replacing the aspartic acid or isoleucine in CPB and CPA, renders this S1' pocket larger as well as more polar. The residues Arg145 and Asn144, which extend into this pocket, anchor the negative C-terminal carboxylate of the substrate. The nearby Tyr248 is found in an "up" conformation in the PCPAHa structure, also seen in the empty active sites of the PCP structures (Coll *et al*., 1991; Garcia-Saez *et al*., 1997; Gomis-Rüth *et al*., 1995), interacting with the side-chains of the residues Val23A and His27A of the pro-segment (see below). The adjacent S1 subsite formed by Arg127 and Glu270 extends, via the S2 and S3 subsites, into the bulk water.



4.2.6) Fig. 3. Section Around the S1' Pocket of PCPAHa (yellow) Compared to PCPA2 (green). The broader substrate specificity of PCPAHa can be explained by the substitution of Ile255 by a Ser at the base of the substrate-binding pocket. This residue change renders the pocket larger and less hydrophobic. The zinc ion is shown as a green sphere. The figure was prepared with Setor (Evans, 1990).

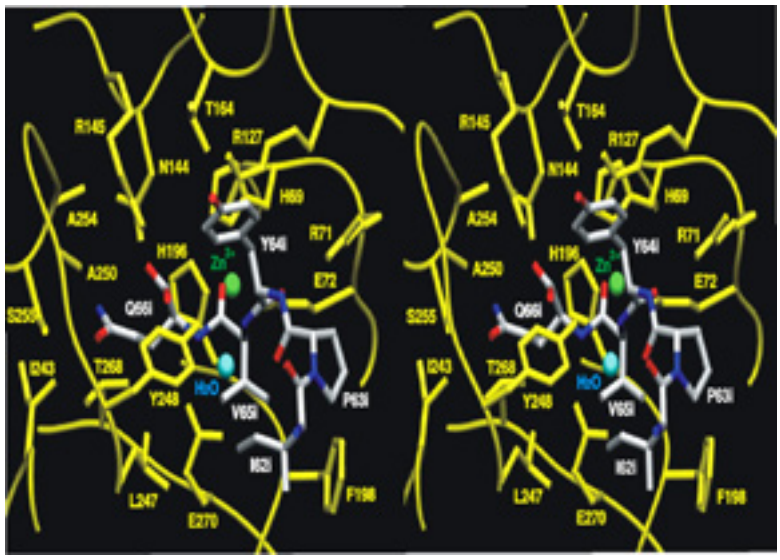
4.2.7) Probable Interaction with Peptidic Substrates

To explore the probable binding geometry of a peptide substrate and the potential enzymatic mechanism of CPAHa, the C-terminal tail Ile62-Pro63-Tyr64-Val65↓Glu66 of LCI (Leech Carboxypeptidase Inhibitor) has been specifically modelled into the active site of PCPAHa based on the structure of LCI in complex with human CPA2 (Reverter *et al.*, 2000). In this structure, the last five C-terminal residues of LCI represent the "primary" interaction region believed to interact with the zinc and the subsites S1', S1, S2 and S3 in a substrate-like manner. The last residue of the inhibitor (Glu66i) is cleaved off by the carboxypeptidase but remains trapped in the S1' cavity. In the PCPAHa model the Glu66i has been substituted by Gln to remove its charge but allowing polar interaction with the lining residues of the pocket. The model was submitted to several

rounds of energy minimisation with XPLOR (Brünger, 1993) in order to avoid clashes and find the most favourable conformation.

The current model of the enzyme complex formed with the LCI C-terminal peptide mimics the optimal geometrical conformation and interactions of the substrate in the active site in the pre-transition state (Fig. 4). The most accepted model of peptide hydrolysis is the so-called water-promoted mechanism (Kim & Lipscomb, 1990). In such a model, the catalytic zinc ion together with the other electrophil, the guanidyl group of the Arg127, polarise the carbonyl group of the scissile peptide bond, and simultaneously, the catalytic water molecule is rendered to a potent nucleophile by donation of one proton to the catalytic Glu270. The resulting hydroxyl can attack the carbonyl carbon of the scissile peptide bond of the substrate.

4.2.8) Figure 4. Probable Interaction of the Ile-Pro-Tyr-Val↓Gln C-terminal Pentapeptide of LCI with the Active Site of PCPAHa. The residues of the active site of PCPAHa are shown in yellow. The modelled C-terminal segment of LCI is shown in a charged-coloured stick representation. The "catalytic" water and zinc atom are shown as a blue and a green spheres, respectively. The figure has been prepared with Setor (Evans, 1990).



In the S1' subsite the accommodation of the C-terminal Gln66i produces not only a slight shift of the Arg145 and Asn144 side-chains, but also a dramatic (and well-documented) conformational change of Tyr248 from the "up" (Fig. 3) to the "down" position (Fig. 4). Arg145 and Asn144 side-chains are at interaction distances to the carboxylate oxygen of Gln66i, and the hydroxyl side-chain oxygen of Tyr248 interacts with the amino group of Gln66i. Such a "down" conformation of the Tyr248 has been frequently reported in CP and PCP structures containing a substrate/inhibitor inside of the active site (Christianson *et al*., 1980; Garcia-Saez *et al*., 1997; Reverter *et al*., 2000). The side-chain of Gln66i fits perfectly to the polar environment created at the bottom of the S1' pocket by Ser255 and Thr268 (Fig. 5).

It is worth mentioning that whilst some discrepancies appeared recently in the literature for the structures of active carboxypeptidases with respect to the "up" and "down" positions of Tyr248 (Bukrinsky *et al.,* 1998; Aalten *et al*., 2000), the view is much more homogeneous in the case of the zymogen structures. Thus, it has been reported the "up" conformation (interacting with residues of the pro-segment) in porcine PCPB and human PCPA2 (Coll *et al*., 1991; Garcia-Saez *et al*., 1997). Only in the case of porcine PCPA1 -with a free valine bound to the S1' subsite- (Guasch *et al*., 1992) and in the benzylsuccinic-complexed PCPA2 (Garcia-Saez *et al*., 1997), the Tyr248 is found in the "down" conformation.

In the S1 subsite the carbonyl group of the P1-Val65i is directed just to the side of the zinc ion, with its oxygen also located close to the guanidinium group of Arg127. The "catalytic" water is sandwiched between the carbonyl carbon of Val65i and the Glu270 carboxylate (Fig. 4). The geometry of the modelled substrate presents a good scenario for the nucleophilic attack of the "catalytic" water on the scissile peptide bond of the substrate (Val65i-Gln66i). The Arg127 and the zinc atom are optimally placed to stabilize the negative charges formed during the transition state. The amino group of Val65i is also in contact with the hydroxyl oxygen of the Tyr248 (as in the S1' subsite), and its side-chain is located inside a small hydrophobic pocket created by Phe198, Phe279 and Leu247.

The major interaction in the S2 subsite is the hydrogen bond between the carbonyl oxygen of the P2-Tyr64i and the guanidinium group of Arg71. This carbonyl group also

interacts with the guanidinium group of Arg127, and the phenyl ring of Phe279, which shapes the S1 and S3 subsites, respectively. The side-chain of Tyr64i is in contact with Arg127, Arg145 and Thr164. In the S3 subsite, the only interactions observed are the contact of the methylene carbons Cγ and Cδ of the P3-Pro63i with the phenyl rings of Phe198 and Phe279.

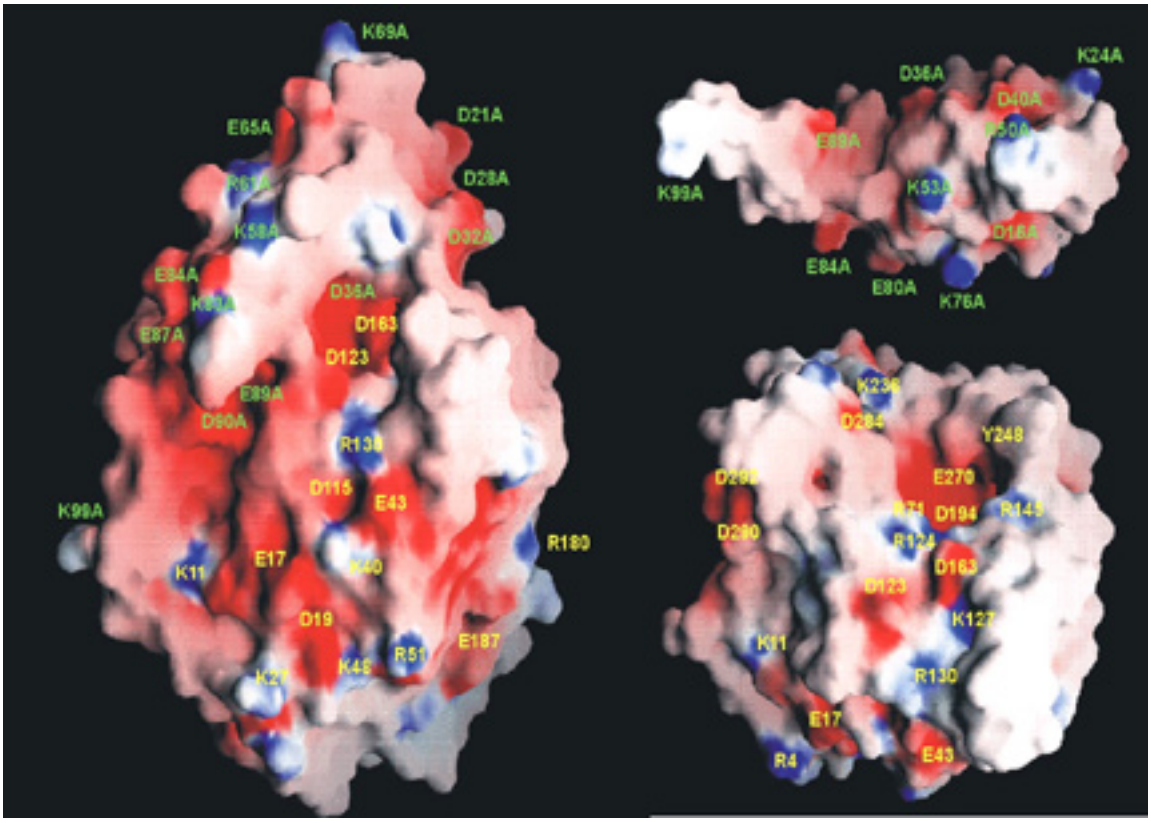
4.2.9) Figure 5 (next page). Solid Surface Representation of PCPAHa.

A. The PCPAHa molecule is shown in an identical orientation as in Figs. 1 and 2, with the exposed Lys99A at the end of the connecting segment being clearly visible. PCPAHa has many acidic residues as indicated by the mainly red colouring of the surface, with colours indicating positive (blue) and negative (red) electrostatic surface potential contoured from +15 kT/e to 15 kT/e.

B. Surface of the pro-segment docking the carboxypeptidase moiety.

C. Surface of the carboxypeptidase moiety interacting with the pro-segment.

Both molecules described in B and C have moved apart and rotated by 90° in opposite directions from each other around the horizontal axis to show the respective interacting surfaces. The salt bridges between residues Asp36A, Lys53A and Glu89A from the activation segment and residues Arg71, Asp284 and Arg124 from the carboxypeptidase moiety, respectively, are observable. The figure was prepared with GRASP (Nicholls *et al*., 1993).



4.2.10) Interaction of the Pro-region with the Enzyme: the Inhibition Mechanism

The inhibition of PCPAHa is accomplished through blockage of the wide active-site depression by the globular part of the activation segment (Figs. 1 & 5). The activation segment occludes the S2, S3 and S4 subsites of the enzyme, preventing substrates from entering into the active site groove of the enzyme. The main interactions in this region are the salt bridge between the side-chain groups of Asp36A with Arg71 (around the S2 subsite), the hydrophobic interactions between the indole group of Trp37A with the phenyl groups of Phe198 and Phe279 (around the S2 and S3 subsites), and the interaction of the His24A and Val20A side-chains with the side chain of Tyr248 (belonging to the S1' subsite, but in the "up" conformation). There is, however, no direct contact of the pro-segment with the Zn^{2+} atom or with the residues involved in the formation of the S1' and the S1 subsites. This inhibition mechanism differs from the substrate-like inhibition mechanism of the CP protein inhibitors LCI and PCI (derived from leech and potato, respectively), where a direct contact with the catalytic residues (S1' and S1 subsites) and the Zn^{2+} atom of the enzyme occurs (Rees & Lipscomb, 1980,

Reverter *et al*., 2000). In figure 5 the electrostatic interactions between the pro-segment and the carboxypeptidase moiety are shown. In contrast to the structure of porcine PCPB (Coll *et al*., 1991) but in agreement with the PCPAs (Guasch *et al*., 1991; García-Sáez *et al.*, 1997), PCPAHa does not possess any salt bridge between the active-site Arg145 and any acidic residue from the pro-segment. But in the case of PCPAHa, no intrinsic activity is detected in contrast to PCPA1 and PCPA2.

Both the globular activation domain and the connecting segment (helix α3) contribute differently to the inhibition of the carboxypeptidase moiety. The interaction surfaces of the globular domain (residues His4A – Val82A) and the connecting segment (Lys83A – Lys99A) with the carboxypeptidase moiety are approximately 730 $Å^2$ and 515 $Å^2$, respectively, which lie within the range of other PCPs structures (Vendrell *et al.*, 2000). Combination of structural information and knowledge about the activation processes of mammalian pancreatic A1, A2 and B forms (Vendrell *et al*., 1990; Reverter *et al*., 1998; Ventura *et al*., 1999), have helped us to understand the contribution of both parts of the pro-segment in the inhibition of the enzyme.

4.3) Conclusions

In this paper we present the three-dimensional structure of a procarboxypeptidase from *Helicoverpa armigera*, a devastating pest of important crop plants. The structure shows the preservation of the overall fold of exopeptidases throughout the animal kingdom. Thus, the characteristic zinc procarboxypeptidase fold, the catalytic mechanism and the inhibition exerted by the pro-segment seem to be universal. However, despite these generals scaffold similarity, structural differences can be found in the loops between the conserved secondary structures, including the loop where the activation processing occurs. The presence of the amino acid region with the sequence $(Ala)_nLys$ located where the proteolytic activation cleavage of the enzyme is likely to occur, is a remarkable feature of this procarboxypeptidase. The fact that this sequence not only appears at the activation-site of the protease but also near the C-terminus, suggests that post-translational protein processing, different from the activation process might occur. We have observed that one of them, from Ala94 to Lys99, is the *in vitro* activation point for a Lys-specific endoproteinase (Bayés *et al.*, unpublished). Most probably the removal

of an octapeptide by the cleavage after the lysine in the second $(Ala)_nLys$ sequence (from Ala303 to Lys309) is a process catalysed *in vivo* by the same enzyme responsible for activation of PCPAHa. Although we still do not have *in vivo* evidence for this cleavage, structural data support this possibility since such a sequence is exposed and thus accessible to this second proteolytic attack.

Trypsin can also *in vitro* activate the pro-enzyme at Arg4, located just five residues behind Lys99A, but this activation is not as efficient as in the case of LysC endopeptidase. The stability and proteolytic activities of the LysC- and trypsin-cleaved CPAHa forms in front of the synthetic substrate N-(3-(2-furyl)acryloyl)-Phe-Phe (FAPP) is similar, at 25 °C. The purification of a carboxypeptidase enzyme in *H. armigera* gut extracts with an N-terminal sequence starting at Leu7, as a result of cleavage at Arg6, confirms that trypsin-like enzymes can sever the unstructured carboxypeptidase N-terminus. Our hypothesis about the *in vivo* activation process would involve a controlled cleavage at the specific $(Ala)_5$-Lys site by the natural activator, followed by the subsequently trimming of the unstructured N-terminus by a trypsin-like action up to the sequence found in the carboxypeptidase purified from the larva extracts.

The modeled structure with the reconstructed C-terminal tail of LCI has revealed the optimal geometrical conformation of the residues involved in the substrate binding subsites of the enzyme. The model provides a good scenario for the water-promoted catalytic mechanism attained before the transition state occurs, including the interaction between the catalytic residues Glu270 and Arg127 with the catalytic water, the zinc atom and the scissile peptide bond of the C-terminal residue of the substrate. The presence of Ser255 in the S1' pocket enlarges the classical aliphatic and aromatic substrate preference found in the A forms. Only the carboxypeptidase sequence from the earthworm *Lumbricus rubellus* possesses a similar residue in its hydrophobic pocket (Accession number Y09625), but no functional information is available in this case. Functional studies performed with recombinant PCPAHa confirm this broad substrate preference for this enzyme, including polar residues (Bayés *et al*., unpublished results).

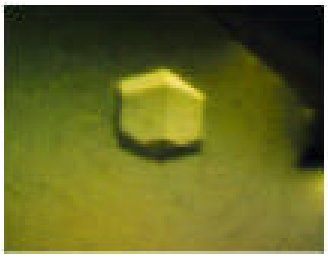
The quests for an effective, safe and lasting pest management program should be the primary target in the development of new products with a minimal undesirable impact on the environment, which could replace conventional toxic pesticides. The development

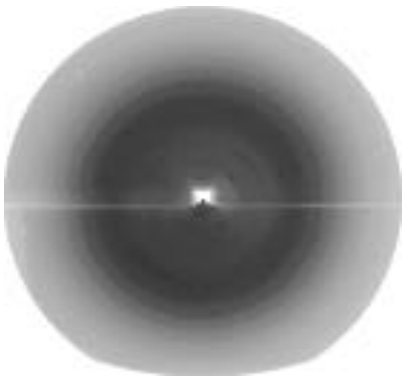
of pesticide resistance in insects causes the need to inhibit more specific enzymatic targets. All attempts to eliminate *H. armigera* pest by conventional means have been unsuccessful to date due the build up of resistance to pesticides (Fitt, 1989). Knowledge of the regulation of digestives proteases at a biochemical and structural level will help to develop new tools to combat this pest.

4.4) Materials and Methods

4.4.1) Crystallization and Structure Determination

A single small crystal grew after two months in 30% PEG 8000, 0.2 M Na-Acetate, 0.1 M Na-cacodylate (pH 5.5) by mixing 2 µl of the protein sample with 2 µl of the crystallization buffer. Using micro and macro-seeding techniques, larger crystals could be produced after a few days, which reached a maximum size (0.4 x 0.3 x 0.5 mm^3) in approximately one week. The largest crystal was harvested using the reservoir solution and mounted in a thin-walled glass capillary tube. A complete data set up to 2.5 Å was collected at room temperature in-house using a 300 mm Mar Research image plate detector. The crystal belonged to space group $P2_1$ with cell constants a=48.03 Å, b=86.39 Å, c=50.55 Å and β=100.70°, and contained one molecule per asymmetric unit. Diffraction data (see Table 1) was evaluated, scaled and reduced with DENZO-SCALEPACK (Otwinowski, 1993). The PCPAHa structure was solved by molecular replacement with AMoRe (Navaza, 1994), using the structure of human procarboxypeptidase A2 (PDB code 1AYE) as a search model.





Diffraction data from 12.0 Å to 3.5 Å was used to calculate the rotation and translation functions. The correlation factor and the R-factor of the best solution after rigid body refinement was 25.5% and 48.5%, respectively. Model building was done on an SGI graphics workstation using Turbo-FRODO (Roussel & Cambillau, 1989). Calculation of the electron density maps and crystallographic refinement were performed with X-PLOR (Brünger, 1993) and REFMAC/CCP4 program (Collaborative Computational Project No. 4, 1994). After several cycles of model building, conjugate gradient minimisation and simulated annealing, a structure with good stereochemistry was obtained. The target parameters of Engh and Huber (Engh & Huber, 1991) were used. A Ramachandran plot calculated using the program PROCHECK (Laskowski *et al*., 1993) shows that all the residues fall into the most favoured (86.6%) or additionally favoured regions (13.4%).

The numbering system for the activation segments of procarboxypeptidases is derived from the alignment of all pancreatic-like procarboxypeptidases as described in García-Sáez *et al.* (1997). The numbers of the amino acid positions at the pro-segment are followed by an "A" to distinguish them from positions in the enzyme moiety. Following this system, the activation segment of PCPAHa spans from His4A to Lys99A even though the real length of the activation segment is 91 residues (see Fig. 1 for deletions). The numbering of bovine carboxypeptidase A1 is kept for the enzyme moiety.

4.5) Data Bank Accession Number.

The coordinates of the *H. armigera* procarboxypeptidase A (PCPAHa) have been deposited with the PDB ID code 1JQG.

4.6) Acknowledgements

Thanks to Carlos Fernández-Catalán, Pablo Fuentes-Prior and John Richardson for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 413 and SFB 469) and by the E.C. Biotechnology Programme (DGXII-E/BH/009294). D. Reverter has been a recipient of a fellowship from the European Community Biotechnology Marie Curie program. E. Estébanez acknowledges the EU "Training and Mobility" programme financial support (COSSAC, FMRX-CT-98-0193).

4.7) Table 1. X-ray Data and Refinement Statistics.

Crystal cell constants	a=48.032 Å, b=86.395 Å, c= 50.550 Å, β=110.7°
Space group	$P2_1$
Limiting resolution (Å)	2.5 Å
Total number of reflections measured	230224
Unique reflections (12.0 – 2.5 Å)	21449
R_{merge} [1], (12.0 – 2.5 Å)	9.2 %
R_{merge} [1], (2.6 – 2.5 Å)	34.4 %
Completeness (%), (12.0 – 2.5Å)	93.2
Completeness (%), (2.6 – 2.5 Å)	93.0
Solvent molecules	157
Zinc ions	1
Reflections used for refinement	18816
Sigma cut-off for refinement	2.0
Resolution range used	12.0 – 2.5 Å
R-value [2], overall	18.01 %
R_{free} [3]	23.81 %
R.m.s standard deviations:	
Bond length	0.011 Å
Bond angles	1.52°

[1] $R_{merge} = [\_{h\_i} | I(h,i) - < I(h) > | / \_{h\_i} | I(h,i)] \times 100$, where I(h,i) is the intensity value of the ith measurement of h and < I(h) > is the corresponding mean value of h for all i measurements of h. The summation is over all measurements.

[2] R-value = $(\_ |F_o-F_c| / \_ F_o) \times 100$.

[3] R_{free} was calculated randomly omitting 7% of the observed reflections from refinement and R-factor calculation.