Introduction

Part 1. Principles of Protein Crystallography

1.1. Protein Crystallography

Protein crystallography is a relatively young science and together with nuclear magnetic resonance (NMR) spectroscopy the main method for the elucidation of three-dimensional macromolecular structures. The use of crystals in structural biology goes back to 1934, when Bernal and Crowfoot (D. Hodgkin) produced the first X-ray diffraction pattern of a protein, that of crystalline pepsin (Bernal and Crowfoot, 1934). Their observation that pepsin crystals gave an X-ray diffraction pattern opened the subject of protein crystallography. Although crystallography, when compared to NMR, gives a more static description of the macromolecular structures, there are no limits in the size of the molecule to be analyzed. This makes X-ray crystallography the method of choice for studying macromolecular complexes at an atomic level.

In X-ray crystal structure analyses the problem is to find not only the amplitudes of all the diffracted X-rays (usually known as the reflections), but also their phases. Knowledge of both the amplitudes and phases allows to reconstitute the electron density of the crystal that gave rise to the diffracted beams. The amplitudes can be deduced from the intensities of the diffracted X-rays but the phases cannot be directly measured. This is known as the "**phase problem**" (see below). For protein crystals this is a particularly difficult problem, and even measurement of X-ray intensities is not straightforward (Blundell,1976).

To determine a protein three-dimensional structure one has to first obtain good diffracting crystals of the protein. To obtain an homogeneous protein sample is crucial for this purpose; crystallizing a protein is mainly a trial-and-error process.

1.2) Crystallization Techniques

Growth of high quality single crystals is the basis of X-ray structure determination and also its limiting step. Protein purity and homogeneity are essential for the growth of single protein crystals. Crystallization of macromolecules is a multi-parametric process involving three main steps: **nucleation**, **growth** and **cessation of growth**. There are many methods to crystallize proteins, all of which aim at bringing the solution of the protein to a **super-saturation state**, which will force the macromolecules into the solid state, the crystal. In the 1960´s the crystallization micro-methods (i.e. dialysis and vapor-phase diffusion) were developed. Vapor phase equilibrium and dialysis techniques are the two methods most used by crystallographers and biochemists; batch and interface diffusion methods are also employed. In the present work only the vapor-phase diffusion techniques will be discussed as it was the method chosen for crystallizing all the proteins presented in this work.

1.2.1) Crystallization by Vapour Diffusion Methods

Among the crystallization micro-methods, vapor diffusion techniques are the most widely used. A micro-drop of "**sitting**" or "**hanging**" protein solution is mixed with buffer, crystallizing agent, and eventually also additives, and equilibrated against a reservoir containing a solution of crystallizing agent at a higher concentration than the droplet (Fig. 1). Equilibration proceeds by diffusion of the volatile species (water or organic solvent) until vapor pressure in the droplet equals the one of the reservoir. The advantages of the sitting drop technique include speed and simplicity. The disadvantages are that crystals can sometimes adhere to the well surface making removal difficult. But sometimes this disadvantage turns into an advantage as this surface can assist in nucleation.

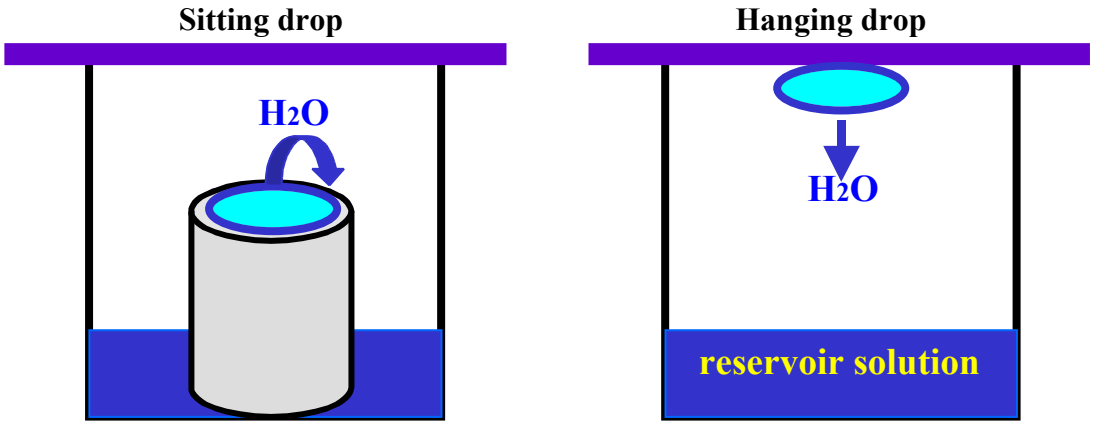


Fig.1. (previous page) Process of Vapour Diffusion.
The initial precipitant concentration in the droplet is less than in the reservoir. Over time the reservoir will pull water from the droplet in a vapor phase until the surface tensions of drop and reservoir are equilibrated. During this equilibration process the sample is also concentrated, increasing the relative supersaturation of the sample in the drop. Equilibration is reached when the reagent in the drop is approximately the same as in the reservoir.

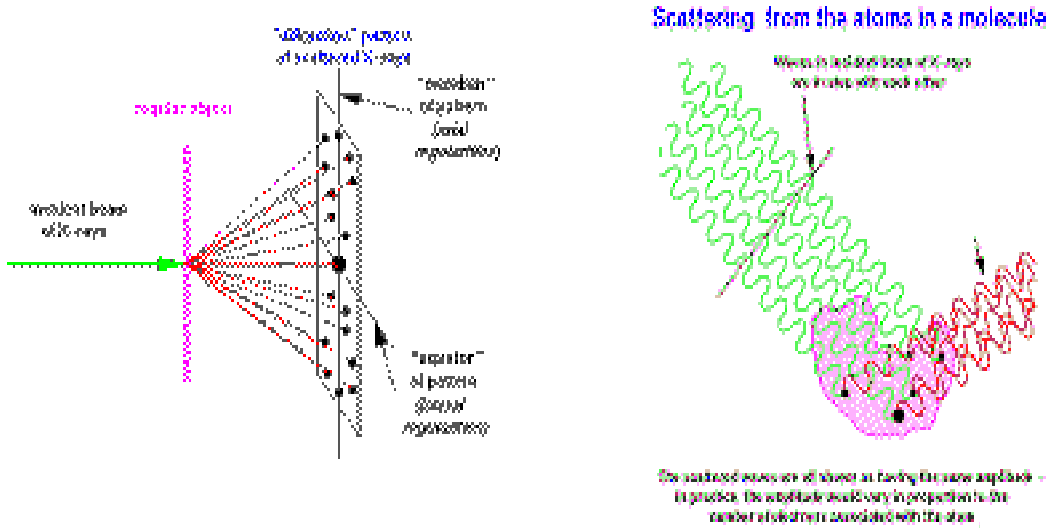
1.2.2) What is a Protein Crystal ?

Crystals are regular, three-dimensional arrays of atoms, ions, molecules or molecular assemblies. Ideally, crystals can be described as infinite arrays in which the building blocks (the asymmetric units) are arranged according to well-defined symmetries (forming 230 space groups) into unit cells that are repeated in three dimensions by translations. Proteins and nucleic acids do not crystallize in space groups with inversion symmetries because they are composed of enantiomers (*L*-amino acids and *D*-sugars, respectively), thus reducing the number of possible space groups to 65.
Why do we need protein crystals for X-ray three dimensional structure determination? In practice, we cannot observe the diffraction pattern of a single molecule, but only that of many in an ordered crystalline array (see below).

1.3) Principles of X-Ray Crystallography

Crystals have the ability to diffract X-rays or neutrons. X-ray diffraction by crystals is a reflection of the periodicity of crystal architecture, so that imperfections in the crystal lattice usually result in poor diffraction properties.
Schematically, a crystal can be described with the aid of a grid or lattice, defined by three axis and the angles between them. Along each axis a point will be repeated at distances referred to as the **unit cell constants**, labelled **a**, **b** and **c**. The angles between **b** and **c**, **a** and **c** and **a** and **b** are denoted _, ß and _, respectively. Within the crystalline lattice, infinite sets of regularly spaced planes can be drawn through lattice points. These planes can be considered as the source of diffraction (see below) and are designated by a set of three numbers called the **Miller indices (hkl)**. The index h is equivalent to the number of

parts into which a set of planes cut the **a** edge of each cell. Similarly, the index k and l define the number of parts into which the **b** and **c** edges of each cell are cut.



Figs.2 and 3: Scattering is the result of interactions between X-rays as electromagnetic waves and the electrons. X-ray diffraction from single crystals results from the interaction of X-rays with the electrons in the crystal lattice.

The elastic scattering of X-rays from free electrons is described by the **Thomson scattering** (Thomson, 1897): $\mathbf{I_{2\_} = I_0 (ne^4/m^2r^2c^4)(1+\cos^2 2\_)}$

where,

$I_{2\_}$ corresponds to the intensity of the scattered radiation at an angle $2\_$ with respect to the beam of unpolarised X-rays,

I_0 is the intensity of the incident radiation,

n is the effective number of independently scattering electrons,

r is the distance between the scatterer and the detector,

e and m are the charge and the mass of the electron, respectively, and

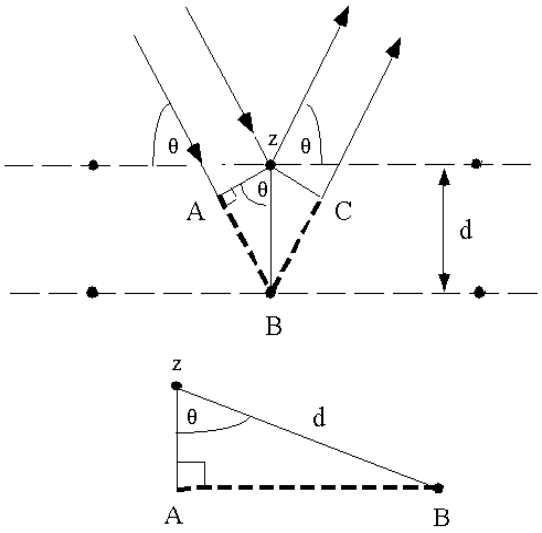
c is the velocity of light.

The term $(1+\cos^2 2\_)$ accounts for the partial polarization of the scattered X-rays. This analysis only holds if the incident X-ray energy is much larger than the binding energies of the electronic states within the atom, because then electrons may be considered as free.

1.3.1) Bragg´s Law

Bragg showed that diffraction from single crystals can be mathematically treated as a reflection from sets of equivalent parallel planes of atoms in the crystals. Constructive interference between the scattered X-rays from successive planes will take place if the path difference between these rays is equivalent to an integral number of wavelengths:

A set of equivalent, parallel planes with index (hkl) and interplanar distance d_{hkl} produces a diffracted beam when X-rays of wavelength _ incide on the planes at an angle _ and are reflected out at the same angle only when _ meets the condition (Bragg and Bragg, 1913):



$$2d_{hkl}\sin\_=n\_$$

Fig. 4. Deriving **Bragg´s law** using the reflection geometry. The lower beam must travel the extra distance (AB+BC) to continue travelling parallel and adjacent to the upper beam.

1.3.2) Ewald´s Sphere

The diffraction principle which follows **Bragg´s law** can be interpreted by a geometrical construction proposed by Ewald in 1921 (Ewald, 1921) (Fig. 5). The theory of the reciprocal lattice and **Ewald´s sphere** are essential tools to construct the diffraction pattern. This rule has been widely used to design instruments for data collection, obtain crystal orientations, determine space groups, predict and index reflections and evaluate diffraction intensities.

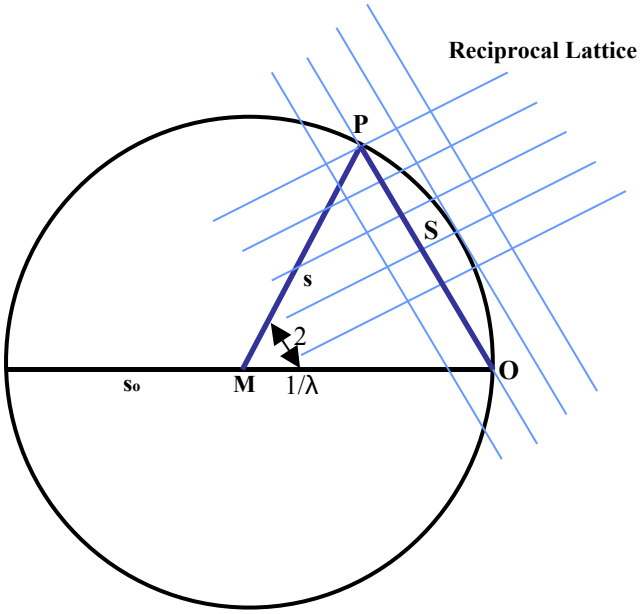


Fig. 5. The Ewald´s sphere is a convenient tool to construct the direction of the scattered beam. The sphere, centred on the crystal M, has a radius of 1/_. The primary beam (s_o) is scattered by the crystal (M). Because P is a reciprocal lattice point OP=$1/d_{hkl}$. A reflection *hkl* occurs in the direction MP (s) when reciprocal lattice point P_{hkl} comes in contact with this sphere. Reflection *hkl* is, according to Bragg´s model, the result of the reflection from the set of equivalent real-space planes (hkl). As the crystal is rotated in the X-ray beam, many reciprocal lattice points come into contact with this sphere producing new reflections. _S_=$2d_{hkl}$sin_/_=$1/d_{hkl}$ when n=1.

A set of planes (hkl) produces the **reflection *hkl*** in the **diffraction pattern** (see Fig. 5). Thus, the diffraction pattern of a lattice is also a lattice, but one whose dimensions are inversely proportional to the dimensions of the real lattice. One can speak of the reciprocal lattice unit cell with axes a*, b* and c*. If the real unit cell angles _, _ and _ are 90°, the reciprocal unit cell has axes a* lying along the real unit cell edge a, b* lying along b, and c* along c. The lengths of edges a*, b* and c* are reciprocals of the lengths of the corresponding cell edges a, b and c: a*=1/a and so forth. Thus, in the reciprocal lattice, the lattice planes are perpendicular to the corresponding planes in the real space and have an interplanar distance of $1/d_{hkl}$, which is reciprocal of the interplanar distance in the real lattice. The reciprocal lattice is spatially linked to the crystal because of the way of defining the lattice points. In this way, when the crystal is rotated the reciprocal lattice rotates with it.

The intensity of a reflection with **Miller indices *hkl*** is proportional to $|\mathbf{F}(hkl)|^2$ where:

$$\mathbf{F(hkl)=\sum_J f_j exp[2\pi i(hx_j+ky_j+lz_j)]}$$

fj is the **atomic scattering factor** for X-rays for the jth atom of co-ordinates (xj,yj,zj) expressed as fractions of the unit cell constants a, b, c. This is the structure factor equation in the absence of thermal motion or disorder. But a crystal is not static as atoms

vibrate around equilibrium positions and the intensity of the diffracted beam is hence weakened. Thus a correction factor has to be usually applied to the atomic scattering factor:

$$\mathbf{T(iso)=exp[-B/4d^2]}$$

The termal vibration is expressed in the **temperature factor (B)**, which is proportional to the mean square atomic displacement _ along the normal to the reflecting planes:

$$\mathbf{B=8\pi^2\_^2}$$

As an example, a B factor of 30 Å corresponds to an average atomic displacement of 0.62 Å.

The amplitude of the structure factors is obtained from the intensity of the diffracted beam, after application of certain correction factors:

$\mathbf{I(hkl)=k\_F(hkl)\_^2}$ being k a constant.

In the absence of anomalous scattering the diffraction pattern has a centre of symmetry because I(h,k,l)=I(-h,-k,-l) according to **Friedel´s law**. Reflections (**hkl**) and (**-h,-k,-l**) are known as **Friedel or Bijvoet pairs**. The structure factors from Friedel pairs have thus the same amplitude but opposite phase angles. The Fourier transform of the structure factor equation is the **electron density (_) equation** that relates the contents of the unit cell to the set of **structure factors F(hkl)**:

$$\mathbf{\_(x,y,z)=(1/V)\sum_{hkl}\_F(hkl)\_exp(i\alpha_{hkl})exp[-2\pi i(hx+ky+lz)]}$$

If the amplitude ½**F**(hkl)½ and the phase α_{hkl} of the structure factor are known for "all" (hkl) planes reflections, then the electron density can be calculated for all points (x,y,z) in the cell and the crystal structure can be solved. The problem of phase determination (**the phase problem**) is the fundamental problem of macromolecular crystallography because no phase information can be directly extracted from the intensities of the diffracted X-rays.

Experimentally the phase problem in protein crystallography can be solved by a number of methods:

- **Multiple isomorphous replacement (MIR)**
- **Multiwavelength anomalous diffraction (MAD)**
- **Patterson search or molecular replacement (MR)**
- **Direct methods**

Solving the phase problem by molecular replacement will be discussed as it was the method used for solving the three-dimensional structures presented in this work

1.4) The Solution of the Phase problem by Molecular Replacement

In the absence of any previous phase information the MIR method is certainly the most important method for the solution of the phase problem in protein crystallography. If a related protein structure is available the method of Patterson search (Molecular replacement) can be used (Hoppe, 1957a; Hoppe, 1957b). For the molecular replacement technique a known structure of a structural homologue is required, which serves as a model for the unknown structure. Homology in the amino acid sequence is an indication of whether a model is suitable.

1.4.1) The Patterson Function

The Patterson function (Patterson, 1934) is the Fourier transform of $|\mathbf{F}(hkl)|^2$ and it may always be calculated from any set of recorded diffraction intensities:

$$\mathbf{P(uvw)=(2/V)\sum_{hkl}|F(hkl)|^2\cos 2\pi(hu+kv+lw)}$$

Patterson functions give information about all possible inter-atomic vectors in the structure, from which an atomic structure of limited complexity can be derived.

1.4.2) Molecular Replacement

Molecular Replacement is a method to obtain a first model of a protein using a homologous protein with known structure as a search model. That is, if the structure of a protein (search molecule) is known that is homologous to the crystallized protein (target molecule) with unknown structure, the former can be used as a model to calculate a starting set of phases that can be iteratively refined. For that, the search molecule must be

oriented and positioned in the unit cell of the target molecule in a way that maximum overlap of the models calculated diffraction pattern with the observed diffraction pattern is achieved. Thus, the related structure (**phasing model**) is used to obtain phase information after orienting the model in the unit cell of the new protein.

Homologous proteins which are derived by divergent evolution from an ancestral protein often have a large proportion of their tertiary structure arranged in similar ways. Haemoglobins and myoglobins form a homologous series as do chymotrypsin, trypsin and elastase. These proteins, despite the similarity in tertiary structures, crystallize in quite different space-groups. In these various situations of identical structures or parts of structures in different crystallographic environments we would expect similarities between their diffraction patterns, because each diffraction pattern is derived from the Fourier transform of the molecule. Thus, we can then derive from the diffraction pattern the relative orientation and position of these molecules. Knowing the structure of one protein in an homologous series, we can find the structures of the others in different crystals. The problem is to place the known protein molecular structure from its crystalline arrangement to the crystal of the protein for which the structure is not yet known.

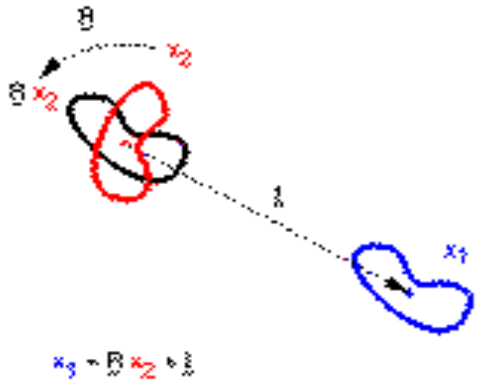


Fig. 6. The solution is the molecular replacement method, which was initiated with the studies of Rossmann and Blow (1962) and Huber and Hoppe (Huber and Hoppe,1962). Placement of the molecule in the target unit cell requires its proper orientation and precise positioning. This involves two steps: **rotation** and **translation**.

In the rotation step the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the next step the translation needed to superimpose the now correctly oriented molecule onto the other is calculated. The basic principle of molecular replacement method can be understood by regarding the Patterson function of a protein crystal structure.

The **Patterson map** is a vector map: vectors between atoms in the real structure show up as vectors from the origin to maxima in the Patterson map. If the pairs of atoms belong to the same molecule, then the corresponding vectors are relatively short and their end-points are found not too far from the origin in the Patterson map; they are called **self-Patterson vectors**. If there were no intermolecular vectors (**cross-Patterson vectors**), this inner region of the Patterson map would be equal for the same molecule in different crystal structures, apart from a rotation difference. For homologous molecules it is not exactly equal but very similar. Therefore, the self-Patterson vectors can supply us with the rotational relationship between the known and the unknown molecular structures.

The solutions of the rotation and translation functions are not always found in a straightforward way. In some cases, it can be necessary to modify the model, for instance, by ignoring the side chains and deletions/insertions in the model, and/or to systematically vary the resolution range of the X-ray data used in the search. With the rapid increase in the number of successful protein structure determinations, molecular replacement has become an extremely useful technique for protein phase angle determination.

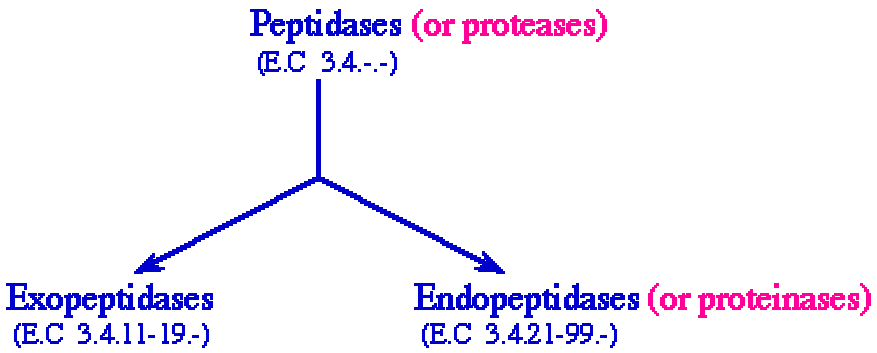
Part 2. Proteinases (Peptidases)

2.1) Introduction to Proteinases (Peptidases)

Proteinases are widely distributed in nature, where they carry out a wide array of functions, i.e. digestive and degradative processes, blood clotting cascade, cellular and humoral immunity, fibrinolysis, fertilization, embryonic development, protein processing and tissue remodeling. They are present in all forms of living organisms (Neurath, 1984). *In vivo* the activity of many proteinases is controlled by endogenous protein inhibitors that complex with the enzymes and block them. The three-dimensional structures of a large number of proteinases have already been determined.

2.1.1) Protease, Proteinase or Peptidase?

The International Union of Biochemistry and Molecular Biology (IUBMB) (1984) has recommended to use the term peptidase for the subset of peptide bond hydrolyses (Subclass E.C 3.4.). The widely used term protease is synonymous with peptidase. Peptidases comprise two groups of enzymes: the **endopeptidases** and the **exopeptidases**, which cleave peptide bonds at points within the protein and remove amino acids sequentially from either N- or C-terminus, respectively. The term proteinase is also used as a synonym for endopeptidase and five mechanistic classes of proteinases are recognized by the IUBMB as detailed below. The modern scheme of nomenclature is thus **(Fig. 7)**:



Five major families of peptidases have been identified so far (Rawlings, N.D., and Barrett, A.J., 1995), in which serine, threonine, cysteine, aspartic or metallo groups play primary roles in catalysis (Fig. 8). They are endopeptidases and are divided on the basis of catalytic mechanism into **serine** endopeptidases, **threonine** endopeptidases, **cysteine** endopeptidases, **aspartic** endopeptidases and **metallo-**endopeptidases. This classification is thus based on a functional criterion, namely, the nature of the most prominent functional group in the active site.

The serine, threonine and cysteine peptidases differ from the aspartic and metallopeptidases in that the **nucleophile of the catalytic site** is an **amino acid atom**, whereas it is an **activated water molecule** in the other families. As a consequence, acyl enzyme intermediates can be only formed in the reactions of the Ser/Thr/Cys peptidases, and only these peptidases can readily act as transferases. **Fig. 8**:

2.2) Proteolytic Enzymes are Synthesized as Inactive Precursors (Zymogens)

In higher organisms proteinases are generally synthesized as larger precursors, also termed **zymogens**, which carry N-terminal extensions (prosegments) that regulate their proteolytic activity, to prevent unwanted protein degradation, and to enable spatial and temporal regulation of enzymatic activity. Upon sorting or appropriate compartmentalization, zymogen conversion to the fully active enzyme typically involves limited proteolysis. There is a broad range of sizes for the different activation segments, from dipeptide units to independently folding domains comprising more than 100 residues.

The activation segment has never been found at the C-terminus of a zymogen, thus precluding the risk of the proenzyme gaining activity before the synthesis of the polypeptide is complete.

Zymogen activation via limited proteolysis plays a central role in a number of important biological processes (Neurath, 1984) such as the blood coagulation (Davie et al, 1991) and apoptotis (programmed cell death) (Yang, 1998). The subject of zymogen activation traces its roots to studies of the pancreatic and gastric enzymes in the 1930s (reviewed by Neurath, 1957; Kunitz and Northop, 1938; Herriot, 1939; Rovery et al., 1953). The development of X-ray crystallographic methods for proteins in the 1960s and 1970s expanded our knowledge of the structural basis for the conversion process in the serine proteinases. Crystallographic studies by Kraut *et al*. (chymotrypsinogen) and Bode and Huber (trypsinogen) provided the first glimpse of the structural basis for the activitation of zymogens (Huber and Bode, 1978; Bode et al., 1979). The contributions of W. Bode and R. Huber to the field of zymogen activation has continued unabated through two decades (Avilés et al., 1997; Khan and James, 1998).

X-ray crystallographic studies of zymogens and comparisons with their active counterparts have identified the structural changes that accompany conversion (for a review see Khan and James (Khan and James, 1998). Most families of proteolytic enzymes have now at least one representative zymogen structure in the protein database, including both chymotrypsin-like (chymotrypsinogen: Wang *et al*., 1985, trypsinogen, Bode *et at*., 1976) and subtilisin-like serine proteinases (prosubtilisin: Gallagher *et al*., 1995), cysteine proteinases (procathepsin B, Cygler *et al*. 1996; Turk *et al*., 1996; apo-calpain, Strobl *et al*., 2000), aspartyl proteinases (pepsinogen, James and Sielecki, 1986), metallo-proteinases (prostromelysin, Becker *et al*., 1995) and carboxypeptidases (procarboxypeptidase B: Coll *et al*., 1991). The chymotrypsin-like serine proteinases have been by far the most extensively family studied.

The inhibitory mechanisms utilized by activation segments are diverse. The positions and conformations of the catalytic triads of the serine and cysteine proteinase zymogens, the two catalytic Asp residues in aspartic proteinase zymogens, and the catalytic Zn^{2+} ion in metalloproteinase zymogens are virtually identical to their corresponding active forms. However, the conversion process often involves significant

conformational changes in regions that are adjacent to the active site, or within the activation segments that are subsequently removed.

2.3) Serine Proteinases

2.3.1) Introduction to Serine Proteinases

Serine proteinases are the most intensively studied group of proteins in biology (Neurath, 1984). Interest in this remarkable family of enzymes is generally due to their well-characterized, widespread and diverse role in a host of physiological and pathological processes (Walker and Lynas, 2001).

Serine peptidases are peptidases in which the catalytic mechanism depends upon the hydroxyl group of a serine residue acting as the nucleophile that attacks the peptide bond. Barret and Rawlings have introduced the nomenclature for the classification of all hydrolytic enzymes, describing species in so-called families and clans. (Barret and Rawlings, 1995). Serin proteinases are subdivided into particular **clans** (**SA**, **SB**, **SC**, **SH**) by comparing the tertiary structures and the order of the catalytic residues in the sequence. The catalytic machinery usually involves, in addition to the serine that carries the nucleophile, a proton donor (or general base). In clans SA, SB, SC, and SH, the proton donor is a histidine residue, and there is a catalytic triad because a third residue is required, probably for orientation of the imidazolium ring of the histidine. The third residue that composes the triad is usually an aspartate, but other residues are possible (i.e. a His in clan SH). Despite these significant differences, serine proteases of clans SA (chymotrypsin-like) (Lesk and Fordham, 1996), SB (subtilisin-like) (Siezen and Leunissem, 1997) and SC (α/β-hydrolase fold) (Ollis *et al*., 1992) mantain a strictly conserved active site geometry among their catalytic Ser, His and Asp residues.

In clan SA the order of the **catalytic triad** is **His**, **Asp**, **Ser**, and the tertiary structure consists mainly of ß-sheet. The members of this clan possess a two domain structure, with each domain containing a ß-barrel, and the active site cleft is located between the domains. All the proteolytic enzymes in this clan are **endopeptidases**. Only **clan SA** will be commented, as Granzyme B and Granzyme K belong to this clan.

2.3.2) Family S1 of Trypsin-like Proteinases (Clan SA)

The **clan SA** is composed by the **family S1 of trypsin-like proteinases** (Barret and Rawlings, 1995). It is the largest family of peptidases, both in terms of the number of sequenced proteins, and in the number of recognizable different peptidase activities. Most of these peptidases are extracellular proteins and have thus a N-terminal signal peptide, and all of them are synthesized as precursors (**zymogens**) with an N-terminal extension that acts as **propeptide**, requiring proteolytic cleavage to generate the active enzyme. The length of this extention varies enormously, with the smallest being two residues in human cathepsin G, and the larger containing hundreds of residues.

2.3.3) Catalytic Mechanism: Serine Proteinases Cleave Peptide Bonds by Forming Tetrahedral Transition States

Serine proteinases have been extensively studied, both by kinetic methods in solution and by X-ray structural studies to high resolution. A serine proteinase cleaves peptide bonds within a polypeptide substrate to produce two new smaller peptides.

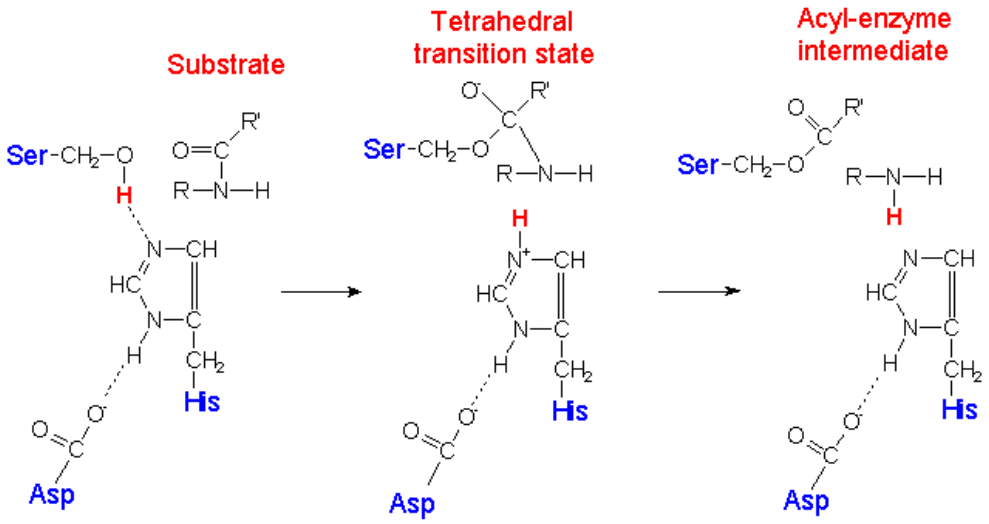
The reaction proceeds en two steps. The first step produces a covalent bond between the carbon **C1 of the substrate** and the hydroxyl group of a **reactive Ser residue** of the enzyme. Production of this **acyl-enzyme intermediate** proceeds through a negatively charged transition state intermediate where the bonds of C1 have tetrahedral geometry in contrast to the planar triangular geometry in the peptide group. During this step the peptide bond is cleaved, one peptide product is attached to the enzyme in the acyl-enzyme intermediate, and the other peptide product rapidly diffuses away.

In the second step of the reaction, **deacylation**, the acyl-enzyme intermediate is hydrolyzed by a water molecule to release the second peptide and to restore the Ser-hydroxyl of the enzyme. This step also proceeds through a negatively charged tetrahedral transition state intermediate. The catalytic triad is no longer thought to act as a "charge-relay system" as was once widely assumed, but instead His57 is thought to act as a general base (Kossiakoff and Spencer, 1981). Asp102 is believed to stabilize the correct tautomer of His57, and to provide compensation for the developing positive charge during the catalytic reaction.

In the following pages the catalytic mechanism of serine proteinases will be schematically shown.

2.3.3.1) Acylation

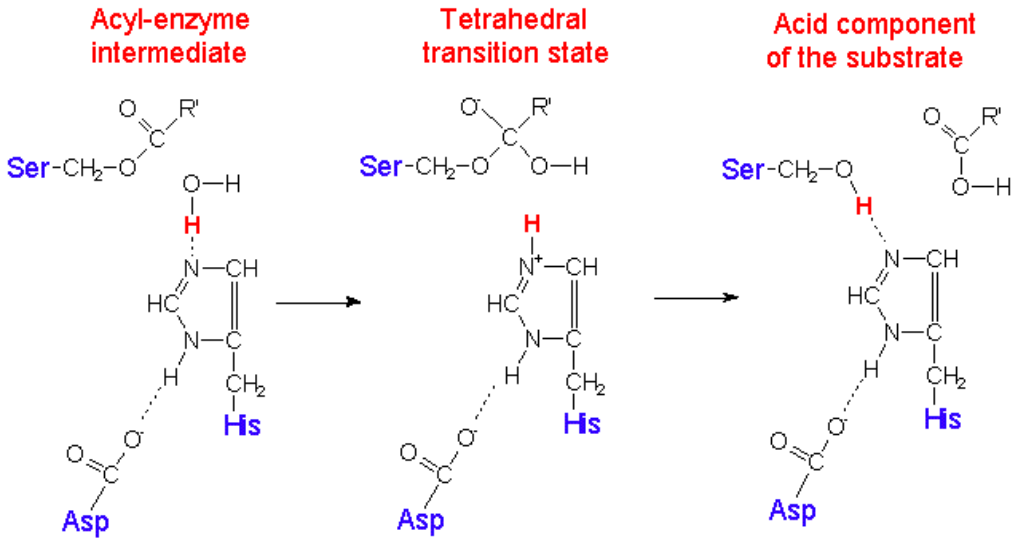
In this step the peptide bond is cleaved and an ester linkage is formed between the peptide carbonyl carbon and the enzyme. The **acylation phase** can be seen in the acylation reactions shown in **Fig. 9**:



In the first step the hidrogen atom of the Ser195 OH group is transferred to the imidazole ring of the histidine, which acts as a **base** (**proton acceptor**), the serine Oγ atom reacts then with the carbonyl group of the substrate peptide amino acid, forming a tetrahedral transition state. The transition state reforms the carbonyl double bond and breaks the nitrogen-carbon peptide bond. The C-terminal portion of the peptide chain accepts the hydrogen from the catalytic His57 and breaks free.

2.3.3.2) Deacylation

In this step the ester linkage is hydrolized and the enzyme regenerated. **Fig. 10**:



The remaining N-terminal part of the peptide substrate (from the acylation step), still attached to the serine, must be removed and the active site regenerated. This second step proceeds through a reaction with water (hydrolysis of the acylated serine residue). The -OH group derived from the activated water molecule reacts with the acyl-enzyme forming a second tetrahedral intermediate, and the hydrogen is bonded to the catalytic histidine. When the carbonyl bond reforms, the bond to the serine oxygen is broken, freeing the N-terminal half of the peptide chain. The hydrogen bonded to the histidine is lost to the serine oxygen, and a hydrogen bond is reformed between the serine and histidine.

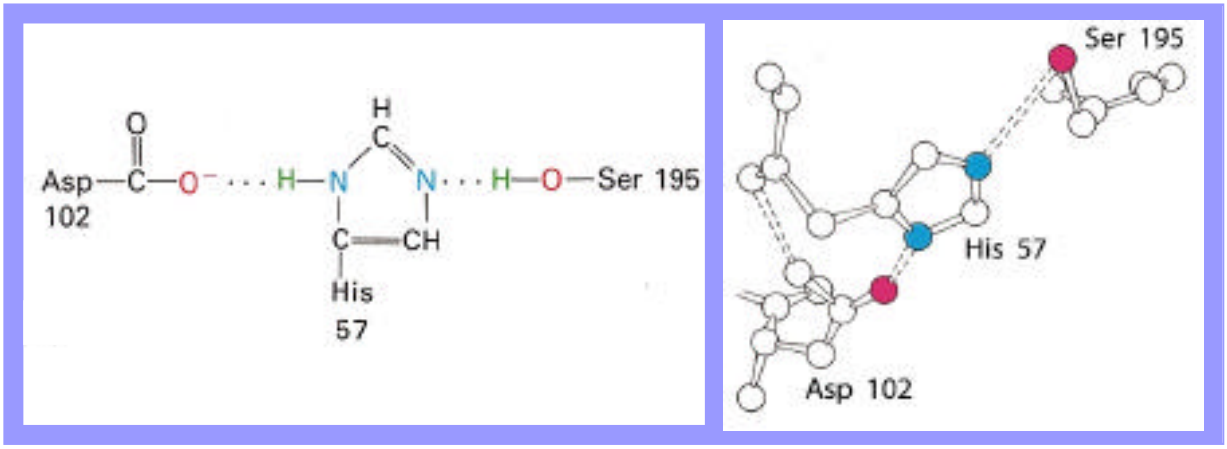
2.3.4) Structural Features of Serine Proteases

There are four important structural features that facilitate the catalytic mechanism of serine proteinases:

2.3.4.1) The Catalytic Triad is Composed by His57, Asp102 and Ser195

The enzyme provides a general base, a His residue, that can accept the proton from the hydroxyl group of the reactive Ser thus facilitating formation of the covalent

tetrahedral transition state. This His residue is part of a catalytic triad consisting of the side chains of **His57**, **Asp102** and **Ser195** (Fig. 11) (chymotrypsinogen numbering, see below) which are close to each other in the active site of the enzyme, although they are far apart in the amino acid sequence of the polypeptide chain. **Fig. 11**:



2.3.4.2) The Oxyanion Hole

The tetrahedral transition state intermediate is tightly bound and stabilized by groups that form hydrogen bonds to the negatively charged oxygen atom attached to C1. These groups are located in a pocket of the enzyme called the **oxyanion hole** (Fig. 12). The positive charge on the protonated His residue also helps to stabilize the negatively charged transition state. It is also believed that these features presumably destabilize binding of substrate in the normal state.

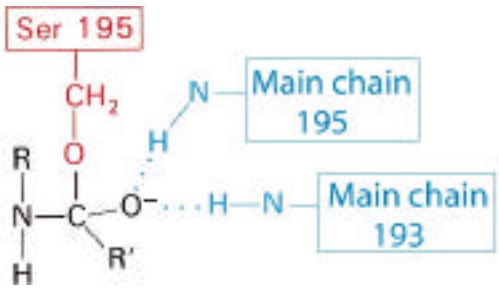


Fig. 12. Main chain nitrogen atoms of residues 195 and 193 stabilize the negative charge of the tetrahedral transition state intermediate.

2.2.4.3) The Unspecific Main-chain Substrate Binding

Most serine proteinases have no absolute substrate specificity and they can cleave peptide bonds with a variety of side chains adjacent to the peptide bond to be cleaved, the **scissile bond** (see Fig. 14). This happens because polypeptide substrates exhibit a nonspecific binding to the enzyme through their main-chain atoms, which form hydrogen bonds in a short antiparallel ß-sheet with main-chain atoms of a loop region in the enzyme. This nonspecific binding therefore also contributes to stabilization of the transition state.

2.3.4.4) The Specificity Pocket

Even though these enzymes have no absolute specificity, many of them show a preference for a particular side chain located before the scissile bond as seen from the amino end of the polypeptide chain. The **preferred side chain** is oriented so as to fit into a pocket of the enzyme called the **specificity pocket**. The **specificity** mainly derives from three residues, **216**, **226** and **189** with additional contributions from residue **190**. Residues 216 and 226 line the sides of the pocket. Residue 189 is at the bottom of the specificity pocket (see Fig. 13). Side chains from these three residues confer preferential cleavage requirements that is, the structural basis for the substrate preference lies in the side chains that line the substrate specificity pocket in the different enzymes.

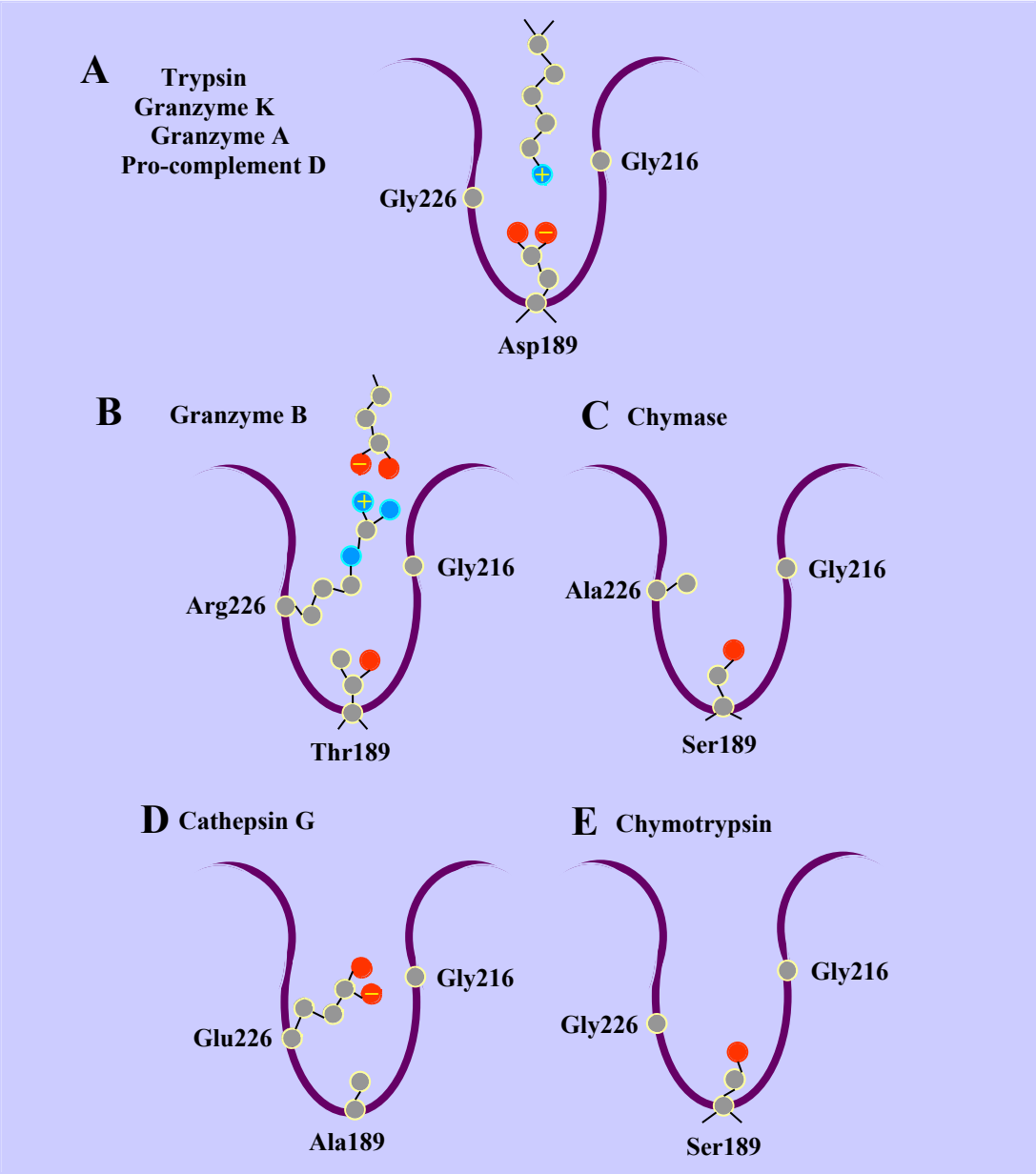


Fig. 13. Schematic diagrams of the specificity pockets of Granzymes K, A and B, Trypsin, Chymase, Cathepsin G, Pro-complement D and Chymotrypsin, illustrating the corresponding residues at positions 189, 216 and 226. As an example, the preference for a Lys and Asp residue at P1 for Granzyme K (together with Trypsin, Granzyme A and Pro-complement D) and Granzyme B, respectively, is shown. Granzyme K, like trypsin, prefers positively charged side chains that can interact with Asp189 at the bottom of the specificity pocket. Granzyme B substrate specificity is conferred by the presence of Arg226, which interact with an entering Asp residue.

The catalytic efficiency of a protease requires specific interactions between the enzyme and the substrate. There is a limited number of residues that interact with the enzyme. The major part of the substrate recognition in the serine proteinases of the trypsin family is determined by the S1-P1 interactions (nomenclature according to Schechter and Berger, 1967) (Fig. 14) which is known as primary specificity. The S' sub-sites are also important for substrate discrimination and catalytic efficiency.

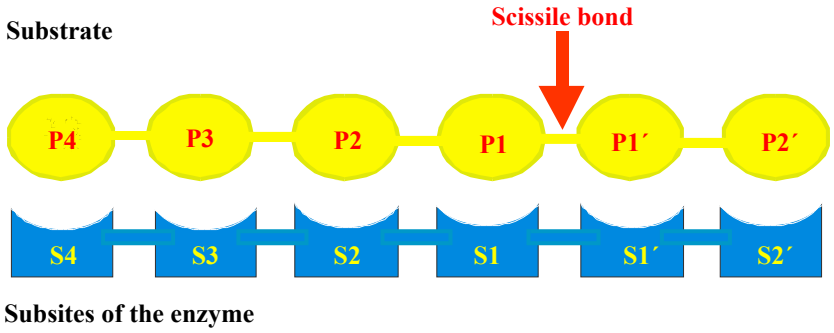


Fig. 14. Schematic representation of protease/substrate interaction, according to the **nomenclature of Schechter and Berger** (1967). The protease is depicted in blue and the substrate in yellow. P4-P2' represent six residues of the substrate around the cleavage site and S4-S2' indicate the corresponding subsites of the enzyme.

2.3.5) Trypsin-like Serine Proteinases Fold

Trypsin was first described and named in 1876 by Kühne as the proteolytic activity in pancreatic secretions (Kühne, 1876; Neurath and Zwilling, 1986). In 1967 David Blow determined the three dimensional structure of chymotrypsin being at this time one of the very first enzyme structures known at high resolution. The three-dimensional structure of cattle trypsin was determined in 1974 (Huber *et al*., 1974, Bode *et al*, 1975, Bode and Schwager, 1975), and this structure has become the **prototype of the S1 family of proteases** as much of the fundamental knowledge about the family has been derived from the study of this enzyme. Structural analyses of eukaryotic and prokaryotic members of this family have revealed a common three-dimensional structure (Delbaere *et al*., 1975, Bode *et al*., 1983).

The polypeptide chain of trypsin-like serine proteinases is folded into **two domains** of the antiparallel ß-barrel type, each containing six ß-strands with the same topology: a **Greek key motif** (strands 1-4) followed by an **antiparallel hairpin motif** (strands 5 and 6). The **active site** is situated in a crevice between the two domains. **Domain 1** contributes two of the residues in the catalytic triad (**His57** and **Asp102**), whereas the reactive Ser195 is part of **domain 2**.

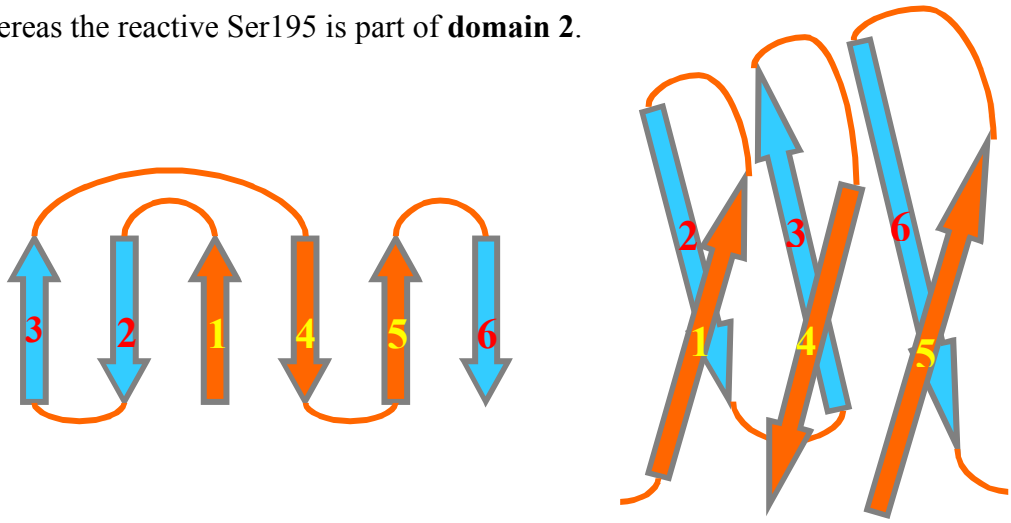


Fig. 15. Topology diagrams of the domain structure of chymotrypsin. The chain is folded into a six-stranded antiparallel ß-barrel arranged as a Greek motif followed by a hairpin motif.

2.3.6) Zymogens of Serine Proteinases

The conversion of chymotrypsin-like zymogens to the active enzymes was the earliest model of the role of limited proteolysis in zymogen activation (Neurath, 1957). The basic principle of zymogen conversion, that structural changes in the zymogen are necessary for activation, was established over 40 years ago (Davie and Neurath, 1955). X-ray crystallographic studies of serine-proteinases in the 1960s and 1970s established that the overall fold of the zymogen and the active enzyme are identical but a small region undergoes large structural changes during conversion. Comparisons of the structures of chymotrypsinogen and α-trypsinogen revealed that the substrate-binding cleft was only partially formed in the zymogen and that the peptide bond Met192-Gly193 was not orientated to contribute a proton to the oxyanion hole (Steitz, 1969; Freer, 1970). Soon afterwards, the detailed crystallographic studies of trypsinogen and trypsin in several crystal forms by Bode and Huber, including an isomorphous trigonal form of the

zymogen and the active enzyme, provided the fundamental insight into the basis for low reactivity of the zymogen (Fehlhammer et al., 1977; Bode and Huber, 1978; Huber and Bode, 1978).

The catalytic domains include partially flexible activation domains, which attain rigid, functional conformation only upon proteolytic liberation of a **highly conserved N-terminus** (Huber and Bode, 1978). It starts typically with the sequence Ile16-Val17-Gly18-Gly19. The activation of chymotrypsinogen, for instance, is initiated by trypsin- or enterokinase-mediated cleavage of the Arg15-Ile16 peptide bond and is influenced by pH and the presence of Ca^{2+} ions (reviewed by Neurath, 1957). Limited proteolysis results in a newly-liberated amino-terminus at Ile16, which forms an ion-pair with the carboxylate side chain of Asp194, triggering the formation of an active enzyme (Freer et al, 1970; Huber and Bode, 1978) (see below).

The active site residues His57, Asp102 and Ser195 remain virtually unchanged during conversion. Together with a preformed catalytic triad, chymotrypsinogen and trypsinogen possess a "**zymogen triad**" composed by **Asp194, His40 and Ser32**, linked by ionic bonds (Bode, 1979; Madison, 1993). The geometry of the "zymogen triad" found in chymotrypsinogen is very similar to that of the catalytic triad, which consists of the same three residues. This stabilizing factor, however, is absent in tPA and other non-pancreatic serine proteinases (Renatus *et al*., 1997).

The **activation domain**, seen in the crystal structures of chymotrypsinogen (Freer, 1970; Wang, 1985), trypsinogen (Bode and Huber, 1976; Bode and Huber, 1978; Walter, 1982) and prethrombin-2 (Vijayalakshmi, 1994) requires a conformational change. The solution of the structure of bovine trypsinogen to 1.8 Å resolution (Wang et al, 1985) showed that part of the substrate binding cleft that includes the segments 213-220 and 225-228 is preformed in the zymogen. The most dramatical conformational changes occur in the segment Ser189-Asp194, which immediately precedes the catalytic Ser195. This segment consists of a loop that points "inwards" in the zymogen, but following activation, the loops moves "outwards" towards the solvent forming the mature substrate binding cleft.

The formation of a **salt bridge between Ile16 and Asp194** creates a **restructured activation domain** with **a functional substrate recognition site**, the

oxyanion hole and a properly shaped **S1 specificity pocket** (Bode, 1976; Huber and Bode, 1978).

2.4) Granzymes are Serine Proteinases

2.4.1) Introduction to Granzymes

Granzymes (Gzm) are chymotrypsin-like serine proteinases present in the lytic granules of activated **cytotoxic T-lymphocytes** ($CD4^+$ and $CD8^+$) and **natural killer cells** (NK-cells) (Hameed, 1988; Nagler-Anderson, 1988; Shi, 1992; Kam, 2000). They play an important role in defending our body against virus-infected cells (Trapani, 1999), tumor cells and cells infected with intracellular bacteria (Mullbacher, 1999; Stenger, 1997; Tschopp, 1996).

The recognition of a virus-infected or tumor cell by lytic cells proceeds via the **T-cell receptor** and induces a raise in the intracellular Ca^{2+} level (Bregestovski, 1986), giving rise to the expression of new granzymes and **Cathepsin C** (Esser, 1998; Brown, 1993), and the parallel releases of already stored granules into the intercellular space (Mabee, 1998). **Perforin** (Podack, 1985), a pore-forming protein, also localized in the granules, forms transmembrane pores in the target cell (Ishiura, 1990). However, GzmB uses the **mannose 6-phosphate/insulin-like growth factor II receptor** to enter the target cells (Motyka, 2000).

2.4.2) Chromosomal Location

The five distinct human granzymes described so far are located on three chromosomes (Smyth, 1993; Smyth, 1996). GzmA and GzmK cluster in chromosome **5q11-q12** (Fink, 1993). The granzymes B and H are located next to each other at chromosome **14q11.2**, clustered together with cathepsin G and mast cell chymase 1 (Jenne, 1989). GzmM is located at **19p13.3**, in neighborhood of azurocidin 1, proteinase 3 and neutral elastase 2 (Hanson, 1990; Pilat, 1994; Jenne, 1994; Jenne, 1991; Jenne, 1988; Jenne, 1988).

2.4.3) Cathepsin C Activates Several Granzymes

Granzymes are synthesized as **pre-pro-enzymes** with an 18 to 24-residue leader peptide that is removed leaving the **pro-peptide**. With the exception of GzmM, which has a hexapeptide (SSFGTQ) as propeptide, the other human granzyme zymogens carry a **dipeptide**.

In mouse, the dipeptides of GzmA and B are processed by the **dipeptidyl peptidase I** (**cathepsin C** / **CatC**) (Pham, 1999). In CatC knock-out mice, which still produce normal levels of granzyme-zymogens, neither GzmA nor GzmB could be detected in their active forms. The cysteine proteinase CatC activates several lysosomal proteinases in a similar way and in humans the loss of its function leads to severe illness (Hart, 2000; Hart, 1999).

In cytotoxic lymphocytes, especially CD8+ thymocytes and myeloid cells, CatC is expressed at higher levels compared to cells without killer-cell function, and co-localizes within specialized granules with granzymes (Thiele, 1990).

2.4.4) Transport and Storage Within the Granules

GzmA and B, both bearing mannose 6-phosphate residues, are partly targeted to the lytic granules via the **mannose 6-phosphate receptor** (Griffiths, 1993). Both GzmK and murine GzmC lack consensus sequences for N-glycosylation, thus these two granzymes have to be transported to the granules via different mechanisms than using the mannose-6-phosphate receptor. Because of the acidic pH of about 5.2 and their tight packaging with proteinase-resistant chondroitin sulfate proteoglycans (Masson, 1990), granzymes are inactive within the granules. Proteoglycans ensure tight packaging of granzymes inside the granules due to favorable ionic interactions (Masson, 1990). In the plasma of patients with activated CTLs and NK-cells, where the level of extracellular GzmA is elevated, part of this extracellular granzyme was found to be still complexed to proteoglycans. The proteoglycans protect the granzyme against inactivation by proteinase inhibitors (Spaeny-Dekking, 2000).

2.4.5) Granzyme Substrates

Once inside the target cell, granzymes activate apoptosis by different pathways (Trapani, 2000). GzmB rapidly triggers apoptosis (Yang, 1998; Van de Craen, 1997) by activating several pro-caspases (Atkinson, 1998; Dall´Acqua, 1999; Darmon, 1995; Medema, 1997, Chynnaiyan, 1996), Bid (Barry, 2000), poly (ADP-ribose) polymerase, and DFF45/ICAD (Thomas, 2000). GzmA is known to bind PhapII, and interleukin-1-β (Beresford, 1997; Pasternak, 1991; Irmler, 1995). Gzms are also transported to the nucleus by a mechanism that has been shown in some respect to be perforin dependent (Blink, 1999; Jans, 1996; Jans, 1998).

Little is known about the physiological substrates of GzmK, GzmH and GzmM during apoptosis *in vivo*. They are constitutively expressed but less abundant than GzmA and GzmB. Substrates for human GzmK have not yet been described. Rat GzmK is identical with fragmentin-3, which was characterized as an inducer of DNA fragmentation in target cells (Shi et al., 1992).

It is also not known whether GzmK and GzmA recognize the same substrates *in vivo* as both granzymes cleave peptide bonds after Arg and Lys P1-residues (see Fig. 13). This substrate specificity is conferred by the presence of Asp189 at the bottom of the specificity pocket (Fig. 13). These proteinases might cleave distinct proteins, being each of them capable of activating the apoptotic pathway independently, resulting in synergistic amplification of the apoptotic cascade in the target cell.

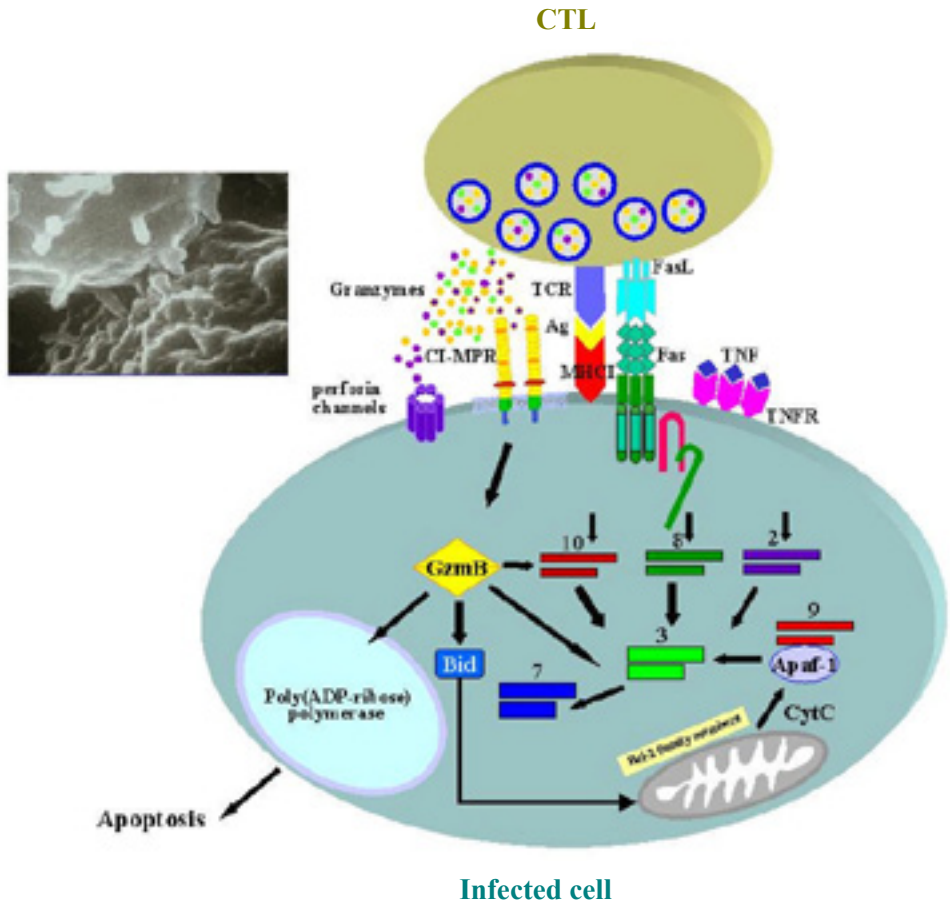


Fig. 16. Schematic representation of a CTL binding to a virus-infected cell. Upon interaction with the TCR-receptor and other membrane receptors, the CTL cytoplasmic vesicles liberate their content: perforin and granzymes (small colored circles) into the space between both cells. GzmB enters the cell through the mannose-6-phosphate receptor, the other granzymes are supposed to enter through the membrane perforin pores. The pro-apoptotic cascade induced by GzmB is shown. GzmB activates several pro-caspases (3,7,10) and the pro-apoptotic Bcl-2 family member, Bid in the cytoplasm. GzmB is also targeted to the nucleus of the infected cell where it cleaves other substrates. As a result of granzyme actions the cell eventually dies by apoptosis (programmed cell death).

2.5) Carboxypeptidases

The hydrolysis of the peptide bond at the C-terminus of peptides and proteins in living organisms is performed by exopeptidases belonging either to the metallo-carboxypeptidase family (Zn-containing class of proteinases) or to the serine-carboxypeptidase family. The latter enzymes contain a reactive Ser residue at the active center and other catalytic residues which resemble and have a similar role to those of the Ser-endopeptidases superfamily (Breddam, 1986; Liao and Remington, 1990; Ghuysen, 1991). The residues and conformation of the active centre of metallocarboxypeptidases are different, as they have one Zn^{2+} ion directly involved in the catalytic mechanism.

Depending on the criteria used, carboxypeptidases (CP) can be classified in different ways: i.e. in accord to their zinc-binding motifs, upon their substrate specificity, or an alternative subdivision is possible based on their specific physiological roles.

Two subfamilies of metallocarboxypeptidases (Family M14 in the Barret-Rawlings-Woessner classification, Barret et al., 1998) can be defined based upon sequence homology and overall structure: the A/B and the E/H (or N/E, based on the first two members identified) subfamilies. The first family has been traditionally referred to as "pancreatic or digestive", whilst the second family has been commonly named "regulatory". However, this is not accurate, because at least one of the members of the A/B family has a regulatory role in fibrinolysis (i.e., carboxypeptidase U, also known as plasma CPB or thrombin activable fibrinolysis inhibitor, TAFI (Hendriks, 1998)).

Several other carboxypeptidases have been identified and are thought to be involved in diverse biological processes. These enzymatic activities include control of vasoactive peptide levels by lysine carboxypeptidase (Skidgel, 1998), and hormone processing in the case of carboxypeptidase M (Skidgel, 1998) and E/H. Carboxypeptidases as well as another family of zinc proteinases (the matrix metalloproteinases or MMPs) have also been implicated in cancer and inflammation (Skidgel and Erdös, 1998; Lochter et al., 1998; Coussens and Werb, 1996).

2.5.1) Carboxypeptidase Subfamilies

2.5.1.1) A/B Subfamily

All the members of the A/B subfamily are soluble, non-glycosylated proteins initially produced as inactive zymogens. Activation requires removal of a pro-peptide segment, in some cases by multiple endopeptidase cleavages. The members of this subfamily that have been characterized are optimally active in neutral pH range, reflecting the pH of the environments where they function.

The digestive enzymes carboxypeptidase A (CPA) and carboxypeptidase B (CPB) are the best known representatives of this protein family. Pancreatic CPA and CPB hydrolyse peptides in the gut following the action of chymotrypsin and trypsin on ingested proteins. They are secreted as zymogens which are activated by trypsin in the duodenum. The process of activation releases the segment known as the "pro-segment" or "activation segment", which plays an inhibitory role by sterically blocking the active site, and thereby prevents binding of substrates to the carboxypeptidase moiety.

The traditional classification of CPs into A and B forms takes their different substrate specificity into account. CPA has a preference for aliphatic and aromatic C-terminal residues (Fig. 17), while CPB cleaves off basic C-terminal residues. Two different isoforms, A1 and A2, were described in the rat (Gardell, 1988; Oppezzo, 1994) and in humans (Pascual, 1989; Catasús, 1995). The A1 form corresponds to the previously named "A" with preference for aliphatic C-terminal residues of peptide substrates, the A2 isoform shows specificity for bulkier aromatic C-terminal residues, being the only isoform that shows specificity towards tryptophan. Bovine pancreatic carboxypeptidase A is one of the best characterized zinc proteases and serves as a paradigm for understanding zinc protease activity (Fig. 17).

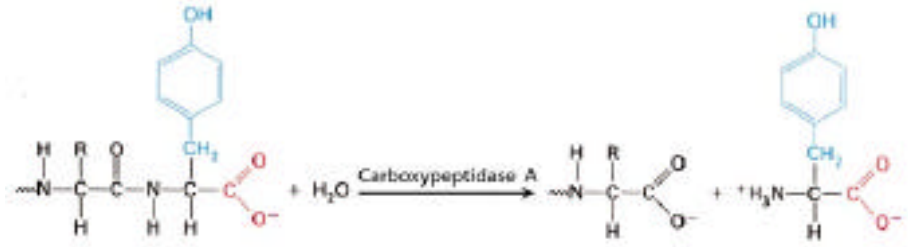


Fig. 17. (previous page) CPA has a preference for aliphatic and aromatic C-terminal residues.

Procarboxypeptidases A from mammals are stored and secreted either as monomers, as binary complexes with proproteinase E or with chymotrypsinogen C, or as a ternary complex, with the last two serine proteinases (Puigserver et al., 1986; Kerfelec et al., 1985). In contrast, procarboxypeptidases B seem to occur only as monomers (Burgos et al., 1989; Pascual et al., 1989; Pascual et al ., 1990, Yoneda, 1980).

2.5.1.1.1) Crystal Structures of Carboxypeptidases from the A/B Family

Crystal structures are now available from pancreatic pro- and carboxypeptidases A1, A2 and B (Rees et al., 1983, Schmid and Herriott, 1976, Famming et al., 1991; Coll et al., 1991, Guasch et al., 1992, Gomis-Rüth et al., 1995, García- Sáez et al., 1997). In the following paragraph the crystal structure of human CPA2 will be briefly described as it is most closely related to the insect procarboxypeptidase presented in this work.

Human Carboxypeptidase A2

The first crystallographic report on a procarboxypeptidase of the A2 type was done by García-Sáez et al. (García-Sáez et al., 1997). They solved both the crystal structure of the human procarboxypeptidase PCPA2 and its active form. Human PCPA2 has a globular shape with two clearly separated moieties:

1. The **pro-** or **activation segment**
2. The **carboxypeptidase moiety**

The activation segment is further divided into two different parts, as was previously described for porcine PCPB (Coll et. al., 1991):

a) The **N-terminal globular domain**
b) A long α-helix or **connecting segment**

The connecting segment links the globular domain with the enzyme and possesses the (trypsin) activation cleavage site.

The N-terminal globular domain has an open sandwich antiparallel-α/antiparallel-β fold, with two α-helices and four β-strands that follow the β1α1β2β3α2β4 topology.

The carboxypeptidase moiety has the same fold as its non-human homologues (Rees et al., 1983): eight α-helices and a mixed eight-stranded β-sheet forming a globular α/β protein.

2.5.1.2) N/E subfamily

In contrast to the A/B family, members of the second CP family commonly referred to as "regulatory" carboxypeptidases range in size and pH optima. This family is more appropriately referred to as the N/E family, based on the first two members that were identified. Unlike the A/B family, the N/E CPs are not produced as inactive precursors that require proteolysis to produce the active form. Instead, the N/E enzymes rely on their substrate specificity and subcellular compartmentalization to prevent inappropriate cleavages that would otherwise damage the cell. Furthermore, all members of the N/E family contain at least one extra domain that is not present in the A/B family. The function of this C-terminal domain remains unclear. It consists of about 80 amino acids and folds into a structure that resembles the β-barrel found in transthyretin. It is possible that this transthyretin-like domain would help in the folding of the carboxypeptidase domain. Besides, it could be involved in protein oligomerization or regulation of the enzyme activity.

2.5.1.2.1) Crystal Structures of Carboxypeptidases of the Regulatory Subfamily

Gomis-Rüth and coworkers (Gomis-Rüth et al., 1999) solved the first crystal structure of domain II of duck carboxypeptidase D (CPD-2), a prototype for members of the regulatory metallocarboxypeptidase subfamily. The CPD-2 polypeptide chain is folded into two distinct subdomains:

1. The catalytic CP subdomain
2. The C-terminal subdomain.

The catalytic CP subdomain presents the α/β-hydrolase fold topology (Ollis et al., 1992), already observed in other carboxypeptidases. The C-terminal subdomain consists of 80 residues and has a rod-like shape, with its N- and C-terminus located on opposite sides of the rod. This subdomain shares topological similarity with transthyretin (prealbumin, Blake and Oatley, 1977). Both subdomains are connected via a single α-helix.

CPD-2 catalytic subdomain displays closest topological similarity to CPA and CPB. The most important differences between these proteases can be observed in the chain segments that form the funnel-like access to the active site cleft. In contrast to the pancreatic CP, CPD-2 does not have a prosegment.

2.5.2) Catalytic Mechanism of Carboxypeptidases: The Water-promoted Pathway

It is currently accepted that the peptide hydrolysis mechanism of pancreatic metallo-carboxypeptidases follows the **water-promoted pathway** (Christianson and Lipscomb, 1989, Kim and Lipscomb, 1990), where a pentacoordinated zinc ion polarizes a water molecule to attack the substrate, leading to a tetrahedral intermediate/transition state. The Zn^{2+} ion is coordinated by three residues, namely **His69** (N_1), **His196** (N_1), and **Glu72** (O_1 and O_2), which are contributed by turns and loops, that connect the secondary structures, and a water molecule. The Zn^{2+} ligands are arranged in a distorted tetrahedral co-ordination (Rees, 1983) (Figs. 18 and 19). The mechanism of peptide hydrolysis involves the activation of a water molecule by **Glu270** (procarboxypeptidase A numbering) and subsequent nucleophilic attack of the scissile bond by the hydroxyde (Matthews, 1988; Hanson, 1989). Positively charged residues, and the Zn^{2+} ion, assist the hydrolytic reaction by neutralising the developing negative charge of the tetrahedral intermediate, analogous to the oxyanion hole of the serine and cysteine proteinases (see Fig. 12). The residues important for substrate binding and catalysis are organized in several subsites based on the substrate residue that they accommodate (see Fig. 14).

S1': Asn144, Arg145, and Tyr248

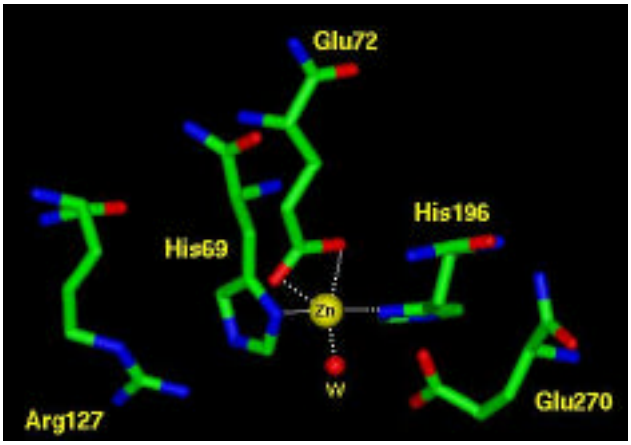
S1: Arg127 and Glu270

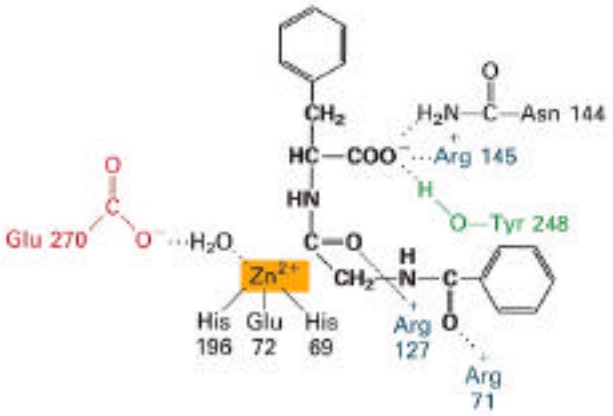
S2: Arg71, Ser197, Tyr198, and Ser199

S3: Phe279

S4: Glu122, Arg124, and Lys128.

The terminal carboxylate group of the peptide substrate is fixed by the side chains of Asn144, Arg145, and Tyr248, while the carbonyl group of the scissile peptide bond is positioned near the Glu270 carboxylate, Arg127, and the zinc ion (Fig. 19).





Figs. 18 and 19. The residues His69, His196 and Glu72 coordinate the zinc ion. The Zn^{2+} polarizes a coordinated water molecule (W) to attack the scissile peptide bond of the substrate. Glu270 and Arg127 assist this attack by, respectively, establishing a hydrogen bond with the water molecule (and abstracting its proton), and by stabilizing the oxyanion of the tetrahedral intermediate/transition state. The residues Arg145, Tyr248 and Asn144 (S1' pocket), are involved in the anchoring and neutralization of the C-terminal carboxylate of the substrate.

2.6) Insect Proteinases

Although the diversity of insects far exceedes that of mammalian systems, it is surprising the absence of sequence, and especially structural information, about insect enzymes. To date only two crystal structures of insect proteinases have been reported; that of a fire ant chymotrypsin and the procarboxypeptidase from *Helicoverpa armigera* presented here.

2.6.1) *Helicoverpa armigera* Taxonomy and Biology

Helicoverpa armigera (Hübner) belongs to the order Lepidoptera, family Noctuidae, being its common name cotton bollworm and it is one of the world´s most important agricultural pests. It is distributed along Africa, Middle East, southern Europe, India, central and southeastern Asia, Australia, New Zealand and eastern Pacific islands.

H. armigera and *H. punctifera* (Wallengreen) are major pests of cotton and other field crops in the major cotton-productive countries (Australia, Pakistan, India, China).

Fig. 20. Female lay eggs on the crop flowering and fruiting structures, where voracious larval feeding leads to substantial economic loss (Reed and Pawar, 1982).

2.6.2) Insect Pest Management

Crop losses due to arthropods, diseases, and weeds have increased world wide from 1965 to 1990 (Oerke, 1994) despite the intensification of pest control. The therapeutic approach of killing pest organisms with toxic chemicals (insecticides) has been the prevailing pest control strategy for over 50 years, replacing plant resistance and cultural practices (Quisenberry, 1994; Lewis, 1997).

The need for effective strategies in resistance management is becoming more pressing as the number of insecticide-resistant species continues to increase world wide while insecticide resources are diminishing (Georghiou, 1994). A reduction in the use of insecticides and an increased use of integrated insect management strategies is necessary. There is thus a need to integrate all appropriate control methods to maximize the effectiveness of pest regulation and also crop production.

The major trend has been toward the use of modern chemistry and molecular biology to replace traditional pesticides with less hazardous chemicals or nontoxic biologically based products; but these means are still therapeutics. New approaches have emerged: Biological Control, and Integrated Pest Management (IPM), for finding better pest control strategies.

Several pest management strategies have been used to control *H. armigera* without success. Organophosphates have been valuable insecticides to control the plague in Australia (Gunning et al., 2001). However, there is an emerging organophosphate-resistance threat *in H. armigera* due to cross-resistance between the compounds used. An insensitive acetylcholinesterase has been identified as the common resistance mechanism (Gunning et al., 1998).

There is thus the need for finding new effective, safe, biologically, ecologically, economically and feasible methods to control *H. armigera* pest.

III) Results

Article 1

Crystal Structure of the Caspase Activator Human Granzyme B, a Proteinase highly Specific for an Asp-P1 Residue

Introduction

The granzymes are trypsin-like serine proteinases localized in the dense cytoplasmic granules of cytotoxic T-lymphocytes (CTLs) and natural killer (NK) cells. These lymphocytes mediate the body's major defence against virus-infected and neoplastically transformed cells, and contribute to transplant rejection and autoimmune diseases (Darmon et al., 1995; Mullbacher et al., 1999). CTLs induce apoptosis via multiple independent mechanisms. CTL-mediated lysis can trigger an extrinsic pathway of apoptosis by calcium-dependent vectorial secretion of perforin and granzymes into the microenvironment between the effector and target cells (Tschopp and Hofmann, 1996). The currently most accepted model of granule-mediated killing by CTLs postulates that granzyme B (GzmB; EC 3.4.21.79) enters the target cell through perforin pores, inducing target cell death via two complementary pathways. The cytosolic pathway involves activation of pro-apoptotic caspases, while the nuclear pathway probably involves cell cycle regulatory proteins and/or kinase Cdc activation. It has also been suggested that GzmB directly triggers a post-caspase cytoplasmic apoptotic death pathway (Froelich et al., 1998).

The genes encoding GzmB, cathepsin G (CatG) and mast cell chymase are located on band q11.2 of chromosome 14 (Jenne and Tschopp, 1989; Smyth et al., 1996). Most granzymes are synthesized as prepro-enzymes and generally contain a two-residue pro-peptide. Upon removal of the Gly-Glu dipeptide by dipeptidyl peptidase I / cathepsin C, GzmB is activated to a 227 amino acid residue single-chain proteinase (Pham and Ley, 1999). GzmB is a very basic glycoprotein, with a pI of about 10; it has been proposed that positively charged patches on its surface could direct the proteinase to the nucleus of the target cell (Trapani et al., 1996). GzmB exhibits the unique specificity among serine

proteinases to cleave substrate peptide bonds after aspartyl residues; it shares this preference only with the caspases, cysteine proteinases that are essential for cytokine activation and apoptosis (Odake et al., 1991; Stennicke and Salvesen, 2000).

CTL or NK cells from GzmB knockout mice are unable to induce rapid DNA fragmentation in target cells *in vitro*, indicating that GzmB has a critical and non-redundant role for rapid induction of apoptosis (Heusel et al., 1994; Shresta et al., 1995). Furthermore, inhibition of GzmB with specific, highly reactive serine proteinase inhibitors (isocoumarins or tripeptide chloromethylketones) correlates with the inhibition of cytolysis mediated by rat NK cells (Kam et al., 2000; Trapani et al., 1996). GzmB can cleave and activate several pro-caspases (Yang et al., 1998) and is the most efficient activator of pro-caspases-3 and 7 *in vitro* (Stennicke and Salvesen, 1998; Zhou and Salvesen, 1997). Pro-caspase-3 appears to be a major physiological substrate of GzmB (Darmon et al., 1995; Quan et al., 1996). Pro-caspase-7, in turn, becomes susceptible to further activation by GzmB after the caspase-3-catalyzed release of its propeptide (Yang et al., 1998). Additionally, GzmB can cleave the inhibitor of an endonuclease (DEF45/ICAD) and the proapoptotic molecule Bid, and can thus initiate also apoptosis in caspase-3 deficient target cells (Medema et al., 1997).

The specificity of subsites S4 to S2' (with P1-Asp and P1'-Xaa flanking the scissile peptide bond; P1, P2, etc. and P1', P2', etc. designate peptide positions in N- and C-terminal direction from the scissile peptide bond, and S1, S2 etc. and S1', S2' etc. the cognate subsites on the proteinase) has been mapped for human and rat GzmB (Harris et al., 1998; Thornberry et al., 1997). Studies conducted with a positional scanning synthetic substrate library and a substrate-phage display library yielded Ile-Glu-Xaa-Asp↓Xaa-Gly (with Xaa denoting any residue) as the optimal substrate sequence for rat GzmB (Harris et al., 1998). This sequence also matches the activation sites of pro-caspases-3 and –10 and the cleavage sites in several other potential protein substrates of GzmB (Kam et al., 2000). The overall specificity profile of human GzmB is similar to that of the rat enzyme, with a much lower acceptance of glutamine residues at P3, however (Thornberry et al., 1997).

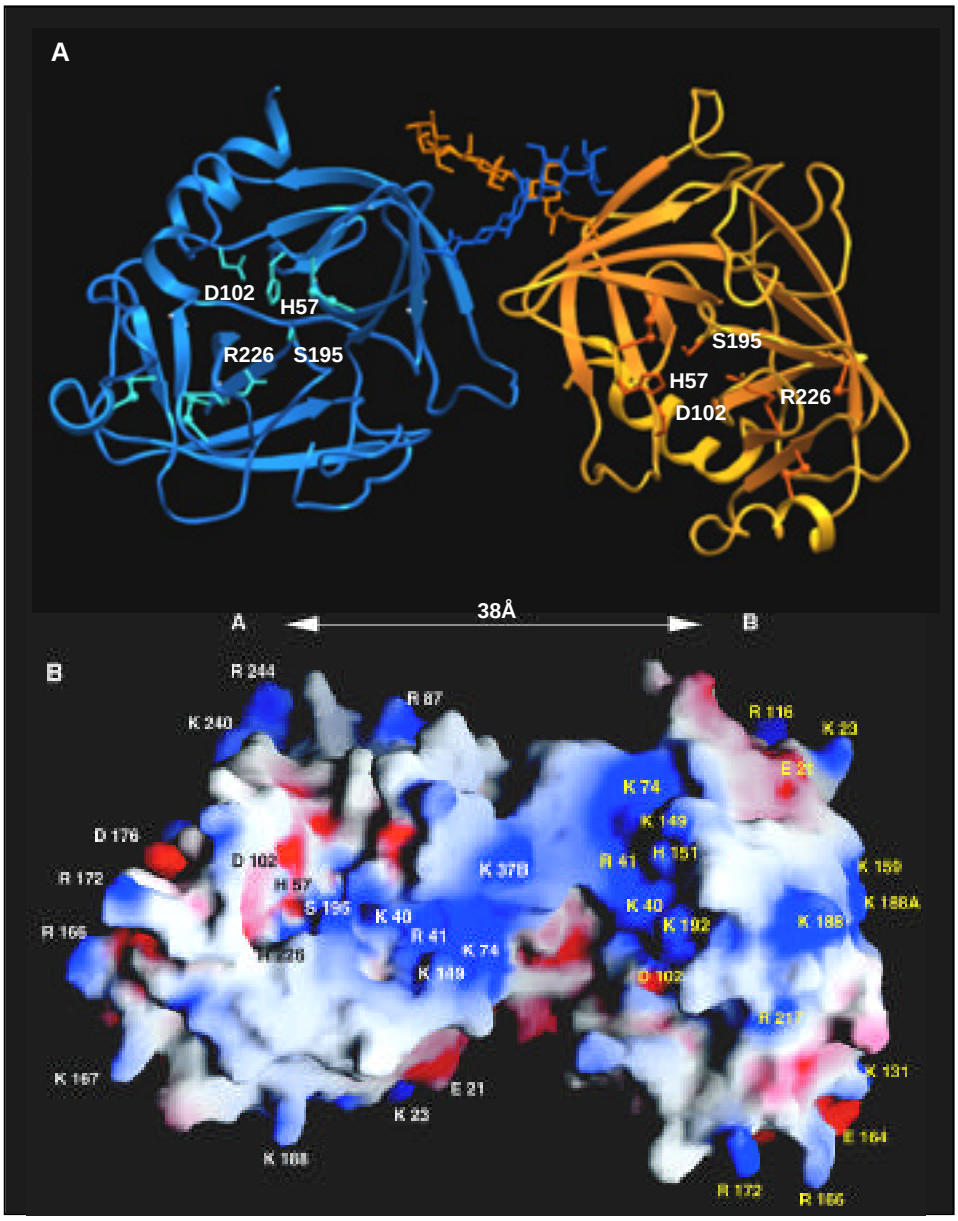
A potent endogenous inhibitor of GzmB, the intracellular serine proteinase inhibitor 9 (PI-9) (Bird et al., 1998), is primarily present in CTLs, probably protecting

them from committing fratricide or undergoing autolysis (Golstein, 1974; Kranz and Eisen, 1987; Nagler-Anderson et al., 1988). PI-9 possesses a glutamic acid residue at the P1 position, and mainly (Bird et al., 1998) but not exclusively (Annand et al., 1999) inhibits GzmB. In contrast, the cytokine response modifier A (CrmA), a serpin-type inhibitor from cowpox virus, has been shown to inhibit both GzmB and caspases, allowing the virus to manipulate the cell death program (Quan et al., 1995). Unlike PI-9, CrmA possesses an aspartate as P1 residue. Surprisingly, the extracellular serpin α_1-protease inhibitor has also been reported to inhibit GzmB (Poe et al., 1991).

GzmB has been implicated in the pathogenesis of rheumatoid arthritis (RA), as markedly elevated levels of soluble GzmB have been detected in plasma and in the synovial fluid of RA patients (Tak et al., 1999). The contribution of GzmB to joint damage in RA has been attributed to its ability to cleave aggrecan, the resident proteoglycan of cartilage (Froelich et al., 1993). Therefore, inhibitors against GzmB might prevent cartilage breakdown and joint destruction.

The GzmB structure has previously been modeled based on the structure of rat mast cell protease II (Murphy et al., 1988), but no experimental structure is as yet available. Until recently (Sun et al., 1999), only rat GzmB has been available as a recombinant protein (Caputo et al., 1994; Harris et al., 1998; Xia et al., 1998). To understand the unique specificity properties of GzmB, we have crystallized and determined the molecular structure of human GzmB. The availability of human recombinant GzmB produced in a baculovirus system has been essential for the elucidation of the current structure. A further crucial step has been the considerable improvement of the scattering power of the human GzmB crystals by controlled humidity manipulation (see Materials and Methods). Due to the accessiblility of the active-site regions of the constituent GzmB molecules, these crystals are ideal for soaking small inhibitors to investigate their binding mode aimed at designing selective inhibitors. Such inhibitors would help to further define the biological role of GzmB, and could also become drugs for medicinal applications.

Fig. 1. Crystal Structure of the Human Granzyme B Dimer



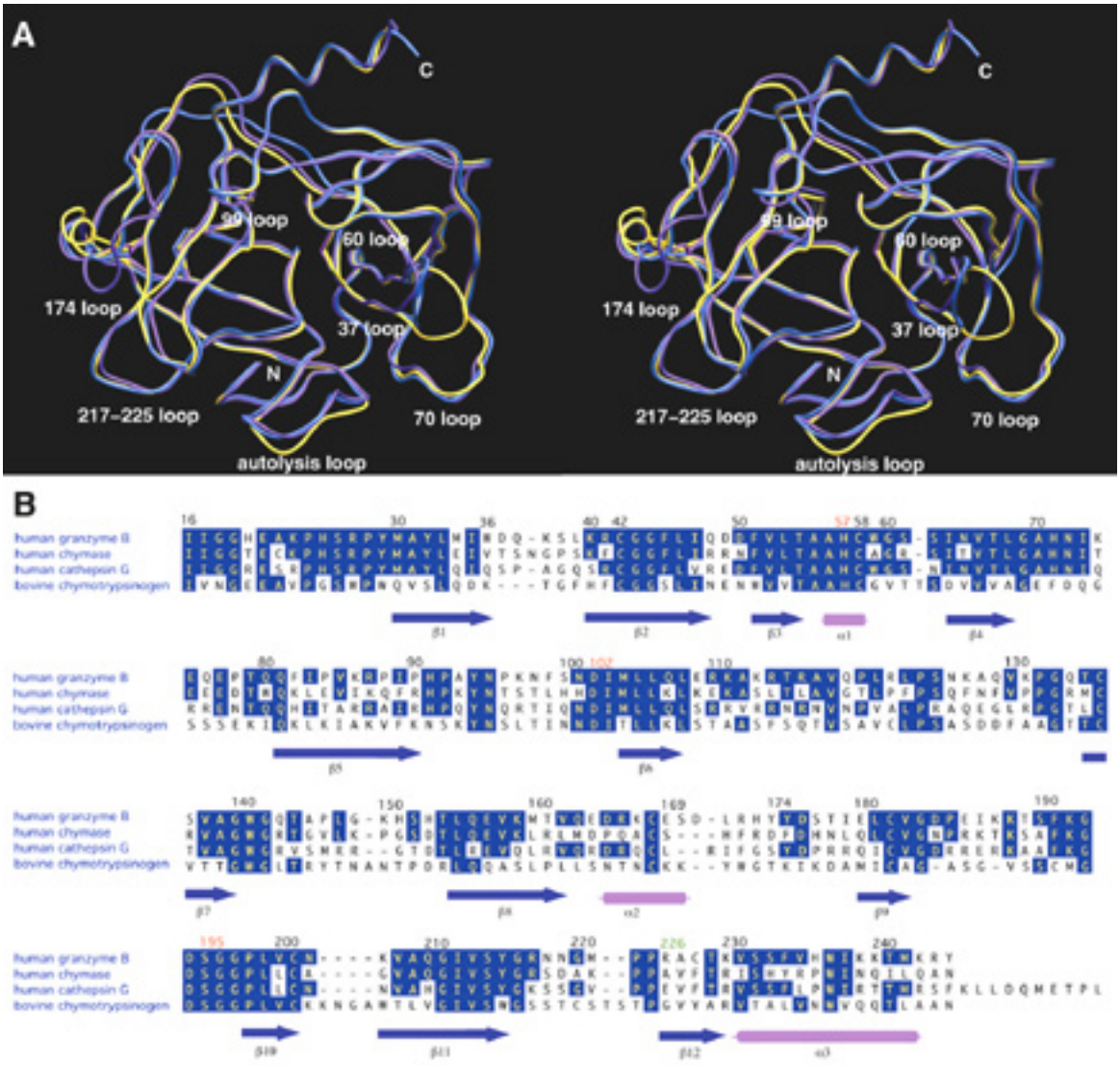
A. Ribbon representation of the GzmB molecules A (blue) and B (yellow). Molecule A is shown in the standard orientation. The carbohydrate chains attached to Asn65 are shown with all non-hydrogen atoms. The catalytic triad residues and the disulfide bridges are represented as blue and orange stick models. The figure was done with Setor (Evans, 1990).

B. Solid Surface Representation of the GzmB Dimer. The orientation is identical as in Fig. 1A. The colors indicate positive (blue) and negative (red) electrostatic potential at the molecular surface contoured at +10 kT/e to -10 kT/e. The figure was made with GRASP (Nicholls et al., 1993).

Fig. 2. (next page) Topological and Sequence Comparison with CatG and Human Chymase

A. Stereo ribbon plot of GzmB molecule B (violet) superimposed with human chymase (blue, (Pereira et al., 1999)) and cathepsin G (yellow, (Hof et al., 1996)). The proteinases are seen in the standard orientation. The figure was done with Setor (Evans, 1990).

B. Structure-based amino acid sequence alignment. The sequence of human GzmB (Caputo et al., 1988) 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 GzmB are indicated by arrows and cylinders, respectively. The catalytic triad residues are indicated with red numbers. The alignment was prepared with the DNASTAR software package (Madison, USA). A proper alignment with bovine chymotrypsinogen requires residue insertions at positions 37, 37A, 37B, 184A, 187A, and deletions at positions 62, 149 and 222-223.



Structure Determination and Crystal Packing

Human GzmB crystallizes as a network of open dimers, in which the two crystallographically independent molecules A and B are not related by a simple two-fold symmetry axis but by an improper screw (Fig. 1a). The active sites of both molecules project away from each other facing large intermolecular solvent spaces. The monomers are tightly packed via the 37 and the 70 loop (for loop nomenclature, see Fig. 2a), with surfaces of 920 $Å^2$ on each molecule removed from contact with bulk water; this number is typical for specific protein-protein interactions and contrasting with interfaces of 530 $Å^2$ or less between other neighbouring molecules. Approximately 40% of the A-B

interaction is mediated via the two equivalent carbohydrate chains *N*-linked to Asn65, which extensively interdigitate with one another but also make contacts with the opposite monomer surface (Figs. 1a and 3). The first Fourier map calculated with the truncated CatG model already displayed appropriate electron density accounting for the first sugar units. The current model includes two N-acetylglucosamine and one mannose unit (Fig. 3), in agreement with the typical structure of *N*-linked oligosaccharides. Faint density that would account for a fourth mannose unit in both molecules has been left uninterpreted. In both cases, only linear arrays of sugar units are visible.

In spite of different crystal packing, the overall structures of molecules A and B are quite similar, corresponding to an rms deviation of 0.58 Å for all 225 defined α-carbon atoms. Significant deviations are limited to the surface loops involved in dimer formation. Molecule B is slightly better defined by electron density, in particular its substrate binding site, so that it will mainly serve for the following structure description.

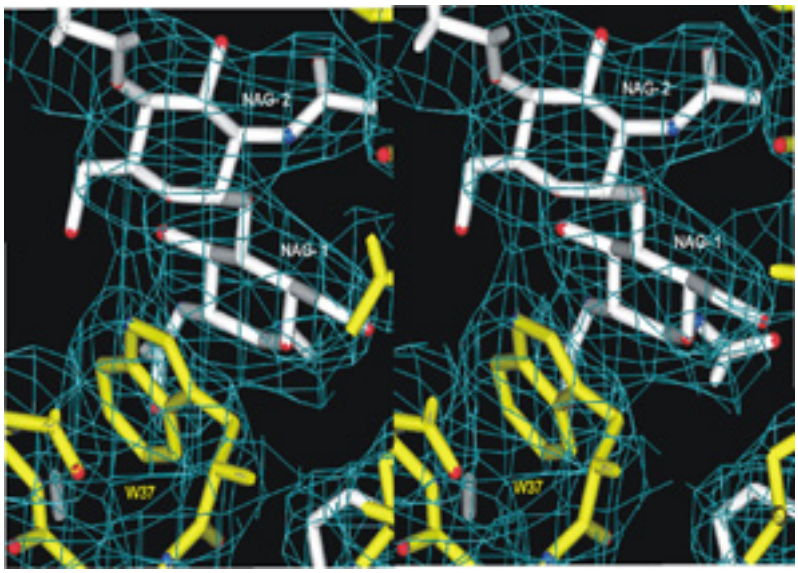


Fig. 3. Electron Density Omit Map Detail of the Dimer Interface Region. Section around Trp37 of molecule B and the first two N-acetyl-glucosamine units of molecule A superimposed with the final electron density calculated under omitting the Trp37 and the sugar units for phase calculation. Contour is at 1.0 σ . The figure has been made with MAIN (Turk, 1992).

Overall Structure of Granzyme B

GzmB resembles an oblate ellipsoid with diameters of about 35 and 50 Å. As in other chymotrypsin-like serine proteinases (Bode and Huber, 1986), the GzmB polypeptide chain is folded into two six stranded β-barrels strapped together by three trans domain segments. The surface is made up of several turn structures, a single turn α-helix from residues Ala56 to Cys58, a two-turn helix from Leu165 to Ser170, and the C-terminal mixed 3_{10}-α-helix (residue Val231 to Lys243) (Figs. 1 and 2). The catalytic residues Ser195, His57 and Asp102 (chymotrypsinogen numbering, see below) are located at the junction of the two β-barrels, and the active-site cleft runs perpendicular to this junction along the surfaces of both barrels.

An optimal superposition of the GzmB structure with those of CatG (Hof et al., 1996), human chymase (McGrath et al., 1997; Pereira et al., 1999), bovine chymotrypsin (Blevins and Tulinsky, 1985), and bovine trypsin (Bode and Schwager, 1975) results in 212, 214, 191 and 198 topologically equivalent α-carbon atoms with rms deviations of 0.74, 0.78, 0.90, and 0.87 Å, respectively, if only α-carbon atom pairs within 2 Å are included. Thus, topologically GzmB mostly resembles the chymases and CatG (Fig. 2a), in agreement with their common chromosomal location and proposed coevolution (Smyth et al., 1996). These close relationships are also reflected by the number of 119 and 126 identical GzmB residues topologically equivalent with human chymase and CatG, respectively, compared to 68 in case of bovine chymotrypsin. The topological equivalence with the latter proteinase forms the basis for the sequence alignment of the four serine proteinases shown in Figure 2b; the numbering of GzmB given is based on the equivalence with chymotrypsinogen (Hartley and Shotton, 1971).

Compared to other serine proteinases, the percentage of polar residues is not particularly high in GzmB. However, the 12 arginine, 21 lysine and 9 histidine residues are not balanced by the 9 glutamate and 9 aspartate residues, making GzmB a very basic molecule. Most charged side chains are, at least partially, exposed to the bulk solvent. Exceptions are the catalytic Asp102; Asp194, which forms an internal salt bridge to generate the functional S1 pocket and oxyanion hole, Arg226 in the S1 pocket (see below), and Glu180 and Asp50, which are located within the surface shell. The juxtaposition of Arg226, Arg217, Lys192, Lys40, His57 and Lys97 in and around the S1-pocket, charge compensated only by Asp102 and Glu180, gives rise to an overall positive

electrostatic potential (see Fig. 4). This contrasts with the strong negative potential of most other serine proteinases, in particular of CatG, at this surface site. In GzmB, this basic region extends towards the east into a larger positively charged surface patch made up mainly of the side chains of Arg41, Lys149, Lys74, Lys37B, Lys86, Arg87, Arg110, Lys111, Lys113, and Arg114 (Fig. 1b). The C-terminal helix also harbours a number of positively charged residues (His236, Lys239, Lys240, Lys243 and Arg244), which are clustered on the northern rim of the GzmB molecule. In contrast, negatively and positively charged side chains are clustered on the surface in the intermediate helix (to the west of molecule A, Fig. 1b) .

Loops Surrounding the Active Site

Although all chymotrypsin-like serine proteinases display a common fold, many of their distinct properties and in particular substrate specificity arise from the surface loops defining the environment of the active site. In the following description of GzmB, these loops (defined according to the residue number of their central residue) will be addressed anticlockwise with respect to the standard orientation displayed by molecule A in Fig. 1.

To the east of the active site (quarters denoted according to the standard orientation) the 37 insertion loop, a quite acute but well-defined β-hairpin, projects out of the molecular surface near the active-site cleft (see Fig. 2a). In molecule B, the external indole side chain of Trp37 stacks to the exposed surface of the 37 loop, further sandwiched by the carbohydrate chain attached to Asn65 from molecule A (Fig.1a). The equivalent Trp37 indole moiety of molecule A is disordered, presumably due to lack of any stabilizing intermolecular contacts.

Between Leu33 and Arg41, the 37 loop segment deviates considerably from the course observed for most other serine proteinases. As in CatG and human chymase, GzmB segment Leu39 to Arg41 is much more exposed, with the Arg41 side chain projecting into the active-site cleft as in CatG and partially obstructing it by forming a bump at its bottom. In contrast Phe41 of chymotrypsin is integrated in the upper wall of this cleft. Also in common with CatG and human chymase but different from the other serine proteinases, GzmB main chain segment 39-41 lines the cleft wall, with the Leu39

carbonyl group and the amino group of Arg41 offering additional hydrogen bonding partners for peptide substrates (see below).

North of and adjacent to the 37 loop resides the 60 loop, which in GzmB, similar to trypsin, CatG and human chymase, runs with an almost extended conformation along the molecular surface (Fig. 2a). The indole moiety of Trp59 extends into the gap between the 37 and the 99 loop (with different orientations in the two GzmB molecules, presumably due to different intermolecular contacts). The side chain of Asn65 and the attached oligosaccharide (see above) are directed away from the active-site cleft and should therefore not interfere with productively bound macromolecular substrates. As in most other chymotrypsin-like serine proteinases, the β-hairpin 99 loop of GzmB (His91-Phe99) protrudes from the north rim like a canopy (Fig. 4). This loop seems to be quite rigid and is presumably stabilized by Pro96. The 174 loop in the west approaches the active-site cleft at Leu171 and Tyr174, while the intervening dipeptide Arg172-His173 bends away with fully undefined side chains.

To the south of the 174 loop, segment 215-220 bulges out in a manner similar to chymotrypsin and virtually identical to that in CatG and in the chymases, but quite different to the shorter and more narrow loops of other trypsin-like proteinases. Similar to the related granule proteinases chymase and CatG, GzmB lacks cysteine residues at positions 191 and 220, which in most other chymotrypsin-like serine proteinases form a disulfide bridge. Instead, GzmB possesses the spatially adjacent residues Gly220 and Phe191 (also conserved in CatG), with the 191 phenyl ring approximately occupying the site of the disulfide bridge. The conformation and position of the active-site segment 189-192 remain largely unaffected in spite of this important structural difference, however. The segment Met221-Pro225 is two residues shorter compared with chymotrypsin, but identical with CatG and the chymases. As in these latter proteinases, the peptide bond between Pro224 and Pro225 is in a *cis* conformation.

The southern boundary of the active-site cleft of GzmB is formed mainly by the 149 or “autolysis” loop, which is fully defined in the electron density. This GzmB loop appears quite rigid and packs against the molecular body by looping around the side chain of Gln143. Within this loop, only the side chain of His151 is poorly defined by density, but the main chain path indicates that it must project into the S2’ subsite (see

below). The 70-80 loop of GzmB resides adjacent to the autolysis loop and behind the 37 loop. It is topologically similar to the calcium-binding loops of trypsin and related proteinases (Bode and Schwager, 1975). In contrast to the calcium-dependent proteinases, however, GzmB lacks the negatively charged residues at positions 70 and 80 that mediate the interaction with calcium. Instead, the calcium site of GzmB molecule B is (as in CatG and in the chymases) occupied by the carboxamide nitrogen of Gln80, which is in hydrogen-bonding distance to the carbonyl oxygens of Asn72 and Glu75 and to one of the carboxylate oxygen atoms of Glu77. Thus, the stability of GzmB should be independent of the presence of calcium. In GzmB molecule A the Gln80 side chain points out of the loop, reflecting its relatively flexible nature.

Active-site Cleft

The residues of the active-site triad, Ser195, His57, Asp102, and other catalytic elements such as the oxyanion hole are arranged at the center of the cleft exactly as in trypsin and chymotrypsin. The specificity pocket (S1) opening to the west of Ser195 is bordered by segments Val213-Gly220 (the entrance frame), Ser190-Ser195 (the basement), Pro225-Cys228 (the back of the pocket), and the Phe191 side chain (closing the pocket towards the south) (Fig. 4). The side chain of Arg226 projects from the 225-228 inner wall in a curved conformation into the S1 pocket, completely blocking access to the Thr189 side chain, an important specificity element in most serine proteinases. The orientation of the guanidyl group of Arg226 differs in GzmB molecules A and B, directed out of the pocket and backwards, respectively. Towards the back of the S1 pocket, this Arg226 side chain borders an apparently empty hydrophobic cavity lined also by the side chains of Val213, Val183, Leu199 and Val138, and the unique unpaired Cys228. Residue 228 replaces an almost strictly conserved aromatic side chain in most other chymotrypsin-like proteinases (e.g. a phenylalanine in CatG or a tyrosine in the chymases). Formation of an empty cavity would be very energy demanding. Therefore, this cavity is probably filled with a water molecule, which is not visible at the current resolution.

To the north of the S1 pocket, a shallow hydrophobic surface depression extends along the side chains of Tyr215 and Leu171, demarcated to the north by the peptide groups 174-175 and 97-98, and bounded to the west by the phenol side chain of Tyr174. Together with the carbonyl groups of Lys97, Asn98, and Tyr174 of the adjacent 174 loop, the Glu180 carboxylate group generates a negative surface potential in the northern end of this S4 subsite (Fig. 4b). To the east, the S4 depression is bordered by the benzyl side chain of Phe99, which together with the flat plane of the His57 imidazol ring demarcates the small S2 pocket. In contrast to most other serine proteinases, where the side chain of residue 41 (often phenylalanine) is buried in the interior of the molecule, the Arg41 side chain of GzmB points towards the active-site cleft. Together with the adjacent imidazole ring of His151, it forms a barrier in the cleft around the S2' subsite.

Probable Peptide Substrate Binding

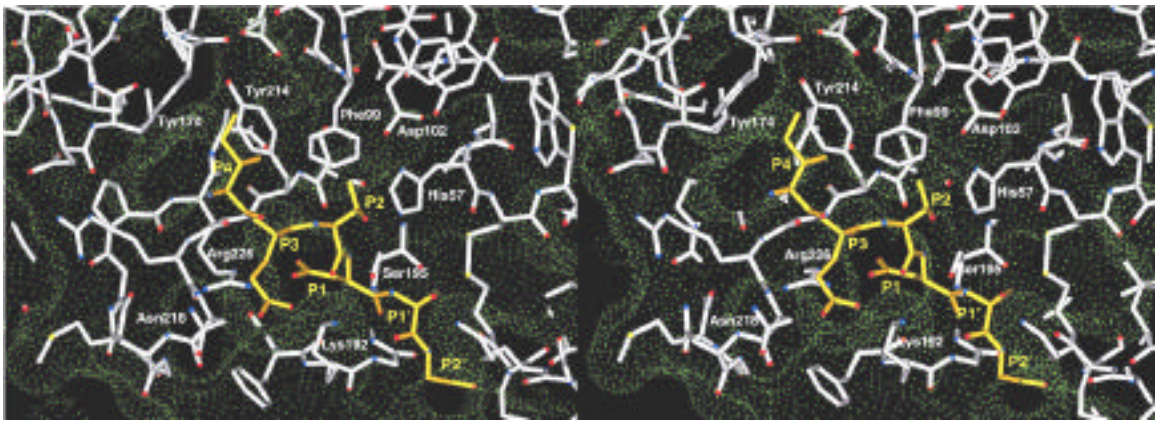
To explore the probable binding geometry of a peptide substrate and its interaction with the GzmB subsites, a productively bound hexapeptide spanning subsites S4 to S2' was modeled into GzmB molecule B (Fig. 4). The conformation of this peptide was derived from the "canonical" (Bode and Huber, 1992) reactive-site loop of turkey ovomucoid inhibitor third domain in complex with human leukocyte elastase (Bode et al., 1986). The chosen sequence Ile-Glu-Thr-Asp↓Ser-Gly (with the scissile bond between P1-Asp and P1'-Ser) corresponds to the interdomain linker activation site of pro-caspase-3 (Earnshaw et al., 1999) and is also in agreement with the experimentally determined preferences of rat (Harris et al., 1998) and human GzmB. As shown in Fig. 4, the N-terminal (P4 to P1) part of this hexapeptide juxtaposes GzmB segment Ser214-Arg217 in an antiparallel, slightly twisted manner, forming favourable inter-main chain hydrogen bonds between the amino and the carbonyl group of P3-Glu and Gly216 O and N, and between the amino group of P1-Asp and Ser214 O. Optimal main chain-main chain interactions would require an almost 45^{o} "up" rotation of the carbonyl groups of Gly216 and Ser214, however. Due to the special course of GzmB segment Ser38-Arg41, the C-terminal (P1' to P2') part of the hexapeptide can engage in a quite extended antiparallel interaction with Leu39-Arg41. In contrast to most other chymotrypsin-like serine proteinases, this not only allows the formation of a hydrogen bond between the P2'-Gly

amide nitrogen and Arg41 O, but an additional one between the P2'-Gly carbonyl group and Arg41 N (and, in a more extended substrate, another one between the P4' amide nitrogen and Leu39 O).

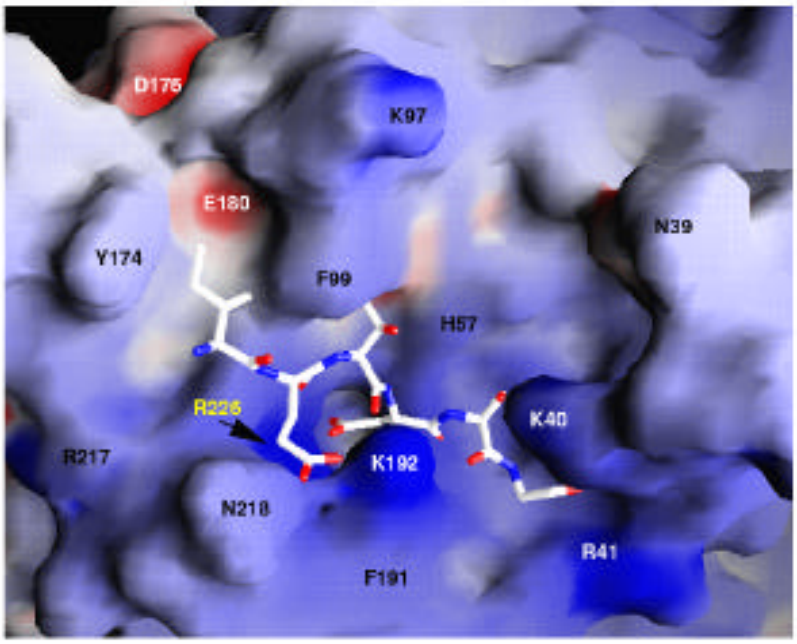
Fixed in this way to the GzmB active-site cleft, the Asp-P1 side chain would insert into the S1 pocket, sandwiched between segments Ser214-Gly216 and Ser190-Lys192. A slight collision of the distal Asp-P1 carboxylate group with the guanidinium group of Arg226 could easily be overcome by a minimal retreat of the Arg226 side chain into the back of the S1 pocket, with the guanidyl group pointing "down" (as in molecule B), anchored via Nη2 to Thr189 Oγ and/or the carbonyl groups of Pro224/Arg217, and facing the carboxyl group of the entering Asp-P1 side chain with its Nη1-Nε side (Fig. 4). This Asp-P1 carboxylate group could (together with Ser190 Oγ and Cys228 Sγ) further coordinate the internal water molecule that would occupy the cavity left around Cys228.

Fig. 4. Probable Interaction of the Ile-Glu-Thr-Asp_Ser-Gly Hexapeptide with Subsites S4 to S2'of GzmB.

A. GzmB is represented as a white stick model, superimposed with a light blue Connolly dot surface. The modeled hexapeptide is shown as a yellow stick structure. The GzmB molecule is shown in standard orientation, i.e. similar to molecule A in Fig. 1a. The figure has been made with MAIN.



B. GzmB is represented by a solid surface with the hexapeptide (white) bound to the active-site cleft. Coloring of the surface is made according to the calculated negative and positive electrostatic potential contoured from –15 kT/e (intense red) to +15 kT/e (intense blue). The figure was done with GRASP.



The S2 subsite of GzmB is a quite hydrophobic, medium-sized pocket lined by the aromatic side chains of Tyr215 and Phe99, and the flat side of the His57 imidazol ring (Fig. 4). For optimal main chain adaptation to the course of the active-site cleft, the bound substrate must attain a kinked polyproline II-like conformation at P2; coupled with a very favourable fit of the S2 pocket, a proline residue would be preferred at this position. Insertion of a threonine side chain, in contrast, requests some further opening of the S2 subsite. This could be brought about by a slight χ1-rotation of Phe99 so that the Cγ atom of the P2-Thr side chain can slip under the benzyl side chain, while the Oγ atom would point towards the bulk water (Fig. 4). The side chain of the P3 residue extends between the side chains of Lys192 and Asn218; a glutamic acid carboxylate group residue would be thus ideally placed to make favourable hydrogen bond contacts with both flanking side chains, in agreement with the strong preference of GzmB for Glu-P3 residues. In both GzmB molecules, the small hydrophobic S4 cleft is based on the side chains of Leu171 and Tyr215, bounded by the side chains of Phe99 and Tyr174.

Suitable placement of the P4-Ile side chain would require a considerable χ1 outwards rotation of the Tyr174 side chain, creating a favourable hydrophobic environment for the P4 side chain (Fig. 4).

On the primed side of the active site, the P1'-Ser side chain points into a large surface cavity mainly bounded by the His57 imidazole side chain, the Cys42-Cys58 disulfide bridge, the side chains of Ile36 and Lys40, and the carbonyl group of His57 (Fig. 4). Small polar side chains such as that of Ser-P1' could interact directly with the ammonium groups of the juxtaposing Lys40 and Lys192.

As a result of these manifold favourable intermolecular interactions, the hexapeptide scissile peptide bond between Asp-P1 and Ser-P1' would be ideally presented to the GzmB active-site, so that the polarized carbonyl could be attacked by the Ser195 Oγ leading to C-N cleavage.

Discussion

Several crystallographic problems had to be overcome to solve this structure of human GzmB. The crystals had to be dehydrated in order to improve their diffraction power. Most of the crystals were twinned, and all of them showed strong decay during X-ray exposure. Unfortunately, none of the cryo-conditions tested so far have been successful, so that two incomplete data sets collected under metastable humidity conditions had to be merged. In spite of the difficult data merging and still relatively low resolution, the electron density obtained after phasing with the truncated CatG model was surprisingly good and allowed the straightforward solution of the GzmB structure. Evaluation of the final model with PROCHECK (Laskowski et al., 1993) reveals that all stereochemical parameters are inside the limits expected for a crystal structure solved at this resolution, or even better. We are therefore confident that the crystal structure of human GzmB described in this paper is essentially correct, providing a structural framework to rationalize the unique substrate specificity exhibited by this granule serine proteinase.

In agreement with the common localization on chromosome 14 and anticipated co-evolution, GzmB shows close overall similarity to CatG and human chymase. Common features are the extremely basic character, the lack of the Cys191-Cys220

disulfide bridge, and an extended 37 loop which includes residues 39-41. GzmB shares the importance of residue 226 as the specificity determinant with CatG, where Glu226 determines the preference for P1-Lys as well as for P1-Phe side chains (see (Hof et al., 1996) for references). With respect to its charge distribution pattern, but in particular with regard to its substrate specificity and binding region, GzmB considerably differs from these related proteinases and displays unique recognition elements.

The unique capability of GzmB to cleave after an aspartate residue is based on its medium-sized basic S1 pocket. In the free enzyme, the anchoring Arg226 guanidyl group at its basement does not seem to be rigidly fixed, but exhibits some flexibility. The insertion of an Asp-P1 side chain would require rearrangement of the Arg226 side chain within the pocket, presumably leading to a hydrogen bond-connected side-on association of its guanidinium group with the P1-Asp carboxylate group. This electrostatic interaction explains the observed strong preference of GzmB for an aspartate at P1 in amide hydrolysis (Harris et al., 1998). The specificity pocket contrasts with the S1 pockets of the caspases that are quite accurately shaped and equipped to accommodate an aspartate side chain. The inferred variability of the GzmB S1 site might be enhanced by the lack of the Cys191-Cys220 disulfide bridge, which in most other chymotrypsin-like serine proteinases clamps the segments forming the basement of this pocket.

Such a plasticity of the S1-pocket might also explain the susceptibility of GzmB for the CTL cytoplasmic serpin PI-9, despite possessing a P1-Glu residue (Bird et al., 1998). PI-9 is non-reactive towards most of the caspases, making it a relatively specific GzmB inhibitor (Annand et al., 1999; Bird et al., 1998). A P1-Asp mutant, in contrast, is an effective caspase inhibitor, but surprisingly is 100-fold less efficient towards GzmB than wild-type PI-9 (Bird et al., 1998). This special serpin behaviour might make conceivable why viral CrmA, with a P1-Asp residue, is a potent caspase inhibitor but only a moderately efficient inhibitor of GzmB (Quan et al., 1995; Tewari et al., 1995). PI-9 has been found to react also with non-aspartate directed proteinases, utilizing other residues of its reactive-site loop, however (Dahlen et al., 1999).

The hydrophobic S2 groove of GzmB is shaped to accommodate small to medium-sized hydrophobic side chains of L-amino acids at P2, with a wide hydrophobic exit towards the bulk solvent to expose longer and more polar side chains, however. This

feature of the S2 pocket explains why GzmB can accept a broad range of amino acids in the P2 position, but still prefers proline over small polar or aliphatic L-amino acids (Harris et al., 1998). D-amino acids, which would extend towards the molecular surface, should not be accepted, in agreement with experimental findings (Harris et al., 1998; Thornberry et al., 1997). Due to replacement of the Ile99 side chain of rat GzmB by the bulkier Phe99 benzyl side chain, the S2 cleft of the human proteinase is significantly smaller, so that in particular a valine or a threonine at P2 would require some displacement of the 99 benzyl moiety to engage the hydrophobic Cγ (see Fig. 4); polar distal groups such as Oγ of threonine, can extend towards the bulk solvent.

The side chain of a P3 residue would run along the hydrophobic molecular surface of GzmB, pointing between the (partially adaptable) side chains of Lys192 and Asn218, whose terminal ammonium and carboxamide groups can become engaged in hydrogen bonds with extended polar/charged side chains at P3 (see Fig. 4). GzmB's strong preference for a glutamic acid residue at P3 seems to result not only from these two putative hydrogen-bonding partners, but also from the neighbouring Lys192 ammonium group and the overall quite positive electrostatic surface potential induced by the underlying Arg226 side chain (Fig. 4). Human GzmB possesses a more restricted specificity than the rat enzyme, clearly favouring glutamic acid at P3 (besides glycine) (Thornberry et al., 1997) probably due to the higher flexibility of the Lys192 side chain, which can more easily hydrogen bond a neighbouring carboxylate compared with the much more restricted Arg192 guanidyl group of the rat enzyme. The importance of the P3-Glu interaction with Lys/Arg192 for the presentation of the scissile peptide bond of a bound substrate is also shown by the drastically reduced hydrolytic efficiency of wild-type GzmB towards a Lys-P3 containing peptide substrate. The GzmB S3 subsite is in good agreement with the generous acceptance of L-alanine, serine or aspartate at this subsite (Thornberry et al., 1997).

The S4 subsite, which is quite narrow in the free GzmB molecule, must first be opened to allow insertion of the P4 side chain of a binding peptide substrate, resulting in an essentially hydrophobic surface cleft of approximately triangular shape (Fig. 4). Relatively rigid β-branched P4 side chains, such as those of isoleucine and valine, might be particularly well suited to displace the Tyr174 phenolic group. The isoleucine side

chain seems to be particularly well shaped to be packed in this hydrophobic cleft. This preference of GzmB for an isoleucine residue at the P4 position is reminiscent of the coagulation factors thrombin (Bode et al., 1992) and factor Xa (Padmanabhan et al., 1993), both of which prefer an isoleucine residue at P4. The small spot of negative electrostatic potential located towards the tip of the S4 triangle of GzmB (see Fig. 4b) is not adequately placed to favourably accommodate P4-Lys or Arg side chains. An equivalent cation hole in factor Xa has been utilized for the design of selective factor Xa inhibitors (Renatus et al., 1998).

The primed side of the GzmB active-site cleft is relatively narrow, but less well contoured (Fig. 4). The essentially hydrophobic cavity that represents the S1' subsite is quite voluminous and could easily accommodate the side chains of bulky P1' amino acids such as tryptophan, phenylalanine or leucine. Small P1' residues like serine could be packed such that their distal polar group contacts the flanking Lys40 (Fig. 4), with the remainder of the cavity (as in human chymase, (Pereira et al., 1999) filled by a bound water molecule. The accumulation of positively charged residues around this subsite provides an explanation for the observed aversion of GzmB towards arginine and lysine, but not towards aspartate and glutamate (Harris et al., 1998). In the case of an extended bound substrate (enforced by GzmB segment Leu39-Arg41), the side chain of the P2' residue would point "inside down" along the molecular surface (Fig. 4); because the S2' subsite of GzmB is flat and limited in size mainly by the (flexible) His151 imidazole ring, small P2' residues should be accommodated primarily, in agreement with GzmB's preference for glycine residues (Harris et al., 1998). Similar to human chymase (Pereira et al., 1999), a longer side chain of the following P3' residue could collide with the Lys40 side chain; again, a glycine at P2' would most likely allow to relieve these constraints, enabling also other P3' residues to bind in a productive manner.

Experimental data show that GzmB requires an extended interaction as well as an optimal side chain fit with the scissile peptide substrate for efficient catalysis (Harris et al., 1998). Such an extended specificity is typical for proteinases with regulatory functions. The primary sequence recognition required for catalysis of chromogenic synthetic peptides reflects the macromolecular specificity of GzmB. Thus, secondary and exosite interactions do not seem to play a decisive role in substrate recognition.

For instance, nuclear poly(ADP-ribose) polymerase, one of the major nuclear targets cleaved during apoptosis (Lazebnik et al., 1994), can be cleaved directly by GzmB, presumably at Val-Asp-Pro-Asp_Ser-Gly and/ or Leu-Glu-Ile-Asp_Tyr-Gly site (Harris et al., 1998).

GzmB shares its specificity profile, in particular the strong preference for Glu-P3, with the caspases-6, -8 and –9 (Thornberry et al., 1997), which activate the pro-caspases-3 and –7 (Earnshaw et al., 1999; Stennicke and Salvesen, 2000). The non-primed residues of the interdomain linker cleavage sites in these latter executioner caspases are Ile-Glu-Thr-Asp and Ile-Gln-Ala-Asp, respectively, i.e. match the specificity profile of GzmB. All mature caspases appear to be homodimers of heterodimers, consisting of two large and two small subunits separated from one another during activation (Wei et al., 2000). Currently, no pro-caspase structure is available, and their association state in solution is under debate, with evidence both in favour and against dimer formation. It is not even clear which of the large and small chains are covalently linked before activation. Several lines of evidence favour the view that, in spite of the more distant N- and C-termini, the hetero pair making up a common substrate binding region in the mature caspases forms a single chain in the pro-caspases, due to the more intimate interdigitation and because of an equivalent connection in the topologically related cysteine proteinase Arg-gingipain (Eichinger et al., 1999).

If pro-caspase-3 indeed would be such a dimer, the linker segments would be arranged on the same molecular surface in an antiparallel fashion, with the cleavage sites about 25 Å apart from one another. These two sites could be then simultaneusly attacked by the GzmB dimer where both active-site clefts run almost antiparallel to one another with active-center distances of about 38 Å (Fig. 1). It is thus tempting to speculate that the GzmB dimer observed here could act on pro-caspase-3 to cleave both linkers simultaneously. However, there are currently no indications for a stable GzmB dimer in solution (D.B. and H. R., unpublished results). The observation that glycosylated GzmB could dimerize under acidic conditions and at high protein concentration seeds the speculation that GzmB could be stored as a dimer in the granules of CTL and NK cells.