The EMBO Journal vol.7 no.8 pp.2593-2599, 1988

The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases

Wolfram Bode, Richard Engh, Djordje Musil, Ulrich Thiele, Robert Huber, Andrej Karshikov[1], Joze Brzin[2], Janko Kos[2] and Vito Turk[2]

Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG and [2]Department of Biochemistry, J.Stefan Institute, Jamova 39, YU-6100 Ljubljana, Yugoslavia

[1]On leave from Institute of Organic Chemistry,.Bulgarian Academy of Sciences, 1040 Sofia, Bulgaria

Communicated by R.Huber

The crystal structure of chicken egg white cystatin has been solved by X-ray diffraction methods using the multiple isomorphous replacement technique. Its structure has been refined to a crystallographic *R* value of 0.19 using X-ray data between 6 and 2.0 Å. The molecule consists mainly of a straight five-turn α-helix, a five-stranded antiparallel β-pleated sheet which is twisted and wrapped around the α-helix and an appending segment of partially α-helical geometry. The 'highly conserved' region from Gln53I to Gly57I implicated with binding to cysteine proteinases folds into a tight β-hairpin loop which on opposite sides is flanked by the amino-terminal segment and by a second hairpin loop made up of the similarly conserved segment Pro103I–Trp104I. These loops and the amino-terminal Gly9I–Ala10I form a wedge-shaped 'edge' which is quite complementary to the 'active site cleft' of papain. Docking experiments suggest a unique model for the interaction of cystatin and papain: according to it both hairpin loops of cystatin make major binding interactions with the highly conserved residues Gly23, Gln 19, Trp177 and Ala136 of papain in the neighbourhood of the reactive site Cys25; the amino-terminal segment Gly9I–Ala10I of bound cystatin is directed towards the substrate subsite S2, but in an inappropriate conformation and too far away to be attacked by the reactive site Cys25. As a consequence, the mechanism of the interaction between cysteine proteinases and their cystatin-like inhibitors seems to be fundamentally different from the 'standard mechanism' defined for serine proteinases and most of their protein inhibitors.

Key words: crystal structure/cystatin/cysteine proteinase inhibitor/docking/kininogens/proteinase–inhibitor complex

Introduction

The cystatins are tight and reversibly binding inhibitors of the papain-like cysteine proteinases. They form a superfamily of sequentially homologous proteins, of which chicken cystatin is representative (for reviews see, for example, Barrett *et al.*, 1986a; Barrett, 1987). On the basis of sequence homology and the presence and position of intradomainal disulphide connections the cystatin superfamily is subdivided into three families (Ohkubo *et al.*, 1984; Barrett *et al.*, 1986b; Turk *et al.*, 1986): (i) the stefins, which have M_r values of ~11 000, lack disulphide bonds and carbohydrate; (ii) the cystatins, which have M_r values of ~13 000, contain two intramolecular disulphide bonds, but lack any carbohydrate; and (iii) the kininogens, single chain plasma proteins, whose amino-terminal 'heavy chains' are constructed of three glycosylated cystatin-like domains with inhibitory activity (reviewed by Müller-Esterl *et al.*, 1986).

Chicken cystatin, a member of the second family, was first isolated from egg white (Fossum and Whitaker, 1968) but has also been found in other chicken tissues (see Barrett *et al.*, 1986a). Its amino acid sequence was independently determined by Turk *et al.* (1983) and Schwabe *et al.* (1984). Chicken cystatin has been resolved into two major fractions characterized by pI values of 6.5 and 5.6 and called forms A and B (Turk *et al.*, 1983) or 1 and 2 (Anastasi *et al.*, 1983). Each peak fraction can be further separated by hydrophobic chromatography (Thiele, 1986) into a long form of 116 amino acid residues starting with an amino-terminal serine (Ser-form, see Table I), and proteolytically degraded shorter forms, which mainly start with Gly9I (Gly-form, containing 108 amino acid residues; in the following, inhibitor residues are indicated with an I after the sequence number to distinguish them from those of the enzyme).

Chicken cystatin was shown to bind with 1:1 stoichiometry and high affinity to various papain-like proteinases from plants and mammalian tissues (reviewed by Barrett *et al.*, 1986a). The dissociation constant for complex formation between the long Ser-form and the papain is <5 pM (Nicklin and Barrett, 1984). Papain inhibition by a shorter form starting with Ala10I is ~10 000-fold weaker (Abrahamson *et al.*, 1987). A similar, also weaker binding has been determined for the Gly-form (W.Machleidt and U.Thiele, personal communication). For other cysteine proteinases this chain length dependence seems to be less strong (Thiele, 1986). Inhibition data obtained with other cystatin species further emphasize the importance of the amino-terminal segment for the interaction with target enzymes (Wakamatsu *et al.*, 1984; Isemura *et al.*, 1986; Lenarcic *et al.*, 1986; Samejima *et al.*, 1986). Abrahamson *et al.* (1987) present evidence that the Gly9I–Ala10I bond of excess (unbound) chicken cystatin is cleaved probably by an 'atypical' cysteine proteinase present in small quantities in papain preparations.

Peptide segment 53I–57I (chicken cystatin nomenclature) reads, in almost all members of the cystatin superfamily, Gln-Val-Val-Ala-Gly (QVVAG) while the chicken sequence is QLVSG (see Figure 2) and represents a rare exception. This conserved segment has been implicated in cystatin

Table I. Amino-terminal sequence of the long Ser-form of chicken egg white cystatin (Turk *et al.*, 1983; Schwabe *et al.*, 1984)

Ser1I	Glu2I	Asp3I	Arg4I	Ser5I	Arg6I	Leu7I	Leu8I	Gly9I	Ala10I

binding to target enzymes (Turk *et al.*, 1985). Indeed, weak inhibitory activity has been observed for synthetic QVVAG peptides (see Teno *et al.*, 1987) and for the second chicken cystatin cyanogen bromide fragment (W.Machleidt, personal communication).

Complex formation also occurs with cysteine proteinases having a carboxy-methylated reactive site cysteine. In cysteine proteinase–cystatin complexes this cysteine is, however, protected against chemical modification. Putting together all these observations Abrahamson *et al.* (1987) concluded that cystatin molecules might exhibit two distinct binding sites for their target enzymes: a substrate-like 'primary' site made up by the amino terminus with the Gly9I–Ala10I bond close to the reactive site cysteine upon complex formation; and a 'secondary' site comprising the 'conserved' QVVAG segment in contact with enzyme groups outside the active site.

In the absence of any data upon the three-dimensional architecture of the cystatin molecules such mechanistic concepts must necessarily remain speculative. In order to elucidate the polypeptide fold of this important and widespread protein family and to allow a reasonable modelling of its interaction with cysteine proteinases we have undertaken a crystal structure analysis of one representative. As previously reported (Bode *et al.*, 1985) we were able to grow large well-diffracting crystals of the short Gly-form of chicken egg white cystatin. Now we have determined and completely refined its crystal structure at high resolution. As subsequently shown, the molecular parts most highly conserved and implicated in binding to cysteine proteinases are adjacent in space and form an exposed, wedge-shaped 'edge' which sterically fits well into the papain reactive site cleft. In the following we present the structure of chicken cystatin and its most probable interaction with papain. A detailed account of the crystallographic analysis, the molecular conformation and the significance of the particular residues for integrity of the molecule and for binding to their target enzymes will be presented elsewhere.

Results and discussion

After extensive refinement, the polypeptide chain of chicken egg white cystatin can be traced unambiguously in the Fourier map from Gly9I to Asp85I and from Ala90I to the carboxy-terminal Gln116I. Between Asp85I and Ala90I there is no continuous electron density which would enable placement of the intervening segment with confidence. The defined electron density is in agreement with the published amino acid sequence of chicken cystatin (Turk *et al.*, 1983; Schwabe *et al.*, 1984).

Figure 1 shows the cystatin molecule seen along the crystallographic *x*-axis. The cystatin molecule has roughly the shape of a prolate ellipsoid, with major and minor axes of 45 and 30 Å, respectively. The molecular structure is dominated by regular structural elements (Figure 2). Five extended strands form an antiparallel, twisted β-pleated sheet, which is partially wrapped around a long, straight α-helix. Relatively separated in space from the main body (on the right-hand side of Figure 1) and aligned almost perpendicular to both the first α-helix and the β-strands is a second, shorter α-helical residue. The polypeptide chain is organized in the following way (Figure 2):

(i) an extended chain residue (A) from the amino-terminal Gly9I to Asp15I with approximate poly-L-proline II conformation;

(ii) a tight turn of type I between Asp15I and Asp18I;

(iii) a straight 3.6_{13} α-helical segment (I) of five turns from Asp18I to Ser36I, with some irregularities in hydrogen bonding in its carboxy-terminal half;

(iv) an open turn, with Asn37I and Tyr40I marking a major change in the polypeptide chain direction;

(v) an extended β-strand (B) from Tyr40I to Gln53I, which is engaged in ladder formation with segment Val12I–Val14I on one side and with segment Lys59I–Lys73I on the other side, but which contains two wide β-bulges which accentuate the right-handed twist of strand and β-sheet (see Figure 1);

(vi) a narrow five-residue β-hairpin loop between Gln53I and Lys59I, with the carbonyl group of Gly57I involved in intra-turn hydrogen bonds with the amide nitrogens of both preceding residues Ser56I and Val55I;

(vii) an extended β-strand (C) from Ile58I to Cys71I of regular β-geometry;

(viii) a slightly flexible tight turn of approximate type III from Pro72I to Ser75I;

(ix) a relatively regular α-helical segment (II) defined for 2.5 turns from Asp77I to Asp85I;

(x) a completely disordered segment from Gln86I to Ala90I;

(xi) an almost ideally extended β-strand (D) from Lys91I to Ile102I which forms regular hydrogen-bond bridges with segments Ile58I–Arg68I and Gln107I–Gln116I, respectively;

(xii) a five-residue B-hairpin loop between Ile102I and Gln107I which contains a type-I turn between Ile102I and Leu105I;

(xiii) an extended β-strand (E) from Gln107I to the carboxy-terminal Gln116I; in the crystal, this strand forms (between Leu111I and Cys115I) an antiparallel ladder with the identical strand of a symmetry-related molecule arranged around a crystallographic 2-fold axis thus giving rise to the formation of a 10-stranded β-pleated sheet in the crystal.

The polypeptide chain of chicken cystatin is covalently interconnected by two disulphide bridges which clamp both the second α-helix (Cys71I–Cys81I) and the carboxy terminus (Cys95I–Cys115I) to the β-pleated sheet. These disulphide bridges are of a left-handed and right-handed conformation, respectively, and are mainly buried in the molecule.

Val14I, segment Leu21I–Met29I of the first helix, and segments Val44I–Ala50I, Leu62I–Ile66I, Thr93I–Val99I and Ile108I–Cys115I of β-strands B–E participate through their side chains in the formation of a hydrophobic molecular core (Figure 1). Most of these hydrophobic residues are either conserved or substituted by residues of similar size in all hitherto known members of the cystatin family. This also indicates a similar polypeptide chain organization. A few hydrophobic residues are exposed to the solvent and form sites of irregularities of secondary structure.

Negatively charged side chains are somewhat clustered at the amino-terminal side of the first helix and around the second helix and its disordered appendix. In contrast, the basic residues are more concentrated in a belt between Lys51I and Lys109I and around the region linking the second α-helix and the main molecular body. This charge distribution gives rise to an electric moment with the positive

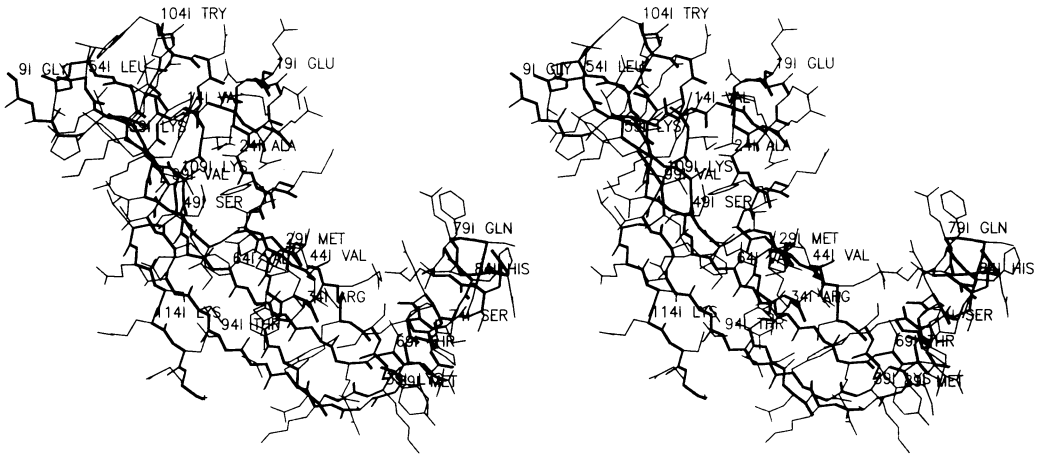

Fig. 1. Structure of chicken egg white cystatin with the polypeptide backbone represented by thick connections. Every 5th residue is labelled.

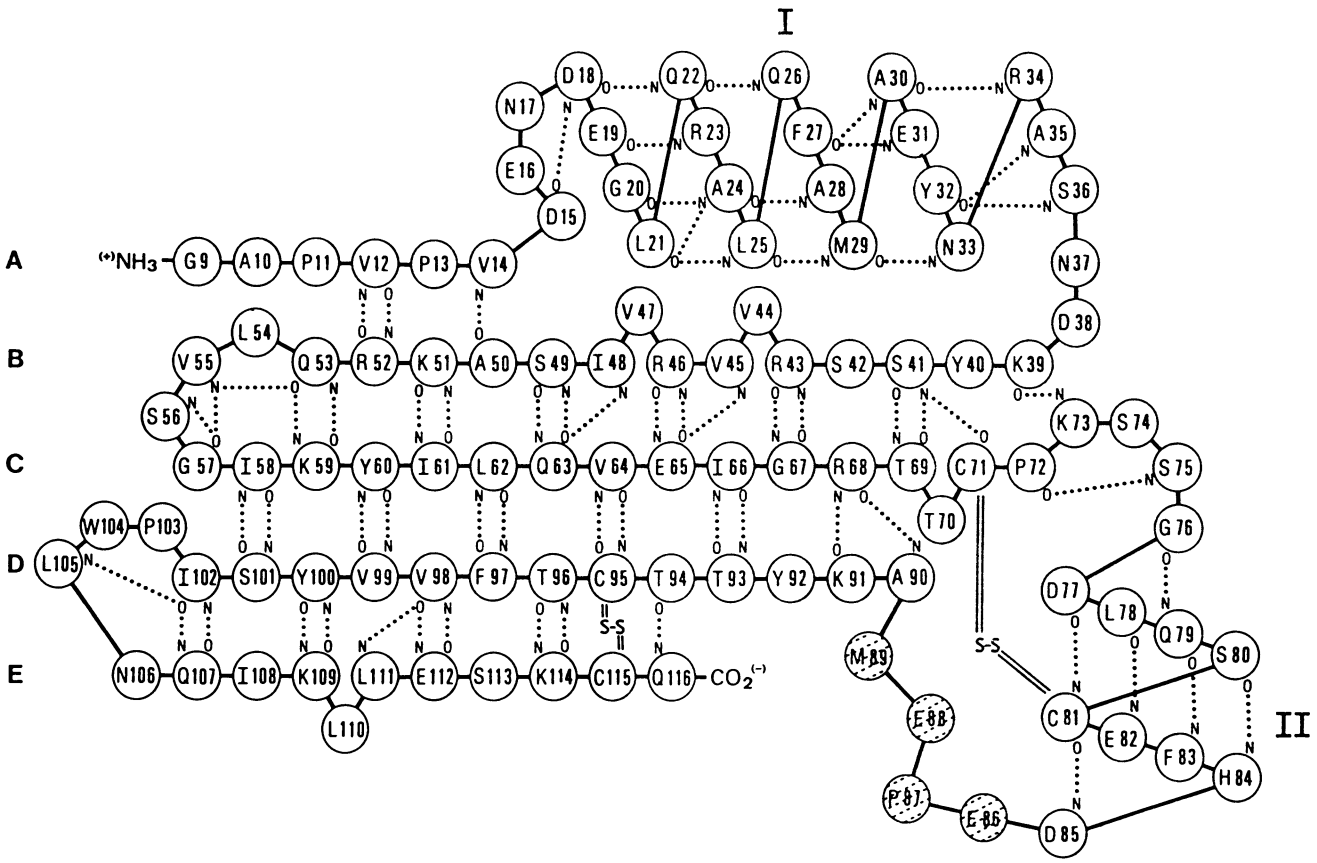

Fig. 2. Schematic representation of the polypeptide fold arrangement of chicken egg white cystatin (Gly-form). Main chain hydrogen bonds displayed by dashed lines were selected according to definitions (using energy cut-off values of below −0.5 kcal/mol) given by Kabsch and Sander (1983). Shadowed circles represent undefined residues. The two disulphide bridges are indicated by double lines.

pole directed towards loop Gln53I−Lys59I and the negative pole towards the second helix (to the left- and right-hand sides in Figure 1 and down and up in Figures 3 and 5). The electrostatic field is determined primarily by ionizable groups, with peptide groups altering it locally (Figure 3).

The amino-terminal segment is in van der Waals contact with the molecular body only from Ala10I onwards. Its Pro-X-Pro sequence induces a poly-proline-II-like conformation which is further stabilized by the hydrogen bond linkage to the β-sheet; this segment is expected to be inert to proteolytic attack. Gly9I is only weakly defined by electron density, possibly due to some amino-terminal heterogeneity detected in the crystalline material. Its position at the extreme corner of the β-sheet suggests that residues 1I to 8I protrude into solution and are relatively accessible to proteolytic attack. In the crystal there is much free intermolecular space around Gly9I which would allow the accommodation of an extended chain also in this crystal form.

The conserved β-hairpin loop segment Gln53I−Lys59I is located adjacent to the amino terminus. The hydrophobic chains of Leu54I, Val55I (and Ser56I) are exposed to the solvent, as are their carbonyl groups (Figures 1 and 3). Enhanced *B* values indicate that part of the loop should easily adapt to different environments. Gly57I is buried so that larger residues could not be accommodated at position 57I.

Of similar architecture and amphiphilic nature is the

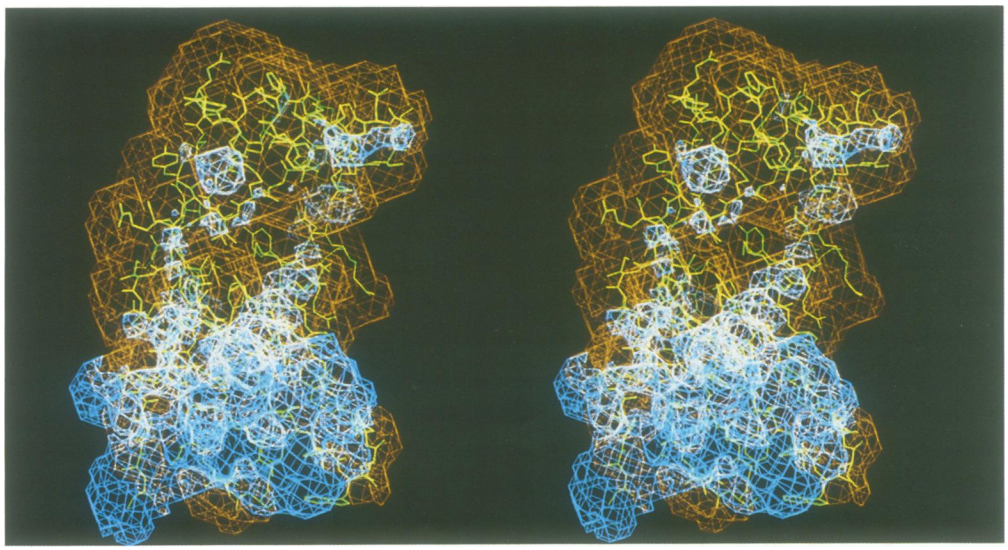

Fig. 3. Cystatin overlaid with electrostatic equipotential surfaces at +2 kcal/mol (blue) and −2 kcal/mol (red). The wedge part is formed by the lower third of cystatin in this view; the amino terminus is on the left, the Gln53I−Lys59I loop in the middle and the Ile102I−Gln107I loop on the right. The wedge is primarily positively charged; along the extreme edge the exposed carbonyl groups give rise to local regions of negative potential as well.

second 'conserved' β-hairpin loop Ile102I−Gln107I, placed alongside loop 53I−58I opposite to the amino-terminal segment (Figures 1 and 2). Again the quite hydrophobic side chains of four residues (Ile102I, Pro103I, Trp104I and Leu105I) are oriented towards the solvent, together with their respective exposed carbonyl groups (Figures 1 and 3). Pro103I and Trp104I are strongly conserved in cystatins and in the third domains of kininogens, emphasizing the importance of these residues for function. There is, however, no counterpart in the stefins, where the organization of the carboxy-terminal chain is obviously different.

The amino terminus and both β-hairpin loops form a contiguous, hydrophobic wedge. Several carbonyl groups protruding from the edge give rise to a restricted region of negative potential (see Figure 3). This 'edge region' is surrounded like a ruffle by a series of polar side chains, in particular those of Lys51I, Gln53I, Tyr100I, Gln107I, Lys109I (directed downwards in Figure 1) and Arg53I (directed upwards). These residues are relatively conserved in all cystatins and are well suited to provide anchoring points for polar interactions with target enzymes.

The four residues Gly9I, Gln53I, Val55I and Gly57I, conserved in all inhibitory cystatin domains elucidated so far (see Barrett *et al.*, 1986a), and the highly conserved residues 54I, 56I, 60I and 100I are all placed in this 'edge' region of the molecule. On the other hand, the more variable regions and the molecular parts present only in the second and third family (as, for example, the second α-helix) are located distant from it. We suggest, therefore, that this 'edge' is the contact region with the target enzymes. On this basis we undertook modelling experiments to dock cystatin with papain.

The reactive site residue Cys25 of papain [the papain numbering of Drenth *et al.* (1971) is used throughout in this paper] is situated at the bottom of a cleft formed between the two domains which essentially make up the papain structure (reviewed by Baker and Drenth, 1987). The walls of the active-site cleft are mainly formed by residues 64−67 on one side and residues 156−159 on the other (Figure 4). Also, the side chain of Trp177 which shields the hydrogen bond between the active site residues His159 (NE2) and Asn175 (OD1) from solvent is said to form a rear wall; the cleft seems further to extend in this direction (towards the back in Figure 4; towards the front in Figure 5). Cys25 resembles a saddle point which subdivides this cleft into two adjacent depressions, a relatively narrow and deep one (the foreground part of the cleft shown in Figure 4, comprising subsites S1 and S2) and a wide, open one (in front of Figure 5). This wide, more shallow part is essentially lined by residues 20−23 and 136−142, and its floor is mainly made up of the side chains of Trp177, Trp181, His159 and Gln19 (Figure 5). As further shown in Figures 4 and 5 by the specific colour coding of the dot surface most of the residues contributing to this more open part of the cleft are identical in papain and actinidin and are highly conserved (to a higher degree than the residues forming the narrow part) in all other papain-like enzymes known so far by their sequence.

As shown in Figures 4 and 5 the active-site cleft of papain is a replica of the wedge-shaped 'edge' of cystatin described above. Docking experiments as demonstrated in Figures 4 and 5 indeed show that cystatin fits almost ideally with its wedge into the papain cleft so that the second loop, around Trp104I, is in intimate contact with the hydrophobic 'floor' of the wide part of the papain cleft (Figure 5), whereas the amino terminus is directed towards its narrow part (in front of Figure 4).

On the basis of this principal model we have examined in detail several related docking models for the intermolecular interactions and have subjected them to energy minimization (which resulted in only small conformational adjustments). These models have the following common features:

(i) the amino-terminal segment is directed towards the narrow cleft, so that Gly9I is in the vicinity of the active site Cys25 of papain;

(ii) the first hairpin loop containing the conserved segment 53I−57I is in close proximity to the depression of the cleft between Cys25 and Trp177. Its hydrophobic side chains

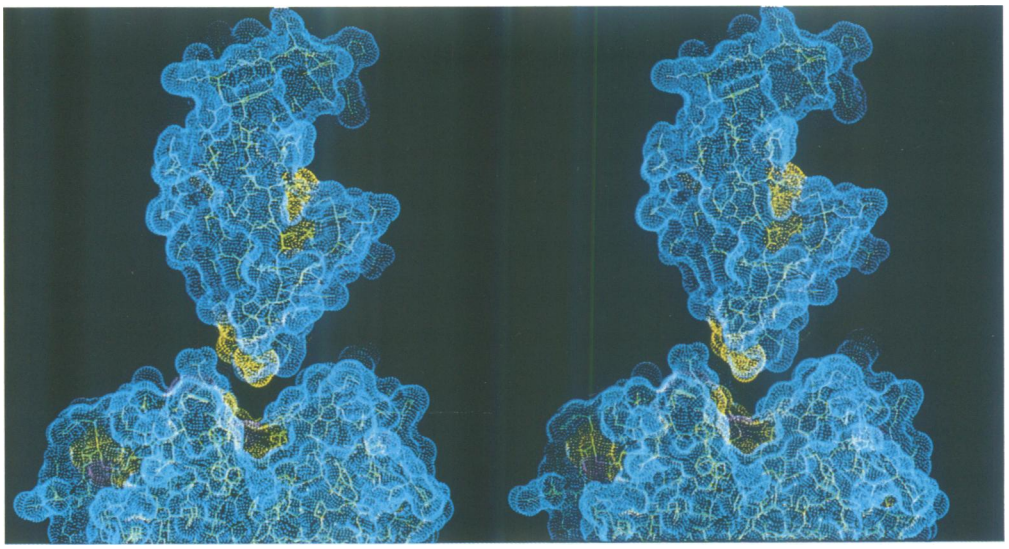

Fig. 4. Docking of cystatin (top) and papain (bottom) in a view from the amino-terminal side of the inhibitor along the edge of the cystatin wedge and the active site cleft of papain (front view); the inhibitor is slightly translated up out of the complex. Both components are displayed with Connolly dot surfaces (Richards, 1977; Connolly, 1985); yellow indicates surfaces of residues conserved in papain, actinidin, and rat cathepsins B and H (Takio *et al.*, 1983) according to the alignment of Barrett *et al.* (1986a); blue indicates less-conserved residues; cystein residues of papain are violet. Conserved residues visible with their surface in the active site cleft of papain are (from front to back) Gly65, Gly66, Ala160, Cys25 (violet), Gly23 and Gln19. The conserved cystatin residues Gly9I, Gln53I and Val55I forming part of the edge are likewise depicted with yellow surfaces; Leu54I (protruding from the edge to the right) varies within the cystatin superfamily but is always hydrophobic and of medium size. The papain coordinates used are those of Drenth *et al.* (1976) and Priestle *et al.* (1984) as deposited at the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977).

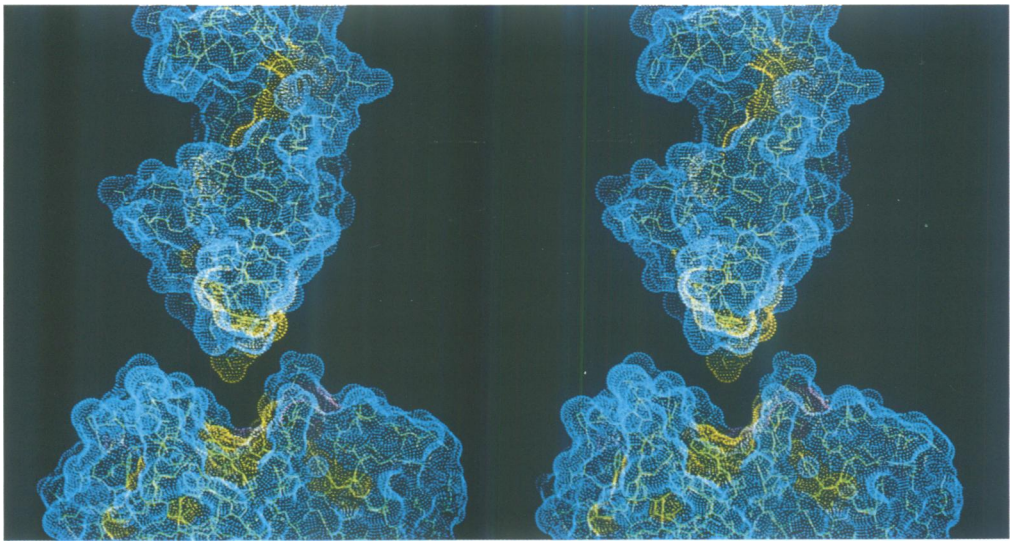

Fig. 5. Docking of cystatin (top) and papain (bottom), with the inhibitor slightly translated up out of the complex (back view); colour code as in Figure 4. Mainly the 'wide', backward part of the papain cleft is visible (from front to back) with Trp181, Trp177, His159, Gln19, Cys25 (violet) and Gly23 depicted as conserved residues. At the cystatin edge, Trp104I (front), Val55I (to the right), Gly9I (back) and Leu54I (to the left, with blue surface colour) are particularly exposed.

Leu54I and Val55I are in contact with the highly conserved hydrophobic side chains of Ala136 and Ala137 and of Gly23 of papain;

(iii) the second hairpin loop around Pro103I–Trp104I is accommodated in the 'wide' part of the cleft. The indole ring of Trp104I stacks on the side chain of Trp177 and lies edge-on with the indole ring of Trp181. These tryptophanes are totally conserved in all related cysteine proteinases (Kamphuis *et al.*, 1985; Ritonja *et al.*, 1988);

(iv) the electrostatic interactions between cystatin and papain are favourable in all models (before and after energy minimization); all polar groups buried in the interface fully exploit their hydrogen bonding capacity.

This model, which is based on structural grounds, explains most of the data known so far about complex formation:

(i) peptide segment Gly9I–Ala10I although close to Cys25 is differently arranged to that expected for the P1–P1′ segment of a papain substrate (Drenth *et al.*, 1976) and thus not suitably located to be cleaved within the complex;

(ii) the loose interaction of the amino-terminal segment leaves (water-filled) space around Cys25 to accommodate a carboxymethyl group but simultaneously (and in co-operation with the conserved loop) shields the reactive site residues from the bulk water and therefore towards the attack of modifying reagents within the complex;

(iii) Gly9I and (even more) Ala10I are only in weak

contact with papain residues, but are still 'buried' enough to be protected against cleavage by 'free' enzyme;

(iv) a preceding residue Leu8I of an elongated egg white cystatin form, in contrast, could bind through its hydrophobic side chain and polar main chain groups into the large hydrophobic S2-subsite and to the main chain amide groups of Gly66 (Drenth *et al.*, 1976) respectively, thus explaining enhanced binding of elongated cystatin forms;

(v) a segment Leu7I–Leu8I–Gly9I–Ala10I utilizing the substrate binding capacity of papain must form a tight turn. Such a conformation is enforced in kininogen domains 2 and 3, where residues 7I and 10I are disulphide-linked cysteines;

(vi) dimer formation through a cysteine at position 8I as observed for rat liver TPI (Wakamatsu *et al.*, 1984) and human stefin B (Lenarcic *et al.*, 1986) or its substitution with bulky groups would interfere with such an interaction at S2 and not allow an unconstrained formation of the complex;

(vii) more amino-terminal residues of elongated cystatin forms would in general run on or away from the papain surface, thus also enabling papain binding to the second and the third domain of intact kininogen chains;

(viii) the most intimate contacts are made through main and side chain groups of the 'conserved' cystatin segment with parts of the papain cleft of the S1′ putative subsite thus preventing binding and cleavage of small substrates;

(ix) the central position in the putative complex of the exposed indole ring of Trp104I makes its total conservation (together with Pro103I) in all cystatins and in kininogen third domains conceivable.

According to the model proposed, the mechanism of binding of cystatin inhibitors to papain-like proteinases is different from that observed for the inhibition of serine proteinases by most of the 'small' protein inhibitors which bind like substrates (Huber *et al.*, 1974; Huber and Bode, 1978; Laskowski and Kato, 1980). In the cystatins the interactions are made by two conserved inhibitor loops which have conformations quite different from a bound substrate, and groups of the enzyme remote from the reactive site Cys25. If elongated beyond Gly9I the amino-terminal segment might, however, utilize the specific subsite interactions of S2 and S3 of the enzyme with residues 8I and 7I. Gly9I and Ala10I, formally representing the P1 and the P1′ residues, are unable to approach the reactive site and cannot be cleaved witin the complex.

The model proposed exhibits several properties which can be experimentally tested, as, for example, by affinity studies with elongated chicken cystatin species with stepwise degraded amino termini. The exact interaction of cystatin-like inhibitors and cysteine proteinases remains, however, to be verified by X-ray diffraction studies of appropriate enzyme-inhibitor crystals.

Materials and methods

Chicken egg white was isolated from egg as described earlier (Turk *et al.*, 1983). Form A, the first peak fraction obtained with increasing NaCl gradient by FPLC (Pharmacia) on Mono Q in 20 mM Tris–HCl at pH 8.2, was used for crystallization. Prismatic crystals, of space group $P3_121$ (as determined later, see below) and cell dimensions $a = b = 47.9$ Å, $c = 87.5$ Å, $\alpha = \beta = 90°$, $\gamma = 120°$ containing one molecule per asymmetric unit were obtained as previously communicated (Bode *et al.*, 1985).

Most of the X-ray intensity data were collected on rotation/precession cameras (Huber, Rimsting, FRG) using graphite monochromatized CuK_α radiation from a rotating anode generator (Rigaku) operated at 4.5 kW. From six native crystals, rotated about the c^* and the a^* axes respectively, 26 500 reflections to 2.0 Å resolution were evaluated, using the FILME program (Schwager *et al.*, 1975) as modified by W.S.Bennett, Jr. These data were merged and scaled by means of the PROTEIN program package (Steigemann, 1974) yielding 6715 unique reflections representing 75% of all data expected to be 2.0 Å resolution (with 30% of all reflections expected in the last shell from 2.2 to 2.0 Å). The final *R*merge (defined as $\Sigma|I - \langle I \rangle|/\Sigma|I|$) is 0.105.

Four heavy-atom derivatives were prepared and used for phasing as follows:

(i) REC6: one crystal soaked for 50 days in 3 mM K_2ReCl_6 in 2.5 M potassium phosphate at pH 4.5.

(ii) PTC6: one crystal soaked for 24 h in 5 mM K_2PtCl_6 in 2.5 M potassium phosphate at pH 4.5.

(iii) MEPT: one crystal soaked for 21 days in 10 mM CH_3PtNO_3 in 2.5 M potassium phosphate at pH 4.5.

(iv) UOAC: one crystal soaked for 2 days in 1 mM uranyl acetate in 3 M ammonium sulphate at pH 3.75.

Data sets of the first three derivatives were collected on cameras, whereas the UOAC data set was (together with an additional 3 Å native data set) collected on a four-circle diffractometer (Siemens, FRG). Difference Patterson maps calculated with isomorphous and anomalous UOAC data respectively, yielded a single UOAC site, which was refined. With phases calculated from this single derivative (UOAC) a 3.5 Å Fourier map was obtained and further improved by truncating the electron density of the solvent region. Back-transformation and combination of these phases with the UOAC phases allowed determination of the correct enantiomorphic space group $P3_121$ (Hoeffken, 1987). The heavy-atom positions of the other three derivatives were derived from difference Patterson maps and cross-phased difference Fourier maps. Heavy-atom parameters and phases were refined with PROTEIN. The refinement was dominated by the UOAC derivative. The overall figure-of-merit for 1352 phases to 2.5 Å resolution was 0.58.

The 3.5 Å MIR Fourier map was improved by solvent flattening. It displayed large parts of the polypeptide chain of cystatin which were modelled on a PS 300 interactive display system (Evans and Sutherland) using the PSFRODO version (Pflugrath *et al.*, 1984) of FRODO (Jones, 1978). This model was refined and completed in a cyclic manner using the energy restraint crystallographic refinement EREF (Jack and Levitt, 1978) with a gradual extension of the data to 2.0 Å resolution. Fourier and difference Fourier maps were repetitively inspected at the graphics display to correct gross model errors. After each of the 12 macrocycles of refinement which consisted of model building, manual intervention and correction, and several (up to 10) cycles of EREF, the Fourier maps were calculated with combined phases obtained from isomorphous and model phases. Finally, Sim-weighted phases were used and individual *B* values refined. Besides 809 non-hydrogen protein atoms, 123 solvent molecules were located in the cell (the relatively large number resulting mainly from the fact that several concatenated density lobes in the 'disordered region' were interpreted as hydrogen-bonded solvent molecules). The current *R* value for 6181 reflections between 6.0 and 2.0 Å (defined as $\Sigma(|F_{obs}| - |F_{calc}|)/\Sigma|F_{obs}|$) is 0.198, the overall *B* value is 20 Å^2. The standard deviations from average bond length and angle values are 0.017 Å and 3.5° respectively.

Docking experiments were performed using the graphics programs HYDRA (R.E.Hubbard, University of York; Polygen Corp., Waltham, MA) and FRODO (Jones, 1978; Pflugrath *et al.*, 1984). Energy minimizations were done with GROMOS (W.F.van Gunsteren and H.J.C. Berendsen, BIOMOS B.V., Groningen, The Netherlands). Photographs and plots were prepared by means of HYDRA and PSFRODO.

The electrostatic interactions were calculated on the basis of the Kirkwood–Tanford theory (Tanford and Kirkwood, 1957) using the iterative procedure of Atanasov and Karshikov (1985). Peptide dipoles are taken into account as two appropriately positioned point charges. Dielectric constants of 2 and 78 were taken for regions inside and outside the protein. The differences in solvent exosure of the protein charges were taken into account according to their solvent accessibilities. The calculations are carried out for pH 7 and an ionic strength of 0.1 M.

Acknowledgements

We thank I.Mayr for skilful help with protein purification and crystallization, Dr F.Lottspeich for sequencing of the crystalline material, and A.-Z.Wei for help with data collection. The receipt of grants from the National Science Foundation, USA (R.E.), and from the Alexander-von-Humboldt-Stiftung (A.K.) is greatly appreciated. This work has been supported by the Research Council of Slovenia (grant C1-0515-106 to V.T.), by the Sonderforschungsbereich 207 of the Deutsche Forschungsgemeinschaft (grants H-1 and H-2) and by the International Bureau of the Kernforschungsanlage Jülich GmbH.

References

Abrahamson,M., Ritonja,A., Brown,M.A., Grubb,A., Machleidt,W. and Barrett,A.J. (1987) *J. Biol. Chem.*, **262**, 9688–9694.

Anastasi,A., Brown,M.A., Kembhavi,A.A., Nicklin,M.J.H., Sayers,C.A., Sunter,D.C. and Barrett,A.J. (1983) *Biochem. J.*, **211**, 129–138.

Atanasov,B.P. and Karshikov,A.D. (1985) *Studia Biophys.*, **105**, 11–22.

Baker,E.N. and Drenth,J. (1987) In Jurnak,F. and McPherson,A. (eds), *Biological Macromolecules and Assemblies*. John Wiley and Sons, New York, Vol. 3, pp. 313–368.

Barrett,A.J. (1987) *Trends Biochem. Sci.*, **12**, 193–196.

Barrett,A.J., Rawlings,N.D., Davies,M.E., Machleidt,W., Salvesen,G. and Turk,V. (1986a) In Barrett,A.J. and Salvesen,G. (eds), *Proteinase Inhibitors*. Elsevier, Amsterdam, pp. 515–569.

Barrett,A.J., Fritz,H., Grubb,A., Isemura,S., Järvinen,M., Katunuma,N., Machleidt,SW., Müller-Esterl,W., Sasaki,M. and Turk,V. (1986b) *Biochem. J.*, **236**, 311–312.

Bode,W., Brzin,J. and Turk,V. (1985) *J. Mol. Biol.*, **181**, 331–332.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Connolly,M.L. (1983) *Science*, **221**, 709.

Drenth,J., Jansonius,J.N., Kockoek,R. and Wolthers,B.G. (1971) In Boyer,P.D. (ed.), *The Enzymes*. Academic Press, New York, Vol. 3, pp. 485–499.

Drenth,J., Kalk,K.H. and Swen,H.M. (1976) *Biochemistry*, **15**, 3731–3738.

Fossum,K. and Whitaker,J.R. (1968) *Arch. Biochem. Biophys.*, **125**, 367–375.

Höffken,W. (1987) Dissertation, TU München, FRG.

Huber,R. and Bode,W. (1978) *Acc. Chem. Res.*, **11**, 114–122.

Huber,R., Kukla,D., Bode,W., Schwager,P., Bartels,K., Deisenhofer,J. and Steigemann,W. (1974) *J. Mol. Biol.*, **89**, 73–101.

Isemura,S., Saitoh,E., Sanada,K., Isemura,M. and Ito,S. (1986) In Turk,V. (ed.), *Cysteine Proteinases and Their Inhibitors*. Walter de Gruyter, Berlin, pp. 497–505.

Jack,A. and Levitt,M. (1978) *Acta Crystallogr., A*, **34**, 931–935.

Jones,A. (1978) *J. Appl. Crystallogr.*, **11**, 268–272.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Kamphuis,I.G., Drenth,J. and Baker,E.N. (1985) *J. Mol. Biol.*, **182**, 317–329.

Laskowski,M., Jr and Kato,I. (1980) *Annu. Rev. Biochem.*, **49**, 593–626.

Lenarcic,B., Ritonja,A., Sali,A., Kotnik,M., Turk,V. and Machleidt,W. (1986) In Turk,V. (ed.), *Cysteine Proteinases and Their Inhibitors*. Walter de Gruyter, Berlin, pp. 473–487.

Müller-Esterl,W., Iwanaga,S. and Nakanishi,S. (1986) *Trends Biochem. Sci.*, **11**, 336–339.

Nicklin,M.J.H. and Barrett,A.J. (1984) *Biochem. J.*, **223**, 245–253.

Ohkubo,I., Kurachi,K., Takasawa,T., Shiokawa,H. and Sasaki,M. (1984) *Biochemistry*, **23**, 5691–5697.

Pflugrath,J.W., Saper,M.A. and Quiocho,F.A. (1984) In Hall,S. and Ashiaka,T. (eds), *Methods and Application in Crystallographic Computing*. Clarendon Press, Oxford, p. 407.

Priestle,J.P., Ford,G.C., Glor,M., Mehler,E.L., Smit,J.D.G., Thaller,C. and Jansonius,J.N. (1984) *Acta Crystallogr., A*, **40** (suppl.), Hamburg Congress Abstracts C-17.

Richards,F.M. (1977) *Annu. Rev. Biophys. Bioeng.*, **6**, 151.

Ritonja,A., Popovic,T., Kotnik,M., Machleidt,W. and Turk,V. (1988) *FEBS Lett.*, **228**, 341–345.

Samejima,T., Kaji,H. and Takeda,A. (1986) In Turk,V. (ed.), *Cysteine Proteinases and Their Inhibitors*. Walter de Gruyter, Berlin, pp. 561–567.

Schwabe,C., Anastasi,A., Crow,H., McDonald,J.K. and Barrett,A.J. (1984) *Biochem. J.*, **217**, 813–817.

Schwager,P., Bartels,K. and Jones,A. (1975) *J. Appl. Crystallogr.*, **8**, 275–280.

Steigemann,W. (1974) Dissertation, TU München, FRG.

Takio,K., Towatari,T., Katunuma,N., Teller,D.C. and Titani,K. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3666–3670.

Tanford,C. and Kirkwood,J.G. (1957) *J. Am. Chem. Soc.*, **79**, 5333–5339.

Teno,N., Tsuboi,S., Itoh,N., Okamoto,H. and Okada,Y. (1987) *Biochem. Biophys. Res. Commun.*, **143**, 749–752.

Thiele,U. (1986) Diplomarbeit, Universität München, FRG.

Turk,V., Brzin,J., Longer,M., Ritonja,A., Eropkin,M., Borchart,W. and Machleidt,W. (1983) *Hoppe-Seyler's Z. Physiol. Chem.*, **364**, 1487–1496.

Turk,V., Brzin,J., Lenarcic,B., Locnikar,P., Popovic,T., Ritonja,A., Babnik,J., Bode,W. and Machleidt,W. (1985) In Khairallah,E. and Bonds,J. (eds.), *Intracellular Protein Catabolism V*. Alan R.Liss, New York, pp. 91–103.

Turk,V., Brzin,J., Kotnik,M., Lenarcic,B., Popovic,T., Ritonja,A., Trstenjak,M., Begic-Odobasic,L. and Machleidt,W. (1986) *Biomed. Biochim. Acta*, **45**, 1375–1384.

Wakamatsu,N., Kominami,E., Takoi,K. and Katunuma,N. (1984) *J. Biol. Chem.*, **259**, 13832–13838.

Received on April 11, 1988; revised on May 16, 1988