Biochem. J. (1993) 291, 123–129 (Printed in Great Britain)

Importance of the evolutionarily conserved glycine residue in the N-terminal region of human cystatin C (Gly-11) for cysteine endopeptidase inhibition

Anders HALL,* Henrik DALBØGE,† Anders GRUBB* and Magnus ABRAHAMSON*‡

*Department of Clinical Chemistry, University Hospital, S-221 85 Lund, Sweden, and †Novo-Nordisk A/S, Novo Allé, DK-2880 Bagsvœrd, Denmark

Human cystatin C variants in which the evolutionarily conserved Gly-11 residue has been replaced by residues with positively charged (Arg), negatively charged (Glu), bulky hydrophobic (Trp), or small (Ser or Ala) side-chains have been produced by site-directed mutagenesis and expression in *Escherichia coli*. The five variants were isolated and structurally verified. Their inhibitory properties were compared with those of wild-type recombinant cystatin C by determination of the equilibrium constants for dissociation (K_i) of their complexes with the cysteine endopeptidases papain and human cathepsin B and with the cysteine exopeptidase dipeptidyl peptidase I. The Ser-11 and Ala-11 cystatin C variants displayed K_i values for the two endopeptidases that were approx. 20-fold higher than those of wild-type cystatin C, while the corresponding values for the Trp-11, Arg-11 and Glu-11 variants were increased by a factor of about 2000. In contrast, the K_i values for the interactions of all five variants with the exopeptidase differed from that of wild-type cystatin C by a factor of less than 10. Wild-type cystatin C and the Ser-11, Ala-11 and Glu-11 variants were incubated with neutrophil elastase, which in all cases resulted in the rapid hydrolysis of a single peptide bond, between amino acid residues 10 and 11. The K_i values for the interactions with papain of these three N-terminal-decapeptide-lacking cystatin C variants were 20–50 nM, just one order of magnitude higher than the value for N-terminally truncated wild-type cystatin C, which in turn was similar to the corresponding values for the full-length Glu-11, Arg-11 and Trp-11 variants. These data indicate that the crucial feature of the conserved Gly residue in position 11 of wild-type cystatin C is that this residue, devoid of a side-chain, will allow the N-terminal segment of cystatin C to adopt a conformation suitable for interaction with the substrate-binding pockets of cysteine endopeptidases, resulting in high-affinity binding and efficient inhibition. The functional properties of the remaining part of the proteinase contact area, which is built from more C-terminal inhibitor segments, are not significantly affected even when amino acids with bulky or charged side-chains replace the Gly-11 residue of the N-terminal segment.

INTRODUCTION

Human cystatin C is a small protein containing 120 amino acid residues in a single polypeptide chain (Grubb and Löfberg, 1982; Abrahamson et al., 1987a). The protein is a cysteine proteinase inhibitor and forms, together with cystatins S, SN, SA and D, the human group in Family 2 of the cystatin superfamily (Barrett et al., 1986a,b; Freije et al., 1991). Of all human cystatins known, including the Family 1 cystatins, cystatin A and B, and the kininogens (Family 3), cystatin C is the tightest-binding inhibitor of the model cysteine proteinase papain, and also of the human cysteine proteinases cathepsins B, H and L. The equilibrium constants for dissociation (K_i) of cystatin C complexes with these enzymes are all lower than 10^{-9} M (Barrett et al., 1984). Since cystatin C is synthesized in most tissues of the body and is widely distributed in body fluids (Abrahamson et al., 1986, 1990), and in addition rapidly forms complexes with all three cathepsins, it appears to be one of the most physiologically important inhibitors of extracellular cysteine proteinase activity in man (Abrahamson et al., 1986, 1991).

Cystatin C forms stoichiometric 1:1 complexes with cysteine proteinases, in competition with the substrate (Abrahamson et al., 1987b). The inhibitor's mode of interaction with target proteinases has not yet been visualized by X-ray crystallography or n.m.r., but functional studies have indicated an important role for amino acid residues on the N-terminal side of the Gly-11 residue in the inhibitory activity against cysteine endopeptidases (i.e. papain, cathepsin B and cathepsin L) (Abrahamson et al., 1987b, 1991). This glycine residue is conserved in all cystatins, and crystallographic studies of the avian Family 2 cystatin, chicken cystatin, indicate that this residue is positioned close to the S_1 substrate-binding subsite when the cystatin forms a complex with papain (Bode et al., 1988). The residues preceding Gly-11 in cystatin C, i.e. Arg-8, Leu-9 and Val-10, probably contribute to enzyme binding through side-chain interactions with the substrate-binding subsites S_4, S_3 and S_2 respectively of papain and cathepsin B (Hall et al., 1992). Corresponding results have been obtained for chicken cystatin, the N-terminal segment of which is important for high-affinity complex formation with papain, with the residues Leu-9 and Leu-10 (cystatin C numbering) playing particularly prominent roles (Abrahamson et al., 1987b; Machleidt et al., 1989). Assuming a common mode of interaction with target proteinases for all cystatins, two more regions of the cystatin C molecule could be proposed to be involved in proteinase binding from sequence similarity studies and X-ray cystallography data for chicken cystatin and a cystatin B–papain complex; the Gln-55–Gly-59 and Pro-105–Trp-106 segments (Abrahamson et al., 1987b; Bode et al., 1988; Stubbs et al., 1990). Direct evidence that the Trp residue in the latter region is involved in proteinase complex formation has also been obtained by fluorescence studies (Lindahl et al., 1992).

The purpose of the present investigation was to elucidate the

Abbreviations used: NHMec, 7-(4-methyl)coumarylamide; pNA, *p*-nitroanilide; Bz, benzoyl; Cbz, benzoyloxycarbonyl; E-64, L-3-carboxy-2,3,-*trans*-epoxypropionyl-leucylamido-(4-guanidino)butane; G11S-, G11A-, G11R-, G11E- and G11W-cystatin C, cystatin C variants in which Gly-11 is replaced with Ser, Ala, Arg, Glu and Trp respectively.

‡ To whom correspondence should be addressed.

importance of the evolutionarily conserved glycine residue in the N-terminal segments of cystatins, by studying the functional effects of amino-acid substitutions of the cystatin C Gly-11 residue. The enzymes chosen for the inhibition studies were papain and human cathepsin B, two well-known cysteine endopeptidases. The former is very tightly bound by cystatin C (K_d 1.1×10^{-14} M; Lindahl et al., 1992) and the latter has a moderately high affinity for the inhibitor (K_i 2.5×10^{-10} M; Barrett et al., 1984). In addition, the aminopeptidase dipeptidyl peptidase I (K_i 3.5×10^{-9} M; Barrett et al., 1984) was used as a representative cysteine exopeptidase.

MATERIALS AND METHODS

Materials

The synthetic substrates Bz-DL-Arg-pNA, Cbz-Phe-Arg-NHMec and H-Gly-Phe-NHMec (where Bz is benzoyl, Cbz is benzoyloxycarbonyl, pNA is *p*-nitroanilide and NHMec is 7-(4-methyl)coumarylamide) were obtained from Bachem Feinchemikalien, Bubendorf, Switzerland. L-3-Carboxy-2,3-*trans*-epoxypropionyl-leucylamido-(4-guanidino)butane (E-64) was from Sigma Chemical Co., St. Louis, MO, U.S.A. Papain (EC 3.4.22.2) isolated from crude papaya latex by affinity chromatography on Gly-Gly-Tyr(Bzl)-Arg–Sepharose (Blumberg et al., 1970), was a gift from Dr. I. Björk, Uppsala, Sweden. Affinity-purified human cathepsin B (EC 3.4.22.1) was purchased from Calbiochem, La Jolla, CA, U.S.A., and bovine dipeptidyl peptidase I (EC 3.4.14.1) was from Sigma. Neutrophil elastase (EC 3.4.21.37), isolated from human leucocytes (Gauthier et al., 1982), was a gift from Dr. K. Ohlsson, Malmö, Sweden. Restriction endonucleases and DNA-modifying enzymes were purchased from Bethesda Research Laboratories, Bethesda, MD, U.S.A. All chemicals used were of analytical grade and were obtained from Sigma.

Generation of cystatin C variants by site-directed mutagenesis and expression in *Escherichia coli*

A cDNA encoding human cystatin C (Abrahamson et al., 1987a), modified by replacement of the DNA segment encoding the wild-type signal peptide with a DNA segment encoding the *E. coli* outer membrane protein A (OmpA) signal peptide, was ligated into a bacterial expression vector. This plasmid construct, called pHD313, results in high-level expression of recombinant cystatin C, with properties identical to those of cystatin C isolated from human urine, when introduced in *E. coli* (Abrahamson et al., 1988; Dalbøge et al., 1989). A fragment from pHD313 containing the entire recombinant cystatin C gene was excised with *Hin*dIII/*Eco*RI, subcloned in M13mp18 and subjected to site-directed mutagenesis according to Zoller and Smith (1982). The oligonucleotides used for the mutagenesis reactions corresponded to the + strand and had the sequences (with bases exchanged in codon 11 underlined): (a) 5′-CGTCTGGTT<u>C</u>GTGGTCCCATGG-3′ (Gly-11 → Arg), (b) 5′-CGTCTGGTTG<u>AA</u>GGTCCCATGG-3′ (Gly-11 → Glu), (c) 5′-CGTCTGGTT<u>T</u>G<u>G</u>GGTCCCATGG-3′ (Gly-11 → Trp), (d) 5′-CGTCTGGTTG<u>C</u>TGGTCCCATGG-3′ (Gly-11 → Ala) and (e) 5′-CGTCTGGTT<u>TC</u>TGGTCCCATGG-3′ (Gly-11 → Ser).

Phage M13 subclones containing mutated cDNA inserts were selected by plaque hybridization with the oligonucleotide used for mutagenesis, plaque-purified and verified by nucleotide sequencing. The double-stranded replicative form of mutation-positive phage clones was isolated and cleaved with *Cla*I/*Eco*RI to yield a fragment that could be re-inserted into the *Cla*I/*Eco*RI-cut expression vector. The mutated vectors were introduced in *E. coli* MC1061 that had been made competent for transformation by treatment with $CaCl_2$ (Sambrook et al., 1989). Subclones of bacteria containing the expression plasmids were selected on ampicillin culture plates. For production of each cystatin C variant, two 0.5 litre bacterial cultures in 2 litre indented flasks at half-maximal densities in the rich medium used (at $A_{555} = 8.5$, in TB medium; Sambrook et al., 1989) were grown at 42 °C for 3 h. The cells were harvested and subjected to osmotic shock, resulting in a 20 ml periplasmic fraction extract. The conditions for induction of expression, as well as for collection of the recombinant-protein-containing periplasmic fraction of the bacteria after cold osmotic shock, have been described in detail previously (Abrahamson et al., 1988; Dalbøge et al., 1989). A proteinase inhibitor cocktail concentrate was added to all recombinant protein extracts (final concentrations 5 mM benzamidinium chloride, 10 mM EDTA, 0.1 % NaN_3) to prevent possible degradation by bacterial enzymes.

Isolation of cystatin C variants

Periplasmic fraction samples of bacteria expressing the G11A-, G11S-, G11W- and G11E-cystatin C variants (i.e. variants in which Gly-11 has been replaced with Ala, Ser, Trp and Gly respectively), typically of 5 ml containing approx. 7.5 mg of recombinant protein, were dialysed against 20 mM ethanolamine buffer, pH 9.5, and applied to a Q-Sepharose (Pharmacia) column (20 cm × 2.5 cm) equilibrated in the same buffer. Chromatography was carried out at a flow rate of 100 ml/h. Under these conditions, the basic recombinant proteins passed through the column unabsorbed, whereas the bulk of bacterial proteins were absorbed. Fractions containing the recombinant protein were concentrated at 4 °C by pressure ultrafiltration using a Diaflo YM2 membrane (Amicon Corp., Danvers, MA, U.S.A.). The samples were then gel-filtered using the Pharmacia f.p.l.c. fast-desalting 10/10 column (containing Sephadex G-25 Superfine) in 50 mM ammonium bicarbonate buffer, pH 8.3, at a flow rate of 1.5 ml/min. Protein-containing fractions were again concentrated by ultrafiltration (Centricon 3; Amicon Corp.) to a concentration of approx. 0.2 mg/ml. The G11R-cystatin C variant (in which Gly-11 has been replaced with Arg) was isolated by preparative electrophoresis in a 1 % agarose gel in 75 mM barbital buffer, pH 8.6 (Johansson, 1972). Gel slices containing the recombinant protein were extracted into 50 mM ammonium bicarbonate buffer, pH 8.3, by repeated freeze–thawing and the protein was then concentrated and desalted as described above. All isolated protein solutions were stored frozen at −20 °C until used.

N-Terminal truncation of cystatin C variants

Cleavage by neutrophil elastase of wild-type recombinant cystatin C and the five variants was performed analytically by incubating samples of the proteins (approx. 1 mg/ml) with the enzyme at a molar ratio of 100:1 in 50 mM ammonium bicarbonate buffer, pH 8.3, at 37 °C. The reactions were stopped by addition of di-isopropyl phosphofluoridate to a final concentration of 1 mM after various incubation times from 5 to 120 min, and the mixtures were analysed by agarose-gel electrophoresis. The rate of proteolytic modification under these conditions was assessed by densitometric scanning, since all truncated recombinant proteins had a more anodal electrophoretic mobility than the uncleaved proteins (Abrahamson et al., 1991). The modified forms of the G11A-, G11S- and G11E-cystatin C variants,

generated by incubation with neutrophil elastase under conditions as described above for 120 min (after which neither the wild-type nor the variant cystatin C samples displayed bands with normal electrophoretic mobility), were isolated and desalted by gel chromatography on the Pharmacia f.p.l.c. Superdex 75 HR 10/30 column in 50 mM ammonium bicarbonate buffer, pH 8.3, at a flow rate of 0.5 ml/min, and concentrated as described above. For truncated wild-type cystatin C, preparative agarose-gel electrophoresis was used as an additional step before gel chromatography, because of the drastic effects that even a very small contamination with the full-length protein would have in the enzyme assays.

Protein characterization

Analytical agarose-gel electrophoresis in barbital buffer, pH 8.6, was performed by the procedure of Jeppsson et al. (1979). SDS/PAGE was carried out as described by Laemmli (1970) with separation gels containing 15% acrylamide, or as described by Schägger and von Jagow (1987) in 16.5% acrylamide gels. N-Terminal sequencing was performed in an Applied Biosystems 470A gas–liquid solid-phase sequenator after blotting of protein bands separated by agarose-gel electrophoresis on to poly(vinyl difluoride) membranes (Trans-blot, Bio-Rad) (Olafsson et al., 1990). The released amino acid phenylthiohydantoin derivatives were identified using an on-line Applied Biosystems Model 120A PTH Analyser. Amino acid composition analysis was carried out using a Beckman High Performance Analyser System 6300 after *in vacuo* hydrolysis of samples for 24 h in 4 M methane sulphonic acid or 6 M HCl. Size determination by gel chromatography was performed on the Superdex 75 column run as described above, with the elution volumes for aprotinin (M_r 6500), cytochrome *c* (M_r 12400), chymotrypsinogen (M_r 23400) and β-lactoglobulin (M_r 35000) used for construction of a calibration curve. Concentrations of isolated wild-type, N-terminally truncated and mutated forms of cystatin C were assessed by A_{280} measurements assuming an absorbance factor ($A_{280}^{0.1\%}$) of 0.83, as for wild-type recombinant cystatin C (Lindahl et al., 1992).

Enzyme inhibition assays

Concentrations of inhibitory cystatins in solutions of isolated recombinant proteins were determined by titration with papain, which previously had been active-site-titrated with E-64 (Barrett et al., 1982). Equal volumes of an approx. 6 μM papain solution and dilutions of the cystatin variant studied were incubated for 30 min at room temperature in 100 mM sodium phosphate buffer, pH 6.5, containing 1 mM dithiothreitol and 2 mM EDTA. Bz-DL-Arg-pNA was then added as substrate (2.5 mM) and the residual enzyme activity was measured as absorbance at 410 nm. For determination of equilibrium constants for dissociation (K_i) of cystatin variant complexes with papain, cathepsin B and dipetidyl peptidase I, continuous rate assays with fluorogenic substrates were carried out as described (Nicklin and Barrett, 1984). Papain was assayed in 100 mM of sodium phosphate buffer, pH 6.5, containing 1 mM dithiothreitol and 2 mM EDTA, with 10 μM Cbz-Phe-Arg-NHMec as substrate. Cathepsin B assays were performed with the same substrate and in the same buffer, but at pH 6.0. Dipeptidyl peptidase I was assayed with 10 μM Gly-Phe-NHMec in 100 mM sodium phosphate buffer, pH 6.5, containing 1 mM dithiothreitol, 2 mM EDTA and 200 mM NaCl. The enzyme concentrations in the assays were 0.05–0.25 nM. Cathepsin B and dipeptidyl peptidase I were preincubated for 20 min in assay buffer at room temperature before use.

Substrate hydrolysis at 37 °C was monitored in a Perkin–Elmer LS50 fluorimeter at excitation and emission wavelengths of 360 and 460 nm respectively. Steady-state velocities before (v_o) and after (v_i) addition of inhibitor were obtained by linear regression analysis using the computer program FLUSYS (Rawlings and Barrett, 1990), which was kindly provided by Neil Rawlings and Alan Barrett (Strangeways Research Laboratory, Cambridge, U.K.). Apparent K_i values [$K_{i(app)}$] were calculated as the slope of the plot of $[I]/(1-v_i/v_o)$ versus v_o/v_i (Henderson, 1972). K_m values for the reaction between enzyme and substrate under the given conditions were taken from the literature [papain, 60 μM (Hall et al., 1992); cathepsin B, 250 μM (Kirscke and Barrett, 1987)]. These were then used to calculate the substrate-independent K_i, using the relationship:

$$K_{i(app)} = K_i(1+[S]/K_m)$$

Compensation for substrate-induced dissociation of inhibitor was not done for dipeptidyl peptidase I reactions, since it has been shown that hydrolysis of Gly-Phe-NHMec by the enzyme has a K_m value $\gg 10\,\mu$M (Gounaris et al., 1984). All determinations of v_o and v_i were based on assays with less than 2% substrate hydrolysis and a linear regression coefficient at steady state greater than 0.990.

RESULTS

Production of cystatin C variants with substitutions at residue 11

A recombinant cystatin C gene, containing the coding sequence for native human cystatin C and the *E. coli* OmpA signal peptide (Dalbøge et al., 1989), was subjected to oligonucleotide-directed mutagenesis in phage M13 to replace the codon for residue Gly-11 with codons for Arg, Glu, Trp, Ala or Ser. DNA sequencing was used to verify that selected phage subclones contained the appropriate mutations. The mutated cystatin C gene fragments were ligated into the same expression vector as used previously for high-level expression of wild-type cystatin C with full biological activity in *E. coli* (Abrahamson et al., 1988; Dalbøge et al., 1989). The plasmids were introduced into *E. coli* MC1061 cells and expression was induced in selected subclones of the bacteria. Since the OmpA signal peptide in the recombinant cystatin C genes should direct the recombinant protein to the periplasmic space of the bacteria, the periplasmic fraction was collected after cold osmotic shock. Starting with 1 litre shaking bottle cultures, the yield of all five recombinant cystatin C variants was approx. 30 mg/litre of culture, similar to the yield obtained for bacteria containing the wild-type cystatin C expression vector pHD313 under the same conditions. The five cystatin C variants were isolated by ion-exchange chromatography using Q-Sepharose at pH 9.5 or by preparative agarose-gel electrophoresis, followed by gel chromatography on an f.p.l.c. Sephadex G-25 column. The purity of all isolated proteins, as assessed by agarose and polyacrylamide gel electrophoreses, was at least 95% (see Figures 1 and 2).

Physiochemical characterization of the cystatin C variants

The expected extra positive and negative charges of the G11R- and G11E-variants respectively, compared with wild-type recombinant cystatin C, could be verified by the mobilities of the variants in agarose-gel electrophoresis (Figure 1). As estimated from SDS/PAGE (Figure 2), all variants had the same M_r as wild-type recombinant cystatin C. The observed M_r was slightly

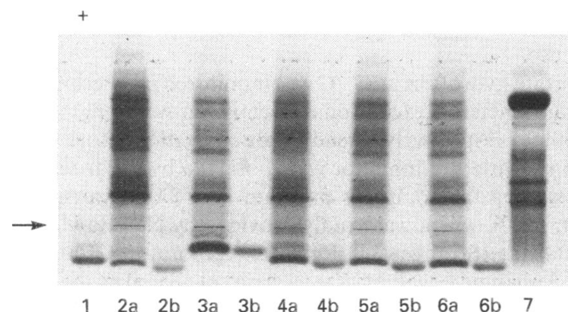

Figure 1 Agarose-gel electrophoresis of cystatin C variants expressed in *E. coli*

Electrophoresis of 8 μl samples was carried out in 1% agarose gel at pH 8.6. The point of sample application and the anode are indicated by an arrow and a plus sign respectively. Lane 1 contains a 1 mg/ml solution of isolated wild-type recombinant cystatin C and lane 7 contains human blood plasma, as references. Lanes 2a and 2b, G11R-cystatin C; 3a and 3b, G11E-cystatin C; 4a and 4b, G11W-cystatin C; 5a and 5b, G11A-cystatin C; 6a and 6b, G11S-cystatin C. Lanes marked a contain the periplasmic extracts used as starting material for purification and lanes marked b contain the isolated protein variants.

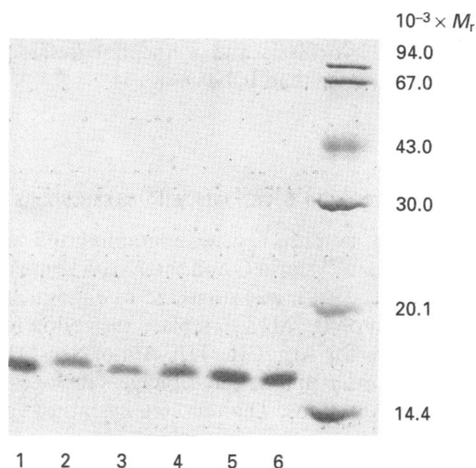

Figure 2 SDS/PAGE of isolated cystatin C variants

The proteins were subjected to electrophoresis in a 15% polyacrylamide gel. Lane 1, wild-type recombinant cystatin C; 2, G11R-cystatin C; 3, G11E-cystatin C; 4, G11W-cystatin C; 5, G11A-cystatin C; 6, G11S-cystatin C. Size markers in the lane to the right are phosphorylase *b* (M_r 94000), BSA (M_r 67000), ovalbumin (M_r 43000), carbonic anhydrase (M_r 30000), soya-bean trypsin inhibitor (M_r 20100) and α-lactalbumin (M_r 14400).

higher than that calculated from the cystatin C sequence (13343), but an identical M_r for the native protein isolated from human urine has been obtained previously using this analysis system (Abrahamson et al., 1988). All variants and the wild-type recombinant cystatin C eluted on gel chromatography on a calibrated Superdex 75 column at a position corresponding to an M_r of 12400 (results not shown).

The five variants were blotted on to poly(vinyl difluoride) membranes and subjected to 15 steps of automated Edman degradation. The released amino acid phenylthiohydantoin derivatives confirmed the expected N-terminal sequences, including the amino acid residues in position 11, for all variants except G11W-cystatin C. The expected tryptophan phenylthiohydantoin derivative in position 11 of this variant could not be identified, probably because of the poor yield of tryptophan phenylthiohydantoin derivatives inherent in the sequencing system used. This variant was therefore subjected to amino acid analysis after hydrolysis in methanesulphonic acid. The observed tryptophan content in the G11W variant was 1.3 mol/mol of protein. The corresponding value obtained for wild-type recombinant cystatin C, which contains a single Trp residue (Abrahamson et al., 1987a), was 0.7 mol/mol of protein. The combined results of the physicochemical characterization of the variants, along with the DNA sequencing of the original expression vector constructs, strongly indicated that the variants produced were those intended, with just single amino acid substitutions of the residue in position 11.

Enzyme-inhibition properties of cystatin C variants

The equilibrium constants for dissociation (K_i) of the complexes formed upon interaction of wild-type recombinant cystatin C and the five residue-11 variants with the two cysteine endopeptidases papain and human cathepsin B and the cysteine exopeptidase dipeptidyl peptidase I were determined from continuous-rate enzyme assays. Cbz-Phe-Arg-NHMec was used as substrate for the endopeptidase activity of cathepsin B at pH 6.0 and of papain at pH 6.5. Dipeptidyl peptidase I activity was assayed with Gly-Phe-NHMec as substrate at pH 6.5. Known K_m values for hydrolysis of these substrates by the enzymes allowed correction of the obtained K_i values for substrate competition. The preparations of wild-type recombinant cystatin C and the five full-length residue-11 variants used in the inhibition experiments were all at least 60% active, as calculated from the ratio between the concentration determined by active-site titration against an approx. 60% active papain preparation, and that obtained by A_{280} measurements (results not shown). The results show that the cystatin C affinity decreases at least 5-fold for papain and about 20-fold for human cathepsin B when the Gly-11 residue is changed to an Ala residue (Table 1). The G11S variant displayed about three times lower affinity than the G11A variant for both endopeptidases. The cystatin C variants with bulky hydrophobic or charged residue-11 substitutions were all such poor inhibitors of cathepsin B that exact K_i values could not be adequately determined with the inhibitor concentrations used ($K_i > 50$ nM), but the G11W variant was observed to inhibit the enzyme to a slightly greater extent than the G11R and G11E variants (results not shown). A similar pattern was observed for papain inhibition by the residue-11 variants; the affinity decreased with substitutions in the order Gly > Ala > Ser > Trp = Arg = Glu. The affinities of the five residue-11 variants for papain were generally about 100-fold higher than for cathepsin B. The affinities for the exopeptidase dipeptidyl peptidase I were similar for all five variants and for wild-type cystatin C, with a 10-fold decreased affinity of the G11E variant as the most notable deviation (Table 1).

Papain-inhibitory properties of cystatin C variants devoid of the N-terminal Ser-1–Val-10 decapeptide

The five residue-11 cystatin C variants were incubated with neutrophil elastase in an attempt to proteolytically remove the N-terminal decapeptides of the proteins, since this enzyme has been shown to hydrolyse a single peptide bond in wild-type recombinant cystatin C, namely the Val-10–Gly-11 bond (Abrahamson et al., 1991). A similar shift in the agarose-gel electrophoresis mobilities of all the variants occurred and indicated removal of charged residues in the N-terminal segments of the proteins. At an enzyme/inhibitor molar ratio of 1:100 the rate of proteolytic modification was, compared with that for wild-type cystatin C, at least 20 times higher for the G11R and G11W variants, about 10 times higher for the G11S variant,

Table 1 Equilibrium constants for dissociation (K_i) of complexes between cysteine proteinases and residue-11-substituted cystatin C variants

The mean K_i values given were calculated from at least three continuous rate assay experiments and corrected for substrate competition, as described in the Materials and methods section. The interaction between normal cystatin C and papain was too tight to allow measurement of complex dissociation with the method used. Addition of cystatin C in a slight molar excess to the lowest papain concentration giving a measurable steady-state activity (0.02 nM) resulted in complete inhibition, indicating that [E] was $\gg K_i$. A K_d value of 11 fM (i.e. 500-fold lower than the limit given) has been calculated for the same complex by determination of association and dissociation rate constants (Lindahl et al., 1992). DPP I, dipeptidyl peptidase I. The S.D.s from the mean values were all in cases $< 10\%$.

	K_i (nM)		
Substitution in cystatin C	Papain	Cathepsin B (human)	DPP I (bovine)
None	< 0.005	0.22	2.8
G11A	0.024	4.4	4.1
G11S	0.077	14	1.4
G11W	2.4	> 50	0.7
G11E	3.5	> 50	21
G11R	3.8	> 50	0.5

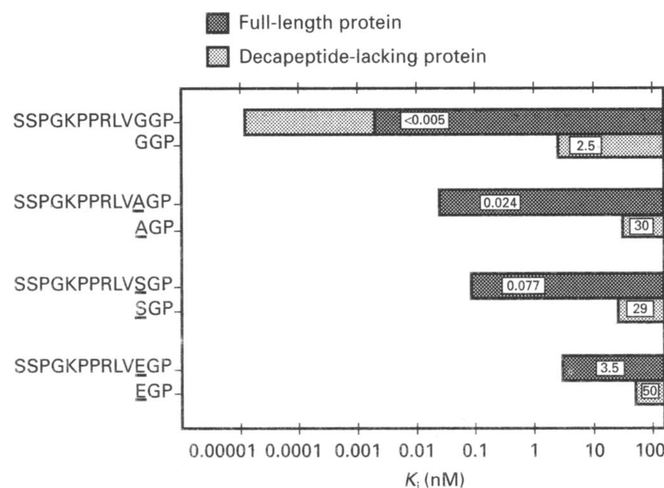

Figure 4 Equilibrium constants for dissociation of papain complexes with full-length and N-terminally truncated cystatin C variants

The K_i values (nM) determined from continuous-rate enzyme assays for full-length and N-terminal-decapeptide-lacking cystatin C variants are represented by pairs of bars. For wild-type cystatin C, the end of the dark part of the bar indicates the limit of sensitivity for the method used. The K_d value for this papain complex, calculated from determined dissociation and association rate constants (Lindahl et al., 1992), is represented by the full length of the bar.

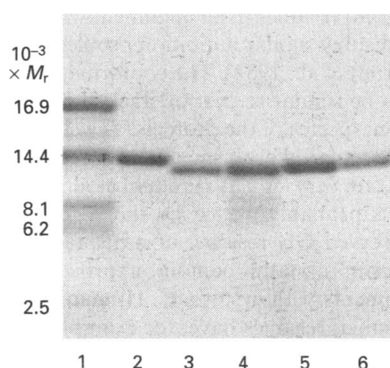

Figure 3 SDS/PAGE of N-terminally truncated cystatin C variants

Modified forms of cystatin C variants, generated by incubation with neutrophil elastase, were subjected to electrophoresis in a 16.5% polyacrylamide gel using the buffer system of Schägger and von Jagow (1987). Lane 2, full-length wild-type recombinant cystatin C; 3, truncated wild-type cystatin C; 4, truncated G11A-cystatin C; 5, truncated G11S-cystatin C; 6, truncated G11E-cystatin C. Size markers in the lane to the left are myoglobin (M_r 16949), myoglobin I and II (M_r 14404), myoglobin I (M_r 8159), myoglobin II (M_r 6214) and myoglobin III (M_r 2512).

about the same for the G11A variant and considerably lower for the G11E variant. The modified proteins resulting from incubation of wild-type recombinant cystatin C and G11A-, G11S- and G11E-cystatin C with neutrophil elastase were isolated by gel chromatography and their N-terminal sequences were determined. The sequences obtained, i.e. Gly-Gly-Pro-Met-, Ala-Gly-Pro-Met-, Ser-Gly-Pro-Met- and Glu-Gly-Pro-Met-, demonstrated that the Val-10–Xaa-11 bond was efficiently hydrolysed by the enzyme not only when residue 11 was Gly but also when it was Ala, Ser or Glu. By using the Schägger and von Jagow (1987) SDS/PAGE system for size analysis, the results obtained were consistent with the loss of the N-terminal decapeptide from all four proteins (Figure 3). The papain-inhibitory properties of the three truncated cystatin C variants were compared with those of truncated and full-length wild-type cystatin C, in an attempt to assess the functional significance of the N-terminal decapeptide in the residue-11-substituted cystatin C variants. The truncated forms of wild-type cystatin C and of the G11A and G11S variants displayed at least a 300-fold decreased affinity compared with the corresponding full-length proteins, whereas the variant with a charged residue 11, (i.e. G11E-cystatin C, showed just a 15-fold loss of affinity upon removal of its N-terminal decapeptide (Figure 4). The K_i values for all truncated cystatin C variants were similar (10^{-8}–10^{-9} M).

DISCUSSION

The eleventh residue in the single polypeptide chain of human cystatin C corresponds to a glycine residue, which is conserved in all inhibitory cystatins (Barrett et al., 1986). Crystallographic data for chicken cystatin (Bode et al., 1988) and a human cystatin B–papain complex (Stubbs et al., 1990) have revealed that this residue is positioned close to the substrate P_1 position when these members of cystatin families 1 and 2 interact with target enzymes. This result is in agreement with results from enzyme-inhibitory studies of N-terminally truncated forms of human cystatin C and chicken cystatin (family 2 cystatins) (Abrahamson et al., 1987b, 1991; Machleidt et al., 1989). Although the residues occupying the tentative P_2, P_3 and P_4 positions, on the immediate N-terminal side of the invariant glycine residue in human cystatin C and chicken cystatin, seem to contribute considerably to the cystatin–cysteine proteinase interaction (Abrahamson et al., 1987b; Machleidt et al., 1989; Hall et al., 1992), no studies have addressed the question of the importance of the Gly residue itself. In the present study we have therefore attempted to elucidate the role of the evolutionarily conserved glycine residue in cystatins by producing cystatin C variants with residue-11 substitutions and investigating their inhibitory capacity against selected target enzymes. The results of our study show that the conserved Gly-11 residue in wild-type cystatin C is required for the physiological inhibition of cathepsin B (Abrahamson et al., 1991), since even exchange of the Gly-11 residue with residues with small side-chains such as Ser or Ala results in K_i values of

the same order as the cystatin C concentration in many human body fluids (Abrahamson et al., 1986).

The G11S- and G11A-cystatin C variants displayed decreased affinities for both cathepsin B and papain. Substitution of a bulky hydrophobic residue (Trp) or charged residues (Arg or Glu) for the Gly-11 residue in cystatin C had even more drastic effects on the inhibitor's affinity for the two endopeptidases. In contrast, inhibition of the aminopeptidase dipeptidyl peptidase I was hardly affected by replacement of the Gly-11 residue even with charged or bulky hydrophobic residues. The results for dipeptidyl peptidase I agree with earlier results obtained from studies of truncated forms of cystatin C devoid of the 11 or 10 most N-terminal residues, demonstrating that the truncations had almost no effect on the inhibition of dipeptidyl peptidase I or of cathepsin H, a human aminopeptidase, respectively (Abrahamson et al., 1987b, 1991).

The initial evidence that the N-terminal segments of human cystatin C and chicken cystatin are important for the inhibitory activity of the proteins against cysteine endopeptidases was derived from studies on the papain-inhibitory properties of N-terminally truncated forms of the proteins starting at residue Gly-12 and Ala-12 respectively (cystatin C numbering) (Abrahamson et al., 1987b). The truncated proteins were produced by the action of papaya proteinase IV (Buttle et al., 1990). Similarly, another N-terminally truncated form of cystatin C, starting at residue Gly-11 and formed upon incubation of the inhibitor with neutrophil elastase, was observed to display more than 1000-fold decreased affinities for papain and human cathepsin L, and about 350-fold decreased affinity for human cathepsin B (Abrahamson et al., 1991). Additional N-terminally truncated forms of cystatin C, starting with N-terminal residues Gly-4, Lys-5, Arg-8, Leu-9 or Val-10, have also been found when the protein has been isolated from human urine (Tonelle et al., 1979; Popovic et al., 1990; M. Abrahamson and A. Grubb, unpublished work). Thus the N-terminal part of the cystatin C polypeptide chain seems to be exposed on the surface of the molecule, and hence will be susceptible to proteolytic attack but also well positioned for interactions with target proteinases. Furthermore, the observed lack of affinity between the N-terminal fragment and the rest of the cystatin C molecule after hydrolysis of the bonds Gly-11–Gly-12 or Val-10–Gly-11 (Abrahamson et al., 1987b, 1991) indicates that the N-terminal segment is linked to the rest of the protein only by the peptide bonds of the polypeptide backbone. In a report on the three-dimensional structure of a truncated form of chicken cystatin, lacking eight N-terminal amino acid residues and starting at the conserved Gly residue, Bode et al. (1988) also concluded that the N-terminal segment probably extends freely in solution in the full-length molecule. Moreover, theoretical docking experiments involving chicken cystatin and papain indicated room for a N-terminally extended chicken cystatin along the active-site cleft of the enzyme. According to the X-ray crystallographic data for chicken cystatin, the conserved Gly residue forms, together with two hairpin loops corresponding to cystatin C residues Gln-55–Gly-59 and Pro-105–Trp-106, a wedge-shaped structure with good complementarity to the remainder of the active-site cleft of papain. These data seem to indicate two possible explanations for the functional importance of the conserved Gly-11 residue: (1) It could be important for cystatin target enzyme binding by being devoid of a side-chain, so that the three-dimensional structure of the inhibitory wedge is kept intact, or (2) it could allow a certain flexibility of the N-terminal cystatin segment, with a concomitant maximal binding contribution by this segment. In an attempt to distinguish between these two possibilities, we have in the present study compared the affinities for papain of full-length residue-11 cystatin C variants and those of the same variants that are devoid of their N-terminal decapeptide segments, thus having the mutated residue as the N-terminal amino acid. Our results show that removal of the N-terminal decapeptide, which results in a decrease in the papain affinity of five orders of magnitude for wild-type cystatin C, has a smaller effect for the residue-11 variants, with a positive correlation between the size of the residue-11 side-chain and the decrease in affinity, down to just a 15-fold decrease for the G11D variant (Figure 4). The papain affinities of the four N-terminally truncated cystatin C variants studied were on the other hand, the same to within a factor of 20. These data indicate that the main significance of the conserved Gly residue in position 11 of wild-type cystatin C is that this residue permits the N-terminal segment to adopt a conformation that allows a maximal interaction between this segment and the substrate-binding pockets of cysteine endopeptidases. The observation that the papain affinities of all four N-terminally truncated cystatin C variants were almost the same also indicates that the part of the proteinase contact area contributed by the two hairpin loops, comprising the polypeptide chain segments Gln-55–Gly-59 and Pro-105–Trp-106, is little influenced by the conformation of the N-terminal segment.

The data presented here for the conserved Gly residue concern human cystatin C and have been discussed together with structural data for chicken cystatin. This comparison should be relevant, since the two proteins are highly similar, with 44% identical amino acid residues in an alignment without gaps in the sequences, and display similar inhibition profiles against cysteine proteinases (Barrett et al., 1984). The conformation of cystatin C is also known to be similar to that of chicken cystatin, since the circular dichroism spectra of the proteins, as well as their changes in absorption, circular dichroism and fluorescence on target enzyme binding, are very similar (Lindahl et al., 1988, 1992). An analogous role is probably played by the N-terminal segment, and by the conserved Gly residue, of other family 2 cystatins, which like chicken cystatin contain approx. 50% identical residues in alignments with cystatin C. Human cystatin S forms devoid of N-terminal residues have, for example, been observed to show decreased affinities for ficin (Isemura et al., 1986). It is also probable that the conserved Gly residue in the N-terminal region of family 1 cystatins serves a similar function as in cystatin C, since a variant of rat cystatin B with a charged substituent added to the residue preceding the Gly residue (Wakamatsu et al., 1984), and N-terminally truncated forms of rat cystatin A (Takeda et al., 1985), all display decreased cysteine proteinase affinities. Furthermore, the X-ray structure of a recombinant human cystatin B (also called stefin B)–papain complex (Stubbs et al., 1990) shows that the N-terminal segment and a first hairpin loop containing the homologue to the cystatin C Gln-55–Gly-59 segment form a large part of the proteinase contact area. However, the second hairpin loop postulated to be involved in family 2 cystatin binding to papain (Bode et al., 1988) is, in the family 1 cystatin–papain complex, formed by a segment with a different primary structure. The length of the polypeptide chain on the N-terminal side of the conserved glycine residue of family 1 cystatins is also in general considerably shorter than in family 2 cystatins, with average length of three and ten residues respectively. This indicates that further studies on the functional importance of the N-terminal segments of cystatins and their conserved glycine residues are warranted, particularly in order to elucidate the relative importance of the residues preceding the glycine residue for binding to different target proteinases.

We thank Ms. Helene Hansson and Mrs. Anne-Cathrine Löfström for skilful technical assistance, and Dr. Ingemar Björk for critical reading of the manuscript. The work

in Lund was supported by grants from T. Nilson's Foundation for Medical Research, A. Österlunds stiftelse, Direktör A. Påhlssons stiftelse, G. & J. Kocks stiftelser, Malmöhus läns landsting, the Medical Faculty of the University of Lund and the Swedish Medical Research Council (project nos. 05196, 09291 and 09915).

REFERENCES

Abrahamson, M., Barrett, A. J., Salvesen, G. and Grubb, A. (1986) J. Biol. Chem. **261**, 11282–11289

Abrahamson, M., Grubb, A., Olafsson, I. and Lundwall, Å. (1987a) FEBS Lett. **216**, 229–233

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

Abrahamson, M., Dalbøge, H., Olafsson, I., Carlsen, S. and Grubb, A. (1988) FEBS Lett. **236**, 14–18

Abrahamson, M., Olafsson, I., Palsdottir, A., Ulvsbäck, M., Lundwall, Å., Jensson, O. and Grubb, A. (1990) Biochem. J. **268**, 287–294

Abrahamson, M., Mason, R. W., Hansson, H., Buttle, D. J., Grubb, A. and Ohlsson, K. (1991) Biochem. J. **273**, 621–626

Barrett, A. J., Kembhavi, A. A., Brown, M. A., Kirschke, H., Knight, C. G., Tamai, M. and Hanada, K. (1982) Biochem. J. **201**, 189–198

Barrett, A. J., Davies, M. E. and Grubb, A. (1984) Biochem. Biophys. Res. Commun. **120**, 631–636

Barrett, A. J., Fritz, H., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W., Sasaki, M. and Turk, V. (1986a) Biochem. J. **236**, 312

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

Blumberg, S., Schechter, I. and Berger, A. (1970) Eur. J. Biochem. **15**, 97–102.

Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. and Turk, V. (1988) EMBO J. **7**, 2593–2599

Buttle, D., Ritonja, A., Dando, P., Abrahamson, M., Shaw, E., Wikström, P., Turk, V. and Barrett, A. (1990) FEBS Lett. **262**, 58–60

Dalbøge, H., Bech Jensen, E., Tøttrup, H., Grubb, A., Abrahamson, M., Olafsson, I. and Carlsen, S. (1989) Gene **79**, 325–332

Received 15 October 1992; accepted 6 November 1992

Freije, J. P., Abrahamson, M., Ólafsson, Í. Velasco, G., Grubb, A. and López-Otín, C. (1991) J. Biol. Chem. **266**, 20538–20543

Gauthier, F., Fryksmark, U., Ohlsson, K. and Bieth, J. G. (1982) Biochim. Biophys. Acta. **700**, 178–183

Gounaris, A. D., Brown, M. A. and Barrett, A. J. (1984) Biochem. J. **221**, 445–452

Grubb, A. and Löfberg, H. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3024–3027

Hall, A., Abrahamson, M., Grubb, A., Trojnar, J., Kania, P., Kasprzykowska, R. and Kasprzykowski, F. (1992) J. Enzyme Inhib. **6**, 113–123

Henderson, P. J. F. (1972) Biochem. J. **127**, 321–333

Isemura, S., Saitho, E., Sanda, K., Isemura, M. and Ito, S. (1986) in Cysteine Proteinases and Their Inhibitors (Turk, V., ed.), pp. 497–505, Walter de Gruyter, Berlin

Jeppsson, J.-O., Laurell, C.-B. and Franzén, B. (1979) Clin. Chem. **25**, 629–638

Johnansson, B. G. (1972) Scand. J. Clin. Lab. Invest. **29** (Suppl. 124), 7–19

Kirschke, H. and Barrett, A. J. (1987) in Lysosomes: Their role in Protein Breakdown (Glauman, H. and Ballard, F. J., eds.), pp. 193–238, Academic Press, London

Laemmli, U. K. (1970) Nature (London) **227**, 680–685

Lindahl, P., Alriksson, E., Jörnvall, H. and Björk, I. (1988) Biochemistry **27**, 5074–5082

Lindahl, P., Abrahamson, M. and Björk, I. (1992) Biochem. J. **281**, 49–55

Machleidt, W., Thiele, U., Laber, B., Assfalg-Machleidt, I., Esterl, A., Wiegand, G., Kos, J., Turk, V. and Bode, W. (1989) FEBS Lett. **243**, 234–238

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

Olafsson, I., Gudmundsson, G., Abrahamson, M., Jensson, O. and Grubb, A. (1990) Scand. J. Clin. Lab. Invest. **50**, 85–93

Popovic, T., Brzin, J., Ritonja, A. and Turk, V. (1990) Biol. Chem. Hoppe-Seyler **371**, 575–580

Rawlings, N. D. and Barrett, A. J. (1990) Comput. Appl. Biosci. **6**, 118–119

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schägger, H. and von Jagow, G. (1987) Anal. Biochem. **166**, 368–379

Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lanarcic, B. and Turk, V. (1990) EMBO J. **9**, 1939–1947

Takeda, A., Kaji, H., Nakaya, K., Aoki, Y., Nakamura, Y. and Samejima, T. (1985) Biochem. Int. **11**, 557–564

Tonnelle, C., Colle, A., Fougereau, M. and Manuel, Y. (1979) Biochem. Biophys. Res. Commun. **86**, 613–619

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

Zoller, M. J. and Smith, M. (1982) Nucleic Acids Res. **10**, 6487–6500