FEBS Letters 539 (2003) 120–124 FEBS 27086

Structural characterization of thyroglobulin type-1 domains of equistatin

Katja Galeša[a], Roger Pain[a], Maarten A. Jongsma[b], Vito Turk[a], Brigita Lenarčič[a,*]

[a] *Department of Biochemistry and Molecular Biology, J. Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*
[b] *Plant Research International, P.O. Box 16, 6700 AA Wageningen, The Netherlands*

Received 13 January 2003; revised 20 February 2003; accepted 20 February 2003

First published online 5 March 2003

Edited by Stuart Ferguson

Abstract Equistatin is a protein composed of three thyroglobulin type-1 domains. It inhibits papain-like cysteine proteinases and the aspartic proteinase, cathepsin D. To determine the structural basis for this inhibition we cloned and expressed the separated domains (eq d-1, eq d-2, eq d-3) in *Pichia pastoris*. Kinetic constants for the interaction of eq d-1 with papain and that of eq d-2 with cathepsin D are of similar order (subnanomolar) and are comparable to the constants obtained for full-length equistatin. The target proteinase for the third domain remains unknown. Thus, we demonstrate here that thyroglobulin type-1 motifs per se are able to support specific structural features that enable them to inhibit proteases from different classes. The overall conformation of three domains in equistatin is such that the interaction of domains 1 or 2 with their respective target enzymes is not hindered sterically by either domain. In addition, we show that the interaction of eq d-2 with cathepsin D results in conformational changes, which is not the case for the eq d-1/papain interaction.

Key words: Cathepsin D; Domain conformation; Equistatin; Papain; Thyroglobulin type-1 domain

1. Introduction

Equistatin is a 22 kDa protein from the sea anemone, *Actinia equina*. It is composed of three ~65-amino acid thyroglobulin type-1 domains (Thyr-1) [1,2]. Similar domains are also present in single or multiple copies in many structurally and functionally diverse proteins [3,4]. Among them, equistatin [1], saxiphilin [5], ECI [6] and p41 invariant chain fragment [7] have been classified, on the basis of the ability to inhibit a variety of papain-related cysteine proteinases, as *thyro*globulin type-1 domain *p*roteinase *in*hibitors, also called thyropins [4]. Of these four inhibitors, equistatin is unusual in also inhibiting an aspartic proteinase, cathepsin D [8].

Thyropins bind reversibly and tightly to cysteine proteinases. The general mechanism of the interaction of Thyr-1 domains with cysteine proteinases was demonstrated by the crystal structure of p41 invariant chain fragment in complex with cathepsin L. The wedge shape of the p41 fragment is anchored into the active site cleft and the interaction is mediated through the three loops of the p41 fragment [9]. In contrast to equistatin, which exhibits broad inhibitory specificity towards cysteine proteinases [1,10], the p41 fragment can discriminate between the two very similar cysteine proteinases, cathepsins L and S [7].

Only a few naturally occurring protein inhibitors of aspartic proteinases have been found. Two completely different mechanisms of inhibition have been described, one based on adopting an α-helical conformation of IA_3 upon interaction with enzyme [11] and the second based on pairing of β-sheets of PI-3 and pepsin near the active site cleft of the enzyme [12]. There are no data on the mechanism of the inhibition of cathepsin D by thyroglobulin type-1 domain.

The existence of the Thyr-1 domains in such a variety of proteins raises questions about their activity and function, and about their possible interactions with neighboring domains. Equistatin serves as a good model for such a study. Our aims were, firstly, to locate the cathepsin D inhibition activity more precisely; secondly, to compare the inhibitory properties of the separate domains with those of equistatin; and thirdly, to look for structural changes upon their interaction with their target enzymes. The individual domains were therefore prepared as recombinant proteins in *Pichia pastoris*, enabling detailed kinetic and Circular dichroism (CD) measurements to be made.

2. Materials and methods

2.1. Materials

Materials for cloning, expression and purification were obtained from the following sources: restriction endonucleases, DNA ligase, *Taq* polymerase, deoxynucleotide from Gibco BRL (The Netherlands), salts and media for *P. pastoris* from Difco (USA), *P. pastoris* expression kit, pPIC9 plasmid and *P. pastoris* strain GS115 from Invitrogen (USA), Sephadex G-50 and Phenyl–Sepharose from Amersham Pharmacia Biotech (UK). Equistatin [13], papain [14] and bovine cathepsin D [15] were purified as described. Pepstatin was from Sigma (Germany), H-Pro-Thr-Glu-Phe-Phe(NO_2)-Arg-Leu-OH from Novabiochem (Germany) and Ac-EE(Edans) KPICFFRLGK(Dabcyl)E-NH_2 from CSPS Pharmaceuticals (USA).

2.2. Cloning and expression of individual domains of equistatin

The individual domains of equistatin, eq d-1 (residues 1–67), eq d-2 (residues 68–135) and eq d-3 (residues 136–199) were created by polymerase chain reaction (PCR) employing a plasmid-encoding full-length equistatin as the template [13] and using 5′-CCCGAACAGAAAAGAGAGGCTGAAGCTAGTCTAACGAAATGCCAACAG-3′- (eq d-1), 5′-CCTCGCGAACTCGAGAAAAGAGAGGCTGAAGCTGCTGCCTTAACACTTTGCCAGT-3′ (eq d-2) or 5′-CCCGAACTCGAGAAAAGAGAGGCTGAAGCTAGCGAATGCGAGGAGGCTCGT-3′, (eq d-3) as the forward primers and 5′-GGATCCGCGGCCGCTTATTTTCTGCGACTGCAGTCTGG-3′ (eq d-1), 5′-GGATCCGCGGCCGCTTATAGGTGTCTTTCGCAGGTTGG-3′ (eq d-2) or 5′-GGATCCGCGGCCGCTTAGCATGTGGGACGTTTGAATCTG-3 (eq d-3) as the reverse primers. All fragments were subcloned into the pPIC9 expression vector by digestion

*Corresponding author. Fax: (386)-1-2573594.
E-mail address: brigita.lenarcic@ijs.si (B. Lenarčič).

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00215-1

with *Xho*I and *Not*I. The expression vectors pPIC9eqd-1, pPIC9eqd-2 and pPIC9eqd-3 were introduced into *P. pastoris* strain GS115 by electroporation with Gene Pulser Electroporator (Bio-Rad, USA). The transformants were selected on MD agar plates, the selected colonies were inoculated in 10 ml of buffered liquid BMG medium and incubated at 30°C with shaking to an A_{600} of 2.0 (~12 h). Cells were recollected by centrifugation (2500×*g* for 5 min at 23°C) and gently resuspended in 10 ml of buffered BMM medium and cultured a further 3 days to induce expression. The production of the cloned inhibitors was monitored by electrophoretic analysis of the supernatant on sodium dodecyl sulfate (SDS) gels according to Laemmli [16]. Large-scale production in bioreactor was carried out as described in [17].

2.3. Purification of the expressed domains

The culture supernatant was processed by two chromatographic steps. Eq d-1 was purified by affinity chromatography on CM-papain Sepharose followed by gel chromatography on Sephadex G-50. Eq d-2 and eq d-3 were purified by hydrophobic interaction chromatography on Phenyl–Sepharose and eluted with a gradient of 1.5–0 M ammonium sulfate as described in [18]. The identities of individual recombinant domains were confirmed by analysis of the N-terminal residues.

2.4. Inhibition kinetics of papain and cathepsin D

The kinetics of inhibition of papain with eq d-1 [1] and the equilibrium dissociation constant for the interaction of cathepsin D with eq d-2 and pepstatin were determined as described in [8]. Association rate constants for the interactions of equistatin, eq d-2 and pepstatin with cathepsin D were determined by continuous measurements under pseudo-first-order conditions at 25°C. The decrease of enzymatic activity was monitored with a DX 17MV stopped-flow apparatus (Applied Photophysics, UK). One syringe was filled with 0.1 M sodium acetate buffer, pH 4.1, inhibitor (final concentration in the range from 24 to 120 nM), and substrate Ac-EE(Edans)KPICFFRLGK(Dabcyl)E-NH_2 (5 μM) while the second one was filled with cathepsin D (3 nM) in the same buffer. 100 μl of solution from each syringe was used per run, and an average of six to eight runs was performed for each inhibitor concentration. The emission of released products was observed using an excitation wavelength of 349 nm and a cut-off filter with ~50% transmission at 420 nm. The progress curves were fitted by non-linear regression according to Morrison [19]: $[P] = v_s t + (v_z - v_s)(1 - e^{-kt})/k$, where $[P]$ is the product concentration, v_z and v_s are the initial and the steady-state velocities, respectively, t is time, and k is the observed rate constant. Association rate constants, k_a, were obtained from the slope of the linear plot of k vs. $[I]$ (slope = $k_a[I]/(1+[S]/K_m)+k_d$). The k_a values were corrected for substrate competition with the use of a K_m value of 4.5 μM [20]. The dissociation rate constants, k_d, were calculated from $k_d = K_i k_a$, where K_i was determined by the equilibrium method.

2.5. CD

Measurements were performed on an Aviv 60DS spectrometer using a quartz cell with 0.2 mm path length. In the wavelength region 250–185 nm, data points were recorded at 1 nm intervals with a dwell time of 10 s. The active concentrations of the enzymes (papain, cathepsin D) and the inhibitors (equistatin, eq d-1, eq d-2) were determined by titration as described in [1,8]. They were in the range of 80–100%. Extinction coefficients for proteins were calculated using the molar extinction coefficients of aromatic amino acids [21]. The ratio between the active concentrations was 1:1 in the case of eq d-1 and papain, and 1.1:1 for eq d-2 and cathepsin D. Eq d-1, papain and complex were scanned in 10 mM sodium phosphate, pH 6.0, and eq d-2, cathepsin D and its complex in 100 mM sodium phosphate, pH 3.2. The CD spectra of equistatin and its individual domains were measured in 100 mM sodium phosphate, pH 7.0. To avoid the autodegradation of enzymes the temperature of scanning was 4°C. In order to confirm the absence of turbidity the absorbance of protein solutions was scanned from 350 to 240 nm (Perkin Elmer lambda 18 spectrophotometer) prior to CD measurements.

3. Results

3.1. Recombinant domains

The yield of expressed eq d-1 domain in the medium, determined by titration with papain, was up to 1 g/l. The amounts of eq d-2 and eq d-3 were in the same range, as estimated from SDS–PAGE gels, where the band of eq d-1 served as the reference. The final yields of the purified domains were ~800, ~300 and ~300 mg/l for eq d-1, eq d-2 and eq d-3, respectively. The main loss of eq d-2 and eq d-3 was observed during Phenyl–Sepharose chromatography, where the proteins tended to aggregate. No major loss occurred during the final purification step on Sephadex-G50. Purified domains were analyzed on SDS–PAGE gel and in all cases a single band corresponding to an M_r of about 7000 was observed (data not shown). The purity of each domain was characterized by N-terminal sequencing. All three domains started with EAEA, followed by their characteristic sequences [13]. The EAEA sequence corresponds to a glutamic acid–alanine repeat which is normally efficiently removed by the yeast STE13 protease [22].

3.2. Kinetic characterization

The purified recombinant domains were tested for their ability to inhibit papain and cathepsin D (Table 1). Eq d-1 inhibited only papain and eq d-2 inhibited only cathepsin D with K_i values closely similar to those determined for equistatin. Association and dissociation rate constants for the inhibition of cathepsin D by eq d-2, native equistatin and pepstatin were determined by stopped-flow and slow-binding kinetic techniques. A linear dependence of k on inhibitor concentration was observed for all three inhibitors investigated (Fig. 1), consistent with simple, competitive inhibition [19]. The rate constants (k_a) for the binding of inhibitors to cathepsin D were calculated from the slopes of these plots. Pepstatin exhibited almost three-fold higher k_a values than those obtained for equistatin and eq d-2. The k_d values could not be determined from these measurements, as the intercept on the ordi-

Table 1
Kinetic data for the interaction of papain and cathepsin D with equistatin, equistatin domains and pepstatin

Enzyme	Inhibitor	$10^{-6} \times k_a$ (M^{-1} s^{-1})	$10^4 \times k_d$ (s^{-1})	K_i (nM)
Papain	equistatin	3.5 ± 0.3[a]	6.3 ± 0.3[a]	0.18 ± 0.01[a,b]
	eq d-1	4.5 ± 1.0	11.6 ± 0.3	0.26 ± 0.06[b]
Cathepsin D	equistatin	2.8 ± 0.1	14.0 ± 6.3	0.5 ± 0.2[a,c]
	eq d-2	3.1 ± 0.2	15.5 ± 8.4	0.5 ± 0.3[c]
	pepstatin	8.0 ± 0.1	0.8 ± 0.02	0.01 ± 0.006[c]

Error estimates for K_i were calculated from the square root of the sum of the relative squared errors for k_a and k_d assuming that variance is additive.
[a]Results previously published [17].
[b]Determined by kinetic measurements.
[c]Determined by equilibrium method.

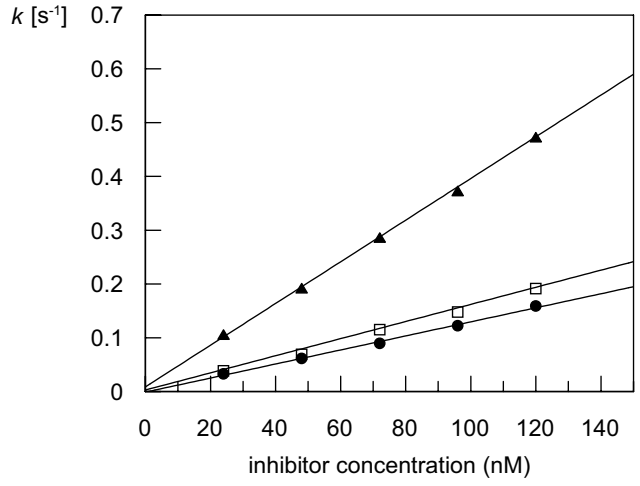

Fig. 1. Dependence of the pseudo-first-order rate constant k on inhibitor concentration for the interaction of cathepsin D with equistatin (●), eq d-2 (□) and pepstatin (▲). Solid lines were obtained by linear regression according to Morrison [27].

nate was indistinguishable from zero. However, the k_d values were calculated from K_i (determined by the equilibrium method) and k_a (Table 1). The much lower K_i value for the interaction of cathepsin D with pepstatin (0.01 nM) than with eq d-2 (0.5 nM) reflects the slower dissociation rate for pepstatin (Table 1).

3.3. Structure analysis by CD spectroscopy

The far-UV CD spectra of equistatin and its domains are shown in Fig. 2. The spectrum of eq d-1 was shown to be independent of ionic strength in the range 10–100 mM (data not shown). The trough at 200 nm of eq d-1 and eq d-2 is typical of proteins that show no regular secondary structure [23], although a β-structure is also possible (see Section 4). The positive peak around 220 nm in eq d-2 is reported to reflect an aromatic contribution [24], and this could also cause a shift of the spectrum to lower wavelengths. It cannot be ruled out, however, that the fold of this domain differs significantly from those of the other two. In addition, the equistatin spectrum was superimposed on the summed spectra of

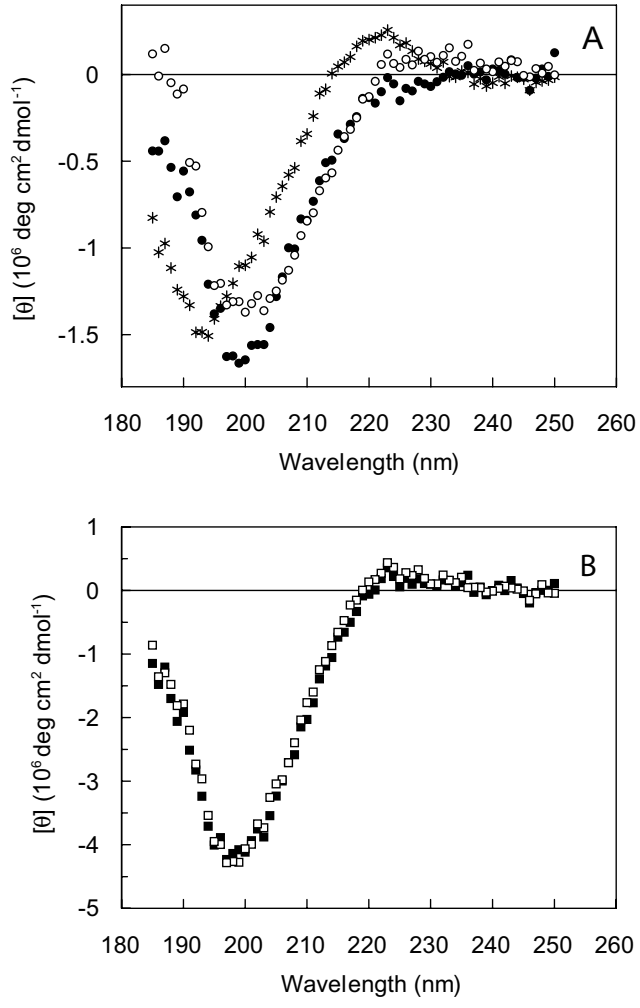

Fig. 2. A: Far-UV CD spectra of eq d-1 (●), eq d-2 (*) and eq d-3 (○). B: Comparison of the CD spectrum of equistatin (□) and the sum of the spectra for eq d-1, eq d-2 and eq d-3 (■).

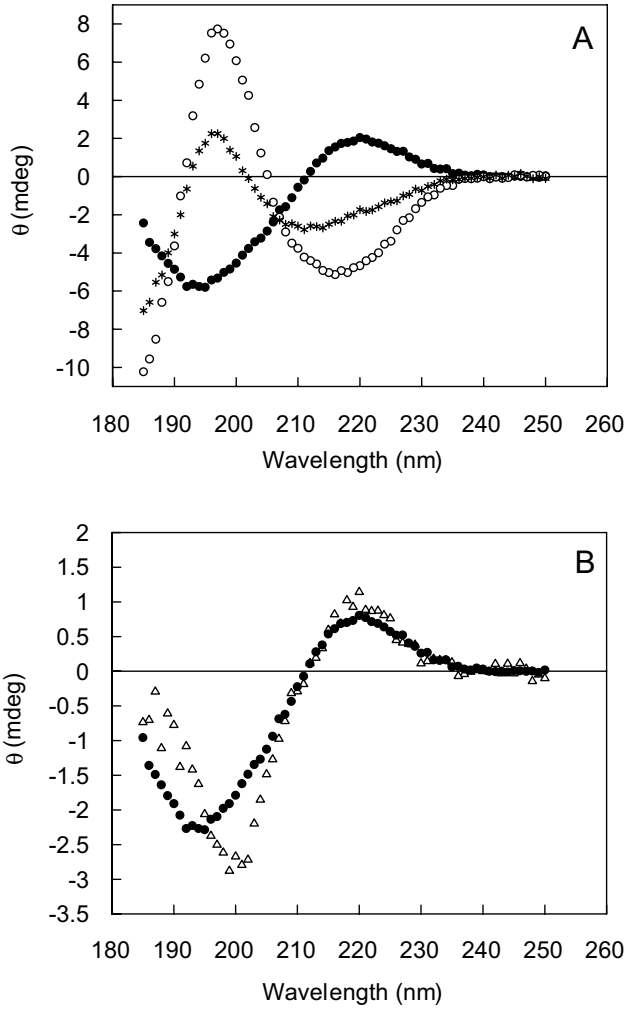

Fig. 3. A: Far-UV CD spectra of eq d-2 (●), cathepsin D (○) and the complex of the two (*). B: Comparison of the observed CD spectrum of eq d-2 (●) with the spectrum obtained by subtraction of the CD spectrum of cathepsin D from that of the complex (△). The CD spectra were recorded at concentrations of 12.8 μM for cathepsin D, 28.1 μM for eq d-2 and the mixed sample contained 7.7 μM of cathepsin D and 11.0 μM of eq d-2.

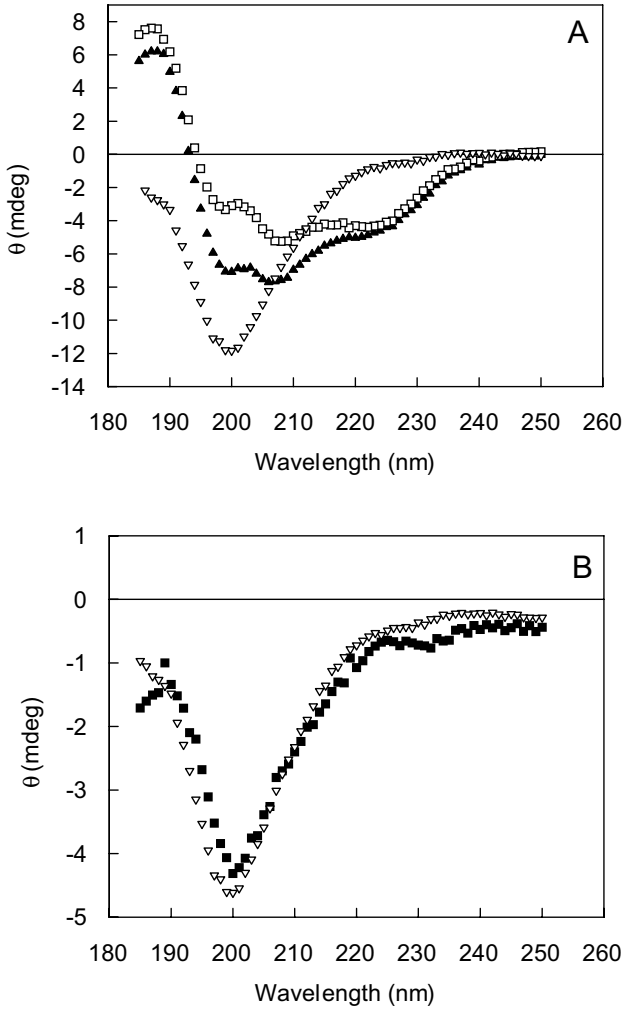

Fig. 4. A: Far-UV CD spectra of eq d-1 (▽), papain (□) and the complex of the two (▲). B: Comparison of the observed CD spectrum of eq d-1 (●) with the spectrum obtained by subtraction of CD spectra of that for papain from that of the complex (△). Spectra were recorded at concentrations of 12.0 μM for papain, 35.0 μM for eq d-1 and the mixed sample contained 12.0 μM of papain and 13.0 μM of eq d-1.

the individual domains (Fig. 2B), showing that the domains in equistatin do not influence each other's conformation.

In order to explore the possibility suggested by the work on the *Saccharomyces* inhibitor IA_3, that the inhibitor–aspartic protease interaction can result in conformational change [11], the spectrum of eq d-2 was compared with the one obtained by subtraction of the spectrum of cathepsin D from that of the complex. There is a significant shift towards the spectra of eq d-1 and eq d-3, characterized by a trough at 200 nm (Fig. 3). The possibility that the change is occurring in cathepsin D, or in both, cannot, at this stage, be ruled out.

The analogous experiment was carried out with papain and eq d-1 (Fig. 4). The only difference between the spectra for the free and complexed domain is in intensity, which could be due to a slight tightening up of their conformations resulting from the expected mutual stabilization.

4. Discussion

The expression system in *P. pastoris* developed here for the individual equistatin domains gave quantities sufficient for their functional and structural characterization. The remarkable expression levels were obtained by using the bioreactor, in which higher cell concentrations and strictly controlled cultivation conditions can be achieved. The expression system adopted was that previously used for the production of full-length equistatin in *P. pastoris*, where the yield of expressed protein was also up to 1 g/l [17,18]. In contrast, the attempted production of individual equistatin domains in *Escherichia coli* resulted in yields of less than 1 mg/l [25].

Previous studies on native equistatin gave a detailed description about the kinetics of the interaction of equistatin and eq d-1 with cysteine proteinases, while the interaction of equistatin with cathepsin D was characterized only by the determination of its equilibrium dissociation constant [1,8]. No information about rate constants for aspartic proteinase–inhibitor interactions is available. The interaction of cathepsin D with equistatin, eq d-2 and pepstatin was studied under pseudo-first-order conditions in the presence of a highly sensitive fluorogenic substrate that enabled us to follow the progressive inhibition. The pseudo-first-order rate constant increases linearly with inhibitor concentration, indicating a simple bimolecular process. The interaction of all the inhibitors tested with cathepsin D resulted in the rapid formation of tight complexes. The higher inhibitory potency of pepstatin is a consequence of lower dissociation rate, which is in agreement with the fact that pepstatin in the complex with cathepsin D is stabilized by the numerous hydrogen bonds between backbone atoms of the inhibitor and both main and side chain atoms of the enzyme [26].

The sequence similarity of known thyroglobulin type-1 domains suggests that they share a common fold. Indeed, most of the conserved residues are in positions that stabilize the fold present in the p41 fragment [4,9]. Although interpretation of the spectra of equistatin domains is ambiguous (see Section 3), those of eq d-1 and eq d-3 are closely similar, indicating similar backbone conformations. The spectra are very similar to those of a group of small β-structure proteins including the WWW domain [27], certain SH3 domains [28,29], and CspA [30], and also of clitocypin, a larger protein with a high β content [31]. The content of α-helix in the p41 invariant chain fragment, and presumably in the equistatin inhibitory domains, is small (nine residues out of 64 total [9]) and would only make a small contribution to the spectrum. Thus the most reasonable interpretation of the far-UV spectra is that the β structure in thyroglobulin type-1 domains is such as to give the same type of CD spectrum as the WWW β-structure domains. The spectrum of free eq d-2 differs somewhat (Fig. 2) in showing a marked aromatic contribution and probably a different backbone conformation. It is interesting that interaction of eq d-2 with cathepsin D induces a change which, when analyzed in terms of the spectrum of the domains, results in a spectrum similar to those of the other two domains (Fig. 3). This may reflect the kind of induced conformation observed with the IA_3 inhibitor [11]. No such change was observed with papain and eq d-1.

Eq d-3 did not inhibit either papain or cathepsin D. This is consistent with the previous finding that equistatin can only bind two enzyme molecules, but it was not clear whether this was due to the topology of the intact molecule which could sterically hinder the approach of an enzyme [8]. Provided that this domain is properly folded – the similarity of its spectrum

to that of the active eq d-1 supports the native fold – this explanation is now excluded. It remains to be seen whether it has another specificity or function.

The evidence now available supports a topological model of the intact equistatin in which the domains do not affect each other's conformation or activity. This is based on the demonstrated additivity of the CD spectra of the individual domains. It is also supported by the similarity of the association rate constants and binding constants of the two domains for the two enzymes to those of the whole equistatin. It will be interesting to discover whether the domains are packed or flexibly linked.

Acknowledgements: The financial support of the Ministry of Education, Science, and Sport is acknowledged.

References

[1] Lenarčič, B., Ritonja, A., Štrukelj, B., Turk, B. and Turk, V. (1997) J. Biol. Chem. 272, 13899–13903.
[2] Lenarčič, B., Ritonja, A., Štrukelj, B., Turk, B. and Turk, V. (1998) J. Biol. Chem. 273, 12682–12683.
[3] Molina, F., Bouanani, M., Pau, B. and Granier, C. (1996) Eur. J. Biochem. 240, 125–133.
[4] Lenarčič, B. and Bevec, T. (1998) Biol. Chem. 379, 105–111.
[5] Lenarčič, B., Krishnan, G., Borukhovich, R., Ruck, B., Turk, V. and Moczydlowski, E. (2000) J. Biol. Chem. 275, 15572–15577.
[6] Yamashita, M. and Konagaya, S. (1996) J. Biol. Chem. 271, 1282–1284.
[7] Bevec, T., Stoka, V., Pungerčič, G., Dolenc, I. and Turk, V. (1996) J. Exp. Med. 183, 1331–1338.
[8] Lenarčič, B. and Turk, V. (1999) J. Biol. Chem. 274, 563–566.
[9] Gunčar, G., Pungerčič, G., Klemenčič, I., Turk, V. and Turk, D. (1999) EMBO J. 18, 793–803.
[10] Stoka, V., Lenarčič, B., Cazzulo, J.J. and Turk, V. (1999) Biol. Chem. 380, 589–592.
[11] Li, M., Phylip, L.H., Lees, E.E., Winther, J.R., Dunn, B.M., Wlodawer, A. and Kay, J. (2000) Nat. Struct. Biol. 7, 113–117.
[12] Ng, K.K.S., Peterson, J., Cherney, M.M., Garen, C., Zalatoris, J.J., Rao-Naik, C., Dunn, B.M., Martzen, M.R., Peanasky, R.J. and James, M.N.G. (2000) Nat. Struct. Biol. 7, 653–657.
[13] Galeša, K., Štrukelj, B., Bavec, S., Turk, V. and Lenarčič, B. (2000) Biol. Chem. 381, 85–88.
[14] Blumberg, S. and Schechter, I. (1970) Eur. J. Biochem. 15, 96–102.
[15] Kregar, I., Urh, I., Umezawa, H. and Turk, V. (1977) Croat. Chem. Acta 49, 587–592.
[16] Laemmli, U.K. (1970) Nature 227, 680–685.
[17] Rogelj, B., Štrukelj, B., Bosch, D. and Jongsma, M.A. (2000) Protein Express. Purif. 19, 329–334.
[18] Outchkourov, N.S., Stiekema, W.J. and Jongsma, M.A. (2002) Protein Express. Purif. 24, 18–24.
[19] Morrison, J.F. (1982) Trends Biochem. Sci. 7, 102–105.
[20] Gulnik, S.V., Suvorov, L.I., Majer, P., Collins, J.R., Kane, B.P., Johnson, D.G. and Erickson, J.W. (1997) FEBS Lett. 413, 379–384.
[21] Peace, C.N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) Protein Sci. 4, 2411–2423.
[22] Cereghino, J.L. and Cregg, J.M. (2000) FEMS Microb. Rev. 24, 45–66.
[23] Sreerama, N., Venyminow, S.Y. and Woody, R.W. (2000) Anal. Biochem. 287, 243–251.
[24] Woody, R.W. (1994) Eur. Biophys. J. 23, 253–262.
[25] Štrukelj, B., Lenarčič, B., Gruden, K., Pungerčar, J., Rogelj, B., Turk, V., Bosch, D. and Jongsma, M.A. (2000) Biochem. Biophys. Res. Commun. 269, 732–736.
[26] Baldwin, E.T., Bhat, T.N., Gulnik, S., Hosur, M.V., Sowder, R.C., Cachau, R.E., Collins, J., Silva, A.M. and Erickson, J.W. (1993) Proc. Natl. Acad. Sci. USA 90, 6796–6800.
[27] Koepf, E.K., Petrassi, H.M., Sudol, M. and Kelly, J.W. (1999) Protein Sci. 8, 841–853.
[28] Viguera, A.R., Arrondo, J.L., Mussachio, A., Saraste, M. and Serrano, L. (1994) Biochemistry 33, 10925–10933.
[29] Knapp, S., Mattson, P.T., Cristova, P., Berndt, K.D., Karshikoff, A., Vihinen, M., Smith, C.I. and Ladenstein, R. (1998) Proteins 31, 309–319.
[30] Reid, K.D., Rodriguez, H.M., Hillier, B.J. and Gregoret, L.M. (1998) Protein Sci. 7, 470–479.
[31] Kidrič, M., Fabian, H., Brzin, J., Popović, T. and Pain, R.H. (2002) Biochem. Biophys. Res. Commun. 297, 962–967.