Structural Basis of Reduction-dependent Activation of Human Cystatin F*

Alexander W. Schüttelkopf†§, Garth Hamilton†, Colin Watts§, and Daan M. F. van Aalten†‡

From the Divisions of †Biological Chemistry and Molecular Microbiology and ‡Cell Biology and Immunology, Wellcome Trust Biocentre, School of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, Scotland

Cystatins are important natural cysteine protease inhibitors targeting primarily papain-like cysteine proteases, including cathepsins and parasitic proteases like cruzipain, but also mammalian asparaginyl endopeptidase. Mammalian cystatin F, which is expressed almost exclusively in hematopoietic cells and accumulates in lysosome-like organelles, has been implicated in the regulation of antigen presentation and other immune processes. It is an unusual cystatin superfamily member with a redox-regulated activation mechanism and a restricted specificity profile. We describe the 2.1 Å crystal structure of human cystatin F in its dimeric “off” state. The two monomers interact in a fashion not seen before for cystatins or cystatin-like proteins that is crucially dependent on an unusual intermolecular disulfide bridge, suggesting how reduction leads to monomer formation and activation. Strikingly, core sugars for one of the two N-linked glycosylation sites of cystatin F are well ordered, and their conformation and interactions with the protein indicate that this unique feature of cystatin F may modulate its inhibitory properties, in particular its reduced affinity toward asparaginyl endopeptidase compared with other cystatins.

The cystatin superfamily of proteins constitutes an important class of natural cysteine protease inhibitors present in a wide variety of organisms. Their primary targets are family C1 cysteine proteases, including plant enzymes like papain, microbial proteases, and mammalian cathepsins (1). Besides playing a major role in lysosomal protein degradation, cathepsins have been shown to have more specific functions in antigen presentation, apoptosis, and bone remodeling (2–4). Correspondingly, aberrant cathepsin function has been implicated in disease processes like inflammation, tumor invasion, or neurodegeneration (5–8). To prevent inappropriate activation, cathepsin function must be tightly regulated. An important component of this regulation is provided by cystatins, inhibitory proteins that act through the formation of tight reversible complexes with cathepsins.

Mammalian cystatins are divided into three classes (9, 10). Stefins (type I cystatins) are predominantly cytosolic single-domain proteins of ~100 amino acids. Type II cystatins are somewhat larger (~120 residues) extracellular proteins that usually contain two conserved intramolecular disulfide bridges. Kininogens (type III cystatins), which are present in blood, consist of three type II-like domains. In addition to inhibiting papain-like enzymes, a subset of type II cystatins has also been shown to reversibly inhibit mammalian asparaginyl endopeptidase (AEP), a family C13 cysteine protease involved in antigen processing, using a binding site distinct from the family C1 interaction site (11).

Structures have been determined for several type II cystatins, including chicken egg white (CEW) cystatin (12, 13) and human cystatin D (14). In addition, several structures have been reported of three-dimensional domain-swapped forms of human cystatin C (15, 16), a pathological structural state associated with amyloid formation. The “cystatin fold” adopted by all these structures consists of a five-stranded antiparallel β-sheet wrapped around a single α-helix, with conserved loops making up structurally distinct cathepsin/AEP binding sites (11, 17, 18).

Cystatin F is a recently identified type II cystatin found in humans and mice (19–21) that possesses a number of unusual properties. The inhibitor shares ~35% sequence identity with other type II cystatins and possesses a unique extension of ~6 amino acids at its N terminus. Compared with other cystatins, the protein exhibits a distinct specificity profile, binding tightly to cathepsins F, K, L, and V, less tightly to cathepsins S and H, and not inhibiting cathepsins B, C, or X (20, 22). Although cystatin F can inhibit AEP, its affinity is reduced compared with other AEP binding cystatins (11). The expression of cystatin F is limited to hematopoietic cells, with the highest expression levels being observed in monocytes, dendritic cells, and certain types of T-cells (19, 20). Furthermore, it has been shown that cystatin F mRNA becomes up-regulated during dendritic cell maturation (23). Taken together, these suggest a specific role for cystatin F in immune response-related processes, even though the details of this role, and indeed its primary enzyme target, remain unknown.

Two N-linked glycosylation sites make cystatin F one of only two known glycosylated human type II cystatins (20). In addition to the two disulfide bridges common to all type II cystatins, mature cystatin F has two additional cysteine residues. Either one or both of these form a third, intermolecular disulfide that allows redox potential-dependent dimerization of cystatin F (19). Cystatin F is produced as a dimer (24) that is inactive as a cathepsin inhibitor. The dimer can be activated by chemical reduction, which is accompanied by a shift from dimeric to monomeric species (22).

Here we present the crystal structure of recombinant human cystatin F. The structure reveals a disulfide-dependent inhibitor dimer with an unusually positioned ordered glycan that appears to protect the intermolecular disulfide. The structure suggests a molecular mechanism for reduction-dependent activation.

EXPERIMENTAL PROCEDURES

Cloning and Expression—Cell lines overproducing full-length, C-terminal His6-tagged human cystatin F were obtained by methotrexate

---

8 This work was supported in part by a Wellcome Trust senior research fellowship and the EMBO Young Investigator Fellowship (to D. M. F. v. A.) and a Wellcome Trust Program grant (to C. W.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

9 The atomic coordinates and structure factors (code 2CH9) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

‡ To whom correspondence should be addressed. Tel.: 44-1382-344979; Fax: 44-1382-345764. E-mail: dav@davapc1.bioch.dundee.ac.uk.

§ Supported by a Biotechnology and Biological Sciences Research Council/AstraZeneca-funded Cooperative Awards in Science and Engineering studentship.

3 The abbreviations used are: AEP, asparaginyl endopeptidase; CEW, chicken egg white; r.m.s.d., root mean square deviation.
selection of dihydrofolate reductase-negative Chinese hamster ovary cells transfected with a vector based on pcDNA DHFR, employing a protocol similar to that previously described for human asparaginyl endopeptidase (25).

Protein Purification—Up to 10 liters of culture medium were adjusted to pH 7.5 and loaded onto a nickel-agarose column that was washed with 500 mM NaCl, 10 mM Tris, pH 7.5. Cystatin F was eluted with washing buffer containing 300 mM imidazole and then dialyzed exhaustively into 100 mM NaCl in 20 mM sodium acetate, pH 5.0. After concentration using Vivaspin 20 concentrators (Vivascience), following exchange chromatography on a HiTrap SP XL column (Amersham Biosciences), and then dialyzed of 2-methyl-2,4-pentanediol as a cryoprotectant. The data were processed using the HKL suite (26) and CCP4 programs (27). Relevant statistics are given in Table 1.

Phasing, Model Building, and Refinement—Using SHELXC/D/E controlled by HKL2MAP (28), five zinc sites were found, giving phases with a figure of merit of 0.40 to 2.5 Å. ARP/wARP (29) was used to build an initial model comprising 102 protein residues. Cycles of model building using O (30) and Coot (31) and refinement with REFMAC5 (32) resulted in a final model with Rwork, 21.7% and Rfree 24.1%, including 126 of 131 possible protein residues. No electron density was visible for the N-terminal 5 residues, whereas the C-terminal pentahistidine tag could be built in defined electron density. It is noteworthy that the ordered His tag, together with two aspartic acid side chains from a symmetry mate, coordinates three of the five zinc atoms found in the asymmetric unit (Table 1). Interactions of cystatins with family C1 cysteine proteases are mediated by three regions, an N-terminal segment and the two hairpin loops L1 and L2 (Fig. 1B). Comparison with cystatin-protease complex structures (17, 18) suggests that for cystatin F the relevant residue stretches are 35–40 at the N-terminal segment and the two hairpin loops L1 and L2 (Fig. 1B). Local superpositions show that, as for the overall structure, cystatin F is most similar to CEW cystatin, which gives an r.m.s.d. of ~0.85 Å for 24 Ca atoms. Given this degree of similarity, the differences in target profile are likely due to the properties of the amino acid side chains in the binding regions. Notably, cystatin F contains several unique basic residues, namely Lys35 and Lys60 in the N-terminal region and Lys49 in the L1 loop. Correspondingly, an electrostatic surface potential calculation reveals a mostly positively charged cathepsin binding edge (Fig. 1C) in contrast to the hydrophobic binding site of CEW cystatin (not shown). Considering the surface potentials of various cathepsins, it becomes obvious why cystatin F would preferentially bind to cathepsin L with its highly negatively charged active site cleft rather than to cathepsin S, which has an almost electroneutral active site surface (Fig. 1C). Interestingly, this preference is inverted for cystatin D (36), which presents a partly negatively charged binding site (14).

Cystatin F has a unique cathepsin inhibition profile, with a clear preference for cathepsins F, K, L, and V (20, 22). Interactions of cystatins with family C1 cysteine proteases are mediated by three regions, an N-terminal segment and the two hairpin loops L1 and L2 (Fig. 1B). Comparison with cystatin-protease complex structures (17, 18) suggests that for cystatin F the relevant residue stretches are 35–40 at the N-terminal segment and the two hairpin loops L1 and L2 (Fig. 1B). Local superpositions show that, as for the overall structure, cystatin F is most similar to CEW cystatin, which gives an r.m.s.d. of ~0.85 Å for 24 Ca atoms. Given this degree of similarity, the differences in target profile are likely due to the properties of the amino acid side chains in the binding regions. Notably, cystatin F contains several unique basic residues, namely Lys35 and Lys60 in the N-terminal region and Lys49 in the L1 loop. Correspondingly, an electrostatic surface potential calculation reveals a mostly positively charged cathepsin binding edge (Fig. 1C) in contrast to the hydrophobic binding site of CEW cystatin (not shown). Considering the surface potentials of various cathepsins, it becomes obvious why cystatin F would preferentially bind to cathepsin L with its highly negatively charged active site cleft rather than to cathepsin S, which has an almost electroneutral active site surface (Fig. 1C). Interestingly, this preference is inverted for cystatin D (36), which presents a partly negatively charged binding site (14).

Cystatin F Is a Covalent Dimer—Cystatin F contains two conserved disulfide bridges connecting residues Cys124 from β4 to Cys144 from β5 and Cys99 to Cys110 from the β3–β4 loop, respectively (Fig. 1B). Interestingly, cysteines Cys26 and Cys51, unique to cystatin F, form an additional intermolecular disulfide bridge across a crystallographic 2-fold
Structure of Human Cystatin F

Experimental data show that dimeric cystatin F is inactive as a family C1 protease inhibitor (22). To understand why the dimer is inactive, a model for the interaction of cystatin F with papain was constructed by superposition of cystatin F dimer on the (monomeric) stefin moiety of the crystallographic stefin B-papain complex (PDB accession number 1ROA) and the active site grooves of cathepsins L (1MHW) and S (1MS6) (45), colored from −10 $kT$ to +10 $kT$. For the cystatins, binding site elements are labeled. All molecular surfaces were calculated with GRASP (46).

Outside the N terminus few direct dimer interactions (Fig. 2B, pale green regions) are observed. In fact, the dimer structure could be described as forming a solvent-filled bowl with the N-terminal extension forming the base and two (low) sides, while the two $\beta$-sheets form the other two sides (Fig. 2A looks into the bowl). Not only are there few interactions between the two cores, but according to a hydrogen bonding network analysis carried out with WHAT IF (38) potential hydrogen bonding interactions remain unsatisfied, which is energetically unfavorable compared with the solvated monomeric state. The N-terminal extension up to residue Pro$^{36}$ forms mostly polar but electroneutral contacts with the other chain, suggesting that it would be enthalpically indifferent to the dimer/(solvated) monomer transition, while the entropic gain on flexibilization of the polypeptide chain should favor the monomeric state. Taken together, this suggests that the cystatin F
dimer, far from being thermodynamically stable, may be "spring loaded" and ready to fall apart as soon as the stabilizing intermolecular disulfide bridges become reduced. This is in agreement with the observation that cystatin F becomes active as an inhibitor only after reduction (22).

**N-linked Glycosylation Modulates Function** — As predicted, cystatin F is modified by N-linked glycosylation on Asn⁶² and Asn¹¹⁵. The latter site lies in a less ordered part of the cystatin F structure, and the glycan is only partly visible in the electron density. In contrast, the sugars bound to Asn⁶² at the C-terminal end of helix α₁ are significantly better ordered; both N-acetylg glucosamine (GlcNAc) residues are clearly visible, and the electron density shows evidence for core fucosylation, i.e. an L-fucose residue α₁–6 linked to the protein-proximal GlcNAc (Fig. 3A). Two of the sugars form hydrogen bonds with the protein; the side chain amide of Asn⁶⁵ donates a hydrogen bond to O₄ of the fucose, while the proximal GlcNAc accepts two hydrogen bonds, one from N₆ of Lys⁶⁹ to the acetyl oxygen, the other from the backbone amide of Thr²⁵ of the second cystatin chain to O₃ (Fig. 3A). This sugar thus contributes to the dimer interface.

Strikingly, the three ordered sugars attached to Asn⁶² almost completely cover the nearby intermolecular disulfide (Fig. 3, B and C). Considering that they are anchored by hydrogen bonds, it is conceivable that this arrangement helps to prevent inappropriate reduction of the cystatin F dimer. This is in agreement with the observation that unusually high concentrations of reductant are required to activate the dimeric form in vitro (22).

The N-glycosylated Asn⁶² immediately precedes the loop between α₁ and β₂, which is thought to be involved in the interaction of human...
Cystatin F has been shown to inhibit AEP, but with a significantly reduced affinity compared with cystatins C and M or CEW cystatin (11). Nevertheless, the local conformation and sequence with a significantly reduced affinity compared with cystatins C and M or cystatins with AEP (11). The present structure of dimeric cystatin F gives new insights into many of the unusual functional properties of this type II cystatin but also raises new questions. The structure explains why dimeric cystatin F is inactive as a cathepsin inhibitor, as well as how the inhibitor becomes activated in a reducing environment, supporting the notion that cystatin F is regulated by changes in redox potential. It is possible that the endosomal cystatin F pool is activated by GILT (γ-interferon-inducible lysosomal thiol-reductase (39)), linking cystatin F activity to antigen processing.

Even so, in vitro activation of cystatin F requires unusually high concentrations of reducing agent (22). This can be rationalized by the presence of the glycan attached to Asn62, which covers the intermolecular disulfide crucial for maintaining the inactive conformation. It is unclear whether glycosylated cystatin F can be reduced in vivo, especially when secreted into a relatively oxidizing environment or whether additional factors are involved in regulation of activity. Not all cystatin F molecules are glycosylated at Asn62 (40), and it is conceivable that this feature defines two species of cystatin F with different functions that, in the case of the glycosylated form, might be unrelated to cathepsin inhibition. Alternatively, it is possible that the activation of cystatin F requires additional steps to make the intermolecular disulfide bridge accessible for reduction. These might include removal or remodeling of the Asn62 glycan or modification of the protein itself, e.g. by proteolytic cleavage. In turn, these modifications are likely to alter the inhibitory profile of cystatin F with respect to AEP and/or cathepsins, making further investigation of this question important for the identification of the as yet elusive in vivo target of cystatin F.

CONCLUSIONS

Figures 3, 4-A, and 4-B show the structure of dimeric cystatin F and 0.85 Å. Given that the Asn62 glycan is unique to cystatin F (Fig. 1A), it is possible that the sugar residues are at least partly responsible for the reduced activity of cystatin F against AEP by altering the shape of the likely AEP-interacting surface. In addition, the direct hydrogen bond between the glycan and the side chain of Asn62 (Fig. 3A), which has been suggested to be crucially involved in cystatin-AEP interactions (11), might fix the conformation of the latter residue in a manner incompatible with AEP binding.

Acknowledgments—We thank the European Synchrotron Radiation Facility for synchrotron beam time, which was funded by the Medical Research Council, and the staff at BM14 for their support.

REFERENCES

1. Turk, B., Turk, V., and Turk, D. (1997) *Biol. Chem.* 378, 141–150
2. Honey, K., and Rudensky, A. Y. (2003) *Nat. Rev. Immunol.* 3, 472–482
3. Turk, B., Stoka, V., Rozman-Pungercar, J., Cirman, T., Droga-Mazovec, G., Oresic, K., and Turk, V. (2002) *Biol. Chem.* 383, 1035–1044
4. Chapman, H. A., Riese, R. J., and Shi, G.-P. (1997) *Annu. Rev. Physiol.* 59, 63–88
5. Lang, A., Horler, D., and Baici, A. (2000) *J. Rheumatol.* 27, 1970–1979
6. Adkinson, A. M., Rapits, S. Z., Kelley, D. G., and Pharm, C. T. N. (2002) *J. Clin. Investig.* 109, 363–371
7. Podgorski, I., and Sloane, B. F. (2003) *Biochem. Soc. Symp.* 70, 263–276
8. Pelbör, U., Kessler, B., Mothes, W., Goebel, H. H., Ploegh, H. L., Bronson, R. T., and Olsen, B. R. (2002) *Proc. Natl. Acad. Sci. U. S. A.* 99, 7883–7888
9. Turk, V., and Bode, W. (1991) *FEBS Lett.* 285, 213–219
10. Abrahamsson, M., Alvarez-Fernandez, M., and Nathanson, C. M. (2003) *Biochem. Soc. Symp.* 70, 179–199
11. Alvarez-Fernandez, M., Barrett, A. J., Gerhartz, B., Dando, P. M., Ni, J., and Abrahamsson, M. (1999) *J. Biol. Chem.* 274, 19195–19203
12. Bode, W., Eghl, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, I., and Turk, V. (1988) *EMBO J.* 7, 2593–2599
13. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J., Turk, V., Auerswald, E. A., Jaenicke, R., and Oschkinat, H. (1993) *J. Mol. Biol.* 234, 1048–1059
14. Alvarez-Fernandez, M., Liang, Y.-H., Abrahamsson, M., and Su, X.-D. (2005) *J. Biol. Chem.* 280, 37891–37897
Structure of Human Cystatin F

JUNE 16, 2006•VOLUME 281•NUMBER 24
JOURNAL OF BIOLOGICAL CHEMISTRY 16575

15. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M., and Jaskolski, M. (2001) Nat. Struct. Biol. 8, 316–320
16. Janowski, R., Abrahamson, M., Grubb, A., and Jaskolski, M. (2004) J. Mol. Biol. 341, 151–160
17. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) EMBO J. 9, 1939–1947
18. Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M., and Turk, D. (2003) J. Mol. Biol. 326, 875–885
19. Halfon, S., Ford, J., Foster, J., Dowling, L., Lucian, L., Sterling, M., Xu, Y., Weiss, M., Ikeda, M., Liggett, D., Helms, A., Caux, C., Lebecque, S., Hannum, C., Menon, S., McClanahan, T., Gorman, D., and Zurawski, G. (1998) J. Biol. Chem. 273, 16400–16408
20. Ni, J., Fernandez, M. A., Danielsson, L., Chillakuru, R. A., Zhang, J., Grubb, A., Su, J., Gentz, R., and Abrahamson, M. (1998) J. Biol. Chem. 273, 24797–24804
21. Morita, M., Yoshiuchi, N., Arakawa, H., and Nishimura, S. (1999) Cancer Res. 59, 151–158
22. Langerholc, T., Zavasnik-Bergant, V., Turk, B., Turk, V., Abrahamson, M., and Kos, J. (2005) FEBS J. 272, 1535–1545
23. Hashimoto, S. I., Suzuki, T., Nagai, S., Yamashita, T., Toyoda, N., and Matsushima, K. (2000) Blood 96, 2206–2214
24. Cappello, F., Gatti, E., Camossetto, V., David, A., Lelouard, H., and Pierre, P. (2004) Exp. Cell Res. 297, 607–618
25. Li, D. N., Matthews, S. P., Antoniou, A. N., Mazzeo, D., and Watts, C. (2003) J. Biol. Chem. 278, 38980–38990
26. Otwinowski, Z., and Minor, V. (1997) Methods Enzymol. 276, 307–326
27. Collaborative Computational Project Number 4 (1994) Acta Crystallogr. Sect. D Biol. Crystallogr. 50, 760–763
28. Pape, T., and Schneider, T. R. (2004) J. Appl. Crystallogr. 37, 843–844
29. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Nat. Struct. Biol. 6, 458–463
30. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjelgaard, M. (1991) Acta Crystallogr. Sect. D Biol. Crystallogr. 47, 110–119
31. Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132
32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255
33. Schuttelkopf, A. W., and van Aalten, D. M. F. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 1355–1363
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) J. Appl. Crystallogr. 26, 283–291
35. Hooft, R. W. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Nature 381, 272
36. Bahadur, R. P., Chakrabarti, P., Rodier, F., and Janin, J. (2003) Proteins 53, 708–719
37. Vriend, G. (1990) J. Mol. Graph. 8, 52–56
38. Arunachalam, B., Phan, U. T., Greuze, H. J., and Cresswell, P. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 745–750
39. Nathanson, C.-M., Wasselli, J., Wallin, H., and Abrahamson, M. (2002) Eur. J. Biochem. 269, 5502–5511
40. Kakub, W., and Sander, C. (1983) Biopolymers 22, 2577–2637
41. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
42. Merrit, E. A., and Bacon, D. J. (1997) Methods Enzymol. 277, 505–524
43. Chowdhury, S. F., Sivaraman, J., Wang, J., Devanathan, G., Lachance, P., Qi, H., Menard, R., Lefebvre, J., Konishi, Y., Cigler, M., Suela, T., and Purisma, E. O. (2002) J. Med. Chem. 45, 5321–5329
44. Ward, Y. D., Thomson, D. S., Frye, L. L., Cywin, C. L., Morwick, T., Emmanuel, M. J., Zindell, R., McNeil, D., Bekkali, Y., Girardot, M., DeTuri, M., Crane, K., White, D., Pav, S., Wang, Y., Yao, M. H., Grygon, C. A., Labadia, M. E., Freeman, D. M., Davidson, W., Hopkins, J. L., Brown, M. L., and Spero, D. M. (2002) J. Med. Chem. 45, 5471–5482
45. Nicholls, A., Sharp, K., and Honig, B. (1991) Proteins 11, 281–296