The crystal structure of human dipeptidyl peptidase I (cathepsin C) in complex with the inhibitor Gly-Phe-CHN₂

ANNE MØLGAARD^{*1,3}, JOSE ARNAU², CONNI LAURITZEN², SINE LARSEN^{1,4}, GITTE PETERSEN² AND JOHN PEDERSEN²

¹Centre for Crystallographic Studies, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

²Prozymex Pharmaceuticals A/S, Dr. Neergaards Vej 17, DK-2970 Hørsholm, Denmark

³Center for Biological Sequence Analysis, Technical University of Denmark, BioCentrum-DTU, Building 208, DK-2800 Lyngby, Denmark

⁴ ESRF, B.P. 220, 38043 Grenoble, France

* Corresponding author

Running title: Human dipeptidyl peptidase I in complex with Gly-Phe-CHN₂

Abstract

Human dipeptidyl peptidase I (hDPPI) is a lysosomal cysteine protease involved in zymogen activation of granule-associated proteases, including granzymes A and B from cytotoxic T lymphocytes and natural killer cells, cathepsin G and neutrophil elastase and mast cells tryptase and chymase. We present here the first crystal structure of an hDPPI-inhibitor complex. The inhibitor Gly-Phe-CHN₂ was co-crystallized with hDPPI and the structure was determined at 2.0 Å resolution. The structure of the native enzyme was also determined to 2.05 Å resolution to resolve apparent discrepancies between the complex structure and the previously published structure of the native enzyme. The new structure of the native enzyme is, within the experimental error, identical to the structure of the enzyme-inhibitor complex presented here. The inhibitor interacts with three subunits of hDPPI, and is covalently bound to Cys234 at the active site. The interaction between the totally conserved Asp1 of hDPPI and the ammonium group of the inhibitor forms an essential interaction that mimics enzyme substrate interactions.

The structure of the inhibitor complex provides an explanation of the substrate specificity of hDPPI, and gives a background for the design of new inhibitors.

Key words: cathepsin C, inhibitor complex, re-refinement, native structure, dipeptidyl peptidase I.

Introduction

Dipeptidyl peptidase I (DPPI, cathepsin C, EC 3.4.14.1) is a lysosomal cysteine protease, which sequentially removes dipeptides from the N-termini of protein and peptide substrates. DPPI homologues have been identified by gene cloning, biochemical characterization or sequence comparison in a variety of species including mammals (man, apes, cow, dog, rabbit, rat and mouse), human parasites (*Plasmodium* and *Schistosoma*), fish (rainbow trout and killifish), reptiles (frog) and aves (chicken), suggesting an important role and a widespread presence of DPPI in nature [1-4].

Increasing evidence of the key role of human DPPI (hDPPI) in different diseases such as sepsis [5], arthritis [6] and other inflammatory disorders using animal models has drawn attention at the potential of DPPI as a drug target. The availability of a mouse strain carrying null mutation has been important to address the central role of DPPI in these diseases [5, 7]. hDPPI is expressed in many tissues and is associated with protein degradation in the lysosomes. Recently, DPPI has also been assigned an important role in the activation of many granule-associated serine proteases, including cathepsin G and elastase from neutrophils, granzymes A and B from cytotoxic T lymphocytes and natural killer cells, together with chymase and β -tryptase from mast cells [5, 6, 8-10]. These inflammatory proteases are translated as inactive zymogens. The final step in the conversion to their active forms is an hDPPI-catalyzed removal of an activation dipeptide from the N-terminus. DPPI knockout mice accumulate exclusively the inactive, dipeptide-extended proforms of chymase together with granzyme A and B providing evidence of the role of DPPI [8].

In recent years, pharmaceutical companies have targeted the inhibition of tryptase and chymase as a drug intervention strategy. However, the active sites and catalytic activities of tryptase and chymase closely resemble a number of other proteases of the same family, including trypsin, chymotrypsin, elastase and cathepsin G. Therefore, the design of inhibitors that combine selectivity, potency, bioavailability and non-toxicity for these enzymes has proven difficult. Furthermore, the large quantities of tryptase and chymase synthesized and released by mast cells poses a challenge to ensure a continuous and suitable supply of inhibitors at the sites of release. Thus, hDPPI is an attractive candidate as a drug target for the treatment of the above diseases.

hDPPI is unique amongst the lysosomal cathepsins due to an elaborated activation mechanism and its tetrameric structure [11, 12]. Evidence has been reported of the likely involvement of cathepsin L and S [11] in hDPPI activation, a process initiated by the proteolytic activation of the proenzyme through the removal of a propeptide (activation peptide) of 87 residues, separating the 119-residue N-terminal exclusion domain from the papain-like domains [12] (Fig. 1A). Further processing splits the papain-like structure into an N-terminal heavy chain and a C-terminal light chain by cleavage of the peptide bond between Arg370 and Asp371 [12]. The structure determinations for the rat and human DPPI showed that while the activation peptide is removed, the residual propeptide (the exclusion domain) remains non-covalently associated to the heavy and light chains forming a heterotrimeric structure (Fig.1A, [12, 13]). The exclusion domain is unique to DPPI and is a determinant for the exopeptidase specificity of this enzyme. The presence of the exclusion domain blocks the active site cleft beyond the S2 site and provides the carboxyl group of the Asp1 side chain as the docking residue (conserved in all known DPPI sequences) for the free amino group of the substrate required for DPPI cleavage (Fig. 1B). Finally, the tetrameric structure of active DPPI displays the active site on the surface of each subunit making it accessible to large protein substrates as well as peptides.

The substrate requirements for effective peptidase cleavage of peptides and proteins by DPPI have been established. DPPI cleaves only substrates with a free N-terminal amino group (Fig. 1B), and does not cleave substrates with P1 or P1' Pro, P1 Ile and P2 ornithine, Lys or Arg [14-16]. Extensive analysis of dipeptide substrates has been reported in an attempt to use this information for the design of better DPPI inhibitors [17].

Natural DPPI inhibitors (*e.g.*, cystatins, E64) display broad specificity towards cysteine proteases [18]. Cystatins inhibit endopeptidases at the picomolar range and exopeptidases at the nanomolar range [12]. Other inhibitors have been discovered using different molecular diversity methods and include several types of dipeptide derivatives (*e.g.*, diazomethyl ketones, vinyl sulfones, semicarbazides, etc) [14, 15], of which some functions by reacting with the catalytic cysteine residue and thereby creating an inactive enzyme (Figure 2).

However, in spite of the availability of different DPPI inhibitors, little is known about the mechanisms involved in the inhibition and information about kinetic studies is scarce. Among the existing inhibitors, Gly-Phe-CHN₂ is one of the most effective compounds, with a $k_{\text{obs}}/[I]$ of

$10^{-4} \text{ M}^{-1} \text{ s}^{-1}$ (Fig. 2) [15, 19], although its use as a therapeutic is inadequate due to instability of the diazomethyl ketone group [15].

In order to increase our knowledge about hDPPI and to enable the rational design of novel hDPPI inhibitors, a detailed structural model of the enzyme-inhibitor complex represents an invaluable tool. Enzyme-inhibitor complexes are known for several cathepsins (refs), but until now only the native structures of human and rat DPPI have been determined (refs). Thus, we present here the first experimental crystal structure of hDPPI in complex with an inhibitor, Gly-Phe-CHN₂. Through a redetermination of the published structure of native hDPPI we also address discrepancies which were found between our hDPPI-inhibitor complex and the previously published structure of native hDPPI. The refined hDPPI structure together with the structure of the inhibitor-enzyme complex represent a significant advance in the quest to develop efficient drugs for the treatment of severe human diseases.

Materials and methods

Crystallization and data collection

Native crystals

The hDPPI protein was expressed in the insect cell/baculovirus system and purified as described previously [11]. Well formed native crystals were obtained from hanging drops by streak seeding drops containing 2 μ l 10 mg/ml hDPPI and 2 μ l reservoir solution. The reservoir solution consisted of 500 μ l 0.2 M potassium/sodium tartrate, 1.8 M ammonium sulfate and 0.1 M sodium citrate at pH 5.6, which is an optimization of the previously published crystallization conditions [12]. The native crystal was initially soaked in reservoir solution containing 20% glycerol and flash cooled in liquid nitrogen. The reflections showed a large mosaicity, and ice rings were formed. The crystal was therefore resoaked in cryoprotectant reservoir solution with 25% glycerol concentration and flash cooled again. This procedure was repeated four times with a clear decrease in the mosaicity of the reflections as well as a slightly higher resolution as

a result. The ice rings were also eliminated by this treatment. X-ray diffraction data were recorded at 120K on a MAR345 image plate detector mounted on a copper rotating anode generator from Rigaku (RU300).

Gly-Phe-CHN₂ complex

1200 μ l hDPPI (0.8 mg/ml in 2-3 mM Na phosphate, 150 mM cysteamine, 50% glycerol, pH 6.8-7.1) was diluted with 1200 μ l 50 mM citric acid, 150 mM NaCl, pH 4.5. 120 μ l 200 mM cysteamine and 120 μ l 0.1 M Gly-Phe-diazomethane dissolved in DMSO was added. The mixture was left for 1 hour. After dialysis and concentration, a yellow precipitate was formed and removed by centrifugation. The protein concentration in the supernatant was ca. 2 mg/ml. Crystals were grown from hanging drops using 2 μ l protein and 1 μ l reservoir solution over 500 μ l reservoir solution consisting of 23% PEG 4000, 0.22 M ammonium acetate and 0.1 M MES pH 6.0. Initially, a trigonal crystal form was obtained with space group $P3_121$, $a = 84.78$ Å, $c = 226.34$ Å. These crystals diffracted to a maximum resolution of 2.7 Å with a mosaicity of 0.89° at the beam line I711 at MAX-lab [20]. The long c-axis combined with the high mosaicity gave problems with low completeness due to overlaps, and therefore the cryo conditions were attempted optimized and a new data set was collected at I711 using 33% sucrose as cryo protectant. Remarkably, a new crystal form was obtained from a drop that also contained crystals of the trigonal crystal form. The new crystal form was identical to the one reported for the native human enzyme, orthorhombic, I222, $a = 87.00$ Å, $b = 89.03$ Å, $c = 115.57$ Å. X-ray diffraction data were recorded at 100 K at beam line I711, MAX-lab, Lund University, Sweden [20] using a CCD detector from MAR Research. The maximum resolution is 2.0 Å, the mosaicity is 0.74° and the overall completeness of the data is 99.3%.

Refinement and structure analysis

As a starting point for the refinement of the complex and the native structures, we used a truncated form of the previously published native structure of hDPPI [12] where all regions which differed from the structure of the rat enzyme [13] were omitted. The missing parts of the protein structure and the inhibitor were built using difference electron density maps. Cycles of

model rebuilding, addition of solvent molecules and N-linked glycosylation and positional and B factor refinement was carried out using *refmac5* [21, 22], *Arp/Warp* [23] and *O* [24]. The final models were validated using *moleman2* [25] and *WHAT_CHECK* [26]. Crystallographic data and refinement and validation statistics are summarized in Table 2. The structures were compared using the program *ESCET* [27], which takes the coordinate precision into account when determining conformationally invariant regions. Figures 3-5 were prepared using the program *PyMol* [28].

Results and discussion

Overall structure

The crystal structure of native DPPI has been determined previously for the human [12] and rat [13] enzymes. Unlike the other papain-like lysosomal cysteine proteases which are all monomeric [29], DPPI exists as a homotetramer, with its four independent active sites exposed to the solvent on the outside of the tetramer (Fig. 3). A comparison of the Gly-Phe-CH₂-hDPPI structure presented here (Fig. 2) with the previously known uncomplexed DPPI structures from rat and human [9,10] showed a good agreement with the structure of the rat enzyme but some distinct differences to the previously published structure of the human enzyme. The discrepancies were a frame shift in a loop from A21 to A29 in the exclusion domain as well as a difference in the orientation of the four C-terminal residues in the papain-like structure, both of which were regions characterized by high B-factors and several Ramachandran plot outliers in the previously published structure of native hDPPI [12] (see supplementary Fig. 1 and 2). In these regions the structure of the Gly-Phe-CH₂-hDPPI complex compare well with the structure of rat DPPI [13]. To address these apparent discrepancies, a re-determination of the native hDPPI structure was performed. After refinement of the native hDPPI structure with the new X-ray data, no significant differences between the native hDPPI structure and the Gly-Phe-CH₂-hDPPI complex were observed (supplementary Fig. 1b). Thus, the observed differences are not caused by the presence of the inhibitor. The complex and free enzyme crystallize in the same space group with similar unit cell dimensions which indicates very similar intermolecular interactions. It is therefore noteworthy that differences are seen in the orientation of the N-

linked carbohydrate structure, which could support the theory that the carbohydrate structure at Asn5 plays a role in the determination of the substrate specificity of the enzyme [12].

The active site

From the available DPPI sequences, several conserved amino acids could be identified. One of these is Asp1, which is a key residue that is involved in docking the substrate via interaction with the amino group. Additionally, Gln228, Ser233, Cys234 (catalytic), Gly277, Asn380 and His381 (catalytic) are conserved amongst all DPPI homologues. These residues are all situated in the active site and take part in the catalytic mechanism or substrate binding (Table 1, Fig. 4).

The active site of hDPPI is blocked beyond the S2 site by the exclusion domain, which determines the exo specificity of the enzyme (Fig. 1B). The structural elements of the exclusion domain responsible for the blocking are the N-terminal residues, the N-linked carbohydrate at Asn5 and a β -hairpin (Lys82-Tyr93), which protrudes from the globular structure of the enzyme, sticking out into the solvent (Fig. 3). In another lysosomal cysteine exo-peptidase, cathepsin H, access to the unprimed substrate binding sites beyond S2 is also blocked, in this case through a covalently bound mini-chain [30]. From the comparison of hDPPI with cathepsin H it would be expected that the C-terminal carboxylate group of the exclusion domain would interact with the N-terminal NH_3^+ group of the substrate. Instead, as predicted from the native structures of rat and human DPPI [12, 13], it is the side chain carboxylate of Asp1, which docks with the N-terminal of the substrate. This interaction is confirmed by the structure of the Gly-Phe-CH₂-hDPPI complex (Fig. 4).

Inhibitor binding to hDPPI

The inhibitor Gly-Phe-CHN₂ binds to the S1 and S2 binding sites of hDPPI and is covalently linked to the active site Cys234 (B) through a thioether bond (Fig. 2). The inhibitor interacts with three different subunits in the hDPPI tetramer as shown in Table 1 and illustrated in Figure 4. All hydrogen bond donors and acceptors from the inhibitor form hydrogen bonds with the enzyme. The strongest interaction is the salt bridge between the N-terminal ammonium group and Asp1 from the exclusion domain (chain A). The two other protons from the terminal ammonium group of the inhibitor are hydrogen bonded to the backbone of the totally conserved Gly277 from chain B, and a water molecule (W112). This water molecule is part of a hydrogen

bonded chain of four conserved water molecules which fills out the S2 binding pocket (Fig. 4b). A water filled channel which is located behind the catalytic Cys234 may provide flexibility in the active site and a means for escape for the water molecules in the S2 binding pocket when the pocket is filled by a substrate (supplementary fig. 3). A similar water filled cavity is also found in the other cathepsins, though in most of the other structures, it does not extend to the surface, as it does in DPPI. The carbonyl group from Gly of the inhibitor forms a hydrogen bond to the NH group of the totally conserved Gly277. The hydrogen bonded ring system formed by the interactions between Gly277 and the N-terminal residue underlines the central role of Gly277 in substrate recognition in the S1 site. Backbone interactions are also the dominant feature of the interactions between hDPPI and the phenylalanine part of the inhibitor. Its NH group is hydrogen bonded to the carbonyl group from the conserved Asn380 from chain C, and its carbonyl group has a similar interaction with the NH group from the catalytic Cys234 from chain B as well as an interaction with the side chain of Gln228 from chain B. Thus the interactions with the inhibitor bound to subsite S1 and S2 involve five of the seven totally conserved residues mentioned above. Of the two remaining, His381 is known to play a role in the activation of the catalytic Cys234, and Ser233 is situated next to Cys234. It is not directly involved in contacts with the inhibitor, but changes orientation of its side chain hydroxyl group upon binding of the inhibitor to form a hydrogen bond to the hydroxyl group of Tyr300 and the back bone carbonyl group of Cys231. Tyr300 is furthermore hydrogen bonded to the back bone carbonyl group of Gln228, which binds to the inhibitor, and thus a network of hydrogen bonds is formed. It should be noted that no side chain atoms of the inhibitor participate in H-bonds to the enzyme, and only two of the enzyme-inhibitor H-bonds involve enzyme side chain atoms. With the exception of the hDPPI-specific N-terminal H-bond to Asp1, all the enzyme-inhibitor H-bonds are also found in other cathepsins. To compare the inhibitor-enzyme interactions in DPPI with those in other cathepsins, the structure of cathepsin K in complex with the non-covalently bound inhibitor tert-butyl(1s)-1-cyclohexyl-2-oxoethylcarbamate [31] was superimposed with the hDPPI-Gly-Phe-CH₂ complex. As shown in Fig. 5, the two structures display very similar backbone interactions, the main differences are in the orientation of the phenyl group and the Asp1 ammonium group hydrogen bond mentioned above.

The backbone interactions in the S1 site gives it a character of a surface rather than a pocket, with the P1 side chain of the inhibitor pointing out toward the solvent [32]. The hydrogen bond that is formed between the backbone N of Phe in the S1 site and the backbone carbonyl group of Asn380 (C) would not be possible with a Pro at the P1 position, which may explain why the

enzyme does not favour peptides with a Pro in this position. The S1 site includes the non conserved Glu275, which is located at a position where the hDPPI may interact favourably with a positively charged side chain at P1, as suggested by Turk et al. [12].

In contrast, the S2 site is a deep pocket, which can accommodate long side chains. It has been suggested that positively charged side chains will dock with the carboxylic group of Asp1 and will therefore not go into the S2 pocket, which can explain why DPPI does not cleave peptides with positively charged amino acids in the N-terminal [12, 13]. The S2 pocket is not occupied by the inhibitor in the Gly-Phe-CH₂-hDPPI complex due to the lack of a side chain of Gly in P2. It is in the inhibitor complex and the native structure filled by four water molecules which are conserved in both structures.

We have reported here the first enzyme-inhibitor structure for DPPI, an unusual cysteine protease with a key role in several human diseases. Our study has also provided a refined structure of the native enzyme, which has clarified two poorly resolved regions in the previously published native structure [12], and may thus serve as a better template in homology modelling experiments. With the data available on the binding mode of Gly-Phe-CHN₂ to hDPPI and the knowledge about the residues involved in the enzyme-inhibitor interaction, we can pursue the rational design of more effective inhibitors to provide drugs for the treatment of *e.g.*, different inflammatory diseases in humans.

Acknowledgements

We thank Flemming Hansen for help with X-ray data collection, Dorthe Boelskifte and Jens-Christian Navarro Poulsen for help with crystallization, and the Danish Natural Science Research Council for financial support. The Maxlab experiments were supported by Dansync and the ARI (Access to Research Infrastructures) programme.

References

- 1 McGuire, M. J., Lipsky, P. E. and Thiele, D. L. (1997) Cloning and characterization of the cDNA encoding mouse dipeptidyl peptidase I (cathepsin C). *Biochim. Biophys. Acta* **1351**, 267-273
- 2 Hola-Jamriska, L., Tort, J. F., Dalton, J. P., Day, S. R., Fan, J., Aaskov, J. and Brindley, P. J. (1998) Cathepsin C from *Schistosoma japonicum*--cDNA encoding the preproenzyme and its phylogenetic relationships. *Eur. J. Biochem.* **255**, 527-534
- 3 Hutchinson, D. W. and Tunnicliffe, A. (1987) The preparation and properties of immobilised dipeptidyl-aminopeptidase I (cathepsin C). *Biochim. Biophys. Acta* **916**, 1-4
- 4 Klemba, M., Gluzman, I. and Goldberg, D. E. (2004) A *Plasmodium falciparum* dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J. Biol. Chem.* **279**, 43000-43007
- 5 Mallen-St Clair, J., Pham, C. T., Villalta, S. A., Caughey, G. H. and Wolters, P. J. (2004) Mast cell dipeptidyl peptidase I mediates survival from sepsis. *J. Clin. Invest.* **113**, 628-634
- 6 Adkison, A. M., Raptis, S. Z., Kelley, D. G. and Pham, C. T. (2002) Dipeptidyl peptidase I activates neutrophil-derived serine proteases and regulates the development of acute experimental arthritis. *J. Clin. Invest.* **109**, 363-371
- 7 Hu, Y. and Pham, C. T. (2005) Dipeptidyl peptidase I regulates the development of collagen-induced arthritis. *Arthritis Rheum.* **52**, 2553-2558
- 8 Pham, C. T. and Ley, T. J. (1999) Dipeptidyl peptidase I is required for the processing and activation of granzymes A and B in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8627-8632
- 9 Wolters, P. J., Pham, C. T., Muilenburg, D. J., Ley, T. J. and Caughey, G. H. (2001) Dipeptidyl peptidase I is essential for activation of mast cell chymases, but not tryptases, in mice. *J. Biol. Chem.* **276**, 18551-18556
- 10 Sheth, P. D., Pedersen, J., Walls, A. F. and McEuen, A. R. (2003) Inhibition of dipeptidyl peptidase I in the human mast cell line HMC-1: blocked activation of tryptase, but not of the predominant chymotryptic activity. *Biochem. Pharmacol.* **66**, 2251-2262
- 11 Dahl, S. W., Halkier, T., Lauritzen, C., Dolenc, I., Pedersen, J., Turk, V. and Turk, B. (2001) Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. *Biochemistry* **40**, 1671-1678
- 12 Turk, D., Janjic, V., Stern, I., Podobnik, M., Lamba, D., Dahl, S. W., Lauritzen, C., Pedersen, J., Turk, V. and Turk, B. (2001) Structure of human dipeptidyl peptidase I (cathepsin C): exclusion domain added to an endopeptidase framework creates the machine for activation of granular serine proteases. *Embo J.* **20**, 6570-6582
- 13 Olsen, J. G., Kadziola, A., Lauritzen, C., Pedersen, J., Larsen, S. and Dahl, S. W. (2001) Tetrameric dipeptidyl peptidase I directs substrate specificity by use of the residual pro-part domain. *FEBS Lett.* **506**, 201-206
- 14 Bondebjerg, J., Fuglsang, H., Valeur, K. R., Kaznelson, D. W., Hansen, J. A., Pedersen, R. O., Krogh, B. O., Jensen, B. S., Lauritzen, C., Petersen, G., Pedersen, J. and Naerum, L. (2005) Novel semicarbazide-derived inhibitors of human dipeptidyl peptidase I (hDPPI). *Bioorg. Med. Chem.* **13**, 4408-4424
- 15 Kam, C. M., Gotz, M. G., Koot, G., McGuire, M., Thiele, D., Hudig, D. and Powers, J. C. (2004) Design and evaluation of inhibitors for dipeptidyl peptidase I (Cathepsin C). *Arch. Biochem. Biophys.* **427**, 123-134

- 16 Pedersen, J., Lauritzen, C., Madsen, M. T. and Weis Dahl, S. (1999) Removal of N-terminal polyhistidine tags from recombinant proteins using engineered aminopeptidases. *Protein Expr. Purif.* **15**, 389-400
- 17 Tran, T. V., Ellis, K. A., Kam, C. M., Hudig, D. and Powers, J. C. (2002) Dipeptidyl peptidase I: importance of proenzyme activation sequences, other dipeptide sequences, and the N-terminal amino group of synthetic substrates for enzyme activity. *Arch. Biochem. Biophys.* **403**, 160-170
- 18 Anastasi, A., Brown, M. A., Kembhavi, A. A., Nicklin, M. J., Sayers, C. A., Sunter, D. C. and Barrett, A. J. (1983) Cystatin, a protein inhibitor of cysteine proteinases. Improved purification from egg white, characterization, and detection in chicken serum. *Biochem. J.* **211**, 129-138
- 19 Green, G. D. and Shaw, E. (1981) Peptidyl diazomethyl ketones are specific inactivators of thiol proteinases. *J. Biol. Chem.* **256**, 1923-1928
- 20 Cerenius, Y., Ståhl, K., Svensson, L. A., Ursby, T., Oskarsson, Å., Albertsson, J. and Liljas, A. (2000) The crystallography beamline I711 at MAX II. *J. Synchrotron Rad.* **7**, 203-208
- 21 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763
- 22 Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240-255
- 23 Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463
- 24 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47 (Pt 2)**, 110-119
- 25 Kleywegt, G. J. and Jones, T. A. (1996) Phi/psi-chology: Ramachandran revisited. *Structure* **4**, 1395-1400
- 26 Hoof, R. W., Vriend, G., Sander, C. and Abola, E. E. (1996) Errors in protein structures. *Nature* **381**, 272
- 27 Schneider, T. R. (2002) A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 195-208
- 28 DeLano, W. L. (2004) The PyMol Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA
- 29 Dolenc, I., Turk, B., Pungercic, G., Ritonja, A. and Turk, V. (1995) Oligomeric structure and substrate induced inhibition of human cathepsin C. *J. Biol. Chem.* **270**, 21626-21631
- 30 Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V. and Turk, D. (1998) Crystal structure of porcine cathepsin H determined at 2.1 Å resolution: location of the mini-chain C-terminal carboxyl group defines cathepsin H aminopeptidase function. *Structure* **6**, 51-61
- 31 Catalano, J. G., Deaton, D. N., Furfine, E. S., Hassell, A. M., McFadyen, R. B., Miller, A. B., Miller, L. R., Shewchuk, L. M., Willard, D. H., Jr. and Wright, L. L. (2004) Exploration of the P1 SAR of aldehyde cathepsin K inhibitors. *Bioorg. Med. Chem. Lett.* **14**, 275-278
- 32 Turk, D., Guncar, G., Podobnik, M. and Turk, B. (1998) Revised definition of substrate binding sites of papain-like cysteine proteases. *Biol. Chem.* **379**, 137-147
- 33 Cruickshank, D. W. (1999) Remarks about protein structure precision. *Acta Crystallogr. D Biol. Crystallogr.* **55 (Pt 3)**, 583-601
- 34 Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162

Tables and figure legends

Table 1. Enzyme-inhibitor interactions

Inhibitor atom	Enzyme	Distance (Å)
CH2 C	Cys234(B) SG	1.86
Phe O	Cys234(B) NH	2.97
Phe O	Gln228(B) NE2	3.03
Phe NH	Asn380(C) O	3.14
Gly O	Gly277(B) NH	3.02
Gly NH	Gly277(B) O	3.01
Gly NH	Asp1(A) OD1	2.82
Gly NH	HOH112	2.89

Table 2. Data collection and refinement statistics

	Gly-Phe-CH ₂ -hDDPI	Native hDPPI
PDB id	2djf	2djg
Wave length (Å)	1.094	1.5418
Space group	I222	I222
Cell parameters (Å)	87.00, 89.03, 115.57	87.48, 88.68, 114.35
Resolution range	28.75-2.0 (2.07-2.0)	24.04-2.05 (2.12-2.05)
No. measured reflections	143822	120690
No. unique reflections	28953	25787
Completeness	99.3 (97.5)	96.1 (96.1)
R _{sym}	0.063 (0.211)	0.051 (0.313)
No. non-H atoms used in refinement	3071	3030
Protein	2748	2720
water	246	211
inhibitor	16	-
Cruickshanks DPI*	0.131 Å	0.157 Å
R factor	0.161	0.176
R _{free}	0.200	0.221

Ramachandran plot outliers†	3.9%	3.0%
Mean B value (overall, Å ²)	26.3	26.1
B values		
Protein	25.0	24.6
Water	32.6	30.2
Inhibitor	37.3	-

* $\sigma(x, B_{\text{avg}}) = 1.0 * (N_i / n_{\text{obs}})^{1/2} * C^{(-1/3)} * R_{\text{free}} * d_{\text{min}}$, where N_i is number of atoms, n_{obs} is number of reflections and C is completeness of the X-ray data [33] † As defined by Kleywegt and Jones [25]. An average better than 2.0Å model has ~0-5% outliers.

Figure legends

Fig. 1:

(A) Overview of the nomenclature of the different parts of the sequence of hDPPI. The letters in front of the sequence numbers refer to the amino acids one-letter code. (B) Schematic drawing of the hDPPI structure in complex with a generic peptide substrate. The amino acids on the peptide bound to the enzyme are numbered P2 and P1 on the N-terminal side of the scissile bond and P1', P2' etc. on the C-terminal side. The corresponding substrate binding sites on the enzyme are numbered S2, S1, S1', S2' etc [32, 34].

Fig. 2:

Chemical reaction between Gly-Phe-CHN₂ and hDPPI. After the reaction, the inhibitor is covalently bound by a thioether bond to the hDPPI active site Cys234, forming a Gly-Phe-CH₂-hDPPI complex.

Fig. 3:

The biologically active tetrameric form of hDPPI in complex with the covalently bound inhibitor Gly-Phe-CHN₂. The monomers are located at the four corners of the tetramer with the exclusion domains shown in blue and the papain-like domains shown in pale yellow. The active sites can be seen on the outside of the tetramer by the location of the inhibitors which are shown with atoms as grey spheres. The structural elements close to the inhibitor binding site, the N-linked carbohydrate at Asn5 and the β -hairpin from Lys82-Tyr93 are shown in pale green. The other N-linked carbohydrates are shown as spheres in the same color as the domains they are linked to.

Fig. 4a:

Closeup of the active site. The carbon atoms of the inhibitor are shown in yellow. For clarity, only atoms within 5 Å of the inhibitor are included in the figure.

Fig. 4b:

The S2 binding pocket is filled by a chain of four hydrogen bonded water molecules which are conserved in the native (blue) and the Gly-Phe-CH₂-hDPPI (orange) structures. The inhibitor is shown as yellow sticks.

Fig. 5:

Superposition of Gly-Phe-CH₂-hDPPI (dark grey) and cathepsin K in complex with the non-covalently bound inhibitor tert-butyl(1s)-1-cyclohexyl-2-oxoethylcarbamate (1q6k) [31] (light grey). The inhibitors are shown in ball-and-stick representation and the enzymes are shown as sticks. The exclusion domain of hDPPI is indicated by a transparent surface.