Tetrameric dipeptidyl peptidase I directs substrate specificity by use of the residual pro-part domain[1]

Johan Gotthardt Olsen[a,2,3], Anders Kadziola[a,2], Conni Lauritzen[b], John Pedersen[b], Sine Larsen[a,*], Søren Weis Dahl[b]

[a] *Centre for Crystallographic Studies, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark*
[b] *Prozymex AlS, Dr. Neergaards Vej 17, DK-2970 Hørsholm, Denmark*

Received 25 July 2001; revised 30 August 2001; accepted 5 September 2001

First published online 21 September 2001

Edited by Amy M. McGough

Abstract The crystal structure of mature dipeptidyl peptidase I reveals insight into the unique tetrameric structure, substrate binding and activation of this atypical papain family peptidase. Each subunit is composed of three peptides. The heavy and light chains form the catalytic domain, which adopts the papain fold. The residual pro-part forms a β-barrel with the carboxylate group of Asp1 pointing towards the substrate amino-terminus. The tetrameric structure appears to stabilize the association of the two domains and encloses a 12 700 Å^3 spherical cavity. The tetramer contains six chloride ions, one buried in each S2 pocket and two at subunit interfaces.

Key words: Cathepsin C; Zymogen activation; Papain fold; Chloride binding; Aminopeptidase; Subunit interaction

1. Introduction

Dipeptidyl peptidase I (DPPI, cathepsin C) appears to serve an essential function in the activation of a number of granule serine peptidases[1–3] involved in cell-mediated apoptosis, inflammation, connective tissue remodelling etc. Recently, patients with deleterious mutations in the DPPI encoding *CTSC* gene were identified and shown to develop the autosomal recessive disorder Papillon-Lefèvre syndrome [4,5].

DPPI is a member of the papain family of cysteine peptidases which includes cathepsins L, S, B, H and K. Each of these peptidases is expressed as a preproenzyme consisting of a signal peptide, a proregion and a catalytic region. A unique characteristic of DPPI is that maturation of the pro-DPPI zymogen involves cleavage of the catalytic region into the heavy and light chains and an unusual non-autocatalytic excision of an internal activation peptide within the proregion [6,7]. The N-terminal, glycosylated part of the proregion, the residual pro-part, remains non-covalently associated with the heavy and light chains in the mature form [8,9]. Further, the residual pro-part and the excision of the activation peptide determine the unique maturation-dependent oligomeric structure of the enzyme. Pro-DPPI has been shown to exist as a dimer [7] and following maturation, a dimer of dimers is formed [6].

Crystal structures of a number of papain family peptidases and their zymogens have been reported as well as the structures of complexes with inhibitors [10–14]. However, the structure and function of the residual pro-part, the oligomeric structures of DPPI and its zymogen, its dipeptidyl aminopeptidase specificity and chloride dependency are distinct characteristics of DPPI that so far have not been characterized at the atomic level. Determination of the crystal structure of mature rat DPPI has revealed detailed information about these remarkable characteristics.

2. Materials and methods

2.1. Purification and crystallization

The protein was expressed using the baculovirus/insect cell system and purified as described [15]. Crystallization was performed using the hanging drop vapor diffusion technique. Drops containing a 1:1 mixture of protein solution (10 mg/ml DPPI in a solution of 20 mM Bis–Tris, 150 mM NaCl, 2 mM dithiothreitol, 2 mM EDTA, pH 7.0) and precipitant solution were equilibrated against 700 μl precipitant solution (0.1 M Bis–Tris propane, pH 7.5, 1.4 M $(NH_4)_2SO_4$). Crystals grew to $0.5\times0.5\times0.2$ mm^3 in 2–4 days. The diffraction data were recorded at the MAX Lab synchrotron facility beam line 711 in Lund, Sweden. Heavy atom derivative data were recorded at 260 K in a glass capillary. The native data set was recorded at 110 K using precipitant solution containing 34% glycerol as a cryoprotectant.

2.2. Structure determination

The structure was solved by multiple isomorphous replacement (MIR) using three heavy atom derivatives: mercury acetate, *para*-hydroxy mercuribenzoic acid (PHMBA) and dipotassium tetrachloro aurate (K_2AuCl_4). Due to radiation damage, the derivative data were not complete (Table 1). Thus the heavy atom sites were located using a reciprocal space method. Heavy atom positions were determined using the molecular replacement program AmoRe [16] as a translation search. Isomorphous differences ($|F_{der}-F_{nat}|$) were used as target for the centered correlation translation function [17] and a mercury atom was used as a 'search model'. In this way two sites were found in the mercury acetate and PHMBA derivatives and four sites in the K_2AuCl_4 derivative. All three derivatives had one site in common while remaining positions differed. Sites were refined using MLPHARE [18] and phases were subjected to density modification in DM [18] using solvent flattening and histogram matching options.

*Corresponding author. Fax: (45)-35-32 02 99.
E-mail address: sine@ccs.ki.ku.dk (S. Larsen).

[1] The structure factors and model coordinates were submitted with the protein data bank (PDB), accession number 1jqp.
[2] These authors contributed equally to this work.
[3] Present address: Carlsberg Laboratory, Department of Chemistry, Gamle Carlsbergvej 10, DK-2500 Valby, Denmark.

Abbreviations: DPPI, dipeptidyl peptidase I; PHMBA, *para*-hydroxy mercuribenzoic acid

PII: S0014-5793(01)02911-8

Table 1
Crystallographic statistics

	Native	$HgCl_2$	K_2AuCl_4	PHMBA
Data collection/phasing statistics				
Space group	$P6_422$			
Unit cell parameters	$a=b=166.2$, $c=80.5$			
Resolution range (Å)	30.0–2.4	15.0–3.3	15.0–3.2	15.0–3.3
Unique reflections	26 061	6 204	6 423	5 681
Completeness (%)	99.2	72.0	75.0	66.0
R^a_{iso} (15–3.5 Å)		0.504	0.512	0.483
Number of sites used for phasing		2	2	2
Refinement statistics				
R^b_{merge} (%) overall/outer res. shell	7.1/32.2			
R_{work}/R^c_{free} (%)	24.5/27.4			
Average *B*-factors (Å^2)				
Main chain atoms	40.3			
Side chain atoms	41.8			
Solvent atoms	36.4			

[a] $R_{iso}=\Sigma_{hkl}|F_{der}-F_{nat}|/\Sigma_{hkl}F_{nat}$.
[b] $R_{merge}=\Sigma_{hkl,i}|I_{hkl,I}-\langle I\rangle_{hkl}|\Sigma_{hkl,i}I_{hkl,I}$.
[c] R_{free} was calculated using reflections (5% of total) not used in refinement.

This resulted in a partially interpretable map where the papain core domain of cathepsin B could be fitted as a rigid body into the electron density and most of the cathepsin C core domain could be build from this initial model. The residual prodomain β-strands were fitted with the Trigonal Planar Pseudo Residues (TPPR) construction option in the graphics display program TURBO-FRODO [19]. As only 80% of the phases had been assigned after MLPHARE because of incomplete derivative data the map was rather poor. The core domain and the TPPR model for the residual propeptide were refined against the map with CNS [20]. The resulting backbone model was used to calculate phase indications for all reflections which in turn were combined with existing MIR phases and subjected to density modification. The resulting map was used to build an initial model for the entire protein.

2.3. Structure refinement

Structure refinement was performed using the synchrotron radiation data extending to 2.4 Å resolution (Table 1) using the program CNS [20]. The final model consists of residues: Asp1–Met118 (pro-part), Leu204–His365 (heavy chain), Pro371–Leu438 (light chain), 1.5 chloride ions and 83 water molecules were refined to an *R*-factor of 24.2% and an R_{free} of 26.9%. The (final) model displays good stereochemistry with root mean square (r.m.s.) deviations from standard values of 0.007 Å for bond distances and 1.4° for bond angles. A Rhamachandran plot calculated with the program PROCHECK [21] showed 99.3% of all non-glycine and non-proline residues in the most favored, additional allowed, and generously allowed regions. The only residue, Tyr64, found in the disallowed region is part of the active site and possesses a well-defined electron density.

3. Results

3.1. Overall structure

Mature rat DPPI crystallizes as a homo-tetramer composed of subunits A–D (Fig. 1C). The crystallographic site symmetry 222 defines the symmetry of the tetramer. Two chloride ions, each interacting with two Arg25 residues, are positioned at the AB and CD interfaces. Together they block what appear to be the only two potential entrances to a 12 700 Å^3 cavity found in the center of the tetramer. The presence of this exceptionally large cavity, theoretically spacious enough to fit 400 water molecules, could explain the low stability of the tetrameric form [15]. The resolution of the structure determination allows identification of structural water molecules at the surface of the protein facing the cavity, but as the bulk of the cavity must be occupied by solvent the energy cost normally associated with cavities or 'packing defects' [22] is not relevant in the present case.

In agreement with biochemical analysis [1,6], the electron density defines three polypeptide chains within each subunit: the residual pro-part (residues Asp1–Met118), the heavy chain (residues Leu204–His365) and the light chain (residues Pro371–Leu438). The subunit is shown in Fig. 1A. The heavy and light chains form the catalytic domain, which is homologous to the papain fold core domain. The catalytic domain superimposes with cathepsins H and B [11,14] with r.m.s. deviations of Cα atoms of around 1 Å and with no major insertions or deletions. However, helix 3, which is well-defined in other members of the family, is distorted and contains a π-helical segment in DPPI. At the N-terminus of this helix we find a chloride ion coordinated to the backbone amide of residue 279 and the Oη of Tyr322.

The residual pro-parts form distinct, globular domains (Fig. 1A,B) in agreement with experimental data [23]. The residual pro-part domain adapts an eight-stranded β-barrel (meander) fold at the core (Fig. 1B). The fold was submitted to the DALI server [24] and structurally similar protein folds were detected in the retinol binding and streptavidin families. Extending from strands rβ6 and rβ10, a two-stranded sheet (strands rβ7 and rβ8) protrudes hovering over the core domain active site. This protrusion is connected to an N-terminal coil via a disulfide bond between Cys6 and Cys94. The rβ7–rβ8 structures point out like spikes from the tetramer (Fig. 1C).

Electron density is missing for five residues at the heavy chain–light chain boundary. However, these residues would not in any conformation be capable of connecting the observed heavy and light chain termini, the His365 Cα–Pro371 Cα distance being 29 Å. Apparently, the N-terminus of the light chain (residues 371–377) has moved away from the core domain following its separation from the heavy chain. After cleavage, it interacts with Thr7, Tyr8 and Tyr64 at the rβ4–rβ5 loop of the residual pro-part within the same subunit. This potentially gives it an important role in positioning the substrate binding Asp1 (see below).

4. Discussion

4.1. Determinants of subunit oligomerization

The DPPI tetramer is defined by the crystallographic 222

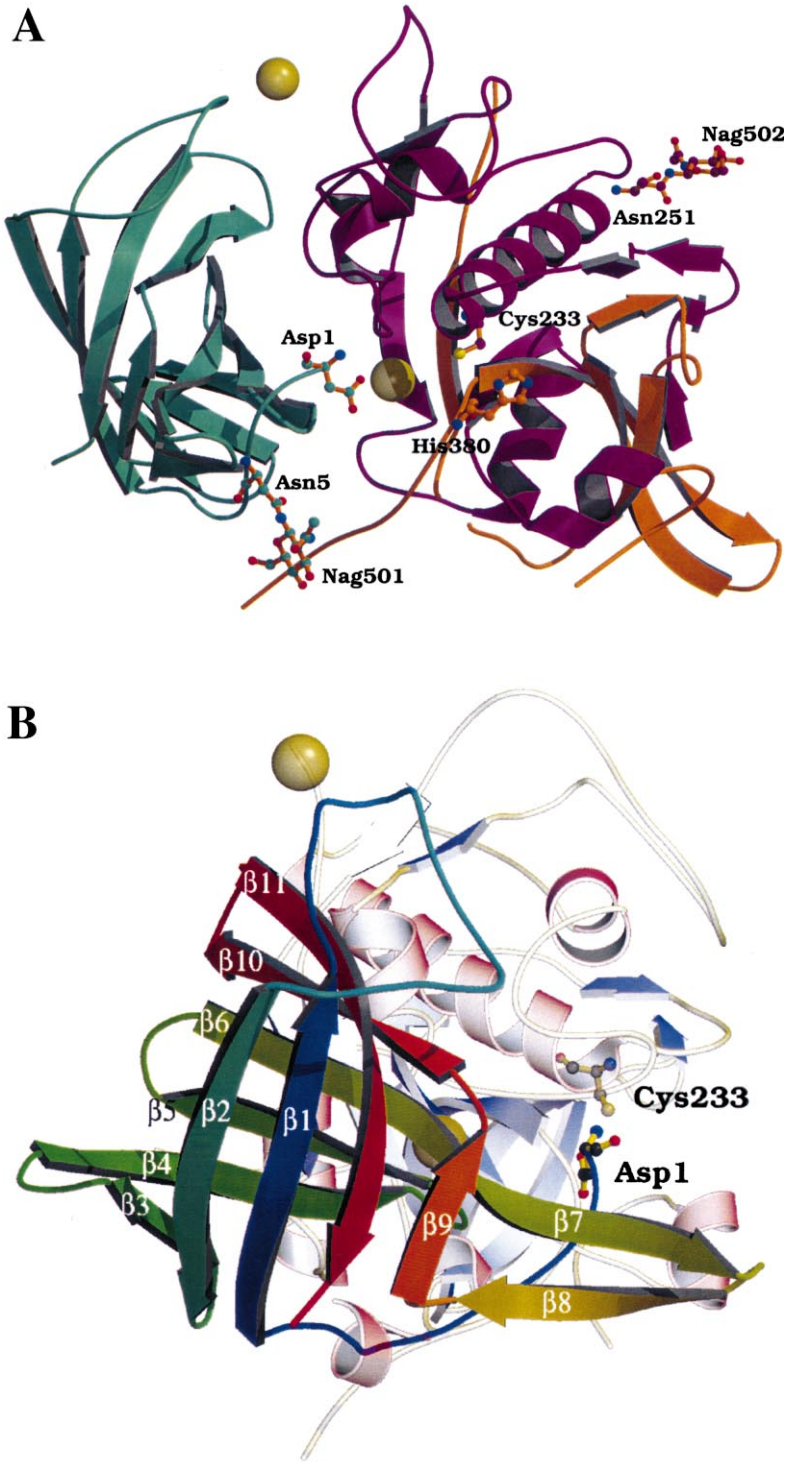

Fig. 1. The fold and domain structure of DPPI. A: A ribbon trace of of the subunit which consists of three polypeptide chains in the activated enzyme: the 118 residue long residual pro-part to the left (shown in turquoise) and the heavy and light chains (shown in magenta and orange, respectively) which constitute the papain fold domain. The pale yellow spheres represent chloride ions. Active site residues Cys233 and His380 as well as *N*-acetylglucosamine residues and their associated asparagines are shown in ball-and-stick representation. B: The eight-stranded meander β-barrel residual pro-part domain is rainbow-colored such that residue 1 is blue and residue 118 is red. Chlorides are represented by pale yellow spheres and Asp1 and the active site nucleophile are shown in ball-and-stick representation. The papain fold core domain is shown in the background. C: The tetrameric structure of DPPI. Residual pro-part domains are shown in gray and the papain fold domain in red, yellow, blue, and green, representing the different subunits. The tetramer assembles around the crystallographic 2-fold axes with one subunit in the asymmetric unit. Figures were prepared in Molscript [28].

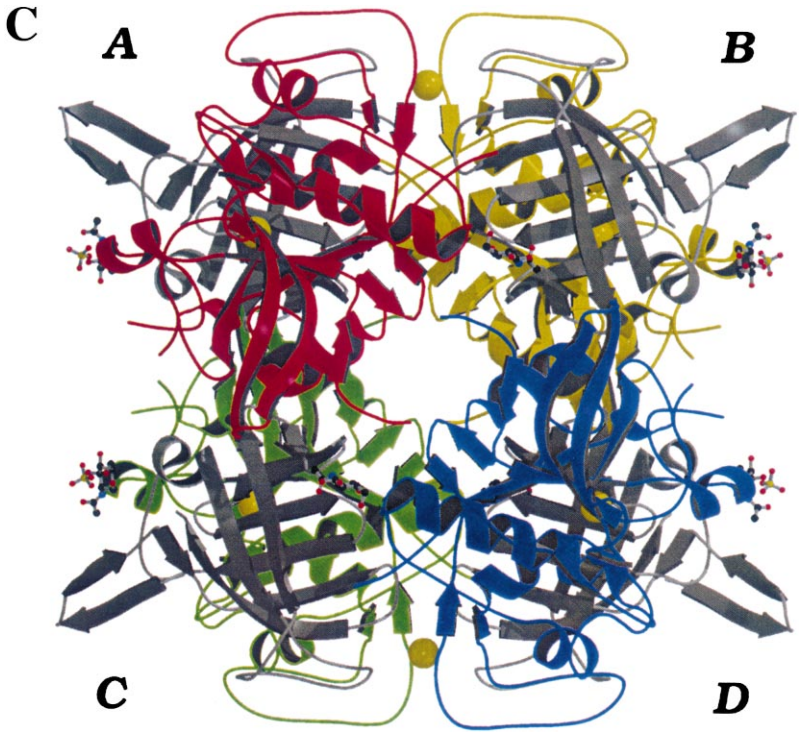

Fig. 1 (*Continued*).

symmetry. There is one contact surface along each symmetry axis, i.e. three per subunit. The four identical subunits are labelled A, B, C, and D (Fig. 1C).

Because the human pro-DPPI is dimeric [7] whereas proteolytically activated DPPI is tetrameric [6,7], the structure of the mature enzyme was analyzed to identify the original pro-DPPI dimer. The AB interface consists of residual pro-part residues of one subunit and core domain residues of the other. It covers 2×934 Å^2. The coil connecting rβ1–rβ2 as well as loops connecting rβ3–rβ4, rβ5–rβ6, and rβ10–rβ interact with the C-terminus of α1 and the C-terminal coil. A chloride ion on a crystallographic 2-fold axis connects arginines 25A and B. The AD interface covers 2×835 Å^2. The first three residues of the core domain plus the α4–β7 connecting loop interact with the rβ2–rβ3 junction. Finally, the AC interface covers the smallest area, 760 Å^2, and mainly involves the first six residues of the heavy chain.

In the zymogen, Leu204 would have to be exposed to allow proteolytic processing. In the tetramer this residue faces the central cavity. Thus, a hypothetical tetrameric proenzyme could not contain the AC and AD contacts that are identical to the ones observed here. The majority of the AD contacts would be permitted in a pro-DPPI dimer but in such a dimer, access for proteolytic enzymes to the heavy chain N-terminus might be sterically restricted. The existence of AB pro-DPPI dimers seems more likely. The AB interaction buries the largest surface of the three. The residues contributing to the AB contact are located on the surface of the catalytic domain opposite the heavy chain–light chain boundary, facilitating proteolytic activation. In addition, hydrophobic/van der Waals interactions dominate along the AB contact surface whereas dipole interactions are relatively more abundant at the AD interface. Notice that compared to the normal situation for dimers of similar mass, the buried surface areas for any dimeric constellation are rather small. This contrasts the values for the tetramer, which lie within the expected range [25].

4.2. The active site

The side chain of the N-terminal aspartyl residue is pointing into the active site cleft (Fig. 2). Superimposing DPPI onto the structure of the aminopeptidase cathepsin H reveals that the Asp1 carboxylate group is within hydrogen bonding distance of the main chain N of Thr83, the last residue in the 'mini-chain' octapeptide of cathepsin H situated in the S2 pocket [14]. The side chain carboxylate group points towards the S2 substrate binding site where it can bind to the N-terminal NH_3^+ group of the substrate, thereby directing dipeptidyl aminopeptidase specificity. Positive charges on lysine and arginine residues could interact with Asp1 resulting in a repositioning of the substrate, and explain why substrates with these side chains at the N-terminal are not cleaved. The introduction of a negative charge for binding the substrate terminal amino group is accomplished very differently in DPPI and cathepsin H. Whereas the negative charge in DPPI is donated by Asp1, the charge in cathepsin H is donated by a C-terminal carboxylate. The orientation of the carboxylate groups is different. The mini-chain in cathepsin H occupies the substrate binding pocket, in DPPI the N-terminus of the pro-part points into the S2 site, oriented perpendicular to the substrate binding cleft.

A chloride ion is found at the N-terminus of core domain helix 3, coordinating to Tyr279 N, Tyr322 Oη, and two water molecules (658 and 678). It is situated 12 Å from the nucleophilic thiol group deep in the core of the enzyme, albeit accessible from the active site. A previous study of the activation of DPPI by chloride ions led to the conclusion that a chloride ion was located in the S2 subsite, and a direct interaction

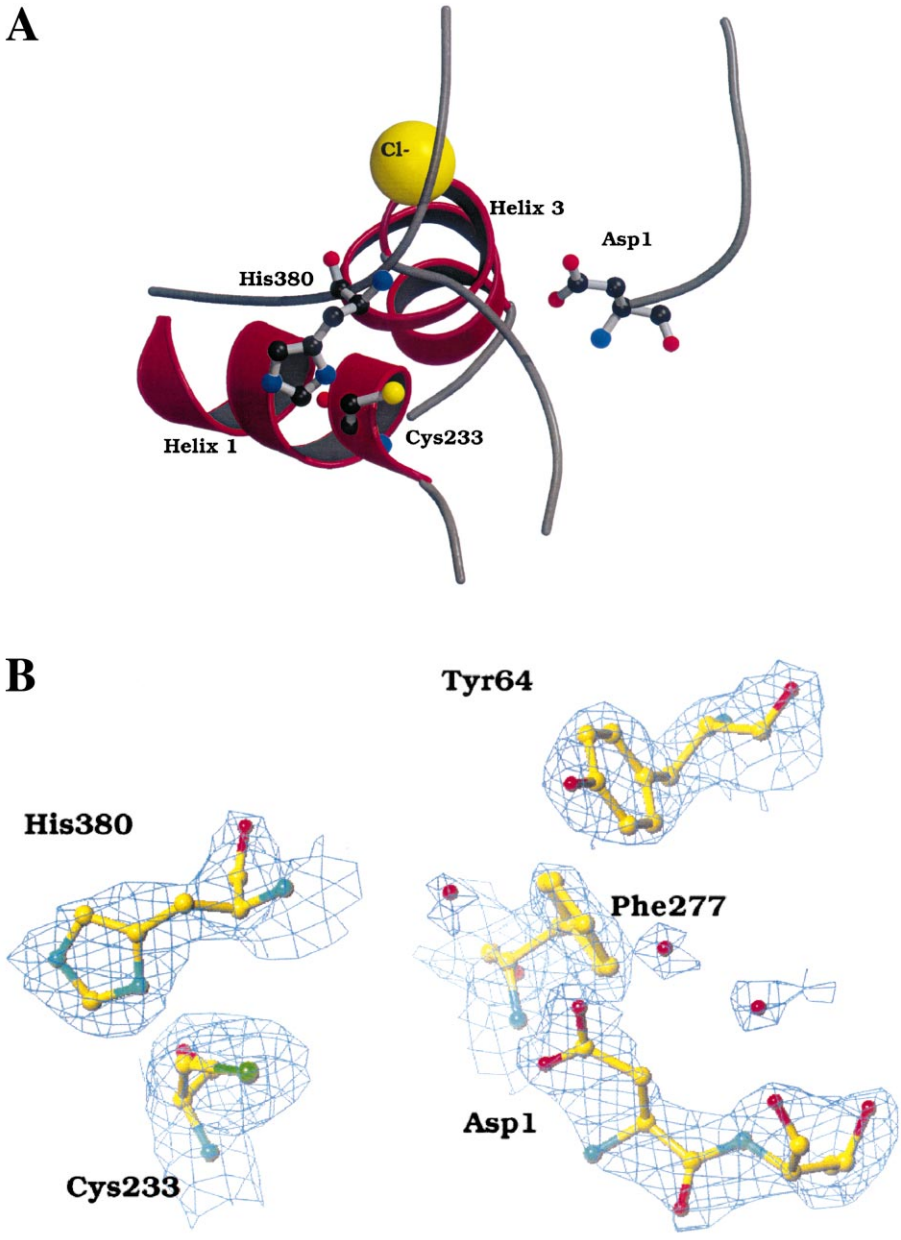

Fig. 2. The DPPI active site. A: The catalytic residues Cys233 and His380 and the substrate binding Asp1 residue are shown in ball-and-stick. Also shown is the chloride ion positioned at the N-terminus of the distorted helix 3. B: Electron density map contoured at 1.2σ superimposed on the model around the active site/substrate binding cleft. Figures prepared in Molscript [28] and TURBO-FRODO [19].

between the ion and the substrate amino-terminus was suggested [23]. Though the chloride ion is localized at the bottom of the S2 site, it is not likely to interact directly with the N-terminus of the substrate (Fig. 2A). The chloride is positioned such that it stabilizes helix 3 by interacting with the dipole [26]. Helix 3 is highly distorted and contains a stretch of π-helical 5/16 hydrogen bonding pattern as well ad a hydrogen bond between Ala282 O and Ser317, creating a bulge in the helix. The side chains of helix 3 residues 279, 280, 284 and 287–289 interact strongly with the residual pro-part rβ4, rβ5, rβ6, rβ10 sheet.

The unique exploitation of the residual pro-part domain in directing dipeptidyl aminopeptidase specificity may well be reflected in the oligomeric structure of DPPI. It was recently shown that the residual pro-part forms a stable domain that can be refolded in vitro independent of the papain core domain [27]. In addition, the core domain is probably a stable unit given its close structural and sequence homology to the other papain family peptidases. On this basis we hypothesize that the unique quarternary structure of DPPI mainly serves to stabilize the interaction between the residual pro-part and the papain core domains.

The structure and localization of the activation peptide in pro-DPPI remain unresolved. Its suggested amino acid sequence homology with the propeptides of the papain family enzymes [7] indicates a similar three-dimensional fold. However, super positioning of procathepsins L and K (PDB entries 1CS8 and 1BY8) onto the structure of DPPI shows that the activation peptide in pro-DPPI will have to obtain a very different conformation for it to be bound to the C-terminus

of the residual pro-part. Furthermore, the activation peptide will probably have to follow a path different from those of the procathepsins L and K propeptides to get around the residual pro-part.

Acknowledgements: This research is supported from the Danish National Research Foundation. We are grateful for beam time at MAX Lab synchrotron radiation facility in Lund, Sweden, and for the help from Drs. Salam Al-Karadaghi, Yngve Cerenius and Anders Svensson during these experiments.

References

[1] McGuire, M.J., Lipsky, P.E. and Thiele, D.L. (1993) J. Biol. Chem. 268, 2458–2467.
[2] Pham, C.T. and Ley, T.J. (1999) Proc. Natl. Acad. Sci. USA 96, 8627–8632.
[3] Sakai, K., Ren, S. and Schwartz, L.B. (1996) J. Clin. Invest. 7, 67–73.
[4] Toomes, C., James, J., Wood, A.J., Wu, C.L., McCormick, D., Lench, N., Hewitt, C., Moynihan, L., Roberts, E., Woods, C.G., Markham, A., Wong, M., Widmer, R., Ghaffar, K.A., Pemberton, M., Hussein, I.R., Temtamy, S.A., Davies, R., Read, A.P., Sloan, P., Dixon, M.J. and Thakker, N.S. (1999) Nat. Genet. 23, 421–424.
[5] Hart, T.C., Hart, P.S., Bowden, D.W., Michalec, M.D., Callison, S.A., Walker, S.J., Zhang, Y. and Firatli, E.J. (1999) Med. Genet. 36, 881–887.
[6] Dolenc, I., Turk, B., Pungercic, G., Ritonja, A. and Turk, V.J. (1995) Biol. Chem. 270, 21626–21631.
[7] Dahl, S.W., Halkier, T., Lauritzen, C., Dolenc, I., Pedersen, J., Turk, V. and Turk, B. (2001) Biochemistry 40, 1671–1678.
[8] Nikawa, T., Towatari, T. and Katunuma, N. (1992) Eur. J. Biochem. 204, 381–393.
[9] Cigic, B., Krizaj, I., Kralj, B., Turk, V. and Pain, R.H. (1998) Biochim. Biophys. Acta 1382, 143–150.
[10] Zhao, B., Janson, C.A., Amegadzie, B.Y., D'Alessio, K., Griffin, C., Hanning, C.R., Jones, C., Kurdyla, J., McQueney, M., Qiu, X., Smith, W.W. and Abdel-Meguid, S.S. (1997) Nat. Struct. Biol. 4, 109–111.
[11] Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J.S., To, R. and Huber, C.P. (1995) J. Biol. Chem. 270, 5527–5533.
[12] Guncar, G., Pungercic, G., Klemencic, I., Turk, V. and Turk, D. (1999) EMBO J. 18, 793–803.
[13] McGrath, M.E., Klaus, J.L., Barnes, M.G. and Bromme, D. (1997) Nat. Struct. Biol. 4, 105–109.
[14] Guncar, G., Podobnik, M., Pungercar, J., Strukelj, B., Turk, V. and Turk, D. (1998) Structure 6, 51–61.
[15] Lauritzen, C., Pedersen, J., Madsen, M.T., Justesen, J., Martensen, P.M. and Dahl, S.W. (1998) Protein Expr. Purif. 14, 434–442.
[16] Navaza, J. (1994) Acta Crystallogr. A 50, 157–163.
[17] Navaza, J. and Vernoslova, E. (1995) Acta Crystallogr. A 51, 445–449.
[18] Collaborative Computational Project number 4 (1991) Acta Crystallogr. D 50, 760–763.
[19] Roussel, A. and Cambillou, C. (1992) 'TURBO-FRODO', Biographics and AFMB (Arcitecture et Fonction des Macromolécules Biologiques), Marseille.
[20] Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Acta Crystallogr. D 54, 905–921.
[21] Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein, R. and Thornton, J.M. (1996) J. Biomol. NMR 8, 477–486.
[22] Vlassi, M., Cesareni, G. and Kokkinidis, M. (1999) J. Mol. Biol. 285, 817–827.
[23] Cigic, B. and Pain, R.H. (1999) Eur. J. Biochem. 264, 944–951.
[24] Holm, L. and Sander, C. (1993) J. Mol. Biol. 233, 123–138.
[25] Miller, S., Lesk, A.M., Janin, J. and Chothia, C. (1987) Nature 328, 834–836.
[26] Kortemme, T. and Creighton, T.E. (1995) J. Mol. Biol. 253, 799–812.
[27] Cigic, B., Dahl, S.W. and Pain, R.H. (2000) Biochemistry 39, 12382–12390.
[28] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.