ELSEVIER

Biochimica et Biophysica Acta 1340 (1997) 215–226

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

A prediction of DPP IV/CD26 domain structure from a physico-chemical investigation of dipeptidyl peptidase IV (CD26) from human seminal plasma

Anne-Marie Lambeir [a,*], José Fernando Díaz Pereira [b], Pablo Chacón [c], Geert Vermeulen [b], Karel Heremans [b], Bart Devreese [d], Jozef Van Beeumen [d], Ingrid De Meester [a], Simon Scharpé [a]

[a] *Laboratory of Medical Biochemistry, Department of Pharmaceutical Sciences, University of Antwerp (U.I.A.), B-2610 Wilrijk, Belgium*
[b] *Laboratory of Chemical and Biological Dynamics, Katholieke Universiteit Leuven, B-3001 Leuven, Belgium*
[c] *Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, E-28006 Madrid, Spain*
[d] *Laboratory of Protein Biochemistry and Protein Engineering, R.U. Gent, B-9000 Ghent, Belgium*

Received 25 November 1996; revised 27 February 1997; accepted 28 February 1997

Abstract

Human DPP IV, isolated from seminal plasma by means of immobilised adenosine deaminase, occurs in different forms which are distinguishable by net charge and native molecular weight. Charge differences arise primarily from different degrees of glycosylation containing various amounts of sialic acid. The majority of DPP IV isolated from total seminal plasma consists of the extracellular part of the protein starting at Gly-31. It is a very stable protein resisting high concentrations of denaturant. Unfolding experiments under reducing conditions are indicative of the existence of at least two domains which function independently. One of these domains is highly stabilised by disulfide bonds. Disruption of the disulfide bonds does not affect the activity, the dimeric state nor the adenosine deaminase binding properties of the protein but renders it more susceptible to proteolysis. The low-angle X-ray scattering spectrum is consistent with a model for a protein containing two subunits, each composed of three domains linked by flexible regions with low average mass. The secondary structure composition, determined by FTIR spectrometry, indicates that 45% of the protein consists of β-sheets, which is higher than expected from computed secondary structure predictions. Our results provide compelling experimental evidence for the three-domain structure of the extracellular part of DPP IV.

Keywords: Dipeptidyl peptidase IV; CD26; Proline; FTIR; Low-angle X-ray scattering; Protein folding

Abbreviations: DPP IV, dipeptidyl peptidase IV EC 3.4.14.5; Gly-Pro-4-OMe-2-NA, glycyl-prolyl-4-methoxy-2-naphthylamide; Lectin SNA, *Sambucus nigra* agglutinin; Lectin MAA, *Maackia amurensis* agglutinin; Lectin DSA, *Datura stramonium* agglutinin; Lectin GNA, *Galanthus nivalis* agglutinin; Lectin PNA, *Arachis hypogaea* (peanut) agglutinin; FTIR, Fourier Transform Infrared Spectroscopy; BSA, bovine serum albumin; ADA, adenosine deaminase EC 3.5.4.4; Maldi-TOF MS, matrix assisted laser desorption ionisation-time of flight mass spectrometry

* Corresponding author. Fax: +32-3-820 2745; E-mail: lambeir@uia.ua.ac.be

PII S0167-4838(97)00045-9

1. Introduction

The interaction with adenosine deaminase (ADA) was shown to be a unique property of dipeptidyl peptidase IV (DPP IV), which is also known as the activation antigen CD26 expressed on the surface of specific subsets of T-lymphocytes [8,20,41]. Some recent studies describe molecules with DPP IV activity which differ in several aspects from DPP IV/CD26, notably in their ability to bind ADA [11,12,19,21]. Previously it was reported that human seminal plasma is a suitable source for the isolation of human DPP IV/CD26 [9,24,43]. In this paper we describe some of the physicochemical properties of human seminal DPP IV/CD26 purified by affinity chromatography on immobilised ADA.

There is evidence from earlier studies that human seminal DPP IV/CD26 exists in different forms. Fractionation of human seminal proteins by anion exchange chromatography and size exclusion chromatography indicated heterogeneity of DPP IV, both in charge and native molecular mass [24]. The intact molecule is an integral membrane protein and as such it was found associated with prostasomes, complex membrane vesicles secreted by the prostate gland [13]. However, after removal of the prostasomes, a significant amount of DPP IV/CD26 remained in the soluble fraction [9].

Although there is good evidence that only one gene product is expressed [1], different molecular forms of DPP IV/CD26 were observed, reflecting differences in post-translational modification which appeared to be tissue specific and/or activation dependent in human lymphocytes [21,25,45]. Furthermore, mature DPP IV/CD26 carries complex hybrid carbohydrate chains containing sialic acid, which were shown to be important for its secretion and, possibly, salvage in epithelial cells [32,33]. The nucleotide sequence of the human DPP IV/CD26 gene and the derived amino acid sequence have been published [28,29,42]. It is a type II transmembrane protein, anchored to the lipid bilayer by a single hydrophobic segment located at the N-terminus (7–29) with a cytoplasmic tail of only 6 amino acids. A flexible 'stalk' links the membrane anchor with a large glycosylated region (48–324), a cysteine-rich region (325–552) and the C-terminal catalytic domain (553–766). The catalytic domain shows sequence homology with the catalytic domains of proteins belonging to the oligopeptidase family. The 3-D structure of this class of peptidases has not yet been determined, but it has been postulated that they possess the α/β hydrolase fold [14]. The residues of the catalytic triad were identified by site-directed mutagenesis [6]. The catalytic activity was not affected by the occupancy of the ADA binding site [8]. Several investigators reported that human DPP IV functions as a dimer in solution [4,7,16,24,27,38]. The interaction of DPP IV/CD26 with collagen was shown to reside in the cysteine-rich region [26,33].

2. Material and methods

2.1. Purification

DPP IV was isolated from pooled seminal plasma obtained after removal of spermatozoa from semen of healthy donors. Prostasomes were isolated by high-speed centrifugation [9]. Membrane bound DPP IV/CD26 was purified after solubilisation of the prostasomes with 1% Triton X-100. Triton X-100 (0.1%) was included in all the purification steps except the last one. Soluble DPP IV/CD26 was isolated from the supernatant after removal of the prostasomes, with or without addition of detergent. When Percoll was used for the isolation of spermatozoa, prostasomes were not isolated; instead seminal plasma was treated with 1% Triton and soluble and membrane bound DPP IV/CD26 were processed together, without including detergent in subsequent steps. The purification procedure [9] consisted of (1) ion exchange chromatography on DEAE-Sepharose fast flow (Pharmacia Biotech, Uppsala, Sweden) and (2) affinity chromatography on ADA coupled to cyanogen bromide activated Sepharose CL6B. A final purification/concentration step on a 1 ml HighTrapQ (Pharmacia Biotech, Uppsala, Sweden) was included before concentration on a centrifugal concentrator (except for the sample prepared for the X-ray scattering study which was concentrated immediately after elution from ADA-Sepharose). The specific activity of the enzyme after purification ranged between 31 and 75 units/mg. Protein concentrations were determined with the Bradford method using BSA as a standard.

2.2. Enzymatic activity

Enzymatic activity was determined with the fluorogenic substrate Gly-Pro-4-OMe-2-NA (Sigma, St. Louis, MO, USA). One unit of activity is defined as the amount of enzyme which cleaves 1 μmol of substrate per minute at 37°C in 50 mM Tris-HCl buffer, pH 8.3, containing 1.4 mM substrate [7]. These assay conditions were chosen to optimise the buffering capacity (pK_a of 8.1) without significant decrease in activity and to eliminate consumption of the substrate by dipeptidyl peptidase II (acidic pH optimum) and aminopeptidase M (inhibited by Tris), which are both present in seminal plasma.

2.3. Ion exchange and gel filtration experiments

For the isolation of DPP IV from total seminal plasma, a DEAE-Sepharose fast flow column of 5-cm diameter and 20-cm length was used. The majority of the Percoll was removed by centrifugation at 10,000 $\times g$ during 1 h, the supernatant was then treated with 1% Triton X-100 and diluted 3-fold in 20 mM Tris-HCl, pH 7.4 (154 units in 300 ml). The column was washed with 5 l of 20 mM Tris-HCl, pH 7.4, containing 70 mM NaCl and eluted with a 2-l linear gradient from 70 to 350 mM NaCl in 20 mM Tris-HCl, pH 7.4. The recovery was 92%. Eluted proteins were pooled as described in Section 3.1 and further purified by ADA-Sepharose affinity chromatography. After loading and washing, the ADA-column (6 ml) was placed in series with a HighTrap Q column (1 ml) so that the eluted protein (80 ml) was immediately bound to the anion exchanger. The HighTrapQ column was washed with 10 ml of 10 mM Tris-HCl, pH 7.4, and eluted with a linear gradient (20 or 40 ml) from 0 to 0.5 M NaCl in 10 mM Tris-HCl, pH 7.4, collecting 1-ml fractions.

Peak fractions of the HighTrapQ column (0.5 ml, typically 10 units) were applied on a Sephacryl S300 HR column (1.6-cm diameter, 57-cm length) equilibrated with 10 mM Tris-HCl, pH 7.4, 150 mM NaCl at 0.25 ml/min. Fractions (1.25 ml) were collected after the void volume (29.5 ml). The column was previously calibrated with thyroglobulin (669 kDa), ferritin (440 kDa) (Pharmacia Biotech) and BSA (67 kDa) (Sigma, St. Louis, MO, USA).

2.4. Unfolding

Unfolding experiments were performed with a sample of DPP IV/CD26 (0.61 mg/ml at 50 U/mg), purified from seminal plasma after removal of the prostasomes, which contained primarily low-salt dimers. Stock solutions of 8 M urea and guanidinium chloride were made in 50 mM Tris-HCl adjusted to pH 7.4 (with or without 10 mM DTT) and diluted in the same buffer. DPP IV (10 or 20 μl) was added to 2 ml of denaturant and incubated at 4°C for at least 24 h. Fluorescence spectra were determined with a Shimadzu RF-5000 fluorimeter. The excitation wavelength was 285 nm, the emission range 300–400 nm, bandwidth 5 and 1.5 nm, respectively. After recording the fluorescence spectra, the samples were diluted 175-fold in 50 mM Tris-HCl, pH 8.3, and the activity was determined without delay.

Prior to proteolysis, samples were reduced and partially denatured by addition of 10 mM DTT (Sigma, St. Louis, MO, USA) in presence of 6 M urea, Molecular Biology grade (Merck, Darmstadt, Germany), and 20 mM Tris buffer, pH 8, for 30 min at 37°C. The reduced thiol groups were alkylated by addition of 50 mM of freshly dissolved iodoacetamide (Sigma, St. Louis, MO, USA) during a 15-min incubation at room temperature. The reduced and alkylated sample was transferred to 10 mM Tris buffer, pH 7.4, by gel permeation chromatography on a PD10 column (Pharmacia Biotech, Uppsala, Sweden). For partial proteolysis experiments, 0.15 mg of reduced and alkylated DPP IV was incubated overnight at 37°C with 1 mg/ml trypsin (Boehringer Mannheim, Germany) in 10 mM Tris buffer, pH 7.4, containing 1 mM $CaCl_2$ or with 1 mg/ml thermolysin or proteinase K (Boehringer Mannheim, Germany) in 10 mM Tris buffer, pH 7.4, containing 10 mM $CaCl_2$ and 4.6 M urea.

2.5. Gel electrophoresis, blotting and immunoassay procedures

SDS polyacrylamide gel electrophoresis and electroblotting on PVDF membranes was carried out using a Mini Protean II apparatus (Biorad, Richmond, CA, USA) following the instructions provided by the manufacturer.

For the detection of glycoproteins by dot-blot, we used the DIG-Glycan differentiation kit from Boehringer Mannheim (Germany). Nitrocellulose blotting paper was obtained from BioRad (Richmond, CA, USA). Neuraminidase was from Sigma (St. Louis, MO, USA) and *N*-glycosidase F/endoglycosidase F from Boehringer Mannheim (Germany).

To determine the relative degree of glycosylation of the different DPP IV/CD26 samples, a microtiterplate immunoassay was used. DPP IV/CD26 was captured by the monoclonal antibody TA5.9 [8] which recognises the ADA binding site without affecting the enzymatic activity. Binding of DIG-labelled lectins (SNA, MAA, DSA, GNA, PNA) was detected by means of an anti-DIG antibody conjugated with alkaline phosphatase (Boehringer Mannheim, Germany). The activity of DPP IV in the wells was measured by addition of 0.5 mM Gly-Pro-*p*-nitroanilide (10 mM Tris, pH 7.5, 150 mM NaCl).

2.6. Mass spectrometry and N-terminal amino acid sequencing

Mass analysis was performed on a Tofspec SE Maldi-TOF analyser (Micromass, Altrincham, UK) using sinappinic acid as the matrix. The instrument is equipped with a N_2-laser and was used in the linear mode using 25-kV voltage in the source [22].

Sequence analysis was performed on a 476A protein sequencer (Perkin-Elmer, Applied Biosystems Division, Foster City, CA, USA).

2.7. Infrared spectrometry

DPP IV was transferred to 2 mM ethylmorpholine acetate buffer, pH 8.3, by passage over a PD10 column and dried by vacuum centrifugation (Speedvac). The residue was dissolved in D_2O at 40 mg/ml. D_2O was used as a solvent to avoid interfering H_2O absorption around 1640 cm^{-1}. The solution was mounted in a stainless-steel gasket of a diamond anvil cell (Diacell Products, Leicester, UK). The initial gasket thickness was 0.075 mm and the hole diameter 0.6 mm. The infrared spectra were obtained with a Bruker IFS66 FTIR spectrometer equipped with a MCT detector. The infrared beam was focussed on the sample by means of a NaCl lens [44]. The sample compartment was continuously purged with dry air. The obtained spectrum was the result of 700 co-added interferograms at a resolution of 2 cm^{-1}. The time needed to record the spectrum was about 15 min.

The secondary structure determination was done with a method which is a combination of self-deconvolution and band fitting [15]. The spectrum was analysed with a full width at half height of 20 cm^{-1} with a triangular squared apodization function, truncating 80% of the interferogram, resulting in a resolution enhancement factor of 1.7. The composition of the secondary structure was found by fitting the deconvoluted spectrum with Gaussian curves. The area of different subcomponents can be associated with the different types of secondary structure [3]. The Fourier self-deconvolution and the band fitting were performed with a program developed in the Laboratory of Chemical and Biological Dynamics [40].

2.8. Synchrotron X-ray scattering in solution

X-ray scattering experiments were performed with the soluble dimeric form of DPP IV at a concentration of 8 mg/ml in 10 mM phosphate buffer, pH 7.4. X-ray scattering data were collected at station 2.1 of the Daresbury Laboratory Synchrotron Radiation Source. Data acquisition and processing were as described previously [2,10]. Absolute values of the scattering vector (defined as $S = 2(\sin\theta)/\lambda$, where 2θ is the angle of incident to scattered radiation and λ is the X-ray wavelength) were calibrated in this work employing all the observable diffraction orders of the 67-nm repeat in wet tail collagen. A 3-m camera length was employed, effectively covering an S range from 0.03 to 0.31 nm^{-1}. The low-resolution X-ray solution scattering patterns were modelled using a software package (Chacón et al., in press), that uses a genetic algorithm based on Debye's formula [17] to fit the experimental data. The DPP IV dimer has been modelled from a box of $12 \times 12 \times 12$ nm containing 752 1.5-nm spheres used as starting point for the search. More refined starting models were constructed successively up to the maximum resolution possible, i.e. starting from 1-nm spheres.

2.9. Secondary structure prediction

Secondary structure and accessibility predictions were done via the E-mail facility of the EMBL

laboratory in Heidelberg according to the PDH method [34–37]. The alignment contained the following sequences: human DPP IV, mouse DPP IV, rat DPP IV, pig DPP IV, human DPP IV-like protein, rat DPP IV-like protein, bovine DPP IV-like protein, yeast dipeptidyl aminopeptidase II.

3. Results

3.1. DPP IV forms differing in net charge and native molecular mass

DPP IV activity in seminal plasma eluted over a broad salt concentration range from an anion exchange column. Pools were made of active fractions eluting between 0.1–0.15 (pool I), 0.15–0.20 (pool II), 0.20–0.25 (pool III) and 0.25–0.32 M salt (pool IV) containing 31%, 35%, 27% and 7% of the total activity eluted from the column. These pools were separately purified on ADA-Sepharose. The resulting DPP IV/CD26 proteins eluted at different salt concentrations on a HighTrapQ column, indicating two distinct populations. We called these populations 'low-salt' (LS) and 'high-salt' (HS) referring to their elution position on both type of anion exchange columns, DEAE and Q-Sepharose. The LS/HS proportion was estimated to be 8:2 in pool I, 3:7 in pool II, 2:8 in pool III (pool IV was not processed). When samples of each population (from pool I and II) were placed on a gel filtration column, the low-salt form was found exclusively in a 290-kDa peak corresponding to the DPP IV dimer. The high-salt form eluted in two peaks with molecular weights of 900 and 290 kDa, roughly in a 4:6 proportion. Therefore, about 75% of the activity eluted from the DEAE column was in a dimeric state (LSD and HSD), the remainder had a higher molecular weight (HSHM). A silver stained SDS-gel of the high-molecular-weight form did not reveal any other protein apart from DPP IV. This heterogeneity was also observed when DPP IV was purified from isolated prostasomes (Fig. 1).

There is no difference in specific activity between the dimeric and high-molecular-weight forms. N-terminal amino acid sequencing of the low- and high-salt eluting forms from total seminal plasma showed no difference in N-terminal sequence, with the main sequence starting at Gly-31 (GTDDA), indicating that both forms had lost their membrane anchor (probably by proteolysis after solubilisation of the membranes with detergent). The sample eluting at higher salt concentration showed a certain degree of proteolytic degradation with fragments starting at Arg-40 (RKTYTL) and at Val-252 (VRVPY). SDS-PAGE indicated that the subunit molecular weight of all forms is very similar, both high-salt forms being marginally larger (Fig. 2). This was confirmed by mass spectrometry giving values of 104 and 106 kDa.

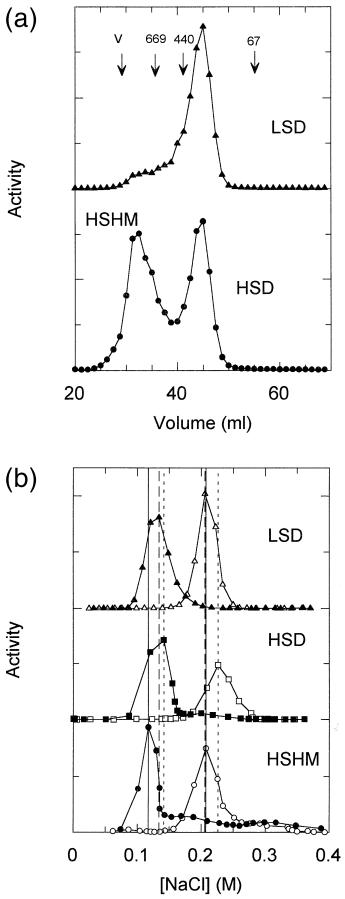

Fig. 1. (a) Sephacryl S300HR gel permeation chromatography of prostasomal DPP IV: (●) 'high-salt' fraction containing the dimeric form (HSD) and a high-molecular-weight form (HSHM), (▲) 'low-salt' fraction consisting mainly of dimeric DPP IV (LSD). The void volume (V) and the position of calibrating proteins is indicated by arrows with their molecular weight in kDa. The two graphs are off-set vertically for clarity. (b) Elution profile of different DPP IV forms on HighTrapQ. Open symbols represent native proteins, closed symbols the same preparation after incubation with neuraminidase and *N*-glycosidase F/endoglycosidase F. (●○) high-molecular-weight form (HSHM), (■ □) 'high-salt' dimeric form (HSD), (▲ △) 'low-salt' dimeric form (LSD). The three graphs are off-set vertically for clarity.

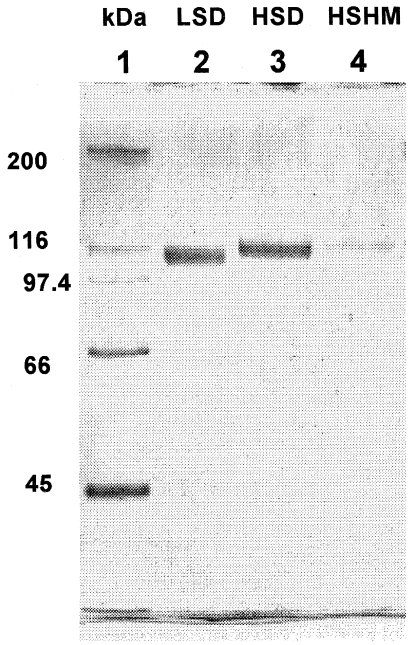

Fig. 2. 7.5% SDS-PAGE of different forms of DPP IV isolated from detergent-treated total seminal plasma. Lane 1, broad-range-molecular-weight standards; lane 2, low-salt dimeric form (LSD); lane 3, high-salt dimeric form (HSD); lane 4, high-salt high-molecular-weight form (HSHM).

The glycosylation pattern of all DPP IV preparations was the same, characterised by binding of lectins SNA, MAA and DSA, indicative of a complex type of glycosylation containing terminal sialic acid. When dilutions of the different DPP IV preparations were spotted on blotting paper, it became evident that the high-molecular-weight form stained more intensely (DSA) than the dimeric forms, indicating a higher degree of glycosylation. This result was confirmed using the microtiterplate assay (Table 1). Incubation (24 h, 37°C) of the various DPP IV samples with neuraminidase (1.2 U/ml), and a mixture of *N*-glucosidase F (140 U/ml) and endoglucosidase F (6 U/ml), prior to HighTrap Q chromatography shifted the elution peaks to lower salt concentrations by 100 mM in all cases except for a fraction of the activity associated with the 900-kDa peak (prostasomal preparation) which appeared resistant (Fig. 1b).

Table 1
Glycosylation pattern of different DPP IV/CD26 forms described in this paper

	Alkaline phosphatase activity/ DPP IV activity				
	DSA	GNA	MAA	PNA	SNA
Prostasomal HSM	0.367	n.d.	0.94	n.d.	4.5
Prostasomal HSD	0.048		0.73		4.9
Prostasomal LSD	0.050		0.84		4.9
Soluble HSM	0.304		0.83		4.3
Soluble HSD	0.044		0.87		3.2
Soluble LSD	0.061		0.57		2.5

Selectivity of the lectins: DSA, galactose-β(1-4)-*N*-acetylglucosamine in complex and hybrid glycan structures; GNA, terminally linked mannose; MAA, sialic acid terminally linked α(2-3) to galactose; PNA, galactose β(1-3)-*N*-acetylglucosamine in the core of O-linked sugars; SNA, α(2-6) linked sialic acids. n.d. = not detectable.

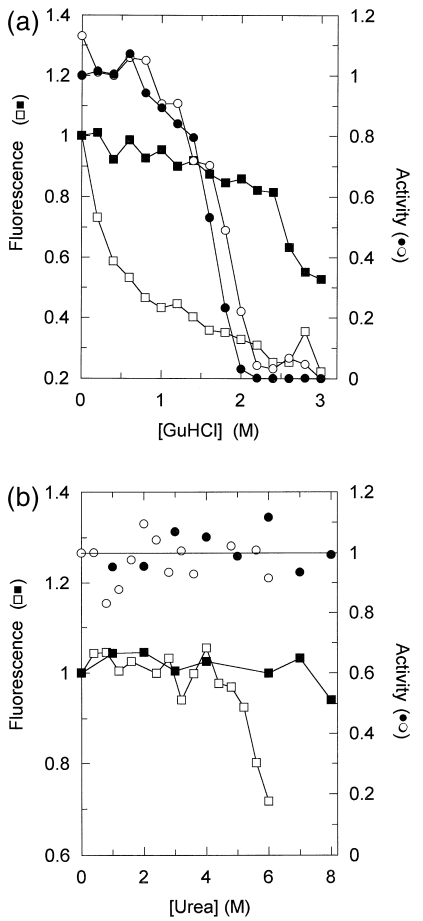

Fig. 3. (a) Unfolding of DPP IV in guanidinium chloride in absence (closed symbols) and in presence of DTT (open symbols). Circles (●○) represent activity after dilution in reaction buffer, squares (■ □) the fluorescence at 330 nm (ex 285 nm). (b) Unfolding in urea in absence (closed symbols) and in presence (opensymbols) of DTT. Circles (●○) represent the relative activity after dilution in reaction buffer, squares the fluorescence at 348 nm (■) or 330 nm (□)(ex 285 nm).

3.2. Unfolding in urea and guanidinium chloride

In absence of reducing agents, DPPIV/CD26 retained its native conformation in 8 M urea: no change in fluorescence spectra occurred and all activity remained after dilution in reaction buffer. In the presence of 10 mM DTT, a transition was observed at 5.6 M urea which was accompanied by a decrease in fluorescence and a shift of the emission spectrum towards higher wavelengths (Fig. 3). The transition appeared to be reversible and was not associated with a permanent loss of activity.

Unfolding of DPP IV/CD26 by guanidinium chloride caused an irreversible transition around 1.8 M, inactivating the enzyme. This transition was characterised mainly by a shift of the emission peak (335 nm) towards higher wavelengths (355 nm). A second transition occurred at 2.5 M guanidinium chloride, associated with a decrease in fluorescence. In the presence of 10 mM DTT, the decrease in fluorescence was observed at much lower concentrations of denaturant (0.4 M) whereas the process associated with the wavelength shift and the loss of activity remained unaffected around 1.8 M (Fig. 3).

When soluble DPP IV was reduced and alkylated in the presence of 6 M urea, the resulting protein was fully active (after removal of the denaturant) and had the same molecular weight as the native dimer as judged from gel filtration results. It was fully retained on ADA-Sepharose.

3.3. Partial proteolysis

Incubation with an excess of trypsin, chymotrypsin, thermolysin, proteinase K, pepsin or papain for extensive lengths of time did not cause degradation of native soluble DPP IV. Trypsin cleaved the reduced and alkylated form to give a major fragment of 92 kDa. N-terminal sequencing of the blotted protein indicated some heterogeneity with a major sequence of K(S)YTADYDI(S)D, consistent with cleavage after Arg-125. In the presence of 4.6 M urea some degradation occurred after incubation with thermolysin, proteinase K and papain with large proteolytic fragments of 78, 76 and 62 kDa visible on silver stained SDS-PAGE gels. Except for papain, the incubation with the proteases did not greatly affect the DPP IV enzymatic activity.

3.4. Structural investigations

Fig. 4 shows the calculated X-ray scattering profile of the best model obtained in comparison with the experimental data. The DPP IV dimer structure in

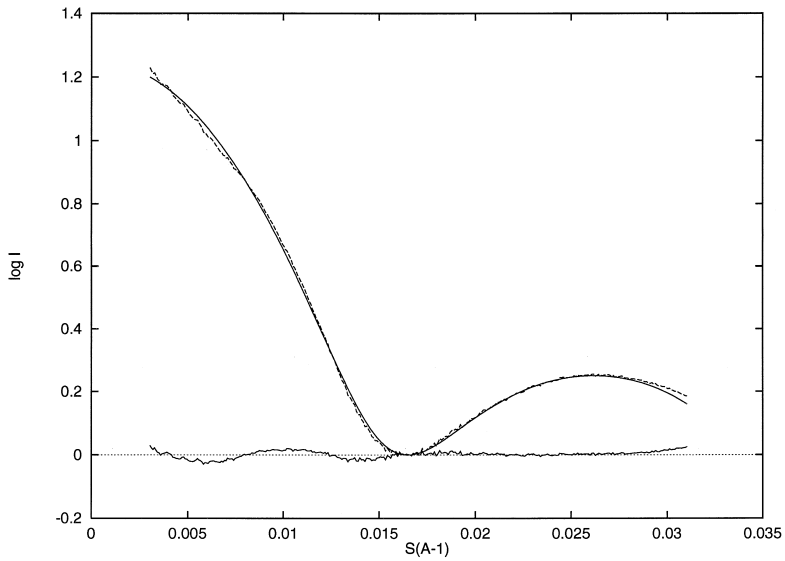

Fig. 4. Fitting of the X-ray scattering profile of DPPIV 8 mg ml^{-1} in 10 mM phosphate buffer pH 7.4. The experimental data (broken line) represent the average of four samples with a total acquisition time of 1 h. The solid line represents the computed X-ray scattering profile of the best model found.

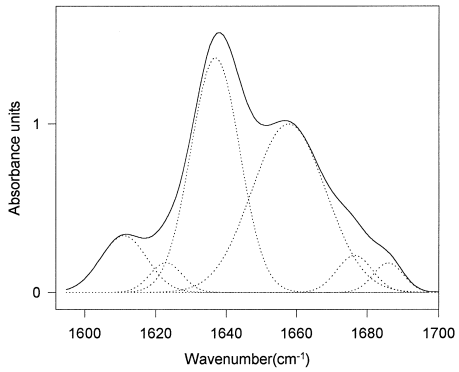

Fig. 5. FTIR spectrum of the amide I′ region of DPP IV. Fitted peaks used for the assignment of secondary structure elements (Table 2) are shown in dotted lines. The peak at 1611 cm^{-1} was assigned to amino acid side-chains.

Table 2
Secondary structure assignment of the FTIR spectrum of DPP IV

Band position (cm^{-1})	Band area (%)	Assignment
1623	3	β-sheet
1637	42	β-sheet
1658	46	α-helix/unordered
1676	5	turn/bend
1686	3	turn/bend

solution could be represented by 38 spheres of 1-nm diameter. The model consisted of two monomers, each composed of three domains — two large and

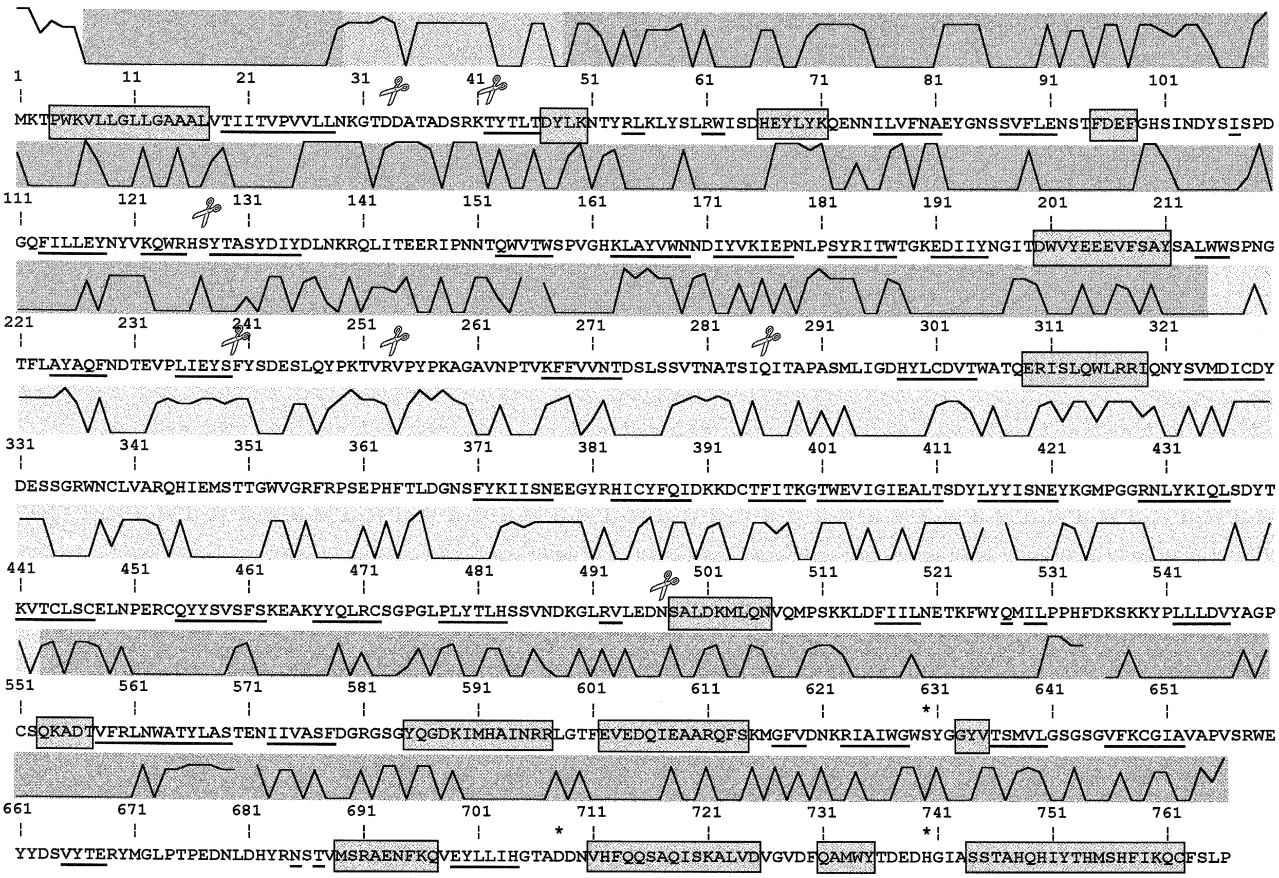

Fig. 6. Amino acid sequence and secondary structure prediction of human DPP IV/CD26. α-helices are placed in rectangles, β-sheets are underlined. The plots located above the primary sequence represent the predicted accessibility of the residues using integer numbers on a scale of 0 to 9. The shaded blocks in the background indicate the different regions: cytoplasmic tail, transmembrane region, flexible stalk region, glycosylated region, cystein-rich domain and protease domain. Scissors are placed near various protease cleavage sites as discussed in the text, the catalytic triad is indicated by $*$.

one small domain — probably connected via flexible regions that could not be detected at the resolution of the data. The radius of giration was 3.2 nm and the dimensions $7.0 \times 7.7 \times 5.2$ nm.

The amide I′ region of the DPP IV infrared spectrum is shown in Fig. 5. The deconvoluted spectrum showed bands at 1611, 1623, 1637, 1658, 1676 and 1686 cm^{-1}. The percentages of the different bands are given in Table 2. The total amount of β-sheet is 45%. The band at 1658 cm^{-1} has been assigned to α-helix overlapping with unordered coil [3]. These two components cannot be separated further. The band at 1611 cm^{-1} has been assigned to amino acid side-chains [5].

The secondary structure prediction for the extracellular domain (starting at Gly-31) of DPP IV was 17% α-helix, 33% β-structure and 50% coil. The results of the secondary structure prediction are shown in Fig. 6 representing the position of the secondary structure elements, the accessibility of individual residues and proteolytic cleavage sites in defined regions of the protein.

4. Discussion

The different molecular forms of human seminal DPP IV derive from two processes: (1) proteolytic action by proteases resulting in the loss of the membrane anchor; and (2) various degrees of glycosylation, intrinsically or due to degradation during the processing of the samples. Seminal plasma contains many degrading enzymes including proteases and glycosidases [39]. The majority of the protein however behaves as a dimer, regardless of the presence of the N-terminal membrane spanning domain. The differences in net charge can be explained as differences in number of sialic acid residues (see Fig. 1 and Table 1, binding of MAA and SNA lectins) although other types of post-translational modifications cannot be excluded [21,25,45]. In our preparations an estimated 25% of the activity is associated with a high-molecular-weight form. This form appears to be highly glycosylated and partially resistant to glycosidases. From Table 1 can be deduced that, overall, the prostasomal samples contained slightly more carbohydrates than DPP IV/CD26 isolated from seminal fluid (after removal of the prostasomes). It seems reasonable to assume that all DPP IV/CD26 in seminal plasma derives from prostasomes and is progressively degraded once it is released from the membrane. However, the binding of DSA lectin by the high molecular weight form is 5 to 7 times higher than by the dimeric forms in both preparations. DSA recognises galactose-β(1-4)-*N*-acetylglucosamine in complex and hybrid glycan structures. The same order of difference is not observed with MAA and SNA, which recognise α(2-3) and α(2-6) linked sialic acid. The origin and potential role of the high-molecular-weight form remains a mystery. It is not contained in intact prostasomes or membrane fragments since it shows only 1 protein band on SDS-PAGE gel. The same argument excludes that it is a complex of DPP IV/CD26 with proteinaceous ligands such as ADA or collagen. It cannot be excluded that it is a non-specific aggregate of DPP IV molecules which stick together because they have not lost their membrane anchor, are partially denatured or remain associated with detergent micelles. One may ask whether this is consistent with a higher negative charge, more sialic acid residues, more exposure of galactose-b(1-4)-N-acetylglucosamine for DSA binding, yet no effect on TA5.9 binding or specific activity. It is interesting to note that in human plasma a small percentage of the DPP IV activity eluted at 550 kDa from a gel filtration column, although the identity of this protein was not confirmed by ADA binding or amino acid sequence analysis [23].

The major form isolated from detergent-treated seminal plasma consists of the extracellular part of DPP IV. It is a highly stable protein requiring high concentrations of denaturant for unfolding. The resistance of DPP IV activity against proteolysis is well known and it was used by some investigators as a hallmark of its identity [19]. Reduction and alkylation of DPP IV's 12 cysteines under mildly denaturing conditions facilitates proteolysis to a certain extent without affecting the affinity for ADA, the activity of the protein or its dimeric state. The results of the unfolding experiments are compatible with the existence of at least two domains, one of which is highly stabilised by disulphide bonds. The catalytic activity is not affected by disturbance of this domain and the intersubunit interactions are mostly intact.

The X-ray scattering analysis shows a molecule

with less mass than expected for a molecule with an apparent molecular weight of 290 kDa, while the radius of giration (3.2 nm) is only slightly smaller than expected. The model obtained by the modelling package shows several areas of high mass density in a symmetrical arrangement – as expected for a dimeric molecule – with large empty spaces between them. This suggests that each of the two subunits is composed of two or three domains, connected by flexible regions of low mass. Flexible regions cannot be detected in a low-angle scattering pattern since only their average distribution of mass scatters the X-rays. If the mass of such fragments is low enough they will not be detected, as in this case. Highly flexible and disordered glycan chains may also escape detection and may account for the apparent mass deficit. The native molecular weight of DPP IV determined in this study is higher than previously reported for the lymphocytic protein (260 kDa), using exactly the same conditions, though probably still within experimental error [7]. Our value agrees with published data for human DPP IV from placenta ($>$ 200 kDa), urine (280 kDa), plasma (250 kDa) and cultured fibroblasts (400 kDa) [4,23,32,38]. Ohkubo et al. [31] reported values of 290 and 310 kDa for DPP IV isolated from porcine seminal plasma.

The results of this study provide compelling additional experimental support that DPP IV is a homodimer, each subunit containing three domains. So far, the domain structure was primarily implied from the position of cysteines, putative glycosylation sites and catalytic residues in the primary structure of the protein. The catalytic domain is most easily delineated since it shows homology with other members of the oligopeptidase family. Goossens et al. [14] started their alignment at Ser-537. It is composed of alternating β-sheets and α-helices connected by loops of variable length, typical for an α/β folded structure. Approximately two-thirds of all residues predicted to be in α-helices are found in the catalytic domain. The remainder of the molecule contains 4 helical segments defined with high accuracy ($>$ 82%) near positions 66–71, 200–211, 309–319 (glycosylated region) and 498–506 (cysteine-rich domain). The FTIR data indicate that the total percentage of β-structure is underestimated in the secondary structure prediction (45% versus 33%). This implies that the glycosylated region and the cysteine-rich domain contain more secondary structure than expected and are essentially all β. Attempts to obtain isolated domains by partial proteolysis of the native protein were not successfull. Our preparations, however, contained a small amount of cleaved material starting at Val-252. The 60-kDa protein which copurified with rat kidney DPP IV started 32 residues further along the polypeptide chain [18]. Digestion of DPP IV with V8 protease under mildly denaturing conditions yielded a 33-kDa fragment, identified as Tyr-238 to Glu-495 which was shown to contain the collagen binding site [26]. McCaughan et al. [27] found a 70–80-kDa fragment in their preparations of DPP IV from rat kidney extracts which could be avoided by adding protease inhibitors in the purification medium. The size of this fragment is reminiscent of the 70–60-kDa fragments we observed on silver stained gels after incubation with several proteases. Coincidentally, the secondary structure prediction of this region contains two long loops at positions 240–266 and 274–297. Perhaps the boundary between the glycosylated region and the cysteine-rich domain should be placed between positions 230 and 290 rather than at position 325 which is predicted to be in the middle of an extended structural element. Another long loop sequence is predicted at position 330–370. This loop is apparently not accessible for proteolytic attack. The same conclusion can be drawn for Arg-125 which is selectively cleaved by trypsin only after reduction and alkylation of the protein. This observation illustrates that modified cysteines some 200 amino acids further in the polypeptide chain alter the local environment of Arg-125 without much effect on the catalytic activity and the ADA binding site. When the protein is further denatured, more cleavage sites for trypsin become accessible. Evidence for this was published by Morrison et al. [30] who identified 5 tryptic peptides of 'ADA-binding protein' after SDS-polyacrylamide electrophoresis. Two of the cleavage sites, Lys-258 and Arg-492, appear to be at the beginning and the end of the cysteine-rich domain as in the 33-kDa V8 protease fragment [26]. Lys-373 is located in an extended structural element following the long loop in this domain. Arg-597 and Arg-658 are located in predicted loop regions in the protease domain. In conclusion, from the proteolysis data and the secondary structure prediction, it would appear that the catalytic domain is smaller than the two

others. Further structural investigation will show whether this assignment is correct.

Acknowledgements

This work received support from the Belgian Program on Interuniversity Poles of Attraction (A.-M.L.), the Research Council of the K.U. Leuven (OT/93/20) and the Flemish Fund for Scientific Research 2.0163.94 (J.F.D.P.), the Belgian National Fund for Scientific Research (NFWO) (K.H. and G.V.) and the Flemish Program for the Concerted Research Action 120.52293 (J.V.B.). P.C. received a pre-doctoral fellowship from MEC of Spain (DGICyT Grant 920007). I.D.M. is a Senior Research Assistant of the Belgian NFWO. We thank Prof. Y. Engelborghs for providing access to instrumentation and computer time. Seminal plasma was generously provided by Dr. E. Bosmans, clinical biologist associated with the Genk Institute for Fertility technology (GIFT), Belgium.

References

[1] C.A. Abbott, E. Baker, G.R. Sutherland, G.W. McCaughan, Immunogenetics 40 (1994) 331–338.
[2] J.M. Andreu, J.F. Díaz, R. Gil, J.M. de Pereda, M. García, V. Peyrot, C. Briand, E. Towns-Andrews, J. Bordas, J. Biol. Chem. 269 (1994) 31785–31792.
[3] D.M. Byler, H. Susi, Biopolymers 25 (1986) 469–487.
[4] T. Chikuma, T. Hama, T. Nagatsu, M. Kumegawa, T. Kato, Biol. Chem. Hoppe Seyler 371 (1990) 325–330.
[5] Y.N. Chirgadze, O.V. Fedorov, N.P. Trushina, Biopolymers 14 (1975) 679–694.
[6] F. David, A.M. Bernard, M. Pierres, D. Marguet, J. Biol. Chem. 268 (1993) 17247–17252.
[7] I. De Meester, G. Vanhoof, D. Hendriks, H.-U. Demuth, A. Yaron, S. Scharpé, Clin. Chim. Acta 210 (1992) 23–34.
[8] I. De Meester, G. Vanham, L. Kestens, G. Vanhoof, E. Bosmans, P. Gigase, S. Scharpé, Eur. J. Immunol. 24 (1994) 566–570.
[9] I. De Meester, G. Vanhoof, A.-M. Lambeir, S. Scharpé, J. Immunol. Meth. 189 (1996) 99–105.
[10] J.F. Díaz, E. Pantos, J. Bordas, J.M. Andreu, J. Mol. Biol. 238 (1994) 214–225.
[11] J.S. Duke-Cohan, C. Morimoto, J.A. Rocker, S.F. Schlossman, J. Biol. Chem. 270 (1995) 14107–14114.
[12] L.S. Duke-Cohan, C. Morimoto, J.A. Rocker, S.F. Schlossman, J. Immunol. 156 (1996) 1714–1721.
[13] R. Fabiani, Upsala J. Med. 99 (1994) 73–112.
[14] F. Goossens, I. de Meester, G. Vanhoof, D. Hendriks, G. Vriend, S. Scharpé, Eur. J. Biochem. 233 (1995) 432–441.
[15] K. Goossens, L. Smeller, J. Frank, K. Heremans, Eur. J. Biochem. 236 (1996) 254–262.
[16] M.D. Gorrell, J. Wickson, G.W. McCaughan, Cell. Immunol. 134 (1991) 205–215.
[17] A. Guinier, G. Fournet, in: Small Angle Scattering of X-rays, Wiley, New York, 1955.
[18] S. Iwaki-Egawa, Y. Watanabe, Y. Fujimoto, Biol. Chem. Hoppe-Seyler 374 (1993) 973–975.
[19] E. Jacotot, C. Callebaut, J. Blanco, B. Krust, K. Neubert, A. Barth, A.G. Hovanessian, Eur. J. Biochem. 239 (1996) 248–258.
[20] J. Kameaoka, T. Tanaka, Y. Nojima, S.F. Schlossman, C. Morimoto, Science 261 (1993) 466–469.
[21] T. Kähne, H. Kröning, U. Thiel, A.J. Ulmer, H.-D. Flad, S. Ansorge, Cell. Immunol. 170 (1996) 63–70.
[22] M. Karas, F. Hillenkamp, Anal. Biochem. 60 (1988) 2299–2301.
[23] E. Krepela, J. Kraml, J. Vicar, L. Kadlecova, E. Kasafirek, Physiol. Bohemslov. 32 (1983) 486–495.
[24] G. Küllertz, M. Nagy, G. Fischer, A. Barth, Biomed. Biochim. Acta 45 (1986) 291–303.
[25] T. Kyouden, M. Himeno, T. Ishikawa, Y. Ohsumi, K. Kato, J. Biochem. 111 (1992) 770–777.
[26] K. Löster, K. Zeilinger, D. Schuppman, W. Reutter, Biochem. Biophys. Res. Commun. 217 (1995) 341–348.
[27] G.W. McCaughan, J.E. Wickson, P.F. Creswick, M.D. Gorrel, Hepatology 11 (1990) 535–544.
[28] Y. Misumi, Y. Hayashi, F. Arakawa, Y. Ikehara, Biochim. Biophys. Acta 1131 (1992) 333–336.
[29] C. Morimoto, S.F. Schlossman, Immunologist 2 (1994) 4–7.
[30] M. Morrison, S. Vijayasaradhi, D. Engelstein, A.P. Albino, A.N. Houghton, J. Exp. Med. 177 (1993) 1135–1143.
[31] I. Ohkubo, K. Huang, Y. Ochiai, M. Takagaki, K. Kani, J. Biochem. 116 (1994) 1182–1186.
[32] G. Püschel, R. Mentlein, E. Heyman, Eur. J. Biochem. 126 (1982) 359–365.
[33] W. Reutter, O. Baum, K. Löster, H. Fan, J.P. Bork, K. Bernt, C. Hanski, R. Tauber, in: B. Fleisher (Ed.), Dipeptidyl Peptidase IV (CD26) in Metabolism and the Immune Response, R.G. Landes Co., 1995, pp. 55–78.
[34] B. Rost, Methods Enzymol. 266 (1996) 525–539.
[35] B. Rost, C. Sander, J. Mol. Biol. 232 (1993) 584–599.
[36] B. Rost, C. Sander, Proteins 19 (1994) 55–72.
[37] B. Rost, C. Sander, Proteins 20 (1994) 216–226.
[38] M. Saison, J. Verlinden, F. Van Leuven, J.-J. Cassiman, H. Van Den Berghe, Biochem. J. 216 (1983) 177–183.
[39] S. Shivaji, K-H. Scheit, P.M. Bhargava, in: Proteins of Seminal Plasma, Wiley, New York, 1990.
[40] L. Smeller, K. Goossens, K. Heremans, Appl. Spectrosc. 49 (1995) 1538–1542.
[41] H. Stein, R. Schwarting, G. Niedobitek, in: W. Knap, B. Dörken, W.R. Gilks, E.P. Rieber, R.E. Schmidt, H. Stein,

A.E.G.Kr. von dem Borne (Eds.), Leucocyte Typing IV, Oxford University Press, New York, 1989, p. 412.
[42] T. Tanaka, D. Camerini, B. Seed, Y. Torimoto, N.H. Dang, J. Kameoka, H.N. Dahlberg, S.F. Schlossmann, C. Morimoto, J. Immunol. 149 (1992) 481–486.
[43] G. Vanhoof, I. De Meester, M. van Sande, S. Scharpé, A. Yaron, Eur. J. Clin. Chem. Clin. Biochem. 30 (1992) 333–338.
[44] P.T.T. Wong, Can. J. Chem. 69 (1991) 1699–1704.
[45] K. Yamashita, Y. Tachibana, Y. Matsuda, N. Katsunuma, N. Kochibe, A. Kobata, Biochemistry 27 (1988) 5565–5573.