The dimeric transmembrane domain of prolyl dipeptidase DPP-IV contributes to its quaternary structure and enzymatic activities

Kuei-Min Chung,¹ Jai-Hong Cheng,¹ Ching-Shu Suen,² Chih-Hsiang Huang,¹ Cheng-Han Tsai,¹ Li-Hao Huang,¹ Yi-Rong Chen,³ Andrew H.-J. Wang,⁴ Weir-Torn Jiaang,¹ Ming-Jing Hwang,² and Xin Chen^{1*}

¹Division of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Miaoli County 350, Taiwan, Republic of China

²Institute of Biomedical Sciences, Academia Sinica, Taipei 115, Taiwan, Republic of China

³Division of Molecular and Genomic Medicine, National Health Research Institutes, Miaoli County 350, Taiwan, Republic of China

⁴Institute of Biological Chemistry, Academia Sinica, Taipei 115, Taiwan, Republic of China

Received 21 April 2010; Revised 13 June 2010; Accepted 16 June 2010

DOI: 10.1002/pro.443

Published online 23 June 2010 proteinscience.org

Abstract: Dipeptidyl peptidase IV (DPP-IV) is a drug target in the treatment of human type II diabetes. It is a type II membrane protein with a single transmembrane domain (TMD) anchoring the extracellular catalytic domain to the membrane. DPP-IV is active as a dimer, with two dimer interacting surfaces located extracellularly. In this study, we demonstrate that the TM of DPP-IV promotes DPP-IV dimerization and rescues monomeric DPP-IV mutants into partial dimers, which is specific and irreplaceable by TMs of other type II membrane proteins. By bioluminescence resonance energy transfer (BRET) and peptide electrophoresis, we found that the TM domain of DPP-IV is dimerized in mammalian cells and *in vitro*. The TM dimer interaction is very stable, based on our results with TM site-directed mutagenesis. None of the mutations, including the introduction of two prolines, resulted in their complete disruption to monomers. However, these TM proline mutations result in a significant reduction of DPP-IV enzymatic activity, comparable to what is found with mutations near the active site. A systematic analysis of TM structures deposited in the Protein Data Bank showed that prolines in the TM generally produce much bigger kinking angles than occur in nonproline-containing TMs. Thus, the proline-dependent reduction in enzyme activity may result from propagated conformational changes from the TM to the

Abbreviations: Ala-Pro-AFC, H-Ala-Pro-7-amino-4-trifluoromethylcoumarin; BRET, bioluminescence resonance energy transfer; DPP, dipeptidyl peptidase; DPP-IV, dipeptidyl peptidase IV; EOM, enzyme overlay membrane; FBS, fetal bovine serum; GFP, green fluorescent protein; Gly-Pro-AMC, H-Gly-Pro-7-amino-4-methylcoumarin; GPCR, G-Protein coupled receptor; PBS, phosphate buffered saline; PDB, protein data bank; PDBTM, protein data bank of transmembrane proteins; PFO, perfluorooctanoate; Rluc, renilla luciferase; SI, sucrase isomaltase; TM, transmembrane; TMD, transmembrane domain; WT, wild-type.

Additional Supporting Information may be found in the online version of this article.

Kuei-Min Chung and Jai-Hong Cheng contributed equally to this article.

Grant sponsor: NRPGM; National Science Council; Grant number: NSC96-3112-B-400-004 and NSC-95-3114-P-002-005-Y; Grant sponsor: National Health Research Institutes, Taiwan, ROC.

*Correspondence to: Dr. Xin Chen, Division of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Zhu Nan town, Miaoli County, Taiwan 350, Republic of China. E-mail: xchen@nhri.org.tw

extracellular active site. Our results demonstrate that TM dimerization and conformation contribute significantly to the structure and activity of DPP-IV. Optimal enzymatic activity of DPP-IV requires an optimal interaction of all three dimer interfaces, including its TM.

Keywords: prolyl dipeptidase; dipeptidyl peptidase IV; transmembrane domain; dimerization; type II membrane protein

Introduction

Dipeptidyl peptidase IV (DPP-IV, also known as CD26; EC 3.4.14.5) belongs to the prolyl dipeptidase family, which preferentially cleaves peptide bond after a penultimate proline residue.¹ Because such prolyl-cleaving activity is not normally present in other cellular proteases, the prolyl dipeptidases are important for various biological functions *in vivo*.¹ DPP-IV is a drug target for the treatment of human type II diabetes, and its inhibitors (Januvia and Galvus) are on the market to treat the disease.

DPP-IV is a type II membrane protein with a short cytoplasmic tail (amino acids 1–6), a transmembrane domain (TMD) (amino acids 7–28), and the catalytic domain, which is located outside the cell (amino acids 29–766) [Fig. 1(A)].² The short cytoplasmic tail of DPP-IV interacts with the protein CARMA-1, which is critical for T-cell signaling and proliferation.³ The TM domain targets DPP-IV to the plasma membrane and anchors the protein to the membrane.^{2,4–6} A number of crystal structures of DPP-IV lacking the 38 amino-terminal amino acids have been solved.^{7–10} The ectodomain of DPP-IV is homodimeric, and we have previously shown that dimerization is required for enzyme activity.^{11,12} The extracellular domain of each monomer of DPP-IV consists of an α/β hydrolase domain and a β -propeller domain, with the active site located between them [Fig. 1(A)].^{7–10} Two dimer interfaces have previously been described: the C-terminal loop in the hydrolase domain and the loop that extends from the propeller domain [Fig. 1(A)].^{7–9} Single-site mutations in the C-terminal loop disrupt the dimers leading to monomers with very low levels of residual enzyme activity.^{11,12} The inactivation of the enzyme is due to the distortion of the catalytic triad close to the C-terminal dimer interface upon monomer formation.^{11,12} A dynamic equilibrium between the DPP-IV dimers and monomers does not appear to exist.^{11,12} Although much work has been carried out on the DPP-IV extracellular domain, little is known of the mechanism by which the TM anchors the protein in the membrane or the potential influence it might have on the catalysis carried out by the enzyme's active site. Because of the potential of this enzyme as a drug target in diabetes, it is imperative that the physical properties of this protein, its mechanism of action, and the manner in which enzyme activity might be regulated are delineated.

In this study, we have assessed the behavior of the DPP-IV TM *in vitro* and in the cell membrane. We have introduced mutations into TM to assess the

effect of potential disruption of this structural element on the physical structure of the enzyme and the enzyme activity. Our results show unexpectedly that the DPP-IV TM dimerizes on its own and promotes dimer formation of the monomeric DPP-IV in mammalian cells. We further show that mutations in TM affect the catalytic activity of the extracellular active site, independent of the ability of TM to promote dimerization. Our results have revealed the critical roles of TM on DPP-IV dimer conformation, structure, and enzymatic activity.

Results

TM promotes the dimerization of monomeric DPP-IVs

In baculovirus-infected insect cells, without TM, soluble DPP-IV is dimeric and enzymatically active.^{11,12} A single site mutation at F713 (F713A) or W734 (W734A) completely disrupts dimerization of the DPP-IV extracellular domain with less than 5% of the original WT activity remaining.^{11,12} We wanted to know whether these two monomeric mutants remain monomers when expressed in the mammalian cells. Transfection of CHO cells was performed because CHO cells do not have the endogenous DPP-IV [Fig. 2(A), Lane 1]. The cell lysate was fractionated on a 6% nonreducing SDS-PAGE gel, shown previously to completely separate the dimeric and monomeric DPP-IV proteins.¹³ Similar to what we observed in insect cells, WT DPP-IV without TM was dimeric [Fig. 2(A), Lane 4], while both F713A and W734A DPP-IV without TM were monomeric [Fig. 2(A), Lanes 5 and 7]. Consistently, these two monomeric mutant proteins had residual enzymatic activities, roughly 5% of the WT activity as measured by hydrolysis of Gly-Pro-pNA (data not shown). Therefore, soluble TM-less F713A and W734A are monomeric when expressed in either CHO cells (this study) or the insect cells.¹²

Surprisingly, we found that when we transfected the plasmid encoding full-length F713A or W734A DPP-IV into CHO cells, the enzymatic activity of the cell lysate was about 40 and 68% of the WT's (Table I), respectively, significantly higher than the activities observed with the TM-less monomers (around 5%). In nonreducing SDS-PAGE gel, full length WT DPP-IV ran as dimers [Fig. 2(A), Lane 2], consistent with previous reports.¹³ In sharp contrast, around 60% and 91% of the full-length F713A and W734A mutant proteins were dimeric, respectively, based on analyses

Figure 1. DPP-IV structure and sequence of its transmembrane domain. **A:** The dimeric structure of DPP-IV. Domains of DPP-IV and their relative orientations to the plasma membrane are shown. **B:** Sequence alignment of the amino termini of DPP-IVs from different species (Vector NT software). The residues *a* and *d*, as predicted by COILS and Paircoil2 are also indicated. The accession numbers for human, porcine, cat, mouse, rat, and bovine DPP-IV are P27487, P22411, Q9N217, P28843, P14740, and P81425, respectively.

using image quantification software [Fig. 2(A), Lanes 3 and 6, and Table I]. The result indicates that the presence of the 38 amino-terminal amino acids (or TM as demonstrated below) significantly restores the ability of the extracellular domains of monomeric F713A and W734A to dimerize.

To determine whether the increased activity measured with the total cell lysate was attributable to the monomeric or dimeric forms of the DPP-IV, we developed an EOM assay to measure qualitatively the relative activity of dimers versus monomers by overlaying a Ala-Pro-AFC-coated nitrocellulose membrane on top of the gel, followed by detecting the cleavage of Ala-Pro-AFC with UV illumination. The activity of the dimers, but not that of the monomers, was readily detectable with the EOM assay [Fig. 2(B)], consistent with our previous finding that DPP-IV must be dimeric to have any substantial enzymatic activity.^{11,12} Therefore, full length F713A and W734A are partially dimeric and only dimers are enzymatically active.

Based on these results, TM of DPP-IV contributes to and promotes the dimerization of DPP-IV. Next, we asked whether other TMs could have similar effects on DPP-IV dimerization. We replaced the TM of DPP-IV with TMs from either aminopeptidase N (also called CD13) or sucrase isomaltase (SI). Similar to DPP-IV, these two enzymes are both dimeric type II membrane proteases located on the plasma membrane by a single TM domain at the amino terminus.¹⁴ CD13-DPP-IV and SI-DPP-IV fusion proteins were dimeric and enzymatically active [Fig. 2(C,D), Lanes 2 and 5]. Interestingly, CD13 TM did

not promote the dimerization of F713A, and the expression level of CD13-F713A was consistently much lower than the wild type [Fig. 2(C), Lane 3]. We failed to detect the expression of SI-F713A after many attempts [Fig. 2(C), Lane 6]. Unlike the complete rescuing effect by DPP-IV TM as demonstrated in Figure 2(A), there was still significant amount of

Figure 2. Transmembrane domain of DPP-IV promotes the dimerization of monomeric DPP-IVs. **A:** DPP-IV TM promotes the dimerization of monomeric F713A and W734A in mammalian cells, detected by 6% nonreducing SDS-PAGE coupled with western blot analysis. Lanes 1, 2, 3, and 6 are the total lysates (20 μ g/lane), and Lanes 4, 5, and 7 are the concentrated media (5 μ g/lane). The cells were transfected with the following plasmids, (1) vector only, (2) full length WT-DPP-IV, (3) full length F713A, (4) CD5-WT (aa 39 to 766), (5) CD5-F713A (aa 39-766), (6) full length W734, or (7) CD5-W734A (aa 39-766), respectively. **B:** Enzymatic activities of DPP-IV detected by EOM, measured with Gly-Pro-AMC coupled to the membrane. The sequence of the lanes is the same as in panel A. **C:** TMs of CD13 or SI are largely ineffective in promoting dimerization of the monomeric DPP-IVs. Lanes 1 to 7 are total cell lysates from the cells transfected with the following plasmids, (1) vector only, (2) CD13-DPP-IV, (3) CD13-F713A, (4) CD13-W734A, (5) SI-DPP-IV, (6) SI-F713A, or (7) SI-W734A, respectively. **D:** Enzymatic activities of CD13-DPP-IVs or SI-CD13-DPP-IVs detected by EOM. The sequence of the lanes is the same as in panel C.

Table I. Enzymatic Activities of Dimeric DPP-IVs Corrected against Dimer Ratios

Name	WT	W734A	F713A	L11P	A14P
Total lysate activity ^a	100.0%	68.0%	40.0%	68.0%	62.0%
Ratio of dimer ^b	98 ± 4%	91 ± 1%	60 ± 9%	97 ± 5%	99 ± 2%
Dimer activity corrected against the ratio of dimers ^c	102% (100%)	75% (74%)	67% (66%)	70% (69%)	63% (62%)

^a The activities were measured and averaged from at least three independent experiments with triplicates for each construct. The activity of the WT full length was taken as 100%.

^b The ratio of dimer was quantified by image quantification software as described in Materials and Methods, from at least three independent repeats. The data were summarized from Figs 2A and 4A.

^c The activity of the full length dimer was determined by correcting the total lysate activity against the ratio of the dimer. The number in parenthesis is corrections after taking the activity of WT as 100%.

monomers present for CD13-W734A. Moreover, only residual amount of the dimeric SI-W734A was detectable suggesting that the replacement of the TM resulted in less stable proteins [Fig. 2(C), Lanes 4 and 7]. The results strongly suggest that the DPP-IV TM for DPP-IV dimerization is rather specific and not readily replaceable. In summary, TM of DPP-IV seems to affect DPP-IV protein folding and biogenesis, and contributes to its dimerization.

The TM domain of DPP-IV dimerizes with itself *in vitro* and in mammalian cells

One plausible explanation for promotion of dimerization by TM is that the amino terminus of DPP-IV is dimerized itself. We aligned the amino terminal region (amino acids 1–39) of DPP-IVs from different species. Both the short cytoplasmic tail (amino acids 1–6) and the TM domain (amino acids 7–28) are highly conserved [Fig. 1(B)]. The short cytoplasmic tail is highly charged, making it unlikely to dimerize with itself. To determine whether TM dimerizes *in vitro*, polypeptides were synthesized corresponding to the TM sequence [Fig. 3(A)]. To increase water solubility, two Lys were added to both the N- and C-termini based on TM Finder program,¹⁵ resulting in a 26mer TM polypeptide [Fig. 3(A)]. As a control, a well-characterized monomeric TM peptide called Hsmr TM1, which has a molecular weight similar to that of DPP-IV TM¹⁶ was included in the analysis [Fig. 3(A)]. As shown in Figure 3(B), by peptide electrophoresis (NuPAGE), WT TM of DPP-IV migrated at around 4 to 5 kD, while the monomeric Hsmr TM migrated at around 2.5 kD, suggesting that TM of DPP-IV forms dimers *in vitro*.

Next, we investigated whether TM forms dimers in mammalian cells. To answer this question, we set up a BRET assay, which is particularly powerful in assaying protein–protein interactions among membrane proteins in living cells.^{17,18} The sequence encoding the 39 amino-terminal amino acids of DPP-IV was fused N-terminally to the sequence encoding GFP or Rluc, and the fusion plasmids were transfected into CHO cells. When the interaction between two TMs brings Rluc and GFP to within a distance of 10–100 Å, resonance energy transfer occurs. The

emission of GFP can be detected with the addition of the Rluc substrate.¹⁷ Thus, the change of the BRET signal could be due to a change in the distance between GFP and Rluc, and/or structural, or conformational changes in the TM region. As determined by confocal microscopy, TM–GFP was properly targeted to the surface of the plasma membrane (data not shown).

Shown in Figure 3(C), neither TM–Rluc with vector control only nor TM–Rluc with soluble GFP resulted in significantly detectable BRET signals (Lanes 1 and 2). Only in the presence of both TM–GFP and TM–Rluc was the BRET signal detected, indicating that the interaction is specific [Fig. 3(C), Lanes 3–6]. The BRET signal was also dose dependent, as shown in the saturation study [Fig. 3(C)]. By increasing the ratio of the donor (TM–Rluc) to the receptor (TM–GFP), we observed an increase in the BRET signal until a plateau was reached [Fig. 3(C) Lanes 5–6]. This saturation study not only confirmed that the BRET signal and the interaction we observed were specific, but also provided the optimal donor–receptor ratio for the experiments presented in Figure 3(D). In summary, these data showed that the TM of DPP-IV forms dimers with itself, both *in vitro* and in mammalian cells. After the previously identified C-terminal and the propeller loop interaction sites, TM is the third dimerization interaction site of DPP-IV.

Robust interaction between DPP-IV TMs

To understand how TM interacts with itself, we first searched for possible dimerization motifs known to contribute to TM dimerization in other proteins. G₁₀XXG₁₃A₁₄ and T₂₂XP₂₄ motifs were found in the TM of human DPP-IV, although the DPP-IVs from other species have hydrophobic residues other than A14 at that position [Fig. 1(B)]. A heptad pattern was also predicted with both COILS¹⁹ and Paircoil2,²⁰ suggesting that the DPP-IV TM α -helices might interact in a manner similar to that of a coiled coil. In this pattern, residues L11, V18, and V25 were predicted to be at the *a* position and residues A14, I21, and L28 at the *d* position [Fig. 1(B)]. To study the interaction between the dimers, TM

Figure 3. Transmembrane domain of DPP-IV dimerizes *in vitro* and in cells. **A:** Sequence and molecular mass of synthesized TM and mutants. **B:** TM of DPP-IV is dimerized by peptide electrophoresis. **C:** TM of DPP-IV dimerizes in live cells by BRET assay. Lane 1: 1 μ g of TM-Rluc and 4 μ g of vector; Lane 2: 1 μ g of TM-Rluc and 4 μ g of pTriEx-GFP (soluble GFP); Lanes 3–6: 1 μ g of TM-Rluc with 1, 2, 4, or 6 μ g of TM-GFP, respectively. **D:** BRET analysis with mutant TMs. Lane 1: cells transfected with 1 μ g of TM-Rluc and 4 μ g of empty vector; Lane 2: cells transfected with 1 μ g of TM-Rluc and 4 μ g of plasmid encoded soluble GFP; Lanes 3 to 14 are cells transfected with 1 μ g of TM-Rluc and 4 μ g of TM-GFP with the following mutations on TM, (3) TM-WT, (4) TM-G10I, (5) TM-L11A, (6) TM-L11P, (7) TM-L12A, (8) TM-G13I, (9) TM-A14I, (10) TM-A14P, (11) TM-L17A, (12) TM-T19A, (13) TM-T22A, or (14) TM-P24A, respectively. The lanes marked with *** are statistically significant with *P* values smaller than 0.001 (Student's *t*-test).

dimerization was assessed with BRET by site-directed mutagenesis on various motifs identified.

First, we investigated the importance of the GXXGA motif and the heptad repeat in TM dimerization. These elements have been previously shown to be important for TM function in other membrane proteins.^{21–26} Single-site mutations were made substituting the bulky hydrophobic amino acid Ile for G10, G13, or A14 of the GXXGA element. Addition-

ally, we also replace hydrophobic residues L11, L12 or L17 with the smaller and less hydrophobic alanine to potentially weaken interactions involving the heptad repeat interactions. This strategy has previously been shown to be effective in determining the contribution of particular residues to the interface interaction.²⁷ The BRET activities of these mutants were all similar to that of the WT TM [Fig. 3(D), Lanes 3–5, 7–9, and 11], indicating that none of the mutations could disrupt TM dimers to monomers. The results suggest that the interaction between these TMs is quite robust in the lipid environment.

In light of the lack of the effect from these mutations, we next chose to mutate one of the residues in each motif to proline which is recognized as a potent helix disrupter.^{28–30} Interestingly, L11P in the heptad repeat and GXXGA motif, and A14P in the GXXGA motif both dramatically reduced the BRET signal to around 31 and 41% of the WT signal [Fig. 3(D), Lanes 6 and 10]. This change of the BRET signal could result from the disruption of dimerization or/and the change of the TM dimer conformation. Next, by peptide electrophoresis, the mutant L11P or A14P peptides run at a position close to and slightly faster than the wild type TM peptides, and they migrated distinctively slower than the monomeric peptide control [Fig. 3(B)]. A14P has a lower intensity on the electrophoretogram because of the trapping of much of the material near the top of the gel (data not shown). Due to the intrinsic limitation associated with either peptide electrophoresis or BRET, the extend of dimerization for L11P or A14P is not clear at this stage. One possible explanation for the data presented is that both L11P and A14P did not disrupt TM dimers to monomers, but the mutations resulted in significant conformational change of TM dimers as detected in both BRET and peptide electrophoresis.

Next, we investigated the importance of the polar residue, T19 and the T₂₂XP₂₄ motif by individually mutating the threonines and the proline to Ala (T19A, T22A, and P24A). The polar residues have been shown to provide a strong driving force in some membrane proteins for the self-association through hydrogen bonding.^{31–34} Neither T19A nor T22A affected the dimerization of the TM, based on results from the BRET assay [Fig. 3(D), Lanes 12 and 13]. However, P24A significantly increased the BRET signal to 129% of the WT signal [Fig. 3(D), Lane 14]. A higher BRET signal represents a greater energy transfer, which could be due to a shorter distance between the donor (Rluc) and the acceptor (GFP), as a result of a P24A induced conformational change in the TM dimer. These results indicate that neither the polar residues T19 or T22, nor the TXP motif plays an essential role in TM dimerization. Our mutational results demonstrating the stability of TM dimers are consistent with the observation that

Figure 4. TM mutations reduce the enzymatic activities of full length DPP-IV. In both panels **A** and **B**, Lane 1 is purified recombinant WT DPP-IV protein after boiling and β -mercaptoethanol treatment as the monomer control. Lanes 2–15 are the lysates of cells transfected with the following plasmids: (2) WT, (3) vector, (4) F713A, (5) G10I, (6) L11P, (7) A14I, (8) A14G, (9) G13I, (10) L11A, (11) L17A, (12) V18A, (13) T19A, (14) T22A, or (15) P24A, respectively. **A:** Full length DPP-IVs with TM mutations are dimeric. The ratio of dimer and monomer was quantified by LabWorks 4.6 Image Acquisition and Analysis Software (UVP Bioimaging Systems) and indicated below the lanes. **B:** Enzymatic activities of full length DPP-IV with TM mutations detected by EOM. **C:** Enzymatic activities of full length DPP-IV with TM mutations. Lanes 1–14 are the activities of total cell lysates from the cells transfected with the following plasmids: (1) vector, (2) WT, (3) G10I, (4) L11G, (5) L11A, (6) L11P, (7) G13I, (8) A14G, (9) A14I, (10) A14P, (11) L17A, (12) V18A, (13) T19A, (14) T22A, or (15) P24A, respectively. The lanes marked with *** are statistically significant, with P values smaller than 0.001 (Student's t test).

TM is capable of rescuing monomeric DPP-IV into dimers as shown previously [Fig. 2(A)].

TM interaction contributes to optimal enzymatic activity of DPP-IV

Next, we investigated whether the TMD can influence catalysis at the remote extracellular active site and at a significant distance from the TM element. Full length DPP-IV containing individual site-directed mutations was expressed in CHO cells, and the quaternary structure of the resulting proteins was determined by nonreducing gel electrophoresis. None of the TM single-site mutations disrupted full-length dimeric DPP-IV into the monomeric form [Fig. 4(A), Lanes 5–15]. All showed over 95% dimers,

similar to the WT protein [(Fig. 4(A) and Supporting information Table 2S)]. Consistently, enzyme activities were only associated with the dimeric forms of the enzyme in the EOM assay [Fig. 4(B)].

We then measured the total activities of the lysates, determined by hydrolysis of Gly-Pro-pNA, after correction for transfection efficiency. Interestingly, only L11P, A14P, T19A, T22A and P24A, resulted in reproducibly lower activities compared with the WT enzyme [Fig. 4(C)]. These reductions in enzymatic activity were not due to the stability of the proteases (data not shown). Interestingly, lysates with L11P or A14P had the most effects, with 68 and 62% of the WT activity left, respectively [Fig. 4(C), Lanes 6 and 10], while T19A, T22A and P24A had about 81 to 86% activities left [Fig. 4(C), Lanes 13–15]. In contrast, lysates with full length F713A and W734A mutant proteins had about 40 and 68% of the wild type activity, respectively (Table I). The difference of the enzymatic activity was not due to the expression level of the proteins (data not shown), and the correction has taken into the account of the transfection efficiency. After correction against the dimer ratios (about 60% and 91% were dimers for F713A and W734A, respectively, when compared to 98% for the TM mutants), the activities of dimeric F713A and W734A were roughly 66 and 74% of the WT's activity, respectively (Table I). Interestingly, dimeric L11P and A14P had 69 and 62% of the WT activity (Table I), indicating that the proline mutations in TM have significant effects on activity, comparable to the C-terminal loop mutations adjacent to the triad in the active sites. Combined with the data from the peptide electrophoresis [Fig. 3(B)] and BRET [Fig. 3(D)], the results suggested that only proline mutations on the TMD alter the conformation of the distant active site and decrease the enzymatic activities of the proteins, independent of dimerization state of the protein. Please note that these proline TM mutations localized to the plasma membrane as determined by alkaline extraction coupled with membrane fractionation (Fig. 1S). To investigate whether this is proline-specific, we have generated two additional mutations, L11G and A14G, and the activities of the full length proteins were measured. None of the mutations affected the activities significantly [Fig. 4(C), Lanes 4 and 8], suggesting that proline has affected the conformation of the dimer differently from other mutations. Thus, the effect of the TM mutations on the enzymatic activity of the full length DPP-IV is independent of its effect on TM dimerization. Other proline mutations such as L17P and I20P were also generated. However, the full length proteins with these mutations could not be produced in mammalian cells, despite of the presence of similar amount of mRNA determined by quantitative RT-PCR (data not shown). Thus, these mutants are not investigated further in this study.

Figure 5. TM mutations further reduce the enzymatic activity of full length DPP-IV with mutation at other dimer interface. In both panels **A** and **B**, Lane 1 is the purified and boiled and reduced recombinant WT-DPP-IV protein as the monomer control, Lanes 2 to 8 are lysates of cells transfected with the following plasmids, (2) vector, (3) WT, (4) F713A, (5) F713A and G10I, (6) F713A and L11P, (7) F713A and A14I, or (8) F713A and A14P, respectively. **A**: Mutations at both TM and other dimer interface result in mixture of dimers and monomers. The ratio of dimer and monomer was quantified by LabWorks 4.6 Image Acquisition and Analysis Software (UVP Bioimaging Systems) and indicated below the lanes. **B**: Enzymatic activities of DPP-IV detected by EOM. **C**: Enzymatic activities of DPP-IV in total cell lysate. Lanes 1 to 7 are the activities of the lysates from the cells transfected with the following plasmids, (1) vector, (2) WT, (3) F713A, (4) F713A and G10I, (5) F713A and L11P, (6) F713A and A14I, or (7) F713A and A14P, respectively. The lanes marked with *** are statistically significant with *P*-values smaller than 0.001 (Student's *t*-test).

Finally, we investigated to what extent simultaneously introducing mutations in two different dimer interfaces, the TM and the C-terminal loop, affected the dimerization and enzymatic activity of the proteins *in vivo*. Significant amounts of the dimeric form were present in these double-mutant proteins, with roughly 60% as dimers and 40% as monomers [Fig. 5(A), Lanes 4–8, and Supporting information Table 2S]. The enzymatic activities remained associated with the dimeric form shown by the EOM assay [Fig. 5(B)]. Interestingly and consistently, lysates of these dimeric double-mutant pro-

teases had about 20% of the WT activity [Fig. 5(C)], less than the activity associated with single F713A mutant (around 40%). Again, the difference of the enzymatic activity was not due to the expression level of the proteins (data not shown), and the correction has taken into the account of the transfection efficiency. The results indicate that TM significantly affected the activity of the dimer by further reducing the activity of F713A enzyme independent of the effects on oligomerization. Our data support the idea that optimal DPP-IV dimer activity requires not only dimerization per se but proper conformation on all three dimer interfaces. Proline mutations in TM result in conformational changes that propagate across long distances to the extracellular active site, resulting in significant reduction of the enzymatic activities of DPP-IV.

Modeling of the DPP-IV with three dimer interfaces

Proline mutations on TM, including L11P, A14P, and P24A, have reduced the enzymatic activity significantly. To investigate the effect of proline on the structure of TM, calculation of kink angles in TM helices using ProKink was performed. Previous measurements of kink angles in TM helices have been made in a database containing the limited number of membrane structures available in or before 2002.^{27,35,36} Because the number of membrane protein structures in the Protein Data Bank (PDB) has increased significantly in recent years, we were able to perform a systematic analysis of the TM structures deposited in the PDB using a much larger dataset (868 TM helices from 216 nonredundant protein chains; see Materials and Methods). Of these 216 chains, 134 contained proline. Our results show that Pro-containing TMs are twice as likely as nonPro-containing TMs to have a kink angle within 20°–45°, whereas kink angles in nonPro-containing TMs are generally around 5°–20° [Fig. 6(A)]. Notably, several TMs containing Pro had kink angles as large as 45°–120° [Fig. 6(A)]. These results, consistent with previous observations,^{27,36} indicate that proline generally induces a significant kink in TM helices in the membrane environment. Moreover, the result is also consistent with the increased of the BRET signal on P24A mutation, which probably eliminates kinks with decreased separation between the donor and acceptor probes [Fig. 6(B) and Fig. 3(D), Lane 14]. Further structural analysis by NMR or crystallization would be needed to confirm the existence of such a change.

Based on our data, we constructed a model in which DPP-IV dimerizes with a helix–helix interaction in the TM region that may be similar to that of a coiled coil along with two other dimerization interfaces located extracellularly [Fig. 6(B)]. The optimal alignment of all three dimer interfaces is essential

Figure 6. Kinking angles induced by proline and computer modeling of DPP-IV with three dimer interfaces. **A:** The frequency of the kinking angles between Pro-containing TMs and nonPro-containing TMs in PDBTM. **B:** The predicted structure of the N-terminal region (residues 1-37), generated by InsightII (Accelrys, San Diego). TM α -helices was manually docked at the GXXGA motif. The β -propeller domain (colored in black) and α/β -hydrolase domain (colored in grey) were adopted from the crystal structure of DPP-IV (PDB code 2BGN). This figure was produced using Molscript⁴⁵ and Raster3D⁴⁶ programs. The possible Pro24 kink is shown.

for optimal activity of DPP-IV, with TM playing a critical role in promoting dimerization and maximizing enzymatic activity.

Discussion

Because of the technical challenges associated with studying polypeptide interactions in a lipid environment, understanding the structure and function of membrane proteins remains one of our biggest challenges. Previously, based on crystal structures, DPP-IV was shown to be a protein with two dimerization interfaces, the C-terminal region and the propeller loop. In this study, we have demonstrated that the TM of DPP-IV dimerizes, actually making this protein one with three different dimerization interfaces [Fig. 6(B)]. Our biochemical studies indicate that TM dimerization con-

tributes to the structure, folding, and enzymatic activity of DPP-IV, and that optimal enzymatic activity requires the proper interactions of all three of these interfaces. It may very well be that other membrane-bound oligomeric proteins use similar TM dimerization strategies to regulate not only protein aggregation but enzyme activity as well. Not only can the presence of TM significantly overcome the disruption of the C-terminal dimerization domain (F713A and W734A mutations), but it can affect, via propagated allosteric conformational changes, the catalytic ability of the far-distant active site. Furthermore, the specificity of the system is demonstrated with constructs carrying the CD13 or SI TMs [Fig. 2(C)]. Overall, with these exogenous TMs, they exhibited a substantially decreased ability to promote dimerization of the protein and enzyme activity, although the extent of the defect was dependent on both the TM and the mutation in the extracellular head portion. The TM dimer might provide the optimal orientation of the ectdomain part of the enzyme. These long-range effects may occur by promoting the proper alignment of all three dimerization domains, leading to the proper orientation of the active site residues. However, gaining such insight will require a great deal of high-resolution structural work.

The interaction mechanism for TM dimerization is robust. TM dimerizes with itself *in vitro* and in live cells (Fig. 3). None of the mutations we introduced, including drastic mutations like proline are capable of breaking the dimers apart. This is different from what was observed for other TM proteins, in which the mutations in general had drastic effect on TM dimerization.^{21–23,37,38} None of the motifs that we identified, including GXXGA, heptad repeat and TXP motifs is dominant in determining DPP-IV TM dimerization. The interaction between membrane and polypeptide could be flexible and dynamic, thereby mitigating the deleterious effect on TM dimerization, as observed with other TM interactions in a lipid environment.³⁹ Whether this kind of interaction mode is adopted by other oligomeric membrane proteins is an interesting possibility unknown. Recently, Walkenhorst *et al.* (2009) has shown that polar residue in TM results in anomalous migration in the SDS-PAGE without dimerization.⁴⁰ It is not known whether this is applicable to other polar residue or nonpolar residue-containing TMs. Thus, it is necessary to interpret the data with caution and support the data with independent methods.⁴⁰ Here in this study, DPP-IV TM dimerization is concluded with both SDS-PAGE and BRET assays, and its ability to promote dimerization of monomeric DPP-IVs in the mammalian cell culture.

TM dimerization might also be important for the biogenesis and function of DPP-IV. We have consistently observed lower expression of CD13TM-DPP-IV and SITM-DPP-IV compared to DPP-IV wild type (Fig. 2). The results suggest that in addition to

affecting the conformation, DPP-IV TM might also contribute to the folding and the stability of the enzyme. DPP-IV has been shown to be an important signaling molecule in T cell activation. Recently, Ohnuma *et al.* showed that the cytoplasmic tail (amino acids 1–6) of DPP-IV interacts with CARMA-1, and that only the cytoplasmic tails from dimeric but not monomeric DPP-IV maintain this interaction.³ The most likely explanation of Ohnuma *et al.*'s findings, as found in our study, is that dimeric DPP-IV, containing dimerized TM domains, brings two cytoplasmic tails together in a conformation suitable for its interaction with CARMA-1. Conversely, monomeric DPP-IV with just one TM and one cytoplasmic tail is not optimal for such an interaction.³

The identification of all three dimer interfaces of DPP-IV provides an important tool for the future study of the biosynthesis, folding, import of oligomeric proteins to the endoplasmic reticulum, and export to the plasma membrane. We have shown previously that dimers and monomers of mutant DPP-IV, lacking the TM, do not equilibrate *in vitro*.^{11,12} It is most likely that, instead of dimerizing through preformed monomers, DPP-IV is assembled and folded cotranslationally as a dimer *in vivo* through the interactions of these three dimer interaction surfaces. Ozkan *et al.* (2007) proposed an interesting protein-folding hypothesis, in which proteins are folded and assembled through the recruitment of neighboring amino acids (in the unfolded chain), which form additional contacts and transiently stable local structures.⁴¹ As a result, the interactions between them reduce the free energy, ultimately producing an ordered structure.⁴¹ Understanding the role of TM dimerization in this folding and assembly process may provide important clues to its regulation.

Materials and Methods

Materials

The fluorogenic dipeptide substrates Gly–Pro–AMC and Ala–Pro–AFC were purchased from Bachem, Switzerland. Fetal bovine serum (FBS) was purchased from HyClone, Lipofectamine was from Invitrogen. The human liver cDNA library was from Clontech. The western detection kit Western Lightning was from PerkinElmer. The Dual-GloTM luciferase assay system was from Promega. DeepBlue C was from PerkinElmer. NuPAGE was from Invitrogen. The anti-DPP-IV antibodies were generated in house by immunizing a mouse or rabbits with purified DPP-IV protein expressed in insect cells.

Plasmid construction

The expression plasmid pEF–DPP-IV was constructed as described below. The full-length gene

encoding DPP-IV was amplified by PCR from a human liver cDNA library with the following two primers: 5'-GGATCCATGAAGACACCGTGGAAG GTTC-3' and 5'-GTCGACCTAAGGTAAAGAGAAAC ATTGT-3'. It was then ligated into pCR-Blunt II-Topo (Invitrogen). The *Bam*HI/*Sal*I fragment containing DPP-IV was released from pCR-Blunt II-Topo and ligated into pBluescript-KS(+) (Stratagene) to produce pBluescript–DPPIV. The mutant full-length DPP-IV constructs were generated as described previously¹¹ using pBluescript–DPPIV as the template with the primers summarized in Supporting information Table 1S. The wild type (WT) or mutant DPP-IV fragments were then released by *Kpn*I digestion, blunt ended with T4 DNA polymerase, and digested with *Xba*I. The *Xba*I–*Kpn*I (blunted) fragments were then ligated into the expression vector pEF–SCM⁴² at the *Xba*I and *Not*I (blunted) sites.

The plasmids pIRES-CD5-Δ39-DPPIV, pIRES-CD5-Δ39-F713A, and pIRES-CD5-Δ39-W734A to express secreted DPP-IV containing amino acids 39 to 766 were constructed as described below. *Bam*HI (blunted) and *Sal*I fragments from the insect cell expression plasmids pBac8-CD5-DPP-IV,¹¹ pBac8-CD5-F713A,¹² and pBac8-CD5-W734A¹² were released and ligated, respectively, to the mammalian cell expression vector pIRES2-EGFP (Clontech) at the *Xho*I (blunted) and *Sal*I sites.

The TM of DPP-IV was replaced with that of CD13 or sucrase-isomaltase (SI), as described below. The TM region from CD13 (amino acids 1–50) were amplified by PCR from a cDNA clone (pCMV-SPORT6 CD13 from Open Biosystems) with the following primers: 5'-CTCATCGCTAGCATGGCCAA GGGCTTCTAT-3', 5'-GAGCTCGAGGGGGTGGT GGAGGCCAC-3'. The TM region from SI (amino acids 1–50) was amplified by PCR from a cDNA library of Caco-2 cell line with the following primers: 5'-TCGCTAGCGCCACCATGGCAAGAAAGAAATTTA GT-3', 5'-CACTCGAGGGAGTAGCTGGAGTTGAAG TAGA-3'. The amplified fragments were then digested with *Nhe*I and *Xho*I, and ligated into pIRES-N-flag plasmids to produce pIRES-CD13 or pIRES-SI plasmids. The *Xho*I/*Sma*I fragments containing Δ39-DPPIV (40–766 amino acids) were released from pIRES-flag-Δ39-DPPIV and ligated into the C-terminus of the TM of pIRES-CD13 or pIRES-SI to generate pCD13-TM-DPP-IV or pSI-TM-DPPIV expression plasmids.

The fusion plasmids for the BRET assays were constructed by ligating the sequence encoding the DPP-IV TM domain to the amino termini of the *Renilla* luciferase (Rluc) and green fluorescent protein (GFP) genes in the pTriEx-1.1 Neo vector (Novagen). The pTriEx plasmids with Rluc or GFP were constructed as described below. First, the *Xho*I/*Sal*I-flanked GFP- and Rluc-encoding sequences

were PCR amplified using pGFP²-C2 and pRluc-C1 (BioSignal Packard, Montreal, Canada) as templates, respectively. The PCR primers for GFP and Rluc were as follows: GFP forward: 5'-GCGCTCGAGTATGGTGTGAGCAAGGGCGAGG-3',

GFP reverse: 5'-CGCGTCGACCTACTTGTACAGCTCGTCCATG-3', Rluc forward: 5'-CGCCTCGAGTATGACTTCGAAAGTTTATG-3', and Rluc reverse: 5'-GCGGTGCGACTTATTGTTTCATTTTGTAG-3'. The amplified fragments were inserted into the pGEM-T Easy vector (Promega, Madison, WI) according to the manufacturer's instructions. The Rluc- and GFP-encoding sequences were then released from the pGEM-T Easy vector by *EcoRI* and *SalI* digestion and were inserted into the pTriEx-1.1 Neo and pTriEx-1.1 Hygro vectors (Merck Biosciences; La Jolla, CA), respectively, between the *EcoRI* and *XhoI* sites. Finally, the region containing the TM domain (amino acids 1–39) of DPP-IV was amplified by PCR with the primers 5'-CCATGGGGACACCGTGAAGGTTTC-3' and 5'-CTCGAGCCACTGTCAGCTGTAGCTG-3' before it was ligated into pCR®-Blunt II-Topo (Invitrogen) to produce pTOPO-TM-DPP-IV. To facilitate cloning, Lys2 was changed to Gly2, creating the restriction site *NcoI*. This change did not affect the sorting of the TM fusion proteins to the plasma membrane (data not shown). The mutations were introduced as described with the primers listed in Supporting information Table 1S using pTOPO-TM-DPP-IV as the template. The TM fragments were released by digestion with *NcoI* and *XhoI* before they were ligated to the amino terminus of GFP2-C2 or Rluc-C1 in the pTriEx expression vector, to produce pTriEx TM GFP²-C2 (TM-GFP) and pTriEx TM Rluc-C1 (TM-Rluc), respectively, which were used in the BRET assays.

Determination of DPP-IV activity and the enzyme-overlay membrane assay

To determine the enzymatic activity of the mutant proteins, the pEF-DPP-IV or pIRES-CD5-Δ39-DPP-IV plasmids encoding full-length DPP-IV or soluble DPP-IV were transfected into Chinese hamster ovary (CHO) cells using Lipofectamine (Invitrogen), as instructed by the manufacturer. The Rluc expression plasmid was cotransfected to quantify the transfection efficiency.⁴² CHO cells were seeded at a density of 8×10^4 cells per well in 24-well plates. Forty-eight hours after transfection, the secreted form of DPP-IV was collected by filtration and concentration from the media free of FBS and penicillin-streptomycin. For full length DPP-IV, CHO cells were washed twice with phosphate-buffered saline (PBS) and detached with PBS containing 5 mM EDTA. The cells were collected by centrifugation and lysed in NP-40 lysis buffer (10 mM HEPES [pH 7.5], 142.5 mM KCl, 5 mM MgCl₂, 1 mM EGTA, 0.2% NP-40). The supernatants were collected and

the protein concentrations were determined with the Bradford assay, the luciferase activities with the Dual-GloTM luciferase assay system (Promega), and the DPP-IV enzymatic activities with Gly-Pro-AMC. In the DPP-IV enzymatic activity assay, the cell lysate was incubated in PBS with 150 μM Gly-Pro-AMC at 37°C for 25 min and then the initial rate was measured by monitoring the absorbance at 460 nm for the first 20 min. The final DPP-IV activities were normalized to the protein concentration and the luciferase activity. In the EOM assay, a nitrocellulose membrane was coated with 2.5 mM Ala-Pro-AFC in a dry bath incubator at 37°C. The concentrated media containing the secreted form of DPP-IV (5 μg) or total cell lysates (20 μg) were fractionated by nonreducing sodium dodecyl sulfate-polyacrylamide gel (6%) electrophoresis (SDS-PAGE) at 4°C, as described previously.¹³ The membrane coated with the substrate Ala-Pro-AFC was placed against the gel and incubated at 37°C in a humidified chamber for 2 h. The membrane was then removed from the gel and air dried before it was photographed under UV light.

BRET assay

The BRET assay was performed essentially as described previously.⁴³ pTriEx TM GFP2-C2 (TM-GFP) (4 μg) and pTriEx TM Rluc-C1 (TM-Rluc) (1 μg) were transfected into the CHO cells using Lipofectamine (Invitrogen), according to the manufacturer's instructions. The CHO cells were seeded at a density of 8×10^5 cells per well in 6 cm culture dish. Forty-eight hours after transfection, the cells were washed twice with PBS and detached with PBS containing 5 mM EDTA. The cells were then collected and resuspended in PBS. The cells (50 μL; ~50,000 cells) were distributed to each well of a 96-well white microplate. DeepBlue C (PerkinElmer), the substrate of Rluc was added to a final concentration of 5 μM per well. Light intensity was measured and sequentially integrated in the 410–480 nm and 515–530 nm windows using a Victor² V multiplate reader (PerkinElmer Life Sciences). The BRET ratio was calculated and defined as described previously⁴³: [(emission at 515–530 nm) – (emission at 410–480 nm) × *Cf*]/(emission at 410–480 nm), where *Cf* corresponds to (emission at 515–530 nm)/(emission at 410–480 nm) for TM-Rluc coexpressed with the pTriEx vector in the same experiment. The BRET ratios are finally presented as the averages of at least four independent experiments, in which each data point was assessed in triplicate.

Peptide electrophoresis

Three microgram of peptides were mixed with SDS sample buffer (50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 10% β-mercaptoethanol, 0.1% bromophenol blue) and incubated at room temperature for

1 h. After incubation, the samples were analyzed by NuPAGE gradient gels containing 4–12% acrylamide in MES buffer (Invitrogen).

Calculation of kink angles in TM helices using ProKink

PDBTM is a database of TM proteins with experimentally determined three-dimensional structures (maintained by the Institute of Enzymology, Hungarian Academy of Sciences, Hungary).⁴⁴ On March 11, 2009, the database contained 868 TM segments from 216 nonredundant protein chains. The kink angles were measured using the ProKink method (55) available in the Simulaid program (Mezei, M. <http://atlas.physbio.mssm.edu/~mezei/simulaid/simulaid.html>). Calculations were made for these 868 TM helices, which contained a total of 16,628 amino acids located within the membrane plane, as defined by PDBTM, of which 425 were proline.

Acknowledgments

We are very grateful for Drs. Peter Rubenstein and Nei-Li Chan for their suggestions. We also thank Drs. Gu-Gang Chang, Hui-Ping Chang, Chi-Yuan Chou, and all members of the lab, including Bill Tseng, Yu-San Han, Jia-Jing Wu, Chia-Hui Chien, Hsiang-Yun Tang, Yue-Rue Chu and Gu-Chuan Chen, for their helps in this study.

References

- Rosenblum JS, Kozarich JW (2003) Prolyl peptidases: a serine protease subfamily with high potential for drug discovery. *Curr Opin Chem Biol* 7:496–504.
- Hong WJ, Doyle D (1988) Membrane orientation of rat gp110 as studied by in vitro translation. *J Biol Chem* 263:16892–16898.
- Ohnuma K, Uchiyama M, Yamochi T, Nishibashi K, Hosono O, Takahashi N, Kina S, Tanaka H, Lin X, Dang NH, Morimoto C (2007) Caveolin-1 triggers T-cell activation via CD26 in association with CARMA1. *J Biol Chem* 282:10117–10131.
- Low SH, Tang BL, Wong SH, Hong W (1994) Golgi retardation in Madin-Darby canine kidney and Chinese hamster ovary cells of a transmembrane chimera of two surface proteins. *J Biol Chem* 269:1985–1994.
- Munro S (1991) Sequences within and adjacent to the transmembrane segment of alpha-2,6-sialyltransferase specify Golgi retention. *EMBO J* 10:3577–3588.
- Hong WJ, Doyle D (1990) Molecular dissection of the NH2-terminal signal/anchor sequence of rat dipeptidyl peptidase IV. *J Cell Biol* 111:323–328.
- Engel M, Hoffmann T, Wagner L, Wermann M, Heiser U, Kiefersauer R, Huber R, Bode W, Demuth HU, Brandstetter H (2003) The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. *Proc Natl Acad Sci USA* 100:5063–5068.
- Rasmussen HB, Branner S, Wiberg FC, Wagtmann N (2003) Crystal structure of human dipeptidyl peptidase IV/CD26 in complex with a substrate analog. *Nat Struct Biol* 10:19–25.

- Thoma R, Löffler B, Stihle M, Huber W, Ruf A, Hennig M (2003) Structural basis of proline-specific exopeptidase activity as observed in human dipeptidyl peptidase-IV. *Structure* 11:947–959.
- Fulop V, Bocskei Z, Polgar L (1998) Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis. *Cell* 94:161–170.
- Chien CH, Huang LH, Chou CY, Chen YS, Han YS, Chang GG, Liang PH, Chen X (2004) One site mutation disrupts dimer formation in human DPP-IV proteins. *J Biol Chem* 279:52338–52345.
- Chien CH, Tsai CH, Lin CH, Chou CY, Chen X (2006) Identification of hydrophobic residues critical for DPP-IV dimerization. *Biochemistry* 45:7006–7012.
- Dobers J, Grams S, Reutter W, Fan H (2000) Roles of cysteines in rat dipeptidyl peptidase IV/CD26 in processing and proteolytic activity. *Eur J Biochem* 267:5093–5100.
- Danielsen EM (1994) Dimeric assembly of enterocyte brush border enzymes. *Biochemistry* 33:1599–1605.
- Deber CM, Wang C, Liu LP, Prior AS, Agrawal S, Muskat BL, Cuticchia AJ (2001) TM Finder: a prediction program for transmembrane protein segments using a combination of hydrophobicity and nonpolar phase helicity scales. *Protein Sci* 10:212–219.
- Rath A, Melnyk RA, Deber CM (2006) Evidence for assembly of small multidrug resistance proteins by a “two-faced” transmembrane helix. *J Biol Chem* 281:15546–15553.
- Xu Y, Piston DW, Johnson CH (1999) A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc Natl Acad Sci USA* 96:151–156.
- Pfleger KD, Eidne KA (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). *Nat Methods* 3:165–174.
- Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences. *Science* 252:1162–1164.
- McDonnell AV, Jiang T, Keating AE, Berger B (2006) Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* 22:356–358.
- Lemmon MA, Flanagan JM, Treutlein HR, Zhang J, Engelman DM (1992) Sequence specificity in the dimerization of transmembrane alpha-helices. *Biochemistry* 31:12719–12725.
- MacKenzie KR, Prestegard JH, Engelman DM (1997) A transmembrane helix dimer: structure and implications. *Science* 276:131–133.
- Petrache HI, Grossfield A, MacKenzie KR, Engelman DM, Woolf TB (2000) Modulation of glycophorin A transmembrane helix interactions by lipid bilayers: molecular dynamics calculations. *J Mol Biol* 302:727–746.
- Gurezka R, Laage R, Brosig B, Langosch D (1999) A heptad motif of leucine residues found in membrane proteins can drive self-assembly of artificial transmembrane segments. *J Biol Chem* 274:9265–9270.
- Gurezka R, Langosch D (2001) In vitro selection of membrane-spanning leucine zipper protein-protein interaction motifs using POSSYCCAT. *J Biol Chem* 276:45580–45587.
- Noordeen NA, Carafoli F, Hohenester E, Horton MA, Leitinger B (2006) A transmembrane leucine zipper is required for activation of the dimeric receptor tyrosine kinase DDR1. *J Biol Chem* 281:22744–22751.
- Senes A, Engel DE, DeGrado WF (2004) Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. *Curr Opin Struct Biol* 14:465–479.
- Monne M, Hermansson M, von Heijne G (1999) A turn propensity scale for transmembrane helices. *J Mol Biol* 288:141–145.

29. Chou PY, Fasman GD (1978) Empirical predictions of protein conformation. *Annu Rev Biochem* 47:251–276.
30. O'Neil KT, DeGrado WF (1990) A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* 250:646–651.
31. Zhou FX, Cocco MJ, Russ WP, Brunger AT, Engelman DM (2000) Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol* 7:154–160.
32. Choma C, Gratkowski H, Lear JD, DeGrado WF (2000) Asparagine-mediated self-association of a model transmembrane helix. *Nat Struct Biol* 7:161–166.
33. Dawson JP, Weinger JS, Engelman DM (2002) Motifs of serine and threonine can drive association of transmembrane helices. *J Mol Biol* 316:799–805.
34. Gratkowski H, Lear JD, DeGrado WF (2001) Polar side chains drive the association of model transmembrane peptides. *Proc Natl Acad Sci USA* 98:880–885.
35. Govaerts C, Blanpain C, Deupi X, Ballet S, Ballesteros JA, Wodak SJ, Vassart G, Pardo L, Parmentier M (2001) The TXP motif in the second transmembrane helix of CCR5. A structural determinant of chemokine-induced activation. *J Biol Chem* 276:13217–13225.
36. Cordes FS, Bright JN, Sansom MS (2002) Proline-induced distortions of transmembrane helices. *J Mol Biol* 323:951–960.
37. Adams PD, Engelman DM, Brunger AT (1996) Improved prediction for the structure of the dimeric transmembrane domain of glycophorin A obtained through global searching. *Proteins* 26:257–261.
38. Fleming KG, Ackerman AL, Engelman DM (1997) The effect of point mutations on the free energy of transmembrane alpha-helix dimerization. *J Mol Biol* 272:266–275.
39. Bowie JU (2005) Solving the membrane protein folding problem. *Nature* 438:581–589.
40. Walkenhorst WF, Merzlyakov M, Hristova K, Wimley WC (2009) Polar residues in transmembrane helices can decrease electrophoretic mobility in polyacrylamide gels without causing helix dimerization. *Biochim Biophys Acta* 1788:1321–1331.
41. Ozkan SB, Wu GA, Chodera JD, Dill KA (2007) Protein folding by zipping and assembly. *Proc Natl Acad Sci USA* 104:11987–11992.
42. Chen X, Wen S, Fukuda MN, Gavva NR, Hsu D, Akama TO, Yang-Feng T, Shen CK (2001) Human ITCH is a coregulator of the hematopoietic transcription factor NF-E2. *Genomics* 73:238–241.
43. Goin JC, Nathanson NM (2006) Quantitative analysis of muscarinic acetylcholine receptor homo- and heterodimerization in live cells: regulation of receptor down-regulation by heterodimerization. *J Biol Chem* 281:5416–5425.
44. Tusnady GE, Dosztanyi Z, Simon I (2005) PDB_TM: selection and membrane localization of transmembrane proteins in the protein data bank. *Nucleic Acids Res* 33:D275–D278.
45. Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Cryst* 24:946–950.
46. Merritt EA, Bacon DJ (1997) Raster3D photorealistic molecular graphics. *Methods Enzymol* 277:505–524.