Docking study of the precursor peptide of matoparan onto its putative processing enzyme, dipeptidyl peptidase IV: a revisit to molecular ticketing

Soonmin Jang · Tse-Yu Chung · Jungho Shin · Kai-Lun Lin · Jason T. C. Tzen · Feng-Yin Li

Abstract Stepwise-cleavage process of promastoparans to reach maturity was investigated theoretically by combining ab initio folding and unbounded docking. The comparison between the structures of the promastoparans both before and after docking were examined along with the hydrogen bonding interaction pattern between the dipetidyl peptidase IV (DPPIV) and promastoparans to reveal how the endpoint of this stepwise cleavage is recognized among these promastoparans with highly resemble amino acid sequences. The current approach of folding and docking study provides structural insight on the stepwise cleavage process.

Keywords Molecular docking · Replica exchange molecular dynamics (REMD) · Mastoparan

Introduction

Mastoparans (MPs), a family of amphiphilic cationic tetradecapeptides isolated from wasp venom, have a variety of biological activities including degranulation of mast cells [1], stimulation of insulin secretion from pancreatic islet

S. Jang · J. Shin
Department of Chemistry and Institute for Chemical Biology, Sejong University, Seoul 143-747, Korea

T.-Y. Chung · J. T. C. Tzen
Graduate Institute of Biotechnology, National Chung Hsing University, Taichung, Taiwan

K.-L. Lin · F.-Y. Li (✉)
Department of Chemistry, National Chung Hsing University, Taichung, Taiwan
e-mail: feng64@yahoo.com

cells [2], regulation of sarcoplasmic reticular Ca^{2+} release by binding to glycogen phosphorylase [3], activation of G_i-type and G_0-type GTP-binding regulatory proteins [4, 5] and stimulation of arachidonic acid release [6]. Furthermore MPs are considered as useful tools in studying signal transduction mechanism [7]. Even with above activities to make MPs a great interest in medicinal applications, their detailed biosynthesis process in animals still remains unknown. Recently, the two gene sequences of the precursor of MPs from two different wasp species were cloned [8, 9]. They also share the same distinct pattern with that of melittin from honeybee [10], i.e., an N-terminal signal sequence followed by a prosequence with the even positions exclusively as proline or alanine, and finally ended with the mature toxin sequence. In addition, a full-length cDNA fragment encoding a putative dipeptidyl peptidase IV (DPPIV) was obtained and sequence comparison reveals that it is homologous to a human DPPIV [8]. This suggests that the maturation processing of MPs is similar to that of melittin [11]. The N-terminal signal sequence is responsible for endoplasmic reticulum (ER) targeting via a signal recognition particle (SRP)-dependent pathway, and cleaved by a signal peptidase after translocation into ER lumen [12]. Unlike regular proteases cleaving proteins in one step, the prosequence is proposed to be removed by a step-wise cleavage of dipeptides by DPPIV to reach the bioactive form [11]. This type of processes, also found in widely different organisms [8, 11, 13–16], was named as molecular ticketing by Kreil in 1990 [17].

DPPIV, a serine protease found in various cell types and in many species [18], selectively removes dipeptides from the N-terminus of peptides with proline or alanine in the penultimate position [19, 20]. Mammalian DPPIV is a multifunctional type II cell surface glycoprotein and physiologically modulates the maturation processing of several

peptide hormones, chemokines and neuropeptides [21, 22]. For example, frog skin DPPIV has been shown to function as a processing enzyme in the liberation of biologically active peptides from their precursor polypeptides. Recently DPPIV received a lot of attention due to its inhibitors being a major new class of oral antidiabetic drug [23–25]. The DPPIV's versatility in regulatory functions is mainly due to the simplicity in its specific proteolytic activity. The enzymatic mechanism of DPPIV was studied through analyzing the crystal structures of DPPIV and its complexes with various substrates [26–29]. However, whether DPPIV is involved in the maturation processing of MPs has yet to be illustrated.

In this study, we examined the possibility of the MP maturation going through the molecular ticketing process by investigating the structures of the MP precursors and the related docking with DPPIV in silico. It is crucial to monitor the change in the precursor structures and at the same time to study the docking process to elucidate the cleavage process as well as how the DPPIV pinpoints the end point of the stepwise-cleavage process. This also provides an excellent opportunity to test the substrate selectivity and specificity of DPPIV since the sequences and conformations of those precursors are highly similar. The results should shed some light on the understanding of the molecular ticketing processes of MP precursors with DPPIV, and also provide an explanation how DPPIV can possess the flexibility in terms of molecular structures for the substrates which it acts on.

Model and method

The MP (INLKAIAALAKKLL) isolated from the venom of wasp *Vespa magnifica* was chosen since its related precursor sequence is available in recent literature [9] and also its highly homologous MP (INLKALAALAKKIL) with its native structure deposited in protein data bank (PDB ID: 1d7n) [30]. To facilitate the discussion, we denoted this MP as MP-C and its related precursor, i.e., promastoparan, as PMP-C, and this sequence is compared with the other two known precursor sequences, promelittin and promastoparan B, as shown in Fig. 1. A series of MP-C promastoparan sequences generated from stepwise cleavage process were studied with ab initio folding to locate their native structures and then followed by protein docking to investigate their cleavage processes.

Conformation of the mastoparan and related promastoparans

Because the promastoparans are short-lived and no native structures are available to perform homology alignment, their structures have to be generated before docking with
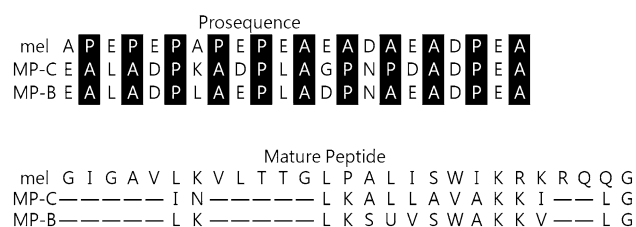
Fig. 1 The sequence comparison between melittin, MP-C and MP-B, where mel stands for mellitin, and MP-B for mastoparan B. Conserved residues in these three sequences are *shaded*. *Broken lines* in the sequences represent gaps introduced for best alignment

DPPIV. Here, we used ab initio folding method to generate native structure (or native-like structure) of each promastoparans. Protein folding is a very slow process (ranging from several mili-seconds to several seconds). Therefore, from the computational point of view, the direct observation of the folding process in silico at an all atom level is rarely feasible due to the tremendous computational cost. To overcome this problem, many researchers have used alternative approaches such as the coarse-grained model and/or efficient simulation techniques. The replica exchange molecular dynamics (REMD) simulation [31, 32] is arguably one of the most efficient sampling schemes for diverse complex molecular systems, especially proteins or peptides. In this study, we have used REMD in an all atom model to analyze structures of 7 different peptides (PMP-Cn, $n = 2, 4, 6, 8, 10$ and 12, where n means n residues attaching to the N-terminus of MP-C and $n = 0$ means the MP-C). Each peptide starts from fully extended conformation and heats up to 700 K for 200 ps, followed by local structural minimization. Then REMD simulation, as implemented in the Amber9 package [33] with all-atom Amber param03 force field and a generalized Born (GB) solvation model [34], was performed up to 60 ns for each replica. The number of replica is 12, with temperatures ranging from 297.5 to 661.0 K, as shown in Table 1. The simulation time step was 2 fs, [35] with bond constraints imposed between a hydrogen atom and heavier atoms using RATTLE [36]. The replica exchange was attempted at every 1.0 ps, and the resulting average acceptance was around 15–25% depending on peptides. The trajectories were saved at every 1 ps interval for further analysis. For the initial 500 ps, the REMD simulation was performed with no trajectory exchange to ensure good structural de-correlation before replica exchange takes place. The Berendsen thermostat, with a coupling constant of 1.0 ps, was used for temperature control.

To ensure the adequateness of the conformation ensemble for structures of these mastoparan-related peptides, free energy landscape approach was used. The free energy, $\Delta G(X,Y)$, in X and Y coordinates can be obtained with equilibrated trajectories via:

Table 1 The temperature distribution of REMD simulations used in current study

Replica index	Temperature (K)
1	297.5
2	313.9
3	334.0
4	357.4
5	384.0
6	413.9
7	447.0
8	483.3
9	522.9
10	565.7
11	611.7
12	661.0

Table 2 The binding affinity energy between the DDPIV and the first four residues of the promastoparans

Promastoparan	Binding energy (kcal/mol)
PMP-C12	−6.616
PMP-C10	−7.259
PMP-C8	−8.548
PMP-C6	−6.642
PMP-C4	−7.292
PMP-C2	−6.078

$$\Delta G(X, Y) = -RT \ln \left[P(X,Y) \big/ P_0 \right] \qquad (1)$$

where R is the gas constant; T is the temperature in Kelvin; $P(X,Y)$ is the population in (X,Y) space, and P_0 is the total population for normalization. From the last 5 ns of the simulation trajectory, we calculated the backbone root mean square deviation (RMSD), radius of gyration (RG), number of inter-strand hydrogen bonds (HB). We obtained free energy landscapes in the spaces of the following pairs of reaction coordinates: (RMSD, HB), (RMSD, RG). We assumed that a hydrogen bond is "formed" if the bond distance between a hydrogen atom and a heavier atom (O or N) is within 3.5 Å, and the angle with the hydrogen as the center atom is larger than 150°. For the RMSD calculations, we used native structure as the reference, but if the native structure is not available, we performed RMSD based hierarchical clustering analysis [37, 38] with conformers at 297.5 K, and selected the center frame of the highest populated cluster as the reference structure. The backbone RMSD of 3.5 Å was used as the clustering criterion. Unless specified otherwise, all the following analyses were done with the set of the convergent trajectory tested with the clustering analysis, i.e., the last 5 ns of the REMD simulation trajectories (5000 conformers) at 297.5 K.

Docking of mastoparan and its related promastoparans

Since there are no structures of wasp DPPIV available in literature, we used human DPPIV as the peptidase for these mastoparan-related peptides. The crystal structure of human DPPIV in complex with the first 10 residues of neuropeptide Y (tNPY) downloaded from the Protein Data Bank (PDB code 1R9 N) [26]. In order to validate this replacement, we performed the sequence alignment between tNPY sequence and the wasp DPPIV sequence take from Genbank as entry ABG57265.1 [8]. The results show that it is 36% identical between these two sequences and the residues in the surface of the binding pocket are almost identical except five residues as indicated in Fig. 2. In order to facilitate docking process, we removed tNPY from the complex, as well as the water molecules, sugars, and N-acetyl-D-glucosamines surrounding the DPPIV. The modified DPPIV structure after hydrogen saturation was minimized with CHARMm force field [39]. The binding pocket of mastoparan for the promastoparan was defined by the residues of the DPPIV which surround the first 4 residues of tNPY as indicated by Aertgeerts et al. [26]. The residues in the DDPIV, other than the binding pocket, were removed to facilitate the docking process. Docking of these promastoparans from the structures obtained by the above folding procedure was performed by employing the rigid body docking program, ZDOCK2.1 [40], which is an initial-stage docking algorithm utilizing a fast Fourier transform to explore all possible binding mode for the proteins. We used an energy function of a pairwise shape complementarity for scoring function. Only the first 4 residues of promastoparan were allowed to enter the binding pocket as indicated by Aertgeerts et al. [26]. The ligand orientation was obtained from 6° rotational sampling, which totaled 54000 predictions were then generated. An all-atom root-mean-square deviation (RMSD) was used to quantify the pairwise similarity between any two ligand orientations produced by ZDOCK to remove structural redundancy. Its cutoff distance was then set to 3.0 Å. The reason for this small cutoff value is mostly due to the small binding pocket.

Only those poses similar to the crystal structure of tNPY interacting with DPPIV[14] were selected to be further refined by CHARMm force field with the integrated DPPIV as well as the promastoparan in full length to be flexible to avoid the possible clash between DDPIV and the docking peptide. All the calculations were implemented in the Discovery Studio 2.0 environment (Accelrys, San Diego, CA).

Table 3 The hydrogen bonding interactions between the DDPIV and the first 4 residues of the promastoparans using tNPY as reference

tNPY[a]	PMP-C12	PMP-C10	PMP-C8	PMP-C6	PMP-C4	PMP-C2
R125HH12-S3O[b]	K122HZ1-P4O	R125HH12-P2O	R125HH12-A4O	R125HH21-A2O	R125HH22-D1O	R125HH12-N4O
R125H22-S3O	R125HH12-A2O	R125HH22-P2O	R125HH22-P2O	E205OE2-D3O	E205OE1-D1HN1	E205OE2-E1HN2
E205OE2-Y1HN2	R125H22-A2O	Y547HH-G1O	Y547HH-N1O	R125HH22-A2O	E206OE2-D1HN3	E206OE2-E1HN3
E206OE1-Y1HN	Y547HH-L1O	S630HG-G1O	Y547OH-A4HN1	S630HG-D1OD1		Y547HH-E1OE1
E206OE2-Y1HN3	Y666OH-L1HT2	Y662OH-G1HN3	P550O-N1HN3	Y662OH-D1HT2		
Y631HN-P2O	S630HG-L1O	H740O-G1HT2	C551O-N1HT1	N710HD22-A2O		
Y662OH-Y1NH3			Y666HH-D3OD1			
N710HD22-Y1O						

[a] Taken from Reference 24

[b] The first letter and the one before dash line are amino acid single letter symbols for receptor and substrate, respectively. The number after the residue symbol is the sequence number for the specific residue. The symbol after the number represents the particular atom involving hydrogen bonding, using the same one as in pdb format

Results and discussion

We have performed REMD simulations of the MP-C and several related prosequences. To locate the native structure from simulation, reasonable simulation convergence is prerequisites even with efficient sample scheme such as REMD used in current study. For this purpose, we have performed independent REMD simulation starting from two different structures, α-helix and fully extended structure. If the simulation is converged, the two REMD simulations starting from different conformation (one from extended structure and the other from helix structure) should give same thermodynamic and structural features in the end. Figures 3 and 4 show the progress of secondary structure as a function of time for MP-C and PMP-C6, respectively. Both helix and hydrogen bonded turn (H-turn) are shown obtained by DSSP analysis program [41]. For both MP-C and PMP-C6, the secondary structure approaches a steady state value roughly after 30 ns. Perhaps more rigorous convergence criteria would be the comparison of free energy landscape for the above two different simulations, one from extended structure and the other from helix structure. Figures 5 and 6 shows the free energy landscape on radius of gyration (RG) and hydrogen bond (HB) space (RG, HB) obtained from last 5 ns of simulation trajectory. One can observe that essentially they are either the same or very similar, indicating that the system has reached reasonable equilibrium state after this REMD simulation. Figure 7 shows the native structure of MP-C [30] and simulated structure. The backbone RMSD between them is 2.75 Å, indicating that the simulation setup can reasonably reproduce the native structure of MP-C. More stringent test regarding to the quality of current simulation is perhaps the comparison of the backbone fluctuation against the NMR results. In Fig. 8, we showed the residue-by-residue fluctuation of backbone alpha-carbon atom both from the simulation and NMR structures. A total of 32 different conformations around the lowest free energy minima are used to obtain this fluctuation. It can be seen that C-terminal side has good agreement while the N-terminal side has some difference (about factor of two). The simulated structure shows to be more flexible structurally compared to the NMR structure, especially around middle of the strand. However, the difference is not large and the overall feature remains about the same. This free energy minimum structure is taken after constructing the free energy landscape in (RMSD, RG) space shown in Fig. 9. Figure 9 further indicates that the folding of MP-C is a rather straightforward downhill helix formation to reach the global energy-minimum structure without any major locally stable structures under the current force field. The global energy-minimum structures of other promastoparans were shown in Fig. 10 by locating the representative structure in global minimum in the free energy landscape, as described in the previous section, indicated in Fig. 11. Essentially, these structures are consistent with the representative structures of the largest cluster from clustering analysis. The simulation indicates not every promastoparan which we studied has a pure helix structure. The degree of helixcity varies depending on promastoparans in their native structure and some promastoparans have a rather flexible structure. Figure 12 further shows the residue by residue detailed secondary structure from DSSP analysis [41] with last 5 ns of trajectories. The global energy-minimum structures of those promastoparans are assumed to represent their native structures and are used as the initial structures for the docking processes. The free energy landscapes show two distinct patterns, only one potential well for MP-C and PMP-C12, and three potential wells for the rest of promastoparans. These structures change as the mastoparan prosequences vary. There is one common feature shared among all the promastoparans, i.e., the first 2 residues from

```
Human   35  WISDHEYLYK--QENNILVFNAEYGNSSVFLENSTFDEFGHSINDYSISPDGQFILLEYN   92
            W SD +Y    +IL F+     +V +++S  D++    ++ Y +SP G+F+L+ Y+
Vespa   69  WTSDTTIVYTDIMTGDILQFDVIKQRLTVIVDSSVMDDY--IVSHYMLSPKGRFLLIGYD  126

Human   93  YVKQWRHSYTASYDIYDLNKRQLITEERIPNNTQ--WVTWSPVGHKLAYVWNNDIY-VKI  149
            + K +RHS   Y IYD+  +L  ++I N     W+P+   L Y+ +NDIY ++
Vespa  127  FQKGFRHSKFMRYVIYDI---ELGGYDKIANGMHIALAKWAPLTDDLIYILDNDIYYMRF  183

Human  150  EPNL--PSYRITWTGKEDIIYNGITDWVYEEEVFSAYSALWWSPNGTFLAYAQFNDTEVP  207
            N    R+T+ G   I+YNG+ DWVYEEEVF  SA+W+SP+G+ LAYA F+D  V
Vespa  184  SNNGFNDVQRVTYDGIVGIVYNGVPDWVYEEEVFHRSSAMWFSPDGSHLAYASFDDRNVQ  243

Human  208  LIEYSFYSDESL---QYPKTVRVPYPKAGAVNPTVKFFVVNTDSLSSVTNATSIQITAPA  264
            I Y Y+      QYP V++ YPKAG+ NP V  +V+ +      I + AP
Vespa  244  EILYLHYGEPGNLVDQYPTEVKIKYPKAGSSNPVVSLTLVDLHN----PTLNKIDLKAPI  299

Human  265  SMLIGDHYLCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWNCLVARQH---IEM  321
            ++ D+ L +V W   + +   W R+QN + +I Y++      L  +H  +E+
Vespa  300  EVVGTDNVLSNVQWKDIDHVIAMWSNRVQNKT--EIVWYNKYGEIVKTLDVVEHEGWVEI  357

Human  322  STTGWVGRF---RPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWEVI  378
            +    F    R +P T G F++  ++  +   Q+D   +T G  EV
Vespa  358  KNLFFYKDFVYIRKLQPSGTKAGR-FHHVTRYDDTLQ------LTPGVTEVQ  408

Human  379  GIEALTSDY--LYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCE-LNPE--RCQYYSVS  433
            I A+   +  +YY++    G P  RNLY + +      K TC+SC  L PE   C Y
Vespa  409  DIRAIDHFHGRIYYLATG-PGKPSQRNLYSVPVDGSEKPTCISCNVLTPEGNACTYADAI  467

Human  434  FSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNSALDKMLQNVQMPSKKLDFIILN  493
            FS  ++Y L C GP  +   + + +K+   EDN +L  L   ++P K ++ N
Vespa  468  FSPPFGQHYVLICQGPDPVIVGIFDNTHKK-VYSWEDNLSLRSKLARRELPLVKDLYVRAN  526

Human  494  ETKFWYQMILPPHFDKSKKYPLLLDVYAGPCSQKADTVFRLNWATYLASTENIIVASFDG  553
            +  ++ LP +FD+SK YP+L+++VYAGP + K  V    + Y+++  ++I A  DG
Vespa  527  GYESKVRLFLPHNFDESKSYPMLVNVYAGPNTAKIIDVASYGYHAYMTTNRSVIYAYIDG  586

Human  554  RGSGYQGDKIMHAINRRLGTFEVEDQIEAAROFSKM-GFVDNKRIAIWGVSYGGYVTSMV  612
            RGS  +G K++  I R+LGT EVEDQI  R+  +M    +D+KR  +WGVSYGG+ T+M+
Vespa  587  RGSSNKGSKMLFEIYRKLGTVEVEDQISVTRKLQEMFPRIDSKRTGVWGVSYGGPCTAMI  646

Human  613  LGSG-SGVFKCGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENFKQV  671
            L   + VFKCGIAVAPVS W YYDS+YTERYMGLPTPEDNL  Y + V R E++
Vespa  647  LAKDLTSVFKCGIAVAPVSSWIYYDSIYTERYMGLPTPEDNLSGYNGTDVSRRVEDIRGK  706

Human  672  EYLLIHGTADDNVHFQQSAQISKALVDVGVDFQAMWYTDEDHGIASSTAHQHIYTHMSHF  731
            +++LIHG+ DDNVH+QQS ++KAL  + F+ +  YTDE H +        H+Y M F
Vespa  707  KFMLIHGSGDDNVHYQQSLALAKALEKADIMFEQITYTDEAHALFGVLP--HLYHTMDRF  764

Human  732  IKQCFSL  738
            CF+L
Vespa  765  WSDCFNL  771
```
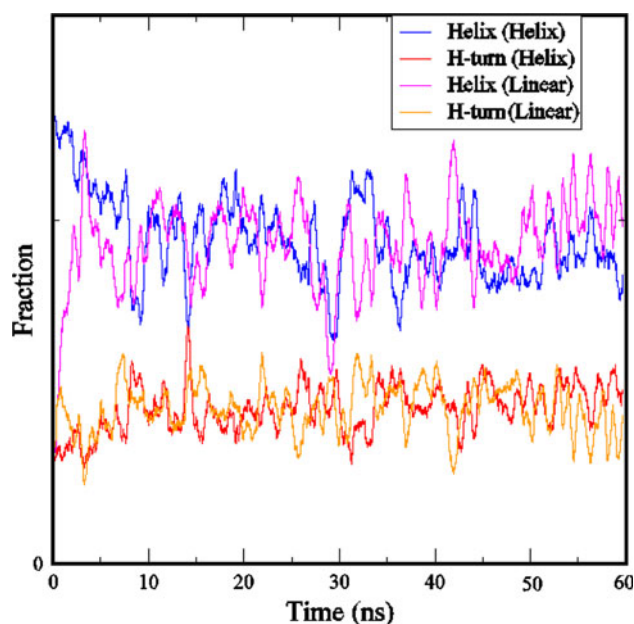
Fig. 2 Sequence alignment of *V. basalis* (Vespa) and human DPP-IVs. Numbers of the first and last amino acid in each line of the protein sequences are listed on the left and right, respectively. *Broken lines* in the sequences represent gaps introduced for best alignment. The residues on the surface of the binding pocket in these two sequences are *shaded*, where the conserved residues are *boldfaced* and the non-conserved residues are in purple. The catalytic triad residues (Ser630, Asp708, His740) responsible for DPP-IV activity are indicated by *red letters*. The two gate residues as discussed in Fig. 15 are highlighted in *blue color*

N-terminal are rather flexible and have no distinct secondary structure. The relatively well defined secondary structure starts from the third residue and the conformation for the third and fourth residues is either bent or turns in all the promastoparans, with small helix content in PMP-C10,

PMP-C6, PMP-C4 and PMP-C2. These features are crucial to the formation of the promastoparans-DPPIV complexes due to the small binding pocket of DPPIV.
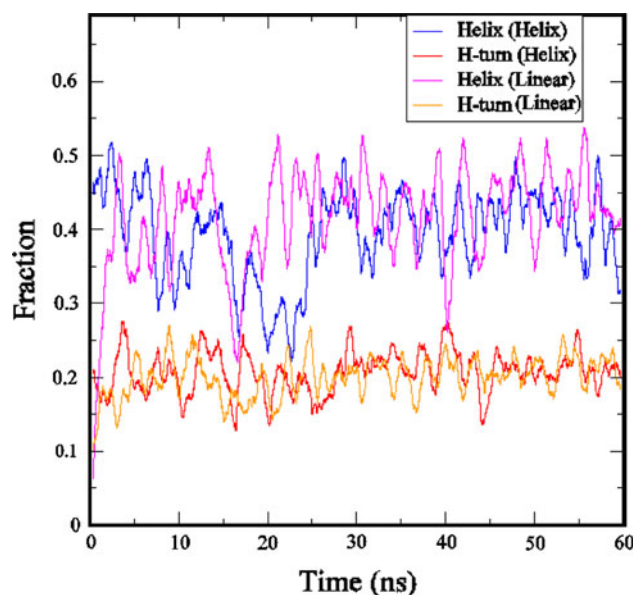

Fig. 3 Time progress of the secondary structure (helix and hydrogen bonded turn) starting from two different initial conformations (Linear and Helix) for MP-C as indicated in the parentheses of the legend. Essentially, the secondary structure reaches steady state after 30 ns. The secondary structure is analyzed by DSSP and the resulting secondary structure is blocked averaged at every 500 data points

Fig. 4 Time progress of the secondary structure (helix and hydrogen bonded turn) starting from two different initial conformations (Linear and Helix) for MP-C6 as indicated in the parentheses of the legend. Essentially, the secondary structure reaches steady state after 30 ns. The secondary structure is analyzed by DSSP and the resulting secondary structure is block averaged at every 500 data points
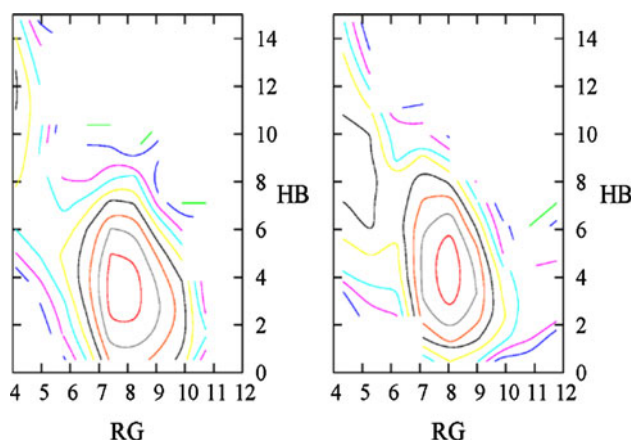
Fig. 5 The free energy landscape of MP-C in (RG, HB) space obtained from last 5 ns of simulation trajectory for two independent runs starting from helix (left) and linear structure (right). We defined a hydrogen bond is "formed" if the distance between heavy atom (N or O) and hydrogen atom is less than 3.5 Å and the angle with the hydrogen as the center atom is larger than 150°
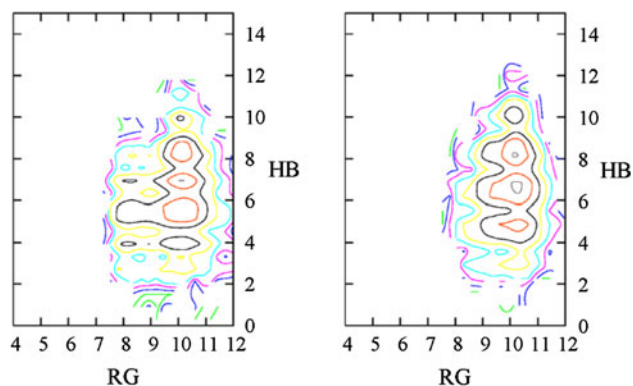

Fig. 6 The free energy landscape of PMP-C6 in (RG, HB) space obtained from last 5 ns of simulation trajectory for two independent runs starting from helix (left) and linear structure (right)

The binding pocket site of DPPIV and several of the complex structures of DPPIV and the promastoparans (PMP-C12) within this site were shown in Fig. 13 and their binding affinity energies calculated with DCOMPLEXE, developed by Liu et al. [42]. There is a narrow channel formed by TYR547 and ARG125, near the active site SER630 and HIS740, and only the first two residues of the promastoparans inside are allowed into the pocket section on the left-handed side. The bond to be cleaved is located around this narrow channel as indicated with an arrow in Fig. 13b. The pocket section has a trapezoidal cleft with the wider part near the narrow channel, and the promastoparans can slide through this cleft to enter the pocket section. In order to obtain a better view on how the promastoparans dock inside the pocket section, we overlap all the docking structures of the promastoparans as shown in Fig. 13d. Clearly there are several different orientations for
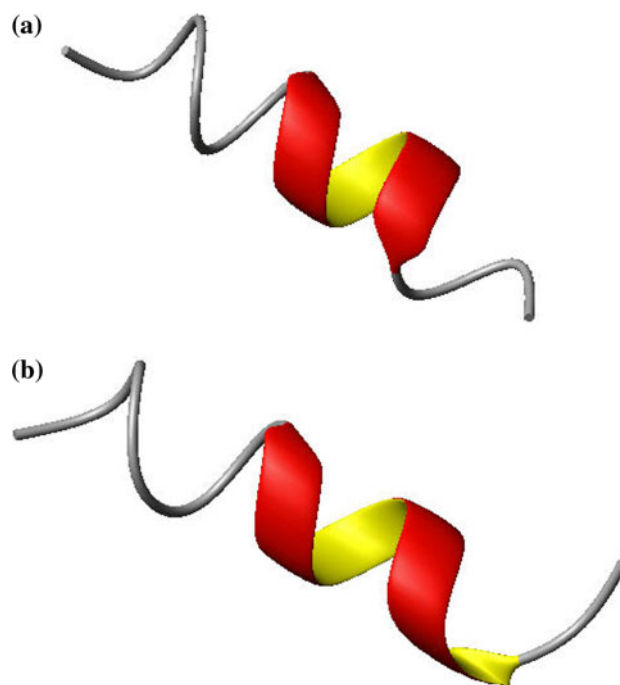

Fig. 7 The comparison between the experimental (**a**) and simulated (**b**) structure of MP-C
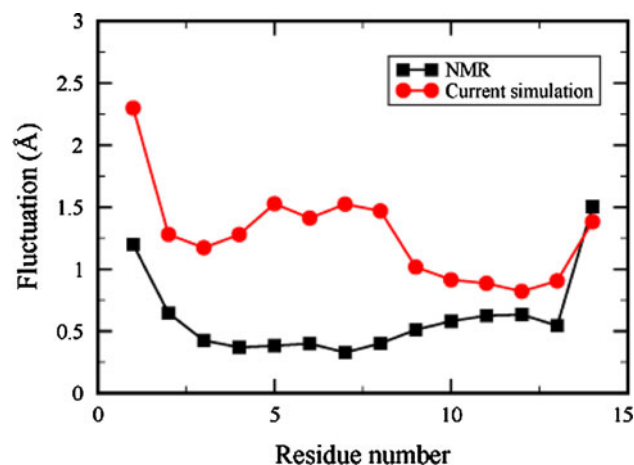
Fig. 8 The residue-by-residue backbone Cα atom fluctuation of MP-C from both current simulation and NMR results. 10 NMR structures and 32 structures obtained from free energy minima region are used for this purpose

the side chain of the first residue among the promastoparans, each of them corresponding to either a cavity or small channel in the inner surface of the pocket section to accommodate the specific side chain according to its propensity. In order to visualize the orientation of the first residue for each promastoparan, we plotted the inner surface of the pocket section in a 2D projection plot, i.e., the angular part of the spherical coordinate, along with the location of the last non-hydrogen atom of the side chain of
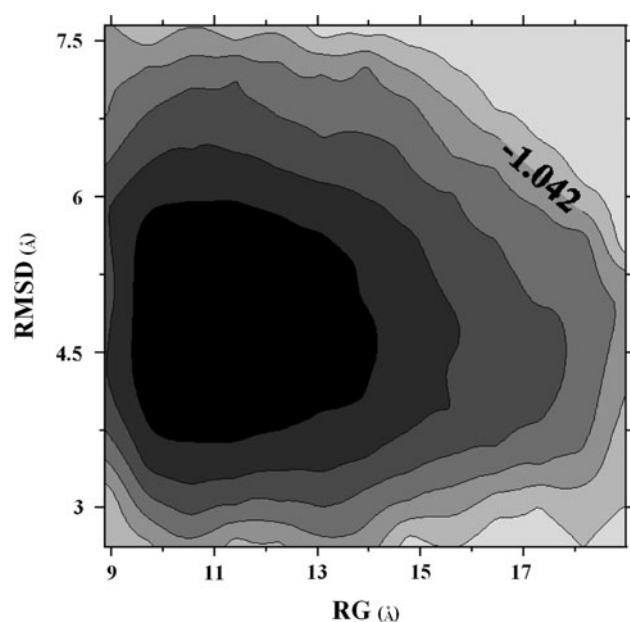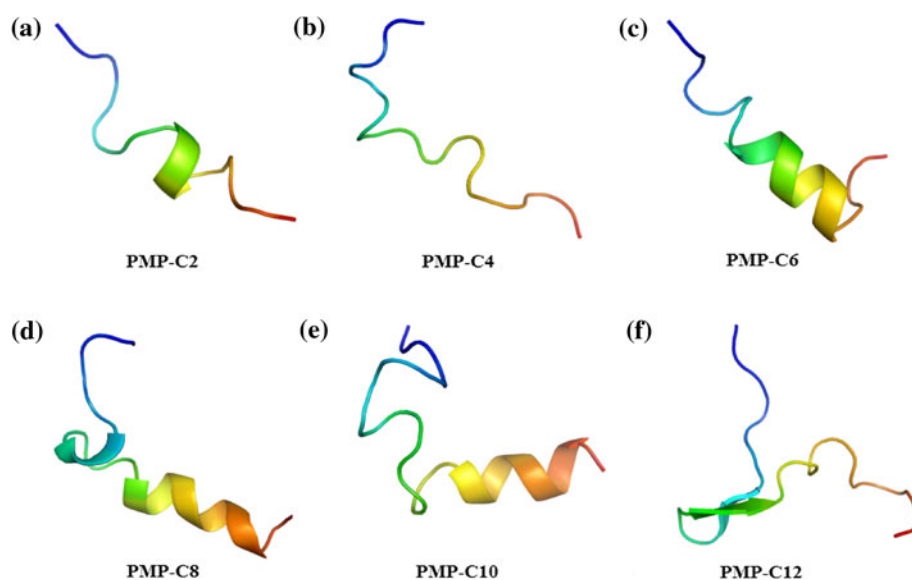
Fig. 9 The free energy landscape on (RG, RMSD) space at 297 K. The contour interval is set to kT

the first residue for each promastoparan. As shown in Fig. 14a, for PMP-C4, PMP-C6 and PMP-C8, the last non-hydrogen atom of the first residue side chain is found around TRP659 of DPPIV; those of PMP-C10 and PMP-C12 are near TYR662; those of PMP-C2 and MP-C are around GLY549 and PHE357, respectively, indicating that the first residues in most of promastoparans are clustered around the narrow region of the DPPIV. Since the first residue is close to the docking pocket, this suggests that the docking mode of promastoparans of different size is essentially similar. The somewhat off site orientation of MP-C and PMP-C2 might be due to their small size. Of

course, the size of side chain is different for different promastoparans. However, the location of the last non-hydrogen atom of the second residue side chain for most of the promastoparans is clustering either around the narrow channel formed by TYR547 and ARG125, or near the catalytic site, SER630 and HIS740, except that of PMP-C2 which points to PHE357. So, the picture suggested here should be understood as rough measure of their binding pattern. According to their corresponding Φ and Ψ angles of the first 4 residues before and after docking as indicated in Table 4, the structures obtained by REMD simulation show no noticeable secondary structure change upon docking with DPPIV. Also, we found no noticeable change of side chain orientation of DPPIV after docking, other than the pocket area (not shown). In order to investigate the stepwise cleavage process, the binding affinity energy between DPPIV and the first four residues of the pro-mastoparans inside the binding pocket was examined. It indicates that the binding affinity energy varies along the cleavage process ranging from -6.6 to -8.5 kcal/mol for PMP-C12 to PMP-C4 without a clear trend, as shown in Table 2. The binding affinity energy of PMP-C2 is the lowest one. This indicates the strength of the binding affinity between DPP IV and the promastoparans decreases significantly when the length of the extra amino acid attached to mastoparan is less than 4. Roughly, this agrees with Fig. 14 in a sense that the side chain orientation begins to show difference from PMP-C2. Therefore the cleavage mechanism of DPPIV should be different between the promastoparans longer than PMP-C2 and those equal to PMP-C2. We also examined the hydrogen bond pattern of the DDPIV-promastoparan complexes as shown in Table 3. There are 8 hydrogen bonds found in the DPPIV-tNPY complex, and 6 or 7 hydrogen bonds found between the

Fig. 10 The global energy-minimum structures of all the promastoparans. Each promastoparan is divided into 6 portions with different colors to show the locations of the secondary structural features related to the N- and C-terminus, where *blue* and *red* colors are the N- and C-terminal sections, respectively

(a) PMP-C2 (b) PMP-C4 (c) PMP-C6
(d) PMP-C8 (e) PMP-C10 (f) PMP-C12

Fig. 11 The free energy landscape on (RG, HB) space at 297 K for prosequence of MP-C. **a** PMP-C12 **b** PMP-C10, **c** PMP-C8, **d** PMP-C6, **e** PMP-C4 and **f** PMP-C2. The contour interval is set to kT, where k is Boltzmann constant and T is temperature at 300 K
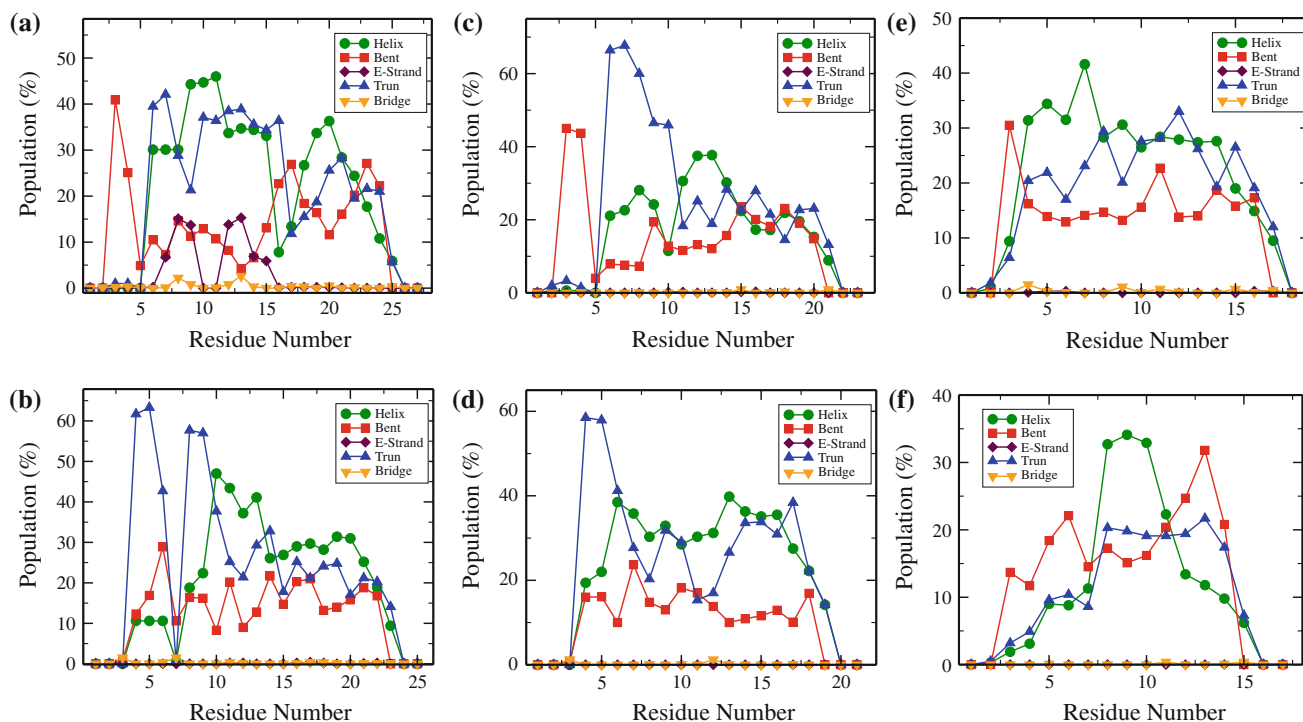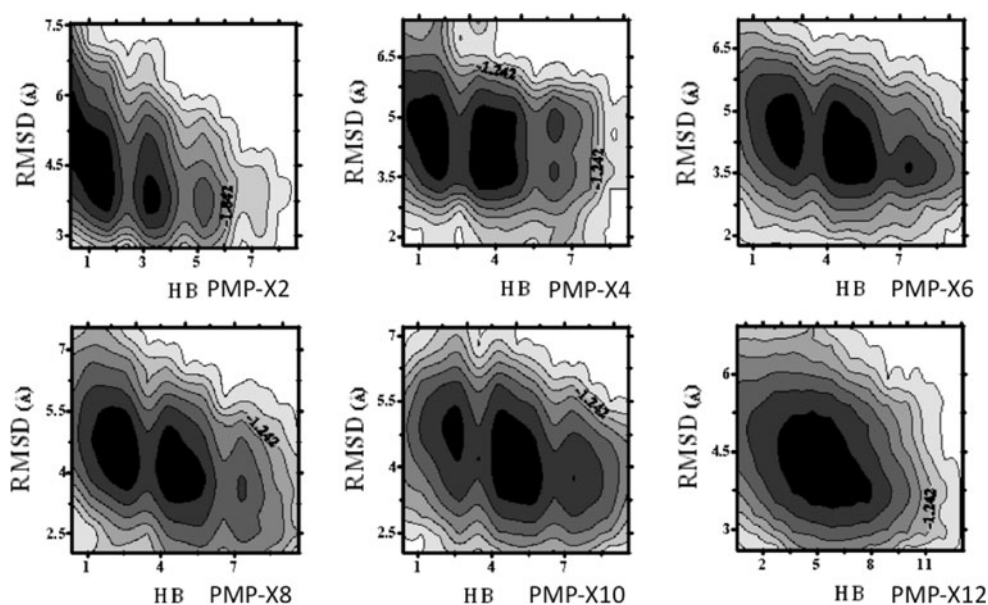

Fig. 12 The contents of secondary structure by DSSP analysis of prosequences at 297 K from (**a**) to (**f**) are PMP-C12, PMP-C10, PMP-C8, PMP-C6, PMP-C4, and PMP-C2, respectively

DPP IV and the promastoparans from PMP-C12 to PMP-C6. However, for the promastoparans PMP-C4 and PMP-C2, there are less hydrogen bonds compared the other cases. The hydrogen bonds for all the promastoparans-DPPIV complexes involve the ARG125 residue of DPPIV, which is also found in the tNPY-complex. However, the residue of the specific substrate involving the hydrogen

bonding is different. It is the third residue of tNPY interacting with the ARG125 residue of DPPIV. On the other hand, they are either the second or fourth residues for all the promastoparans except PMP-C4, which uses the first residue. The TYR547 residue of DPPIV is found in the hydrogen-bonding interaction with PMP-C12, PMP-C10, PMP-C8 and PMP-C2. Interestingly, all of hydrogen

Fig. 13 The local structure of the binding site and with PMP docked inside. **a** is the structure of the binding site including several important measurements to characterize the pocket section. **b** is the docking structure of the PMP-C12 while the pink arrow pointing the bond to be cleaved. **c** is the docking structure of the PMP-C10. **d** is the overlap of the docking structures of all the promastoparans with a different viewing angle, where the dumbbell shape represents the cleavage bond
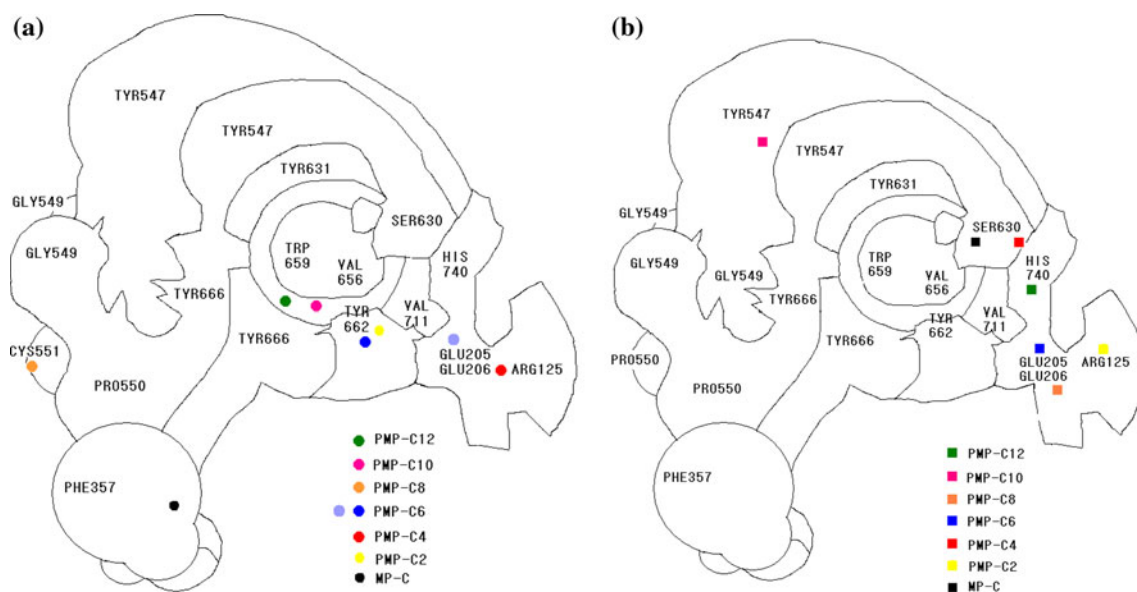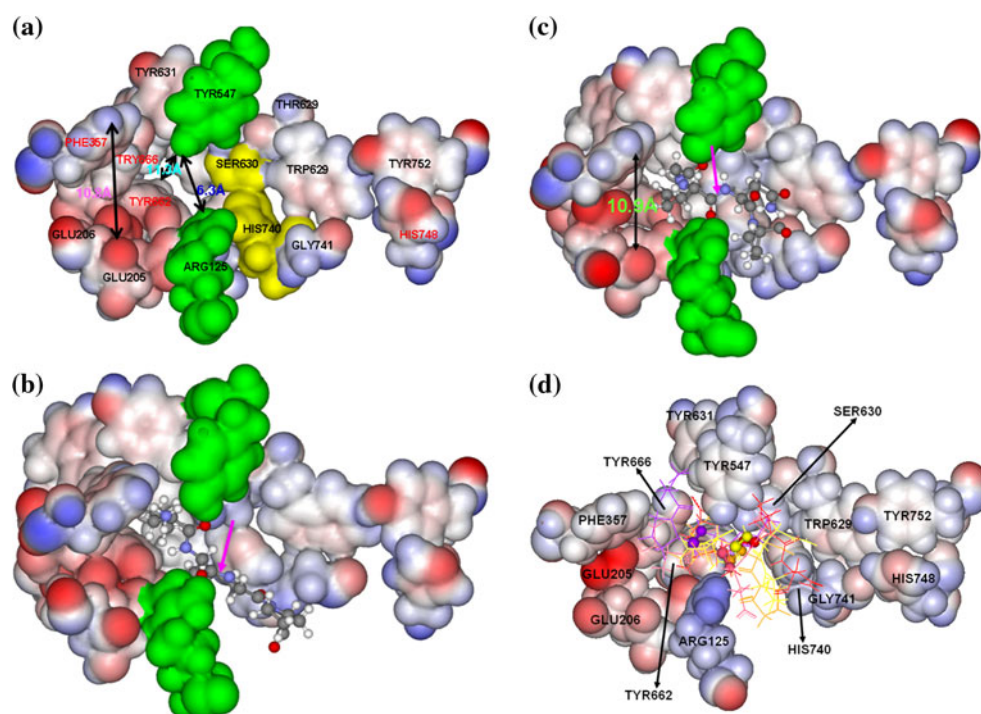
Fig. 14 The 2D projection plot of the inner surface of the pocket section in DPPIV. Each color dot presents the location of the last non-hydrogen atom of the side chain for **a** the first residue and **b** the second residue in a given promastoparan. The two points for the first residue in PMP-C6 is because their distances are the same

bonding in the DDPIV-promastoparan complexes in this study are found in the tNPY-DPPIV complex, except the CYS551 interacting with the first residue of PMP-C8.

With the above analysis, the binding pocket site can be roughly divided into three sections: pocket, gate and adsorption wall sections, as shown in Fig. 15. The gate section is constructed by ARG125 and TYR547 of DPPIV to form a narrow gate which separates the wall and pocket

sections. Different binding affinities to DPP IV can be explained by the structures of the promastoparans. The common characteristics of those docking features are that the relative orientation between promastoparans inside the docking pocket is relatively well preserved (Fig. 16) and every N-terminus of promastoparans are aligned to the pocket site of DPPIV. The pocket, blocked by two gate residues, can effectively accommodate the two residues of

Table 4 The Φ, Ψ angles of the first 4 residues before and after docking for the promastoparans

	Before docking	After docking
PMP-C12		
L1[a]	(0.0, 66.9)[b]	(0.0, 102.2)
A2	(92.1, −164.4)	(62.3, −158.8)
G3	(167.5, 160.7)	(142.9, 179.9)
P4	(−45.6, 0.0)	(−62.8, 0.0)
PMP-C10		
G1	(0.0, −151.8)	(0.0, −84.6)
P2	(−53.5, 128.0)	(−44.8, 135.1)
N3	(−76.4, 118.9)	(−44.8, 135.1)
P4	(−76.8, 0.0)	(−51.8, 0.0)
PMP-C8		
N1	(0.0, 118.0)	(0.0, 115.7)
P2	(−101.0, −8.9)	(−82.8, 3.3)
D3	(−77.2, −19.7)	(−61.4, −79.5)
A4	(−112.7, 0.0)	(−72.7, 0.0)
PMP-C6		
D1	(0.0, 72.0)	(0.0, 91.1)
A2	(−175.4, 131.5)	(−163.4, 137.3)
D3	(−64.0, 145.9)	(−60.4, 140.0)
P4	(−81.0, 0.0)	(−70.6, 0.0)
PMP-C4		
D1	(0.0, 147.3)	(0.0, 99.8)
P2	(−70.5, 157.7)	(−55.3, 130.3)
E3	(−87.7, 7.3)	(−57.3, −9.1)
A4	(−75.8, 0.0)	(−50.0, 0.0)
PMP-C2		
E1	(0.0, 94.2)	(0.0, 138.2)
A2	(−113.4, 164.4)	(−161.4, −172.5)
L3	(−106.7, −8.9)	(−137.9, 13.4)
N4	(−162.4, 0.0)	(−166.4, 0.0)

[a] The number after the single letter representation of amino acid is the sequence number

[b] The first value in the parenthesis is Φ angle and the second one is Ψ angle
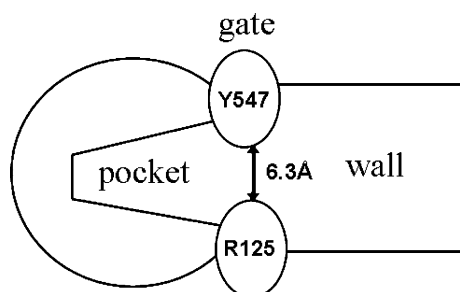
Fig. 15 A schematic plot for the binding site of promastoparan in DPPIV
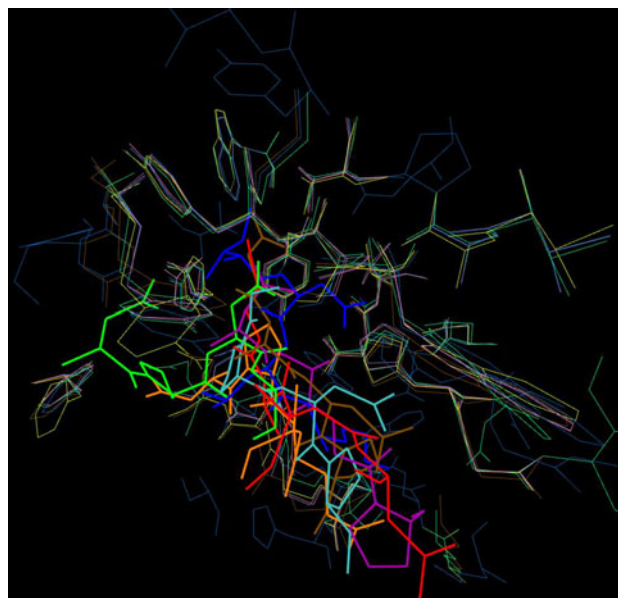
Fig. 16 Overlapped view of docked promastoparans. The promastoparans are indicated as thick line. One can observe the relative orientation is well preserved

promastoparans yet not deep enough to handle more than that. The promastoparans with a size less than or equal than PMP-C4 experience less steric hindrance for effective cleavage while the other PMP-Cn, with $n > 4$, require rather large structural change due to more extensive hydrogen bonds. After the last cleavage, i.e., releasing the dipeptide from the N terminus of the PMP-C2 cut by DPPIV to generate mastoparan, the 1st and 2nd amino acid residues of mastoparan, ILE and ASN, are both with a long side chain, which cannot fit in the narrow gate formed by ARG125 and TYR547 of DPPIV. We speculated that the cleavage by DPPIV, after scissoring full prosequence, were stopped because mastoparan was blocked by ILE and ASN from getting into the binding pocket. As to the determination of the endpoint for the step-wise cleavage process, it can be explained by our proposed model of DPPIV cleavage site in Fig. 15, which has a narrow opening and is shallow but has a wide binding pocket. Based on these observations, the step-wise cleavage process to reach the maturation of mastoparan can be proposed as to the following: due to the shallow binding pocket, only the first two resides in the N terminus of each promastoparan are allowed to enter with the rest of the residues adsorbed onto the wall section to stabilize the docking conformation inside the pocket site, and then undergoing cleavage followed by releasing the corresponding processed product. Because of the cleavage site with a narrow opening and wide binding pocket, only the species with the residue in the penultimate position with a short side chain, such as Ala or Pro, can slide into the binding pocket. The width of

the binding pocket gives the flexibility for the first residue in the N terminus of each promastoparan, whose side chain can point to the residues in the pocket site of DPPIV with favorite interaction. Between the two consecutive insertions to undergo cleavages process, the promastoparans will undergo folding process, usually with a coli conformation in the N-terminal region. This cleavage process reaches its endpoint, when the side chain of the residue in the penultimate position is too big to slide through the narrow gate.

Conclusion

The prosequences of MP-C remain an extended conformation as long as there are extra residues attached in front of the N-terminus of mature MP-C. The step-wise cleavage tendency could be differentiated into two classes depending on the length of substrates residues, i.e. shorter or equal then MP-C4 or even larger. Interestingly, after removal of all the pro part of the prosequence, the native structure of MP-C appears at the end of the repetitive cleavage process. The current study indicates that the specificity of DPPIV toward the promastoparans can be understood in terms of the pocket size, coupled with the action of two gate residues as an effective blocker of deep insertion of substrate toward the pocket. Moreover, our results show that this step-wise cleavage process has a well-defined endpoint when the helical structure of MP-C forms. The success of this ticketing mechanism strongly replies on the pro part of the residues to stop the helix formation. We speculate that the flexibility of the first two residues, as indicated in Fig. 12, dynamically facilitates accommodation of promastoparans. Of course, the current approach has its limitation both in a sense that only static nature of docked promastoparans was investigated and the peptide level simulation model lacks explicit water. Perhaps, the dynamic docking study might further elucidate the detailed docking behaviors and the accompanying cleavage process.

References

1. Hirai Y, Yasuhara T, Yoshida H, Nakajima T, Fujino M et al (1979) A new mast cell degranulating peptide "mastoparan" in the venom of Vespula lewisii. Chem Pharm Bull (Tokyo) 27:3

2. Yokokawa N, Komatsu M, Takeda T, Aizawa T, Tamada T (1989) Mastoparan, a wasp venom, stimulates insulin release by pancreatic islets through pertussis toxin sensitive GTP-binding protein. Biochem Biophys Res Commun 158:712–716

3. Hirata Y, Atsumi M, Ohizumi Y, Nakahata N (2003) Mastoparan binds to glycogen phosphorylase to regulate sarcoplasmic reticular Ca2+ release in skeletal muscle. Biochem J 371:81–88

4. Higashijima T, Uzu S, Nakajima T, Ross EM (1988) Mastoparan, a peptide toxin from wasp venom, mimics receptors by activating GTP-binding regulatory proteins (G proteins). J Biol Chem 263: 6491–6494

5. Higashijima T, Burnier J, Ross EM (1990) Regulation of Gi and Go by mastoparan, related amphiphilic peptides, and hydrophobic amines. Mechanism and structural determinants of activity. J Biol Chem 265:14176–14186

6. Gil J, Higgins T, Rozengurt E (1991) Mastoparan, a novel mitogen for Swiss 3T3 cells, stimulates pertussis toxin-sensitive arachidonic acid release without inositol phosphate accumulation. J Cell Biol 113:943–950

7. Chahdi A, Choi WS, Kim YM, Beaven MA (2003) Mastoparan selectively activates Phospholipase D2 in cell membranes. J Biol Chem 278:12039–12045

8. Lee VSY, Tu WC, Jinn TR, Peng CC, Lin LJ et al (2007) Molecular cloning of the precursor polypeptide of mastoparan B and its putative processing enzyme, dipeptidyl peptidase IV, from the black-bellied hornet, Vespa basalis. Insect Mol Biol 16:231–237

9. Xu X, Yang H, Yu H, Li J, Lai R (2006) The mastoparanogen from wasp. Peptides 27:3053–3057

10. Habermann E (1972) Bee and wasp venoms. Science 177:314–322

11. Kreil G, Haiml L, Suchanek G (1980) Stepwise cleavage of the pro part of promelittin by dipeptidylpeptidase IV. Evidence for a new type of precursor–product conversion. Eur J Biochem 111:49–58

12. Suchanek G, Kreil G, Hermodson MA (1978) Amino acid sequence of honeybee prepromelittin synthesized in vitro. Proc Natl Acad Sci USA 75:701–704

13. Boman HC, Boman IA, Andreu D, Li ZQ, Merrifield RB et al (1989) Chemical synthesis and enzymic processing of precursor forms of cecropins A and B. J Biol Chem 264:5852–5860

14. Gibson BW, Poulter L, Williams DH, Maggio JE (1986) Novel peptide fragments originating from PGLa and the caerulein and xenopsin precursors from Xenopus laevis. J Biol Chem 261:5341–5349

15. Julius D, Blair L, Brake A, Sprague G, Thorner J (1983) Yeast [alpha] factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. Cell 32:839–852

16. Richter K, Egger R, Kreil G (1986) Sequence of preprocaerulein cDNAs cloned from skin of Xenopus laevis. A small family of precursors containin one, three, or four copies of the final product. J Biol Chem 261:3676–3680

17. Kreil G (1990) Processing of precursors by dipeptidylaminopeptidases: a case of molecular ticketing. Trends Biochem Sci 15:23–26

18. Lambeir A-M, Durinx C, Scharpe S, De Meester I (2003) Dipeptidyl-peptidase IV from bench to bedside: an update on structural properties, functions, and clinical aspects of the enzyme DPP IV. Crit Rev Clin Lab Sci 40: 209–294

19. Mentlein R (1999) Dipeptidyl-peptidase IV (CD26)-role in the inactivation of regulatory peptides. Regul Pept 85:9–24

20. McDonald J (1985) An overview of protease specificity and catalytic mechanisms: aspects related to nomenclature and classification. Histochem J 17:773–785

21. Gorrell MD (2005) Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. Clin Sci 108:277–292

22. Kikkawa F, Kajiyama H, Shibata K, Ino K, Nomura S et al (2005) Dipeptidyl peptidase IV in tumor progression. Biochim Biophys Acta (BBA)–Proteins & Proteomics 1751:45–51

23. Drucker DJ, Nauck MA (2006) The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. Lancet 368:1696–1705

24. Idris I, Donnelly R (2007) Dipeptidyl peptidase-IV inhibitors: a major new class of oral antidiabetic drug. Diabetes Obes Metab 9:153–165

25. McIntosh CHS, Demuth H-U, Kim S-J, Pospisilik JA, Pederson RA (2006) Applications of dipeptidyl peptidase IV inhibitors in diabetes mellitus. Int J Biochem Cell Biol 38:860–872

26. Aertgeerts K, Ye S, Tennant MG, Kraus ML, Rogers J et al (2004) Crystal structure of human dipeptidyl peptidase IV in complex with a decapeptide reveals details on substrate specificity and tetrahedral intermediate formation. Protein Sci 13:412–421

27. Engel M, Hoffmann T, Wagner L, Wermann M, Heiser U et al (2003) The crystal structure of dipeptidyl peptidase IV (CD26) reveals its functional regulation and enzymatic mechanism. Proc Natl Acad Sci 100:5063–5068

28. Hiramatsu H, Kyono K, Higashiyama Y, Fukushima C, Shima H et al (2003) The structure and function of human dipeptidyl peptidase IV, possessing a unique eight-bladed [beta]-propeller fold. Biochem Biophys Res Commun 302:849–854

29. Oefner C, D'Arcy A, Mac Sweeney A, Pierau S, Gardiner R et al (2003) High-resolution structure of human apo dipeptidyl peptidase IV/CD26 and its complex with 1-[({2-[(5-iodopyridin-2-yl)amino]-ethyl}amino)-acetyl]-2-cyano-(S)-pyrrolidine. Acta Crystallogr D 59:1206–1212

30. Hori Y, Demura M, Iwadate M, Ulrich AS, Niidome T et al (2001) Interaction of mastoparan with membranes studied by 1H-NMR spectroscopy in detergent micelles and by solid-state 2H-NMR and 15 N-NMR spectroscopy in oriented lipid bilayers. Eur J Biochem 268:302–309

31. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151

32. Swendsen RH, Wang J-S (1986) Replica monte carlo simulation of spin-glasses. Phys Rev Lett 57:2607

33. Case DA, Darden TA, Cheatham TE I, Simmerling CL, Wang J, et al. (2006) AMBER9 University of California, San Francisco

34. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized born model. Proteins: Struct Funct Bioinform 55:383–394

35. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420:102–106

36. Palmer BJ (1993) Direct application of shake to the velocity verlet algorithm. J Comput Phys 104:470–472

37. Daura X, Gademann K, Jaun B, Seebach D, van Gunsteren WF et al (1999) Peptide folding: when simulation meets experiment. Angew Chem Int Ed 38:236–240

38. Shao J, Tanner SW, Thompson N, Cheatham TE (2007) III. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. J Chem Theory Comput 3:2312–2334

39. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S et al (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217

40. Chen R, Li L, Weng Z (2003) ZDOCK: an initial-stage protein-docking algorithm. Proteins 52:80–87

41. Kabsch W, Sander C (1983) Secondary structure definition by the program DSSP. Biopolymers 22:2577–2637

42. Liu S, Zhang C, Zhou H, Zhou Y (2004) A physical reference state unifies the structure-derived potential of mean force for protein folding and binding. Proteins 56:93–101