

ARTICLE

Computational insight into dynamic properties of dipeptidyl peptidase III from *Bacteroides Thetaiotaomicon* and the structural basis for its substrate specificity

Cite this: DOI: 10.1039/x0xx00000x

Received

Accepted

DOI:

www.rsc.org/

M. Tomin^a, and S. Tomić^{a*}

Dipeptidyl peptidase III (DPP III) from the human gut symbiont *Bacteroides thetaiotaomicron* (Bt) is the first identified prokaryotic DPP III orthologue. It has low sequence similarity to the thoroughly studied human DPP III, and differently from eukaryotic orthologues it has a cysteine (Cys450) residue in the zinc-binding motif HEXXGH (HECLGH). Recently determined crystal structure of Bt.DPP III showed that its 3D structure, similarly to the structure of the human DPP III, consists of two domains with a wide cleft inbetween. Although such a striking similarity of the 3D structures of orthologues with low sequence similarity is not surprising, it is no guarantee for similarity of their dynamic properties and the catalytic performance. Here we report results of the molecular modelling study of Bt.DPP III, wild type and its C450S mutant as well as their complexes with characteristic DPP III substrates Arg-Arg-2-naphthylamide (RRNA) and Lys-Ala-2-naphthylamide (KANA). During several hundred nanoseconds of all atom MD simulations of the wild type protein the long range conformational changes which can be described as protein 'closing', have been traced. Similar conformational change we had determined for the human orthologue as well. However, the amplitude of the change is lower for Bt.DPP III than for the human DPP III. The MD simulations have been performed using ff03, ff12SB and ff14SB force fields wherein the results of the last two better fit to the experimental results. The calculated free energies combined with the hydrogen bond analysis indicate reasons for higher substrate specificity of Bt.DPP III towards RRNA than KANA as well as for decrease of the RRNA hydrolysis rate induced by the Cys450 to Ser mutation. The obtained results are in line with the experimental data.

1. Introduction

Peptidases are a type of hydrolases that primarily hydrolyse peptide bonds in proteins and peptides to release amino acids or shorter peptides. Classification of peptidase into families is based on their sequence homology while subdividing of the families into clans is based on similarities in three-dimensional structures, and the sequences comprising catalytic amino acid residues¹.

Dipeptidyl peptidases III (DPP III) or M49 peptidases (according to the MEROPS database, <http://merops.sanger.ac.uk/>)² are zinc-dependent peptidases that cleave dipeptides from the amino-end of their substrates, peptides consisting of 3 or more amino acid residues (preferable 3-8) and have a role in protein catabolism, pain

modulation, as well as in defense against oxidative stress.^{1, 3, 4} The first 3D structure determined for a DPP III was the crystal structure of yeast DPP III (pdb code: 3csk).⁵ Crystal structures of human DPP III followed (pdb codes: 3T6B and 3FVY).⁶ The 3D structure of DPP III from *Bacteroides thetaiotaomicron* was just recently uploaded to PDB. All these structures have the unique protein fold resembling two domains separated by a wide cleft. The zinc-binding site is in the upper, catalytic domain, next to the cleft, and represents a part of the enzyme active site. Both, experiments and calculations revealed that the eukaryotic DPP IIIs exhibit long range conformational motion which can be described as protein closure/opening (two domains approaching each other). During the motion, the conformation of each

protein domain remains conserved.^{6,7} DPP III from *Bacteroides thetaiotaomicron* (Bt.DPP III) is the first bacterial DPP III for which the crystal structure was solved. *Bacteroides thetaiotaomicron* is a Gram-negative anaerobe, a predominant member of the normal intestinal microbiota of mammals.^{8,9}

Sequencing of the *B. thetaiotaomicron* genome revealed 4779 protein coding genes and a highly adaptable organism able to compete with hundreds of the other bacterial species for survival in the colon environment.⁹ *Bacteroides thetaiotaomicron* is of great importance for study of the symbiotic host-bacterial relationship within the human intestine.¹⁰ However, this bacterium is known also as an opportunistic pathogen, resistant to a wide variety of antibiotics,¹¹ which is of clinical interest.

To investigate structural and dynamic properties of the bacterial Bt.DPP III and its complexes with characteristic DPP III substrates, Arg-Arg-2-naphthylamide (RRNA) and Lys-Ala-2-naphthylamide (KANA), we performed a detailed molecular modelling study. Docking, accompanied with the long all atom MD simulations (conventional and accelerated), was performed in order to determine the way of substrates binding into the Bt.DPP III active site as well as to learn about the influence of ligand on protein structure and dynamics. Conventional MD simulations (cMD) powered by the accelerated MD^{12,13} (aMD) simulation enabled efficient searching of the conformational space. Besides just tracing the interconversion from the open to the closed protein form we have been interested in the mechanism driving this conversion. For this purpose the conformational changes determined during simulations of different Bt.DPP III variants as well as their complexes with the characteristic DPP III ligands, were compared and analysed in the light of changes determined for human DPP III.

The present study is the first computational study of a bacterial DPP III. We showed that Bt.DPP III, similarly to the eukaryotic DPP IIIs, experiences the long range conformational motion (considered as protein closure) which is promoted by binding of a substrate into the enzyme active site. However, the extent of this motion determined for Bt.DPP III is smaller than for the human orthologue. Analysis of the intermolecular hydrogen bonds, combined with the MM-PBSA free energy calculations, enabled us to rationalize the experimentally determined difference in the Bt.DPP III catalysed hydrolysis of the synthetic substrates RRNA and KANA as well as to elucidate reasons for significant drop of the RRNA hydrolysis induced by the Cys450 to Ser mutation.¹⁴ The simulations were performed using three

different force fields, ff03, ff12SB and ff14SB. Comparison with the available experimental data suggests that ff12SB and ff14SB are more suitable for study of the zinc-dependent peptidases than the ff03 force field.

2. Methods

2.1. System preparation

The experimentally determined structure of the ligand free Bt.DPP III fourfold mutant was used as the initial structure for our molecular modeling study. In the mutant Cys residues (143, 174, 419 and 435) are replaced by Ser, and Met residues (83, 103, 122, 165, 207, 513, 517, 534, 556 and 660) by selenomethionine (Se-Met). For the purpose of molecular modelling of the wild type enzyme Ser and Se-Met were reverted back to Cys and Met, respectively. The amino acids residues missing at the N terminal of the experimentally determined structure, Met1 – Lys23, were partially (Thr16 – Lys23) reconstructed using the program Modeller 9.14.¹⁵ The first 15 amino acids were omitted because there are strong evidences that they constitute a signalling sequence, as predicted by the SignalP¹⁶ server (Figure S1). Both Se-Met to Met and Cys to Ser transformations were performed by editing the pdb structure and rebuilding the residues with tleap, a basic preparation program for Amber simulations available within the AMBER14 package (<http://ambermd.org>).^{17,18} In the model all Arg and Lys residues are positively charged (+1e) while Glu and Asp residues are negatively charged (-1e), as expected at the physiological (experimental) conditions. The protonation of histidines was checked according to their ability to form hydrogen bonds with neighbouring amino acid residues or to coordinate the metal ion. The protein parametrization was performed within the three AMBER force fields: ff03,¹⁹ ff12SB²⁰ and ff14SB,²¹ while the substrate was parameterized within the generalized amber force field (gaff)²² and the missing parameters were derived using Antechamber²³ module within the Amber14 suit of programs. For the zinc cation, Zn²⁺, parameters derived in previous work were used.²⁴

The proteins and protein-substrate complexes, were placed into the truncated octahedron box filled with TIP3P water molecules,²⁵ and Na⁺ ions²⁶ were added in order to neutralize the systems.

2.2 Simulations.

Conventional MD (cMD) simulations.

System preparation, geometry optimization and equilibration were as followed. Prior to molecular dynamics simulations, the

protein geometry was optimized in three cycles with different constraints. In the first cycle (1500 steps), only water molecules were relaxed, while the protein and zinc atoms were restrained by the harmonic potential with a force constant of 32 kcalmol⁻¹Å⁻¹. In the second (2500 steps) cycle, the same force was applied to the zinc atom while the backbone atoms were relaxed. In the third cycle (1500 steps), all atoms were free to move. During the first period of equilibration (50 ps of gentle heating from 0 to 300K), the *NVT* ensemble was used, while all of the following simulations were performed at constant temperature and pressure (300K and 1 atm, the *NpT* ensemble).

In order to reduce any bias that the force field could have on the simulation results, we performed 200 ns of *NpT* MD simulations (300 K, 1 atm, time step 2 fs) using ff03, ff12SB and ff14SB force fields. The temperature was held constant using Langevin dynamics with a collision frequency of 1 ps⁻¹. Pressure was regulated by a Berendsen barostat. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm.

Accelerated MD (aMD) simulations. The solvated systems (ligand-free DPP III and DPP III-RRNA complex) obtained after 50 ps of heating from 0 to 300K were used as the starting structures for dual boost accelerated MD simulations (aMD). Differently from some alternative methods, the aMD simulations do not require prior knowledge of the potential energy landscape, *i.e.* a reaction coordinate does not need to be defined prior to running simulations. In the dual boosting approach, one boost potential is applied to the torsional terms and the other, different, boost potential to the total potential. The latter diminishes the moment of inertia of bulky water and thus facilitates diffusion of the explicit solvent molecules around the biomolecule. The average total potential energy, $\overline{E_{pot}}$, and the average torsional potential energy, $\overline{E_{dih}}$, for the simulated systems, were obtained from the 200-ns-long cMD simulation of the open DPP III structure (see Table 1). E_r and E_a are the parameters (energy per residue and atom, respectively) used to calculate the potential energy boost $E_T(r)$ and the torsional potential energy boost $E_t(r)$, respectively, as well as the parameters controlling roughness of the boost potentials, α_T and α_t , respectively.^{12, 27} The simulations (*NpT*, time step 2 fs combined with the SHAKE algorithm, Langevin thermostat, Berendsen barostat) were conducted with the AMBER12 package (GPU version)²⁸ using the ff14SB force field. An overview of all simulations performed within this work is given in Table S1.

Table 1. Average energy values from a 200 ns long MD simulation using ff14SB and the parameters required for aMD simulations. All values are given in kcal mol⁻¹.

$\overline{E_{pot}}^a$	-263551.9
$\overline{E_{dih}}^a$	8463.7
E_r^b	1.0
E_a^b	0.1
$E_{threshP}$	-254890
α_P	8662
$E_{threshD}$	9123
α_D	132

^aAverage potential and dihedral energies. ^b E_r and E_a are the energy per residue and atom, respectively, used to calculate threshold energies and α .²⁷

Docking. Recent simulations of the human DPP III – RRNA complex had shown that RRNA forms strong and persistent interactions with the binding site when the enzyme is in the more compact form.⁷ In order to let the substrate achieve the optimal conformation, the partially closed structure obtained from MD simulations of the ligand free protein was used. Docking was performed manually using the structure of the h.DPP III – RRNA complex as a template²⁹. Initial Bt.DPP III – RRNA and Bt.DPP III – KANA complexes were geometry optimized, equilibrated and simulated for 200 ns. The Bt.DPP III – KANA complex was prepared from the structure of the Bt.DPP III – RRNA complex obtained after 100 ns. After minimization and equilibration the Bt.DPP III – KANA complex was simulated for 100 ns. The simulation procedure was same as for the ligand free enzyme. Classical MD using ff12SB and ff14SB, as well as accelerated MD using ff14SB were chosen for fine tuning of the initial, manually performed docking and to trace the influence of ligand on protein dynamics.

MM-PBSA calculations. The substrates binding free energies were obtained using the MM-PBSA approach as implemented in AMBER12 program.³⁰ MM-PBSA calculations were performed for the complexes with the wild type enzyme, DPP III – RRNA and DPP III – KANA, as well as for the complex with its mutant, C450S DPP III, C450S – RRNA. For each complex the binding free energies have been calculated on 5 ns long intervals sampled throughout the trajectory. The calculations were performed using a salt concentration of 0.15M. Additionally the MM-GBSA calculations, utilizing GB model of Onufriev et al.³¹ have been performed as well.

2.2. Data analysis

In order to analyze and to characterize the conformational space that Bt.DPP III structures (experimental and those generated during MD simulations) span, as well as to determine the most relevant motions associated with protein closure, several types of data analysis were performed. All calculations were performed with CPPTRAJ module of the AmberTools program package.³²

Geometric parameters. In order to quantify differences among the Bt.DPP III forms, either generated computationally or

obtained by X-ray diffraction, several geometric parameters, given in Table 2, were analyzed.

In general, the radius of gyration (R_g) is a good measure of the long ranged conformational changes. However, in the case of Bt.DPP III, which has a significant number of unstructured regions (loops) it is not a perfect measure of protein closure. Therefore we used 3 additional geometrical parameters, namely $d_1 - d_3$, as given in Figure 1. Equivalent distances were traced during MD simulations of human DPP III.⁷

Hydrogen bonds. Intermolecular hydrogen bond analysis was performed on same sections of the trajectories as MM-PBSA calculations in order to closely examine the ligand-protein interactions. The cutoff distance and angle for hydrogen bonds was set on a default values of 3.0 Å and 135°, respectively. These relatively tight criteria ensured that only the most relevant interactions were taken into account.

Principal component analysis. As the main criteria for selecting the appropriate computational procedure we considered its ability to adequately sample the protein conformational space. In order to quantify this, two principal components were calculated using the trajectories sampled during entire set of simulations (Table S1), in a way described by Roe et al.³³ and all the data was projected on two PCs.

3. Results and discussion

3.1. Protein flexibility

The wide interdomain cleft, for example distance between the Cα atoms of D14 and N15 residues at the outer edge of the cleft and belonging to the upper and lower domain, respectively, is 42 Å, large enzyme promiscuity, and the presence of two of the five highly conserved regions in the lower protein domain, suggest that the Bt.DPP III, similarly to human and yeast orthologues could experience long-ranged inter-domain motion. In order to thoroughly investigate the long range conformational changes of Bt.DPP III, and to find out how substrate influences these changes, we performed a series of cMD and aMD simulations of ligand free Bt.DPP III and its complexes with RRNA and KANA. The final structures obtained by simulations of the ligand free wild type enzyme are shown in Figure 1 and geometrical parameters determined for these structures are given in Tables 2 and S2. An increase in protein globularity traced during MD simulations, is accompanied by R_g decrease. The R_g values calculated for the initial all-atom crystallographic structure, x-ray, and the closed enzyme structure obtained after 200 ns of MD simulations performed using different force fields (ff03, ff12SB and ff14SB force fields) are given in Tables 2 and S2. The highest and the lowest R_g values are associated with the x-ray structures of Bt.DPP III, oXR, and cWT, respectively (Table 2). However, distances between Cα atoms of D193-K468, D208-K468 and A394-K468, d_1 , d_2 and d_3 , respectively are more appropriate than R_g to describe the mutual orientation of domains, and therefore represent good parameters for the structures differentiation, see Table 2 and Figure 1. D193, D208 and A394 belong to sheet 1 and helices 9 and 16 from the lower domain while K468 belongs to the helix 20 from the upper domain (Fig. 1).

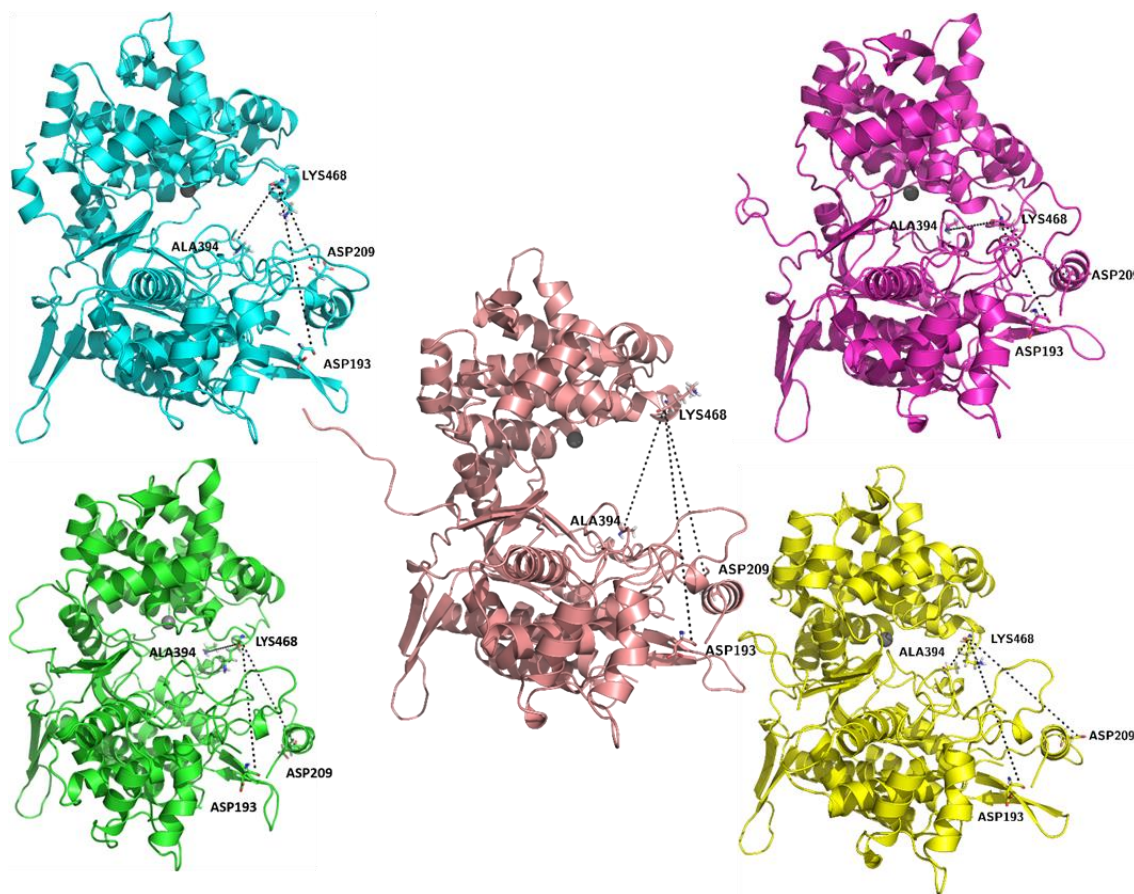


Figure 1. Cartoon representations of four different conformations of Bt.DPP III: open crystal structure, oXR (center), closed structure obtained after 200 ns of MD simulations using ff03 (top left), ff12SB (top right), ff14SB (bottom left) and ff14SB with accelerated MD (bottom right). Residues D193, D209, A394 and K468 are shown in stick representation, the zinc ion as a gray sphere, and d_1 (D193-K468), d_2 (D209-K468) and d_3 (A394-K468) C α distances are indicated by the black dashed lines.

Variability of the conformational space sampled during MD simulations was assessed by similarity of the structures from the most distant conformational regions with the crystallographically determined structures of the ‘open’ and ‘closed’ of Bt.DPP III forms (oXR and cWT, respectively). The shape of the protein domains in the crystallographically determined structures is conserved (*RMSD* between domains in the oXR and cWT 0.435 Å and 0.259 Å for the upper and lower domain, respectively). During MD simulations domains also have not changed their conformation significantly so the *RMSD* calculated for the entire structure mostly represents the difference in mutual orientation of these domains (Figures S2 – S5). However, the other four geometric parameters, R_g , and the $C\alpha$ distances D193-K468 (d_1), D208-K468 (d_2) and A394-K468 (d_3) (Fig. 1), are more appropriate to quantify differences among the sampled conformations. While the closed structure obtained by aMD simulations of the ligand free protein has d_3 (A394-K46) significantly shorter than the crystallographically determined structures oXR and cWT, see Table 2, d_3 of the structures obtained by simulation of the complexes (WT-RRNA and WT-KANA) is similar to that of the cWT structure.

Table 2. Values of geometric parameters used to describe the degree and type of Bt.DPP III closure determined in the most distinct enzyme structures. The oXR and cWT structures are obtained by X-ray diffraction, while structures in the last three columns, WT_{MD}, WT-RRNA and WT-KANA are obtained by molecular modelling. The radius of gyration (R_g) was calculated for the protein backbone atoms. The distances $d_1 - d_3$ were calculated for $C\alpha$ atoms.

geometric parameters	oXR	cWT	WT _{MD} ^a	WT-RRNA ^b	WT-KANA ^b
$R_g / \text{\AA}$	26.7	24.8	25.6	25.3	25.2
$d_1(\text{D193-K468}) / \text{\AA}$	41.2	24.1	33.8	24.0	25.2
$d_2(\text{D208-K468}) / \text{\AA}$	32.9	20.6	25.2	21.8	20.7
$d_3(\text{A394-K468}) / \text{\AA}$	26.0	22.6	7.5	19.9	19.4

^a aMD-ff14SB, ^b cMD-ff14SB + aMD-ff12SB

It should be noted that the experimentally determined so called ‘open’ structure is for the fourfold Cys to Ser mutant while the simulations were performed for the wild type enzyme. In order to more efficiently explore the conformational space, besides conventional, the accelerated MD simulations using the ff14SB force field were performed as well. The added dual-boost potential boosted the protein closure (noticed as the sudden decrease of the radius of gyration, as well as of the d_3 value, Figure 2) and aMD simulations sampled much larger region of the enzyme conformational space (see Fig. 3) than conventional, cMD, simulations. However, the modes of protein closure obtained by aMD and cMD simulations differ (Figure S6). The mode of closure determined by aMD simulations is in agreement with our earlier MD simulations performed for the human DPP III. The mutual orientation of the domains in the WT_{MD} and oXR structures slightly differs. In the WT_{MD} structure domains are

facing each other while in oXT they shifted are in opposite directions in the plain orthogonal to the direction of closure.^{7, 27}

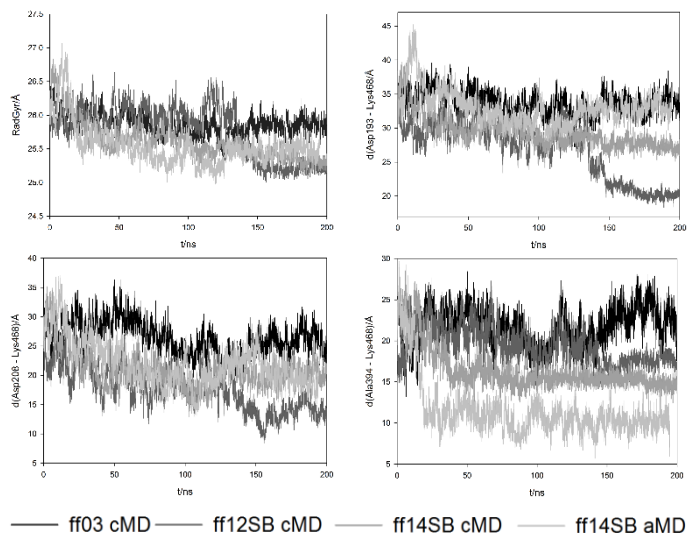


Figure 2. The protein radius of gyration and the selected $C\alpha$ atom distance profiles obtained from 200 ns long MD simulations.

The principal component analysis supports the conclusions based on the comparison of the geometrical parameters (R_g , d_1 , d_2 and d_3). As shown in Figure 3, classical MD simulations using either, ff12SB or ff14SB force field, as well as the accelerated MD with ff14SB force field, explored larger part of conformational space than cMD-ff03 simulations. The conformation region obtained by the latter is limited to the space of the open and semi-closed conformations. Based on these results, cMD-ff12SB and aMD-ff14SB simulations were chosen for the substrate docking.

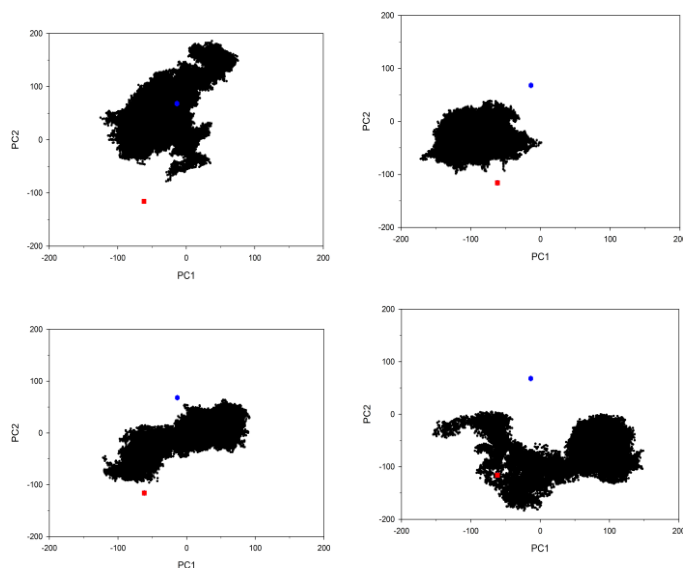


Figure 3. Sampling of the conformational space (defined by the first two principal components) by MD simulations, cMD-ff12SB (upper left), ff03 (upper right), ff14SB (lower left) and aMD-ff14SB (lower right). Position of the experimentally determined structures oXR and cWT is represented by red square and blue hexagon, respectively.

The principal component analysis was performed on the protein backbone atom, omitting the first 10 amino acid residues of the N-terminus, for the set of trajectories sampled during either set of cMD or aMD simulations accomplished with different force fields. The first two principal components describe approximately 73% of total variance, with PC1 accounting for ca. 43% and PC2 for ca. 30% of variability. The concentrated motions described by these principal components correspond to the protein domains' movement, as illustrated in Figure 4. The modes of motion are similar to those described for human DPP III.⁷

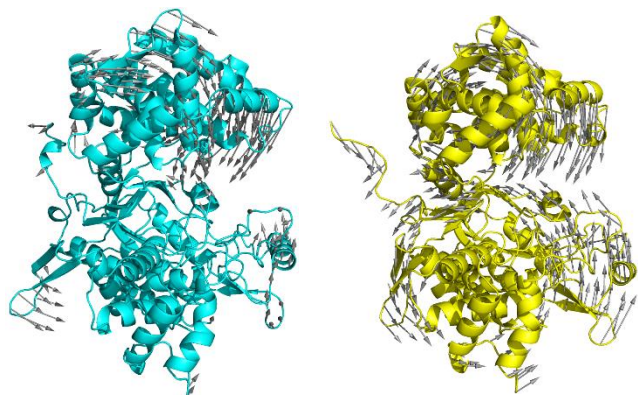


Figure 4. Concentrated protein motions corresponding to the first (left) and the second (right) principal component.

Normal mode analysis was performed using the Delarue group server,³⁴ and the obtained modes agree well with the principal components calculated from our trajectories (results not given). RMSF values calculated for the individual residues of APO form revealed that the residues from the upper (Leu328 – Lys365; Ser455 – Ala477; Val521 – Ala619) and lower (Val175 – Val235) domain which constitute the outer rim of the interdomain cleft correspond to the most flexible regions (Figures S7 – S10) indicating that the long range protein closure is indeed the dominant structural change noticed during MD simulations. Similar results have been previously reported for the human orthologue as well.⁷

3.2. Zinc ion coordination

The crystal structure of the ligand-free Bt.DPP III revealed the six-coordinated zinc ion coordinated by two histidines (His463 and His458), two glutamates (Glu449 and Glu476) and two water molecules (Figure 5) wherein both glutamates coordinate Zn^{2+} monodentately. These His and Glu residues are a part of two conserved regions in the DPP III family, HEXXGH(HG) and EECR(K)A.³⁵

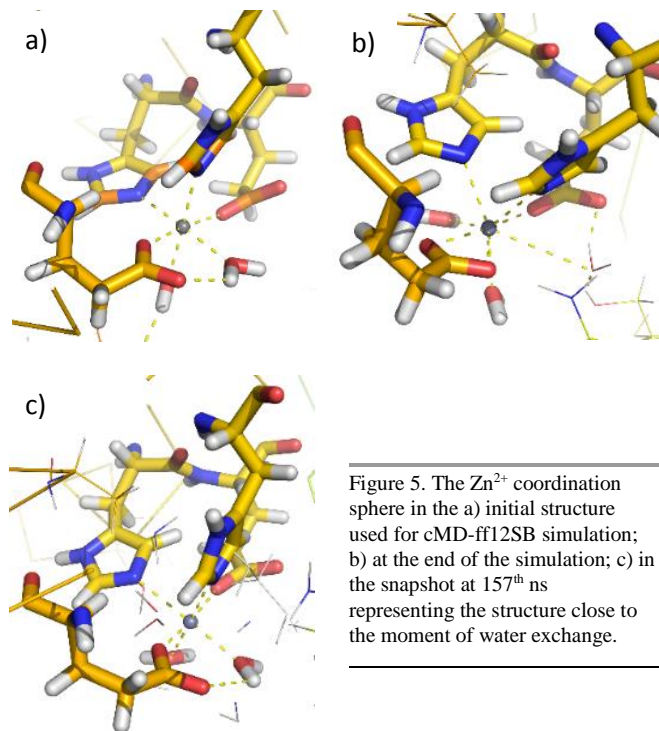


Figure 5. The Zn^{2+} coordination sphere in the a) initial structure used for cMD-ff12SB simulation; b) at the end of the simulation; c) in the snapshot at 157th ns representing the structure close to the moment of water exchange.

During MD simulations His463, His458, Glu449 and Glu476 remained in the Zn^{2+} coordination sphere, even more, differently from the initial structure, glutamates occasionally coordinated the metal ion bidentately. However, Glu449 and Glu476 never coordinate Zn^{2+} at the same time, see for example Figure 6). Such behavior of glutamates aids to the zinc ion stabilization during the water exchange. During aMD-ff14SB simulations, exchange of the carbonyl oxygens in the zinc coordination sphere was noticed only in the case of Glu476 (Figures S11 and S12).

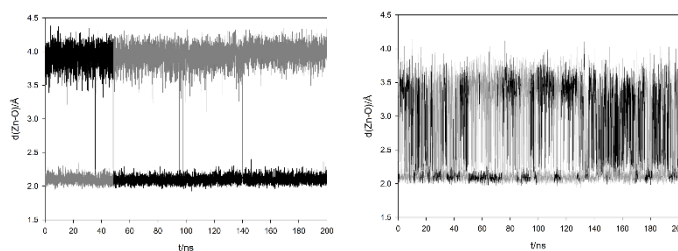


Figure 6. Distance between Zn^{2+} and the carboxyl oxygens of coordinating Glu449 (left) and Glu476 (right) residues during aMD-ff14SB simulation.

During the first 100 ns of simulations the water molecules exchanged fairly often within the Zn^{2+} coordination sphere (Figure 7) but by the protein closure exchange of water molecules from the coordination sphere with those in the bulk becomes less frequent, and the two molecules trapped in the Zn^{2+} coordination sphere at 150th ns remain there until the end of simulation. The number of water molecules in the Bt.DPP III active site and close to it significantly decreased during the last 50 ns of MD simulations.

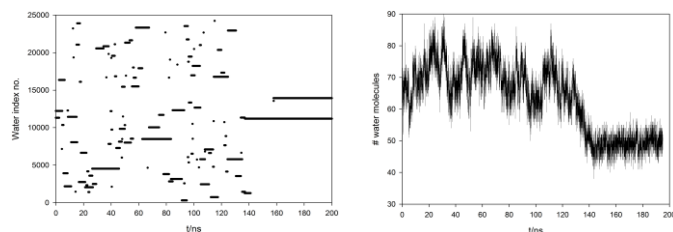


Figure 7. Water molecules in the Bt.DPP III active site, a) “Resistance time” for the water molecules coordinating Zn^{2+} in the unligated enzyme during 200 ns of cMD-ff12SB simulation. b) Number of water molecules occupying the 10 Å sphere around the zinc ion over the course of the 200 ns of cMD-ff12SB simulation

3.3. The role of Ser340 and Tyr386

The computational studies revealed important role of Ser340 for the protein closure. Namely the Ser340 – Asp446 hydrogen bond formed at the early stage of MD simulations (5th – 10th ns) facilitates the interdomain motion through the Ser363-Asp446-Arg333-Glu432 chain of interactions. Similar role was determined for its analogue Asp372 in human DPP III.³⁶ Tyr286 is 100% conserved within the M49 family and is presumed to be a part of the S1 active subsite.³⁷ Both experimental³⁸ and computational⁷ studies have shown its functional role in the human orthologue (Tyr318). The magnitude of its translocation during MS simulations of the ligand free Bt.DPP III is fairly large, with distance from its C α atom to Zn^{2+} decreasing from 29.7 Å in the open crystal structure to 20.6 Å after at 200th ns of MD simulations. In the experimentally determined ‘closed’ structure of the Bt.DPP III this distance is 22.2 Å. During the MD simulations resulting with the protein closure Tyr286 forms hydrogen bonds with Lys468, Gly461 and, most notably, Ser472 which is thought to be a part of the S2 subsite.³⁷

3.4. MD simulations of the DPP III – RRNA complex

The structure of the human DPP III – Arg2-2NA (RRNA) complex from our previous computational works^{7, 36} was used as a template to construct the Bt.DPP III – substrate complexes. The substrates were docked into the semi-closed Bt.DPP III conformation determined during MD simulations of the ligand free enzyme.

The initial binding mode of the preferred substrate, RRNA closely resembles the binding mode of the peptide tynorphin in the crystal structure of human DPP III-tynorphin complex (pdb code: 3T6B). In this structure tynorphin has the form of a β -strand and binds to the five-stranded β -core of the lower domain in an anti-parallel fashion.⁶ Our MD simulations suggest that RRNA binds to Bt.DPP III in the same manner, antiparallel to the Gly383 – Leu386 strand of the β -sheet 5. This is in agreement with RRNA binding to the human ortholog.^{7, 27}

In order to study the effect of the ligand binding on the rate, degree and type of the protein closure, we simulated the Bt.DPP III – RRNA complex for 200 ns using cMD-ff12SB and aMD-ff14SB approaches. By comparing the radii of gyration and the

geometrical parameters $d_1 - d_3$ (Figure 1), we can conclude that the ligand boosts the protein closure (Figures 2 and 8). A sudden drop in the radius of gyration during the first 50 ns of MD simulations is clearly visible in the case of the complex, while the unliganded enzyme needed around 80 ns to achieve the similar R_g value. The distance between the Asp193 and Lys468 C α atoms oscillates around 28 Å over the first 130 ns of the simulation in the case of unliganded enzyme, whereas the same distance dropped to 25 Å within the first 80 ns of MD simulations of the complex.

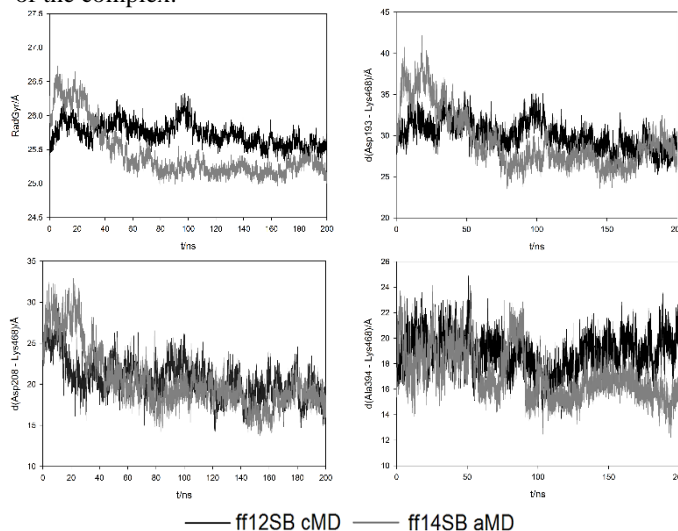


Figure 8. The protein radius of gyration and the selected C α atom distance profiles determined for the Bt.DPP III – RRNA complex during 200 ns long MD simulations.

When considering water population in the active site (Figures S13 and S14) we can notice several fast water exchanges in the zinc coordination sphere in the period of 40th - 110th ns of the cMD-ff12SB simulation time. This corresponds to the substrate reorientation within the enzyme active site. During the second half of the simulation time fluctuations of RRNA molecule decreased (RMSF value decreased from 1.97 Å during the first 100 ns of the simulation to 1.28 Å during the last 100 ns) and it sterically protects the Zn^{2+} ion from bulky water molecules. The average number of water molecules within 10 Å of the zinc ion determined for the last 50 ns of simulations was only about 47 (± 4) wherein the number of water molecules coordinating Zn^{2+} was two during most of the simulation time. Similar trend of substrate stabilization has been observed during the aMD-ff14SB simulations as well (RMSF decreased from 2.21 Å to 1.54) Å. However, differently from the cMD-ff12SB simulation, Zn^{2+} was coordinated by only one water molecule during aMD-ff14SB simulation (Figure S15) and the Zn^{2+} coordination number varied between 6 and 7 depending on the Glu476 orientation and its participation in the metal ion coordination as a mono- or bidentate ligand (Figure S16). During entire cMD-ff12SB simulation of the Bt.DPP III – RRNA complex Glu476 coordinated the metal ion monodentately with exchange of the carboxyl oxygens in the zinc coordination sphere (Figure S17). All these findings are in accord with previously determined high

plasticity of the zinc ion coordination. Glu449 coordinates Zn^{2+} exclusively in a monodentate manner through O2 in both types of simulations (Figures S18 and S19).

The principal component analysis (Figure 9) clearly indicates that the presence of substrate influences the Bt.DPP III closure in the direction of cWT form. Differently from the aMD-ff14SB simulation of the ligand free enzyme, the protein conformational change observed during aMD-ff14SB simulation of the Bt.DPP III – RRNA complex corresponds to the protein closure to the experimentally determined cWT form, i.e. form in which domains are with respect to the initial, open structure shifted in opposite directions in the plain orthogonal to the direction of closure (Figure S20). This finding is in agreement with our previous results on the human orthologue that ligand binding into the inter-domain cleft enhances the protein closure and domains shifting.

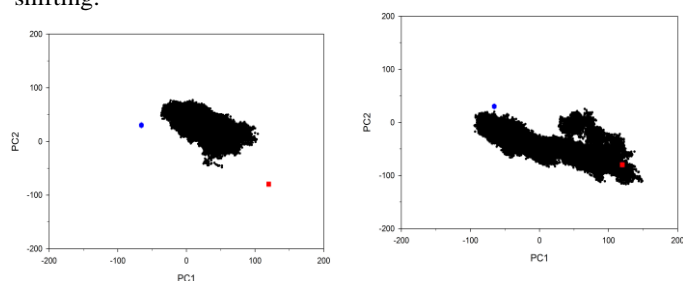


Figure 9. The Bt.DPP III conformational space, defined by the first two principal components, sampled during the cMD-ff12SB (left) and aMD-ff14SB (right) simulations of Bt.DPP III – RRNA complex. Position of the experimentally determined structures oXR and cWT is represented by red square and blue hexagon, respectively.

During the simulations RRNA accommodated into the active site in the similar manner as tynorphin in the human orthologue, with the naphthalene ring located in a hydrophobic pocket (Figure 10).⁷ The most notable difference between the binding modes obtained by cMD-ff12SB and aMD-ff14SB simulation procedures is the position of the substrate aromatic ring. During the aMD simulation, it enters deeper into the hydrophobic pocket than during cMD simulation. In both cases, RRNA is stabilized by hydrogen bonds with Glu320, Gly383 and Glu476. The final structure of the complex obtained by cMD-ff12SB simulation contains an additional hydrogen bond between RRNA and Asn385, while in the structure obtained by aMD-ff14SB simulation RRNA is H-bonded to Glu531 (Figure 10).

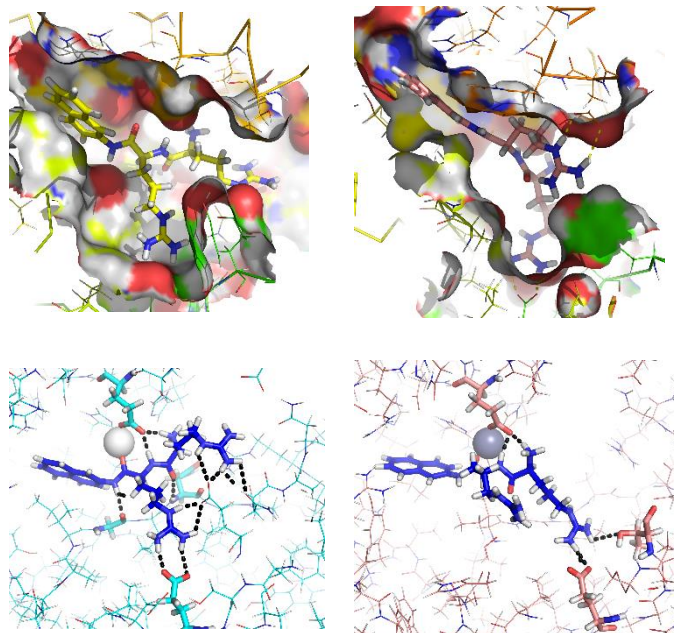


Figure 10. Electrostatic potential surface of the active site calculated by APBS plugin for PyMOL³⁹ and intermolecular Bt.DPP III - RRNA hydrogen bonds (blue) in the lowest energy conformations obtained by cMD-ff12SB (left) and aMD-ff14SB simulations (right).

In order to determine the most stable complex conformation, the lowest-energy conformation obtained by the aMD-ff14SB simulation was taken as a starting point for additional 100 ns of cMD-ff12SB simulation. As a result, a conformation given in Figure 11 was obtained. In addition to the 4 hydrogen bonds that RRNA formed in previously acquired structures (with Glu320, Gly383, Glu476 and Glu531), in this complex RRNA is further stabilized by hydrogen bonds with His119, Thr306 and Glu307 (Table 4 and Figures S21 and S22). Consequently, the binding free energy decreased.

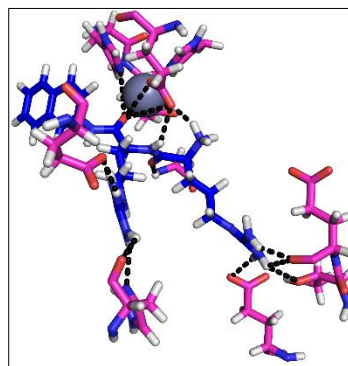


Figure 11. Hydrogen bond network in the Bt.DPP III - RRNA obtained after 100 ns of aMD-ff14SB followed by 100 ns of cMD-ff12SB simulation.

Surface-accessible solvent area (SASA) was calculated using NACCESS.^{40, 41} Large decrease of per-residue all-atom SASA ($>45 \text{ \AA}^2$) in the final Bt.DPP III – RRNA complex in comparison to the unliganded enzyme indicates that RRNA hinders the solvent access to the cleft-lining amino acid residues, as well as to the “hinges” connecting the two domains (Figure S23 and S24). Similar

behaviour has been reported for human DPP III complexes with RRNA.⁷

3.4. MD simulations of the DPP III – KANA complex

The experimental data available for the wide variety of dipeptidyl-2-naphthylamides shows that Bt.DPP III most efficiently catalyzes hydrolysis of RRNA while the hydrolysis rate for Lys-Ala-2NA (KANA) is an order of magnitude lower.¹⁴ In order to rationalize this data, KANA was docked into a semi-closed enzyme structure obtained after 100 ns of simulations of the Bt.DPP III – RRNA complex. The system was heated and minimized as described in system preparation and simulated for 100 ns using classical MD with ff12SB and accelerated MD with ff14SB force field. After these simulations, the same procedure as for the Bt.DPP III – RRNA complex was used. The binding free energies obtained by MM-PBSA and given in Tables 3 and S3 are in line with the experimental data reported by Vukelić et al.¹⁴

Table 3. The binding free energies for the Bt.DPP III-RRNA and Bt.DPP III-KANA complexes obtained by MM-PBSA calculations.

Complex	Method	$\Delta E_{PB}/\text{kcalmol}^{-1}$
DPP III - RRNA	cMD-ff12SB	-49.76
DPP III - KANA	cMD-ff12SB	-41.41
DPP III - RRNA	aMD-ff14SB	-57.83
DPP III - KANA	aMD-ff14SB	-35.33
DPP III - RRNA	aMD-ff14SB + cMD-ff12SB	-67.27
DPP III - KANA	aMD-ff14SB + cMD-ff12SB	-47.68

The difference in the binding affinities of RRNA and KANA can partly be explained by the hydrogen bond analysis, see Table 4. Apparently, the population of the intermolecular substrate-protein hydrogen bonds is significantly higher in the Bt.DPP III – RRNA complex than in the Bt.DPP III – KANA complex. For example the substrate stabilization by the hydrogen bonds with Gly383 and Glu320, commonly present in the complex with RRNA, was not observed in the Bt.DPP III – KANA complex. According to the structural differences between RRNA and KANA such result was expected.

Table 4. Hydrogen bonds population (%). The analysis was performed for the lowest-energy 5 ns long fragments of the 200 ns long trajectories used to calculate the binding energies. Only the hydrogen bonds occurring in >30% of the sampled structures are given.

Acceptor	RRNA ^a	RRNA ^b	KANA ^a	KANA ^b
Glu307 (I)	31.08	43.88	47.48	-
Glu307 (II)	-	40.84	-	-
Glu320 (I)	70.88	59.44	-	-
Glu320 (II)	-	32.40	-	-
Asp375	-	-	38.88	-
Gly383	71.84	-	-	-
Glu476 (I)	98.72	35.08	66.76	97.56
Glu476 (II)	33.36	34.64	55.84	75.56
Glu476 (III)	-	33.24	31.28	-

Glu531	-	62.72	-	54.40
--------	---	-------	---	-------

Extracted from the trajectory obtained during ^a200 ns of cMD-ff12SB simulation and ^b100 ns of cMD-ff12SB simulation following 200 ns of aMD-ff14SB simulation.

3.5. MD simulations of the C450S DPP III – RRNA complex

The earlier performed kinetic measurements clearly indicate that Cys450 to Ser mutation significantly reduces catalytic performances of Bt.DPP III. For example, in comparison with the wild type enzyme C450S mutant has 1.4 times higher K_m and an order of magnitude lower k_{cat}/K_m ($\mu\text{M}^{-1}\cdot\text{s}^{-1}$) for the RRNA hydrolysis.¹⁴ In order to rationalize this data we performed MD simulations of the C450S DPP III – RRNA complex using the same starting structure as in the wt-Bt.DPP III – RRNA simulations wherein the mutations were performed using the AMBER module tleap. The binding free energies calculated from the obtained trajectories are in qualitative agreement with the experimental data (Table 5). Besides, the hydrogen bond analysis also supports the results of kinetic measurements (see Table 6).

Table 5. The binding free energies for wild-type Bt.DPP III and C450S Bt.DPP III-RRNA complexes obtained by MM-PBSA calculations.

Complex	Method	$\Delta E_{PB}/\text{kcalmol}^{-1}$
wt-DPP III	cMD-ff12SB	-49.76
C450S DPP III	cMD-ff12SB	-44.24
wt-DPP III	aMD-ff14SB	-57.83
C450S DPP III	aMD-ff14SB	-34.27

The substrate stabilization by Glu476 is similar in both complexes, but Glu307, Glu320 and Gly383 stabilize RRNA stronger in the complex with the wild type enzyme than in the complex with the mutant. On the other hand RRNA stabilization by Glu449 is tighter in the C450S – RRNA than in the wt-Bt.DPP III – RRNA complex (Table 6). In both complexes the Glu449 carboxyl oxygen participates in Zn^{2+} coordination.

Table 6. Hydrogen bond analysis on lowest-energy 5ns fragments of used to calculate the binding energies

Acceptor	% of frames WT - RRNA	% of frames C450S - RRNA
Glu476 (I)	98.72	99.00
Gly383	71.84	14.44
Glu320	70.88	-
Glu476 (II)	33.36	21.60
Glu307	31.08	-
Glu449	-	77.84

Conclusions

In this work, we performed a detailed MD study (cMD and aMD) of bacterial, *Bacteroides thetaiotaomicron*, DPP III, a two-domain metallopeptidase, in order to study its structural and dynamical properties as well as the influence of Cys to Ser mutation and ligand binding to these properties.

In agreement with the experimental evidence (*i.e.* X-ray structures of the open and closed Bt.DPP III, oXR and cWT, respectively), MD simulations revealed that Bt.DPP III can transform from the extended structure determined by X-ray diffraction, oXR, to a more compact form. However, this transformation is not unambiguous; namely there are two distinct compact structures, cWT and WT_{MD}, that Bt.DPP III could adopt as well as a series of intermediate, so called semi-closed structures. The common feature of all these conformations is the conserved structure of each protein domain, *i.e.* the long range conformational changes of Bt.DPP III can be described as inter-domain movements. The main difference between the two compact structures is in mutual orientation of its domains. While in the WT_{MD} DPP III structure the protein domains are ‘facing’ each other similarly as in the open structure, in the cWT structure they are shifted in opposite direction in the plain orthogonal to the direction of closure (Figure S20).

Similar dynamical behavior was earlier found in human DPP III.²⁷ However, the amplitude of the interdomain separation is significantly larger in human DPP III than in DPP III from *Bacteroides t.* Apparently, the binding site plasticity is higher in human DPP III than in bacterial, Bt.DPP III resulting with higher substrate promiscuity of the previous.

Earlier studies revealed that the closing of human DPP III is an entropy-driven process with the major contribution being the release of ordered water molecules from the cleft.^{6, 7, 27} Similarly during the simulations presented in this work release of water molecules from the inter-domain cleft of Bt.DPP III was noticed. Further on the MD simulations presented in this work indicated that, like in the case of human orthologue, substrate binding into the enzyme active site dictates the protein closure and shifts the equilibrium in the direction of the active, cWT, form.

Finally, the MD simulations combined with the binding free energy calculations and the hydrogen bond analysis revealed that affinity of KANA binding into the Bt.DPP III active site is significantly lower than that of RRNA. Further on, stabilization of the preferred RRNA substrate in the Bt.DPP III active site by hydrogen bonds is disturbed by the Cys450 to Ser mutation. In conclusion, the present study corroborates the experimental findings¹⁴ that Bt.DPP III much more efficiently hydrolyzes RRNA than KANA, as well as the results of the kinetic measurements that revealed significant drop of the RRNA hydrolysis by the Cys450 to Ser mutation.

Acknowledgements

The work has been supported by Croatian Science Foundation under the project “7235 Flexibility, activity and structure correlations in the dipeptidyl peptidase III family”. Some of the calculations have been performed at the Croatian National Grid Infrastructure (CRO NGI, <http://www.cro-ngi.hr/>). We are

grateful to Dr. Maja Abramić for her valuable suggestions and fruitful discussion.

Notes and references

^a Department of Physical Chemistry, Rudjer Boskovic Institute, Bijenička cesta 54, 10000, Zagreb.

[†] Electronic Supplementary Information (ESI) available. Additional tables and figures of the MD (cMD, aMD, and sMD) simulations as described in text. See DOI: 10.1039/b000000x/

1. M. Chen and A. J. Barrett, *Handbook of Proteolytic Enzymes*, Elsevier, London, second edn., 2004.
2. S. C. Prajapati and S. S. Chauhan, *FEBS J*, 2011, **278**, 3256–3276.
3. T. Chiba, Y.-H. Li, T. Yamane, O. Ogikubo, M. Fukuoka, R. Arai, S. Takahashi, T. Ohtsuka, I. Ohkubo and N. Matsui, *Peptides*, 2003, **24**, 773–778.
4. Y. Liu, J. T. Kern, J. R. Walker, J. A. Johnson, P. G. Schultz and H. Luesch, *Proc Natl Acad Sci U S A*, 2007, **104**, 5205–5210.
5. P. K. Baral, N. Jajcanin-Jozic, S. Deller, P. Macheroux, M. Abramic and K. Gruber, *J Biol Chem*, 2008, **283**, 22316–22324.
6. G. A. Bezerra, E. Dobrovetsky, R. Viertlmayr, A. Dong, A. Binter, M. Abramic, P. Macheroux, S. Dhe-Paganon and K. Gruber, *Proc Natl Acad Sci U S A*, 2012, **109**, 6525–6530.
7. A. Tomic, M. Gonzalez and S. Tomic, *J Chem Inf Model*, 2012, **52**, 1583–1594.
8. L. E. Comstock and M. J. Coyne, *Bioessays*, 2003, **25**, 926–929.
9. J. Xu, M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper and G. J. L., *Science*, 2003, **299**, 2074–2076.
10. J. Xu and J. I. Gordon, *Proc Natl Acad Sci U S A*, 2003, **100**, 10452–10459.
11. I. Brook, *Expert Rev. Anti Infect. Ther.*, 2007, **5**, 991–1006.
12. L. C. Pierce, R. Salomon-Ferrer, F. d. O. C. Augusto, J. A. McCammon and R. C. Walker, *J Chem Theory Comput*, 2012, **8**, 2997–3002.
13. D. Hamelberg, C. A. de Oliveira and J. A. McCammon, *J Chem Phys*, 2007, **127**, 155102.
14. B. Vukelic, B. Salopek-Sondi, J. Spoljaric, I. Sabljic, N. Mestrovic, D. Agic and M. Abramic, *Biol Chem*, 2012, **393**, 37–46.
15. A. Sali and T. Blundell, *Journal of Molecular Biology*, 1993, **234**, 779–815.
16. T. N. Petersen, S. Brunak, G. von Heijne and H. Nielsen, *Nat Methods*, 2011, **8**, 785–786.
17. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J Comput Chem*, 2005, **26**, 1668–1688.
18. R. D. A. Case, J. T. Berryman, R. M. Betz, D. S. Cerutti, T. E. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. M. Merz, G. Monard, P. Needham, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, R. Salomon-Ferrer, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, Y. D.M. and P. A. Kollman, *Journal*, 2015.
19. Y. Duan, C. Wu, S. Chowdhury, C.-M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, L. Taisung, J. Caldwell, J. Wang and P. A. Kollman, *Journal of Computational Chemistry*, 2003, **24**, 1999–2012.
20. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
21. J. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.
22. J. Wang, R. M. Wolf, J. Caldwell, P. A. Kollman and D. A. Case, *J Comput Chem.*, 2004, **25**, 1157–1174.
23. J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J Mol Graph Model*, 2006, **25**, 247–260.
24. B. Bertosa, B. Kojic-Prodic, R. C. Wade and S. Tomic, *Biophys J*, 2008, **94**, 27–37.
25. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926.

26. I. S. Joung and T. E. Cheatham, *J Phys Chem B*, 2008, **112**, 9020-9041.
27. A. Tomic, M. Berynskyy, R. C. Wade and S. Tomic, *Mol Biosyst*, 2015, **11**, 3068-3080.
28. R. Salomon-Ferrer, A. W. Gotz, D. Poole, S. Le Grand and R. C. Walker, *J Chem Theory Comput*, 2013, **9**, 3878-3888.
29. A. Tomic and S. Tomic, *Dalton Trans*, 2014, **43**, 15503-15514.
30. B. R. Miller, T. D. McGee, J. Swails, N. Homeyer, H. Gohlke and A. Roitberg, *J Chem Theory Comput*, 2012, **8**, 3314-3321.
31. A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383-394.
32. D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
33. D. R. Roe, C. Bergonzo and T. E. Cheatham, *J. Phys. Chem. B*, 2014, DOI: 10.1021/jp4125099, 3543-3552.
34. E. Lindahl, C. Azuara and D. M., *Nucleic Acids Res*, 2006, **34**, W52-W56.
35. M. Abramic, J. Spoljaric and S. Simaga, *Period biol*, 2004, **106**, 161-168.
36. A. Tomic, M. Abramic, J. Spoljaric, D. Agic, D. M. Smith and S. Tomic, *J Mol Recognit*, 2011, **24**, 804-814.
37. A. Cvitesic, I. Sabljic, J. Makarevic and M. Abramic, *J Enzyme Inhib Med Chem*, 2016, **0**, 1-6.
38. B. Salopek-Sondi, B. Vukelic, J. Spoljaric, S. Simaga, D. Vujaklija, J. Makarevic, N. Jajcanin and M. Abramic, *Biol Chem*, 2008, **389**, 163-167.
39. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, *Proc Natl Acad Sci U S A*, 2001, **98**, 10037-10041.
40. B. Lee and M. Richards, *J. Mol. Biol.*, 1971, **55**, 379-400.
41. S. J. Hubbard and J. M. Thornton, *Journal*, 1993.