Dalton Transactions

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/dalton

Journal Name

RSCPublishing

ARTICLE

Hunting the human DPP III active conformation: combined thermodynamic and QM/MM calculations

Cite this: DOI: 10.1039/x0xx00000x

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Antonija Tomić[a] and Sanja Tomić[b]

Multiple choices of the protein active conformations in flexible metalloenzymes complicate study of their catalytic mechanism. We used three different conformations of human dipeptidyl-peptidase III (DPP III) to investigate influence of the protein environment on the ligand binding and the Zn^{2+} coordination. MD simulations followed by calculations of binding free energy components accomplished for a series of DPP III substrates, synthetic and natural, revealed that binding of the β-strand shaped substrate to the five stranded β-core of the compact DPP III form (in antiparallel fashion) is the preferred binding mode in agreement with the experimentally determined structure of the DPP III inactive mutant-tynorphin complex (Bezerra *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**). Previously it was proposed that the catalytic mechanism of DPP III is similar to that of thermolysin, which assumes exchange of five and four coordinated Zn^{2+}, and activation of Zn-bound water by a nearby Glu. Our QM/MM calculations, performed for altogether 18 protein structures with different zinc ion environment, revealed that only in the most compact DPP III form, the 5-coordinated metal ion is more favourable than the 6-coordinated one. Besides, in this structure E451 is H-bonded to the metal ion coordinating water. Also our study revealed two constrains for the broad substrate specificity of DPP III. One is the possibility of substrate to adopt the β-strand shape and the other is its charged N-terminus. Altogether, we assume that human DPP III active conformation would be the most compact form, similar to the "closed X-ray" DPP III structure.

1 Introduction

The dipeptidyl-peptidase III (DPP III; EC 3.4.14.4) is a two domain zinc-exopeptidase from the peptidase family M49 (according to MEROPS database, www.merops.ac.uk) that cleaves dipeptidyl residues from an unsubstituted N-terminus of its substrates[1–3]. The motifs HEXXGH and E*EXR(K)AE(D) are considered as their trademark[2,4], wherein histidines of the first motif and Glu* of the second one take part in the zinc ion coordination, while Glu from the first motif is crucial for the DPP III catalytic activity (found at 3.9 Å from the zinc ion, see Figure 1)[5]. Members of M49 family, characterized by five highly conserved amino acid regions, are involved in intracellular protein catabolism, pain modulation and defense against oxidative stress[2,6,7]. However, the broad DPP III specificity toward peptides of varying lengths and compositions still has not been properly understood.

The crystal structures of human DPP III (PDB code 3FVY, and hereinafter referred as "open X-ray" structure) and the C130S mutant from yeast DPP III (PDB code 3CSK)[8] revealed tetrahedral zinc coordination, made by H450, H455, E508 (numbering according to human DPP III) and water. One additional water molecule is found at 2.6 Å from the Zn^{2+} ion in the structure of the human DPP III (see Figure 1). Differently, the MD simulations of the human DPP III indicated the octahedral zinc ion coordination, contributed by one additional water molecule and Glu451 carboxyl oxygen, as the most preferred one[9,10].

Discrepancy between the experimental data and the results obtained by MD simulations prompted quantum mechanical researches on zinc coordination in DPP III described in this study. The research is additionally justified by the problems associated with the parameterization of the metal ions, and lack of proper parameters for the zinc cation in the specific environment defined by the structure of the M49 family of enzymes.

Besides, many metalloenzymes, like alcohol dehydrogenase, carbonic anhydrase[11], metallopeptidases, like carboxypeptidase, thermolysin, neprilysin and matrix metalloproteinases,[12] rely on Zn^{2+} for their catalytic activities. Analysis that we have performed using the BioMe server (http://metals.zesoi.fer.hr/metals/) for the protein

structures in the Protein Data Bank (March 10th 2013) showed that the majority of the zinc sites are 4-coordinated (61.4 %), while 24 % and 12.4 % of them are 5- and 6-coordinated, respectively. By excluding the zinc-fingers, tetrahedral zinc coordination remained the most abundant (52 %), but share of five and six coordinated zinc ion rose to 29.7 % and 15.7 %, respectively.

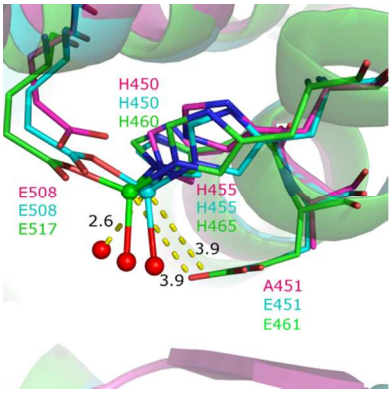


Figure 1. Overlay of the zinc-binding sites from three DPP III crystal structures: the yeast C130S mutant (PDB code 3CSK, carbon and zinc atoms are colored green) and human, wild-type (PDB code 3FVY, carbon and zinc atoms are colored cyan), and the E451A mutant (PDB code 3T6B, carbon atoms are colored purple). Water molecules are shown as red spheres. Distances between some of the pairs discussed within the article are shown with yellow dashed lines.

A recently determined structure of the human DPP III E451A mutant, cocrystallized with the opioid peptide tynorphin, (PDB codes 3T6B and 3T6J and hereinafter referred as "closed X-ray" structure)[13] is more globular than the previously determined one. Although the electron density for the central zinc ion is missing in this structure, positions of the amino acid residues involved in metal ion coordination are preserved (see Figure 1). The increased protein globularity, which is a result of large scale conformational change that can be described as protein closure, have been observed during molecular dynamic (MD) simulations of the ligand-free human DPP III[10]. However, the two "closed" protein structures, the experimental one and the one obtained by long MD simulation, differ (see Fig.S1)[10,13]. One of the differences is the relative position of the two α-helixes colored black in Fig.S1, one from the "upper" and the other from the "lower" protein domain. In the "closed" structure obtained from long MD simulations (hereinafter referred as "closed MD" structure) these helices are positioned just one above the other, while in the "closed X-ray" DPP III structure they are "next" to each other. Consequently, the Zn^{2+} coordinating histidines and glutamic acid residue from "upper" protein domain are in the "closed MD" structure positioned just above the five-stranded β-core that forms the core of the DPP III "lower" domain, while in the "closed X-ray" they are slightly shifted.

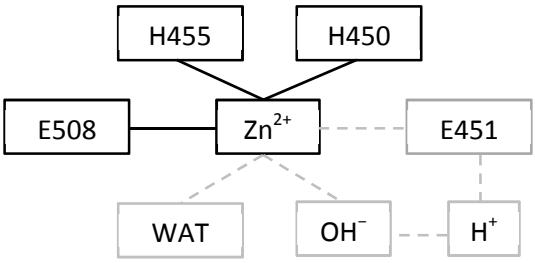
In order to find out which of the available DPP III structures is the most likely to be the active one we conducted a multiscale computational study.

Thermodynamic data calculated for two different ways of ligand binding (determined experimentally and by computational method) into different protein conformations were combined with the QM/MM calculation of Zn^{2+} coordination. The results enabled us to explain the experimentally determined wide substrate specificity of DPP III and to suggest the preferred mode for a ligand binding. The zinc ion coordination flexibility, as well as pronounced ligand binding stability (both, natural and synthetic), enabled us to declare the "closed" human DPP III as the active one.

2 Methods

2.1 Unbound enzyme

Initial Structures Preparation - MM and MD Calculations. We used three different DPP III structures in our studies, "open X-ray" [*o*WT], "closed X-ray" [*c*WT] and "closed MD" [$c\mathrm{WT}_{\mathrm{MD}}$] structure, representing three different Zn-coordination environments (clarification of the names and codes is given in Table S1). In each, three different types of the zinc coordination; tetrahedral, distorted trigonal bipyramidal and octahedral, were considered (see Figure S2 and Scheme 1).



Scheme 1. The zinc ion coordination environment considered in QM/MM calculations. The residues shown in black square coordinate the zinc ion in all initial structures, while the residues shown in gray square exchange in coordination.

The initial structures used for quantum mechanical-molecular mechanical (QM/MM) studies (shown in Figure S2) were prepared from publicly available enzyme crystal structures ("open" and "closed" DPP III forms, PDB codes 3FVY and 3T6B, respectively) subjected to energy minimization and molecular dynamics (MD) simulations.

Since the "closed" enzyme structure from the PDB (code: 3T6B) represents the Glu451Ala mutant of human DPP III and lacks the electron density for the zinc ion in its active site, before the simulations we mutated Ala451 back to Glu and added the zinc ion using the "open" DPP III crystal structure (3FVY) as a template. Thus obtained, wild type, "closed" DPP III as well as the crystallographically determined "open" human DPP III, with all crystallographic

water molecules preserved, were parametrized using AMBER ff03 force field of Duan *et al.*[14]. All Glu, Asp, Arg and Lys residues were protonated according to the respective protonation state in aqueous solution at pH=7. Protonation of the histidines was checked according to their ability to form hydrogen bonds with neighboring amino acid residues or to coordinate the metal ion. For the zinc cation, Zn^{2+}, the non-bonding parameters derived in our previous work[15] and modified according to the PDB survey were used (charge 2.0 *e*, while the VdW radius and energy well are 1.22 Å and 0.250 kcal/mol, respectively)[16]. Structures were placed in the truncated octahedron filled with TIP3P water molecules (additional 18061 and 8216 water molecules, respectively) and in order to neutralize the systems, 24 sodium ions were placed in the vicinity of the negatively charged amino acid residues at the protein surface. The resulting systems were simulated using periodic boundary conditions. The electrostatic interactions were calculated using the particle-mesh Ewald method[17,18]. Prior to MD simulations, the protein geometry was optimized in three cycles with different constraints. In the first cycle (1500 steps), only water molecules were relaxed, while protein and zinc atom were constrained using the harmonic potential with a force constant of 32 kcal/(mol Å^2). In the second (2500 steps) and the third cycle (1500 steps), the same force was applied to the zinc atom while the protein backbone was constrained with 32 and 10 kcal/(molÅ^2), respectively. During the first period of equilibration (50 ps of gentle heating from 0 to 300 K), the *NVT* ensemble was used, while following 100 ps of classical MD simulations (water density adjustment and productive simulations) were performed at constant temperature and pressure (300 K and 1 atm, the *NpT* ensemble). The temperature was held constant using Langevin thermostat with a collision frequency of 1 ps^{-1}. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm.

The initial structures for study of the hexa coordinated zinc ion were obtained as a result of unconstrained MD simulations: 100 ns long [*c*WT$_{MD}$] and 0.1 ns long [*o*WT and *c*WT]. Namely after only 100 ps simulations the octahedral zinc coordination (made by H450, E451, H455, E508 and two water molecules), was established in both, the "open X-ray" and "closed X-ray" structures. The QM/MM calculations were performed on the partially solvated protein structures, namely only 52 water molecules closest to the zinc ion (that correspond to 10 Å water sphere around the zinc ion in "open" enzyme structure) were retained. From these, five and four coordinated zinc ion structures were prepared in which the ionization states for the Zn-bound water molecules and the neighbor glutamic acid (E451) residue were varied (see Scheme 1). Altogether, eighteen structures, 6 different types of zinc coordination in three different protein conformations, each containing 52 water molecules, were used to examine the zinc coordination. The structures are named according to protein form, *o*WT, *c*WT, *c*WT$_{MD}$, the Zn^{2+} coordination number and presence of the ligands E451, water(s) and OH^- (see Figure S2. for detailed description).

ONIOM CALCULATIONS. Quantum mechanics-molecular mechanics (QM/MM) geometry optimization of the zinc coordination sphere in the DPP III active site were performed using the *2-layer* ONIOM (*Our own N-layered Integrated molecular Orbital and molecular Mechanics*) methodology implemented in program GAUSSIAN09:

$$E^{\mathrm{ONIOM}}=E_{\mathrm{MM}}(\mathrm{S})+E_{\mathrm{QM}}(\mathrm{SM})-E_{\mathrm{MM}}(\mathrm{SM}) \quad (1)$$

where S represents the whole system and SM, the smaller, quantum mechanically treated part of it[19,20]. Empirically calculated interactions between the two regions SM (QM) and S-SM (MM) could be considered as a low level energy perturbation of the SM region by rest of the system.
At the places where borders between two regions (layers) "cut" covalent bonds (Cα-Cβ), link atoms were placed. The QM layer, consisting of H450, E451, H455 and E508 side chains, the zinc ion and four water molecules, was treated by the DFT method (M06[21] and B3LYP[22,23] functional) and two different basis sets; 6-31+g(d,p) for the H, N, C and O atoms, and LANL2DZ-ECP for the Zn^{2+} atom[24]. The MM part (the rest of the protein) was treated by the AMBER force field (param96)[25]. To take into account solvation, additional 48 water molecules, closest to the central zinc ion, were considered as a part of the MM layer as well. During QM/MM geometry optimization, only 16 water molecules closest to the zinc ion (correspond to number of water within 7Å of the Zn^{2+} ion in the "open X-ray" structure) were allowed to move. The net charge of the whole system was –24 *e*, while the QM region had a zero charge. Initial cycle of geometry optimizations was performed using mechanical embedding (ME). In this way the interactions between two layers are treated only at MM level; however, the final geometry optimization was performed using electronic embedding (EE)[20]. In the EE approach the electrostatic interactions between QM and MM regions (Coulombic interaction between the QM electrons and nuclei with the partial charges of the atoms in the MM region) are explicitly built in QM Hamiltonian and therefore allow polarization of the electron wave function.

2.2 Ligated enzyme

CLASSICAL MD SIMULATIONS OF THE DPP III IN COMPLEX WITH ITS SUBSTRATES. Recently determined structure of the "closed" DPP III-tynorphin complex (PDB codes 3T6B and 3T6J)[13] revealed ligand binding mode different from the one obtained computationally (RRNA$_{MD}$)[9,10]. The simulations performed for Arg-Arg-2-naphthylamide (RRNA) molecule bound into both, "open" (ligand-free enzyme, PDB code 3FVY) and "closed" (obtained after 72 ns of the MD simulations of the "open" form) enzyme forms, *o*WT-

$RRNA_{MD}$ and cWT_{MD}-$RRNA_{MD}$ complexes, respectively (see Figures 2a-b), had shown that RRNA made stronger and more persistent interactions with the enzyme binding site when it was in more compact form.[10] In order to explore all relevant binding modes we performed a new set of MD simulations. For this purpose three additional DPP III-RRNA complexes were prepared, *o*WT-RRNA, cWT_{MD}-RRNA and *c*WT-RRNA complexes. In all of them RRNA was bound in a way to mimic the tynorphin binding (RRNA), *i.e.* it was docked into the active site in an extended conformation, with its first two arginine residues forming a β-strand bounded to the five-stranded β-core of the DPP III in the antiparallel fashion, see Figures 2c-e.

For complex parametrization, geometry optimization, equilibration and MD simulations the same procedure was used as in our earlier work[10]. The only difference was that during the equilibration, cWT_{MD}-RRNA complex was shortly heated to 400 K and then cooled back to 300K, altogether within 40 ps, in order to allow substrate molecule to find more natural position in the binding site. The complexes *o*WT-RRNA, *c*WT-RRNA and cWT_{MD}-RRNA were simulated at 300K for 50, 80 and 30 ns, respectively, with a time step of 1 fs and structures were sampled every 1 ps. In order to examine convergence of ligand binding modes obtained after 30 (or 50 and 80) ns of MD simulations at room temperature, the complexes were subjected to additional 10 ns of MD simulation at 350 K with the time step of 1 fs. The simulations at the increased temperature should facilitate and speed up the conformational transitions of a system that is not stable.

In order to get better insight into the DPP III substrate specificity, and since it is already known that DPP III also cleaves some important opioid peptides[26], Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu), endomorphin-1 (Tyr-Pro-Trp-Phe-NH2) and endomorphin-2 (Tyr-Pro-Phe-Phe-NH2) were bound in the "closed X-ray" DPP III structure in a way to mimic the tynorphin binding, *c*WT-L-EN, *c*WT-EN1 and *c*WT-EN2 complexes, respectively. The same computational procedure was used as in the complexes with RRNA. Complexes with the peptides were simulated for 30 ns.

MM-PBSA CALCULATIONS. The substrates binding free energies were calculated using the MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) approach[27] as implemented in AMBER12 program. MM-PBSA calculations were performed for all five DPP III-RRNA complexes (shown in Figure 2), and three DPP III-peptide complexes using a single trajectory approximation. For each complex, three 5 ns long sections of the trajectories were considered.

The calculations were accomplished for the enzyme with the dielectric constants 2.0 immersed into the solvent with dielectric constant 80. The ion concentration was 0.05 M. The polar component of solvation enthalpy was calculated using the Poisson-Boltzmann method and the non-polar component was determined by $\Delta H_{nonpol} = \gamma SASA + \beta$, where the solvent accessible surface area (SASA) was calculated with the MolSurf program[28]. The surface tension γ and the offset β were set to the standard values of 5.42×10^{-3} kcal/(mol Å^2) and 0.92 kcal/mol, respectively. The zinc charge was +1.5 *e*. For more details see our previous publications[9,10]. The entropic contribution to the binding free energy (translational, rotational and vibrational) was calculated for the period from the 15th till 30th ns (considering every 200th ps) using Nmode in Amber12. Since the normal-mode analysis is computationally expensive, we only considered the residues lining the enzyme binding site, as this part of the enzyme experiences the largest changes upon the ligand binding. Namely, instead of the whole protein the truncated, either enzyme or enzyme-substrate complex, *i.e.* protein residues Val308–Cys519, were considered.

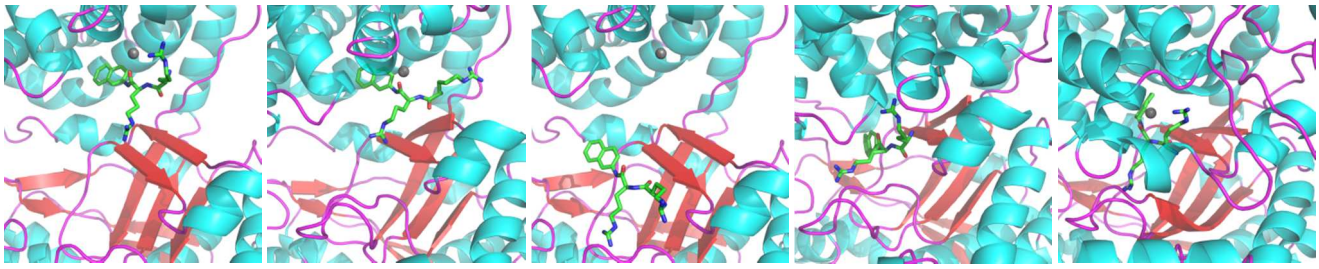
3 Results and discussion

3.1 The active protein conformation - assumption based on DPP III-RRNA complexes simulation.

In order to elucidate the relative stability of distinct ligand binding modes, 30 ns long MD simulations of five different enzyme-RRNA complexes, shown in Figure 2, were performed. For this purpose three different protein conformations (*o*WT, *c*WT and cWT_{MD}) and two RRNA binding modes (by X-ray diffraction, RRNA, and computationally determined ones, $RRNA_{MD}$) were considered.

Contributions to the binding free energy, enthalpy calculated with MM-PBSA approach, and entropy normal mode based approximation, are given in Table 1.

The calculations agree with the findings proposed by Tomic *et al.*[10] that DPP III-RRNA complex stability increases by protein compactness. In addition, the results revealed binding of the ligand in an extended conformation, as indicated by Bezzera *et al.*[13] for tynorphin, as the preferred one. In this binding mode two arginine residues (of Arg-Arg-2-naphthylamide substrate) form a β-strand that in the antiparallel fashion binds to the five-stranded β-core of DPP III (as shown in Figure 3a). It should be noted that the absolute free energy values are meaningless and that we were interested primarily in the binding free energy difference. To be more precise, the most relevant here is the trend of the binding free energy changes.



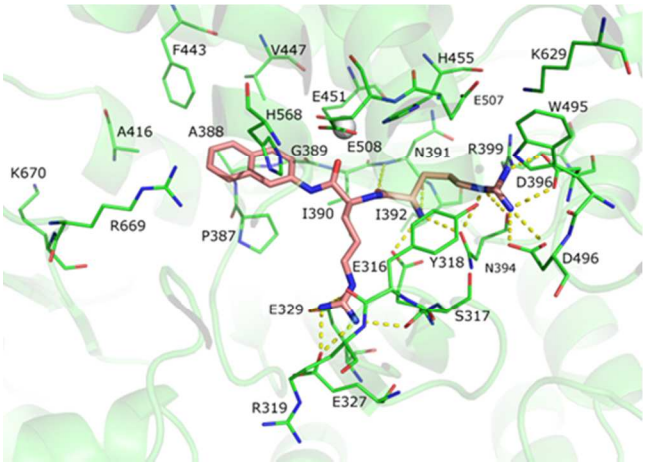
a) b) c) d) e)

Figure 2. Five different DPP III-RRNA complexes used as starting structures for classical MD simulations: a) *o*WT-RRNA$_{MD}$, b) *c*WT$_{MD}$-RRNA$_{MD}$, c) *o*WT-RRNA, d) *c*WT$_{MD}$-RRNA, and e) *c*WT-RRNA. The substrate molecule (shown in stick representation) and central zinc ion (shown as a sphere) are colored by the atom type. The hydrogen atoms are not shown for clarity.

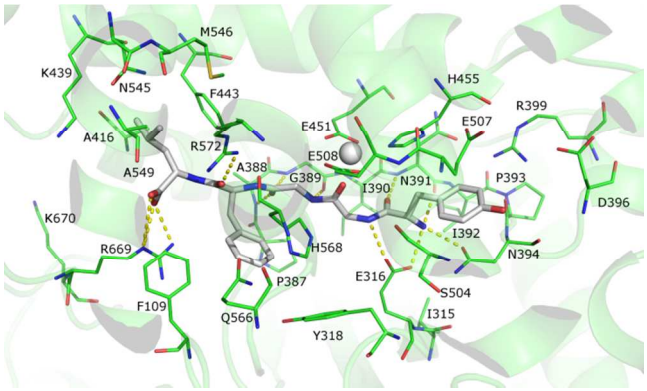
Table 1. The enthalpy and entropy contributions to the substrate (Arg-Arg-2-naphthylamide, Leu-enkephalin, endomorphin-1 and endomorphin-2, RRNA, L-EN, EN-1 and EN-2, respectively) binding free energy calculated using the MM-PBSA approach (the solute dielectric constant was 2 and that of the solvent 80). The zinc charge used for the calculations was +1.5 *e*. Standard deviations are given in brackets.

	15-20 ns	20-25 ns	25-30ns	15-30ns			
system	ΔH/ kcal mol^{-1}	ΔH/ kcal mol^{-1}	ΔH/ kcal mol^{-1}	$T\Delta S_{TRA}$/ kcal mol^{-1}	$T\Delta S_{ROT}$/ kcal mol^{-1}	$T\Delta S_{VIB}$/ kcal mol^{-1}	$T\Delta S_{TOT}$/ kcal mol^{-1}
*o*WT-RRNA$_{MD}$	−37.2 (6.50)*	−43.74 (5.32)*	−43.95 (5.60)*	−13.25 (0.00)	−11.52 (0.05)	1.19 (9.98)	−23.58 (9.99)
*c*WT$_{MD}$-RRNA$_{MD}$	−56.04 (5.31)*	−55.54 (5.07)*	−58.60 (4.52)*	−13.25 (0.00)	−11.59 (0.02)	−1.77 (7.64)	−26.61 (7.64)
*o*WT-RRNA	−69.55 (3.73)	−70.11 (4.12)	−69.07 (3.46)	−13.25 (0.00)	−11.55 (0.05)	2.89 (7.19)	−21.92 (7.18)
*c*WT$_{MD}$-RRNA	−65.13 (4.82)	−67.77 (5.03)	−64.82 (6.04)	−13.25 (0.00)	−11.46 (0.05)	−7.63 (7.54)	−32.34 (7.55)
*c*WT-RRNA	−119.69 (5.21)	−115.85 (4.95)	−116.48 (5.04)	−13.25 (0.00)	−11.62 (0.01)	−11.24 (6.1)	−36.12 (6.10)
*c*WT-L-EN	−68.06 (4.27)	−65.41 (4.29)	−63.18 (3.79)	−13.42 (0.00)	−11.90 (0.02)	0.49 (8.99)	−24.82 (8.98)
*c*WT-EN1	−72.84 (4.24)	−73.05 (4.02)	−73.05 (4.56)	−13.50 (0.00)	−11.81 (0.01)	1.83 (6.21)	−23.48 (6.21)
*c*WT-EN2	−60.69 (4.26)	−62.09 (4.23)	−62.18 (3.61)	−13.45 (0.00)	−11.71 (0.01)	−2.16 (10.21)	−27.31 (10.22)

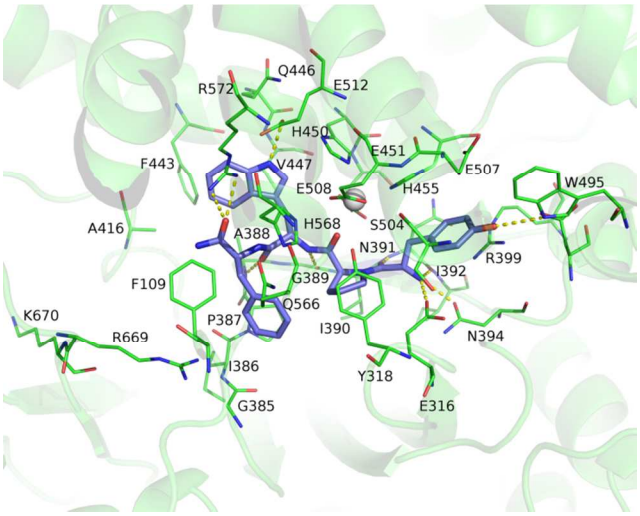
* Tomić *et. al.*[10]



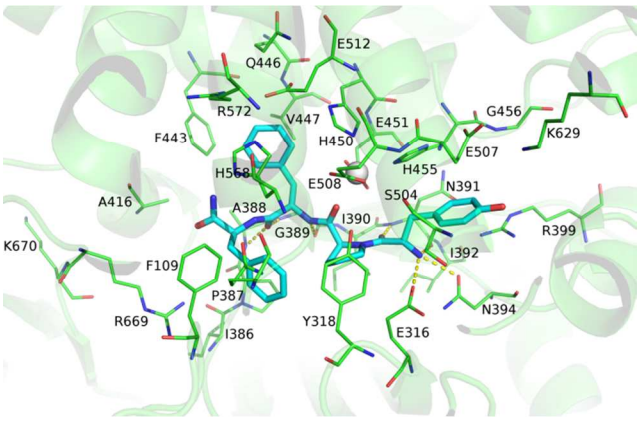
a)



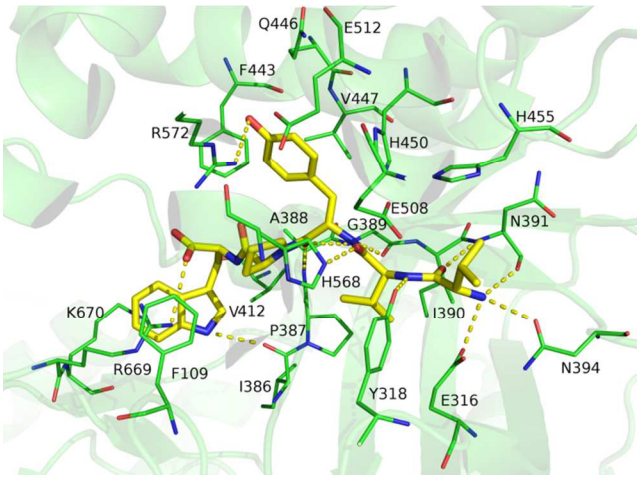
b)



c)



d)



e)

Figure 3. DPP III (*c*WT) complexes with five different substrates: a) Arg-Arg-2-naphthylamide (the substrate is shown in stick representation with carbon atoms colored pink), b) Leu-enkephalin (the substrate is shown in stick representation with carbon atoms colored gray), c) endomorphin-1 (the substrate is shown in stick representation with carbon atoms colored blue) and d) endomorphin-2 (the substrate is shown in stick representation with carbon atoms colored cyan) after 30 ns of MD simulations (followed by 5000 cycles of geometry optimization), and e) tynorphin (the substrate is shown in stick representation with carbon atoms colored yellow) in the X-ray structure (3T6B complex structure). Residues within 4 Å of the substrates are displayed. Polar contacts are shown with yellow dashed lines. The zinc ion is shown as a white sphere. Hydrogen atoms are not shown for clarity. Figures were made in PyMol (http://www.pymol.org/)[29].

The free energy decomposition analysis performed by MM-GBSA method (Table S2), enabled us to detect the strongest protein-substrate interactions. In the case of RRNA the hydrogen bonds between the substrate N-terminus and its first carbonyl group and protein residues (E316, N391 and N394, N391, respectively), as well as the hydrogen bonds between the side chains of RRNA arginines, 1st and 2nd, and D396 and D496, and E316, E327 and E329, respectively, were present during entire simulations of *c*WT-RRNA complex. Majority of these hydrogen bonds were also present during simulations of the *o*WT-RRNA and cWT_{MD}-RRNA complexes, like the hydrogen bond between substrate N-terminus and E316, and hydrogen bonds that substrate 2nd arginine side chain makes with E327 and E329. The significant contribution to substrate (RRNA) stabilization by the zinc ion is observed only in the case of *c*WT-RRNA complex where it as about 2.1 Å from the scissile peptide bond (Table S2). In the cWT_{MD}-RRNA complex this distance is ~10 Å. In addition, the hydrogen bond network between the substrate backbone and backbone of the five-stranded β-core present in *c*WT-RRNA and *o*WT-RRNA complexes, is lost in the cWT_{MD}-RRNA complex. On the other hand, in cWT_{MD}-RRNA, the 1st arginine side chain is hydrogen bonded to E508, while this is not the case in *c*WT-RRNA.

According to RMSF, per atom, calculations (see Figure 4a), RRNA has the lowest flexibility when bound in the extended conformation to the "closed" DPP III, either of two forms. For the RRNA bound in the same conformation to the "open X-ray" DPP III structure, significantly higher flexibility of the 1st arginine side chain (atom numbers 1-7) as well as of the naphthylamide group (atom numbers 21-33), was determined. Apparently, in all complexes except *c*WT-RRNA, the 1st arginine was more flexible than the 2nd one during MD simulations. Pronounced flexibility of the 2nd arginine side chain (atom numbers 14-20) was observed in the complexes with RRNA bound in a way determined by MD simulations, *i.e.* cWT_{MD}-$RRNA_{MD}$ and *o*WT-$RRNA_{MD}$. As expected, the protein globularity and the ligand flexibility are negatively correlated. During simulations at increased temperature neither the substrate flexibility profiles nor its mobility have changed significantly (Figs.4b and S3). Only flexibility of the substrate 1st arginine side chain and of naphthylamide group slightly increased in the case of *c*WT-RRNA and cWT_{MD}-RRNA, and cWT_{MD}-RRNA and cWT_{MD}-$RRNA_{MD}$ complexes, respectively. This suggests that the obtained ligand binding modes are well defined, converged (aligned RRNA binding modes obtained after MD simulations performed at 300 and 350 K are shown in Fig.S4).

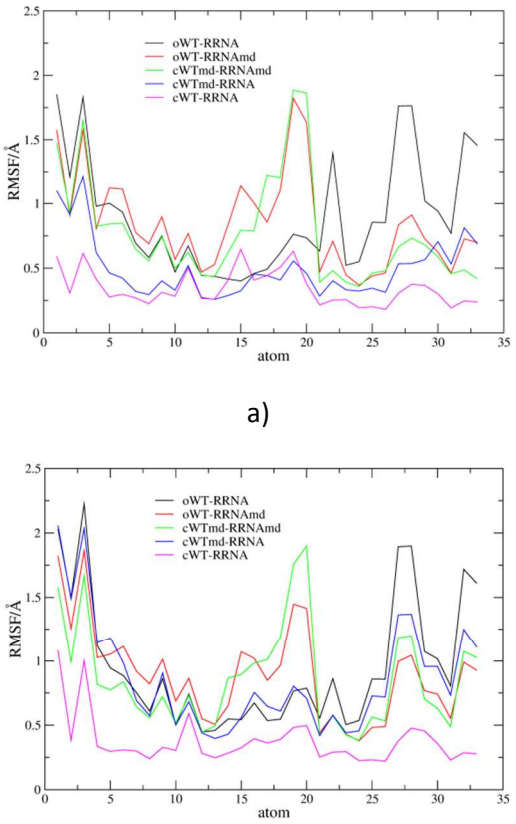


Figure 4. The atom based RMSF profiles calculate for the substrate (RRNA) molecule in all five DPP III-RRNA complexes: a) for the period from 15th till 30th ns (or 50 and 80 ns) of MD simulations, and b) during the last 10 ns at elevated temperature of 350 K. Hydrogen atoms were not considered. Atom numbers 1-7 and 14-20 constitute 1st and 2nd arginine side chains, respectively, atoms 8-13 and 21-23 substrate backbone and atoms 24-33 naphthyl group.

When RRNA binds in form of β-strand antiparallely to the five-stranded β-core of DPP III, its 1st Arg residue is stabilized by numerous interactions with amino acids from the both protein domains, while the 2nd Arg interacts mainly with amino acid residues from the "lower" protein domain; finally the naphthylamide group is equally stabilized with amino acids from both domains (see Figure 3a and Table 2). Such a pattern is expected for the amino acids that form a β-strand.

In the X-ray structure of the DPP III E451A-tynorphin complex, the substrate first and third amino acid residues form the backbone hydrogen bonds with N391 and G389, respectively, while in the *c*WT-RRNA complex, the backbone hydrogen bonds are established just between the first Arg and N391, as indicated by yellow dashed lines in Figure 3e and 3a, respectively.

Table 2. Amino acids residues found within 4 Å of the substrate (Arg-Arg-2-naphthylamide, Leu-enkephalin, endomorphin-1 and endomorphin-2, RRNA, L-EN, EN-1 and EN-2, respectively) in the structure of *c*WT-substrate complex obtained after 30 ns MD simulations followed by 5000 cycles of geometry optimization. P1, P2, P1', P2' and P3' are substrate residues/groups which interact with corresponding enzyme subsites (S1, S2, S1', S2' and S3')[30]. Residues that according to MM-GBSA calculations have a significant impact (± 1 kcal mol^{-1}) to the substrate stabilization (see Table S2) are show in italic. Underlined residues according to MM-GBSA calculations destabilize the DPP III-RRNA complex.

*c*WT-RRNA	**P2 Arg**	**P1 Arg**	**P1' NA***		
	S2	**S1**	**S1'**		
	E316	R319	*P387*		
	S317	*Y318*	*A388*		
	Y318	*E327*	*G389*		
	N391	*E329*	A416		
	I392	*I390*	F443		
	N394	E451	V447		
	D396	E508	*H568*		
	R399		R669		
	H455				
	W495				
	D496				
	E507				
	K629				
	P2	**P1**	**P1'**	**P2'**	**P3'**
*c*WT-L-EN	**Tyr**	**Gly**	**Gly**	**Phe**	**Leu**
	S2	**S1**	**S1'**	**S2'**	**S3'**
	E316	*Y318*	*G389*	*F109*	A416
	N391	*I390*		*Y318*	K439
	N394	*E451*		*P387*	*F443*
	I392	E508		*A388*	N545
	P393	*H568*		Q566	M546
	D396			*H568*	A549
	R399			*R572*	*R669*
	H455			*R669*	
	E507				
	S504				
	P2	**P1**	**P1'**	**P2'**	
*c*WT-EN1	**Tyr**	**Pro**	**Trp**	**Phe-NH$_2$**	
	S2	**S1**	**S1'**	**S2'**	
	E316	Y318	A388	F109	
	N391	I390	G389	G385	
	N394	E451	A416	I386	
	R399	E508	F443	P387	
	H455	H568	V447	Q566	
	W495		H450	H568	
	S504		E512	R572	
	E507		R572	R669	
	P2	**P1**	**P1'**	**P2'**	
*c*WT-EN2	**Tyr**	**Pro**	**Phe**	**Phe-NH$_2$**	
	S2	**S1**	**S1'**	**S2'**	
	E316	Y318	A388	F109	
	N391	I390	G389	I386	
	N394	E451	A416	P387	
	R399	E508	F443	Q566	
	H455		V447	R572	
	S504		H450	R669	
	E507		E512		
			H568		
			R572		

* NA is abbreviation of naphthylamide

Strong electrostatic stabilization of the positively charged RRNA guanidine groups and N-terminus with the negatively charged residues, mostly from the protein "lower" domain

(*i.e.* E316, E327, E329 and D396; Figure 3 and Table 2), is clearly demonstrated by visualization of the electrostatic potential surface calculated by APBS (Adaptive Poisson-Boltzmann Solver)[31] module implemented in program PyMol[29] (Figure 5). Apparently, the RRNA binding into the protein cleft with arginines oriented toward the negative region, next to the five-stranded β-core of the DPP III "lower" domain, is highly favorable. Additionally the naphthylamide group accommodated into the hydrophobic pocket is stabilized by numerous van der Waals interactions made by P387, A388, G389, I390, A416, F443 and V447 (see Fig. 3a, Fig.5 and Table 2) and CH-π interactions with H568 and R669. Since the bottom of the protein "upper" domain is mostly made of hydrophobic residues (regions colored white in Figure 5) the protein closure additionally aids to hydrophobicity of the enzyme S1' subsite. The negatively charged regions of the "upper" protein domain, comprised by E451, E507, E508 and D496, constitute the enzyme S1 and S2 subsites.

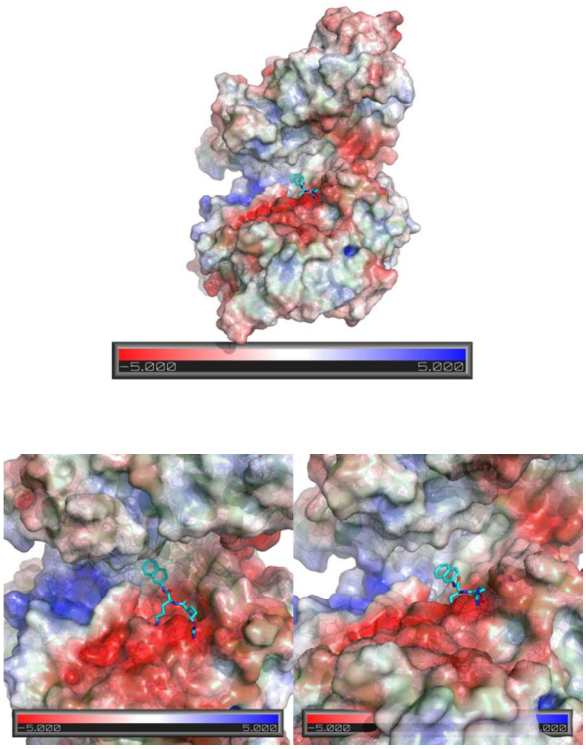


Figure 5. Electrostatic potential surface calculated by the APBS module as implemented in PyMol for the optimized structure of the *o*WT-RRNA complex obtained after 30 ns of MD simulations (different views are shown). The ion concentration, positive (radius 2.0 Å) and negative (radius 1.8 Å), was 0.05 M, and the protein and solvent (radius 1.4 Å) dielectric constants were set to 2 and 78, respectively. Blue and red protein surfaces correspond to potentials > 5 k_BT and < −5 k_BT, respectively. Substrate, RRNA, is shown in stick representation.

Simulations with RRNA molecule bound in the extended conformation showed that conserved, positively charged Arg-669 makes CH-π interactions with the RRNA naphthylamide group. Bezzera *el al.*[13] in their paper refer to the role of Arg-669 and Lys-670 in stabilization of the tynorphin C-terminus. They suggested that protein conformation change, described as protein closure, is a result of hydrogen bond breakage between Lys-670 and amino acid residues from the protein hinge region (protein residues 409-420) induced by substrate binding. However, by replacement of tynorphin with RRNA (*c*WT-RRNA complex, Figure 3a), this assumption is not any longer valid. The naphthylamide group of RRNA molecule makes CH-π interactions with Arg-669, but Lys-670 is too far to interact with it. On the other hand, according to the MD simulations of the ligand-free protein and the free energy calculated for different conformers[12], we proposed that the protein closing could occur even in the absence of a ligand. We assumed that ligand binding shifts equilibrium toward "closed" protein form, and supports closure in the final stage.

3.2 Substrate specificity explained by MD simulation of different DPP III complexes. Besides Arg-Arg-2-naphthylamide, DPP III was found to cleave the other substrates, synthetic, such are Phe-Arg-2-naphthylamide and Ala-Arg-2-naphthylamide (see Jajčanin-Jozić and Abramić[32]) and natural peptides, endomorphin-1 (Tyr-Pro-Trp-Phe-NH_2), endomorphin-2 (Tyr-Pro-Phe-Phe-NH_2), Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu) and tynorphin (Val-Val-Typ-Pro-Trp)[26]. In order to better understand this broad substrate specificity we simulated binding of these peptides to the DPP III "closed" form. The final structures are shown in Figure 3b-d.

The enthalpy and entropic contributions to the binding free energies calculated for Leu-enkephalin, endomorphin-1 and endomorphin-2 binding to the "closed X-ray" DPP III structure are given in Table 1, while the results of energy decomposition analysis performed by MM-GBSA method are given in Table S2. As indicated in Table 1, binding of the peptide substrates is less exergonic than binding of RRNA. These results agree with experimentally determined K_m values of 22.2 μM, 8.1 μM and 6.5 μM for RRNA, endomorphin-1 and Leu-enkephalin binding to the human DPP III, respectively[26] (resulting with the correlation coefficient of 0.99, see Supporting Inf.). However, caution is needed when comparing Michaelis (K_m) and dissociation constants (K_d) since the previous, besides substrate affinity, incorporates effects of transition state stabilization and product release rates. According to the free energy decomposition results (Table S2), all simulated peptide substrates bind into the same enzyme subsites as RRNA, *i.e.* they are stabilized by the same amino acid residues. However, differently from the electrostatically stabilized RRNA, Leu-enkephalin and endomorphins are mostly stabilized by van der Waales, CH-π and stacking interactions (see Table 2).

The most striking is different stabilization of the substrate P1 subsite. While the RRNA arginine is stabilized by numerous hydrogen bonds with the "lower" DPP III domain, the glycine/proline in the peptides P1 substrate is stabilized almost exclusively by van der Waals interaction with five-

stranded β-core of the same domain. Although the side chains of arginine and tyrosine, residues in the P2 subsite of RRNA and peptide substrates, respectively, interact with the same amino acids, the significantly higher stabilization energy of RRNA is due to large number of hydrogen bonds. At P1' subsite the large aromatic residues make numerous van der Waals and CH-π interactions with the protein hydrophobic pocket, while glycine in Leu-enkephalin, has only van der Waals interaction with G389 from the "lower" protein domain. All three peptide substrates have phenylalanine residue bound in the enzyme S2' subsite that is stabilized mostly by van der Waals, CH-π and by the stacking interactions with F109, P387, H568, R572 and R669, and frequently with Y318. Leu-enkephalin leucine in S3' position is stabilized by the hydrogen bonds between its C-terminus carboxylate group and the R669 side chain (see Figure 3b) and by van der Waales and CH-π interactions with the protein hydrophobic pocket. Similar interactions has tynorphin in the "closed X-ray" DPP III structure (see Figure 3e). The C-terminal carbonyl group in tetrapeptide endomorphin-1 makes the hydrogen bond with the R572 from the "upper" protein domain.

Similarly to the *c*WT-RRNA complex, Leu-enkephalin and endomorphins are evenly stabilized by the amino acid residues from the both protein domains, as well as with the central metal ion (see Table S2 and Table 2). Most of the hydrogen bonds present in DPP III complexes with peptide substrates originate from substrate binding in the antiparallel fashion to the five-stranded β-core of DPP III. Besides the "backbone" stabilization each arginine side chain of the preferred RRNA substrate makes in the *c*WT-RRNA complex additional 4-5 hydrogen bonds with the negatively charged amino acid residues from the "lower" protein domain (see Figure S5).

During MD simulations of the complexes DPP III remained close to its initial conformation (e.g. radgyr changed up to 5%, see radgyr profiles in Fig.S6), except in the case of *o*WT-RRNA complex where the protein closure was observed, first gentle (after ~10 ns of MD simulations at 300 K), and then more pronounced (during the last 10 ns of MD simulations at elevated temperature). For the binding free energy calculations the snapshots with the protein radgyr within 5% of the initial were used (the initial part of the MD trajectory).

Taking all this together, it seems the substrate specificity of the human DPP III is defined by two factors: a) the substrate ability to form the β-sheet secondary structure and b) its positively charged N-terminus. One must mention that the protein β-strand to which substrate molecule binds in an antiparallel fashion and amino acid residues participating in the N-terminus stabilization are part of the first and the second highly conserved DPP III family[33] regions clearly indicating their importance for the enzyme activity. It has been proven that different amino acids have different propensity for the β-strand conformation, wherein Val, Ile, Leu, Cys, Phe, Tyr, Trp, Thr and Met are the best candidates[34]. Indeed, the DPP III natural substrates like. endomorphin-1 (Tyr-Pro-Trp-Phe-NH_2), endomorphin-2 (Tyr-Pro-Phe-Phe-NH_2) and Leu-enkephalin (Tyr-Gly-Gly-Phe-Leu) are mostly made of these amino acid residues. We propose that a substrate molecule in order to bind in an active mode should have the amino acid residues that show propensity for the β-strand formation in the first three positions from the ligand N-terminus. Furthermore, our study revealed that the aromatic amino acid residues that have a tendency for β-strand conformation, like Tyr, Phe and Trp, at P2, P1' and P2' positions due to their possibility to establish strong van der Waals, CH-π and stacking interactions with the residues constituting S2, S1' and S2' enzyme subsites significantly influence a ligand binding affinity.

It is intriguing that endomorphins and Leu-enkephalin are DPP III substrates although they have proline/glycine residue, amino acids considered to be β-strand breaker, in second or/and third position from the substrate N-terminus. How to explain this in a light of the above made assumption on the DPP III substrate specificity? First, the amino acid residue at the 2nd position from the ligand N-terminus does not participate in the hydrogen bond network between the substrate β-strand and the five-stranded β-core of DPP III. Second, the secondary structure of a region is a consensus of the amino acids residues. For example a minimum of 4 amino acids out of 6 should show alpha preferences, or 3 out of 5 beta preferences, to set the secondary structure in a region. An individual misfit adopts the secondary structure of the neighbors.

Already in 1967 Ellis and Nuenke[3] showed that of dipeptidyl-2-naphthylamides, RRNA is by far the best DPP III substrate, while Ala-Ala-2-naphthylamide, Lys-Lys-2-naphthylamide, Gly-Arg-2-naphthylamide, Ser-Tyr-2-naphthylamide, Leu-Ala-2-naphthylamide, and Lys-Ala-2-naphthylamide are pure. Although Arg like the most of the other peptide building amino acids (Lys, Ala, Ser) shows rather low propensity for β-strand, arginines, in the *c*WT-RRNA complex, compensate this by numerous hydrogen bonds that they make with the enzyme binding site residues. As a result, RRNA remains, in an antiparallel fashion, connected to the five-stranded β-core of DPP III through all 80 ns of MD simulations (see Fig.S3.). While this is not so with other naphthylamide-dipeptides.

The substrate specificity model based on the β-strand amino acid propensities narrows the DPP III substrate specificity but also could explain the product expulsion from the enzyme binding site. Namely, after the peptide bond is broken, the β-strand stabilization is destroyed and the reaction products are released.

3.3 The zinc ion coordination. To investigate effects of the environment, protein and solvent, on the zinc ion coordination, we have performed the QM/MM energy minimizations of 18 different systems consisting of partly solvated DPP III with Zn^{2+} bound in the active site (see

Figure S2). The geometrical details of the zinc ion first coordination sphere together with the protonation state of its ligands, and their relative energies are given in Table 3 and Figure 6 and 7.

Consistently with the published data[16] and the results of the PDB search, the calculations revealed the high plasticity of the Zn^{2+} coordination. ONIOM calculations, with M06 functional, revealed that, in all three different DPP III structures ("open X-ray", "closed X-ray" and "closed MD" conformers), the octahedral coordination, accomplished by four amino acid residues (H450, E451, H455 and E508) and two water molecules, is the most stable.

In the "open", probably inactive (as suggested by Tomić *et al.*[10] and Bezzera *et al.*[13]), protein conformation, the structure with 5-coordinated zinc ion, where beside H450, H455, E508 and water molecule, the fifth ligand is either the E451 carboxyl oxygen or an additional water molecule, is according to the ONIOM calculations about 28 kcal/mol and 43 kcal/mol respectively, higher in energy than the 6-coordinated one (see Table S3).

Table 3. Energies[a] (in kcal mol^{-1}) of the partly solvated DPP III (52 water molecules) obtained by *2-layer* ONIOM optimization (M06 or B3LYP). Only the energies obtained for the same protein conformation are compared. Two different basis sets; 6-31+g(d,p) for the H, N, C and O atoms, and LANL2DZ-ECP for the Zn^{2+} atom, were used. The MM part was treated by the AMBER force field (param96). The energies calculated using electrostatic embedding protocols and within 20 kcal mol^{-1} of referent are shown. Numbers in square brackets indicate in which proportion E451, water molecule(s) and OH^- ion, respectively, coordinates the zinc ion.

initial structures		M06		B3LYP	
name	CN[E451:WAT:OH]	CN[E451:WAT:OH]	ΔE^{ONIOM}	CN[E451:WAT:OH]	ΔE^{ONIOM}
*c*WT	6[1:2:0]	6[1:2:0]	0.00	6[1:2:0]	2.38
	5[0:2:0]	6[1:2:0]	7.10	5[0:2:0]	3.71
	5[1:1:0]	5[1:1:0]	4.60	5[1:1:0]	0.00
	4[0:1:0]	4[0:0:1]	12.29	4[0:0:1]	8.85
	4[0:0:1]	4[0:0:1]	8.84	4[0:0:1]	5.57
cWT_{MD}	6[1:2:0]	6[1:2:0]	0.00		
	5[0:2:0]	6[1:2:0]	0.00		
	5[0:1:1]	5[0:2:0]	17.31		
	5[1:1:0]	5[1:1:0]	13.96		
	4[0:1:0]	6[1:2:0]	0.00		
	4[0:0:1]	5[1:1:0]	11.85		
*o*WT	6[1:2:0]	6[1:2:0]	3.78		
	4[0:0:1]	6[1:2:0]	0.00		

[a] $E^{ONIOM}=E_{MM}(S)+E_{QM}(SM)-E_{MM}(SM)$

However, the energy difference between six- and five-coordinated zinc ion structures is much smaller in the case of "closed", both "X-ray" and "MD", protein conformations (results obtained by both, B3LYP and M06, functional), than in the "open" one. For the "closed MD" protein conformation this difference, calculated utilizing M06 functional, is around 12 kcal/mol, in the case when E451coordinates the zinc ion, and 17.31 kcal/mol, when one additional water molecule coordinates the metal ion. Even smaller energy difference of 4.60 kcal/mol between the 5- and 6-coordinated Zn^{2+} was obtained for the "closed X-ray" DPP III structure with M06 functional (Table 3).

The "open X-ray", human and yeast, DPP III structures represent the only experimental data on the zinc ion coordination in the DPP III active site. Although they suggest that the zinc ion in the enzyme active site is tetra coordinated, by H450, H455, E508 and one water molecule (with one additional water molecule at 2.63 Å from the zinc ion in the human ortholog) as shown in Figure 1, ONIOM calculations predicted possibility of the tetrahedral zinc coordination only in the active site of the "closed X-ray" DPP III structure. Energy of this 4-coordinated metal ion is ~9 kcal/mol higher than those of the 6-coordinated metal ion in the same protein environment. In this particular structure the fourth zinc ligand is the hydroxide ion formed from the water molecule deprotonated by E451.

The subsequential optimization (of the *c*WT structure) with B3LYP revealed even higher plasticity of the Zn^{2+} coordination sphere with 5-coordinated metal ion, 5[1:1:0], representing energetically most favorable geometry. However the 6[1:2:0], and 4[0:0:1] coordinated Zn^{2+} geometries have only about 2.4 and 5.6 kcal mol^{-1}, respectively, higher energy. Further on, according to the B3LYP results, the 5[0:2:0] type of coordination is also probable, with the energy of 3.71 kcal mol^{-1} above the lowest one.

Coordination 6[1:2:0] has octahedral geometry in the "open X-ray" and the "closed MD" DPP III, while in the "closed X-ray" (obtained using both, M06 and B3LYP, functionals) this is slightly distorted octahedral coordination. Geometry of 5[1:1:0] and 5[0:2:0] coordinations is distorted trigonal bypiramidal in all the protein forms. 4[0:0:1] Zn^{2+} coordination type have tetrahedral geometry (B3LYP functional) or slightly distorted tetrahedral (M06 functional).

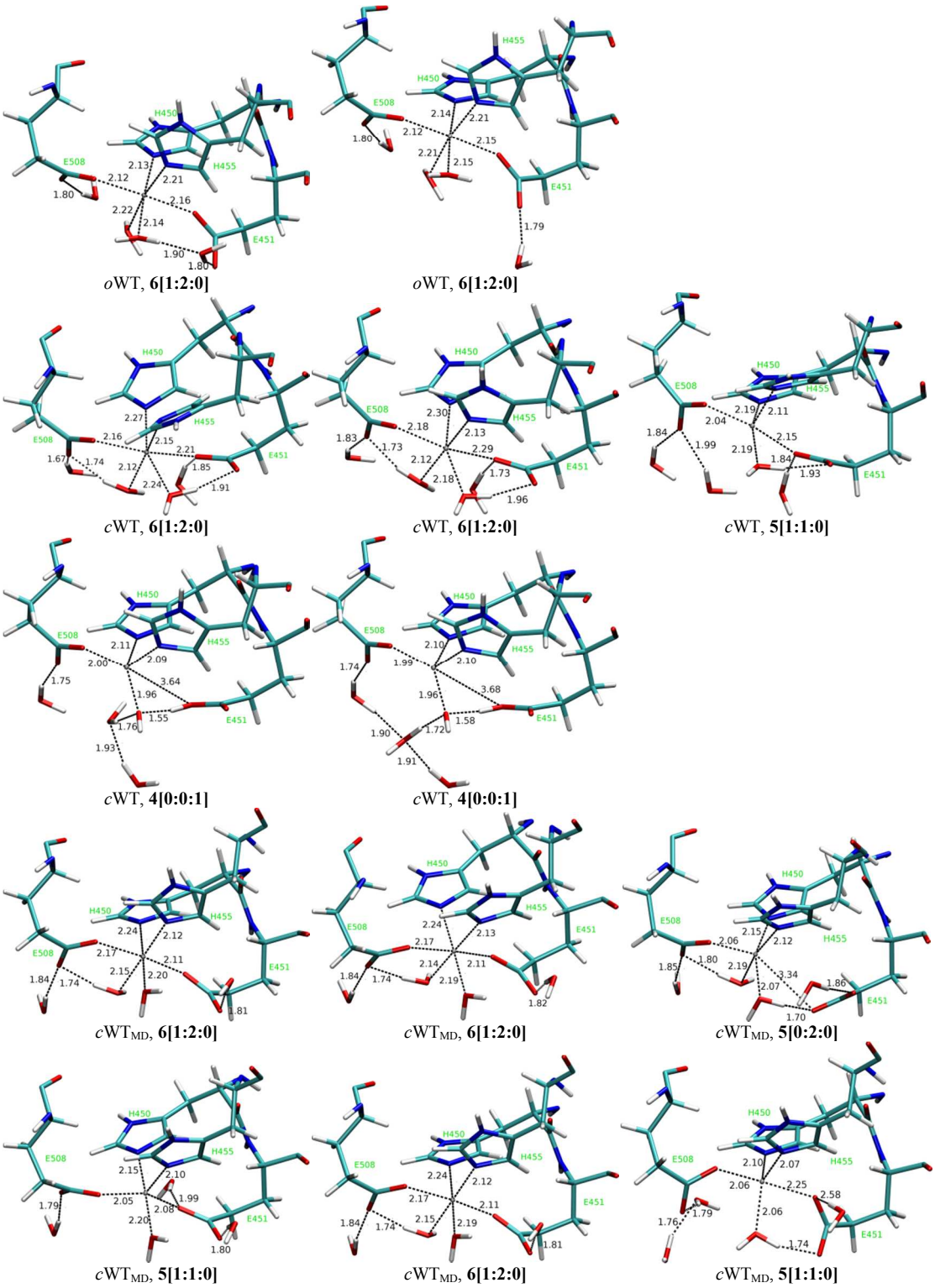


Figure 6. Coordination spheres around the catalytic zinc ion as extracted from the QM/MM (M06) optimized partly solvated (52 water molecules) DPP III structures. The prefixes *o*WT, *c*WT and *c*WT$_{MD}$ refer to different enzyme conformation, *i.e.* "open x-ray", "closed x-ray" and "closed MD", respectively. The structures with ΔE^{ONIOM} within 20 kcal mol^{-1} of the lowest one are shown. Distances are in Å.

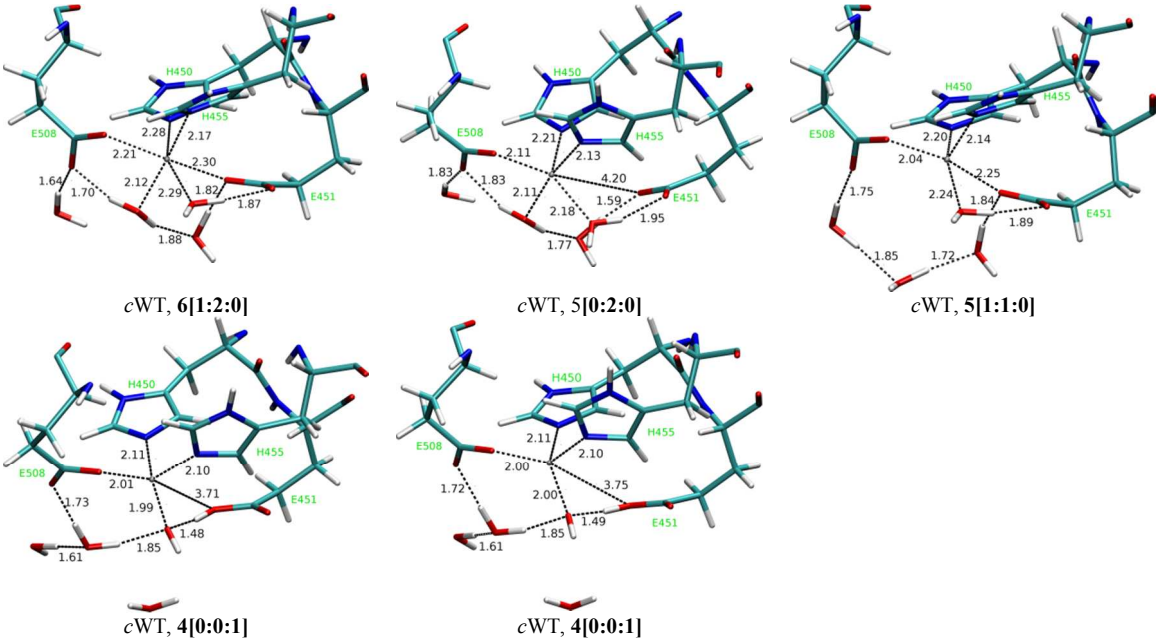


Figure 7. The zinc ion coordination spheres extracted from the QM/MM (B3LYP) optimized partly solvated (52 water molecules) DPP III structures. Prefix *c*WT refers to "closed X-ray" protein conformation. Structure with ΔE^{ONIOM} within 20 kcal mol^{-1} of the lowest one are shown. Distances are in Å.

QM and QM/MM calculations on the thermolysin catalyzed peptide hydrolysis revealed the five[35] and exchange of five and four[36] coordinated zinc ion, respectively, during the reaction as well as activation of the Zn-bound water molecules by Glu. If we assume that the proposed similarity between the catalytic mechanisms of DPP III and thermolysin[36] is correct, we should search for the equivalent coordination of the zinc ion in the DPP III. Among the zinc ion coordination geometries that we have obtained by the QM/MM calculations, the hydrogen bond between the metal coordinated water, either neutral or deprotonated, and E451 is found only in the "closed" DPP III structures. Formation of such hydrogen bond, but also the pronounced plasticity of Zn^{2+} coordination, energetically is the most feasible in the "closed X-ray" structure. In this structure the 6[1:2:0], 5[1:1:0] and 4[0:0:1] zinc coordination geometries have relative energies of 0.0, 4.6 and ~9 kcal mol^{-1}, respectively, according to the M06 calculations. Wherein B3LYP calculations revealed even lower relative energies of 0.0, 2.38, 3.71 and 5.57 kcal mol^{-1} for the 5[1:1:0], 6[1:2:0], 5[0:2:0] and 4[0:0:1] zinc coordination geometries, respectively. Based on the proposed similarity of the DPP III catalytic mechanism with those of thermolysin (glutamate assisted water addition mechanism)[8], and our present findings on the zinc ion coordination, as well as the ligand binding study, we assume that DPP III would be active when it is in the most compact form, *i.e.* in "closed X-ray" DPP III structure.

4 Conclusions

The present study provides detail insight into the zinc ion coordination and substrates binding in the three different human DPP III conformations: two obtained experimentally and one computationally. According to our computational studies the both compact forms "closed MD" and "closed X-ray" were stable during the MD simulations[37]. However, the binding affinities calculated for the preferred DPP III substrate (Arg-Arg-2NA) revealed its stronger binding to the "closed X-ray" than to the "closed MD" form, indicating that the first one might be the active form. This assumption is in agreement with the correlation between the zinc ion coordination flexibility and the protein compactness indicated by the QM/MM calculations. Namely, according to the proposed similarity between catalytic mechanisms of DPP III and of thermolysin[36] the Zn ion should be either four or five coordinated during the reaction. The QM/MM calculations showed that these types of the zinc coordination, as well as transition between them, are the most feasible in the "closed X-ray" structure, see Table 3 and Figures 6 and 7. Furthermore, the experimentally determined importance of E451 for the enzymatic activity of DPPII[5] is in line with the proposed similarity of the catalytic mechanisms since the equivalent glutamate in thermolysin participate in proton transfer. To have such a role E451 should be hydrogen bonded to the water molecule coordinating (5- or 4-coordinated) Zn^{2+}. The particular constellation (rearrangement) of the zinc ion environment was determined only in the "closed", low energy (the both "closed MD" and "closed X-ray") DPP III forms.

The MD and PBSA calculations suggested that substrates preferably bind in a form of β-strand, antiparallelly to the five-stranded β-core of DPP III, wherein binding of peptides (endomorphins 1 and 2 and Leu-enkephalin) is less exergonic than binding of RRNA, in agreement with the experimental measurements. The substrate binding study

revealed possible constrains on substrate specificity of human DPP III: a substrate ability to form the β-strand secondary structure and to have the positively charged N-terminus. Ligands that have amino acids residues with high propensity for β-strand at the first four positions (P2, P1, P1' and P2') from the N-terminus, should be able to bind in the DPP III binding site. Those with the bulky, non-polar residues (like Phe, Tyr and Trp) at P2 and P1'-P3' positions would be additionally preferred.

Acknowledgements

The work has been supported in part by Croatian Science Foundation under the project "7235 Flexibility, activity and structure correlations in the dipeptidyl peptidase III family", the Croatian National Grid Infrastructure (CRO NGI, http://www.cro-ngi.hr/) and Alexander von Humboldt foundation (project name: „Study of plant enzymes from metallopeptidase families M20 and M49"). For critically reading the manuscript the authors also thank Dr. Marija Abramić.

Notes and references

[a] Division of Physical Chemistry, Ruđer Bošković Institute, Bijenička 54, 10 000 Zagreb, Croatia. E-mail: atomic@irb.hr.

[b] Division of Physical Chemistry, Ruđer Bošković Institute, Bijenička 54, 10 000 Zagreb, Croatia. E-mail: sanja.tomic@irb.hr.

† Electronic Supplementary Information (ESI) available: additional tables and figures of the MD, MM-GBSA and QM/MM calculations can be found in the supporting information. See DOI: 10.1039/b000000x/

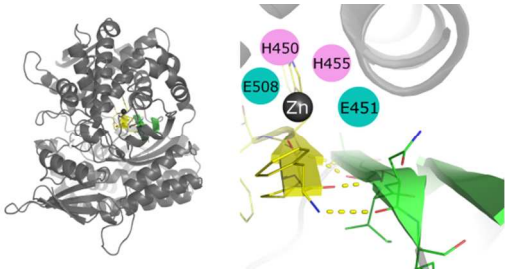
1. M. Abramić, M. Zubanović, and L. Vitale, *Biological chemistry Hoppe-Seyler*, 1988, **369**, 29–38.
2. J. M. Chen and A. J. Barrett, in *Handbook of Proteolytic Enzymes*, eds. A. J. Barrett, N. D. Rawlings, and J. F. Woessner, Elsevier, Academic Press, London, 2004, pp. 809–812.
3. S. Ellis and J. M. Nuenke, *The Journal of biological chemistry*, 1967, **242**, 4623–4629.
4. K. Fukasawa, K. M. Fukasawa, M. Kanai, S. Fujii, J. Hirose, and M. Harada, *The Biochemical journal*, 1998, **329**, 275–82.
5. K. Fukasawa, K. M. Fukasawa, H. Iwamoto, J. Hirose, and M. Harada, *Biochemistry*, 1999, **38**, 8299–303.
6. 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.
7. Y. Liu, J. T. Kern, J. R. Walker, J. A. Johnson, P. G. Schultz, and H. Luesch, *Proceedings of the National Academy of Sciences of the United States of America*, 2007, **104**, 5205–10.
8. P. K. Baral, N. Jajcanin-Jozić, S. Deller, P. Macheroux, M. Abramić, and K. Gruber, *The Journal of biological chemistry*, 2008, **283**, 22316–24.
9. A. Tomić, M. Abramić, J. Spoljarić, D. Agić, D. M. Smith, and S. Tomić, *Journal of molecular recognition : JMR*, 2011, **24**, 804–14.
10. A. Tomić, M. González, and S. Tomić, *Journal of chemical information and modeling*, 2012, **52**, 1583–94.
11. D. P. Martin, Z. S. Hann, and S. M. Cohen, *Inorganic chemistry*, 2013, **52**, 12207–15.
12. I. Bertini, A. Sigel, and H. Sigel, *Handbook on metalloprotein*, CRC Press, New York, USA, 1 edition., 2001.
13. G. A. Bezerra, E. Dobrovetsky, R. Viertlmayr, A. Dong, A. Binter, M. Abramic, P. Macheroux, S. Dhe-Paganon, and K. Gruber, *Proceedings of the National Academy of Sciences of the United States of America*, 2012, **109**, 6525–30.
14. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, *Journal of computational chemistry*, 2003, **24**, 1999–2012.
15. B. Bertosa, B. Kojić-Prodić, R. C. Wade, and S. Tomić, *Biophysical journal*, 2008, **94**, 27–37.
16. I. Dokmanić, M. Sikić, and S. Tomić, *Acta crystallographica. Section D, Biological crystallography*, 2008, **64**, 257–63.
17. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *The Journal of Chemical Physics*, 1995, **103**, 8577.
18. T. Darden, D. York, and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089.
19. S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma, and M. J. Frisch, *Journal of Molecular Structure: THEOCHEM*, 1999, **461-462**, 1–21.
20. T. Vreven, K. Morokuma, O. Farkas, H. B. Schlegel, and M. J. Frisch, *Journal of computational chemistry*, 2003, **24**, 760–9.
21. Y. Zhao and D. G. Truhlar, *Theoretical Chemistry Accounts*, 2007, **120**, 215–241.
22. A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648.
23. A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 1372.
24. P. J. Hay and W. R. Wadt, *The Journal of Chemical Physics*, 1985, **82**, 299.
25. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. a. Kollman, *Journal of the American Chemical Society*, 1995, **117**, 5179–5197.
26. M. Barsun, N. Jajcanin, B. Vukelić, J. Spoljarić, and M. Abramić, *Biological chemistry*, 2007, **388**, 343–8.
27. J. M. J. Swanson, R. H. Henchman, and J. A. McCammon, *Biophysical journal*, 2004, **86**, 67–74.
28. M. L. Connolly, *Journal of Applied Crystallography*, 1983, **16**, 548–558.
29. L. Schrödinger, 2010.
30. I. Schechter and A. Berger, *Biochemical and biophysical research communications*, 1967, **27**, 157–62.
31. N. a Baker, D. Sept, S. Joseph, M. J. Holst, and J. a McCammon, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 10037–41.
32. N. Jajčanin-Jozić and M. Abramić, *Biological chemistry*, 2013, **394**, 767–71.
33. M. Abramić, J. Špoljarić, and Š. Šimaga, *Periodicum biologorum*, 2004, **106**, 161–168.
34. K. Fujiwara, H. Toda, and M. Ikeguchi, *BMC structural biology*, 2012, **12**, 18.
35. V. Pelmenschikov, M. R. A. Blomberg, and P. E. M. Siegbahn, *Journal of biological inorganic chemistry : JBIC : a publication*

of the Society of Biological Inorganic Chemistry, 2002, **7**, 284–98.
36. J. Blumberger, G. Lamoureux, and M. L. Klein, *Journal of Chemical Theory and Computation*, 2007, **3**, 1837–1850.
37. A. Tomić, R. C. Wade, and S. Tomić, *in preparation*.

for Table of Contents use only

Hunting the human DPP III active conformation: combined thermodynamic and QM/MM calculations

Antonija Tomić, Sanja Tomić



It was found that the DPP III active conformation is the compact one. The substrate should be able to form the β-strand secondary structure.