

# Synthesis, QSAR, and Molecular Dynamics Simulation of Amidino-substituted Benzimidazoles as Dipeptidyl Peptidase III Inhibitors

**Vesna Rastija,<sup>1,\*</sup> Dejan Agić,<sup>1</sup> Sanja Tomić,<sup>2</sup> Sonja Nikolić,<sup>2</sup> Marijana Hranjec,<sup>3</sup> Grace Karminski-Zamola<sup>3</sup> and Marija Abramic<sup>2</sup>**

<sup>1</sup> Josip Juraj Strossmayer University of Osijek, Faculty of Agriculture in Osijek, Petra Svačića 1d, 31 000 Osijek, Croatia

<sup>2</sup> The Ruđer Bošković Institute, P.O. Box 180, Zagreb, 10 002, Croatia

<sup>3</sup> University of Zagreb, Faculty of Chemical Engineering and Technology, P.O. Box 177, Zagreb, 10 000, Croatia

\* Corresponding author: E-mail: vrastija@pfos.hr.  
Phone: +385 31 554 903 (V. Rastija).

Received: 09-04-2015

## Abstract

A molecular modeling study has been performed on a series of 16 benzimidazole-based inhibitors of human dipeptidyl peptidase III (DPP III). Eight of these were novel compounds synthesized in excellent yields using green chemistry approach. This study aims at elucidating the structural features of benzimidazole derivatives required for the antagonism of human DPP III activity using Quantitative Structure-Activity Relationship (QSAR) analysis, and at understanding the mechanism of one of the most potent inhibitor bindings into the active site of this enzyme, namely by molecular dynamics (MD) simulations. The best model obtained includes S3K and RDF045m descriptors, which have explained 89.4% of inhibitory activity. The depicted moiety for the strong inhibition activity matches the structure of the most potent compound. MD simulation has revealed the importance of imidazolinyl and phenyl groups in the mechanism of binding into the active site of human DPP III.

**Keywords:** Molecular modeling; dipeptidyl peptidase III; benzimidazole derivatives

## 1. Introduction

Dipeptidyl peptidases (DPPs) are a subclass of exopeptidases including enzymes that cleave two amino acids from the amino-end of a peptide chain. Dipeptidyl peptidase III (DPP III; EC 3.4.14.4) is the sole member of the M49 family of metallopeptidases.<sup>1</sup> DPP III sequentially hydrolyses dipeptides from the N-terminal of various oligopeptides, but has restricted specificity on dipeptidyl arylamide substrates, preferring Arg-Arg-2-naphthylamide. The ligand-free crystal structure of yeast DPP III has revealed the unique protein fold and zinc-binding site, enabling an insight into the possible catalytic mechanism.<sup>2</sup> DPP III from *Saccharomyces cerevisiae* has a two-lobe structure with a wide cleft through the middle containing the catalytic zinc ion, which is the possible substrate-binding site. The recently resolved crystal structure of human

DPP III (PDB ID: 3FVY)<sup>3</sup> shows close resemblance to the yeast DPP III (PDB ID: 3CSK).<sup>2</sup> DPP III has been implicated in intracellular protein catabolism, oxidative stress response,<sup>4</sup> and various physiological processes, such as blood pressure and pain regulation mechanism,<sup>5,6</sup> related to the hydrolysis of bioactive peptides,<sup>1</sup> as well as in some pathological processes – the cataractogenesis<sup>7</sup> and progression of ovarian cancer.<sup>8</sup> Due to its biological significance, human DPP III is interesting as a potential drug target. Specific inhibitors have proven important for determining the mechanism of action and the physiological role of several mammalian metallopeptidases.<sup>9</sup> However, efforts are still rarely invested in designing and synthesizing the specific inhibitors of DPP III. It is known that inhibition by non-specific chelating and thiol reagents is a feature of DPP III originating from various species, and that opioid pentapeptide tynorphin is a potent inhibi-

tor of DPP III isolated from the monkey brain.<sup>10</sup> In addition, synthetic hemorphin-like pentapeptides have been shown to possess strong inhibitory activity towards the recombinant rat DPP III.<sup>11</sup> Since it has been reported that the peptidic inhibitors of DPP III are degraded in the human body by (amino) peptidases,<sup>10,12</sup> we have investigated a series of benzimidazole derivatives in order to find new and more potent non-peptidic inhibitors of human DPP III.<sup>13</sup>

To improve the inhibitory potency, a cyclobutane ring was included in the structure of diamidino-substituted dibenzimidazole. A structural feature important for the enhanced inhibitory activity of these compounds seemed to be the presence of cyclized amidino group coupled to benzimidazole moiety and the presence of an additional aromatic ring.<sup>13</sup>

Since the benzimidazole derivatives investigated in our previous study showed wide variation in their structures, our present study aims at elucidating the structural features of benzimidazole derivatives required for the antagonism of dipeptidyl peptidase III using the Quantitative Structure-Activity Relationship (QSAR) analysis. In order to understand the mechanism of the most potent inhibitor binding into the active site of human DPP III, we have performed molecular dynamic simulation.

## 2. Experiment

### 2. 1. General Method for Preparing Compounds 1–16

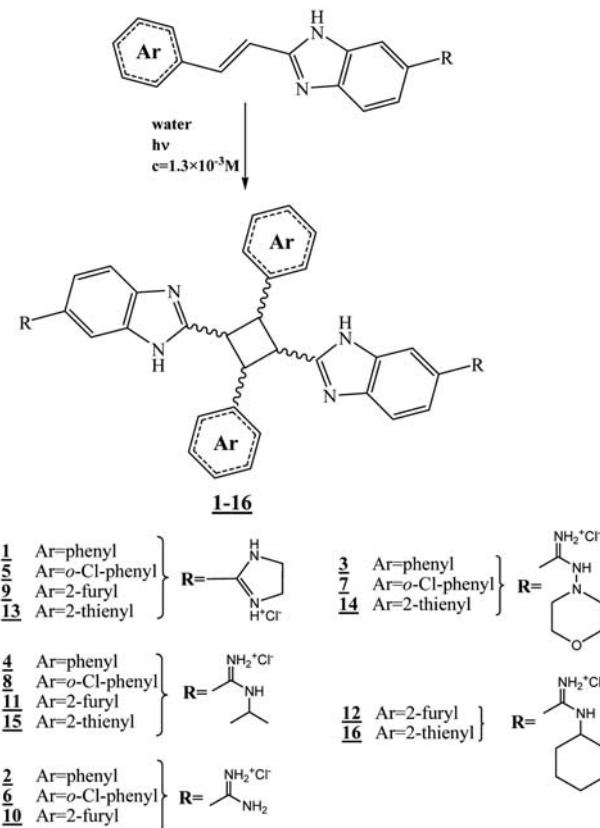
A solution ( $c = 1.3 \times 10^{-2}$  M) of corresponding *E*-2-styryl-5(6)-*N*-amidino-substituted-benzimidazole hydrochlorides or *E*-*N*-substituted-2-(2-thiophen/furyl-2-yl-vinyl)-3*H*-benzimidazole-5(6)-carboxamidine hydrochlorides in water was irradiated at room temperature with a 400 W high-pressure mercury lamp using a Pyrex filter for 2 h. The air was bubbled through the solution. The solution was concentrated, precipitated with acetone, and the resulting solid was filtered off. The product was dissolved in ethanol and precipitated with diethyl-ether, filtered off, and dried. The procedure was repeated two times until the product was analytically pure. The synthesis of compounds **1–2**, **4–6**, **8–9** and **13** had been previously published.<sup>13</sup>

The obtained cyclobutane derivatives are of the head-to-tail type, which was confirmed by mass spectrometry, namely by the presence of the same m/z fragment ion in mass spectrometry.

Melting points were determined on a Kofler hot-stage microscope and are uncorrected. IR spectra were recorded on the Nicolet magna 760, Perkin-Elmer 297, and Perkin-Elmer Spectrum 1 spectrophotometers, with KBr disks. <sup>1</sup>H and <sup>13</sup>C NMR spectra were recorded on the Varian Gemini 300, Bruker Avance DPX 300, and Bruker

Avance DRX 500 spectrometers using TMS as an internal standard in *DMSO-d*<sub>6</sub>. Mass spectra were recorded on the Agilent 1100 series LC/MSD Trap SL. Elemental analyses for carbon, hydrogen, and nitrogen were performed on a Perkin-Elmer 2400 elemental analyzer and a Perkin-Elmer, Series II, CHNS Analyzer 2400. Where analyses are indicated only as symbols of elements, analytical results obtained are within 0.4% of the theoretical value. In preparative photochemical experiments, irradiation was performed at room temperature with a water-cooled immersion well and an “Origin Hanau” 400-W high-pressure mercury arc lamp, using Pyrex glass as a filter.

All 2-benzimidazolyl-substituted cyclobutane derivatives **1–16** were synthesized according to Scheme 1 by reaction of photochemical [2+2] cycloaddition in water, starting from the corresponding, previously prepared 2-styrylbenzimidazoles.



Scheme 1. Synthetic procedure of a series of benzimidazole derivatives.

The corresponding *N*-amidino-substituted *E*-2-styryl-1*H*-benzimidazoles and *E*-*N*-substituted-2-(2-thiophen/furyl-2-yl-vinyl)-3*H*-benzimidazole-5(6)-carboxamidines were prepared by the condensation reaction of *o*-substituted-3-phenyl-propenals or 3-thiophen(furan)-2-yl-propenals with 4-*N*-amidino-substituted 1,2-phenylenediamines and *p*-benzoquinone.<sup>14,15</sup> Cyclobutane derivatives appeared as a constant reaction mixture containing

two stereoisomers, with a very high amount of one isomer (> 98%). The mixture of isomers could not be separated either by column chromatography or by the HPLC/MS method. According to the mass spectra, which showed small molecular ion and high base peaks  $m/z$  M<sup>+2</sup>/2, we assigned them as head-to-tail adducts. The <sup>1</sup>H and <sup>13</sup>C NMR data confirmed the symmetrical cyclobutane structure. The direct elucidation of cyclobutane ring stereochemistry was very difficult due to the highly overlapped aryl and benzimidazole signals in <sup>1</sup>H NMR spectra. In comparison to the starting benzimidazole derivatives, the <sup>1</sup>H NMR spectra of cyclobutane derivatives showed the disappearance of signals related to the trans-ethylenic double bond, with the coupling constants of 16 Hz, while two signals for four cyclobutane protons appeared in the aliphatic part of the spectra in 1:1 ratio, which confirmed their symmetrical structure. The 2D NOESY NMR spectra, as well as HETCOR, confirmed the suggested stereochemistry of cyclobutane derivatives due to the NOE interactions between cyclobutane protons and the NH of benzimidazole nuclei, as well as those between cyclobutane protons and aryl protons.

**1,3-di-[5(6)-(N-4-morpholinylamidino)2-benzimidazolyl]-2,4-di-phenylcyclobutane dihydrochloride (3)** was prepared using the above-described method from *E*-N-Morpholin-4-yl-2-styryl-3H-benzimidazole-5(6)-carboxamidine hydrochloride (0.15 g; 0.39 mmol) after 2 hours of irradiation. 0.114 g (76%) of yellow powder obtained; mp = 259–261 °C.

MS  $m/z$ : 695 (M<sup>+1</sup>(-2HCl) 25%), 263 (M<sup>+2</sup>/2, 100%); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm) 13.20–13.09 (m, 2H, NH<sub>benzimid.</sub>), 11.20 (s, 2H, NH<sub>amid.</sub>), 9.65 (s, 2H, NH<sub>amid.</sub>), 9.02 (brs, 2H, NH<sub>amid.</sub>), 7.86 (s, 2H, H<sub>arom.</sub>), 7.81 (d, 2H, J = 8.70 Hz, H<sub>arom.</sub>), 7.68 (d, 2H, J = 8.66 Hz, H<sub>arom.</sub>), 7.64 (dd, 4H, J<sub>1</sub> = 7.86 Hz, J<sub>2</sub> = 1.46 Hz, H<sub>arom.</sub>), 7.43–7.28 (m, 6H, H<sub>arom.</sub>), 5.46 (d, 2H, J<sub>2</sub> = 7.88 Hz, H<sub>cyclobut.</sub>), 5.19 (d, 2H, J<sub>2</sub> = 7.94 Hz, H<sub>cyclobut.</sub>); <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm): 162.23 (2C), 156.74 (2C), 142.45 (2C), 139.10 (2C), 138.10 (2C), 130.01 (2C), 128.69 (2C), 127.85 (4C), 127.41 (2C), 126.33 (2C), 120.93 (2C), 118.36 (2C), 118.32 (2C), 65.49 (4C), 53.83 (4C), 44.37 (2C), 41.41 (2C); Anal. Calcd for (C<sub>40</sub>H<sub>44</sub>Cl<sub>2</sub>N<sub>10</sub>O<sub>2</sub>): C 62.58, 5.78, N 18.24; Found: C 62.73, H 5.40, N 18.56.

**1,3-di-[5(6)-(N-4-morpholinylamidino)2-benzimidazolyl]-2,4-di-2-chlorophenyl-cyclobutane dihydrochloride (7)** was prepared using the above-described method from *E*-2-[2-(2-chloro-phenyl)-vinyl]-N-morpholin-4-yl-3H-benzimidazole-5(6)-carboxamidine hydrochloride (0.15 g; 0.36 mmol) after 2 hours of irradiation. 0.112 g (75%) of light violet powder obtained; mp = 267–269 °C. MS  $m/z$ : 763 (M<sup>+1</sup>(-2HCl) 30%), 383 (M<sup>+2</sup>/2, 100%); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm) 13.20 (brs, 2H, NH<sub>benzimid.</sub>), 11.24 (s, 2H, NH<sub>amid.</sub>), 9.80 (s, 2H, NH<sub>amid.</sub>), 9.10 (bs, 2H,

NH<sub>amid.</sub>), 7.93 (s, 2H, H<sub>arom.</sub>), 7.84 (d, 2H, J = 8.64 Hz, H<sub>arom.</sub>), 7.72 (d, 2H, J = 8.68 Hz, H<sub>arom.</sub>), 7.64 (t, 2H, J = 7.80 Hz, H<sub>arom.</sub>), 7.48–7.33 (m, 6H, H<sub>arom.</sub>), 5.20 (t, 2H, J = 7.15 Hz, H<sub>cyclobut.</sub>), 4.99 (t, 2H, J = 7.10 Hz, H<sub>cyclobut.</sub>); <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm): 162.78 (2C), 157.00 (2C), 142.74 (2C), 139.18 (2C), 138.22 (2C), 134.65 (2C), 130.07 (2C), 128.81 (2C), 127.89 (4C), 127.45 (2C), 126.36 (2C), 120.93 (2C), 118.65 (2C), 68.77 (4C), 63.89 (4C), 44.64 (2C), 41.77 (2C); Anal. Calcd for (C<sub>40</sub>H<sub>42</sub>Cl<sub>2</sub>N<sub>10</sub>O<sub>2</sub>): C 57.40, 5.06, N 16.74; Found: C 57.65, H 5.38, N 6.51.

**1,3-di-[5(6)-N-amidino-2-benzimidazolyl]-2,4-di-(2-furyl)cyclobutane dihydrochloride (10)** was prepared using the above-described method from *E*-2-(2-furan-2-yl-vinyl)-3H-benzimidazole-5-carboxamidine hydrochloride (0.09 g; 0.31 mmol) after 2 hours of irradiation. 0.080 g (80%) of grey powder obtained; mp = >300 °C. MS  $m/z$ : 505 (M<sup>+1</sup>(-2HCl) 85%), 253 (M<sup>+2</sup>/2, 100%); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm) 13.05 (brs, 2H, NH<sub>benzimid.</sub>), 9.30 (brs, 8H, NH<sub>amid.</sub>), 8.17 (s, 2H, H<sub>arom.</sub>), 7.81 (d, 2H, J = 8.66 Hz, H<sub>arom.</sub>), 7.64 (d, 2H, J = 8.60 Hz, H<sub>arom.</sub>), 7.30 (d, 2H, J = 3.10 Hz, H<sub>furyl</sub>), 6.16 (d, 4H, J = 3.00 Hz, H<sub>furyl</sub>), 5.11 (t, 2H, J = 8.10 Hz, H<sub>cyclobut.</sub>), 5.01 (t, 2H, J = 8.12 Hz, H<sub>cyclobut.</sub>); <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm): <sup>13</sup>C NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm): 166.77 (2C), 156.17 (2C), 153.14 (2C), 142.99 (2C), 142.68 (2C), 138.94 (2C), 122.13 (2C), 120.85 (2C), 119.56 (2C), 112.25 (2C), 110.83 (2C), 107.52 (2C), 45.22 (2C), 40.79 (2C); Anal. Calcd for (C<sub>28</sub>H<sub>26</sub>N<sub>8</sub>Cl<sub>2</sub>O<sub>2</sub>): C 58.24, H 4.54, N 19.40; Found: C 58.50, H 4.27, N 19.68.

**1,3-di-[5(6)-N-isopropylamidino-2-benzimidazolyl]-2,4-di-(2-furyl)cyclobutane dihydrochloride (11)** was prepared using the above-described method from *E*-N-isopropyl-2-(2-furan-2-yl-vinyl)-3H-benzimidazole-5-carboxamidine hydrochloride (0.10 g; 0.30 mmol) after 2 hours of irradiation. 0.090 g (90%) of grey powder obtained; mp = 274–276 °C. MS  $m/z$ : 589 (M<sup>+1</sup>(-2HCl) 15%), 253 (M<sup>+2</sup>/2, 100%); <sup>1</sup>H NMR (DMSO-d<sub>6</sub>) ( $\delta$  /ppm) 13.10 (brs, 2H, NH<sub>benzimid.</sub>), 9.40 (brs, 4H, NH<sub>amid.</sub>), 9.10 (brs, 2H, NH<sub>amid.</sub>), 7.95 (s, 2H, H<sub>arom.</sub>), 7.76 (d, 2H, J = 8.60 Hz, H<sub>arom.</sub>), 7.62 (d, 2H, J = 8.56 Hz, H<sub>arom.</sub>), 7.34 (d, 2H, J = 3.00 Hz, H<sub>furyl</sub>), 6.24 (d, 4H, J = 3.05 Hz, H<sub>furyl</sub>), 5.15 (t, 2H, J = 8.14 Hz, H<sub>cyclobut.</sub>), 5.09 (t, 2H, J = 8.16 Hz, H<sub>cyclobut.</sub>), 4.15–4.12 (m, 2H, CH), 1.27 (d, 12H, J = 6.26 Hz, 4CH<sub>3</sub>); 162.32 (2C), 155.72 (2C), 152.55 (2C), 142.28 (2C), 134.78 (2C), 122.24 (2C), 121.81 (2C), 120.99 (2C), 119.56 (2C), 110.41 (2C), 107.33 (2C), 107.19 (2C), 45.01 (2C), 40.78 (2C), 38.28 (2C), 21.33 (4C); Anal. Calcd for (C<sub>34</sub>H<sub>38</sub>N<sub>8</sub>Cl<sub>2</sub>O<sub>2</sub>): C 61.72, H 5.79, N 16.94; Found: C 61.88, H 5.99, N 16.77.

**1,3-di-[5(6)-N-cyclohexylamidino-2-benzimidazolyl]-2,4-di-(2-furyl)cyclobutane dihydrochloride (12)** was prepared using the above-described method from *E*-N-

cyclohexyl-2-(2-furan-2-yl-vinyl)-3H-benzimidazole-5-carboxamidine hydrochloride (0.100 g; 0.27 mmol) after 2 hours of irradiation. 0.070 g (70%) of grey powder obtained; mp = 283–285 °C; MS *m/z*: 669 ( $M^{+1}$ (–2HCl) 89%), 335 ( $M^{+2}/2$ , 100%);  $^1\text{H}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm) 13.20 (brs, 2H, NH<sub>benzimid.</sub>), 9.62 (s, 2H, NH<sub>amid.</sub>), 9.49 (s, 2H, NH<sub>amid.</sub>), 9.23 (s, 2H, NH<sub>amid.</sub>), 8.06 (s, 2H, H<sub>arom.</sub>), 7.80 (d, 2H,  $J$  = 8.40 Hz, H<sub>arom.</sub>), 7.67 (d, 2H,  $J$  = 8.90 Hz, H<sub>arom.</sub>), 7.33 (s, 2H, H<sub>arom.</sub>), 6.40 (d, 2H,  $J$  = 3.10 Hz, H<sub>furyl</sub>), 6.22 (d, 2H,  $J$  = 3.00 Hz, H<sub>furyl</sub>), 5.27 (t, 2H,  $J$  = 8.20 Hz, H<sub>cyclobut.</sub>), 5.05 (t, 2H,  $J$  = 8.30 Hz, H<sub>cyclobut.</sub>), 4.10–4.03 (m, 2H, CH), 1.99–1.70 (m, 10H, 5CH<sub>2</sub>); 1.59–1.40 (m, 10H, 5CH<sub>2</sub>);  $^{13}\text{C}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm): 162.88 (2C), 155.16 (2C), 151.96 (2C), 151.27 (2C), 151.14 (2C), 143.15 (2C), 125.47 (2C), 124.71 (2C), 117.52 (2C), 113.09 (2C), 110.96 (2C), 108.52 (2C), 52.55 (2C), 40.36 (2C), 38.50 (2C), 31.55 (4C), 25.22 (2C), 23.50 (2C); Anal. Calcd for (C<sub>34</sub>H<sub>38</sub>N<sub>8</sub>Cl<sub>2</sub>S<sub>2</sub>): C 58.86, H 5.52, N 16.15; Found: C 58.60, H 5.26, N 16.38.

**1,3-di-[5(6)-N-4-morpholinylamidino-2-benzoimidazolyl]-2,4-di-(2-thienyl)cyclobutane dihydrochloride (14)** was prepared using the above-described method from *E*-*N*-Morpholin-4-yl-2-(2-thiophen-2-yl-vinyl)-3H-benzimidazole-5(6)-carboxamidine hydrochloride (0.120 g; 0.31 mmol) after 3 hours of irradiation. 0.090 g (75%) of grey powder obtained; mp = 271–274 °C. MS *m/z*: 707 ( $M^{+1}$ (–2HCl) 15%), 354 ( $M^{+2}/2$ , 100%);  $^1\text{H}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm)

13.20 (brs, 2H, NH<sub>benzimid.</sub>), 11.14 (bs, 2H, NH<sub>amid.</sub>), 9.67 (bs, 2H, NH<sub>amid.</sub>), 8.99 (bs, 2H, NH<sub>amid.</sub>), 8.14 (s, 2H, H<sub>arom.</sub>), 7.72 (s, 2H, H<sub>arom.</sub>), 7.45 (d, 2H,  $J$  = 8.78 Hz, H<sub>arom.</sub>), 7.21 (d, 2H,  $J$  = 5.00 Hz, H<sub>thienyl</sub>), 7.11 (d, 2H,  $J$  = 8.72 Hz, H<sub>arom.</sub>), 6.81 (d, 2H,  $J$  = 5.04 Hz, H<sub>thienyl</sub>), 5.25 (t, 2H,  $J$  = 7.80 Hz, H<sub>cyclobut.</sub>), 4.78 (t, 2H,  $J$  = 7.80 Hz, H<sub>cyclobut.</sub>), 3.78 (s, 8H, 4CH<sub>2</sub>), 2.95 (s, 8H, 4CH<sub>2</sub>);  $^{13}\text{C}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm): 162.88 (2C), 156.00 (2C), 142.98 (2C), 142.19 (2C), 138.72 (2C), 134.37 (2C), 128.30 (2C), 127.30 (2C), 127.18 (2C), 119.74 (2C), 116.10 (2C), 113.90 (2C), 68.02 (4C), 54.34 (4C), 45.55 (2C), 42.80 (2C); Anal. Calcd for (C<sub>36</sub>H<sub>40</sub>N<sub>10</sub>Cl<sub>2</sub>S<sub>2</sub>O<sub>2</sub>): C 55.45, H 5.17, N 17.96; Found: C 55.25, H 5.41, N 17.69.

**1,3-di-[5(6)-N-isopropylamidino-2-benzoimidazolyl]-2,4-di-(2-thienyl)cyclobutane dihydrochloride (15)** was prepared using the above-described method from *E*-*N*-isopropyl-2-(2-thiophen-2-yl-vinyl)-3H-benzimidazole-5(6)-carboxamidine hydrochloride (0.135 g; 0.39 mmol) after 3 hours of irradiation. 0.110 g (81%) of grey powder obtained; mp = 273–276 °C. MS *m/z*: 621 ( $M^{+1}$ (–2HCl) 10%), 311 ( $M^{+2}/2$ , 100%);  $^1\text{H}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm) 13.20 (brs, 2H, NH<sub>benzimid.</sub>), 9.50 (brs, 4H, NH<sub>amid.</sub>), 9.02 (s, 2H, NH<sub>amid.</sub>), 8.03 (s, 2H, H<sub>arom.</sub>), 7.95 (d, 2H,  $J$  = 8.85 Hz, H<sub>arom.</sub>), 7.50 (d, 2H,  $J$  = 8.80 Hz, H<sub>arom.</sub>), 7.20 (d, 2H,  $J$  = 5.20 Hz, H<sub>thienyl</sub>), 7.10–6.79 (d, 2H,  $J$  = 4.80 Hz,

H<sub>thienyl</sub>), 5.23 (t, 2H,  $J$  = 7.40 Hz, H<sub>cyclobut.</sub>), 4.78 (t, 2H,  $J$  = 7.60 Hz, H<sub>cyclobut.</sub>), 4.11–4.06 (m, 2H, CH), 1.30 (d, 12H,  $J$  = 6.20 Hz, (CH<sub>3</sub>)<sub>2</sub>);  $^{13}\text{C}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm): 161.89 (2C), 155.78 (2C), 147.14 (2C), 142.34 (2C), 135.01 (2C), 127.99 (2C), 126.04 (2C), 125.34 (2C), 122.55 (2C), 119.71 (2C), 118.94 (2C), 112.06 (2C), 44.26 (2C), 30.42 (4C), 29.27 (4C); Anal. Calcd for (C<sub>34</sub>H<sub>38</sub>N<sub>8</sub>Cl<sub>2</sub>S<sub>2</sub>): C 58.86, H 5.52, N 16.15; Found: C 58.60, H 5.26, N 16.38.

**1,3-di-[5(6)-N-cyclohexylamidino-2-benzoimidazolyl]-2,4-di-(2-thienyl)cyclobutane dihydrochloride (16)** was prepared using the above-described method from *E*-*N*-cyclohexyl-2-(2-thiophen-2-yl-vinyl)-3H-benzimidazole-5(6)-carboxamidine hydrochloride (0.100 g; 0.26 mmol) after 3 hours of irradiation. 0.075 g (75%) of grey powder obtained; mp = 261–264 °C. MS *m/z*: 669 ( $M^{+1}$ (–2HCl) 75%), 335 ( $M^{+2}/2$ , 100%);  $^1\text{H}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm) 13.25 (brs, 2H, NH<sub>benzimid.</sub>), 9.60 (brs, 4H, NH<sub>amid.</sub>), 9.16 (s, 2H, NH<sub>amid.</sub>), 8.06 (s, 2H, H<sub>arom.</sub>), 7.84–7.50 (d, 2H,  $J$  = 8.70 Hz, H<sub>arom.</sub>), 7.36 (d, 2H,  $J$  = 8.76 Hz, H<sub>arom.</sub>), 7.19 (d, 2H,  $J$  = 4.95 Hz, H<sub>thienyl</sub>), 6.80 (t, 2H,  $J$  = 4.80 Hz, H<sub>thienyl</sub>), 5.23 (t, 2H,  $J$  = 7.80 Hz, H<sub>cyclobut.</sub>), 4.76 (t, 2H,  $J$  = 7.70 Hz, H<sub>cyclobut.</sub>), 3.77–3.72 (m, 2H, CH), 1.94–1.75 (m, 10H, 5CH<sub>2</sub>), 1.60–1.51 (m, 10H, 5CH<sub>2</sub>);  $^{13}\text{C}$  NMR (DMSO-*d*<sub>6</sub>) ( $\delta$  /ppm): 162.80 (2C), 155.77 (2C), 146.83 (2C), 142.19 (2C), 134.42 (2C), 127.90 (2C), 125.77 (2C), 125.30 (2C), 122.38 (2C), 119.59 (2C), 118.90 (2C), 111.96 (2C), 52.26 (2C), 45.60 (2C), 42.80 (2C), 31.64 (4C), 25.22 (2C), 24.76 (4C); Anal. Calcd for (C<sub>40</sub>H<sub>46</sub>N<sub>8</sub>Cl<sub>2</sub>S<sub>2</sub>): C 62.08, H 5.99, N 14.48; Found: C 62.37, H 6.21, N 14.67.

## 2. 2. Biochemical Methods

DPP III was purified to homogeneity from human erythrocytes as described earlier.<sup>16</sup> The inhibitory activity of benzimidazole derivatives towards human DPP III was assayed in 50 mM Tris-HCl buffer, pH 7.4. In brief, enzyme (2.7 nM) was incubated with an inhibitor (10  $\mu\text{M}$ ) for 10 min at 25 °C, followed by 5 min at 37 °C. The enzymatic reaction was started with Arg-Arg-2NA (40  $\mu\text{M}$ ) as a substrate, and after the 15 min incubation at 37 °C in a water bath, the reaction was stopped and the absorbance was measured using the spectrophotometric method described above.<sup>17</sup> The percentage of enzyme inhibition (%) inh.) was calculated by comparing the enzymatic activity without (normal activity) and with an inhibitor (inhibited activity), using the following formula:

$$\% \text{ inh.} = [(\text{normal activity} - \text{inhibited activity}) / (\text{normal activity})] \times 100\% \quad (1)$$

## 2. 3. Computational Methods

*Quantitative Structure-Activity Relationship (QSAR) analysis*

### 2.3.1. Descriptor Calculation

The 3D structure of 16 benzimidazole derivatives (Table 1) were optimized applying the HyperChem 7.0 (HyperCube, Inc., Gainesville, FL), using the molecular mechanics force field (MM+).<sup>18</sup> Subsequently, all structures were submitted to geometry optimization using the semi-empirical AM1 method.<sup>19</sup> Molecular structures were optimized using the Polak-Ribiere algorithm until the root mean square gradient was 0.1 kcal mol<sup>-1</sup>.

After geometry optimization, several physical and chemical parameters were calculated with HyperChem: the energy of the highest occupied molecular orbital ( $E_{\text{HOMO}}$ ), the energy of the lowest unoccupied molecular orbital ( $E_{\text{LUMO}}$ ), the difference between  $E_{\text{HOMO}}$  and  $E_{\text{LUMO}}$  (GAP), the heat of formation ( $H_f$ ), the hydration energy ( $E_{\text{HYDR}}$ ), and the volume ( $V$ ) of the molecule. The 2D and 3D molecular descriptors used in this study have been calculated by applying the online software Parameter Client (Virtual Computational Chemistry Laboratory, <http://146.107.217.178/lab/pclient/>), an electronic remote version of the Dragon program.<sup>20</sup> These descriptors may be grouped according to their dimensionality:

1. Zero-dimensional (0D): constitutional descriptors;
2. One-dimensional (1D): charge, molecular descriptors;
3. Two-dimensional (2D): topological, topological charge, 2D autocorrelation, BCUT, eigenvalue-based, and E-state descriptors;
4. Three-dimensional (3D): radic molecular profiles, geometrical, RDF (Radial Distribution Function), Molecule Representation of Structures based on Electron diffraction (3D-MoRSE), WHIM (WeigHted Covariance Matrices), and GETAWAY (Geometry, Topology, and Atom Weights AssemblY) descriptors.

### 2.3.2. Regression Analysis

The selection of descriptors based on the best-subset method and the multiple regression analysis (MLR) was performed with the use of STATISTICA 12 (StatSoft, Inc. Tulsa, USA, 2013). The number of descriptors ( $I$ ) in the multiple regression equation was limited to two due to the small number of compounds. The initial number of 1370 calculated molecular indices and physicochemical properties was reduced to 121 descriptors using the following procedure:

The initial set of 1370 descriptors was divided into 7 groups. In every group, descriptors in which values were degenerated were eliminated. A further selection of descriptors was performed by using the best-subset method. The best-subset regression is a model-building technique that finds the subset of predictor variables that best predicts the response on a dependent variable by linear regression. The criterion for the best model is based on the  $R^2$  values of the obtained models. The descriptor included

in the ten best two-variables and in the ten best one-variable models was selected for further analysis. In the final space of 121 selected descriptors, the best models were found by using the best-subset method.

The best models were selected based on the squared correlation coefficient ( $R^2$ ), the adjusted standard deviation of regression ( $S$ ), and the Fisher ratio values ( $F$ ). In order to avoid overfitting, the terminal selection of models was based on an inter-correlation study between the variables included in the equation. Models with highly inter-correlated ( $|R| \geq 0.70$ ) descriptors were discarded.

### 2.3.3. Validation of QSAR Models

Models have been assessed by internal cross-validation, using the leave-one-out (LOO) method and Y-scrambling, performed by QSARINS.<sup>21</sup> The compound number is insufficient for external validation. Internal validation included the following criteria: the coefficient of determination ( $R^2$ ), adjusted ( $R^2_{\text{adj}}$ ) coefficient, leave-one-out coefficient ( $Q^2_{\text{LOO}}$ ), the global correlation among the descriptors ( $K_{xx}$ ), the difference between the global correlation between the molecular descriptors and  $y$  as the response variable, the global correlation among the descriptors ( $\Delta K$ ), and the Fisher ratio ( $F$ ).

Internal validations also included the following parameters: root-mean-square error ( $RMSE$ ); root-mean-square error determined by using the cross-validated LOO method ( $RMSE_{cv}$ ), the concordance correlation coefficient ( $CCC$ ), the LOO cross-validation ( $CCC_{cv}$ ), the mean absolute error of the training set ( $MAE$ ), the internal validation set ( $MAE_{cv}$ ), and the predictive residual sum of squares determined by using the cross-validated LOO method ( $PRESS_{cv}$ ).

The robustness of QSAR models was tested by means of a Y-randomization test, where robust QSAR models should have low  $R^2_y$ scr and low  $Q^2_y$ scr values.<sup>22</sup>

### 2.3.4. Definition of the Applicability Domain of a Model

Investigation of the applicability domain of the prediction model was performed by means of leverage plot (plotting residuals vs. leverage of training compounds). Detection of outliers was carried out for compounds that had values of standardized residuals higher than two standard deviation units. The leverage  $h^*$  of a compound is the measure of its influence on the model and is defined as:

$$h^* = 3 \times p' / n \quad (2)$$

where  $n$  is the number of training compounds and  $p'$  is the number of model-adjustable parameters.

#### 2.3.4.1. Dynamics simulations

The three-dimensional structure of ligand-free human dipeptidyl peptidase III was obtained from the

Brookhaven Protein Databank<sup>23</sup> (ID: 3FVY). The residues of amino acids missing in the crystal structure of DPP III, Pro224 – Asp227, were added using the program *t leap*, a basic preparation program for Amber simulations available in the AMBER10 package (<http://ambermd.org>). In the model, all Arg and Lys residues are positively charged (+1) and all Glu and Asp residues are negatively charged (-1), as expected in physiological (experimental) conditions. The initial protonation of the protein was determined by the WHATIF-server (at pH 7.0; <http://swift.cmbi.ru.nl/servers/html/index.html>) and the protonation of histidines was checked according to their ability to form hydrogen bonds with the neighboring amino acid residues. Protein parametrization was performed within the AMBER ff03 force field of Duan et al.,<sup>24</sup> while the inhibitor parameters were derived using the Antechamber in the Amber10 program. For the zinc cation, Zn<sup>2+</sup>, parameters were derived from previous work<sup>25</sup> and modified according to a PDB survey.<sup>26</sup>

The ligand-free DPP III and the protein-inhibitor complex were placed in the truncated octahedron box filled with TIP3P water molecules, and Na<sup>+</sup> ions were added in order to neutralize the systems. The resulting system consisted of ~ 80 700 atoms (~ 69 000 atoms of water). Electrostatic interactions were calculated using the particle-mesh Ewald method. The same molecular dynamics (MD) simulation procedure was used as in our previous publication.<sup>27</sup> The structure of ligand-free DPP III obtained after 30 ns of MD simulations was selected as a starting point and a complex between the protein and the inhibitor was built. Prior to the MD simulations of the protein-inhibitor complex, protein geometry was optimized in four cycles with different constraints. In the first (300 steps) and second (300 steps) cycles, water and the inhibitor molecule were relaxed, while the protein and the zinc atom were constrained using a harmonic potential with a force constant of 133.98 kJ mol<sup>-1</sup> Å<sup>-2</sup>. In the third (300 steps) and fourth (300 steps) cycles, the same constant was applied to the zinc atom while only the protein backbone was constrained with 133.98 kJ mol<sup>-1</sup> Å<sup>-2</sup> and 41.87 kJ mol<sup>-1</sup> Å<sup>-2</sup>, respectively. The energy minimization procedure, consisting of 150 steps of steepest descent followed by the conjugate gradient algorithm for the remaining optimization steps, was the same in all four cycles. During the first period of equilibration (200 ps of gentle heating from 0 to 300 K), the NVT ensemble was used while all of the following simulations (water density adjustment and productive simulations) were performed at a constant temperature and pressure (300 K and 1 atm, the NPT ensemble). The temperature was held constant using the Langevin dynamics, with a collision frequency of 1 ps<sup>-1</sup>. Bonds involving hydrogen atoms were constrained using the SHAKE algorithm. The complex was simulated for 20 ns with a time step of 2 fs, and structures were sampled every 1 ps. After 20 ns of MD simulation, in order to enable a more efficient conformational space search, the

temperature of the protein-inhibitor complex was elevated to 500 K, followed by its cooling back to 300 K. Both the heating and cooling intervals were 300 ps long, and the time step was 1 fs. In between, the system was simulated 200 ps at 500 K. The rest of MD simulations (20 ns) were carried out using the NPT ensemble and a time step of 2 fs. Altogether, MD simulations resulted in a 40.8 ns long trajectory of the protein-inhibitor complex.

The radius of gyration (Rg), the root-mean square deviation (RMSD), and the root-mean square coordinate fluctuation (RMSF) of each residue in the system were calculated using the »ptraj« module of the AMBER10 program.<sup>28,29</sup>

### 3. Results and Discussion

#### 3. 1. Inhibitory Activity of Benzimidazole Derivatives Towards Human DPP III

Inhibitory activities of the analyzed compounds are shown in Table 1. Fifteen of the sixteen analyzed compounds at the concentration of 10 µM have shown inhibitory effects against human DPP III. The strongest inhibition (% inh. > 90) has been obtained with compounds **1**, **2**, **5** and **9**. A slightly lower inhibition has been observed for the compounds **6**, **8**, **10** and **13**, while others have shown an inhibition of less than 30%. We have reported earlier on that the inhibition of dipeptidyl peptidase IV (DPP IV) has not detected in this type of benzimidazole derivatives.<sup>13</sup>

**Table 1.** Inhibitory activities of benzimidazole derivatives against human DPP III determined experimentally (% inh.<sub>exp</sub>), the calculated inhibitory activities (% inh.<sub>calc</sub>) obtained using model III, the residuals, and the values of descriptors involved in the model.

Compound No.	S3K*	RDF045m**	% inh. <sub>exp</sub> ***	% inh. <sub>calc</sub>	Residuals
<b>1</b>	4.422	23.359	97.8	93.5	4.3
<b>2</b>	4.242	16.934	90.3	75.5	14.8
<b>3</b>	6.701	19.893	13.4	7.4	6.0
<b>4</b>	5.911	15.699	10.7	17.2	-6.5
<b>5</b>	4.801	29.106	97.2	102.7	-5.5
<b>6</b>	4.627	23.717	89.3	88.3	1.0
<b>7</b>	7.094	29.741	18.3	31.3	-13.0
<b>8</b>	6.304	30.541	69.4	59.7	9.7
<b>9</b>	4.003	15.225	94.2	76.8	17.4
<b>10</b>	3.8	15.537	67.6	84.5	-16.9
<b>11</b>	5.401	13.957	19.4	27.1	-7.7
<b>12</b>	6.245	18.962	14.7	18.6	-3.9
<b>13</b>	4.21	13.836	51.9	65.0	-13.1
<b>14</b>	6.446	19.952	0	15.8	-15.8
<b>15</b>	6.253	16.514	20.8	9.2	11.6
<b>16</b>	6.496	19.46	29.9	12.4	17.5

\* S3K = topological descriptor, 3-path Kier alpha-modified shape index \*\* RDF045m = RDF (Radial Distribution Function) group of descriptors, \*\*\* at the 10 µmol L<sup>-1</sup> concentration of the compound

### 3. 2. Results of the Multiple Regression Analysis (MLR)

To elucidate the structural features of the experimentally examined benzimidazole derivatives required for the antagonism of human DPP III activity, several MLR models were built after the variable selection carried out by the genetic algorithm. The models were initially developed by using the all-subset-procedure, until two variable models achieved the best statistical parameters. Considering the small data set of 16 compounds, it was not split into training and test. Models were validated by means of internal cross-validation (LOO) and Y-scrambling. The best QSAR model with one descriptor ( $I = 1$ ) was:

#### Model I

$$\% \text{ inh.} = 84.031 (\pm 7.075) - 31.094 (\pm 6.592) N\text{-}067$$

The best two QSAR models with two descriptors ( $I = 2$ ) generated are:

#### Model II

$$\% \text{ inh.} = 243.246 (\pm 48.043) - 64.435 (\pm 16.863)$$

$$Mor24p - 6.725 (\pm 0.950) F08[C\text{-}N]$$

#### Model III

$$\begin{aligned} \% \text{ inh.} = & 148.991 (\pm 17.812) - 32.159 (\pm 3.170) S3K \\ & + 3.713 (\pm 0.628) RDF045m \end{aligned}$$

Statistical parameters of the obtained models are given in Table 2.

A comparison of various statistical parameters given in Table 2 reveals that the performance of model III, which includes two descriptors,  $S3K$  and  $RDF045m$ , is better than the other two models. The fitting power of all the obtained models is acceptable as it satisfies the following criteria:  $R^2 > 0.6$ ;  $F \geq 0.70$ . The coefficient of determination ( $R^2$ ) is the amount of dependent variable variance explained by the regression mode.

The low co-linearity of the selected molecular descriptors in models II and III has been verified by the low  $K_{xx}$  values and  $\Delta K \geq 0.05$ . The internal predictivity of the obtained models has been verified by cross-validation, using the Leave-One-Out (LOO) procedure. The cross-validated explained variance ( $Q^2_{\text{loo}}$ ) is a measure of the validity of internal predictivity.<sup>22</sup>

The fitting concordance correlation coefficient ( $CCC_{\text{tr}}$ ) and the LOO correlation coefficient ( $CCC_{\text{cv}}$ ) are acceptable in all models since they are higher than 0.85, but the highest values have been identified in model III.  $RMSE_{\text{tr}}$ ,  $MAE_{\text{tr}}$ ,  $RMSE_{\text{cv}}$  and  $MAE_{\text{cv}}$  should be as low as possible, and in model III, the values of these parameters are lower than in the rest of the models.

Randomization testing is a technique for checking the robustness of a QSAR model. It is expected that the new QSAR models resulting from a Y-randomization

**Table 2.** Statistical parameters of the obtained QSAR models.

Statistical parameters	Model I	Model II	Model III
$R^2$	0.756	0.860	0.894
$R^2_{\text{adj}}$	0.739	0.838	0.878
$R^2 - R^2_{\text{adj}}$	0.017	0.022	0.016
$K_{xx}$	0.000	0.212	0.344
$\Delta K$	0.870	0.346	0.134
$RMSE_{\text{tr}}$	17.510	13.283	11.553
$MAE_{\text{tr}}$	13.020	11.323	10.295
$CCC_{\text{tr}}$	0.861	0.925	0.944
$s$	18.719	14.737	12.817
$F$	43.459	39.855	54.780
$Q^2_{\text{loo}}$	0.683	0.795	0.838
$RMSE_{\text{cv}}$	19.975	16.057	14.293
$MAE_{\text{cv}}$	14.900	13.792	12.759
$CCC_{\text{cv}}$	0.820	0.891	0.915
$R^2_{\text{y,scr}}$	0.067	0.135	0.130
$Q^2_{\text{y,scr}}$	-0.220	-0.322	-0.329

test should have low  $R^2_{\text{y,scr}}$  and low  $Q^2_{\text{y,scr}}$  values. This is a strong proof that the proposed model is well-founded, rather than a result of chance correlation. Low  $R^2_{\text{y,scr}}$  and  $Q^2_{\text{y,scr}}$  values of the reported models prove that they were not obtained by chance correlation.

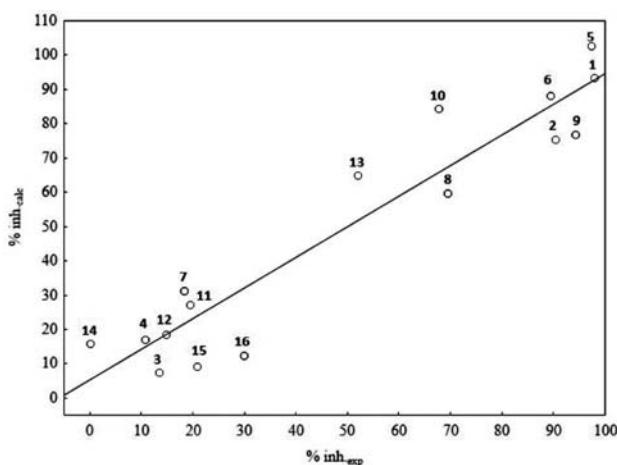
Comparing the statistical parameters of all the reported models, it is evident that model III has obtained the best fitting and predictive power.

Model II includes the two-dimensional descriptor (2D)  $F08[C\text{-}N]$  and the three-dimensional descriptor (3D)  $Mor24p$ . The best model obtained (model III) was using  $S3K$  and  $RDF045m$ , which explains 89.4% of the inhibitory activity. The inhibitory activities of benzimidazole derivatives against human DPP III have been determined experimentally. The calculated inhibitory activities obtained using model III, the residuals, and the values of  $S3K$  and  $RDF045m$  descriptors are given in Table 1. A scatter plot of the experimentally obtained percentages of enzyme inhibition versus values calculated by model III is shown in Fig. 1.

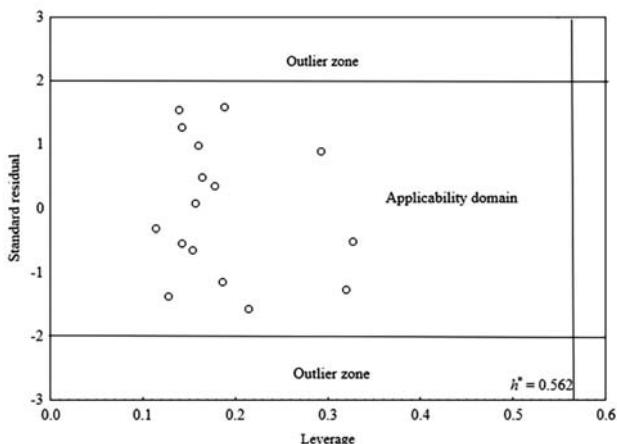
Two groups of clustered compounds are clearly visible in Fig. 1. Highly active compounds are clustered into the upper right part of the graph (compounds **1**, **5**, **6**, **2**, **9**), while the less active ones are clustered into the lower left part of the graph (compounds **3**, **4**, **7**, **11**, **12**, **14**, **15**, **16**).

In order to investigate the applicability of prediction model 3 and to detect the possible outliers, the leverage of the training set was plotted against the residuals (Fig. 2).

As it can be seen from the plot, there are no compounds outside the domain of applicability of the model, since their leverage values are not greater than the warning leverage ( $h^* = 0.308$ ). Moreover, no analyzed compounds were considered as outliers because their standardized residuals were not greater than  $\pm 2$ .



**Fig. 1.** A scatter plot of experimentally obtained enzyme inhibition percentages (% inh.<sub>exp</sub>) versus values calculated by model III (% inh.<sub>calc</sub>).



**Fig. 2.** Applicability domain of the QSAR model for% inh. calculated by model III.QSAR models interpretation

The best one-variable model has been obtained by descriptor *N*-067, which belongs to the Ghose-Crippen atom-centered fragments molecular descriptors.<sup>30</sup> The given descriptor presents a number of secondary amino groups bonded by two aliphatic carbon or nitrogen atoms. In the studied benzimidazole derivatives, these amino groups are present in *R* substituents of compounds **3**, **4**, **7**, **8**, **11**, **12**, **14**, **15**, and **16**. According to model I, an increased number of these groups in a molecule decreases the inhibition of benzimidazoles.

The bi-parametric model II, containing the *F08[C-N]* and *Mor24p* descriptors, gives good results in the prediction of inhibitory activity. *F08[C-N]* belongs to the 2D frequency fingerprint group of descriptors. *F08[C-N]* is defined by the presence (1) or absence (0) of specific atom pairs at different topological distances (from 1 to 10).<sup>31</sup> The negative coefficient of *F08[C-N]* in model 2 implies that a higher frequency of presence of C and N atoms at

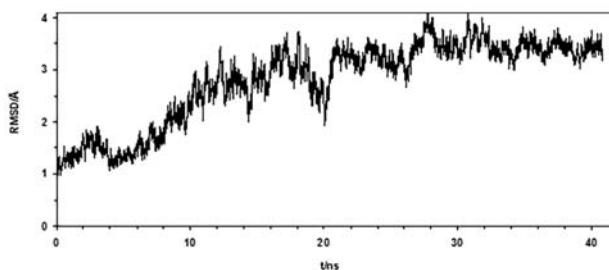
the topological distance 8 decreases the inhibitory activities. The second variable in model II, *Mor24p*, belongs to the 3D molecular representation of structures based on the electron diffraction code and weighted by atomic polarizabilities. The descriptors reflect the three-dimensional arrangement of atoms in a molecule and are not in correlation with chemical bonds.<sup>32</sup> Judging from the negative coefficient of the given descriptor in model 2, an increased number of atoms with higher polarizabilities (O, N, Cl) at the interatomic distance of 24 Å is related to weaker inhibition.

The best multiple linear regression model, model III, includes two descriptors, *RDF045m* and *S3K*. *RDF045m*, which belongs to the RDF (Radial Distribution Function) group of descriptors, and offers information about the three-dimensional distribution of polarizabilities and mass in molecules calculated at the radius of 4.5 Å from the geometrical center of each molecule.<sup>33</sup> It is evident from their positive coefficient in model 3 that, for higher inhibitory activities, the benzimidazole derivatives should have the atom mass (O, N, Cl) at the mentioned radius as a tendency. The second descriptor in model III, *S3K*, is a topological descriptor that belongs to the Kier shape indices. This group of indices is defined in terms of the number of graph vertices (*N*) and the number of paths <sup>m</sup>*P* with the length *m* (*m* = 1, 2, 3) in the H-depleted molecular graph, and therefore encodes the information on the centrality of branching. *S3K* is a Kier alpha-modified shape index that takes into consideration the contribution of covalent radii and hybridization states (*sp*<sup>3</sup>) to the shape of the molecule.<sup>34</sup> Model III reveals that the lower value of *S3K* improves the inhibition activity of compounds. Therefore, the benzimidazole derivatives with highly branched substituents that possess *sp*<sup>3</sup>-hybridized nitrogen atoms (**4**, **11**, **12**, **15**), especially in combination with *sp*<sup>3</sup>-hybridized oxygen atoms (**3**, **7**, **14**), have low inhibitory activities.

The results of QSAR analyses elucidate the reasons of strong inhibition activity of compounds **1** and **5**. Both compounds contain two imidazolinyl groups, with only one *sp*<sup>3</sup>-hybridized nitrogen atom each. Compound **5** has slightly lower inhibition potency than compound **1** because of the presence of chloro-substituted phenyl groups instead of phenyl groups on a cyclobutane ring.

### 3. 3. Results of the Molecular Dynamics Simulation of the DPP III Inhibitor Complex

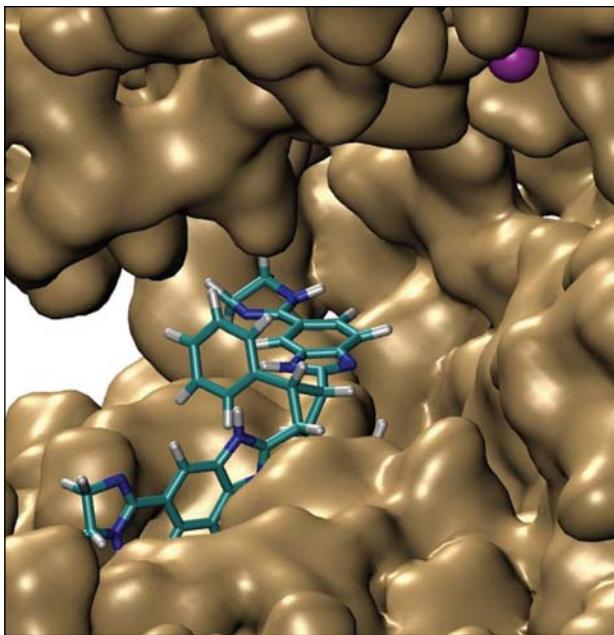
In order to investigate the influence of a strong inhibitor (**1**) on the structure of human DPP III and to better understand the mechanism of its inhibitory activity, we have performed a molecular modeling study. For this purpose, we have prepared a model of human DPP III in complex with compound **1** and performed a 40.8 ns long molecular dynamics (MD) simulation. According to the



**Fig. 3.** Root Mean Square Deviations (RMSD) of the DPP III obtained during 40.8 ns of MD simulation, in complex with inhibitor **1**.

RMSD profile (Fig. 3), the protein structure stabilized after about 22 ns of MD simulation (almost immediately after the short-time heating).

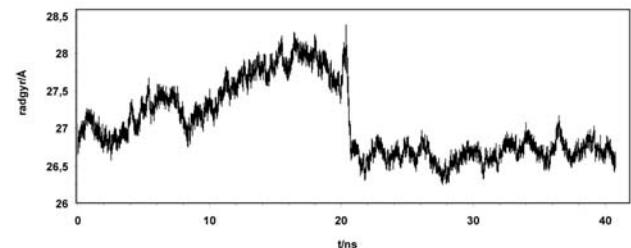
Immediately after the system optimization, compound **1** settled in the area of the central enzyme cleft, at the distance of ~ 16 Å from the catalytic Zn(II) ion (measured from the cyclobutane ring) (Fig. 4).



**Fig. 4.** Surface representation of a segment of the wide DPP III cleft, with inhibitor **1** in the optimized structure of the complex. The Zn<sup>2+</sup> is presented as a magenta sphere and the inhibitor is shown in stick representation.

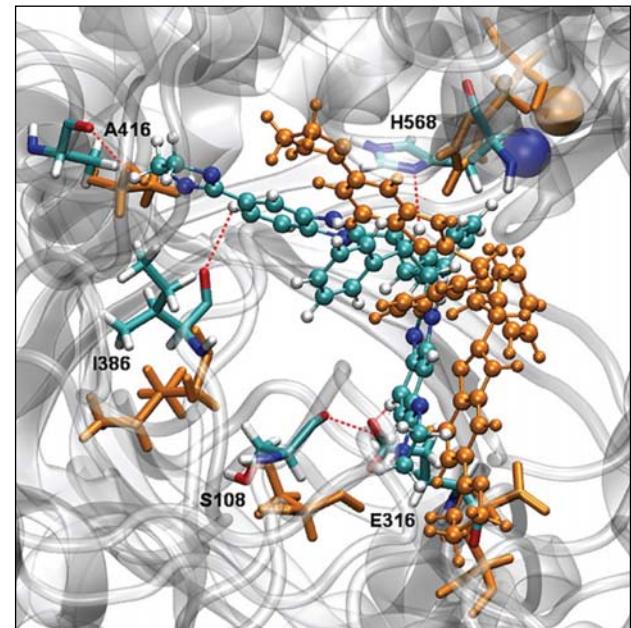
In this position, the inhibitor interacts by means of electrostatic and van der Waals forces with the residues of the “lower” domain (residues Ser108, Phe109, Asp111, Leu161, Tyr318, Glu327) and the “upper” domain (residues Phe556 and Arg669). However, whereas during the simulation of ligand-free protein<sup>35</sup> the domains progressively approached each other and the Rg decreased (the initial value was about 28.5 Å and the final one about 26.5

Å), during the MD simulation of the complex, the Rg was increasing during the first 20 ns. After short-time heating, the protein domains experienced a significant shift, accompanied by a drastic increase in protein compactness (the Rg decreased for about 1 Å and reached its minimum of 26.5 Å) (Fig. 5).



**Fig. 5.** Protein backbone radius of gyration obtained during 40.8 ns of productive MD simulation of DPP III in complex with the inhibitor.

As a result of the domains motion, several hydrogen bonds were formed between the inhibitor and the residue Ser108, Glu316, Ile386, Ala416, and His565 (Fig. 6). Apparently, the inhibitor reorientation sped up the protein “closure” (Fig. S1).



**Fig. 6.** Overlay of inhibitor **1** orientation in the DPP III binding site before (orange) and after short-time heating (colored by the atom type). The five red dotted lines represent the hydrogen bonds between the inhibitor and the residue Ser108, Glu316, Ile386, Ala416, and His565. Zn<sup>2+</sup> is shown as a sphere.

The RMSDs of the “lower” (4-335, 374-416, 670-726) and “upper” (336-373, 417-669) protein domains also increased during the heating period. However, du-

ring the remaining 20 ns they remained unchanged (Fig. S2). Apparently, during simulations at an increased temperature, the protein and the ligand finely adapt to each other.

Besides the tertiary structure changes, the protein's secondary structure also changed during the simulation. A portion of  $\alpha$ -helices reduced for ~8%; however, the contribution of  $\beta$ -sheets remained nearly unchanged during the MD simulation (Table 3). Similar structural changes were observed during the 101 ns long simulations of ligand-free human DPPIII.<sup>34</sup>

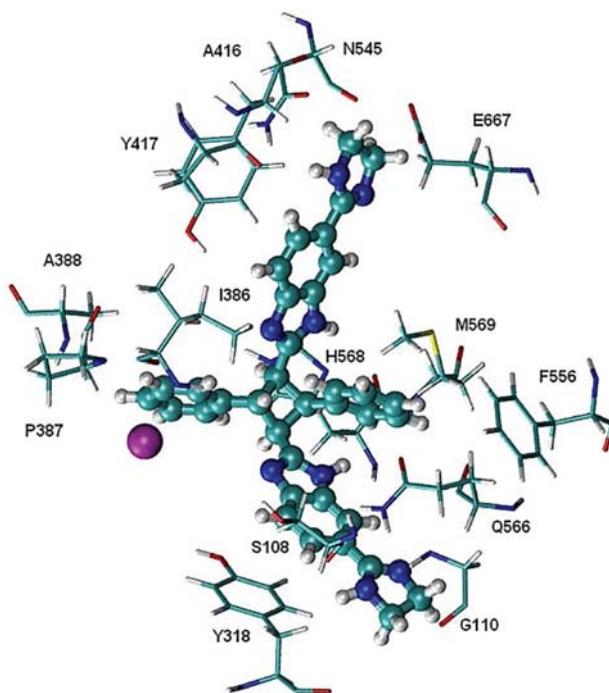
**Table 3.** Secondary structure elements calculated using »2Struc: the secondary structure server« (<http://2struc.cryst.bbk.ac.uk/>).<sup>36</sup>

Structure	Secondary structure elements		
	Helix (%)	Sheet (%)	Other (%)
After the optimization steps	49.0	13.4	37.6
After 20 ns	45.1	13.3	41.6
After heating to 500 K	39.3	12.3	48.4
After 40.8 ns	41.4	13.0	45.6

During the MD simulation, the zinc cation,  $Zn^{2+}$ , was mostly six-coordinated by two histidines (His450 and His455), two glutamates (Glu451 and Glu508), and two water molecules, as also determined for the ligand-free DPPIII<sup>27</sup>. These results imply that the binding of inhibitor **1** has no significant effect on the zinc ion coordination.

An analysis of protein-ligand interactions during the last ns of the simulation (Fig. 7) revealed that compound **1** interacted with DPP III mostly electrostatically. The strongest interactions took place between imidazolinyl groups and the amino acid residues of the “lower” (Ser108, Gly110, Tyr318, and Ala416) and “upper” (Tyr417, Asn545, and Glu667) domains. Benzimidazole groups interacted with Tyr318 from the “lower” domain residue and with Tyr417 from the “upper” one. The interaction between the phenyl groups from compound **1** and with Ile386, Pro387, Ala388, Phe556, Gln566, and Met569 was a combination of van der Waals and electrostatic interactions, as well as of the cyclobutane ring and His568.

Although the binding of compound **1** did not disturb the coordination geometry of the catalytic zinc ion, it engaged two amino acid residues (Tyr318 and His568), evolutionary conserved in the M49 family, which are crucial for the DPP III activity. Interestingly, the binding mode and the multiple interactions established during the MD simulations of this potent benzimidazole-based inhibitor in complex with DPP III, differ from those established for the reversible inhibitor Tyr-Phe-NHOH<sup>27</sup> and the opioide peptide tynorphin (Val-Val-Tyr-Pro-Trp)<sup>3</sup> (see Fig. 7. and compare with the Supplemental Fig. S3 and Fig. S4). Al-



**Fig. 7.** Orientation and interactions of inhibitor **1** in complex with human DPP III obtained after 40.8 ns of MD simulations. The zinc cation is shown as a magenta sphere, the inhibitor in ball-and-stick, and the amino acid residues in stick representation.

together, this could explain the experimentally observed irreversible (tight binding) inhibition of DPP III by compound **1**.<sup>13</sup>

#### 4. Conclusions

The purpose of this study was to investigate the inhibition effect of benzimidazole derivatives against human dipeptidyl peptidase III and to find the key features responsible for their inhibitory activity. The results of the QSAR analysis obtained by multiple linear regressions showed that, for higher inhibitory activities, benzimidazole derivatives should have atoms with a higher mass (O, N, Cl) at the radius of 4.5 Å from the geometrical centers of each molecule as a tendency. Also, an increased number of amino groups, the presence of C and N atoms at the topological distance 8, and the highly branched substituents that possess sp<sup>3</sup>-hybridized nitrogen atoms in combination with sp<sup>3</sup>-hybridized oxygen atoms, decrease the inhibitory activities of benzimidazole derivatives. The depicted moiety for strong inhibition activity matches the structure of compound **1**, which possesses two imidazolinyl and two phenyl groups as substituents.

The importance of imidazolinyl and phenyl groups in the mechanism of binding into the active site of human DPP III, in one of the most potent inhibitors, compound **1**, has been confirmed by molecular dynamic simulation.

Most of the interactions of imidazolinyl groups were achieved by electrostatic forces with the residuals Ser108, Gly110, Tyr318, and Ala416 of the “lower” domain, and with the residuals Tyr417, Asn545, and Glu667 of the “upper” domain of enzymes, while benzimidazole groups were less relevant for the interactions. The phenyl groups of compound **1** interacted mainly by means of van der Waals and electrostatic forces with the residuals Ile386, Pro387, Ala388, Phe556, Gln566, and Met569, as well as the cyclobutane ring with His568. The results of MD also imply that the binding of compound **1** has no important effect on the coordination geometry of the catalytic zinc ion.

The present study can guide the rational synthesis of novel inhibitors with an increased affinity and efficacy of inhibition in DPP III.

## 5. Acknowledgements

This work was supported by a grant awarded by the Josip Juraj Strossmayer University of Osijek (V.R.), and by the Croatian Ministry of Science, Education, and Sports (Projects 098-1191344-2938, 098-1191344-2860 to M.A. and S.T. and 125-0982464-1356 to M.H.).

## 6. References

- S. C. Prajapati, S. S. Chauhan, *FEBS J.* **2011**, *278*, 3256–3276. <http://dx.doi.org/10.1111/j.1742-4658.2011.08275.x>
- P. K. Baral, N. Jajčanin-Jozić, S. Deller, P. Macheroux, M. Abramić, K. Gruber, *J. Biol. Chem.* **2008**, *283*, 22316–22324. <http://dx.doi.org/10.1074/jbc.M803522200>
- G. A. Bezerra, E. Dobrovetsky, R. Viertlmayr, A. Dong, A. Binter, M. Abramic, P. Macheroux, S. Dhe-Paganon, K. Gruber, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 6525–6530. <http://dx.doi.org/10.1073/pnas.1118005109>
- Y. Liu, J. T. Kern, J. R. Walker, J. A. Johnson, P. G. Schultz, H. Luesch, *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 5205–5210. <http://dx.doi.org/10.1073/pnas.0700898104>
- H. Sato, K. Kimura, Y. Yamamoto, T. Hazato, *Masui* **2003**, *52*, 257–263.
- M. Baršun, A. N. Jajčanin, B. Vukelić, J. Špoljarić, M. Abramić, *Biol. Chem.* **2007**, *79*, 343–348.
- H. Zhang, Y. Yamamoto, S. Shumiya, M. Kunimatsu, K. Nishi, I. Ohkubo, *Histochem. J.* **2001**, *33*, 511–521. <http://dx.doi.org/10.1023/A:1014943522613>
- Š. Šimaga, D. Babić, M. Osmak, M. Šprem, M. Abramić, *Gynecol. Oncol.* **2003**, *91*, 194–200. [http://dx.doi.org/10.1016/S0090-8258\(03\)00462-1](http://dx.doi.org/10.1016/S0090-8258(03)00462-1)
- A. Taylor, *FASEB J.* **1993**, *7*, 290–298.
- Y. Yamamoto, J.-I. Hashimoto, M. Shimamura, T. Yamaguchi, T. Hazato, *Peptides* **2000**, *21*, 503–508. [http://dx.doi.org/10.1016/S0196-9781\(00\)00174-1](http://dx.doi.org/10.1016/S0196-9781(00)00174-1)
- T. Chiba, Y.-H. Li, T. Yamane, O. Ogikubo, M. Fukuoka, R. Arai, S. Takahashi, T. Ohtsuka, I. Ohkubo, N. Matsui, *Peptides* **2003**, *24*, 773–778. [http://dx.doi.org/10.1016/S0196-9781\(03\)00119-0](http://dx.doi.org/10.1016/S0196-9781(03)00119-0)
- Y. Yamamoto, H. Kanazawa, M. Shimamura, M. Ueki, T. Hazato, *Life Sci.* **1998**, *62*, 1767–1773. [http://dx.doi.org/10.1016/S0024-3205\(98\)00138-6](http://dx.doi.org/10.1016/S0024-3205(98)00138-6)
- D. Agić, M. Hranjec, N. Jajčanin, K. Starčević, G. Karminski-Zamola, M. Abramić, *Bioorg. Chem.* **2007**, *35*, 153–169. <http://dx.doi.org/10.1016/j.bioorg.2006.11.002>
- M. Hranjec, M. Kralj, I. Piantanida, M. Sedić, L. Šuman, K. Pavelić, G. Karminski-Zamola, *J. Med. Chem.* **2007**, *50*, 5696–5711. <http://dx.doi.org/10.1021/jm070876h>
- M. Hranjec, I. Piantanida, M. Kralj, L. Šuman, K. Pavelić, G. Karminski-Zamola, *J. Med. Chem.* **2008**, *51*, 4899–4910. <http://dx.doi.org/10.1021/jm8000423>
- M. Abramić, D. Schleuder, Lj. Dolovčak, W. Schröder, K. Strupat, D. Šagi, J. Peter-Katalinić, Lj. Vitale, *Biol. Chem.* **2000**, *381*, 1233–1243. <http://dx.doi.org/10.1515/BC.2000.151>
- M. Abramić, M. Zubanović, L. Vitale, *Biol. Chem Hoppe Seyler* **1988**, *369*, 29–38. <http://dx.doi.org/10.1515/bchm3.1988.369.1.29>
- A. Hocquet, M. Langgård, *J. Mol. Model.* **1998**, *4*, 94–112. <http://dx.doi.org/10.1007/s008940050128>
- M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909. <http://dx.doi.org/10.1021/ja00299a024>
- I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, *J. Comput. Aided Mol. Des.* **2005**, *19*, 453–463. <http://dx.doi.org/10.1007/s10822-005-8694-y>
- P. Gramatica, N. Chitico, N. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* **2013**, *34*, 2121–2132. <http://dx.doi.org/10.1002/jcc.23361>
- P. Gramatica, *QSAR Comb. Sci.* **2007**, *26*, 694–701. <http://dx.doi.org/10.1002/qsar.200610151>
- H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242. <http://dx.doi.org/10.1093/nar/28.1.235>
- 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, P. Kollman, *J. Comput. Chem.* **2003**, *24*, 1999–2012. <http://dx.doi.org/10.1002/jcc.10349>
- B. Bertoša, B. Kojić-Prodić, R. C. Wade, S. Tomić, *Biophys. J.* **2008**, *94*, 27–37. <http://dx.doi.org/10.1529/biophysj.107.109025>
- I. Dokmaniæ, M. Šikić, S. Tomić, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *64*, 257–263. <http://dx.doi.org/10.1107/S090744490706595X>
- A. Tomić, M. Abramić, J. Špoljarić, D. Agić, D. M. Smith, S. Tomić, *J. Mol. Recognit.* **2011**, *24*, 804–814. <http://dx.doi.org/10.1002/jmr.1115>
- D. A. Case, T. E. III Cheatham, T. A. Darden, H. Gohlke, R. Luo, K. M. Jr. Merz, A. Onufriev, C. Simmerling, B. Wang,

- R. J. Woods, *J. Computat. Chem.* **2005**, *26*, 1668–1688.  
<http://dx.doi.org/10.1002/jcc.20290>
29. D. A. Case, T. A. Darden, T. E. III Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, D. H. Mathews, M. G. Seetin, C. Saguí, V. Babin, P. A. Kollman, AMBER 10, University of California, San Francisco, **2008**.
30. V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, R. K. Robins, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172.  
<http://dx.doi.org/10.1021/ci00063a006>
31. R. Todeschini, V. Consonni, Handbook of molecular descriptors, M. Manhold, H. Kubinyi, H. Temmerman, eds., Wiley-VCH, Weinheim, **2000**, p. 667–673.  
<http://dx.doi.org/10.1002/9783527613106>
32. J. Gasteiger, J. Sadowski, J. Schuur, P. Selzer, L. Steinhauer, V. Steinhauer, *J. Chem Inf. Comput. Sci.* **1996**, *36*, 1030–1037. <http://dx.doi.org/10.1021/ci960343+>
33. M. C. Hemmer, V. Steinhauer, J. Gasteiger, *Vibrat. Spect.* **1999**, *19*, 151–164.  
[http://dx.doi.org/10.1016/S0924-2031\(99\)00014-4](http://dx.doi.org/10.1016/S0924-2031(99)00014-4)
34. L. B. Kier, *Quant. Struct.-Act. Relat.* **1985**, *4*, 109–116.  
<http://dx.doi.org/10.1002/qsar.19850040303>
35. A. Tomić, M. González, S. Tomić, *J. Chem. Inf. Model.* **2012**, *52*, 1583–1594.  
<http://dx.doi.org/10.1021/ci300141k>
36. D. P. Klose, B. A. Wallace, R. W. Janes, *Bioinformatics* **2010**, *26*, 2624–2625.  
<http://dx.doi.org/10.1093/bioinformatics/btq480>

## Povzetek

Študija molekularnega modeliranja je bila izvedena na 16 inhibitorjih na osnovi benzimidazola, ki inhibirajo človeško dipeptidil peptidazo III (DPP III). Osem od teh so nove spojine, sintetizirane z odličnimi izkoristki in uporabo zelenih pristopov kemije. Namen te študije je pojasnjevanje strukturnih značilnosti derivatov benzimidazola, potrebnih za antagonizem delovanja človeškega DPP III, ob uporabi analize povezave med strukturo in aktivnostjo (QSAR). Da bi razumeli mehanizme, kako se tovrstni močni inhibitorji vežejo v aktivno mesto encima, smo uporabili simulacijske metode molekularne dinamike (MD). Najboljši model vključuje S3K in RDF045m deskriptorje, ki razložijo 89,4 % zaviralnega delovanja. Opisani del z močno aktivnostjo inhibicije se ujema s strukturo najmočnejše spojine. MD simulacija je pokazala pomen imidazolinov in fenilnih skupin v mehanizmu vezavev aktivno mesto človeškega DPP III.