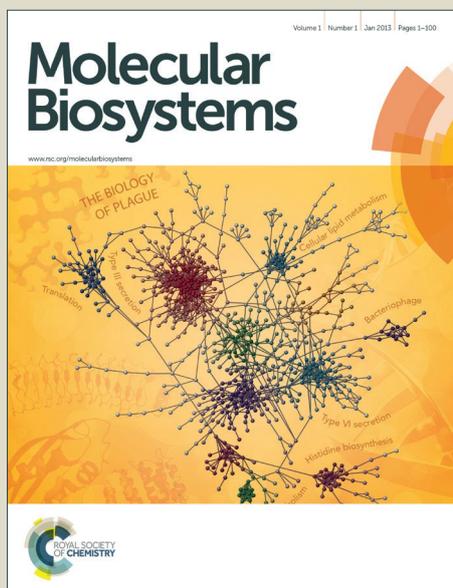


# Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

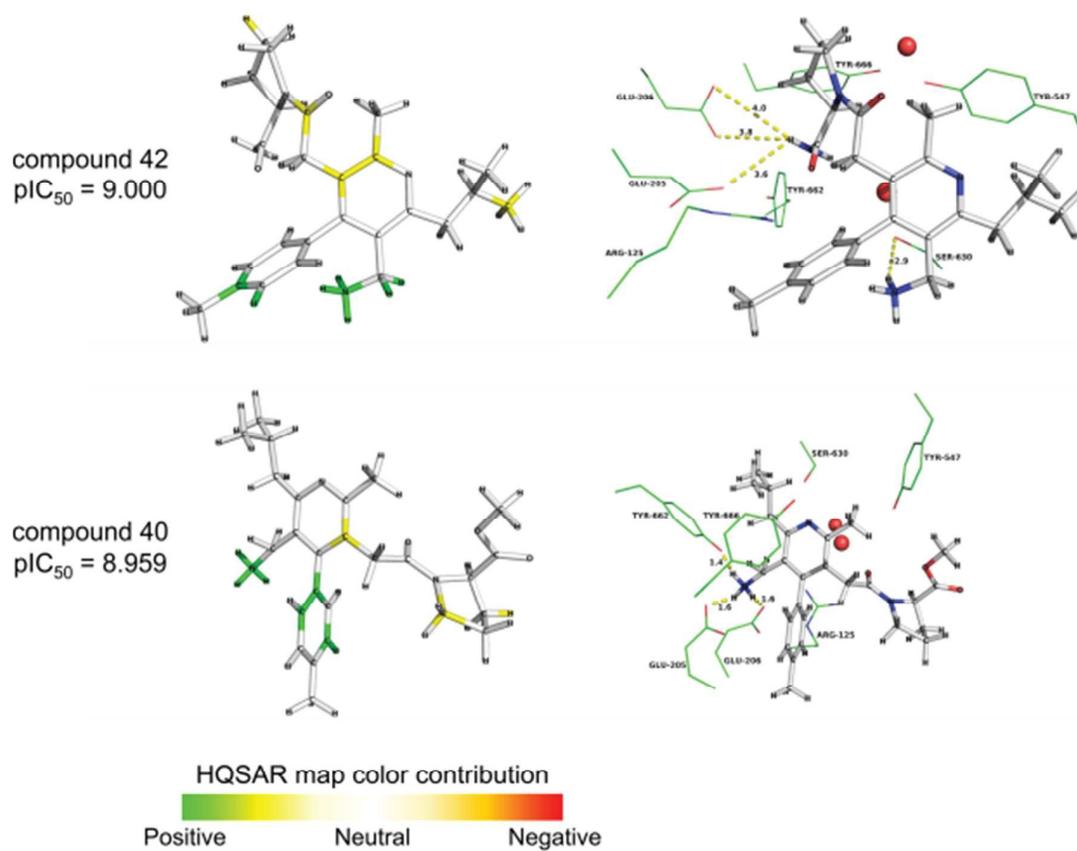


[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

## Graphical Abstract

**Molecular docking studies and 2D analyses of DPP-4 inhibitors as candidates to the treatment of diabetes**

Simone Queiroz Pantaleão, Vinicius Gonçalves Maltarollo, Sheila Cruz Araujo, Jadson Castro Gertrudes and Kathia Maria Honorio



Received 00th January  
20xx,

## Molecular docking studies and 2D analyses of DPP-4 inhibitors as candidates to the treatment of diabetes

Simone Queiroz Pantaleão,<sup>a</sup> Vinicius Gonçalves Maltarollo<sup>b</sup>, Sheila Cruz Araujo<sup>a</sup>, Jadson Castro Gertrudes<sup>c</sup> and Kathia Maria Honorio<sup>a,c\*</sup>

Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

[www.rsc.org/](http://www.rsc.org/)

Dipeptidyl peptidase-4 (DPP-4) is an important biological target related to the treatment of diabetes as DPP-4 inhibitors can lead to an increase in the insulin levels and a prolonged activity of glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP), being effective in the glycemia control. Thus, this study analyses the main molecular interactions between DPP-4 and a series of bioactive ligands. The methodology used here employed molecular modeling methods, such as HQSAR (Hologram Quantitative Structure-Activity) analyses and molecular docking, with the aim of understanding the main structural features of the compound series are essential for the biological activity. Analyses of the main interactions in the active site of DPP-4, in particular, the contribution of the hydroxyl coordination between Tyr547 and Ser630 by the water molecule, which is described in the literature as important for the coordinated interactions in the active site, were performed. Significant correlation coefficients of the best 2D model ( $r^2 = 0.942$  and  $q^2 = 0.836$ ) were obtained, indicating the predictive power of this model for untested compounds. Therefore, the final model constructed in this study, along with the information from the contribution maps, could be useful for the design of novel DPP-4 ligands with improved activity.

### 1 Introduction

Diabetes is characterized by a metabolic disorder in the protein synthesis and storage, cell uptake and use of carbohydrates. The occurrence of this imbalance may originate in insulin resistance.<sup>1</sup> Metabolic diseases such as diabetes and obesity are considered as the century epidemic.<sup>1-3</sup> According to the International Health Organization, there are over 347 million people with diabetes and it is predicted that, around 2030, it will become the seventh leading cause of death worldwide. The lack of awareness about diabetes, combined with poor access to health services and essential medicines, can lead to severe complications such as blindness, amputation and kidney failure.<sup>4</sup> The most common complications of diabetes include: (1) hypoglycemia that occurs when there is large amount of released insulin, (2) inflammatory reactions, which may contribute to the development of vascular complications and atherosclerosis, (3) microangiopathy that can lead to myocardial infarction, stroke and gangrene of lower members. Other complications involve retinopathy, edema, microaneurysms, nephropathy, symmetrical peripheral neuropathy affecting motor and sensory nerves of the lower limbs, and others.<sup>1,5,6</sup>

For the diabetes treatment, there are the following drugs available in the market: insulin, secretagogues (sulfonylureas and incretins) and hypoglycemics (biguanides, thiazolidinediones and  $\alpha$ -glucosidase inhibitors), as well as the combination of different drug classes.<sup>7-13</sup> These drugs are effective but can result in side effects such as cardiac complications, bone density loss, fluid retention, weight gain, digestive problems and urinary tract. From these consequences, the progression of diabetes constantly challenges the search for new drugs with tolerable side effects and easy acquisition by population. Inhibitors of dipeptidyl peptidase-4 (DPP-4) enzyme are considered as a recent pharmacological class to treat diabetes, which lead to an increase in the insulin levels and a prolonged activity of glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide (GIP), being effective in the glycemia control.<sup>9,10</sup> This control mechanism of insulin, associated with the distribution of energy for the cells of body tissues, is responsible for the glucose transport across the cell wall.<sup>14</sup>

DPP-4 is also known as adenosine deaminase complexing protein 2 or CD26 (cluster of differentiation 26) and is encoded by DPP4 gene in humans. DPP-4 enzyme is expressed on the surface of most cell types and is associated with immune regulation, signal transduction and apoptosis. It is responsible for the degradation of incretins such as GLP-1. DPP-4 inhibition prevents the inactivation of glucagon-like peptide 1 (GLP-1), which increases the insulin secretion, thereby lowering glucose levels. It is important to mention that, in the case of serine peptidases, the region of the catalytic domain that comprises the catalytic triad (Ser630, Asp708 and His740), the subdomain

<sup>a</sup> Center for Natural and Human Sciences, Federal University of ABC, 09210-170, Santo André, SP, Brazil.

<sup>b</sup> Faculty of Pharmaceutical Sciences, University of São Paulo, 05508-000, São Paulo, SP, Brazil.

<sup>c</sup> School of Arts, Sciences and Humanities, University of São Paulo, 03828-0000, São Paulo, SP, Brazil. E-mail: kmhonorio@usp.br

\*Electronic Supplementary Information (ESI): Table S1, S2 and S3 and Figure S1. See DOI: 10.1039/x0xx00000x

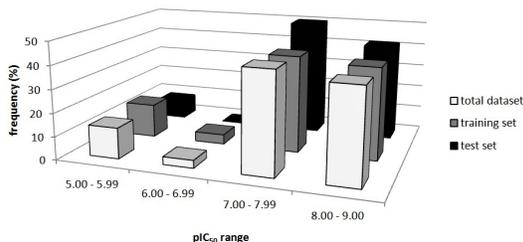
region with the presence of Tyr547, Tyr666, Ser630 (the latter residue is responsible for the interaction of the subdomain-field regions), as well as structural water molecules, that establish the coordination with these aminoacids, are essential to understand the main interactions between the biological target and the bioactive ligands.<sup>11,12</sup>

With the aim of analyzing the main molecular fragments related to the DPP-4 inactivation, we employed HQSAR (Hologram Quantitative Structure-Activity Relationships) technique, which is capable of generating 2D maps with positive and negative fragment contributions for the biological activity, allowing the planning of new ligands with better biological potential.<sup>15-18</sup> Other technique used in this study was molecular docking, which helps the understanding of the main interactions between the studied compounds and the residues in the active site of DPP-4. The combination of HQSAR and molecular docking analyses allows us to understand the main ligand-receptor interactions, as well as indicating molecular modifications for the design of new DPP-4 inhibitors.

## 2 Methodology

### 2.1 Compound set

60 inhibitors of the DPP-4 enzyme, derived from 3-aminomethyl-1,2-dihydro-4-phenyl-1-isoquinolones; quinoline; 3-pyridylacetamide and 3-pyridylacetic acid, were employed in all molecular modeling studies. These compounds were synthesized and the biological activity ( $IC_{50}$  concentration) was measured under the same experimental conditions, as performed by Banno *et al.*<sup>19</sup>, Maezaki *et al.*<sup>20</sup> and Miyamoto *et al.*<sup>21,22</sup>. For the construction of the statistical models, the  $IC_{50}$  values were converted to  $pIC_{50}$  ( $-\log IC_{50}$ ) and the compounds were grouped in training and test sets, containing 80 and 20% of the total number of compounds, respectively. In order to perform the training and test splitting, the compounds were divided in 4 subsets according to the  $pIC_{50}$  values: subset 1 ( $pIC_{50} \geq 5$  and  $< 6$ ); subset 2 ( $pIC_{50} \geq 6$  and  $< 7$ ); subset 3 ( $pIC_{50} \geq 7$  and  $< 8$ ) and subset 4 ( $pIC_{50} \geq 8$  and  $\leq 9$ ). After the splitting of the compounds, we employed MATLAB software<sup>23</sup> to perform a random selection of the test compounds, considering each  $pIC_{50}$  range and respecting an homogeneous distribution (chemical diversity) in both training and test sets (Figure 1). The structure of the studied compounds and the  $pIC_{50}$  values are presented in Table S1.



**Figure 1.** Distribution of the compounds in the training and test sets according to the  $pIC_{50}$  range.

### 2.2. HQSAR

Using the 60 selected compounds, we employed the HQSAR technique to generate a 2D model that correlates the molecular hologram (fragments derived from the 2D structure and its respective contributions) with the biological activity (expressed as  $pIC_{50}$  values). For the generation of the molecular hologram, some parameters are varied during this procedure, such as the distinction of fragments using information on atoms (A), bonds (B), connections (C), hydrogen atoms (H) chirality (Ch) and hydrogen bonds acceptor/donor (DA), as well as the fragment size (defined by the number of atoms) and the fragment length (amount of binary descriptors related to fragment count). The molecular holograms are mathematical representations of the fragments present in each compound of the dataset and can be related to the biological response from the use of multivariate techniques, such as partial least square (PLS) regression.<sup>24</sup>

After the construction of the HQSAR model, the holograms can graphically be converted in contribution maps using a range of specific colors that represent the contribution of the fragments to the biological activity.<sup>12</sup> In this study, the quality of the models was evaluated taking into account the internal validation coefficient ( $q^2$ ) values obtained from leave-one-out (LOO) cross validation method, as well as the regression coefficient ( $r^2$ ) values and the respective errors. After the construction of the best 2D model and the respective internal validation, other validation procedures were carried out with the aim of verifying the robustness, the predictability and the applicability domain of the constructed HQSAR model. The construction and validation of the HQSAR model were performed employing the computational package Sybyl 8.1.<sup>25</sup>

### 2.3. Molecular Docking

After all HQSAR analyses, molecular docking studies were performed in order to combine the information provided from the HQSAR model and the binding mode of the studied DPP-4 ligands. For all molecular docking studies, we used GOLD 5.0 software<sup>26</sup> which employs genetic algorithm to generate the poses (conformations) of the ligands at the active site of the biological target.

The docking protocol employed in this study consisted in the following steps: (i) preparation of the ligands (construction of 3D structures, setting the protonation state according to the physiological pH and calculation of the atomic charges employing PM3 semi-empirical method<sup>27</sup> from the MOPAC software<sup>28</sup>, implemented in Sybyl 8.1 package<sup>25</sup>); (ii) selection of the 3D structure of the DPP-4 enzyme (PDB ID: 4A5S<sup>29,30</sup>, that has the best resolution (1,52Å)); (iii) inclusion of 2 structural waters in the docking calculations according to experimental data from the literature<sup>11,29-33</sup>; (iv) addition of hydrogen atoms and protonation setting of the main residues in the active site; (v) setting the flexibility of the residues Ser630, Tyr547 and Tyr666 according to the previous validation (Table S2); (vi) GOLDScore<sup>26</sup> was employed as scoring function; (vii) definition of the active site as 5Å around the crystallographic ligand; (viii) validation of the docking

protocol based on root mean squared deviation (RMSD) values (calculated with Chimera software<sup>34</sup>) from the redocking poses. The validation step was performed varying the use of structural waters and the flexibility of some residues in the docking calculations (Table S2 and Figure S1). The conditions adopted for the docking simulations were evaluated according to several studies in the literature<sup>11,19-22, 31, 33, 35</sup> and also the redocking analysis. The structural water positions at the active site were obtained according to the alignment of several DPP-4 structures (PDB IDs: 4A5S<sup>29, 30</sup>, 3SWW<sup>36</sup>, 3D4L<sup>37</sup>, 3EIO<sup>38</sup>, 3C45<sup>39</sup>, 2OGZ<sup>40</sup>, 2QT9<sup>41</sup>, 1X70<sup>42</sup>, 2AJ8<sup>43</sup>, 2IIV<sup>44</sup>, 2P8S<sup>45</sup>, 2QJR<sup>46</sup>, 1RWQ<sup>47</sup>, 3QBJ<sup>48</sup>, 3HAC<sup>49</sup>), which contain the conserved water molecules at the active site.

### 3 Results and Discussion

Initially, we constructed 13 HQSAR models varying the fragment distinction and maintaining the default fragment size (4-7 atoms) with the aim of assessing the influence of the descriptors on the robustness of the models (Table 1). For each constructed model, we also evaluated the influence of the hologram length (HL) by constructing models varying HL as: 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353 and 401 bins. All constructed models showed acceptable LOO validation coefficients ( $q^2 > 0.7$ ) indicating that the HQSAR method is suitable to generate robust statistical models and can be used to predict the biological activity of new bioactive ligands. The best 2D model, according to the highest  $q^2$  value, was constructed with atoms, bonds, H atoms, chirality and H-bond donor and acceptor as fragment distinction (A/B/H/Ch/DA, model 11). The model 11 had higher  $q^2$  value and lower standard error of prediction than the two second best models (models 5 and 8).

**Table 1.** Statistical results of the 13 initial models obtained from the variation of the fragment distinction and maintaining the default fragment size (4 - 7 atoms)

Model	Fdist	$q^2$	SEP	$r^2$	SEE	HL	PCs
1	a/b	0.789	0.496	0.866	0.395	97	3
2	a/b/c	0.771	0.528	0.919	0.315	307	5
3	a/b/c/h	0.761	0.547	0.937	0.280	353	6
4	a/b/c/h/ch	0.769	0.538	0.911	0.333	61	6
5	a/b/h	0.817	0.479	0.927	0.302	61	6
6	a/b/c/ch	0.769	0.519	0.857	0.409	199	3
7	a/b/da	0.798	0.503	0.939	0.276	401	6
8	a/b/c/da	0.817	0.479	0.934	0.288	53	6
9	a/b/h/da	0.802	0.481	0.852	0.416	61	3
10	a/b/c/ch/da	0.791	0.511	0.928	0.301	53	6
<b>11</b>	<b>a/b/H/ch/da</b>	<b>0.829</b>	<b>0.463</b>	<b>0.937</b>	<b>0.282</b>	<b>257</b>	<b>6</b>
12	a/b/c/h/ch/da	0.794	0.502	0.921	0.311	71	5
13	a/b/c/h/Da	0.804	0.495	0.947	0.258	97	6

Fdist: fragment distinction;  $q^2$ : cross-validated coefficient; SEP: standard error of validation;  $r^2$ : non-validated coefficient; SEE: standard error of estimation; HL: hologram length; PCs: number of principal components

After this first analysis (variation of the fragment distinction), we constructed more six models by varying the

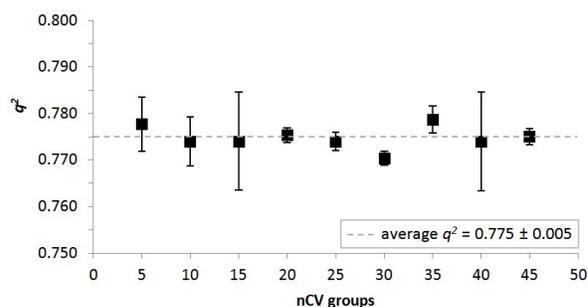
fragment size and maintaining both fragment distinction (A/B/H/Ch/DA) and hologram length (257 bins) (see Table 2). From this procedure, we found that the variation of the fragment size provided an improvement in the statistical quality of the model 11, i.e. the model with fragment distinction as A/B/H/Ch/DA, hologram length equals to 257 bins and fragment size as 3 to 6 atoms was the best one (model 16, Table 2).

**Table 2.** Variation of the fragment size for the model containing A/B/H/Ch/DA as fragment distinction

Fsize	$q^2$	SEP	$r^2$	SEE	HL	PCs
1 to 4	0.806	0.493	0.923	0.310	257	6
2 to 5	0.818	0.471	0.918	0.316	257	5
<b>3 to 6</b>	<b>0.836</b>	<b>0.453</b>	<b>0.942</b>	<b>0.270</b>	<b>257</b>	<b>6</b>
5 to 8	0.799	0.501	0.944	0.266	257	6
6 to 9	0.773	0.532	0.935	0.286	257	6
7 to 10	0.761	0.547	0.934	0.288	257	6

Fsize: fragment size;  $q^2$ : cross-validated coefficient; SEP: standard error of validation;  $r^2$ : non-validated coefficient; SEE: standard error of estimation; HL: hologram length; PCs: number of principal components

We also evaluated the internal quality of the model 16 by carrying out a robustness test. We validated the model with leave-N-out technique varying the number of cross-validation groups, in triplicate experiments, in order to check the sensibility of the model robustness by varying the number of compounds employed in the model construction. All  $q^2$  values for each N cross-validation groups were higher than 0.770, indicating the robustness of the model 16. The average  $q^2$  value was equal to  $0.775 \pm 0.005$  (Figure 2)



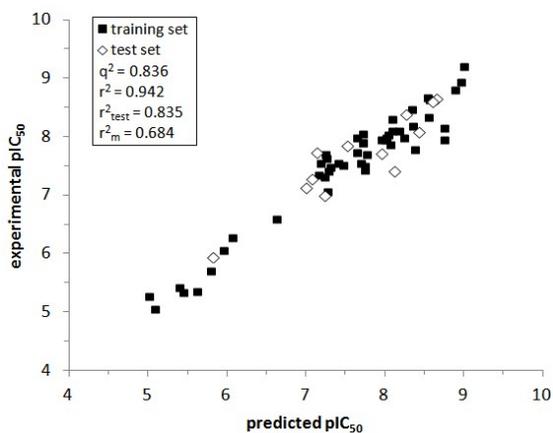
**Figure 2.** Robustness test for the HQSAR model with variation in the number of cross-validation groups from 5 to 45.

After all internal validations, the model 16 was submitted to external validations employing the test compounds. In this step, the obtained results were very promising taking into account some data: (i) the predicted  $pIC_{50}$  values for all test compounds showed a residual error lower than 1 log unit (Table 3); (ii) the external validation coefficient ( $r^2_{test}$ ) was equal to 0.835; (iii) the predictive potential for the test set ( $r^2_m$ ) was equal to 0.684 (Figure 3) and; (iv) the predictive potential for all compounds ( $r^2_{m,overall}$ ) was equal to 0.691. We also calculated the percentage of missing fragments for the test compounds and all molecules showed missing

information equals to 0%, indicating that all dataset is inside the applicability domain. These results indicate that the model 16 has a high predictive power and robustness. Therefore, this model is reliable to predict the biological activity of new compounds.

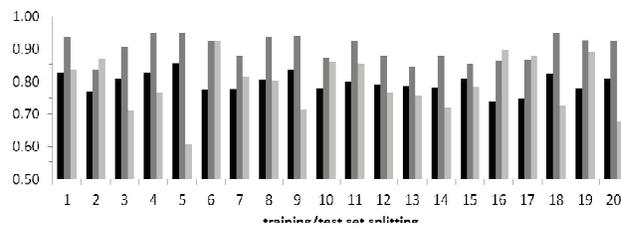
**Table 3.** Experimental and predicted  $pIC_{50}$  for the test compounds

Compound	Experimental $pIC_{50}$	Predicted $pIC_{50}$	Residual
10	5.824	5.952	0.128
15	7.076	7.288	0.212
17	8.125	7.417	0.708
20	7.000	7.135	-0.135
23	8.658	8.666	-0.008
28	7.137	7.733	-0.596
29	8.432	8.084	0.348
35	7.523	7.850	-0.327
38	7.959	7.727	0.232
51	7.244	7.006	0.238
58	8.602	8.610	-0.008
60	8.276	8.389	-0.113



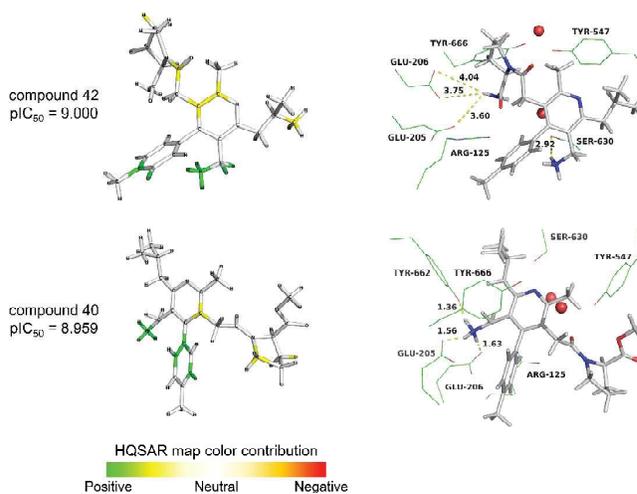
**Figure 3.** Experimental versus predicted  $pIC_{50}$  for the training and test sets.

Finally, we performed 19 more training/test set splittings using the same strategy (random selection taking into account each  $pIC_{50}$  range) in order to evaluate the influence of the composition in each compound subset on the statistical results. All test compounds in the 20 subsets are displayed in Table S2 and this test indicated that all statistical parameters ( $q^2$ ,  $r^2$  and  $r^2_{test}$ ) are acceptable according to the literature (Figure 4), suggesting that the model was not obtained by chance.



**Figure 4.** Training and test set splittings (20 subsets). The dark gray bars represent  $q^2$  values, light gray bars indicate  $r^2$  values and black bars represent  $r^2_{test}$  values.

After the construction of the 2D model, we calculated the contribution maps for the most active compounds (Figure 5) and the least active ones (Figure 6). The green and yellow regions indicate essential groups for the biological activity, while regions in orange and red indicate groups that negatively contribute to the biological activity and could be modified to increase the biological potential.

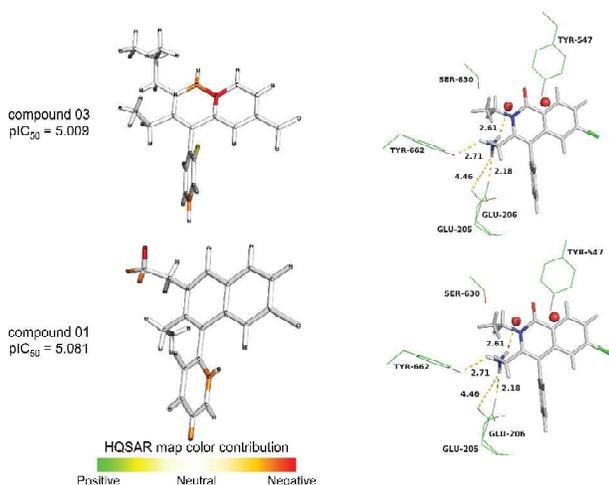


**Figure 5.** HQSAR map and docking analysis of the most active compounds.

Analysing the HQSAR maps, we can see that some regions can be considered as essential for the biological activity, such as: (1) pyrrolidine ring; (2) isobutyl group; (3) anilide group, which are responsible for a significant number of primary interactions with residues in the active site. From the docking studies, we analysed the main interactions of some ligands with the main residues at the active site (Ser630, Tyr547, Tyr666, His740 and Asp708), taking into account the presence of structural water molecules. The nitrogen atom of the amine group ( $NH_3$ ) in the compound **42** performs hydrogen bonding with Ser630, which interacts with a water molecule (W1) and the amide group interacts with Glu205 and Glu206. It is also important to mention that t-butyl group of the compound **42** (colored as yellow in the HQSAR contribution map) could perform hydrophobic interactions with the phenyl ring of Tyr547. Finally, the compound **42** can perform a  $\pi$ -stacking

interaction with the phenyl ring (predicted by the HQSAR model as positive contributor to the biological activity) and Arg125.

Compound **40** can also perform hydrophobic interaction with Tyr547, which performs hydrogen bond with a structural water (W1). The amine group (NH<sub>3</sub>) of the compound **40** (colored as green in the HQSAR contribution map) performs a hydrogen bond with Tyr662 and ionic interactions with Glu205 and Glu206. A possible cation- $\pi$  interaction between Arg125 and the phenyl group (also colored as green in the HQSAR map) contributes for the biological activity.



**Figure 6.** HQSAR map and docking analysis for the least active compounds.

The main interactions performed by the compounds **3** and **1** (the least active ones, Figure 6) are ionic interactions with Glu206, as well as hydrogen bond with Tyr662, indicating that these interactions are important to explain the biological activity of the DPP-4 inhibitors studied in this work. Furthermore, both amine groups of these compounds are highlighted as neutral contribution in the HQSAR contribution map. The compound **3** can also perform a hydrogen bond with the nitrogen atom of the main chain of Tyr547 via carboxyl moiety of the quinoline group. The benzene ring of both compounds (**3** and **1**) are located at Phe357 region and is colored as orange in the HQSAR map, indicating that hydrophobic interactions at this region do not favor the biological activity.

## Conclusions

From the HQSAR analyses, it was possible to obtain a significant statistical model ( $q^2=0.836$ ,  $r^2=0.942$  and  $r^2_{\text{test}}=0.835$ ) and can be used to predict the activity of novel ligands with unknown activity. Furthermore, the contribution 2D maps indicate regions with positive contribution to the biological activity, for example, the presence of pyrrolidine, isobutyl and anilide groups, favoring a significant number of

primary interactions with residues in the active site. It was found that not only the catalytic triad is important for the biological activity, but also the residues Tyr547 and Tyr666, can provide essential interactions in the active site. Structural water molecules were also crucial to understand the main interactions between the studied compounds and the biological target. Furthermore, interactions with Tyr547 and Arg125 can improve the potency of the DPP-4 inhibitors. The polar and hydrophobic interactions are very important for the molecular recognition of the network interactions formed between ligands, water molecules and the main residues in the active site. Therefore, the results obtained in this study can contribute to the understanding of the fundamental requirements for the interaction between the bioactive substances and the DPP-4 enzyme, helping the design of new DPP-4 ligands.

## Acknowledgements

The authors would like to thank CNPq, FAPESP, CAPES and UFABC for the financial support.

## References

1. A. D. Association, *Diabetes care*, 2014, **37**, S14.
2. E. Nilsson, P. A. Jansson, A. Perfilyev, P. Volkov, M. Pedersen, M. K. Svensson, P. Poulsen, R. Ribell-Madsen, N. L. Pedersen and P. Almgren, *Diabetes*, 2014, **63**, 2962-2976.
3. Diabetes Prevention Program Research, *Diabetes Care*, 2003, **26**, 36-47.
4. W. W. H. Organization 10 facts about diabetes, <http://www.who.int/features/factfiles/diabetes/en/> (accessed May 2015)
5. B. Ahrén, *Diabetes Metab. Syndr. Obes.* 2010, **3**, 31-41.
6. A. J. Dunkley, D. H. Bodicoat, C. J. Greaves, C. Russell, T. Yates, M. J. Davies and K. Khunti, *Diabetes Care*, 2014, **37**, 922-933.
7. R. V. C. Guido, A. D. Andricopulo and G. Oliva, *Est. Av.* 2010, **24**, 81-98.
8. H. Duez, B. Cariou and B. Staels, *Biochem. Pharmacol.* 2012, **83**, 823-832.
9. J. F. Gautier, S. Fetita, E. Sobngwi and C. Salaün-Martin, *Diabetes Metab.* 2005, **31**, 233-242.
10. M. A. Elrishi, K. Khunti, J. Jarvis and M. J. Davies, *Pract. Diabetes Int.* 2007, **24**, 474-482.
11. J. R. Bjelke, J. Christensen, S. Branner, N. Wagtmann, C. Olsen, A. B. Kanstrup and H. B. Rasmussen, *J. Biol. Chem.* 2004, **279**, 34691-34697.
12. J. L. Gross, S. P. Silveiro, J. L. Camargo, A. J. Reichelt and M. J. d. Azevedo, *Arq. Bras. Endocrinol. Metabol.* 2002, **46**, 16-26.
13. R. Kahn and M. B. Davidson, *Diabetes Care*, 2014, **37**, 943-949.
14. M. A. Nauck, *Diabetes Care*, 2013, **36**, 2126-2132.
15. S. C. Araujo, V. G. Maltarollo and K. M. Honorio, *Eur. J. Pharm. Sci.* 2013, **49**, 542-549.

16. T. S. Garcia and K. M. Honório, *J. Braz. Chem. Soc.* 2011, **22**, 65-72.
17. A. G. Ugarkar, P. K. Ambre, E. C. Coutinho, S. Nandan and R. R. Pissurlenkar, *Can. J. Chem.* 2014, **92**, 670-676.
18. J. Xu, S. Huang, T. Zhang, N. Wu, H. Kang, S. Cai and W. Shen, *Med. Chem. Res.* 2015, **24**, 1744-1752.
19. Y. Banno, Y. Miyamoto, M. Sasaki, S. Oi, T. Asakawa, O. Kataoka, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, S. Tsubotani, A. Tani, M. Funami, M. Tawada, Y. Yamamoto, K. Aertgeerts, J. Yano and H. Maezaki, *Bioorg. Med. Chem.* 2011, **19**, 4953-4970.
20. H. Maezaki, Y. Banno, Y. Miyamoto, Y. Moritou, T. Asakawa, O. Kataoka, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, M. Sasaki, S. Tsubotani, A. Tani, M. Funami, Y. Yamamoto, M. Tawada, K. Aertgeerts, J. Yano and S. Oi, *Bioorg. Med. Chem.* 2011, **19**, 4482-4498.
21. Y. Miyamoto, Y. Banno, T. Yamashita, T. Fujimoto, S. Oi, Y. Moritoh, T. Asakawa, O. Kataoka, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, S. Tsubotani, A. Tani, M. Funami, M. Amano, Y. Yamamoto, K. Aertgeerts, J. Yano and H. Maezaki, *Bioorg. Med. Chem.* 2011, **19**, 172-185.
22. Y. Miyamoto, Y. Banno, T. Yamashita, T. Fujimoto, S. Oi, Y. Moritoh, T. Asakawa, O. Kataoka, H. Yashiro, K. Takeuchi, N. Suzuki, K. Ikedo, T. Kosaka, S. Tsubotani, A. Tani, M. Sasaki, M. Funami, M. Amano, Y. Yamamoto, K. Aertgeerts, J. Yano and H. Maezaki, *J. Med. Chem.* 2011, **54**, 831-850.
23. Matlab, version 7.8.0 (R2009a). The MathWorks Inc.: Natick, MA, 2009.
24. B. Nadler and R. R. Coifman, *J. Chemometr.* 2005, **19**, 45-54.
25. *Sybyl, version 8.1, Tripos Inc.: St. Louis, MO, 2008.*
26. M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins*, 2003, **52**, 609-623.
27. J. P. Stewart, *J. Mol. Model.* 2004, **10**, 155-164.
28. M. Yináyi, *J. Chem. Soc.* 1991, **5**, 635-637.
29. J. M. Sutton, D. E. Clark, S. J. Dunsdon, G. Fenton, A. Fillmore, N. V. Harris, C. Higgs, C. A. Hurley, S. L. Krintel, R. E. MacKenzie, A. Duttaroy, E. Gangl, W. Maniara, R. Sedrani, K. Namoto, N. Ostermann, B. Gerhartz, F. Sirockin, J. Trappe, U. Hassiepen and D. K. Baeschlin, *Bioorg. Med. Chem. Lett.* 2012, **22**, 2359.
30. J. M. Sutton, D. E. Clark, S. J. Dunsdon, G. Fenton, A. Fillmore, N. V. Harris, C. Higgs, C. A. Hurley, S. L. Krintel, R. E. MacKenzie, A. Duttaroy, E. Gangl, W. Maniara, R. Sedrani, K. Namoto, N. Ostermann, B. Gerhartz, F. Sirockin, J. Trappe, U. Hassiepen and D. K. Baeschlin, *Bioorg. Med. Chem. Lett.* 2012, **22**, 1464-1468.
31. R. Thoma, B. Löffler, M. Stihle, W. Huber, A. Ruf and M. Hennig, *Structure*, 2003, **11**, 947-959.
32. A. J. Barrett, N. D. Rawlings and J. F. Woessner, *Handbook of Proteolytic Enzymes: Cysteine, serine and threonine peptidases.*, 2. edn., 2004.
33. R. Mentlein, *Regul. Peptides*. 1999, **85**, 9-24.
34. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.* 2004, **25**, 1605-1612.
35. N. D. Rawlings, M. Waller, A. J. Barrett and A. Bateman, *Nucleic Acids Res.* 2014, **42**, D503-D509.
36. W. Wang, P. Devasthale, A. Wang, T. Harrity, D. Egan, N. Morgan, M. Cap, A. Fura, H. E. Klei, K. Kish, C. Weigelt, L. Sun, P. Levesque, Y.-X. Li, R. Zahler, M. S. Kirby and L. G. Hamann, *Bioorg. Med. Chem. Lett.* 2011, **21**, 6646-6651.
37. G.-B. Liang, X. Qian, T. Biftu, S. Singh, Y.-D. Gao, G. Scapin, S. Patel, B. Leiting, R. Patel, J. Wuc, X. Zhang, N. A. Thornberry and A. E. Weber, *Bioorg. Med. Chem. Lett.* 2008, **18**, 3706-3710.
38. J. H. Ahn, W. S. Park, M. A. Jun, M. S. Shin, S. K. Kang, K. Y. Kim, S. Dal Rhee, M. A. Bae, K. R. Kim, S. G. Kim, S. Y. Kim, S. K. Sohn, N. S. Kang, J. O. Lee, D. H. Lee, H. G. Cheon and S. S. Kim, *Bioorg. Med. Chem. Lett.* 2008, **18**, 6525-6529.
39. S. D. Edmondson, L. Wei, J. Xu, J. Shang, S. Xu, J. Pang, A. Chaudhary, D. C. Dean, H. He, B. Leiting, K. A. Lyons, R. A. Patel, S. B. Patel, G. Scapin, J. K. Wu, M. G. Beconi, N. A. Thornberry and A. E. Weber, *Bioorg. Med. Chem. Lett.* 2008, **18**, 2409-2413.
40. S. M. Sheehan, H.-J. Mest, B. M. Watson, V. J. Klimkowski, D. E. Timm, A. Cauvin, S. H. Parsons, Q. Shi, E. J. Canada, M. R. Wiley, G. Ruehler, B. Evers, S. Petersen, L. C. Blaszcak, S. R. Pulley, B. J. Margolis, G. N. Wishart, B. Renson, D. Hankotius, M. Mohr, J.-C. Zechel, J. M. Kalbfleisch, E. A. Dingess-Hammond, A. Boelke and A. G. Weichert, *Bioorg. Med. Chem. Lett.* 2007, **17**, 1765-1768.
41. D. E. Kaelin, A. L. Smenton, G. J. Elermann, H. He, B. Leiting, K. A. Lyons, R. A. Patel, S. B. Patel, A. Petrov, G. Scapin, J. K. Wu, N. A. Thornberry, A. E. Weber and J. L. Duffy, *Bioorg. Med. Chem. Lett.* 2007, **17**, 5806-5811.
42. D. Kim, L. P. Wang, M. Beconi, G. J. Eiermann, M. H. Fisher, H. B. He, G. J. Hickey, J. E. Kowalchick, B. Leiting, K. Lyons, F. Marsilio, M. E. McCann, R. A. Patel, A. Petrov, G. Scapin, S. B. Patel, R. S. Roy, J. K. Wu, M. J. Wyratt, B. B. Zhang, L. Zhu, N. A. Thornberry and A. E. Weber, *J. Med. Chem.* 2005, **48**, 141-151.
43. M. Engel, T. Hoffmann, S. Manhart, U. Heiser, S. Chambre, R. Huber, H. U. Demuth and W. Bode, *J. Mol. Biol.* 2006, **355**, 768-783.
44. T. Biftu, D. Feng, X. Qian, G.-B. Liang, G. Kieczkowski, G. Eiermann, H. Huaibing, B. Leiting, K. Lyons, A. Petrov, R. Sinha-Roy, B. Zhang, G. Scapin, S. Patel, Y.-D. Gao, S. Singh, J. Wu, X. Zhang, N. A. Thornberry and A. E. Weber, *Bioorg. Med. Chem. Lett.* 2007, **17**, 49-52.
45. T. Biftu, G. Scapin, S. Singh, D. Feng, J. W. Becker, G. Eiermann, H. He, K. Lyons, S. Patel, A. Petrov, R. Sinha-Roy, B. Zhang, J. Wu, X. Zhang, G. A. Doss, N. A. Thornberry and A. E. Weber, *Bioorg. Med. Chem. Lett.* 2007, **17**, 3384-3387.
46. S. W. Wright, M. J. Ammirati, K. M. Andrews, A. M. Brodeur, D. E. Danley, S. D. Doran, J. S. Lillquist, S. Liu, L. D. McClure, R. K. McPherson, T. V. Olson, S. J. Orena, J. C. Parker, B. N. Rocke, W. C. Soeller, C. B. Soglia, J. L. Treadway, M. A. VanVolkenburg, Z. Zhao and E. D. Cox, *Bioorg. Med. Chem. Lett.* 2007, **17**, 5638-5642.

## Journal Name

## ARTICLE

47. J. U. Peters, S. Weber, S. Kritter, P. Weiss, A. Wallier, M. Boehringer, M. Hennig, B. Kuhn and B. M. Loeffler, *Bioorg. Med. Chem. Lett.* 2004, **14**, 1491-1493.
48. S. P. LIU, *To be Published*.
49. S. D. Edmondson, A. Mastracchio, J. M. Cox, G. J. Eiermann, H. He, K. A. Lyons, R. A. Patel, S. B. Patel, A. Petrov, G. Scapin, J. K. Wu, S. Xu, B. Zhu, N. A. Thornberry, R. S. Roy and A. E. Weber, *Bioorg. Med. Chem. Lett.* 2009, **19**, 4097-4101.