



**HAL**  
open science

## Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors

Georgiana Surpateanu, Bogdan Iorga

► **To cite this version:**

Georgiana Surpateanu, Bogdan Iorga. Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors. *Journal of Computer-Aided Molecular Design*, 2012, 26 (5), pp.595-601. 10.1007/s10822-011-9526-x . hal-02377117

**HAL Id: hal-02377117**

**<https://hal.science/hal-02377117>**

Submitted on 22 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors

Georgiana Surpateanu and Bogdan I. Iorga\*

*Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Centre de Recherche de Gif-sur-Yvette, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France*

Phone: +33 1 69 82 30 94

Fax: +33 1 69 07 72 47

E-mail: bogdan.iorga@icsn.cnrs-gif.fr

**Abstract:** In this study, we have "blindly" assessed the ability of several combinations of docking software and scoring functions to predict the binding of a fragment-like library of bovine trypsin inhibitors. The most suitable protocols (involving Gold software and GoldScore scoring function, with or without rescoring) were selected for this purpose using a training set of compounds with known biological activities. The selected virtual screening protocols provided good results with the SAMPL3-VS dataset, showing enrichment factors of about 10 for Top 20 compounds. This methodology should be useful in difficult cases of docking, with a special emphasis on the fragment-based virtual screening campaigns.

**Keywords:** *virtual screening; docking; scoring function; fragment-like compounds; trypsin inhibitors*

Abbreviations:

RMSD – Root Mean Square Deviation

ROC – receiver operating characteristic

AUC – Area Under Curve

CI – Confidence Interval

QPLD – QM-Polarized Ligand Docking

## Introduction

The use of fragment-based approaches in drug design and particularly in virtual screening has known an explosive development in recent years [1-7]. In some instances, this approach has been successfully used as a less expensive alternative

to the NMR and X-ray crystallography techniques for the assessment of binding site druggability [8]. The key usefulness of fragment-based techniques is provided by the maximal exploration of the chemical space using a limited number of compounds [9, 10].

Docking of these small-sized molecules represents a very efficient tool for the identification of new weak, fragment-like binders that will represent starting points for the elaboration of complex and more potent drug-like compounds [11-14]. One major limitation of this approach is considered to be the performance of scoring functions to correctly predict the binding modes of fragments [13, 15, 16]. In this context, the SAMPL3-VS challenge intended to provide an opportunity for the evaluation of the current knowledge in this exciting field. This year, the dataset contained 500 fragment-like compounds that were potential inhibitors of bovine trypsin and it was required to submit a ranked list of these compounds.

## Materials and Methods

### Construction of the training set of bovine trypsin inhibitors

A training set of 25 compounds, with known biological activities against bovine trypsin and for which crystal structures in complex with this enzyme were available in the Protein Data Bank (PDB [17, 18]), were selected in the first step (Table 1). These compounds were further used for the evaluation of two docking software (Gold and Glide) for the ability of pose prediction and ranking of fragments using different docking protocols.

**Table 1.** Compounds with known biological activities ( $K_i$ ) for bovine trypsin and crystallographic structures available in the PDB which were used as training test in this study

PDB code	Ligand code	Resolution (Å)	$K_i$ (μM)	Reference
1K1N	CCR	2.00	0.068	[19]
1G36	R11	1.90	0.067	[20]
1OYQ	T87	1.90	0.11	[20]
2ZQ2	13U	1.40	0.102	[21]
1BJV	GP8	1.80	2.9	[22]
3GY2	BRN	1.57	1.8	[23]
1V2J	BEN	1.90	4.8	[24]

1GI0	BMZ	1.42	3.4	[25]
1GI4	122	1.37	1.9	[25]
1GI5	123	1.60	1.2	[25]
1O2S	CR4	1.65	3.4	[26]
1O30	693	1.55	0.17	[26]
1O38	653	1.38	0.15	[26]
1O36	607	1.70	1.1	[26]
1O33	801	1.46	1.8	[26]
1O2Z	312	1.65	0.78	[26]
1O2P	972	1.47	0.44	[26]
1O2N	762	1.50	0.81	[26]
1O2W	847	1.38	1.4	[26]
1QB9	806	1.80	0.036	[27]
1QB1	974	1.80	0.17	[27]
1QBO	711	1.80	0.018	[27]
1QB6	623	1.80	0.87	[27]
1G3B	108	1.80	4.2	[28]
1G3C	109	1.80	2.8	[28]

---

### Ligand preparation

All 544 structures from the SAMPL3-VS dataset and the 25 structures from the training set were prepared using the LigPrep module from Schrödinger Suite (<http://www.schrodinger.com>). When applicable, all possible stereoisomers and protonation forms at  $\text{pH } 7.0 \pm 2.0$  were considered, and the one showing the best correlation with the experimental data (in the case of the training set) or the best docking score (for SAMPL3-VS dataset) was retained. In this way, an extended SAMPL3-VS dataset (containing 647 structures) was obtained and used for virtual screening.

### Docking and virtual screening protocols

The bovine trypsin structure used in this study (PDB code 2AYW [29]) was selected considering its very good resolution (0.97 Å) and the length of the amino acid sequence.

Docking and virtual screening calculations were performed using Gold 5.0 [30] and Glide 5.5 (<http://www.schrodinger.com>) software. Gold employs a genetic algorithm to explore the full range of conformational flexibility of the ligand in the protein binding site, whereas Glide was described to perform a complete systematic search of the conformational, orientational, and positional space of the docked ligand, using a hierarchical series of filters to search for possible locations of the ligand in the active-site region of the receptor.

With Gold, five different protocols were used (Table 2). Protocols 1, 4 and 5 involved the GoldScore, ChemScore and ChemPLP scoring functions, without rescoring, whereas for the protocols 2 and 3 GoldScore was used for pose prediction, followed by rescoring with ChemScore and ChemPLP. For each docking calculation (training set and SAMPL3-VS dataset) 50 conformations were generated using 100.000 GA operations, the search efficiency was increased to 110% and the algorithm was not allowed to terminate early when the same solution was produced repeatedly. All other parameters had default values. The binding site was defined as a sphere with a 15 Å radius centered on the center of mass of the ligand originally present in the structure 2AYW. Computing time used was about one minute per ligand using a single core on a recent workstation running with Linux CentOS.

With Glide, four protocols (denoted as 6-9 in Table 2) were evaluated, involving the standard SP and XP parameters alone (entries 6 and 7) or in combination with the QPLD procedure (entries 8 and 9). For each docking, 50 conformations were generated using the default values of all parameters. The binding site was defined as a box with the size of 20 Å, centered on the center of mass of the ligand originally present in the structure 2AYW. Computing time used was about a few seconds (SP), a few minutes (XP) or a few hours (QPLD) per ligand using a single core on a recent workstation running with Linux CentOS.

**Table 2.** Docking protocols used in this study

Protocol	Software	Scoring Function	Re-scoring Function
1	Gold	GoldScore	-
2	Gold	GoldScore	ChemScore
3	Gold	GoldScore	ChemPLP
4	Gold	ChemScore	-
5	Gold	ChemPLP	-

6	Glide	GlideScore SP	-
7	Glide	GlideScore XP	-
8	Glide	GlideScore SP	QPLD
9	Glide	GlideScore XP	QPLD

---

## Results and Discussion

The methodology that we followed in this study, in order to propose the best prediction for the SAMPL3-VS challenge, involved the following three steps: i) selection of the most appropriate docking protocols for this task, using the experimental data presently available; ii) fragment virtual screening, which provided a ranked list of compounds from this dataset; iii) evaluation of the virtual screening results in the light of experimental data released by the SAMPL3 organizers at the end of the submission period.

### **Preliminary evaluation of the most appropriate docking protocols for fragment virtual screening on trypsin using available experimental data**

Different docking software generally show unequal ability to predict the bioactive conformation of ligands, depending on the docking protocol that is implemented and on the protein target. From this point of view, the docking of fragment-like compounds is very challenging, given their low binding affinity and the multiple possibilities of positioning in the binding site. In our approach, the selection of the most suitable tools for correct docking predictions represents the key step, which will ensure the quality of the data to be generated.

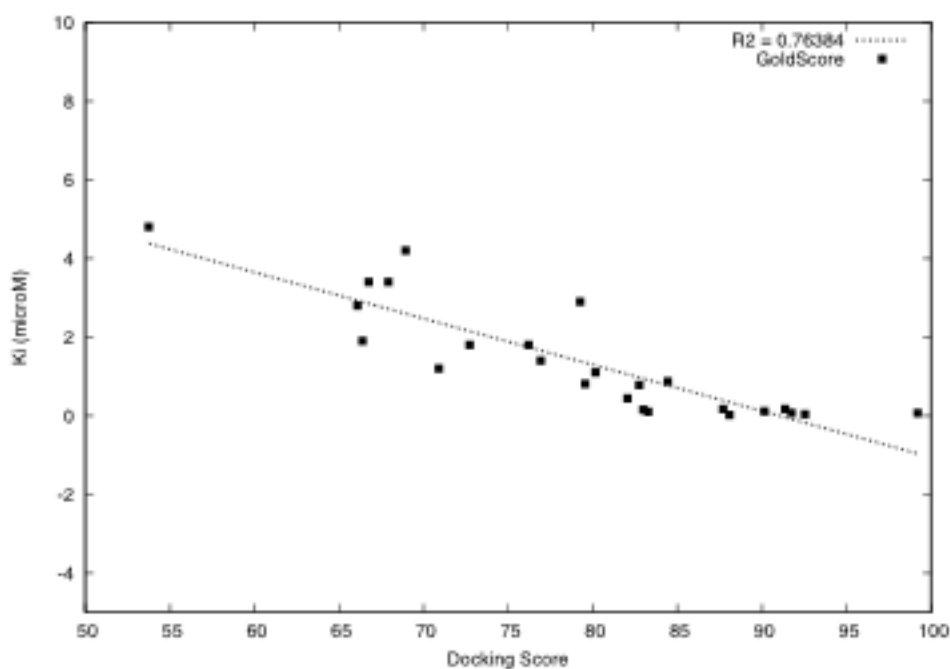
The ability of nine docking protocols (Table 2) in reproducing experimental data has been evaluated using a cross-docking approach. All 25 ligands forming the training set were docked on the crystal structure of bovine trypsin (PDB code 2AYW). The resulting docking (or rescoring) scores were further correlated using linear regression with the experimental inhibition constants ( $K_i$ ) available.

Good results (correlation coefficients  $R^2$  of 0.7-0.8) were obtained only for protocols involving GoldScore in the pose prediction, without rescoring (protocol 1, Figure 1) or with rescoring using ChemScore (protocol 2, Figure 2) or

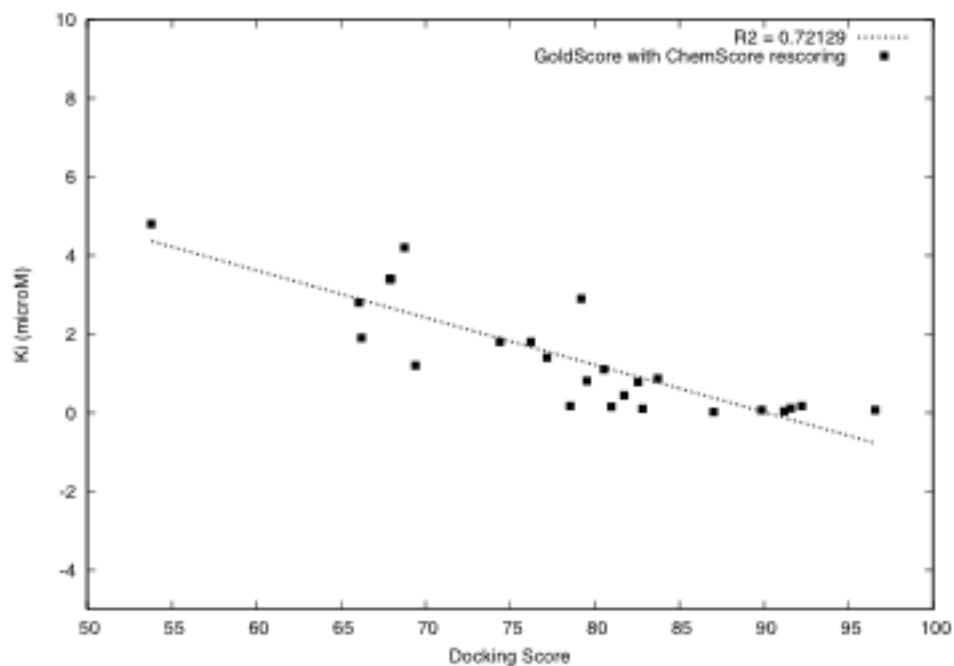
ChemPLP functions (protocol 3, Figure 3). Based on these encouraging results, the three protocols were then used in the subsequent virtual screening step.

Average predictions were obtained with protocols 4 and 5, which employ ChemScore and ChemPLP functions for docking, without rescoring. The protocols 6-9, making use of Glide scoring functions, and even with the computationally expensive protocol QPLD, were unable to predict the order of affinity for these trypsin ligands and no correlations with the experimental data were found (see Figures S1-S6 in Supplementary Information for the correlation plots). In all cases, similar correlation coefficients were obtained using  $\log K_i$  instead of  $K_i$  (data not shown).

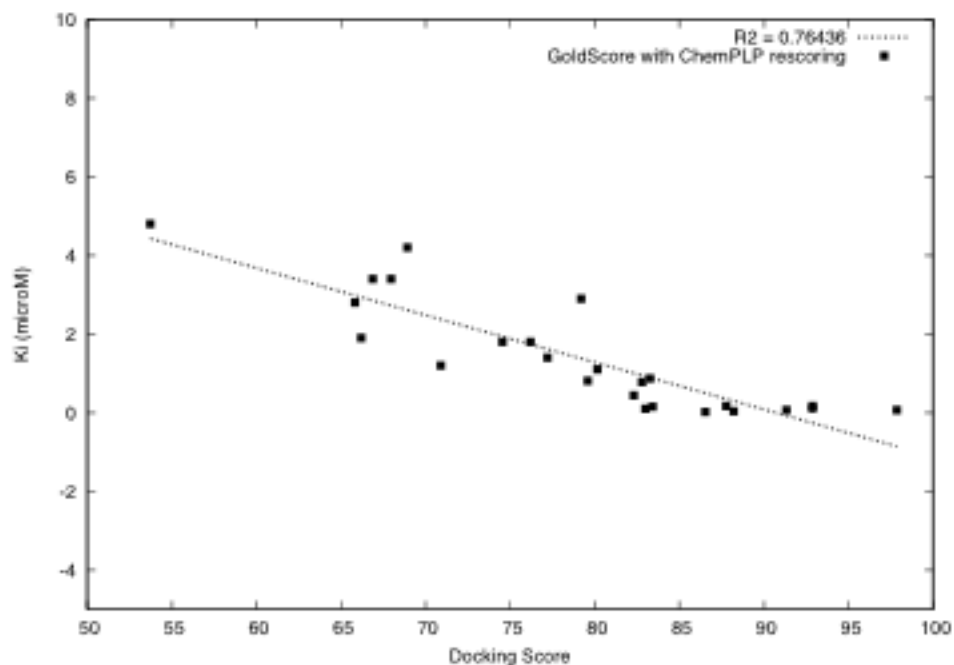
A rapid visual inspection, the comparison of ligand binding modes and the RMSD calculation showed that in most cases Glide is not able to predict the correct pose, whereas in the same conditions Gold succeeds. Once the ligand is correctly positioned in the binding site using the GoldScore function, the differences after rescoring are minor.



**Figure 1.** Correlation between biological activities and cross-docking results (using protocol 1, with GoldScore scoring function and no rescoring) for a training set of trypsin inhibitors.



**Figure 2.** Correlation between biological activities and cross-docking results (using protocol 2, with GoldScore scoring function followed by rescoring with ChemScore) for a training set of trypsin inhibitors.



**Figure 3.** Correlation between biological activities and cross-docking results (using protocol 3, with GoldScore scoring function followed by rescoring with ChemPLP) for a training set of trypsin inhibitors.



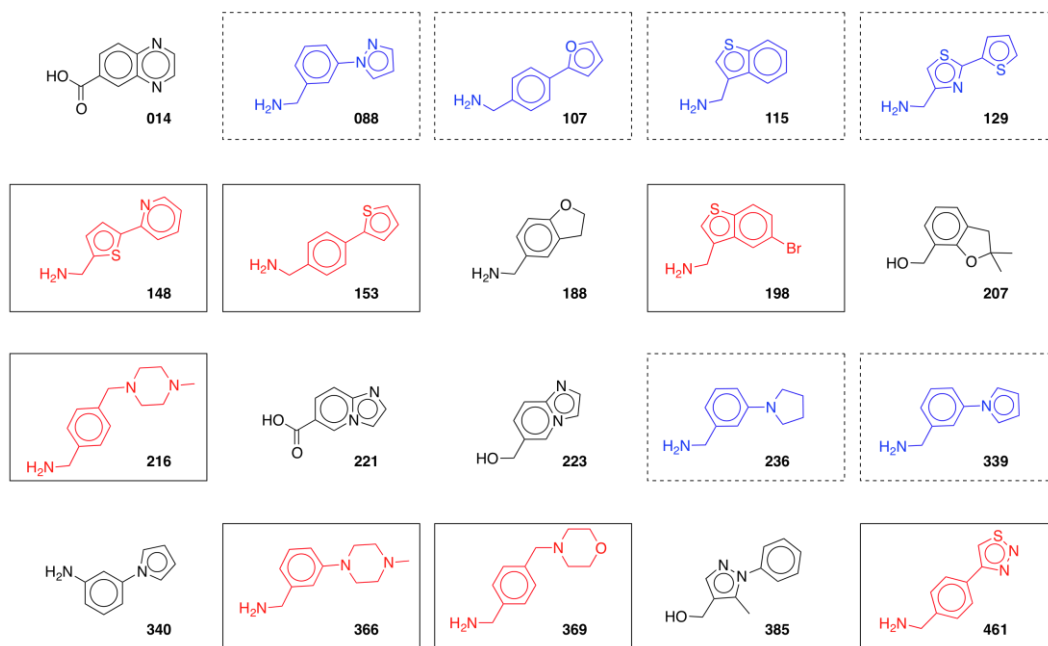
## Fragment virtual screening

The first three protocols (1-3), which were previously identified using the training set as the most suitable for ranking fragment-like trypsin inhibitors, were employed for the virtual screening of the SAMPL3-VS dataset. The resulting three predictions were submitted (as ranked lists) for the SAMPL3 challenge, in order to obtain a "blind" assessment for the accuracy of our selected docking protocols.

Although based on the training set results we were not expecting better predictions, for the sake of completeness the remaining six protocols (4-9) were also evaluated in the virtual screening of the SAMPL3-VS dataset.

## Evaluation of virtual screening results

The structures of the 20 active compounds included in the SAMPL3-VS dataset are depicted in Figure 4. They present a high degree of similarity, with two rings, aromatic or aliphatic, which can be fused or not. Most of them contain an aminomethyl or hydroxymethyl group connected to an aromatic ring, and a few compounds present a carboxyl or amino substituent on an aromatic ring.



**Figure 4.** The 20 active compounds included in the SAMPL3-VS dataset. The compounds colored in red (subset A, plain contour line) were correctly identified in Top 20, whereas compounds colored in blue (subset B, dotted contour line) were ranked in the position range 21-60 by our

virtual screening protocols. The other compounds, colored in black (subset C), were not identified as actives and ranked below the 60<sup>th</sup> position.

Table 3 shows an overview of the virtual screening results obtained using the nine protocols described previously (see also Table S1 in Supplementary Information for more detailed statistics of the results obtained with protocols 1-3). We were pleased to observe that protocols 1-3, which were selected based on the good prediction accuracy for the training set, gave very good results, being able to identify correctly 7 active compounds (out of 20) in Top 20. Additional 6 active compounds were ranked by these protocols in the position range 21-60. In these conditions, the enrichment factor is 9.52 and 5.89 for Top 20 and Top 60, respectively.

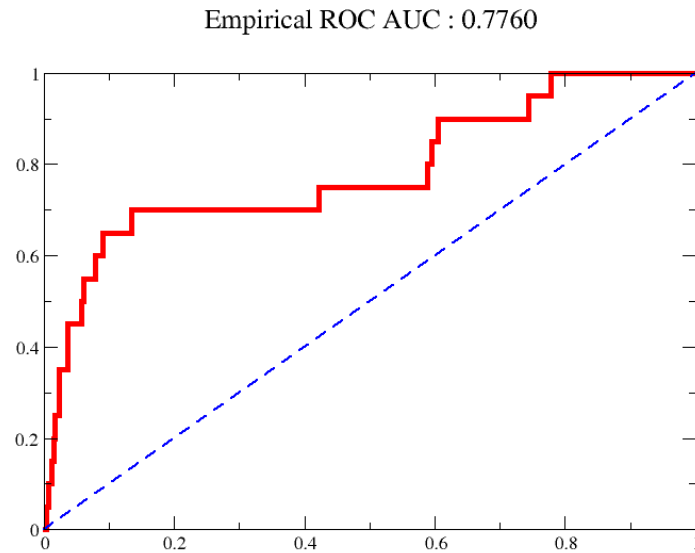
Unexpectedly, even better results were obtained using the protocol 5, which was able to identify correctly 8 and 13 active compounds in Top 20 and Top 60, respectively. All other protocols showed prediction accuracies similar to those obtained with the training dataset.

**Table 3.** General overview of the virtual screening results obtained in this study using the SAMPL3-VS dataset.

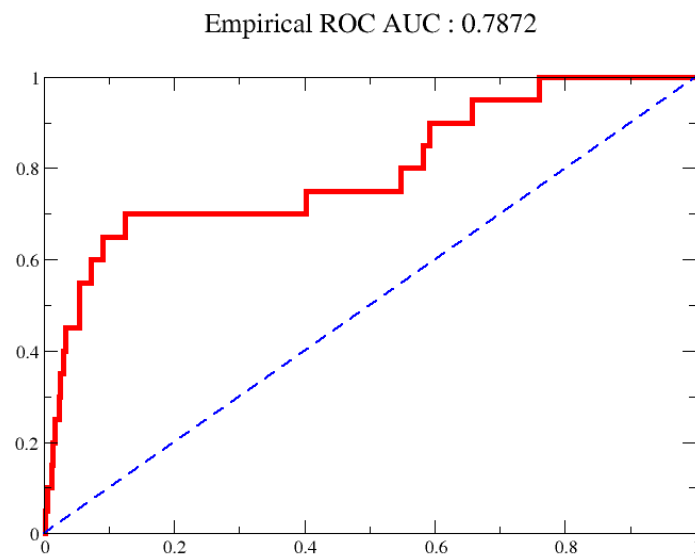
Protocol	Empirical ROC AUC	Top 20		Top 60	
		Actives found	Enrichment factor	Actives found	Enrichment factor
1	0.776	7	9.52	13	5.89
2	0.787	7	9.52	13	5.89
3	0.786	7	9.52	13	5.89
4	0.793	5	6.80	12	5.44
5	0.810	8	10.88	13	5.89
6	0.807	3	4.08	8	3.63
7	0.721	2	2.72	7	3.17
8	0.758	4	5.44	9	4.08
9	0.717	0	0.00	4	1.81

The plots of empirical ROC AUC for the results obtained with protocols 1-3 are shown in Figures 5-7 (see Figures S7-S12 in Supplementary Material for the analogous plots with protocols 4-9). However, from Table 3 it can be observed

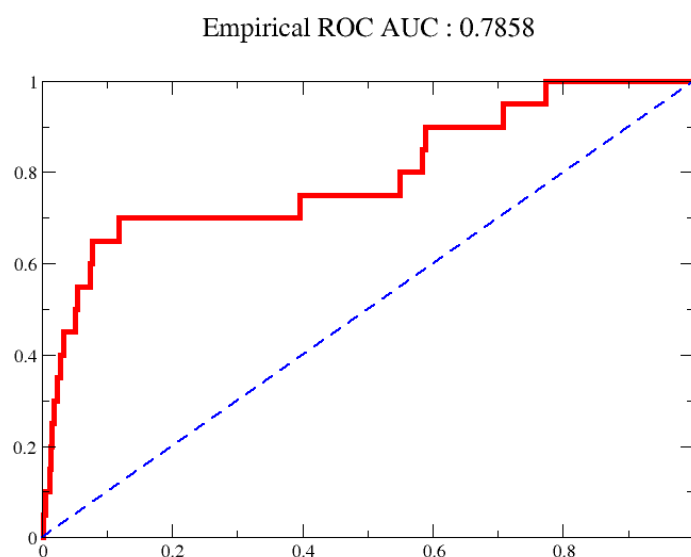
that this metric is not able to evaluate correctly the comparative performance of different protocols, at least not for the Top 20 or Top 60, which are the most important in real life studies. We believe that the enrichment factor could another possible evaluation criteria, which can be possibly used for the assessment of prediction quality in future SAMPL3 challenges.



**Figure 5.** Empirical ROC AUC calculated for results obtained using protocol 1, with GoldScore scoring function (no rescoring).

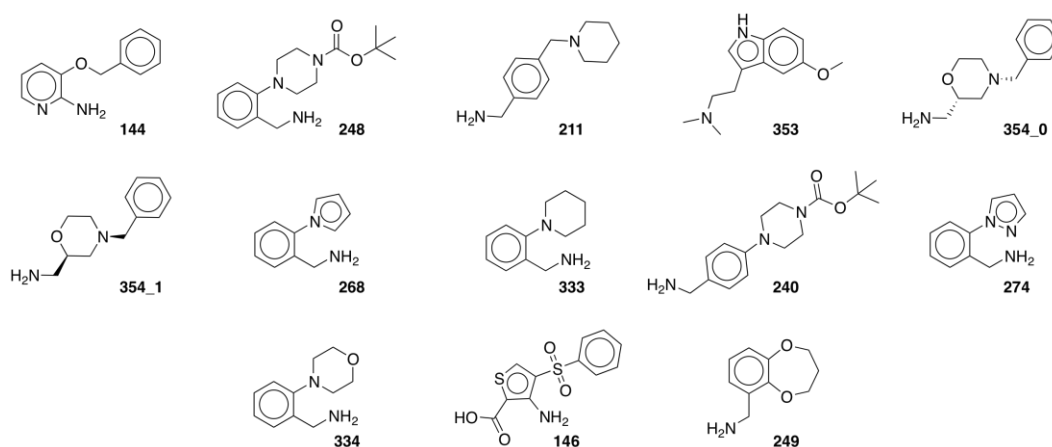


**Figure 6.** Empirical ROC AUC calculated for results obtained using protocol 2, with GoldScore scoring function followed by rescoring with ChemScore.



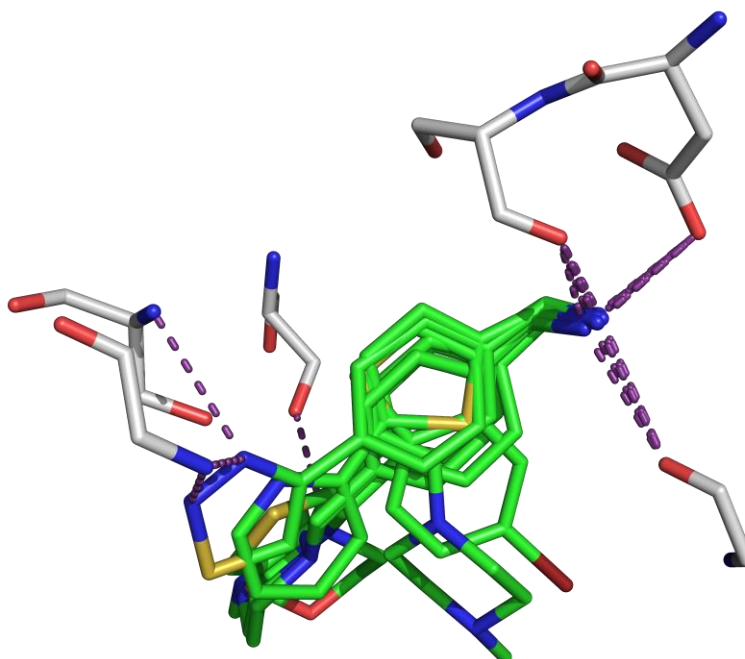
**Figure 7.** Empirical ROC AUC calculated for results obtained using protocol 3, with GoldScore scoring function followed by rescoring with ChemPLP.

A special remark should be made for the compounds that have been incorrectly predicted as actives in Top 20 (Figure 8). We can observe that the structures of these inactive compounds are very similar to those of the active compounds (Figure 4), so it would be an extremely difficult task for a scoring function to discriminate between these two subsets.



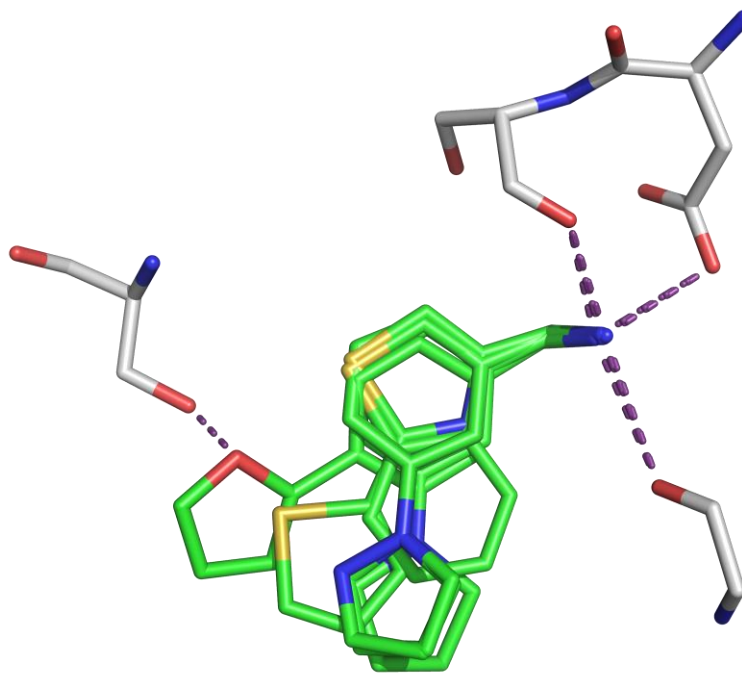
**Figure 8.** The 13 compounds (subset D) incorrectly predicted as actives in Top 20 by the virtual screening protocols used in this study.

From Figure 4 it can be observed that all the active compounds from the subset A contain a terminal amine group, which is protonated under physiological conditions. These compounds were correctly positioned by the protocol 1 in the trypsin S1 pocket, in the vicinity of the residues Asp189, Ser190 and Gly219 (Figure 9).



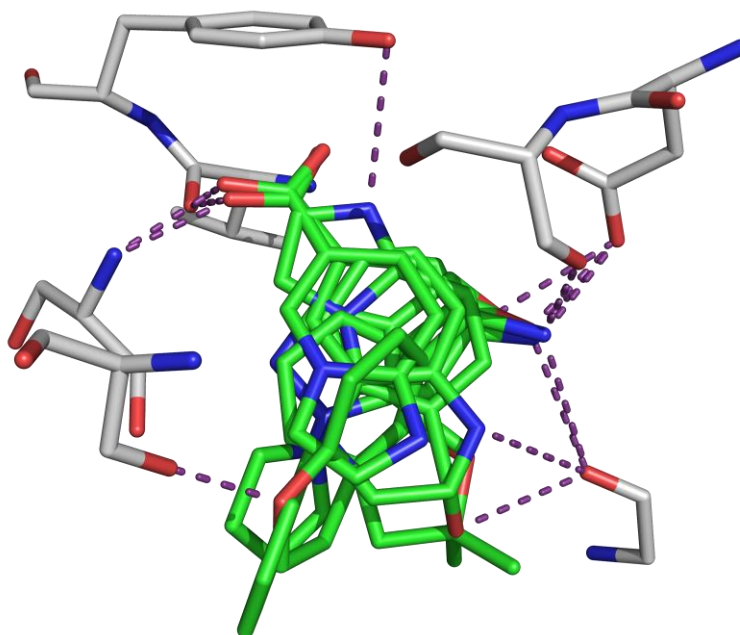
**Figure 9.** Predicted binding modes of compounds that were correctly identified in Top 20 (subset A) using protocol 1

The second group of active compounds, ranked in the position range 21-60 (subset B), was also correctly positioned in the binding site. These compounds were not correctly ranked as active due to the inherent imprecision of the scoring functions.



**Figure 10.** Predicted binding modes of compounds that were ranked in the position range 21-60 (subset B) using protocol 1

From the analysis of the binding modes for the compounds that were not identified as actives and ranked below the 60th position (subset C, Figure 11), several cases can be distinguished: i) compound **188**, which is very similar to all the active compounds that were correctly identified in Top 20, was well positioned in the binding site. One possible explanation for this inaccurate ranking might be the compactness of its structure and the smaller volume, which will require a small rearrangement of the binding site residues to optimize the protein-ligand interactions. This rearrangement is not possible under the conditions used in this study (rigid protein); ii) some of the compounds bearing the hydroxymethyl group (**207**, **385**) and the aniline derivative **340** are positioned in a similar way as the aminomethyl derivatives. The inaccurate ranking is probably due in this case to the error in the estimation of the interaction energy between the hydroxyl group or the aromatic (unprotonated) amine and the trypsin S1 pocket, compared to the positively charged amino group in the case of the other compounds; iii) the compounds **014**, **221** and **223** are positioned in different regions of the binding pocket.



**Figure 11.** Predicted binding modes of compounds that were not identified as actives and ranked below the 60<sup>th</sup> position (subset C) using protocol 1

## Conclusion

In this study, we have "blindly" assessed the ability of several combinations of docking software and scoring functions to predict the binding of a fragment-like library of bovine trypsin inhibitors. The most suitable protocols (involving Gold software and GoldScore scoring function, with or without rescoring) were selected for this purpose using a training set of compounds with known biological activities. The selected virtual screening protocols provided good results with the SAMPL3-VS dataset, showing enrichment factors of about 10 for Top 20 compounds. This methodology should be useful in difficult cases of docking, with a special emphasis on the fragment-based virtual screening campaigns.

## References

1. Huang D, Caflisch A (2011) Fragment-based approaches in virtual screening In: Sotriffer C (ed) Virtual screening, Wiley-VCH Verlag, Weinheim, pp. 467-489.

2. Kawatkar S, Wang H, Czerminski R, Joseph-McCarthy D (2009) Virtual fragment screening: An exploration of various docking and scoring protocols for fragments using glide. *J Comput Aided Mol Des* 23:527-539
3. Sandor M, Kiss R, Keseru GM (2010) Virtual fragment docking by glide: A validation study on 190 protein-fragment complexes. *J Chem Inf Model* 50:1165-1172
4. Joseph-McCarthy D (2009) Challenges of fragment screening. *J Comput Aided Mol Des* 23:449-451
5. Law R, Barker O, Barker J, Hesterkamp T, Godemann R, Andersen O, Fryatt T, Courtney S, Hallett D, Whittaker M (2009) The multiple roles of computational chemistry in fragment-based drug design. *J Comput Aided Mol Des* 23:459-473
6. Warr W (2011) Fragment-based drug discovery: What really works. *J Comput Aided Mol Des* 25:599-605
7. Rabal O, Urbano-Cuadrado M, Oyarzabal J (2010) Computational medicinal chemistry in fragment-based drug discovery: What, how and when. *Future Med Chem* 3:95-134
8. Huang N, Jacobson MP (2010) Binding-site assessment by virtual fragment screening. *PLoS ONE* 5:e10109
9. Schulz M, Landström J, Bright K, Hubbard R (2011) Design of a fragment library that maximally represents available chemical space. *J Comput Aided Mol Des* 25:611-620
10. Chen IJ, Hubbard R (2009) Lessons for fragment library design: Analysis of output from multiple screening campaigns. *J Comput Aided Mol Des* 23:603-620
11. Verdonk ML, Berdini V, Hartshorn MJ, Mooij WT, Murray CW, Taylor RD, Watson P (2004) Virtual screening using protein-ligand docking: Avoiding artificial enrichment. *J Chem Inf Comput Sci* 44:793-806
12. Chen Y, Shoichet BK (2009) Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nature chemical biology* 5:358-364
13. Chen Y, Pohlhaus DT (2010) In silico docking and scoring of fragments. *Drug Discovery Today: Technologies* 7:e149-e156
14. Rummey C, Nordhoff S, Thiemann M, Metz G (2006) In silico fragment-based discovery of dpp-iv s1 pocket binders. *Bioorg Med Chem Lett* 16:1405-1409
15. Friedman R, Caflisch A (2009) Discovery of plasmepsin inhibitors by fragment-based docking and consensus scoring. *ChemMedChem* 4:1317-1326
16. Verdonk ML, Giangreco I, Hall RJ, Korb O, Mortenson PN, Murray CW (2011) Docking performance of fragments and druglike compounds. *J Med Chem* 54:5422-5431
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235-242
18. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide protein data bank. *Nat Struct Mol Biol* 10:980-980
19. Dullweber F, Stubbs MT, Musil Đ, Stürzebecher J, Klebe G (2001) Factorising ligand affinity: A combined thermodynamic and crystallographic study of trypsin and thrombin inhibition. *J Mol Biol* 313:593-614



20. Nar H, Bauer M, Schmid A, Stassen J-M, Wienen W, Pripke HWM, Kauffmann IK, Ries UJ, Huel NH (2001) Structural basis for inhibition promiscuity of dual specific thrombin and factor xa blood coagulation inhibitors. *Structure* 9:29-37
21. Brandt T, Holzmann N, Muley L, Khayat M, Wegscheid-Gerlach C, Baum B, Heine A, Hangauer D, Klebe G (2011) Congeneric but still distinct: How closely related trypsin ligands exhibit different thermodynamic and structural properties. *J Mol Biol* 405:1170-1187
22. Presnell SR, Patil GS, Mura C, Jude KM, Conley JM, Bertrand JA, Kam C-M, Powers JC, Williams LD (1998) Oxyanion-mediated inhibition of serine proteases. *Biochemistry* 37:17068-17081
23. Perilo CS, Pereira MT, Santoro MM, Nagem RAP (2010) Structural binding evidence of the trypanocidal drugs berenil® and pentacarinate® active principles to a serine protease model. *Int J Biol Macromol* 46:502-511
24. Rauh D, Klebe G, Stubbs MT (2004) Understanding protein–ligand interactions: The price of protein flexibility. *J Mol Biol* 335:1325-1341
25. Katz BA, Elrod K, Luong C, Rice MJ, Mackman RL, Sprengeler PA, Spencer J, Hataye J, Janc J, Link J, Litvak J, Rai R, Rice K, Sideris S, Verner E, Young W (2001) A novel serine protease inhibition motif involving a multi-centered short hydrogen bonding network at the active site. *J Mol Biol* 307:1451-1486
26. Katz BA, Elrod K, Verner E, Mackman RL, Luong C, Shrader WD, Sendzik M, Spencer JR, Sprengeler PA, Kolesnikov A, Tai VWF, Hui HC, Breitenbucher JG, Allen D, Janc JW (2003) Elaborate manifold of short hydrogen bond arrays mediating binding of active site-directed serine protease inhibitors. *J Mol Biol* 329:93-120
27. Whitlow M, Arnaiz DO, Buckman BO, Davey DD, Griedel B, Guilford WJ, Koovakkat SK, Liang A, Mohan R, Phillips GB, Seto M, Shaw KJ, Xu W, Zhao Z, Light DR, Morrissey MM (1999) Crystallographic analysis of potent and selective factor xa inhibitors complexed to bovine trypsin. *Acta Crystallogr Sect D* 55:1395-1404
28. Toyota E, Ng KKS, Sekizaki H, Itoh K, Tanizawa K, James MNG (2001) X-ray crystallographic analyses of complexes between bovine  $\beta$ -trypsin and schiff base copper(ii) or iron(iii) chelates. *J Mol Biol* 305:471-479
29. Sherawat M, Kaur P, Perbandt M, Betzel C, Slusarchyk WA, Bisacchi GS, Chang C, Jacobson BL, Einspahr HM, Singh TP (2007) Structure of the complex of trypsin with a highly potent synthetic inhibitor at 0.97 Å resolution. *Acta Crystallogr Sect D* 63:500-507
30. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003) Improved protein–ligand docking using gold. *Protein Struct Funct Bioinf* 52:609-623

## **Supplementary Information**

# **Evaluation of docking performance in a blinded virtual screening of fragment-like trypsin inhibitors**

Georgiana Surpateanu and Bogdan I. Iorga\*

*Institut de Chimie des Substances Naturelles, CNRS UPR 2301, Centre de Recherche de Gif-sur-Yvette, Avenue de la Terrasse, 91198 Gif-sur-Yvette, France*

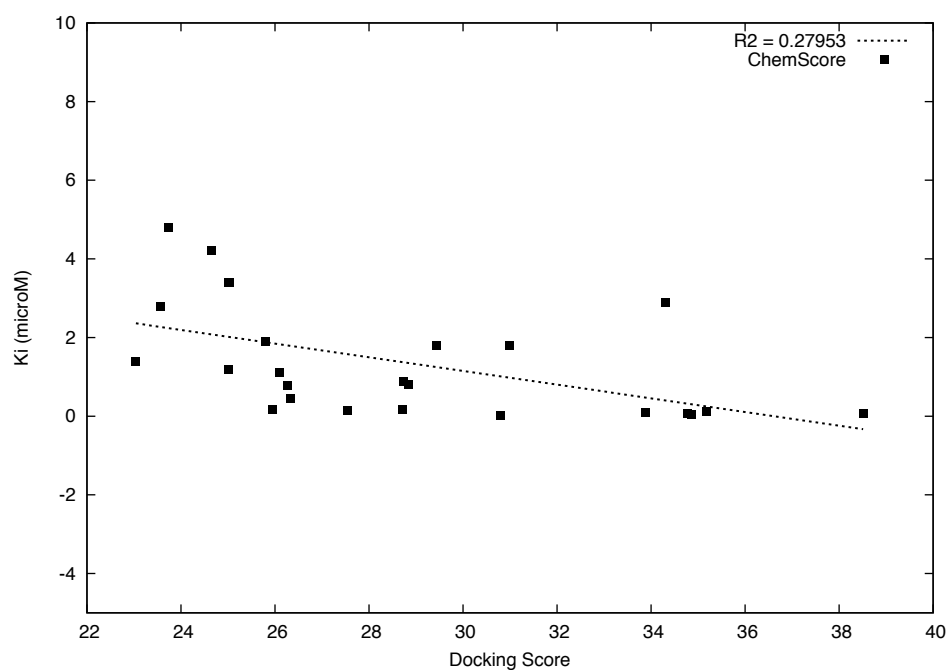
Phone: +33 1 69 82 30 94

Fax: +33 1 69 07 72 47

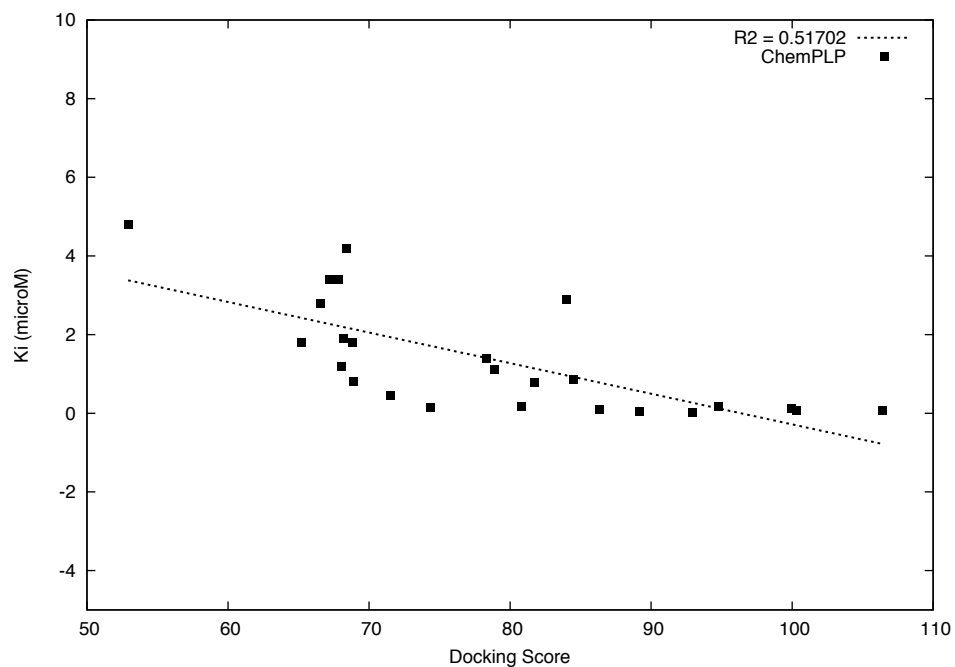
E-mail: bogdan.iorga@icsn.cnrs-gif.fr

## **Table of contents**

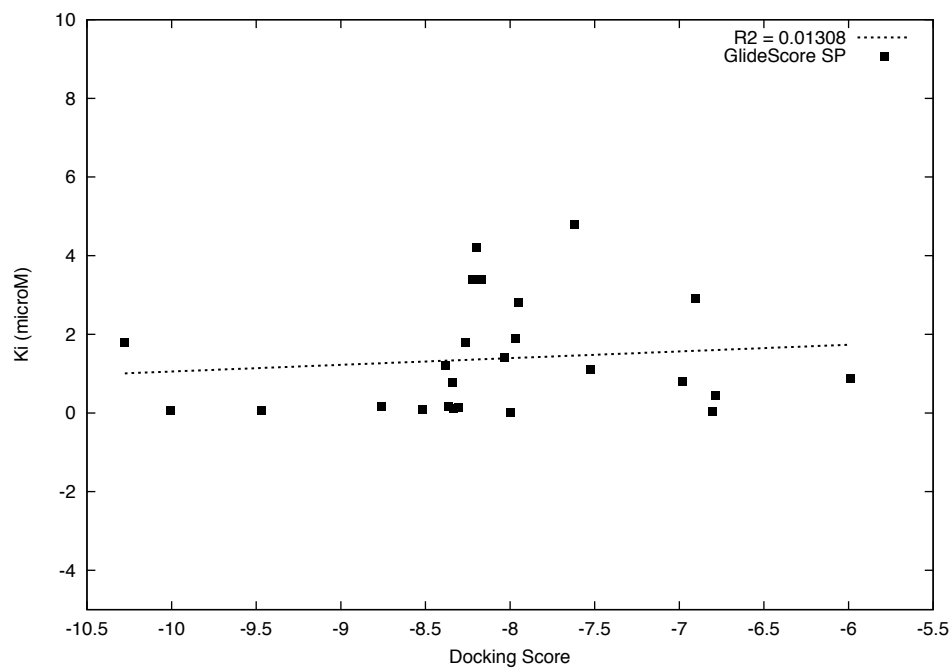
Correlation plots between biological activities and cross-docking results for a training set of trypsin inhibitors using different docking protocols	S2
Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using different docking protocols	S5
Statistics for virtual screening results obtained with protocols 1–3	S8



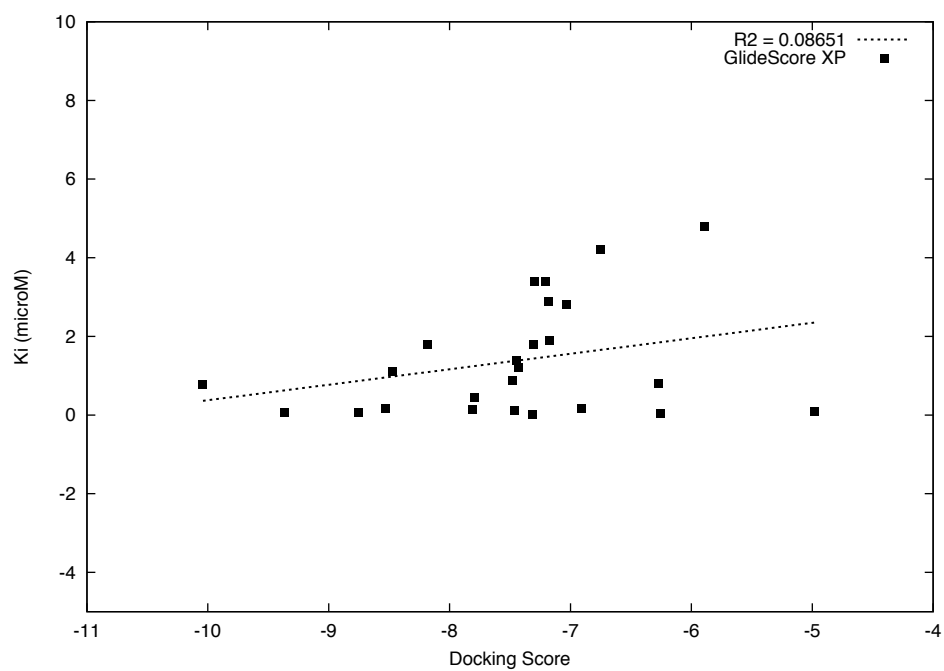
**Figure S1.** Correlation between biological activities and cross-docking results (using protocol 4, with ChemScore scoring function and no rescoring) for a training set of trypsin inhibitors.



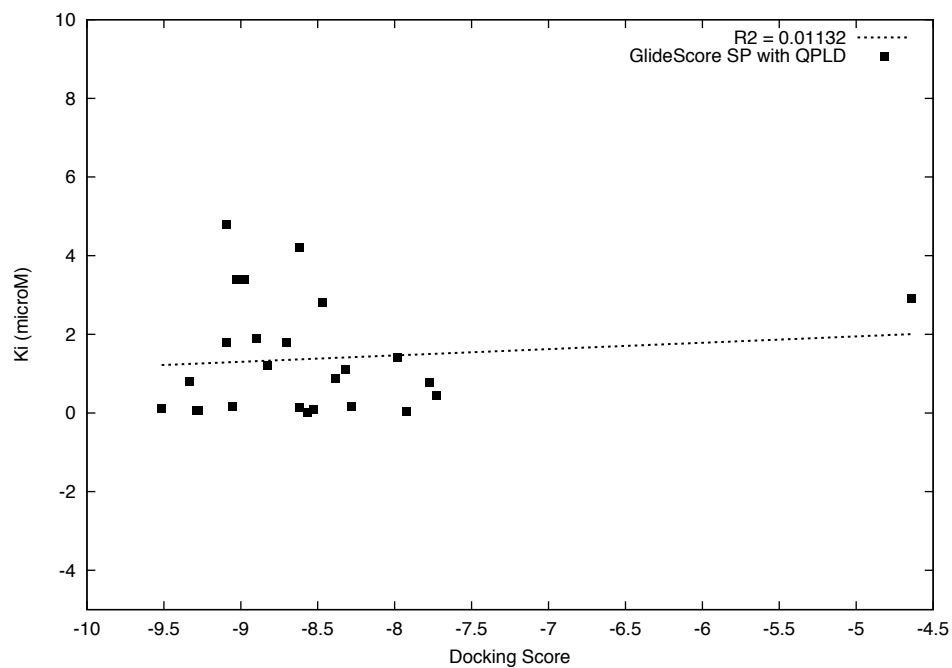
**Figure S2.** Correlation between biological activities and cross-docking results (using protocol 5, with ChemPLP scoring function and no rescoring) for a training set of trypsin inhibitors.



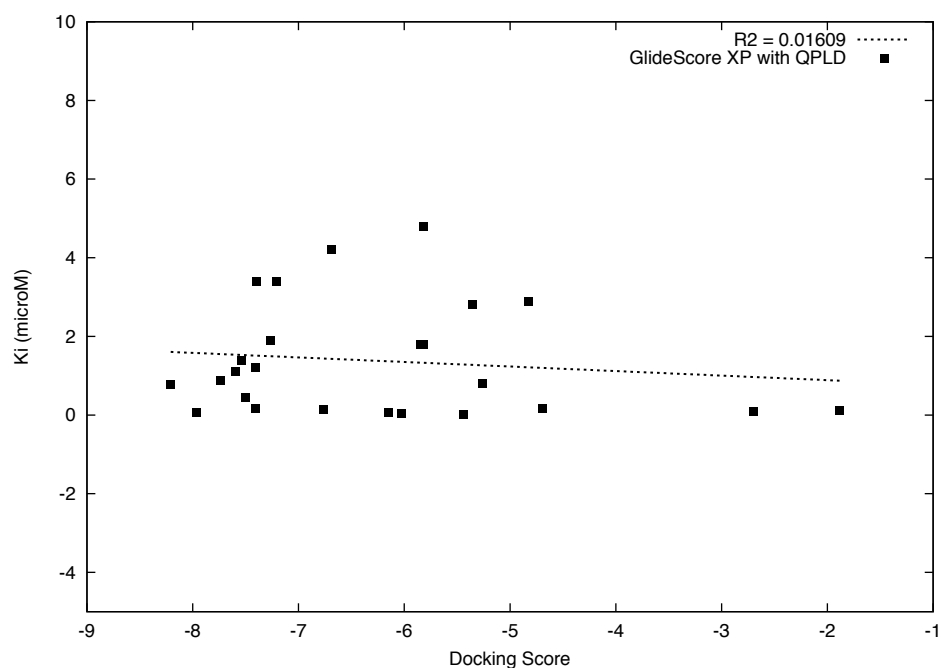
**Figure S3.** Correlation between biological activities and cross-docking results (using protocol 6, with Glide SP scoring function and no rescoring) for a training set of trypsin inhibitors.



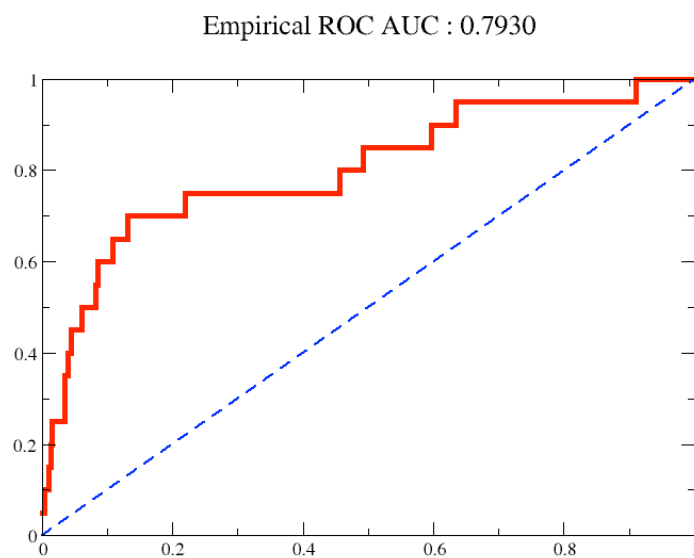
**Figure S4.** Correlation between biological activities and cross-docking results (using protocol 7, with Glide XP scoring function and no rescoring) for a training set of trypsin inhibitors.



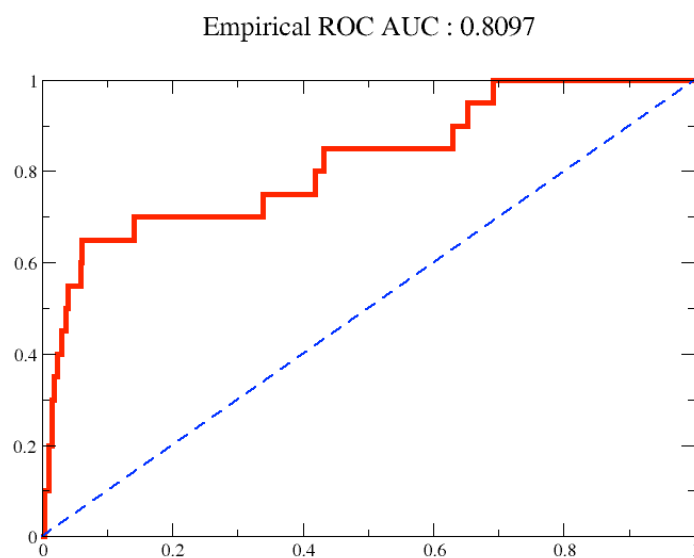
**Figure S5.** Correlation between biological activities and cross-docking results (using protocol 8, with Glide SP scoring function and QPLD) for a training set of trypsin inhibitors.



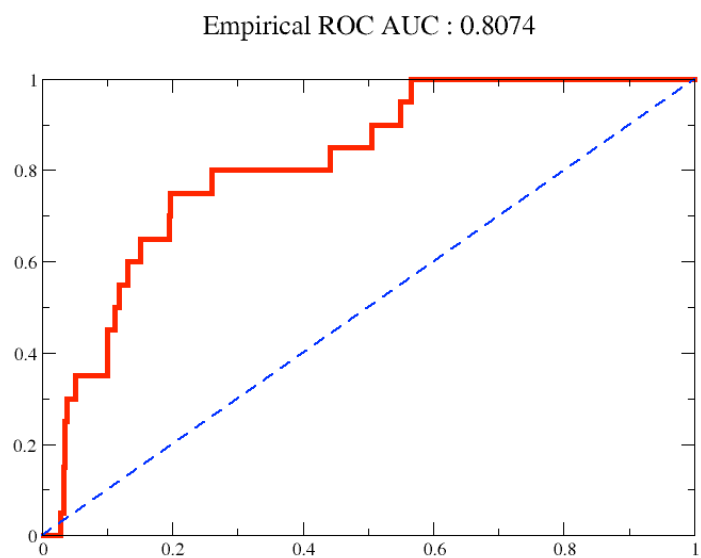
**Figure S6.** Correlation between biological activities and cross-docking results (using protocol 9, with Glide XP scoring function and QPLD) for a training set of trypsin inhibitors.



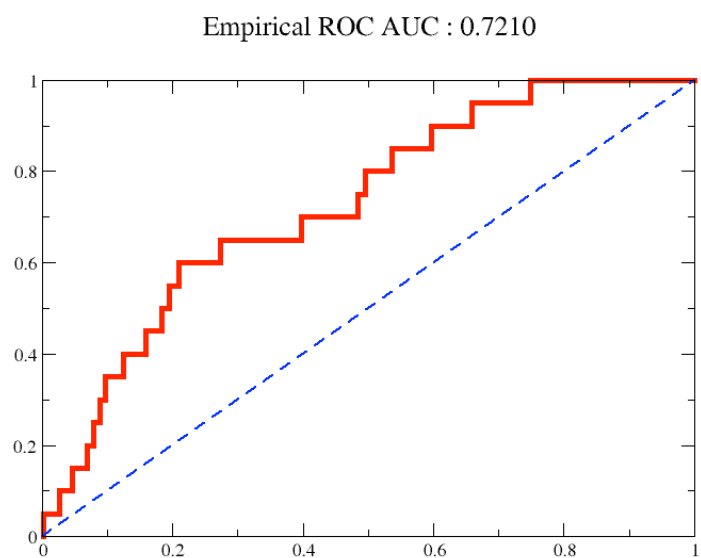
**Figure S7.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 4, with ChemScore scoring function (no rescoring).



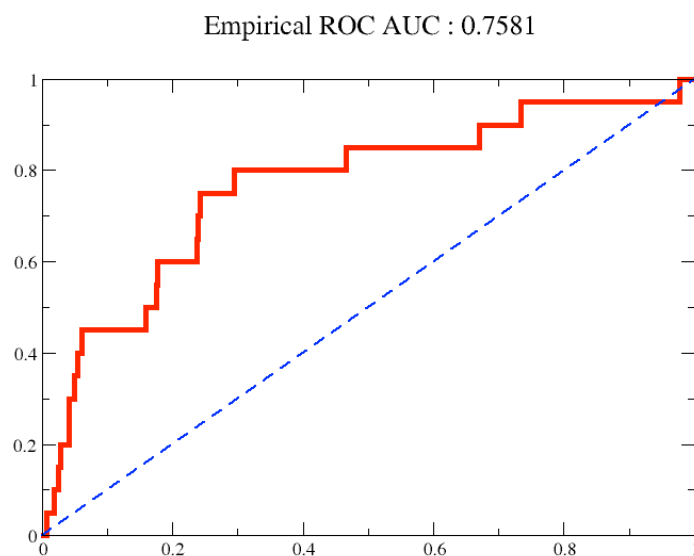
**Figure S8.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 5, with ChemPLP scoring function (no rescoring).



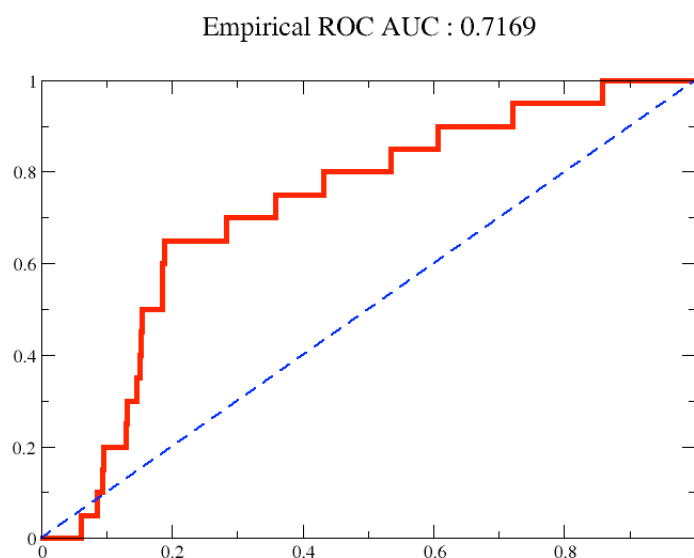
**Figure S9.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 6, with Glide SP scoring function (no rescoring).



**Figure S10.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 7, with Glide XP scoring function (no rescoring).



**Figure S11.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 8, with Glide SP scoring function and QPLD.



**Figure S12.** Empirical ROC AUC calculated for SAMPL3 dataset after virtual screening using protocol 9, with Glide XP scoring function and QPLD.



**Table S1.** Statistics for virtual screening results obtained with protocols 1–3, provided by the SAMPL3 organizers.

	Protocol 1	Protocol 2	Protocol 3
Empirical ROC AUC	0.776	0.788	0.787
ROC AUC Standard Deviation	0.068	0.065	0.066
Empirical 95% CI	0.64 < AUC < 0.91	0.66 < AUC < 0.92	0.66 < AUC < 0.92
Bootstrapped ROC AUC*	0.760 (0.727,0.806)	0.772 (0.741,0.815)	0.772 (0.739,0.816)

\* median (95% confidence values)