View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Universidade do Minho: RepositoriUM

Selective flexibility of side-chain residues improves VEGFR-2 docking score using AutoDock Vina

Rui M. V. Abreu[1,2,*], Hugo J. C. Froufe[1], Maria-João R. P. Queiroz[3] and Isabel C. F. R. Ferreira[1]

Adress:

[1]CIMO-ESA, Instituto Politécnico de Bragança, Campus de Sta Apolónia, Apartado 1172, 5301-855 Bragança, Portugal. [2]Instituto de Biotecnologia e Bioengenharia, Centro de Genómica e Biotecnologia, Universidade de Trás-os-Montes e Alto Douro (CGB-UTAD/IBB), Apartado 1013, 5001-801, Vila Real, Portugal. [3]Centro de Química, Universidade do Minho, Campus de Gualtar 4710-057 Braga, Portugal.

*corresponding author

Rui MV Abreu (adress: CIMO-ESA, Instituto Politécnico de Bragança, Campus de Sta Apolónia, Apartado 1172, 5301-855 Bragança, Portugal, e-mail: ruiabreu@ipb.pt; tel.: +351 273303219; fax: +351 273325405).

Keywords: drug design, virtual screening, aa residues flexibility; VEGFR-2

ABSTRACT

Selective side-chain residue flexibility is an option available on AutoDock Vina docking software. This approach is promising as it attempts to provide a more realistic ligand-protein interaction environment, without an unmanageable increase in computer processing time. However, studies validating this approach are still scarce. VEGFR-2 (vascular endothelial growth factor receptor 2), a known protein target for anti-angiogenic agents, was used in this study. Four residues present in the VEGFR-2 kinase site were selected and made flexible: Lys866, Glu885, Cys917 and Asp1044. The docking scores for all possible combinations of flexible residues were compared to the docking scores using a rigid conformation. The best overall docking scores were obtained using the Glu883 flexible conformation, with pearson and spearman rank correlation values of 0.568 and 0.543, respectively, and a 51% increase in computer processing time. Using different VEGFR-2 X-ray structures a similar trend was observed with Glu885 flexible conformation presenting the best scores. This study demonstrates that careful use of selective side-chain residue flexibility can improve AutoDock Vina docking score accuracy, without a significant increase in computer processing time. This methodology proved to be a valuable tool in drug design when using VEGFR-2 but will also probably be useful if applied to other protein targets.

Introduction

Angiogenesis is the process of new blood vessel formation from pre-existing vascular networks by capillary sprouting (1) and plays an important role in the pathogenesis of several disorders including cancer, proliferative retinopathies and rheumatoid arthritis (2). A key regulatory pathway of angiogenesis is mediated by the vascular endothelial growth factor (VEGF) and its cell membrane tyrosine kinase receptor VEGFR-2 (also know as KDR kinase) (3). Several VEGFR-2 inhibitors have emerged as promising anti-angiogenic agents for possible treatment against a wide variety of cancers. Sorafenib (Bay 43-9006), sunitinib (SU-11248) and pazopanib (GW786034) are VEGFR-2 inhibitors that have been approved for the treatment of advanced renal cell carcinoma (4).

Molecular docking (henceforth referred as docking) and virtual screening are structure-based drug design (SBDD) methods routinely used in modern drug discovery (5). The existence of VEGFR-2 kinase domain crystal structures enabled the use of SBDD tools to investigate new potential inhibitors. Several studies have been published with modeling studies using VEGFR-2 structures, mostly developing pharmacophore based 3D-QSAR models in order to predict the activity of new synthesized compounds (6-9). However the use of docking has been limited to pose inspection of known inhibitors and not to predict VEGFR-2 inhibition activity of small compounds. This is probably because, despite its potential for a wide range of applications, docking continues to face methodological issues and is still not considered accurate enough for virtual screening, as recently reported by Plewczynski *et al.* (10). Probably the most notable difficulty of current docking applications is the use of rigid receptor conformations. Receptor flexibility is still a difficult problem due to the computational challenges posed by the numerous degrees of freedom involved in incorporating protein flexibility when performing docking (11). In fact, it has been shown that, when only a rigid receptor conformation is considered, state-of-the-art docking algorithms predict incorrect binding pose for about 50–70% of all ligands (12). Several docking programs are currently available including DOCK, FlexX, GOLD, AutoDock4 and AutoDock Vina, just to name the more widely used (10). Most of these programs can model full ligand flexibility but only a few take receptor flexibility into account and even then only to a limited extent (13).

In the present work AutoDock Vina (henceforth referred to as ADVina), a freely available program for academic and commercial applications, was used. ADVina provides some degree of protein flexibility by allowing predefined residue side-chains to be flexible during docking (14). An analysis of VEGFR-2 selective residue side-chain flexibility was performed using this program. A dataset of 123 compounds with known VEGFR-2 inhibition activity was used and the Glu885 flexible conformation provided the best results with a significant docking score accuracy improvement over the rigid conformation. A demonstration that Glu885 flexibility improves scoring accuracy is also provided, even when using different VEGFR-2 X-ray structures. This work proves that a careful selection of residue side-chain flexibility may improve robustness of ADVina docking scores, without an unmanageable increase in computer processing time.

Methods and Materials

Dataset of VEGFR-2 inhibitors

A dataset of 123 compounds, covering four log units (pIC_{50} = 5.66–9.70) of VEGFR-2 inhibitory activity, was taken from the literature (15-18) (experimental IC_{50} values for the 123 compounds are available in the results.xls file as supporting information). The 2D structures of the dataset compounds were drawn using ACD/ChemSketch Freeware 12.0 software (19). VegaZZ 2.3.1 (20) software was then used to: convert 2D structures to 3D structures, perform energy minimization and record files in PDB format. AutoDockTools1.5.2 (ADT) (21) was then used to merge nonpolar hydrogens, add gasteiger charges, and set up rotatable bonds. Finally all ligands were recorded in the PDBQT file format used by ADVina (structures in SDF format of the 123 dataset compounds are available in the dataset.sdf file as supporting information).

Preparation of VEGFR-2 flexible conformations

All VEGFR-2 X-ray crystal structures used were extracted from the Protein Data Bank (PDB): 1YWN, 3BE2, 3EWH and 2P2H. Flexible conformations of VEGFR-2 were prepared using the respective crystal structure. The software AutoDockTools was used to extract the co-crystallized ligands from the PDB files, assign polar hydrogens, add gasteiger charges, select the flexible residues side-chains and save the structures in PDBQT file format required to use ADVina.

Docking with ADVina

ADVina 1.0.2 (14) was used to perform docking using all the 15 VEGFR-2 flexible conformations and the rigid conformation. A docking grid with a size of 24 Åx24 Åx24 Å and centred on the coordinates x=1.361, y= 38.562, z= 14.311 was used. The centre coordinate was obtained form the central atom of the co-crystallized inhibitor of each PDB structure. The grid size was selected in order to encompass the co-crystallized inhibitors plus 15 Å on each direction. Due to the large number of docking runs, MOLA software (22) was used to automate all docking runs using a 8 Intel Dual-Core 2.8 GHz computer cluster. ADVina scoring function presents the results as free energy of binding (ΔG). The predicted inhibition constant (predicted Ki) for all docking runs was calculated from the ΔG value as follows: Ki = exp((ΔG*1000)/(Rcal*TK)) where Rcal

is 1.98719 and TK is 298.15 (the predicted Ki values are presented in the results.xls file as supporting information). The R statistical software (23) and the R Commander graphical user interface (24) were used to calculated the Pearson and Spearman rank correlations by comparing the experimental IC_{50} values with the predicted Ki docking scores (Table 1). All the figures with structure representations were produced using the PyMOL software (2**Results and Discussion**

Selection of VEGFR-2 flexible residues

A statistical analysis of the PDB databank revealed that 85% of proteins contain one to three flexible residues in the active site (26), and the inability to consider this flexibility is probably related to the limitations of the flexible ligand-rigid receptor approach. ADVina docking tool provides the possibility to select specific residues as flexible, allowing rotation around torsional degrees of freedom (27). In this work, we set out to analyze if the flexibility options on ADVina can improve docking accuracy using VEGFR-2 as protein target. VEGFR-2 can be considered as a good candidate for this type of flexibility study as VEGFR-2 structure presents a relatively rigid conformation when comparing the available X-ray structures. Furthermore, VEGFR-2 does not engage in any major conformational alterations when binding to different known inhibitors (Figure 1).

By comparing the kinase binding pocket of different VEGFR-2 crystal structures, we started by selecting residues with some degree of cross-structure flexibility (Figure 1). The main issue to be aware of at this stage is an increase in degrees of freedom, resulting in more computational time need to perform the docking runs (table 1). From our experience using ADVina, flexibilizing many residue side-chains can result in unpredictable results as it tends to introduce a bias towards lower energies of binding, with a consequent increase in false positives. Also, more flexible residue side-chains results in an increase in the computational processing time need to perform the docking runs. Taken these factors into account, a maximum of 4 flexible residue side-chains was established. After inspecting the kinase active site of several VEGFR-2 structures, 4 residues were selected: Lys868, Glu885, Cys919 and Asp1046. As can be seen by the superimposition of 4 VEGFR-2 crystal structures, the selected residues present some degree of flexibility (Figure 1). Also these residues interact

consistently, by forming hydrogen bonds (H-bonds) with the inhibitors that are co-crystallized in the inspected VEGFR-2 structures.

After an extensive survey of the available PDB structures, the VEGFR-2 crystal structure PDB:1YWN was selected (28). This structure was chosen as it presents low resolution (1.72 A) and has been consistently used in several SBDD studies (6,7). Furthermore, the co-crystallized inhibitor 4-amino-furo[2,3-*d*]pyrimidine interacts with all the selected residues (Figure 1).

Evaluation of ADVina scoring using VEGFR-2 rigid and flexible conformations

A standard docking study yields two main results: a docking pose of the interaction between the ligand and the protein target, and a docking score that estimates the strength of the protein–ligand interaction. This study focuses on the docking score element and evaluates if VEGFR-2 residue side-chain flexibilization is able to produce a significant enrichment in ADVina scoring. A dataset of 123 compounds, with known VEGFR-2 inhibition activity spanning 4 orders of magnitude, was docked against the rigid conformation and against 15 combinations of VEGFR-2 flexible residue side-chains conformations (flexible conformations). The docking scores were then compared and correlated with the experimental binding affinity values of the dataset (Table 1). The docking score enrichment was evaluated using several parameters: pearson correlation, spearman rank correlation and docking computational processing time. Especially in virtual screening, the spearman rank correlation is an important statistical parameter as the position of the compounds within an ordered list is usually more valuable information than the actual docking scores. As more flexible residues are selected, the time need to perform the docking runs should be taken in consideration as there is an increase in the number of rotatable bonds and consequent increase in computational processing time (Table 1).

The pearson and spearman rank coefficients obtained using the rigid conformation were 0.223 and 0.456, respectively. These values are in agreement with a recent study using different proteins and docking softwares (10), were eHits software (averaging the results for all the proteins used in the study) was the top performer, with pearson and spearman rank coefficients of 0.380 and 0.470, respectively. When using single flexible residue side-chain conformations, the Glu885 flexible conformation presented the best scores with a sharp pearson coefficient increase (0.568) and a significant spearman rank coefficient increase (0.543). As expected, due to an increase

in three rotatable bonds, the Glu885 flexible conformation resulted in a 51 % increase in processing time (table 1). Single Lys868, Cys919 or Asp1046 flexible conformations did not produce a consistent increase in scoring enrichment. When using two, three or all four flexible residue side-chain conformations, only conformations including the flexible Glu885 improved pearson or spearman rank correlations. The Lys868-Glu885 flexible conformation presented higher pearson correlation (0.634) but lower spearman rank correlation (0.417), while Glu885-Cys919 flexible conformation presented higher spearman rank correlation (0.580) but lower pearson correlation (0,560), although with a 112% increase in processing time. Of note is that the all four residue flexible conformation presented lower spearman rank coefficients (0,436) when compared to the rigid conformation. This analysis demonstrates that careful selection of flexible residue side-chains can increase the quality of ADVina docking scoring, with an acceptable payoff in processing time. Nevertheless, a blind increase in the number of flexible residue side-chains does not translate into an automatic increase in ADVina docking score enrichment. In fact, the results confirm our earlier impression when using ADVina that, as the number of flexible residue side-chains increase, there is a threshold where the quality of the docking scores does not increase or even decrease. This is probably due to an increase in false positives as a consequence of an easier accommodation of compounds made possible by the flexible residue side-chains. Although this work has been performed using VEGFR-2, we believe that this flexible docking methodology can be applied using other protein targets to improve docking poses and docking scores.

Table 1. Results for ADVina scoring using rigid and flexible conformations.

Flexible Side-Chain Residues	Pearson Correlation	Spearman Rank Correlation	Spearman p-value	Processing Time % (a)	Rotatable bonds
RIGID	0,23	0,456	1,13E-007	100	0
ASP	0,219	0,381	1,39E-005	154	2
CYS	0,389	0,421	1,22E-006	126	1
LYS	0,307	0,356	5,36E-005	232	4
GLU	0,568	0,543	8,78E-011	151	3
ASP CYS	0,201	0,247	0.005914	147	3
ASP LYS	0,352	0,335	0.0001488	327	6
ASP GLU	0,410	0,555	2,68E-011	258	5
CYS LYS	0,204	0,322	0.0002823	275	5
CYS GLU	0,560	0,584	1,28E-012	212	4
LYS GLU	0,634	0,477	2,52E-008	378	7
CYS LYS GLU	0,549	0,531	2,68E-010	437	8
ASP LYS GLU	0,249	0,525	4,65E-010	525	9

ASP CYS GLU	0,518	0,427	8,65E-007	311	6
ASP CYS LYS	0,315	0,293	0.001023	374	7
ASP CYS LYS GLU	0,350	0,436	4,56E-007	616	10

(a) Average processing time for the rigid conformation was 1 minute per ligand that corresponds to 100% processing time.

To investigate if flexible Glu885 docking scores improvement was independent of the crystal structure used, the same methodology was applied to 3 different VEGFR-2 crystal structures available: 2P2H, 3BE2 and 3EWH (Table 2). The structures were selected according to: structure resolution (lower than 2 Å) and presence of different inhibitors. In all crystal structures used, a consistent improvement in pearson and spearman rank correlations was observed when using the Glu885 flexible conformation compared to the rigid conformation (Table 2). On the other hand, no consistent trend was observed with Glu885-Cys919 and Glu885-Lys868 flexible conformations. In fact all structures presented lower pearson and spearman rank correlation, with the notable exception of Glu885-Lys868 flexible conformations when using the 3BE2 structure. This study demonstrates that the Glu885 flexible conformation provides the best overall docking scores, using less computer processing time and in a crystal structure independent manner.

Table 2. Results for ADVina scoring using different VEGFR-2 X-ray structures: 1YWN, 2P2H, 3BE2 and 3EWH; and different flexible residue conformations: Rigid, GLU and Glu885-Cys919 and Glu885-Lys868 flexible conformations in.

	Pearson Correlation				**Spearman Correlation**			
PDB entry	**Rigid**	**GLU**	**GLU CYS**	**GLU LYS**	**Rigid**	**GLU**	**GLU CYS**	**GLU LYS**
1YWN	0,230	0,568	0,560	0,634	0,456	0,543	0,584	0,477
2P2H	0,342	0,51	0,476	0,266	0,296	0,382	0,374	0,197
3BE2	0,466	0,498	0,472	0,53	0,421	0,467	0,438	0,542
3EWH	0,256	0,625	0,522	0,372	0,469	0,514	0,507	0,449

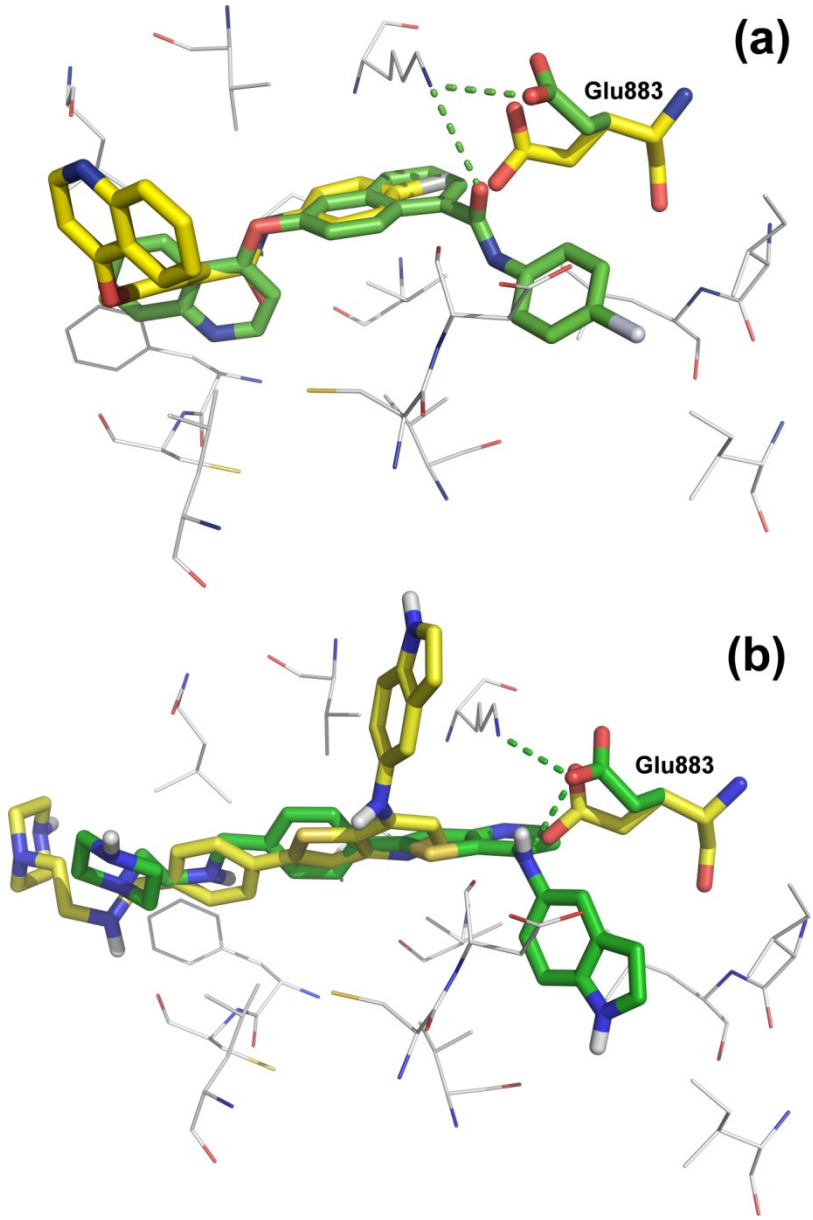

Figure 2. Docking pose superimposition of two dataset compounds, (a) A8 and (b) A117, using the rigid (yellow) and Glu885 flexible (green) VEGFR-2 conformations. H-bonds represented in traced green.

To better understand the scoring enrichment promoted by flexibilization of Glu885, docking pose inspections of the dataset compounds were performed. In general it was observed that Glu883 flexibilization improved docking score accuracy by enabling compounds to completely enter the VEGFR-2 kinase binding pocket. This was made possible by the accommodation of the Glu883 side-chain and, in almost all cases, by the formation of new H-bonds. Figure 2 shows two representative dataset compounds, A8 and A117, were Glu885 flexibilization enabled the positioning of both compounds completely inside the VEGFR-2 binding pocket. These compounds

represent the two different H-bonds configurations observed in the dataset: an H-bond between the compounds and the amino group of the rigid Lys868 side-chain (Fig. 2a), or an H-bond between the compounds and the carboxyl group of the flexible Glu885 (Fig. 2b).

Conclusions and future work

In this work, a docking study of selective residue side-chain flexibility was performed using the tyrosine kinase receptor VEGFR-2 as protein target, and ADVina as docking software. It was demonstrated that an important docking score improvement can be obtained with careful binding site analysis and selection of relevant residues. The Glu885 flexible conformation provided the best docking scores and this improvement was observed across all the X-ray structures used. This methodology may be a valuable tool in drug design when performing virtual screening, either using VEGFR-2 or applied to other therapeutic protein targets. To our knowledge it is the first time that an in-depth receptor-flexibility docking study software was performed using ADVina.

Acknowledgments

The authors are grateful to Foundation for Science and Technology (Portugal) and COMPETE/QREN/EU for financial support through research project PTDC/QUI-QUI/111060/2009 and Rui M.V. Abreu thanks to SFRH/PROTEC/49450/2009 grant.

Figures

Figure 1. Superimposition of the VEGFR-2 selected residues from 4 crystal structures: 1YWN (sticks and balls representation), 3BE2, 3EWH, 2P2H (line representation). Light grey the cartoon representation of 1YWN, line green the 1YWN co-crystallized ligand.

References

1. Patan S. (2001) Vasculogenesis and angiogenesis as mechanisms of vascular network formation, growth and remodelling. J Neurooncol 50: 501-15.

2. Folkman J., Shing Y. (1992) Angiogenesis. J Biol Chem 267: 10931-10934.

3. Lemmon M., Schlessinger J. (2010) Cell signalling by receptor tyrosine kinases. Cell 141: 1117-34.

4. Eichholz A., Merchant S., Gaya A. M. (2010) Anti-angiogenesis therapies: their potential in cancer management. Onco Targets Ther 24: 369-82.

5. Kitchen D. B., Decornez H., Furr J. R., Bajorath J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3: 935-49.

6. Lee K., Jeong K.-W., Lee Y., Song J. Y., Kim M. S., Lee G. S., Kim Y. (2010) Pharmacophore modeling and virtual screening studies for new VEGFR-2 kinase inhibitors. Eur J Med Chem 45: 5420-5427.

7. Pasha F., Muddassar M., Neaz M. M., Cho S. J. (2009) Pharmacophore and docking-based combined in-silico study of KDR inhibitors. J Mol Graph Model 28: 54-61.

8. Neaz M. M., Pasha F. A., Muddassar M., Lee S. H., Sim T., Hah J.-M., Cho S. J. (2009) Pharmacophore based 3D-QSAR study of VEGFR-2 inhibitors. Med Chem Res 18: 127-142.

9. Du J., Lei B., Qin J., Liu H., Yao X. (2009) Molecular modelling studies of vascular endothelial growth factor receptor tyrosine kinase inhibitors using QSAR and docking. J Mol Graph Model 27: 642-54.

10. Plewczynski D, Laźniewski M, Augustyniak R, Ginalski K. (2010) Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem 32: 742-55.

11. B-Rao C., Subramanian J., Sharma S. D. (2009) Managing protein flexibility in docking and its applications. Drug Discov Today 14: 394-400.

12. Totrov M., Abagyan R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 18: 178-84.

13. Cavasotto C. N., Abagyan R. A. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 337: 209-25.

14. Trott O., Olson A. J. (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31: 455-461.

15. Munchhof M. J., Beebe J. S., Casavant J. M. , Cooper B. A., Doty J. L., Higdon R. C., Hillerman S. M., Soderstrom C. I., Knauth E. A., Marx M. A., Rossi A. M., Sobolov S. B., Sun J. (2004) Design and SAR of thienopyrimidine and thienopyridine inhibitors of VEGFR-2 kinase . Bioorg Med Chem Lett 14: 21-24.

16. Weiss M. M., Harmange J.-C., Polverino A. J., Bauer D., Berry L., Berry V., Borg G., Bready J., Chen D., Choquette D., Coxon A., DeMelfi T., Doerr N., Estrada J., Flynn J. *et al.* (2008) Evaluation of a series of naphthamides as potent, orally active vascular endothelial growth factor receptor-2 tyrosine kinase inhibitors. J Med Chem 51:1668-80.

17. La D. S., Belzile J., Bready J. V., Coxon A., DeMelfi T., Doerr N., Estrada J., Flynn J. C., Flynn S. R., Graceffa R. F., Harriman S. P., Larrow J. F., Long A. M., Martin M. W., Morrison M. J. *et al.* (2008) Novel 2,3-dihydro-1,4-benzoxazines as potent and orally bioavailable inhibitors of tumor-driven angiogenesis. J Med Chem 51: 1695-705.

18. Harmange J.-C., Weiss M. M., Germain J., Polverino A. J., Borg G., Bready J., Chen D., Choquette D., Coxon A., DeMelfi T., DiPietro L., Doerr N., Estrada J., Flynn J., Graceffa R. F. *et al.* (2008) Naphthamides as novel and potent vascular endothelial growth factor receptor tyrosine kinase inhibitors: design, synthesis, and evaluation. J Med Chem 51: 1649-67.

19. ACD/ChemSketch Freeware 12.0 software. [Internet]. Available from http://www.acdlabs.com/resources/freeware/.

20. Pedretti A., Villa L., Vistoli G. (2004) VEGA--an open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming. J Comput Aided Mol Des 18: 167-73.

21. Sanner M. F. (2005) A component-based software environment for visualizing large macromolecular assemblies. Structure 13: 447-62.

22. Abreu R. V., Froufe H. J. C., Queiroz M.-J. R. P., Ferreira I. C. F. R. (2010) MOLA: a bootable, self-configuring system for virtual screening using AutoDock4/Vina on computer clusters. J Cheminform 2: 10.

23. The R Project for Statistical Computing [Internet]. Available from: http://www.r-project.org/.

24. Fox J. (2005) The R Commander: A Basic-Statistics Graphical User Interface to R. J Statistic Soft 14: 1-42.

25. The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC. [Internet]. Available from: http://www.pymol.org/.

26. Najmanovich R., Kuttner J., Sobolev V., Edelman M. (2000) Side-chain flexibility in proteins upon ligand binding Proteins 39: 261-8.

27. Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D.S., Olson A. J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem 30: 2785-91.

28. Miyazaki Y., Matsunaga S., Tang J., Maeda Y., Nakano M., Philippe R. J., Shibahara M., Liu W., Sato H., Wang L., Nolte R. T. (2010) Novel 4-amino-furo[2,3-*d*]pyrimidines as Tie-2 and VEGFR-2 dual inhibitors. Bioorg Med Chem Lett 15: 2203-7.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

File dataset.sdf. Chemical structures of all the 123 known VEGFR-2 inhibitors in the dataset, provided in SDF file format.

File results.xls. Spreadsheet presenting the experimental IC_{50} of all the 123 known VEGFR-2 inhibitors in the dataset. The predicted Ki of the dataset compounds for the fifteen flexible conformations and the rigid conformation are also presented.