SpringerPlus
a SpringerOpen Journal

SOFTWARE

Open Access

CrossMark

CrossDocker: a tool for performing cross-docking using Autodock Vina

Jamal Shamsara[*]

Abstract

Background: Cross-docking is an approach to find the best holo structures among multiple structures available for a target protein.

Results: CrossDocker significantly decreases the time needed for setting parameters and inputs for performing multiple dockings, data collection and subsequent analysis.

Conclusion: CrossDocker was written in Python language and is available as executable binary for Windows operating system. It is available at http://www.pharm-sbg.com. Some example data sets were also provided.

Keywords: Autodock Vina, Cross-docking, Virtual screening
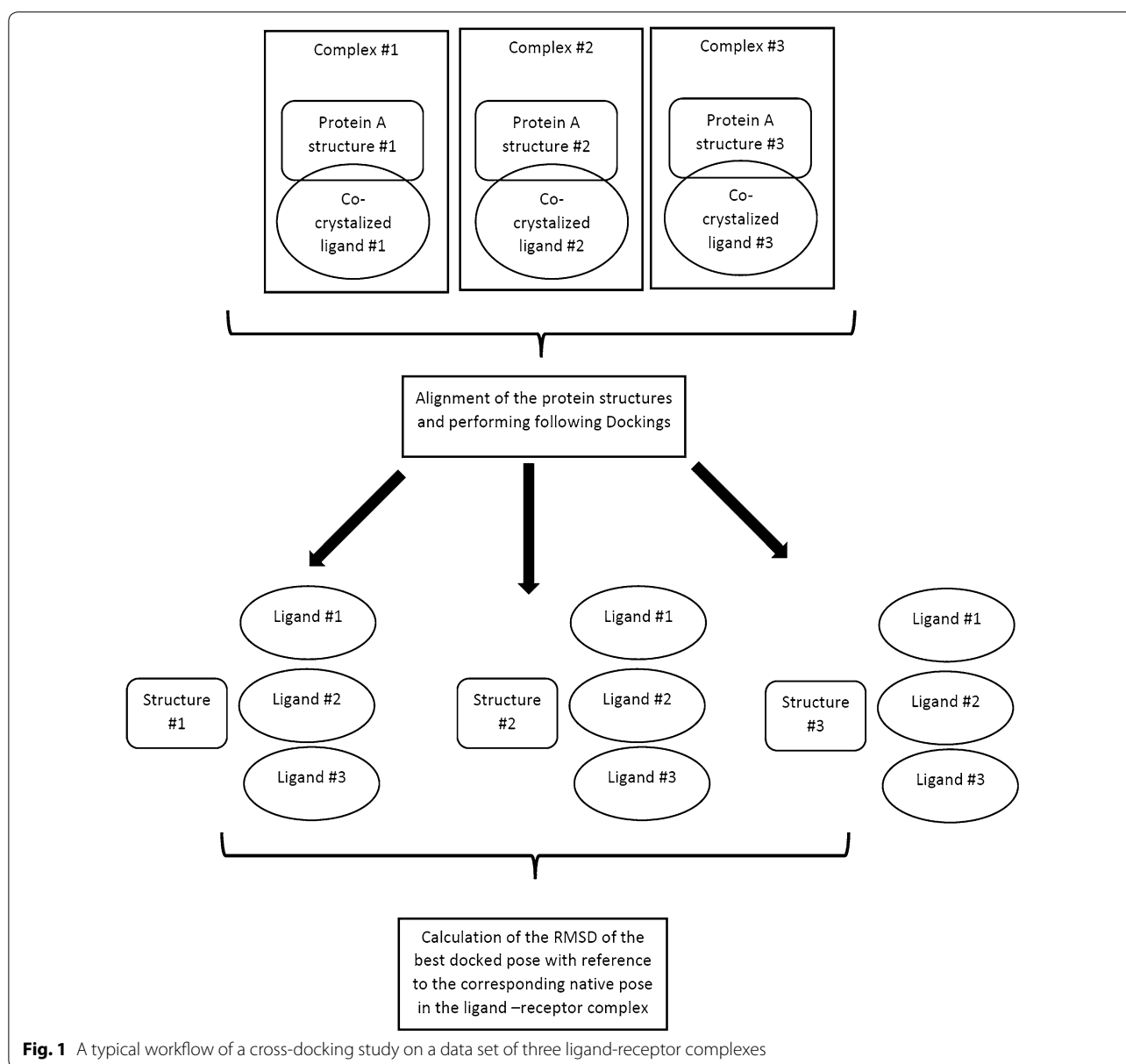
Background

Structural based virtual screenings are widely used for identification of new lead compounds for specific targets that their experimental 3D structure is available (Kubinyi 2006). The successfulness of such virtual screenings greatly depends on the quality of the available 3D structure of the receptor (Pitt et al. 2013). In general, holo structures perform better than apo structures. In case of multiple available 3D structures for a receptor, selection of the best structure for pose prediction and virtual screening is an important issue (Mohan et al. 2005). They are several suggested methods for selection of the best structure (Hawkins et al. 2008). One is the selection according to the specification of the X-ray crystallography. The most prominent one is the resolution of the X-ray crystal structure. Others are R-factor and average B-factor. For example it was proposed that the quality of a crystal structure can be estimated considering resolution and R-factor (Sacan et al. 2012).

However, these metrics are not absolute and it was proposed (Vinh et al. 2012; Ramezani and Shamsara 2015) that solely considering these structural parameters of an x-ray

crystal structure cannot properly predict the performance of a 3D structure in virtual screenings, especially in case of receptors with flexible active site. Flexible receptors could exhibit either intrinsic or induced flexibility (Chandrika et al. 2009). Different ligand can induce different conformational changes in the active site residue of a receptor upon binding. Thus, a single ligand-receptor complex solved in the presence of a specific ligand may have a lower affinity for another ligand (with different scaffold). In the other words, the chemical characteristics (size, functional groups, etc.) of the co-crystalized ligand is also important for the applicability of a flexible 3D structure in virtual screenings (Ramezani and Shamsara 2015; Zhang et al. 2014). For example, a flexible binding pocket which is arranged to have interactions with a small ligand (in a crystal structure) cannot easily accommodate binding of larger ligands (in a virtual screening). These can be indirectly determined by a retrospective virtual screening on a predefined set of active compounds and decoys or a cross-docking (Fig. 1) study. It was suggested that the structures that have the best ability to dock non-native ligands with lower RMSD with reference to the crystalled pose of the ligands are probably more successful in prediction of binding pose (Zhang et al. 2014) of the ligands correctly and virtual screenings (Ramezani and Shamsara 2015).

On another hand, the performance of docking algorithms and scoring functions are varying for different targets (Cheng

*Correspondence: shamsaraj@mums.ac.ir
Pharmaceutical Research Center, School of Pharmacy, Mashhad University
of Medical Sciences, 91775-1365 Mashhad, Iran

Springer

Fig. 1 A typical workflow of a cross-docking study on a data set of three ligand-receptor complexes

et al. [2012](#); Warren et al. [2006](#); Shamsara [2014](#)). This can also be assessed by self and cross-docking. If the proper pose of a ligand in the active site is well modeled by a method and ranked higher among other possible poses it can be an indicator of suitability of the method for a given target protein.

In this paper we described CrossDocker which can significantly accelerate performing multiple dockings, data collection and subsequent analysis.

Implementation and preparation of inputs

The whole process of cross-docking was implemented as a computer program using the python language. As a case study 12 holo 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase structures were retrieved from PDB

(PDB codes: 1HW8, 1HW9, 1HWJ, 1HWK, 1HWL, 2Q1L, 2Q6B, 2Q6C, 2R4F, 3BGL, 3CCT, 3CCW, 3CCZ, 3CD0, 3CD5, 3CD7, 3CDA and 3CDB). The retrieved holo crystal structures from the PDB were aligned before docking to make the RMSD calculation possible after dockings. The protein alignment can be done by several open source or commercial tools such as PyMol or Chimera. All the structures were aligned by PyMol using align command. After structural alignment, the first two chains (A and B) of the each PDB file, adenosine-5′-diphosphate (ADP) and ligands were retained. The co-crystalized ligand and protein were saved as two individual files. CrossDocker detects the corresponding ligand-receptor by their file names. The name of the protein can be anything but should end with "-p" and

the name of the native ligand should be same as the protein name and ends with "-l". Thus, each Receptor/ligand pairs were separated into two pdb files with same root-name (e.g. their PDB codes) and different suffix, "-p" and "-l" respectively. CrossDocker reads structures in mol2, pdb or pdbqt formats and for receptors it considers all cofactors and coenzymes as a part of receptor molecule and removes all water molecules. Then ligands and receptors were divided into two separate folders. There is a configuration file (config.txt) that the path to the receptors and ligands folders can be set there. Some parameters for the run of Autodock Vina (Trott and Olson 2010) can also be set in this file (see the explanations in config.txt file) such as grid box dimensions and maximum number of binding poses to be generate for each dock (coordination will be set automatically by CrossDocker according to the coordination of the ligand). The initial conformation of the ligands was randomized by CrossDocker using "randomize_only" option of Autodock Vina prior to docking to avoid bias toward conformation of ligands in the crystal structure.

Results

"Output.xlsx" contains RMSD and energy calculated for each pose of each docked ligand in every receptor. The best RMSD for each dock and its docking energy is reported in "Output_the_best_RMSD.xlsx". "Table_the_best_RMSD. xlsx" contains best RMSD obtained for each ligand for each receptor. It shows the number with conditional formatting: green represents the 10 ‰, yellow the 50 ‰ and red the 90 ‰. The values for self-dockings are in bold faces. Thus, the interpretation of the results would be very easy. The average RMSD for each 3D structure is reported. The number of docks with RMSD <2.0 Å is another parameter that is reported in "Table_the_best_RMSD.xlsx" file. "Table_ the_best_energy.xlsx" contains the best energy obtained for each ligand docked in each receptor. "Table_RMSD_for_ the_best_energy.xlsx" includes the RMSD that is obtained for a docked pose of a ligand which has a best docked energy among generated modes by Autodock Vina. In the most cross-docking studies the best RMSDs (which can be found in Table-RMSD.xls file) were considered by authors. However, it seems that the calculated RMSD for the docked pose with the lowest energy which are reported in "Table_ energy_for_the_best_RMSD.xlsx" can also be important for analysis, because in a typical virtual screening the docked poses with lowest energy are only considered for analysis. Tables 1 and 2 shows contents of the "Table_the_best_ RMSD.xlsx" file and "Table_RMSD_for_the_best_energy.

Table 1 RMSD of each ligand pose which has smallest RMSD docking to each PDB code protein was shown

|	1HW8	1HW9	1HWJ	1HWK	1HWL	2Q1L	2Q6B	2Q6C	2R4F	3BGL	3CCT	3CCW	3CCZ	3CD0	3CD5	3CD7	3CDA	3CDB
1HW8	**1.60**	1.42	1.72	1.77	1.71	2.11	1.61	2.32	1.78	1.89	1.71	1.87	1.47	1.57	1.96	1.71	1.71	1.66
1HW9	1.67	**1.56**	1.74	1.72	1.67	1.65	1.45	1.60	1.62	1.69	1.75	2.06	2.02	1.55	2.02	2.01	1.96	2.12
1HWJ	1.67	1.54	**1.19**	1.46	1.21	1.20	1.90	1.45	2.22	1.42	1.17	1.67	1.48	1.97	1.53	1.30	1.32	1.21
1HWK	1.91	1.74	1.04	**1.29**	0.87	1.09	4.60	1.54	1.56	1.14	1.30	1.13	1.10	1.02	4.74	0.97	0.84	1.32
1HWL	2.71	1.77	1.94	2.90	**1.59**	2.24	1.92	1.82	2.65	2.72	2.89	2.80	1.57	2.77	1.95	2.93	2.75	2.88
2Q1L	2.18	2.06	1.10	0.84	0.88	**1.15**	4.60	1.43	1.90	1.56	1.17	2.42	2.03	2.08	5.81	1.29	2.02	1.56
2Q6B	1.74	1.88	0.84	1.22	1.18	1.60	**1.34**	0.97	1.41	1.57	1.79	1.57	1.46	1.20	1.81	1.57	1.62	1.28
2Q6C	2.90	1.68	1.27	1.42	1.27	1.32	4.95	**1.37**	1.58	0.91	1.30	1.92	1.73	1.43	1.87	1.21	1.29	1.22
2R4F	1.45	1.01	0.96	1.13	1.26	1.91	1.66	4.68	**1.22**	1.51	1.28	1.08	0.94	1.19	1.15	1.58	1.24	1.02
3BGL	3.95	4.77	1.31	1.11	1.19	1.31	7.10	6.44	5.42	**1.34**	1.37	5.42	5.38	4.37	1.64	0.95	1.16	1.19
3CCT	1.70	1.74	1.62	1.46	1.26	1.41	4.48	1.72	2.10	1.27	**1.52**	1.54	2.13	2.22	1.93	1.29	1.41	1.21
3CCW	1.73	1.53	1.68	1.43	1.64	4.84	1.78	5.03	1.45	2.03	1.29	**1.87**	1.70	1.42	4.21	1.52	1.37	1.76
3CCZ	1.35	1.73	1.91	1.54	1.85	1.38	2.08	4.92	1.29	1.38	0.87	1.61	**1.42**	1.62	3.90	2.98	5.27	1.82
3CD0	4.53	1.15	1.03	1.03	1.10	1.83	1.84	1.55	1.82	4.54	1.07	1.45	1.15	**4.72**	1.21	2.83	5.23	1.87
3CD5	3.96	2.77	4.80	4.04	2.68	1.02	1.66	1.49	0.92	1.21	1.46	1.45	1.29	1.46	**1.34**	1.11	1.69	1.99
3CD7	1.36	1.91	1.21	0.91	1.33	1.16	4.77	1.31	2.09	1.52	1.20	2.48	1.97	2.01	4.89	**0.97**	1.55	1.45
3CDA	1.87	1.42	0.83	1.12	1.50	1.00	4.47	1.66	2.01	1.41	1.17	0.99	1.95	1.17	1.57	1.49	**1.50**	1.23
3CDB	3.90	4.27	1.25	3.20	1.75	3.73	4.62	3.54	3.74	3.64	3.66	5.69	1.73	3.47	3.99	3.73	5.38	**1.09**
Average	2.34	2.00	1.52	1.64	1.44	1.77	3.16	2.49	2.04	1.82	1.55	2.17	1.81	2.07	2.64	1.75	2.18	1.55
RMSD < 2.0	11	14	17	15	17	14	9	12	11	14	16	12	14	11	11	13	13	16

Receptors are in columns while co-crystalized ligands are in rows. The average of the calculated RMSDs for a receptor is presented. The number of successful docks (RMSD <2.0) is also presented for each receptor in the last row of the table. The RMSDs of self-dockings are in *bold* faces

Table 2 RMSD of each ligand pose which has smallest energy docking to each PDB code protein

	Lowest value				50th percentile				Highest value							

	1HW8	1HW9	1HWJ	1HWK	1HWL	2Q1L	2Q6B	2Q6C	2R4F	3BGL	3CCT	3CCW	3CCZ	3CD0	3CD5	3CD7	3CDA	3CDB
1HW8	**1.60**	1.42	1.72	1.85	7.13	7.16	1.61	7.18	2.01	7.05	7.39	1.87	1.47	1.57	1.96	7.23	7.46	7.28
1HW9	1.67	**1.56**	1.74	1.97	1.67	1.65	2.46	1.60	1.62	7.00	1.75	2.06	2.02	1.89	2.02	2.01	1.96	2.12
1HWJ	7.96	1.54	**1.19**	1.46	1.21	1.20	1.98	1.45	2.22	1.42	1.17	1.67	1.48	6.17	1.53	1.30	1.32	1.21
1HWK	8.55	8.89	1.04	**1.29**	0.87	1.09	8.09	1.54	5.27	1.14	1.30	1.13	5.04	1.02	5.58	0.97	0.84	1.32
1HWL	4.88	1.77	4.91	5.67	**5.59**	4.56	1.92	1.82	2.65	3.43	2.89	2.80	3.10	2.77	2.59	3.50	2.99	4.57
2Q1L	2.18	2.06	1.10	0.84	0.88	**1.15**	7.52	1.43	1.90	1.56	1.17	8.18	5.30	2.08	9.49	1.29	8.31	1.56
2Q6B	1.74	1.88	0.84	1.22	1.18	1.60	**1.34**	0.97	1.41	1.62	1.79	1.84	1.46	1.20	1.81	1.57	1.62	2.47
2Q6C	2.90	1.68	1.27	1.42	1.27	1.32	6.57	**1.37**	1.58	0.91	1.30	1.92	1.73	1.43	1.87	1.21	1.29	1.22
2R4F	7.41	1.01	0.96	1.13	1.26	5.27	1.66	7.26	**7.51**	10.20	1.28	1.08	0.94	1.19	1.15	1.94	7.40	1.02
3BGL	8.12	4.77	1.31	1.11	1.19	8.00	8.46	8.26	8.25	**1.34**	1.37	7.93	8.39	7.96	13.37	0.95	1.16	1.19
3CCT	8.58	1.74	1.62	1.46	1.26	1.41	9.00	1.72	2.10	1.27	**1.52**	1.54	4.82	8.54	5.46	1.29	1.41	1.21
3CCW	9.38	1.53	1.68	10.12	1.64	8.65	1.78	7.86	8.65	8.53	7.32	**1.87**	8.55	8.37	4.42	6.40	8.60	8.05
3CCZ	7.60	7.58	1.91	10.88	1.85	10.53	2.08	7.77	7.58	8.72	0.87	10.88	**1.42**	8.63	5.69	8.45	7.46	8.94
3CD0	8.19	1.15	1.03	1.03	1.10	6.49	1.84	7.79	7.98	8.94	8.86	10.82	8.91	**8.29**	4.33	8.15	9.03	8.84
3CD5	10.56	3.17	10.53	10.48	10.55	10.86	1.66	1.49	0.92	10.73	10.57	11.15	10.86	10.49	**1.34**	10.57	10.64	10.38
3CD7	1.36	5.20	1.21	0.91	1.33	1.16	6.92	1.31	2.09	1.52	1.20	8.21	1.97	8.24	13.44	**0.97**	8.32	1.45
3CDA	6.04	2.24	0.83	1.12	4.58	4.71	10.41	1.66	5.03	1.41	1.27	0.99	5.95	1.17	5.30	4.71	**1.50**	1.23
3CDB	8.63	8.73	7.89	7.96	7.88	8.53	8.89	3.95	8.41	3.64	7.32	7.25	8.60	8.26	5.22	7.81	7.95	**3.56**
Average	5.97	3.22	2.38	3.44	2.91	4.74	4.68	3.69	4.29	4.47	3.35	4.62	4.56	4.96	4.81	3.91	4.96	3.76
RMSD < 2.0	4	10	15	13	13	8	8	11	5	9	12	9	7	7	6	9	8	9

Receptors are in columns while co-crystalized ligands are in rows. The average of the calculated RMSDs for a receptor is presented. The number of successful docks (RMSD <2.0) is also presented for each receptor in the last row of the table. The RMSDs of self-dockings are in *bold* faces

xlsx" generated by CrossDocker from HMG-CoA reductase data set. Figure 2 shows reliability of the docked pose of a ligand with reference to the native conformation in the active site of the structure with PDB code 1HWJ.
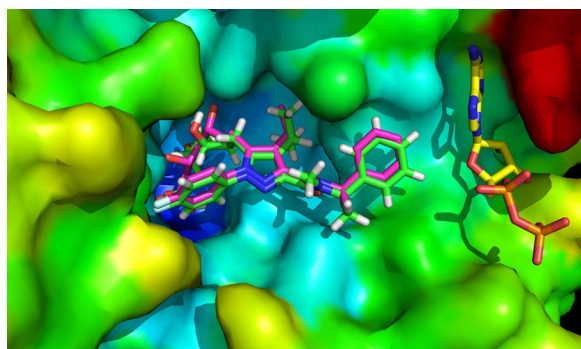
Fig. 2 Docked pose (*green*) of the ligand of 2R4F PDB code in the active site of HMG-CoA reductade structure with PDB code 1HWJ. The experimental pose of the ligand in PDB code 2R4F is shown in *magenta*. ADP is shown in *yellow* color. The calculated RMSD is 0.96 Å. The atoms in the 1HWJ active site were color coded by their B-factors. *Blue* is for low B-factor and *red* is for high B-factor value. Higher B-factor may indicate flexibility of the residues (inaccuracy in crystallography for some part of the protein also causes the higher B-factor)

Discussion and conclusions

CrossDocker provides a good opportunity to perform cross-docking easily on a series of 3D structure of a same receptor and subsequent data collection and analysis. As it was suggested previously the structures with lower average RMSD and/or higher number of docked poses with RMSD <2.0 Å have higher probability for good performance in virtual screening study and pose prediction (Vinh et al. 2012; Ramezani and Shamsara 2015; Zhang et al. 2014). Thus the reported results by CrossDocker can be used to achieve the best structures for computational drug design studies. Furthermore, if the most of the calculated average RMSD for a specific target is high and/or self-docking RMSDs are also high, it can indicate poor performance of Autodock Vina on a specific target and vice versa. In some cases, it can be improved by increasing exhaustiveness parameter in configuration file that would also increase the computation time. In the example above the performance of Autodock Vina on the set of HMG-CoA reductase structures was reasonable with default setting (exhaustiveness = 8) (see Fig. 2 and calculated self-docking RMSDs in Tables 1, 2). According to the results of the obtained best RMSDs, structures 1HWJ, 1HWL, 3CCT and 3CDB would be more promising to use for binding pose prediction and virtual

screening studies. As the docking poses with the lowest energy normally get more attention in the analysis step of a docking study the contents of the output table entitled "Table_RMSD_for_the_best_energy.xlsx" should also be considered for selection of the best PDB structure. Thus, according to Table 2, 1HWJ is the best one among those four structures. In this paper the applicability of Cross-Docker was shown. CrossDocker was written in Python language and available as executable binary for Windows operating system. All examples input and output files as well as CrossDocker are available at http://www.pharm-sbg.com.

Availability and requirements

Project name: CrossDocker project;

Project home page: http://www.pharm-sbg.com;

Operating system(s): Windows;

Programming language: Python;

Other requirements: Openbabel 2.3 or higher.

Competing interests
The author declare that he has no competing interests.

References

Chandrika C, Subramanian J, Sharma SD (2009) Managing protein flexibility in docking and its applications. Drug Discov Today 14(7–8):394–400. doi:10.1016/j.drudis.2009.01.003

Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH (2012) Structure-based virtual screening for drug discovery: a problem-centric review. AAPS J 14(1):133–141. doi:10.1208/s12248-012-9322-0

Hawkins PC, Warren GL, Skillman AG, Nicholls A (2008) How to do an evaluation: pitfalls and traps. J Comput Aided Mol Des 22(3–4):179–190. doi:10.1007/s10822-007-9166-3

Kubinyi H (2006) Success stories of computer-aided design. In: Wang B, Ekins S (eds) Computer applications in pharmaceutical research and development. Wiley, Hoboken, pp 377–424. doi:10.1002/0470037237.ch16

Mohan V, Gibbs AC, Cummings MD, Jaeger EP, DesJarlais RL (2005) Docking: successes and challenges. Curr Pharm Des 11(3):323–333

Pitt WR, Calmiano MD, Kroeplien B, Taylor RD, Turner JP, King MA (2013) Structure-based virtual screening for novel ligands. Methods Mol Biol 1008:501–519. doi:10.1007/978-1-62703-398-5_19

Ramezani M, Shamsara J (2015) A cross-docking study on matrix metalloproteinase family. AntiInflamm Antiallergy Agents Med Chem 14(3):164–171

Sacan A, Ekins S, Kortagere S (2012) Applications and limitations of in silico models in drug discovery. Methods Mol Biol 910:87–124. doi:10.1007/978-1-61779-965-5_6

Shamsara J (2014) Evaluation of 11 scoring functions performance on matrix metalloproteinases. Int J Med Chem 2014:162150-1–162150-9. doi:10.1155/2014/162150

Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multi-threading. J Comput Chem 31(2):455–461. doi:10.1002/jcc.21334

Vinh NB, Simpson JS, Scammells PJ, Chalmers DK (2012) Virtual screening using a conformationally flexible target protein: models for ligand binding to p38alpha MAPK. J Comput Aided Mol Des 26(4):409–423. doi:10.1007/s10822-012-9569-7

Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) A critical assessment of docking programs and scoring functions. J Med Chem 49(20):5912–5931. doi:10.1021/jm050362n

Zhang H, Wang Y, Xu F (2014) Impact of the subtle differences in MMP-12 structure on Glide-based molecular docking for pose prediction of inhibitors. J Mol Struct 1076:153–159. doi:10.1016/j.molstruc.2014.06.002