HHS Public Access

Author manuscript

J Comput Aided Mol Des. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

J Comput Aided Mol Des. 2018 January ; 32(1): 225–230. doi:10.1007/s10822-017-0069-7.

Protein - ligand docking using FFT based sampling: D3R case study

Dzmitry Padhorny[1,2], **David R. Hall**[3,*], **Hanieh Mirzaei**[4], **Artem B. Mamonov**[4], **Mohammad Moghadasi**[4], **Andrey Alekseenko**[5,6], **Dmitri Beglov**[4,**], and **Dima Kozakov**[1,2,***]

[1]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA

[2]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA

[3]Acpharis Inc., Holliston, MA 01746, USA

[4]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA

[5]Moscow Institute of Physics and Technology (State University), Institutskii per. 9, Dolgoprudnyi, Moscow oblast, 141700 Russia

[6]Institute of Computer Aided Design of the Russian Academy of Sciences, 19/18, 2-nd Brestskaya st, Moscow, 123056 Russia

Abstract

Fast Fourier Transform (FFT) based approaches have been successful in application to modeling of relatively rigid protein-protein complexes. Recently, we have been able to adapt the FFT methodology to treatment of flexible protein-peptide interactions. Here, we report our latest attempt to expand the capabilities of the FFT approach to treatment of flexible protein-ligand interactions in application to the D3R PL-2016-1 challenge. Based on the D3R assessment, our FFT approach in conjunction with Monte Carlo Minimization (MCM) off-grid refinement was among the top performing methods in the challenge. The potential advantage of our method is its ability to globally sample the protein-ligand interaction landscape, which will be explored in further applications.

Introduction

Development of the Fast Fourier Transform (FFT) based sampling methods for macromolecular interactions has been the focus of our group for a number of years. The major advantages of FFT methods are their globality and speed: by calculating the interaction energy of two rigid bodies as a sum of correlation functions it becomes possible to quickly sample the interaction landscape, evaluating billions of putative interactions on a grid in a matter of minutes.

*,**,***To whom correspondence should be addressed: david@acpharis.com, dbeglov@bu.edu, midas@laufercenter.org.

Unsurprisingly, FFT-based methods became popular in the field of protein-protein docking [1–5], where the two interacting macromolecules can be approximately viewed as rigid. Our own protein docking server ClusPro has repeatedly performed as the top automated server in CAPRI - a community-wide protein docking competition [6–8].

Recently we have successfully expanded the capabilities of our FFT docking method to treatment of flexible peptide interactions [9]. This was achieved by docking multiple peptide conformations drawn from a library of PDB-derived fragments. This implies that an FFT-based approach can be suitable for modeling of protein-small molecule interactions even in the case when ligand flexibility needs to be taken into account.

In this work, we describe our first steps to construction of such docking protocol in application to the targets from the D3R PL-2016-1 challenge. We show that an FFT-based approach in conjunction with short Monte Carlo Minimization (MCM) refinement simulations is able to provide accurate pose predictions for the various targets in the challenge. The top-ranked poses produced by our approach for the 2 challenge datasets have mean RMSDs of 0.559 and 1.420 Angstrom, and were among the best predicted poses in the D3R PL-2016-1 round according to the assessors. This work demonstrates the feasibility of using FFT-based sampling approaches for modeling of flexible protein-ligand interactions.

Methods

D3R 2016 challenge and test datasets

The D3R 2016 PL-2016-1 Challenge was based on the two datasets provided by Barry Stoddard and David Baker. Each of these datasets contained a series of closely related proteins (sequence identity >90% within each series) artificially designed to bind a single ligand, 17-hydroxyprogesterone (17-OHP) or 25-hydroxycholecalciferol (25-D3) respectively. Throughout the text, we refer to the datasets using the names of the target ligands.

The 17-OHP binding dataset consisted of two protein-ligand co-crystal structures: PL-2016-1-O-1 and PL-2016-1-O-2, resolved with resolutions of 2.5 and 2.0 Angstrom, respectively, and provided to the challenge participants with ligands removed. The goal of the challenge was to predict the ligand binding poses for each of the two protein-ligand complexes and rank proteins within this series by their affinity to 17-OHP.

The 25-D3 binding dataset contained three protein-ligand co-crystal structures referred as PL-2016-1-C-1, PL-2016-1-C-2 and PL-2016-1-C-3, with resolutions of 1.9, 2.1 and 1.9 Angstrom and, again, with ligands removed. Similarly to the previous case, the D3R challenge participants were expected to predict the ligand binding pose for each complex. The affinity prediction challenge, however, was different: the goal was to predict the relative affinity of PL-2016-1-C-1 to 25-D3 and its other ligand, vitamin D3.

Notably, for both datasets the water-mediated interactions were assumed to play an important role, and the crystal structures provided by the organizers included crystal waters.

The ligands were available in 2D representations as SDFiles, and, therefore, an additional challenge was to correctly predict the puckers of nonplanar rings for 17-OHP.

Up to 5 ligand poses per protein were allowed for submission in each docking prediction. Affinity prediction was to be submitted as absolute affinity in nM, as relative affinity, or as a simple ranking of proteins by their affinity.

Docking and scoring protocol

The overall structure of the docking protocol is presented in Fig. 1.

The whole protocol consisted of the 5 major steps:

First, multiple possible conformers of the ligand featuring various possible torsion angle orientations and different nonplanar ring puckers were generated. For this purpose we employed Confab - a systematic generator of low energy conformers [10]. At the end of this stage 10 lowest-energy conformers for 25-D3 and vitamin D were retained for future processing. For 17-OHP only 5 low conformers were generated, likely due to the rigidity of the molecule.

Pre-generated conformers were then used as inputs for our FFT-based rigid body global docking program PIPER [1]. Using an FFT-based approach greatly accelerates the solution of the rigid body search problem, and allows to perform global sampling of protein-ligand interaction landscape in affordable time, evaluating billions of putative protein-ligand orientations on a grid. Performing docking of a rigid protein to multiple pre-generated ligand conformers allowed to account for ligand flexibility within the FFT based docking framework, similar in style to our recently developed FFT based peptide docking protocol [9].

Docking of each ligand conformer constituted a single FFT docking run; the resulting low-energy poses from different runs were merged to form a pool of 1000 poses and then clustered together. Docking poses serving as cluster centers were retained for later optimization stages.

Rigid docking poses generated with FFT served as starting points for an in-house Monte Carlo minimization algorithm [11–13]. Since all proteins present in the two datasets had a large central cavity, in each case such cavity was assumed to be the target binding site, and thus only poses located within the cavity were further refined. During the MCM stage, all ligand torsions were treated as rotatable, except for those localized within closed rings. CHARMM19-based energy function augmented with knowledge-based hydrogen bonding and GBSA terms was used for scoring of generated conformations as previously described in [12]. Each MCM trajectory consisted of 10,000 Monte Carlo steps, and conformations corresponding to all accepted moves were retained.

In the final stage of the docking pipeline, all accepted Monte Carlo conformations were rescored with crystal waters using a Vina [14] based energy function. Structural water molecules in the binding site are known to mediate protein-ligand interactions, and correct placement of waters was shown to improve the scoring of ligand poses [15]. The

minimization step in MCM drives ligands toward local minima and greatly reduces the number of poses for rescoring. Rescored poses were clustered with a 1.0 Angstrom RMSD cutoff after which the lowest-energy poses from each cluster were reported as final docking predictions. These final predictions were ranked according to their Vina scores.

For the purpose of scoring in D3R, we simply used the Vina scoring function of the top docking models.

FFT based rigid body docking

FFT docking was performed with PIPER [1] using a variation of the protocol described previously [16]. In short, given a protein receptor and a ligand conformer, we explore the full discretized conformational space of receptor-ligand rigid body orientations using the Fast Fourier Transform correlation approach. In the process of sampling the protein remains fixed, while the ligand is rotated and translated. Rotational degrees of freedom are sampled from a semi-uniform set of 70,000 rotations constructed using layered Sukharev grid sequence, and translations are sampled on a 3D grid with 1.0 Angstrom spacing.

The advantage of the FFT approach is that if the energy function is written down as a sum of several correlation functions, then for a single rotation of the ligand (or, in our case, ligand conformer), energy values for all possible translations can be evaluated very efficiently using P forward and one inverse Fourier transforms, where P is the number of correlation terms in the energy function. Thus if N denotes the size of translation grid in each dimension, then the complexity of the FFT approach per one rotation is $O(N^3 ln(N))$, as opposed to $O(N^6)$ for the direct approach. In PIPER, the FFT calculations are repeated for all 70,000 rotations of the ligand, and a single lowest-energy pose is retained for each rotation, resulting in a total of 70,000 low energy poses.

For the purpose of small molecule docking, we use an energy function composed of attractive (E_{attr}) and repulsive (E_{rep}) Van der Waals contributions, as well as an electrostatic term with Born correction (E_{elec}) [17]: $E = E_{vdw} + w_2 E_{elec}$. Here individual terms are computed as:

$$E_{vdw} = E_{attr} + w_1 E_{rep},$$
$$E_{elec} = \sum_i \sum_j \frac{q_i q_j}{r^2 + D^2 exp(-r^2/D^2)},$$

where r is the distance between atoms i,j and D is an Born radius. For consistency, we use a standard set of weights previously developed for protein docking applications: $w_1 = 1$, $w_2 = 750$.

When docking the D3R targets, a separate docking run was performed for each ligand conformer, resulting in 10 docking runs for for each target from the 25-D3 dataset and 5 docking runs for those from the 17-OHP dataset. An equal number (100 for 25-D3 dataset targets and 200 for 17-OHP) of lowest-energy poses from each run were taken to form a pool of 1000 poses, which were then clustered together with a clustering radius of 1.0 Angstrom and cluster centers of the 10 largest clusters retained as final FFT docking models.

Monte Carlo Minimization refinement

For the purpose of flexible docking pose optimization, we have updated the Monte Carlo minimization (MCM) package previously developed in our lab for protein docking refinement[11–13], with capabilities for sampling ligand torsional degrees of freedom. Additionally the algorithm provides full support for sampling ligand rigid body degrees of freedom, and allows for side-chain flexibility on the receptor side. Random moves are coupled with fast local minimization of the generated poses carried out using a manifold-based optimization algorithm. In the latest version of the code, our previously reported custom energy function [12] and Vina [14] energy function can be used for pose evaluation, and pose acceptance is guided by a Metropolis criterion. Here we provide a brief overview of the major algorithmic features of our MCM package.

In our approach, the treatment of ligand flexibility is based on the common assumption that the changes in covalent bond-lengths and bond-angles can be neglected. In this context, the whole ligand can be viewed as a set of rigid molecular clusters (aromatic rings, methyl groups, etc.) interconnected by rotatable bonds. The only internal degrees of freedom of the ligand are, therefore, associated with torsional moves around such rotatable bonds, and description of the ligand as a whole can be given in terms of 6 rigid body and d torsional degrees of freedom, where d is the number of torsions. Overall, such ligand representation is typical for small molecule docking methods (e.g. [14, 18]. To correctly handle the changes in the molecule upon torsional moves, we employ a torsion tree data structure as described in previous work [13].

An interesting observation that comes from representing molecular configurations in terms of rigid body orientations and torsions is the fact that the underlying space of such representation is, a manifold, that is, a locally Euclidean topological space, which is formed as a direct product of the rigid motion space, $SO(3) \times R^3$, and the internal motion space, $T^d = (S^1 \times S^1 \times ... \times S^1)_{d\ times}$. This allows us to perform local energy minimization directly on this 6+d -dimensional space, as opposed to doing full-atomic minimization. We have recently shown that manifold optimization can achieve a speedup with a factor of 5 over the full-atomic minimization, which provides a considerable speedup of the protocol.

Protein flexibility is taken into account by allowing movable side-chains. In practice, this is achieved by sampling a predefined set of rotameric states for each residue in the binding site. Possible rotamers are taken from the Dunbrack rotamer library [19]. Additionally, we allow off-rotamer flexibility by employing the same torsion tree representation of side-chains as that used for the ligand.

Each step of MCM, consists of 3 stages: (1) perturbation of ligand conformation, (2) sliding receptor and ligand into contact with simultaneous side chain repacking, and (3) local refinement using a manifold optimization algorithm. Figure 2 provides a brief visual summary of a single MCM step.

When docking the D3R targets, each simulation was started from a rigid docking model generated in the FFT docking stage, and only the models found in the central cavity were refined (3 models for all targets except PL-2016-1-O-2, for which 2 models were processed).

Translation steps of up to 1.0 Angstrom, rotations of up to 5 degrees, dihedral steps of up to 0.2 radian and sliding steps of 0.1 Angstrom were used. Since bound protein structures were available, side chain flexibility was disabled. Simulations of 10,000 steps were performed for each target.

Results and discussion

Predicted docking poses and affinities

The summary of docking results for the 2 datasets is presented in Table 1. Results are assessed in terms of averaged RMSD for the top-1 ranked poses, best submitted poses and all submitted poses. For the 17-OHP dataset the reference X-ray structure had 4 copies of the protein, with bound ligand conformations showing differences among the copies, so the evaluations were done using the copy which provided the lowest RMSD. For visual reference, top-1 docking poses for all submissions are presented in FIgure 3, overlaid onto ligand X-ray structures.

As can be seen from the table, we obtained good docking predictions for both datasets, featuring sub-Angstrom mean best pose RMSDs. Notably, for all targets in the 17-OHP dataset our best pose predictions happened to be the ones that we ranked first. A somewhat lower mean quality of top ranked poses for the 25-D3 dataset was mainly caused by suboptimal top docking pose quality for the PL-2016-1-C-1 case (see Fig 2A, A*). A better model was, however, present in the top-5 poses submitted, so the mean best pose RMSD is significantly lower. Overall, in terms of mean pose 1 RMSD our pose predictions were ranked first for 17-OHP dataset and fourth for the 25-D3 dataset out of 13 groups that participated (in terms of mean best pose RMSD it's first and third accordingly). However, it should be noted that the results from best performing groups were very close to each other.

Scoring results are summarized in Table 2. Note that for the 17-OHP series the values correspond to affinities of two different proteins to the same ligand, while for the 25-D3 series the values are for the affinities of the same protein (PL-2016-1-C-1) to different ligands. In both cases our predicted affinity ranking was agreeing with experimental data.

Conclusion

In this work, we report on the feasibility of a multistage docking pipeline consisting of an FFT-based sampling stage followed by MCM refinement, applied to small molecule docking in the D3R PL-2016-1 challenge. In this application, with the location of the binding site known beforehand, the quality of pose predictions produced by the method was on par with the top performing methods, as judged by the D3R assessment. This result demonstrates that FFT sampling with off-grid minimization has potential to achieve sufficient accuracy for successful flexible small molecular docking. The advantage of the proposed approach is the possibility to develop global systematic sampling of the protein-ligand configurational space, which would require incorporation of accurate knowledge-based scoring functions [17, 20–25] and further validations on docking benchmarks [26, 27].

Acknowledgments

This work was supported by grants NSF CCF AF 1527292, NIH R43 GM109555, RSF No 14-11-00877

References

1. Kozakov D, Brenke R, Comeau SR, Vajda S. PIPER: an FFT-based protein docking program with pairwise potentials. Proteins. 2006; 65:392–406. [PubMed: 16933295]
2. Gabb HA, Jackson RM, Sternberg MJ. Modelling protein docking using shape complementarity, electrostatics and biochemical information. J Mol Biol. 1997; 272:106–120. [PubMed: 9299341]
3. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins. 2000; 39:178–194. [PubMed: 10737939]
4. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. 2006; 34:W310–4. [PubMed: 16845016]
5. Chen R, Li L, Weng Z. ZDOCK: an initial-stage protein-docking algorithm. Proteins. 2003; 52:80–87. [PubMed: 12784371]
6. Lensink MF, Velankar S, Wodak SJ. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. Proteins. 2017; 85:359–377. [PubMed: 27865038]
7. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins. 2013; 81:2082–2095. [PubMed: 24115211]
8. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. Proteins. 2010; 78:3073–3084. [PubMed: 20806235]
9. Porter KA, Xia B, Beglov D, Bohnuud T, Alam N, Schueler-Furman O, Kozakov D. ClusPro PeptiDock: Efficient global docking of peptide recognition motifs using FFT. Bioinformatics. 2017; doi: 10.1093/bioinformatics/btx216
10. O'Boyle NM, Vandermeersch T, Flynn CJ, Maguire AR, Hutchison GR. Confab - Systematic generation of diverse low-energy conformers. J Cheminform. 2011; 3:8. [PubMed: 21410983]
11. Mamonov AB, Moghadasi M, Mirzaei H, Zarbafian S, Grove LE, Bohnuud T, Vakili P, Ch Paschalidis I, Vajda S, Kozakov D. Focused grid-based resampling for protein docking and mapping. J Comput Chem. 2016; 37:961–970. [PubMed: 26837000]
12. Moghadasi M, Mirzaei H, Mamonov A, Vakili P, Vajda S, Paschalidis IC, Kozakov D. The impact of side-chain packing on protein docking refinement. J Chem Inf Model. 2015; 55:872–881. [PubMed: 25714358]
13. Mirzaei H, Zarbafian S, Villar E, Mottarella S, Beglov D, Vajda S, Paschalidis IC, Vakili P, Kozakov D. Energy Minimization on Manifolds for Docking Flexible Molecules. J Chem Theory Comput. 2015; 11:1063–1076. [PubMed: 26478722]
14. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31:455–461. [PubMed: 19499576]
15. Huggins DJ, Tidor B. Systematic placement of structural water molecules for improved scoring of protein-ligand interactions. Protein Eng Des Sel. 2011; 24:777–789. [PubMed: 21771870]
16. Kozakov D, Beglov D, Bohnuud T, Mottarella SE, Xia B, Hall DR, Vajda S. How good is automated protein docking? Proteins. 2013; 81:2159–2166. [PubMed: 23996272]
17. Chuang G-Y, Kozakov D, Brenke R, Comeau SR, Vajda S. DARS (Decoys As the Reference State) potentials for protein-protein docking. Biophys J. 2008; 95:4217–4227. [PubMed: 18676649]
18. Meiler J, Baker D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins. 2006; 65:538–548. [PubMed: 16972285]
19. Shapovalov MV, Dunbrack RL Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure. 2011; 19:844–858. [PubMed: 21645855]
20. Grudinin S, Kadukova M, Eisenbarth A, Marillet S, Cazals F. Predicting binding poses and affinities for protein - ligand complexes in the 2015 D3R Grand Challenge using a physical model

with a statistical parameter estimation. J Comput Aided Mol Des. 2016; 30:791–804. [PubMed: 27718029]

21. Yan C, Grinter SZ, Merideth BR, Ma Z, Zou X. Iterative Knowledge-Based Scoring Functions Derived from Rigid and Flexible Decoy Structures: Evaluation with the 2013 and 2014 CSAR Benchmarks. J Chem Inf Model. 2016; 56:1013–1021. [PubMed: 26389744]
22. Ballester PJ, Schreyer A, Blundell TL. Does a More Precise Chemical Description of Protein–Ligand Complexes Lead to More Accurate Prediction of Binding Affinity? J Chem Inf Model. 2014; 54:944–955. [PubMed: 24528282]
23. Wang C, Zhang Y. Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. J Comput Chem. 2017; 38:169–177. [PubMed: 27859414]
24. Debroise T, Shakhnovich EI, Chéron N. A Hybrid Knowledge-Based and Empirical Scoring Function for Protein-Ligand Interaction: SMoG2016. J Chem Inf Model. 2017; 57:584–593. [PubMed: 28191941]
25. Pires DEV, Ascher DB. CSM-lig: a web server for assessing and comparing protein-small molecule affinities. Nucleic Acids Res. 2016; 44:W557–61. [PubMed: 27151202]
26. Cheng T, Li X, Li Y, Liu Z, Wang R. Comparative assessment of scoring functions on a diverse test set. J Chem Inf Model. 2009; 49:1079–1093. [PubMed: 19358517]
27. Li Y, Liu Z, Li J, Han L, Liu J, Zhao Z, Wang R. Comparative assessment of scoring functions on an updated benchmark: 1. Compilation of the test set. J Chem Inf Model. 2014; 54:1700–1716. [PubMed: 24716849]

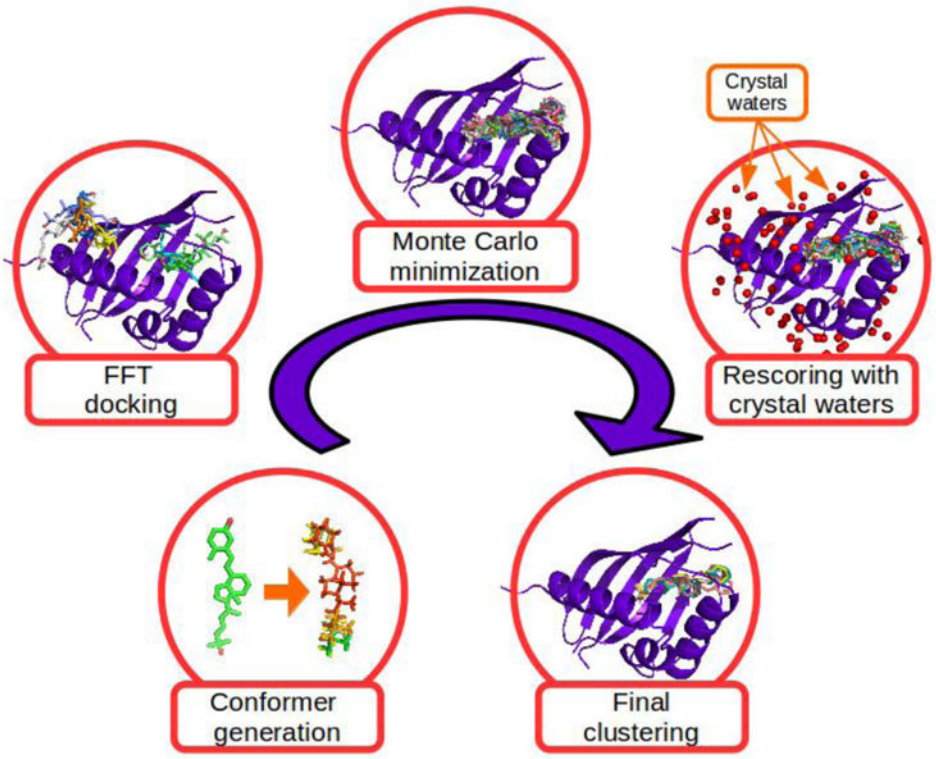

Figure 1.
Ligand pose prediction protocol.

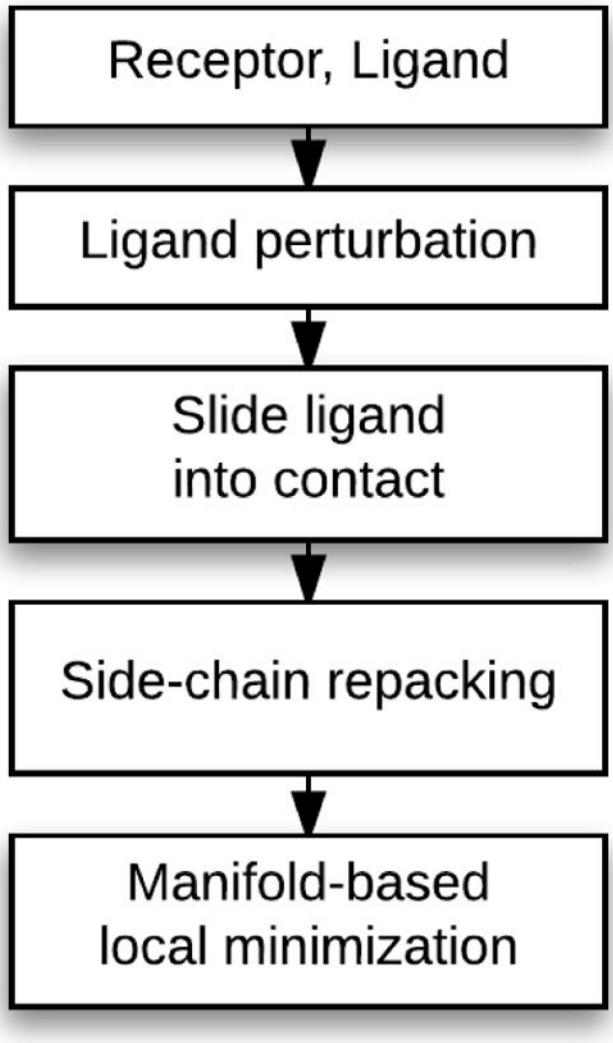

Figure 2.
The structure of a single MCM step.

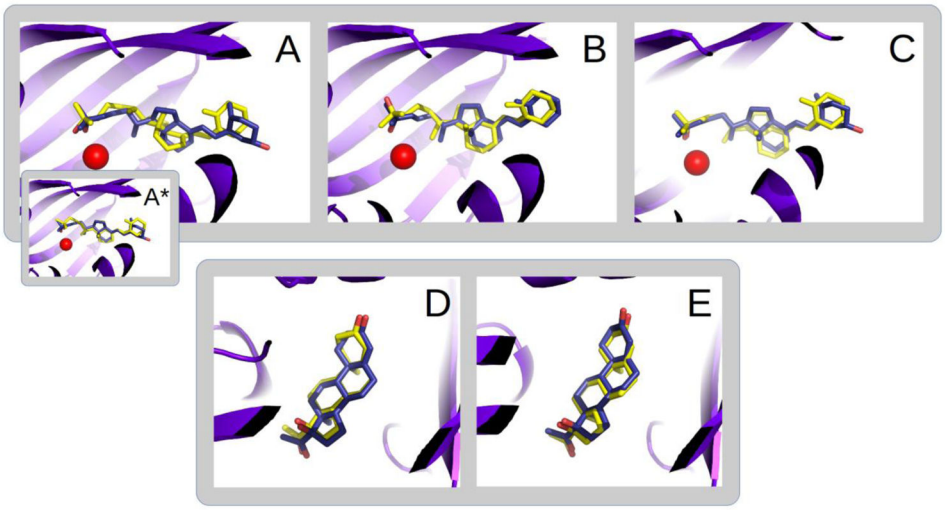

Figure 3.

Top ranked poses submitted in the D3R 2016 Grand Challenge for 25-D3 (A, B, C) and 17-OHP (D, E) datasets. X-ray structures of the ligands are presented in blue, docking models in yellow. Crystal waters interacting with ligands are depicted as red spheres. For PL-2016-1-C-1 case (A) our top-ranked model was not the best out of the 5 models submitted, so the best model ranked 3 is presented in the inset (A*).

Table 1

Assessment of ligand docking pose predictions for the 17-OHP and 25-D3 datasets. Predictions were evaluated in terms of ligand root mean square deviations (RMSD) relative to its position in the X-ray structure. The values presented in the table have been averaged over the protein targets in each dataset (2 for 17-OHP and 3 for 25-D3).

		RMSD		
Dataset	**# of poses submitted**	**Mean pose 1**	**Mean best pose**	**Mean all poses**
17-OHP	10	0.559	0.559	4.390
25-D3	15	1.420	0.908	1.320

Table 2

Affinity predictions for the 17-OHP and 25-D3 datasets. Predictions were made based on top-ranked ligand docking pose using Vina-based scoring function.

Dataset	Receptor	Ligand	Experimental affinity (μM)	Predicted affinity (Vina based score)
17-OHP	PL-2016-1-O-1	17-OHP	0.060 ± 0.008	−11.57187
	PL-2016-1-O-2	17-OHP	15 ± 2	−10.96171
25-D3	PL-2016-1-C-1	25-D3	0.300 ± 0.040	−10.40789
	PL-2016-1-C-1	D3	~2	−10.29677