Post-print of: Wojciechowski M.: Simplified AutoDock force field for hydrated binding sites. JOURNAL OF MOLECULAR GRAPHICS & MODELLING. Vol. 78, (2017), s. 74-80. DOI: 10.1016/j.jmgm.2017.09.016

Simplified AutoDock force field for hydrated binding sites

Marek Wojciechowski

Department of Pharmaceutical Technology and Biochemistry, Gdansk University of Technology, ul. Narutowicza 11/12, 80-233 Gdańsk, Poland

A B S T R A C T

A set of high quality structures of protein-ligand complexes with experimentally determined binding affinities has been extracted from the Protein Data Bank and used to test and recalibrate AutoDock force field. Since for some binding sites water molecules are crucial for bridging the receptor-ligand interactions, they have to be included in the analysis. To simplify the process of incorporating water molecules into the binding sites and make it less ambiguous, new simple water model was created. After recalibration of the force field on the new dataset much better correlation between the computed and experimentally determined binding affinities was achieved and the quality of pose prediction improved even more.

Keywords: Molecular docking simulations Binding sites, X-ray crystallography Ligands, Molecular models, Protein binding, Water

1. Introduction

AutoDock [1] is one of the most popular programs used for molecular docking and virtual screening. It has been developed and improved for several years and its force field was also improved over the time. Since this force field is the empirical one its quality is heavily dependent on the quality of the structures used to develop and calibrate the force field parameters. Due to the advances in experimental methods both, the quality and the number of available experimental structures is constantly growing giving the opportunity to extend and improve existing empirical force fields and to develop the new ones [2–4].

One of the important issues one faces while developing the empirical force fields dedicated to docking and virtual screening is the availability of the reliable experimental affinity data required for the proper calibration of the force field parameters. Recently a few databases (such as PDBBind-CN [5], MOAD [6] or BioLiP [7]) were created, aimed at collecting together structural information about various high quality protein-ligand complexes derived from PDB associated with the experimental affinity data extracted from the original literature.

Yet another important factor determining the quality of the developed model is the treatment of the receptor molecule and particularly the binding site itself [8]. In many cases water molecules can be found in the binding sites of the enzymes. Typically in the process of receptor preparation these water molecules are removed and only the protein receptor and ligand molecules are considered in docking. However, in many cases such approach leads to incorrect binding modes since some water molecules, tightly bound in the biding site, are responsible for bridging the ligand-receptor interactions and their omitting in the model leads to incorrect results. Various ways of including such water molecules in the docking protocols were approached by various researchers but there is no generally accepted way to do so [3,9–12]. One of the main issues, especially for virtual screening and new inhibitors design, is the strong bias toward particular ligand binding mode, induced by the choice of the specific water molecules and this effect is even enhanced after completing the waters with hydrogen atoms in the process of the receptor preparation. This may be less of a problem when performing a virtual screening of new ligands for the well-known receptor, with well recognized all the essential elements of its binding site, including water molecules. But even in this case sticking to one orientation of water molecules may limit the number of recognized ligands. The task becomes even more complicated when it is little known about the binding site of the particular receptor and one has to decide if any of the water molecules should be included at all.

The goal of this work was to select the most reliable subset of good quality structures of protein-small molecule complexes from the PDB, with reliable affinity data available and use this data to develop simple consistent procedure of the automatic data processing in such a way, that it could be used to recalibrate AutoDock force field and improve the quality of the binding mode and affinity prediction, especially for the hydrated binding sites.

https://doi.org/10.1016/j.jmgm.2017.09.016

2. Methods

2.1. Dataset compilation

The set of receptor-ligand complexes was selected on the bases of the information extracted from the BioLiP database. To obtain as numerous dataset as possible with the best possible quality of the structures, only PDB files with the resolution 2.3 Å or better were accepted. Since structures with resolution above 2.0 Å might be considered as less reliable, additional structure quality criterion was used. Namely, if the R-free value for such structure exceeded 2.3 or R-free and R-value difference was greater than 0.05 then this structure was rejected. Also multimeric receptors and structures with multiple binding sites or multiple ligands were removed from the dataset. Particularly if any other molecules, than just water, were present within a distance of 8 Å from the biologically relevant ligand, such structure was also rejected.

Because the reliability of the affinity data is an important factor for the force field parameters estimation, structures without such information available in the BioLiP database, or with inconsistent information were rejected as well. Affinity data were considered inconsistent and not reliable when the information from MOAD and PDBBind-CN databases, concerning the same receptor-ligand pair, differed significantly. During the preliminary data processing it turned out that it is the case for numerous PDB files. This inconsistency in most cases seems to arise from the incorrect unit assignment, particularly confusion of μM with nM by the scripts responsible for parsing the information from the primary literature.

After applying all these restrictions, the final dataset consisted of 310 structures.

2.2. Dataset processing

The structures selected in the protocol described above were then prepared for calculations in the following way. Since, as mentioned earlier, some water molecules found in the binding sites are important for ligand binding, they were preserved as a part of the receptor if they were bound to at least 2 receptor atoms (the distance between water oxygen and any of the protein's heteroatoms was below 3.3 Å) and its distance to any of the ligand's atoms was less than 3.5 Å. This way two subsets were compiled. The first one comprising 225 complexes with water molecules fixed as part of the receptor (referred later as hydrated receptors) and the remaining 85 pure protein-ligand complexes.

If there were any atoms missing in the receptor structure incomplete residues were fixed by means of the MODELER software [13]. Since, on one hand, correct identification of all receptor-ligand interactions inside the binding site is crucial for the proper force field calibration, but on the other hand, it is well known, that even 20% of high quality structures in Protein Data Bank (PDB) have incorrectly assigned rotameric states of Gln and Asn residues [14,15], in the next step of the receptor preparation it was processed by the Reduce tool [14]. Reduce is a program that adds hydrogens to a PDB molecular structure files and optimizes the orientations of Asn and Gln sidechain amides, as well as His rings.

In the last step of the molecular data files preparation all receptors as well as all ligand files were processed with standard AutoDock utility scripts prepare_recepto4.py and prepare_ligand4.py. At this stage, to all atoms of the receptor and ligand, appropriate Autodock's force field atom types were assigned, Gasteiger partial charges were calculated and finally nonpolar hydrogens were merged with carbon atoms.

2.3. Force field modifications and parametrization

Water molecules present in the binding sites of some receptors are in fact mediating the interactions with the ligand, and are essential for its proper positioning in the binding site and as such should be taken into account in the docking calculations. Unfortunately, since in most PDB structures water molecules are represented by oxygen atom only, adding hydrogens to these atoms in appropriate orientation is not a trivial task since it can hardly influence the predicted ligand's binding mode. To simplify the process of incorporating water molecules into the binding sites and make it less ambiguous, new water model was created. The main goal was to define a model that is compatible with the current AutoDock's force field and the source code. That on one hand would be able to mimic the bridging interactions between ligand and receptor, but on the other hand would not be sensitive to the orientation of added hydrogens. To achieve this two new atom types were defined: water oxygen OW and water hydrogen HW. Their initial parameters were copied from AutoDock's atom types OS and HS thus both these new atoms were defined as spherical hydrogen bond acceptor and donor respectively. To avoid ambiguity with proper positioning of hydrogens around the oxygen atoms for directional hydrogen bonds, water molecules were described as two concentric spheres with hydrogen bond accepting and hydrogen bond donating properties respectively (Fig. 1). In comparison to the original parameters of the OS and HS atoms, for atoms OW and HW van der Walls well depths were set to 0 and sum of the van der Waals radii (R_{ii}) for HW atom was set to 4.0 thus both atoms will be participating only in hydrogen bonding interactions.

The coefficients of the force field were estimated with an iterative least-squares procedure. First the crystallographic structures of complexes were minimized using Solis-Wet local search algorithm implemented in AutoDock, then the coefficients of the force field's energy terms were modified and the minimization procedure was repeated until convergence. The final optimized values of the free energy coefficients for van der Waals, hydrogen bonds, electrostatic, desolvation and torsional terms are summarized in Table 1.

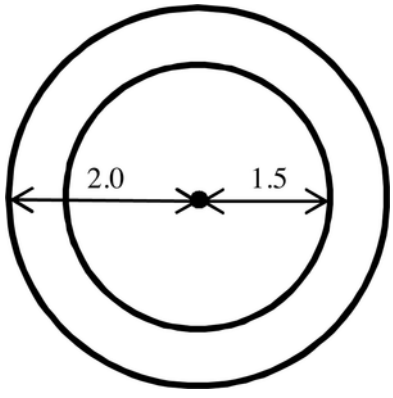

Fig. 1. Definition of the proposed, water model. Hydrogen bond acceptor atom (OW) is represented by the inner sphere ($R_i = 1.5$). Hydrogen bond donor (HW) is represented by the outer sphere ($R_i = 2.0$). Van der Waals well depths were set to 0 for both atoms thus they do not participate in this type of interactions.

Table 1
Comparison of the coefficients for the original AutoDock's "AD4.1_bound" force field and their recalibrated values.

	Coefficient				
Force field	vdW	hbond	estat	desolv	tors
AD4_bound	0.1662	0.1209	0.1406	0.1322	0.2983
After recalibration	0.1538	0.0668	0.0357	0.1003	0.0873

2.4. Docking protocol

The docking environment was defined in the same way for both, the local optimization procedure used for the force field calibration and for all redocking experiments.

The semi flexible docking simulations (with flexible ligands and rigid receptors) were carried out with AutoDock 4.2 suite of programs. For all complexes parameters were set as follows. For each receptor a grid box 20 × 120 × 120 points was defined with default resolution of 0.375 Å and the center of the box set to the geometric center of the bound ligand. Since the grid box of this size was large enough to cover not only the binding site but, in most cases, the entire receptor molecule, additional bias toward the actual binding pocket in the redocking experiments was minimized.

For the force field calibration procedure only the binding energy after Solis & Wets local optimization of the ligand position was taken into account.

For the full redocking experiments 50 independent runs for each ligand were performed with the initial population of 150 random solutions. The maximum number of 25,000,000 energy evaluations, and 27,000 generations were set for the Lamarckian genetic algorithm based search procedure. Mutation and crossover were applied to the population at rates 0.02 and 0.80, respectively and the probability of performing the local search was set to 0.06. Resulting docked ligand poses were clustered with the 2.0 Å tolerance.

2.5. Virtual screening efficiency

To evaluate the performance of the new scoring function in discriminating true binders among decoys, it was tested on the set of ligands obtained from the DUD-e database [16]. To compare the performance of the original and recalibrated force field, the receiver operating characteristic (ROC) curves were plotted for both of them and the area under curves (AUC) as well as the enrichment factors (EF) were then calculated. The area under the ROC curve is often used as a measure of the discrimination quality between true ligands and decoys. Its value varies from 0 to 1, with 0.5 indicating that the method perform no better than the random selection, while the value of 1 means perfect separation of active compounds from inactive decoys.

In virtual screening (VS) experiments it is essential, for practical reasons, that in a large set of tested compounds the active ones should be ranked by the screening algorithm at the very beginning of the ordered list. This is so called "early recognition" problem that is usually expressed as enrichment factor (EF):

$$EF_{x\%} = \frac{Act_x/N_x}{Act_t/N_t}$$

Where $EF_{x\%}$ is the enrichment factor for the subset of the first x% of scored compounds, Act_x is the number of actives found in this subset, Act_t is the total number of actives in the entire database, N_t represents the total number of compounds in this database and N_x is the number of compounds in the subset.

Since only 102 targets are represented in the DUD-e database, the receptor that is present in both, the DUD-e database and the hydrated data set was selected for VS efficiency testing (pdb id 1e66). For this particular target sets of 20 random active ligands and 80 random inactive decoys were chosen for the ROC curves preparation. The docking protocol as well as the data preprocessing procedure were identical to the ones described above. All ligands were docked and scored by the original AD4.1_bound force field as well as by the recalibrated one. On the bases of the results the ROC curves, AUC and EF_{10} were calculated.

3. Results and discussion

Improvement in the docked ligand pose and its binding affinity prediction by the modified force field was tested via comparing them to the experimental affinities and the crystal complex geometries. First it was tested how well the default AutoDock force field is able to predict the binding affinities of all PDB complexes collected in the complete dataset of 310 structures. Geometry of each complex was optimized by performing Solis-Wet local search procedure and the resulting binding energies were compared to the experimental affinities. The result is shown in Fig. 2. Surprisingly Pearson's correlation between binding energies predicted by AutoDock and experimental affinities is relatively high (0.68) with standard deviation 1.95 kcal/mol. It might be surprising, since previous reports pointed out that not only AutoDock's but most other docking programs' default scoring functions are worse in this task [17–20]. This result might be attributed to the very restrictive criteria of selecting complexes to the dataset, and processing them before actual docking, but also to the strict selection of complexes with very consistent experimental affinity data available.

To test how much binding site water molecules contribute to the calculated affinities, the calculations were repeated for the same set of complexes but with all water molecules removed from the binding sites. Again, correlation coefficient as well as the standard deviation were calculated and they turned out to be very similar to the previous values. It is known, that scoring functions are, to some extent, blind to the specific interactions [20]. This observation may indicate that, although for many complexes binding site water molecules are important, only for a few of them these water molecules are really crucial for the interactions, and scoring function is not specific enough to value the difference.

The subset of 225 hydrated complexes was used to recalibrate the AutoDock force field as described in the methods section. Similarly, the binding energies were calculated for the entire set of 310 complexes by local optimization, using the recalibrated force field and the results were compared to the experimental values (Fig. 3). The calculated Pearson's correlation coefficient improved significantly to 0.82 with the standard deviation being in this case 1.41 kcal/mol. It should be noted, that despite the fact that the force field parameters were optimized for the subset of complexes comprising only the hydrated binding sites, the improvement of the binding affinity prediction is more general. Testing the predicted/experimental affinities correlation for the subset of pure protein-ligand complexes revealed the same level of improvement (Fig. 3 right).

Another important ability of docking programs is to recreate the experimental geometry of the known complex. This can be independent from the ability of the scoring function to predict the correct binding affinity, since it is related also to the capability of the program to efficiently search the conformational space of the ligand-receptor system as well as to the specificity of the scoring function. In

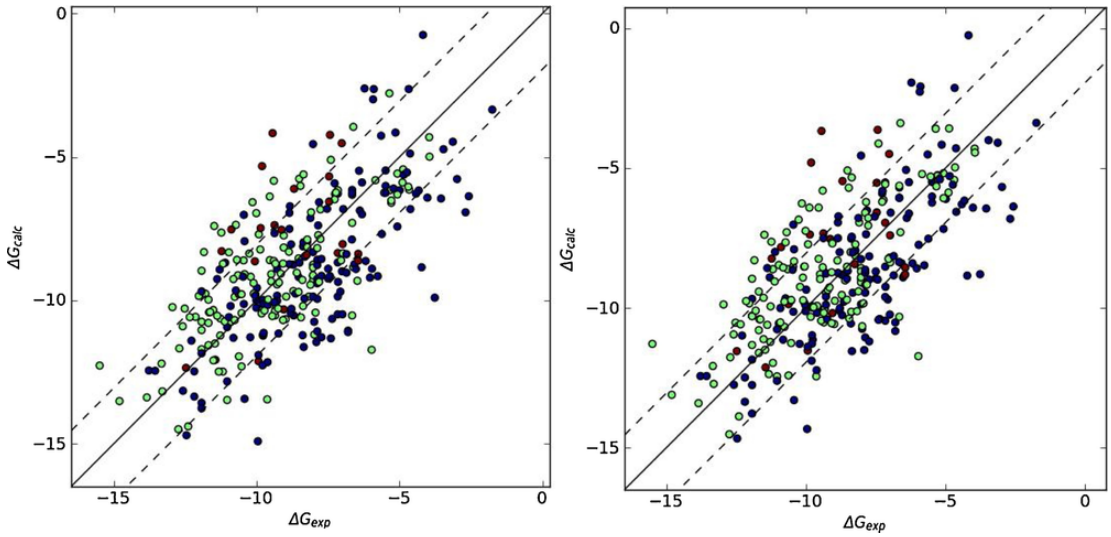

Fig. 2. Comparison of predicted versus observed binding free energies. The solid line represents a perfect fit. Dashed lines show the area of ±1 standard deviation. Colors represent the number of torsional degrees of freedom in the ligand (*n*). Blue dots $n <= 5$, green dots $5 > n <= 10$, red dots $n > 10$. The left chart shows the correlation of experimental and predicted binding affinities for the entire set of 310 complexes with 225 of them having water molecules as part of the binding site ("hydrated receptors"). The right hand chart shows the result for the same set but with all receptors deprived of waters molecules. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

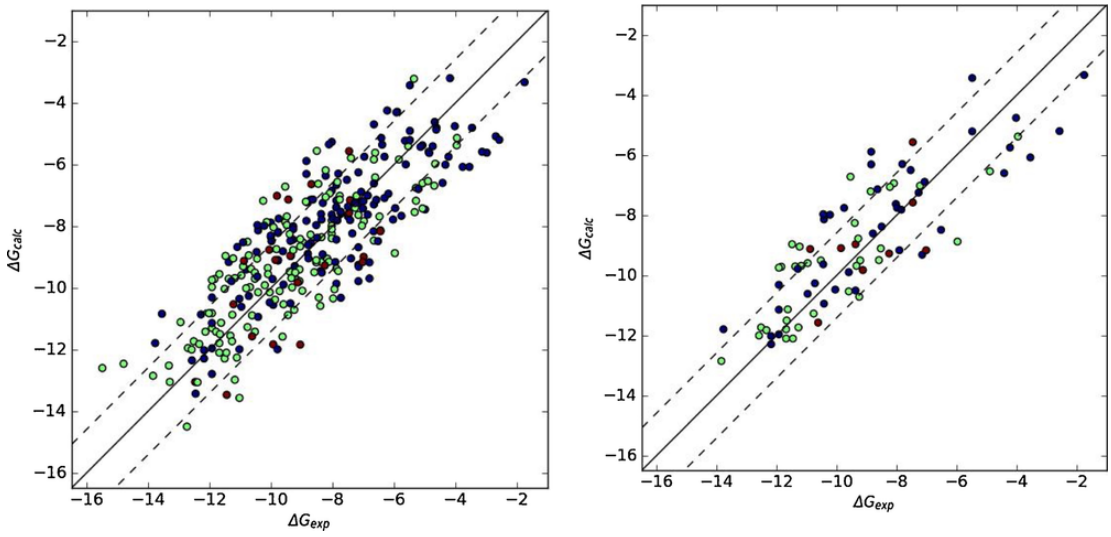

Fig. 3. The left chart shows the correlation of experimental and predicted binding affinities for the entire set of 310 complexes with 225 of them having water molecules as part of the binding site ("hydrated receptors"). The right chart shows the results for the set of 85 molecules with unhydrated binding sites.

order to test the effectiveness of both force fields in this task, complete redocking experiments were performed on the entire set of 310 complexes. Resulting binding affinities of the lowest energy poses as well as the geometries of the resulting complexes were again compared to the experimental ones. Results are shown in Figs. 4 and 5.

In this test the standard autodock's force field performed much worse. Calculated Pearson's correlation coefficient between predicted and experimental affinities was only 0.42 with standard deviation of 2.35 kcal/mol. Surprisingly it can be noted that calculated binding energy of ligands with small (5 or less) number of rotatable bonds is generally overestimated while for ligands with 6 and more rotatable bonds is underestimated. It is unexpected since, because of their additive nature, scoring functions usually tend to favor larger ligands. Yet in general these results are still slightly better than the values reported previously for much larger sets of complexes [18,20].

Examination of the lowest energy ligand poses revealed that 52% of them are true positives as their RMSD to the original crystal geometry of the complex fits into the limit of 2 Å (Fig. 5). These can be regarded as successful dockings and would be easily picked up in the blind docking experiment, even though for many of these complexes predicted binding affinity differs from the experimental one by more than 2 kcal/mol. Again, it confirms the known fact that AutoDock is much better in finding the correct ligand poses than in scoring them according to the binding affinities [18].

Analogous analysis was performed for the recalibrated force field and its results are presented in Fig. 4 (bottom) and Fig. 5 (right). Significant improvement can be noticed in both areas. First, redocking of the entire set of 310 complexes showed that the binding energies of the lowest energy ligand poses were in much better agreement with the experimental ones, with Pearson's correlation 0.72 and standard

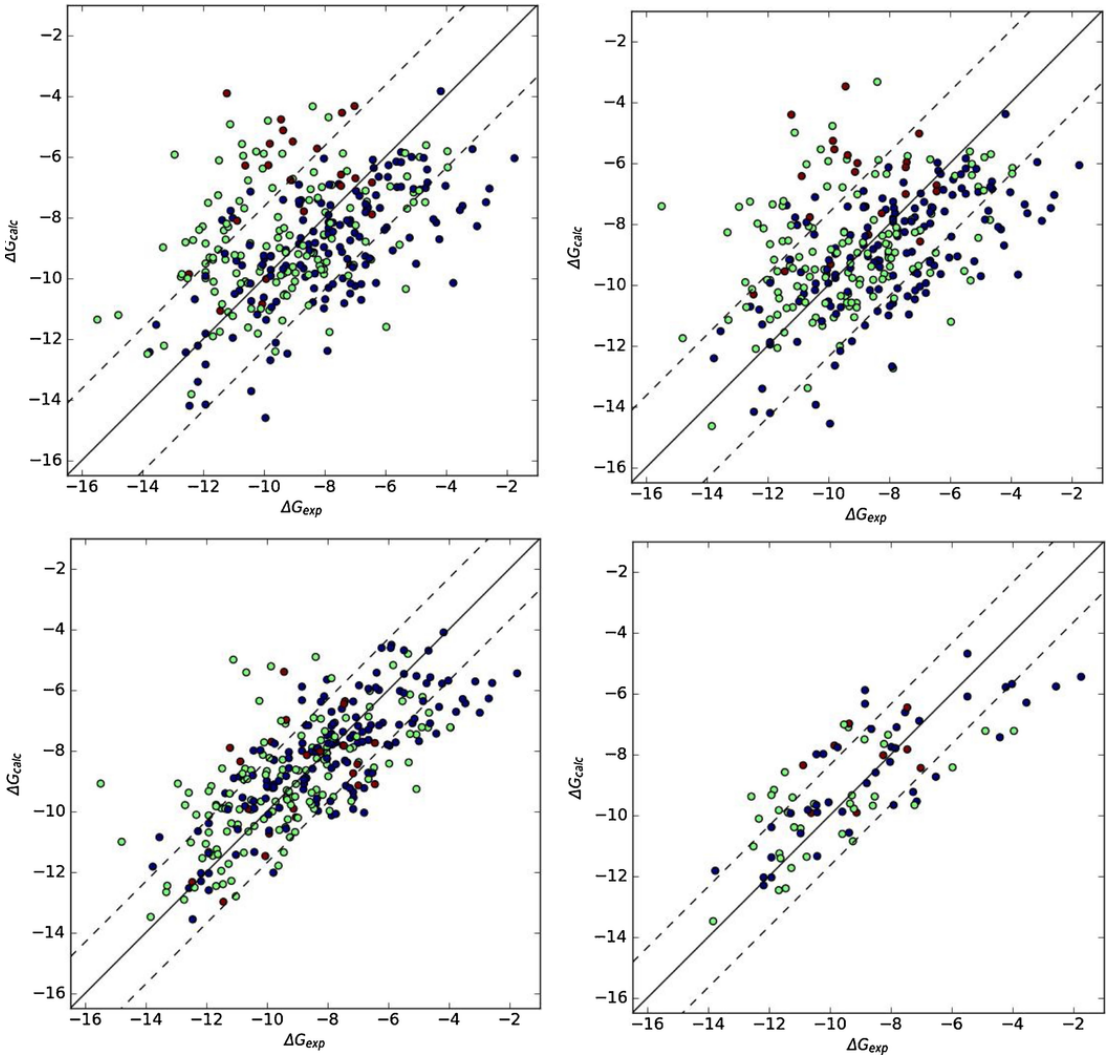

Fig. 4. The upper charts show the results of redocking for the original AutoDock force field. The top left hand chart shows the results for the complete set of 310 complexes, while the upper right hand shows the results for the same set with all waters removed from the binding site. The bottom charts show the results of redocking for the recalibrated force field. The left hand one shows the results for the entire set of 310 complexes. The bottom right hand chart shows the results of redocking for the 85 unhydrated receptors.

deviation of 1.69 kcal/mol. It is important to notice that the same level of improvement is again achieved for the complexes with and without hydrated binding sites. Second, the number of true positives, namely the lowest energy ligand poses that were different from their crystal geometries by less than 2 Å (as measured by their RMSD) also improved significantly reaching over 60% of results (Fig. 5).

Comparison of the results obtained for the standard AutoDock force field and for the recalibrated one showed that, there still seems to be room for improvement and a better set of empirical data with improved model of interactions can still result in development of a better force field.

To check how much room is actually available for the force field improvement with the current state of the AutoDock's search engine, one additional analysis was performed. Because during the previous analyses it turned out that, for some complexes, the difference of the calculated binding energy among the few top scoring poses was very low − despite the significant differences between the geometries of the resembling complexes − instead of picking simply the lowest energy pose, the ligand pose with the lowest RMSD and predicted binding energy not higher than 1 kcal/mol from the lowest one was selected (Fig. 5, bottom).

In this case the number of acceptable solutions, i.e. resulting ligand poses with RMSD from crystal pose below 2 Å, increased by about 20%, up to 72% for the standard AutoDock force field and even to 81% for the recalibrated one.

An important measure of the docking algorithm performance is its ability of picking true binders among the large dataset and grouping as many of them as possible in the top scoring molecules. To check how recalibrated force field perform in this task in comparison to the standard AutoDock's AD4.1_bound force field, the ROC plots of docking results were prepared and are shown in Fig. 6.

The recalibrated force field performed better than the original one. Not only the value of AUC improved from 0.73 to 0.89 but also the enrichment factor for 10% subset increased from 3.18 to 4.09 indicating that the new force field should perform better in the task of enriching the dataset in the virtual screening experiment for receptors, that have water molecules bound as part of their binding sites.

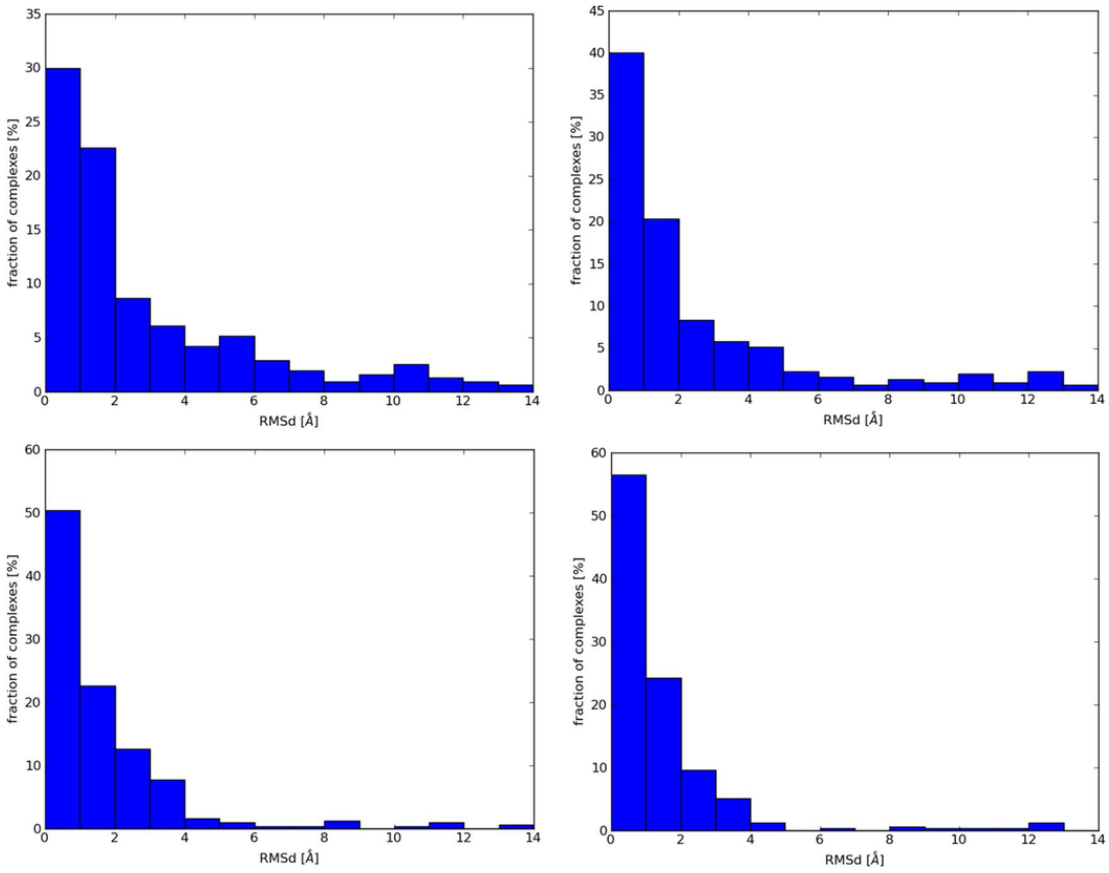

Fig. 5. Histograms showing the number of complexes redocked with particular quality. The left column shows the number of lowest energy complexes redocked with the standard AutoDock force field. The right one shows the results for the recalibrated force field. The top two histograms depict the number of the lowest energy ligand poses at particular RMS distance from the crystal structure. The bottom two show the number of complexes with particular RMS distance from the crystal structure, that have estimated binding energy not higher than 1 kcal/mol from the lowest one.

4. Summary and conclusion

A set of 310 high quality structures of protein-ligand complexes with experimentally determined binding affinities has been extracted from the Protein Data Bank and used to test and recalibrate AutoDock force field. It is known that the quality of the results obtained from docking experiments may be dependent on many factors including the data preparation procedure and even the personal preferences and experience of the scientist performing the calculations with particular docking software [8,21]. This, results in sometimes inconsistent conclusions drawn by various researchers, even when referring to the same programs and datasets [18,22].

The goal of this work was to compile a set of PDB files with the best possible quality and for which reliable experimental affinity data is available. Then, a simple and easy to reproduce procedure of processing these files and performing a docking experiment on a large number of complexes was formulated. It is clear that very restrictive choice of the structures for the training set and appropriate further processing them helps to develop a better model of intermolecular interactions and later improve the results of the docking experiments. It turned out, that the standard AutoDock force field performed, on the compiled set of complexes, as good as it was reported in the previous pose prediction test, but slightly better in the affinity prediction [18].

However, after recalibration of the force field much better correlation between the computed and experimentally determined binding affinities was achieved and the quality of pose prediction improved even more. Closer examination of the pose prediction results, revealed that despite the imperfections of the AutoDock scoring function, its search engine is quite effective and even if the lowest energy pose is not the correct one, the correct solution can be found in most cases, as one of the low energy (within a 1 kcal/mol range from the top) poses. This solution cannot be easily found by simple “lowest energy” criterion, but can be identified by a researcher, based on his experience or a priori knowledge of the particular receptor.

On the other hand this 20% gap between the number of top scoring and the lowest RMSD solutions clearly shows that there is still some space for the further improvement of the AutoDock force field. Availability of a large number of very good quality structural and experimental affinity data should allow to create a force field that will be more sensitive to the specific ligand-receptor interactions.

Acknowledgements

All plots were prepared with in-house written python scripts and matplotlib library [23].

All calculations were performed using the computing facilities of the Center of Excellence ChemBioFarm at Gdansk University of Technology as well as the resources of the Academic Computer Centre in Gdańsk (TASK).

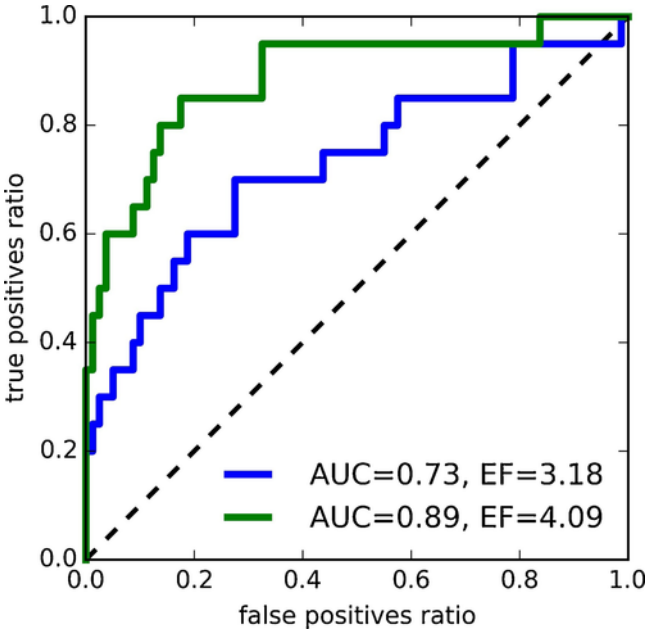

Fig. 6. ROC curves plotted for the results of docking with the original AD4.1_bound force field (blue) and recalibrated one (green). Enrichment factors are calculated for the top 10% of the data set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

[1] G. Morris, R. Huey, AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256, (AutoDock4).

[2] D. Santos-Martins, S. Forli, M.J. Ramos, A.J. Olson, AutoDock4Zn: An improved AutoDock force field for small-molecule docking to zinc metalloproteins, J. Chem. Inf. Model. 54 (2014) 2371–2379, https://doi.org/10.1021/ci500209e.

[3] S. Uehara, S. Tanaka, AutoDock-GIST: Incorporating thermodynamics of active-site water into scoring function for accurate protein-ligand docking, Molecules 21 (2016) https://doi.org/10.3390/molecules21111604.

[4] V.Y. Tanchuk, V.O. Tanin, A.I. Vovk, G. Poda, A new, improved hybrid scoring function for molecular docking and scoring based on AutoDock and AutoDock vina, Chem. Biol. Drug Des. 87 (2016) 618–625, https://doi.org/10.1111/cbdd.12697.

[5] R. Wang, X. Fang, Y. Lu, S. Wang, The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures, J. Med. Chem. 47 (2004) 2977–2980, https://doi.org/10.1021/jm030580l.

[6] L. Hu, M.L. Benson, R.D. Smith, M.G. Lerner, H.A. Carlson, Binding MOAD (Mother of all databases), Proteins Struct. Funct. Genet. 60 (2005) 333–340, https://doi.org/10.1002/prot.20512.

[7] J. Yang, A. Roy, Y. Zhang, BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions, Nucleic Acids Res. 41 (2013) https://doi.org/10.1093/nar/gks966.

[8] C.R. Corbeil, C.I. Williams, P. Labute, Variability in docking success rates due to dataset preparation, J. Comput. Aided Mol. Des. 26 (2012) 775–786, https://doi.org/10.1007/s10822-012-9570-1.

[9] P. Cozzini, M. Fornabaio, A. Marabotti, D.J. Abraham, G.E. Kellogg, A. Mozzarelli, Free energy of ligand binding to protein: evaluation of the contribution of water molecules by computational methods, Curr. Med. Chem. 11 (2004) 3093–3118, https://doi.org/10.2174/0929867043363929.

[10] S. Forli, A.J. Olson, A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking, J. Med. Chem. 55 (2012) 623–638, https://doi.org/10.1021/jm2005145.

[11] F. Österberg, G.M. Morris, M.F. Sanner, A.J. Olson, D.S. Goodsell, Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in autodock, Proteins Struct. Funct. Genet. 46 (2002) 34–40, https://doi.org/10.1002/prot.10028.

[12] H. Claußen, C. Buning, M. Rarey, T. Lengauer, FlexE: efficient molecular docking considering protein structure variations1 1Edited by J. Thornton, J. Mol. Biol. 308 (2001) 377–395, https://doi.org/10.1006/jmbi.2001.4551.

[13] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815, https://doi.org/10.1006/jmbi.1993.1626.

[14] J.M. Word, S.C. Lovell, J.S. Richardson, D.C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285 (1999) 1735–1747, https://doi.org/10.1006/jmbi.1998.2401.

[15] C.X. Weichenberger, M.J. Sippl, Self-consistent assignment of asparagine and glutamine amide rotamers in protein crystal structures, Structure 14 (2006) 967–972, https://doi.org/10.1016/j.str.2006.04.002.

[16] M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, Directory of useful decoys, enhanced (DUD-E): Better ligands and decoys for better benchmarking, J. Med. Chem. 55 (2012) 6582–6594, https://doi.org/10.1021/jm300687e.

[17] R. Wang, Y. Lu, X. Fang, S. Wang, An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes, J. Chem. Inf. Comput. Sci. 44 (2004) 2114–2125, https://doi.org/10.1021/ci049733j.

[18] D. Plewczynski, M. Łaźniewski, R. Augustyniak, K. Ginalski, Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, J. Comput. Chem. 32 (2011) 742–755, https://doi.org/10.1002/jcc.21643.

[19] G.L. Warren, C.W. Andrews, A.M. Capelli, B. Clarke, J. LaLonde, M.H. Lambert, M. Lindvall, N. Nevins, S.F. Semus, S. Senger, G. Tedesco, I.D. Wall, J.M. Woolven, C.E. Peishoff, M.S. Head, A critical assessment of docking programs and scoring functions, J. Med. Chem. 49 (2006) 5912–5931, https://doi.org/10.1021/jm050362n.

[20] R. Kim, J. Skolnick, Assessment of programs for ligand binding affinity prediction, J. Comput. Chem. 29 (2008) 1316–1331, https://doi.org/10.1002/jcc.20893.

[21] M.J. Hartshorn, M.L. Verdonk, G. Chessari, S.C. Brewerton, W.T.M. Mooij, P.N. Mortenson, C.W. Murray, Diverse, high-quality test set for the validation of protein-ligand docking performance, J. Med. Chem. 50 (2007) 726–741, https://doi.org/10.1021/jm061277y.

[22] J.C. Cole, C.W. Murray, J.W.M. Nissink, R.D. Taylor, R. Taylor, Comparing protein-ligand docking programs is difficult, Proteins Struct. Funct. Genet. 60 (2005) 325–332, https://doi.org/10.1002/prot.20497.

[23] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (2007) 99–104, https://doi.org/10.1109/MCSE.2007.55.