Supercomputer Docking: Investigation of Low Energy Minima of Protein-Ligand Complexes

Danil C. Kutov[1,2]*, Alexey V. Sulimov*[1,2]*, Vladimir B. Sulimov*[1,2]

It is shown that the global energy minimum of a protein-ligand complex, when the energy is calculated by the PM7 quantum-chemical semiempirical method with the COSMO implicit solvent model, can be determined as follows. First, the low energy minima are found by a docking program when the protein-ligand energy is calculated with the MMFF94 force field in vacuum. Second, energies of all these minima are recalculated with the PM7 method and the COSMO implicit solvent model. Third, among these recalculated energies the minimal energy is determined and the respective minimum is the global energy minimum when the energy is calculated with the PM7 method and the COSMO implicit solvent model. The optimal width of the spectrum of low energy minima found with MMFF94 in vacuum is determined to perform minimal quantity of quantum-chemical recalculations. The proposed approach allows to perform docking in solvent with the quantum-chemical method and to increase the docking positioning accuracy.

Keywords: docking, protein-ligand, global minimum, force field, quantum-chemical method.

Introduction

Docking is the most demanded software for the development of new medicine. Docking performs positioning of ligands (small molecules) in protein and computes the protein-ligand binding energy. For the effective application of docking to the drug discovery, the docking accuracy must be improved. The satisfactory positioning accuracy demonstrated by many docking programs is insufficient to obtain the high enough accuracy of the binding energy calculation.

To overcome this problem, supercomputer docking programs have been recently developed [4, 9]. These programs do not use most simplifications worsening the accuracy of many existing docking programs, and they are based on the docking paradigm [4] assuming that the global minimum (GM) of the energy of the protein-ligand complex corresponds to the experimentally detected ligand pose in the protein. The docking accuracy depends on the energy calculation method, and best results [8] are obtained with the PM7 quantum-chemical semiempirical method [7] and the COSMO implicit solvent model [3]. Unfortunately the PM7 method demands much larger computational resources than classical force fields, as well as docking in vacuum is faster than docking in solvent. Nevertheless, docking with PM7 can be realized using the quasi-docking procedure [8] when a sufficiently broad spectrum of low energy minima is found using a force field, and then energies of all these minima are recalculated with PM7 and COSMO. By this procedure we are able to identify the ligand pose corresponding to the global energy minimum for PM7 with solvent, and the goal of docking is reached. The quasi-docking procedure saves a great deal of computational resources because it is necessary to fulfil many hundreds of thousands of local optimizations for the reliable detection of low energy minima [4], but for the energy recalculation several thousand low energy minima found with the force field are enough for the detection of the global minimum with the PM7+COSMO energy [8]. However, for practical application of the quasi-docking procedure we should answer the following question. How many low energy minima found with a given force field must be saved and then

recalculated with PM7 to detect the global PM7 energy minimum? The answer is found, and respective results are presented below.

1. Materials and Methods

The FLM docking program [4] is used to search for low energy minima of protein-ligand complexes. FLM-0.05 utilizes the MMFF94 force field [2] in vacuum, and docking is performed by randomly tossing the ligand into the protein followed by the local optimization of the energy of the protein-ligand system with variations of Cartesian coordinates of all ligand atoms and keeping fixed all protein atoms. FLM finds and saves a pool of a given number of unique low energy minima. This pool consists of the global minimum and every successive minimum above it. FLM performs parallel processing until the set of low energy minima stops to get renewed. FLM-0.10 selects low energy minima on the base of their MMFF94 energy in water with the PCM implicit solvent model [10]. FLM-0.05 is much faster than FLM-0.10 because the latter consumes additional computing resources for PCM computations. Energy calculations by the PM7 method with the COSMO solvent model [3] are performed by the MOPAC program [6]. 16 protein-ligand complexes [8] with good structures taken from Protein Data Bank [1] are utilized.

2. Results

The quasi-docking procedure is performed on the base of 8192 low energy minima found by FLM-0.10 for every test protein-ligand complex [8]. This set of minima is designated as {2}MMFF94+PCM. Then, energies of all these minima are recalculated with PM7+COSMO. The ligand pose corresponding to the global PM7+COSMO energy minimum is near the crystallized native ligand pose for almost all test complexes [8], and therefore, the docking paradigm is fulfilled for them.

Is it possible to accelerate the quasi-docking procedure and to find the pool of low energy minima with MMFF94 in vacuum using the much faster FLM-0.05 program? In order to answer this question, pools of 8192 low energy minima are found (we designate them as {1}MMFF94) by the FLM-0.05 program. Next, for every complex we perform the search in {1}MMFF94 for the minimum corresponding to the ligand pose which is closest (by RMSD) to the pose corresponding to the global minimum of PM7+COSMO energy in {2}MMFF94+PCM. Values of such minimal RMSD are presented in Fig. 1. The PDB id is the identification of the 3D-structure from Protein Data Bank [1]. Values in bars are differences of PM7+COSMO energies in respective minima.

As we can see, for 10 complexes (out of 16) among {1}MMFF94 minima it is possible to identify a minimum which corresponds to the ligand pose in the close proximity of the ligand pose corresponding to the global minimum of PM7+COSMO energy from the {2}MMFF94+PCM set. It can also be seen that the close proximity can be defined by the condition RMSD < 0.1Å, because the respective difference in PM7+COSMO energies is sufficiently small less than 0.5 kcal/mol. However, for other 6 complexes the respective global minimum of PM7+COSMO energy is not revealed among {1}MMFF94 minima: for these complexes RMSD > 0.1Å, and the respective difference of PM7+COSMO energies is larger than 1 kcal/mol. One reason for this can be related to the insufficient quantity of low energy minima stored by FLM-0.05.

To support this explanation, the MMFF94 in vacuum energy bands are presented in Fig. 2: the energy band (green) occupied by 4096 unique low energy minima from the {1}MMFF94 set and the energy band (beige) between the energy of the ligand pose corresponding to the global
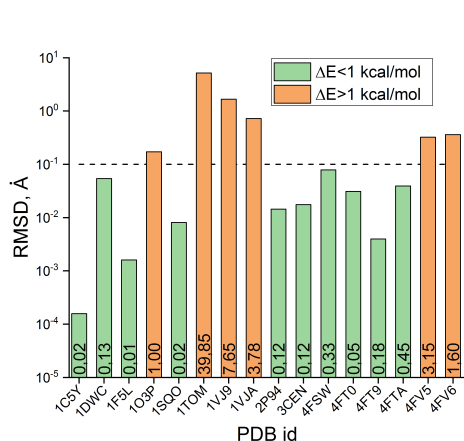
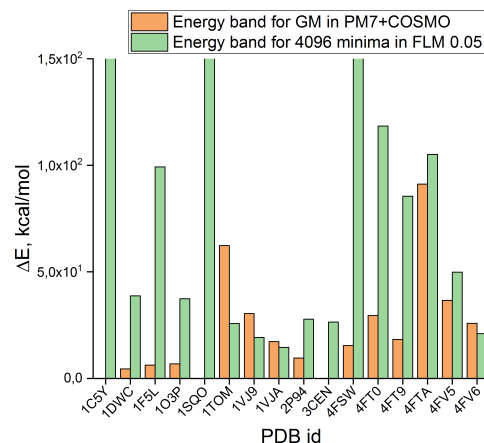**Figure 1.** The minimal RMSD of the PM7+COSMO GM from MMFF94 minima

Figure 2. Energy bands calculated with MMFF94 in vacuum

PM7+COSMO energy minimum (from the {2}MMFF94+PCM set) and the energy of the global minimum from the {1}MMFF94 set.

As we can see for four complexes (1TOM, 1VJ9, 1VJA and 4FV6), the beige bars are higher than the green ones. This means that the energy bands occupied by minima found by FLM-0.05 with MMFF94 in vacuum do not cover the energy range needed to detect the global PM7+COSMO energy minimum from the {2}MMFF94+PCM set. For all these 4 complexes RMSD > 0.1Å in Fig. 1 and the global minimum of PM7+COSMO energy is not detected by FLM-0.05. For the remaining two complexes (1O3P, 4FV5) with RMSD > 0.1Å in Fig. 1 the possible reason can be related to incompleteness of the low energy minima set found by the FLM-0.05 program due to casual nature of the minima search algorithm. It is worth to note that for three complexes (1C5Y, 1SQO and 3CEN) the global minimum of the energy computed with MMFF94 in vacuum coincides with the global minimum of the PM7+COSMO energy, and the respective energy band is zero.

Conclusion

The presented results show that the quasi-docking procedure with the PM7 quantum-chemical method in solvent described by the COSMO model can be performed on the base of low energy minima found by docking when the energy is calculated with the MMFF94 force field in vacuum. For reliable finding of the global minimum of the protein-ligand energy calculated with PM7 and COSMO methods, sufficiently large number of unique low energy minima computed with the MMFF94 force field in vacuum must be identified and stored by docking. This number is specific for a given complex, it is determined by the energy band occupied by these minima, and this number is less than 4096 for most of test complexes. The energy band (MMFF94 in vacuum) occupied by these minima can reach several dozen kcal/mol for some complexes.

Acknowledgments

facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011 [5].

References

1. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The protein data bank. Nucleic Acids Research 28(1), 235–242 (2000), DOI: 10.1093/nar/28.1.235

2. Halgren, T.A.: Merck molecular force field. Journal of Computational Chemistry 17(5-6), 490–519 (1996), DOI: 10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

3. Klamt, A., Schuurmann, G.: COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. J. Chem. Soc-Perkin Trans. 2 (5), 799–805 (1993), DOI: 10.1039/P29930000799

4. Oferkin, I.V., Katkova, E.V., Sulimov, A.V., Kutov, D.C., Sobolev, S.I., Voevodin, V.V., Sulimov, V.B.: Evaluation of Docking Target Functions by the Comprehensive Investigation of Protein-ligand Energy Minima. Adv Bioinformatics 2015, 126858 (2015)

5. Sadovnichy, V., Tikhonravov, A., Voevodin, V., Opanasenko, V.: "Lomonosov": Supercomputing at Moscow State University. In: Contemporary High Performance Computing: From Petascale toward Exascale. pp. 283–307. Chapman & Hall/CRC Computational Science, CRC Press, Boca Raton, United States (2013)

6. Stewart, J.J.: MOPAC2016. `http://OpenMOPAC.net` (2016), accessed: 2018-09-06

7. Stewart, J.: Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. J. Mol. Model. 19(1), 1–32 (2013), DOI: 10.1007/s00894-012-1667-x

8. Sulimov, A.V., Kutov, D.C., Katkova, E.V., Ilin, I.S., Sulimov, V.B.: New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking. J. Mol. Graph. Model. 78, 139–147 (2017), DOI: 10.1016/j.jmgm.2017.10.007

9. Sulimov, A.V., Zheltkov, D.A., Oferkin, I.V., Kutov, D.C., Katkova, E.V., Tyrtyshnikov, E.E., Sulimov, V.B.: Evaluation of the novel algorithm of flexible ligand docking with moveable target-protein atoms. Comput. Struct. Biotechnol. J. 15, 275–285 (2017), DOI: 10.1016/j.csbj.2017.02.004

10. Sulimov, V.B., Mikhalev, A.Y., Oferkin, I.V., Oseledets, I.V., Sulimov, A.V., Kutov, D.C., Katkova, E.V., Tyrtyshnikov, E.E.: Polarized continuum solvent model: Considerable acceleration with the multicharge matrix approximation. Int. J. Appl. Eng. Res. 10(24), 44815–44830 (2015)