NOVEL GRIDLESS PROGRAM SOL-P FOR FLEXIBLE LIGAND DOCKING WITH MOVEABLE PROTEIN ATOMS

Alexey Sulimov (1,2), Dmitry Zheltkov (3), Igor Oferkin (1), Danil Kutov (1,2), Eugene Tyrtyshnikov (3,4),

Vladimir Sulimov (1,2)

1) Dimonta, Ltd., Nagornaya Street 15, Building 8, Moscow 117186, Russia
2) Research Computer Center of Lomonosov Moscow State University, Leninskie Gory 1, Building 4, Moscow 119992, Russia
3) Faculty of Computational Mathematics and Cybernetics of Lomonosov Moscow State University, Leninskie Gory 1, Building 52, Moscow 119992, Russia
4) Institute of Numerical Mathematics of Russian Academy of Sciences, Gubkin str., 8, Moscow, 119333, Russia

The docking algorithm is based on the Tensor Train (TT) decomposition [1] and the TT-CROSS method [2] for multidimensional arrays (tensors), the novel TT-CROSS-based global optimization method [3] and the local energy optimization by the simplex method with respect to ligand atoms as well as some target-protein atoms. There is no the preliminary calculated grid of the ligand probe atoms energies in the field of the target protein atoms and the energy of the protein-ligand complex is calculated directly in the frame of MMFF94 force field for any complex conformation. The conformation space of the system coordinates is formed by translations and rotations of the ligand as a whole, by the ligand torsions and also by Cartesian coordinates of selected target-protein atoms. Low energy minima of the protein-ligand complex including the global minimum are found as a result of the SOL-P docking program performance. The parallel MPI (message passing interface) based SOL-P program is written on C++ with usage of BLAS and LAPACK libraries, and its main parameters are: the maximal rank r_{max} of the matrix cross approximation method, the power m of the discretization degree of the search space (the initial grid size is equal to 2^m) and the number of iterations n_i. The initial grid is introduced in the d-dimensional search space to transform the continuous global optimization problem to the discrete one: finding the maximal in magnitude element of a d-dimensional tensor. The discrete problem is solved by a technique based on the TT-cross interpolation machinery applied cleverly, although heuristically, to selected submatrices in the unfolding matrices of the given tensor. The docking algorithm performance has been validated by comparing the revealed low energy minima with those found by the exhaustive supercomputer low energy minima search for a set of 30 protein-ligand complexes in the frame of rigid protein approximation and fully flexible ligands [4]. It was shown that SOL-P for the rigid target-protein was able to find the global minimum for 50% of investigated protein-ligand complexes. The moveable protein atoms are determined by the ligand poses and by the radius defining spheres with centers at the ligand atoms: all protein atoms inside the spheres are considered moveable. Mobility of protein and ligand atoms is taken into account in the docking process simultaneously and equally. The dependence of the docking time on the number of the protein moveable atoms is linear up to at least 60 atoms (200 degrees of freedom of the protein-ligand complex). It is shown that SOL-P is able to dock ligands with several torsions and several dozen moveable protein atoms over a reasonable time, e.g. for r_{max}=4, m=8 and n_i=10 the docking time of the native ligand (7 torsions) into the target protein (PDB ID 3CEN) is 100 minutes at 256 computational cores of Lomonosov supercomputer [5]. The Dependence of docking positioning quality is investigated as a function of the docking algorithm parameters as well as the number of moveable protein atoms. Ability to reproduce with docking calculations experimental values of the protein-ligand binding energy is investigated as a function of the number of moveable protein atoms. Acceleration and efficiency of the docking procedure as function of the number of computational cores is investigated. The work was financially supported by the Russian Science Foundation, Agreement no. 15-11-00025.

References

1) Oseledets I.V., Tyrtyshnikov E.E. Breaking the Curse of Dimensionality, or How to Use SVD in Many Dimensions. SIAM Journal on Scientific Computing, 2009, vol. 31, no.5, pp.3744-3759.
2) Oseledets I.V., Tyrtyshnikov E.E. TT-Cross Approximation for Multidimensional Arrays. Linear Algebra and its Applications, 2010, vol. 432, no. 1, pp. 70-88.
3) Zheltkov D.A., Oferkin I.V., Katkova E.V., Sulimov A.V., Sulimov V.B., Tyrtyshnikov E.E. TTDock: Docking Method Based on Tensor Train. Numerical methods and programming, 2013, vol. 14, pp. 279-291. (in Russian)
4) Oferkin I.V., Zheltkov D.A., Tyrtyshnikov E.E., Sulimov A.V., Kutov D.C., Sulimov V.B. Evaluation of the docking algorithm based on Tensor Train global optimization. Bulletin of the South Ural State University. Ser. Mathematical Modelling, Programming & Computer Software (Bulletin SUSU MMCS), 2015, vol. 8, no. 4, pp. 83-99.
5) Sadovnichy V., Tikhonravov A., Voevodin V., and Opanasenko V. "Lomonosov": Supercomputing at Moscow State University. In Contemporary High Performance Computing: From Petascale toward Exascale (Chapman & Hall/CRC Computational Science), pp.283-307, Boca Raton, USA, CRC Press, 2013