DOI: 10.14529/jsfi180405

Benchmarking Quantum Chemistry Methods in Calculations of Electronic Excitations

Vladimir A. Mironov[1], ***Bella L. Grigorenko***[1,2], ***Igor V. Polyakov***[1], ***Alexander V. Nemukhin***[1,2]

Quantum chemistry methods are applied to obtain numerical solutions of the Schrödinger equation for molecular systems. Calculations of transitions between electronic states of large molecules present one of the greatest challenges in this field which require the use of supercomputer resources. In this work we describe the results of benchmark calculations of electronic excitation in the protein domains which were designed to engineer novel fluorescent markers operating in the near-infrared region. We demonstrate that such complex systems can be efficiently modeled with the hybrid qunatum mechanics/molecular mechanics approach (QM/MM) using the modern supercomputers. More specifically, the time-dependent density functional theory (TD-DFT) method was primarily tested with respect to its performance and accuracy. GAMESS (US) and NWChem software were benchmarked in direct and storage-based TDDFT calculations with the hybrid B3LYP density functional, both showing good scaling up to 32 nodes. We note that conventional SCF calculations greatly outperform direct SCF calculations for our test system. Accuracy of TD-DFT excitation energies was estimated by a comparison to the more accurate *ab initio* XMCQDPT2 method.

Keywords: quantum chemistry, multi-scale approaches, parallel algorithms, fluorescent proteins.

Introduction

Over the past few decades, advancements in supercomputer technology led to a dramatic rise of computational resources available to scientists in chemistry and physics. Modern computational chemistry methods achieve numeric solution of the Schrödinger equation for molecular systems with different kind of approximations. The ultimate goal is to have an accurate enough description of a very large molecular system with as little computational effort as possible. Complexity of the computational chemistry methods ranges from $\mathcal{O}(N)$ to $\mathcal{O}(N!)$ where N is the size of a molecular system, while usually the most accurate methods are the most expensive ones. A popular compromise which allows conducting a routine study of large biomolecules is the use of fragmentation techniques and the density functional theory approach. Such methods are available in a wide number of modern quantum chemistry software which is more or less adapted to efficient use of modern computer clusters of multicore nodes.

In this communication, performance and accuracy of the time-dependent density functional theory (TD-DFT) in calculations of electronic excitation in a complex molecular model system (described in the next subsection) were studied. The TD-DFT implementations in two popular open-source quantum chemistry packages were used for benchmarking, namely NWChem [5] and GAMESS (US) [4, 7]. Both timings and the computed vertical excitation parameters (energies and oscillator strengths) were collected for analysis.

[1]Department of Chemistry, M.V. Lomonosov Moscow State University, Moscow, Russia
[2]N.M. Emanuel Institute of Biochemical Physics, Russian Academy of Sciences, Moscow, Russia

1. Methodology

The scaling and performance benchmarks were carried out on "Lomonosov-2" [3] supercomputer (1696 nodes, single-socket Intel Haswell-EP E5-2697v3, 64 Gb RAM, NVidia Tesla K40M). We used the NWChem v6.6 package installed on the supercomputer. The GAMESS (US) 2016r1 package was compiled manually using Intel Parallel Studio 2015. Both packages were linked with OpenMPI 1.8.4 message-passing library.

The model chemical system shown in Fig. 1 was used for benchmarking of quantum chemical calculations of electronic excitations.

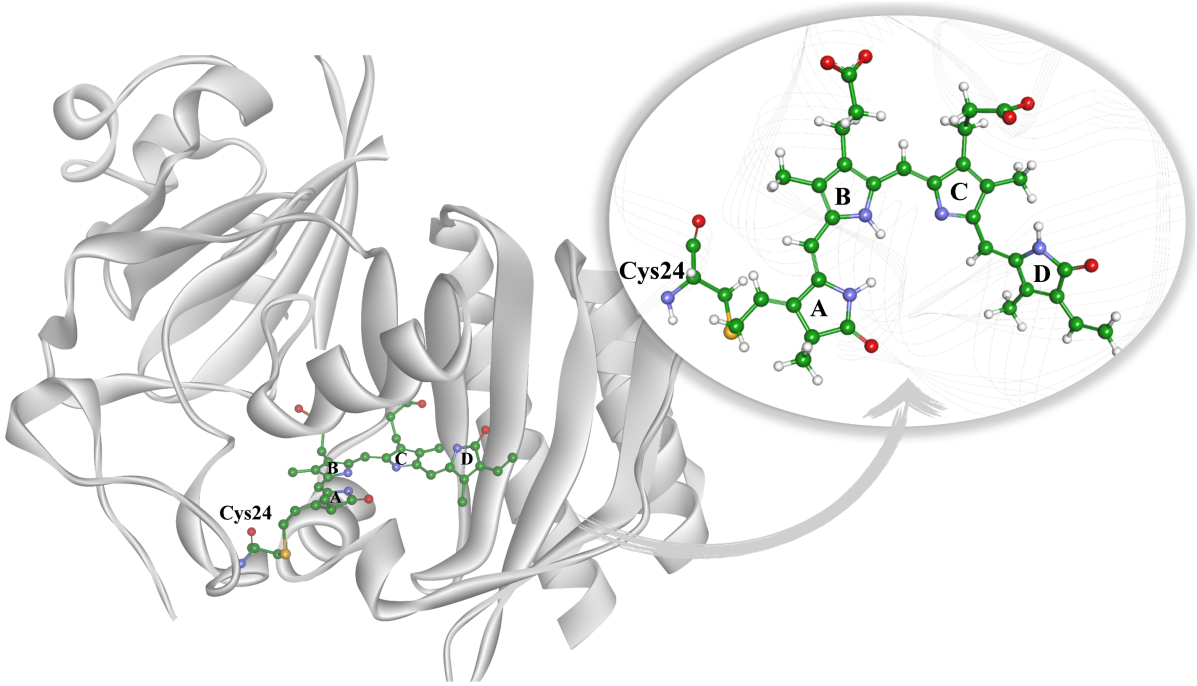

Figure 1. Molecular model of the near-infrared fluorescent protein (left) and the protein-bound biliverdin (BV) chromophore BV (right inset). Carbon atoms are shown in green, oxygen in red, nitrogen in blue, sulfur in yellow, hydrogen in white

Infrared and near-infrared fluorescent proteins are highly in demand for in vivo imaging because their absorption and emission bands fall into the optical transparency window of biological materials [2]. These proteins are engineered from the bacterial phytochrome domains. Detailed characterization of structural and spectral properties of the chromophore-containing protein domains is a necessary step when designing novel variants of these efficient fluorescent markers in living cells.

The equilibrium geometry parameters of this model system were computed using the quantum mechanics/molecular mechanics (QM/MM) approach. All QM/MM calculations were carried out using NWChem quantum chemistry package [5]. The QM subsystem of the optimized protein structure was used in TD-DFT benchmarks. The B3LYP functional [6] and a cc-pVDZ basis set were used in these TD-DFT calculations.

2. Results and Discussion

Each single point TD-DFT calculation has two steps. The first step is a regular ground-state DFT calculation. On this step the Kohn–Sham equations are solved through the usual iterative

self-consistent field (SCF) procedure. The computed density is used on the second step which is actual TD-TFT calculation. Results of the second step are excitation parameters.

There are two algorithms of the SCF procedure. According to the storage-based SCF algorithm, all the required intermediate data are computed prior the first SCF iteration and reused on each following iterations. However, the amount of these data is usually very high. According to another SCF algorithm ("direct" SCF), the intermediate data are not stored at all but recomputed on every SCF iteration. In this case the overhead of recomputing intermediate data increases linearly with the number of iterations. The amount of SCF iterations required to solve the Kohn–Sham equations for the BV chemical system is quite large: it is ≈ 50 for both packages. Thus, it it not surprising that the storage-based SCF algorithm outperforms the direct one (see Tab. 1). The drawback of the storage-based SCF algorithm is its high demand for storage

Table 1. Strong scaling of single-point TD-DFT (B3LYP) calculation of the BV chemical system in NWChem quantum chemistry package. The data is applicable for both direct and storage-based SCF algorithms

Number of nodes	Time, s		Scaling efficiency		
	direct	storage-based	direct, rel. to 1 node	direct, rel. to 8 nodes	storage-based, rel. to 8 nodes
			NWChem		
1	16 455	-	100%	-	-
4	7344	-	56%	-	-
8	3907	2310	53%	100%	100%
16	2204	1288	47%	89%	90%
24	1619	1047	42%	80%	74%
32	1487	995	35%	66%	58%
			GAMESS (US)		
1	10 738	-	100%	-	-
4	2776	-	97%	-	-
8	1446	959	93%	100%	100%
16	788	539	85%	92%	89%
24	581	414	77%	83%	77%
32	474	349	71%	76%	69%

bandwidth and capacity. In this study, a `tmpfs` in-memory partition was used to hold intermediate data. The calculations were therefore limited by the number of nodes used and the amount of system memory on each node. In the particular setup of BV chromophore benchmark, the limit was 8 nodes having 32 GB memory per node. Additionally, the lack of I/O in direct SCF results in a somewhat better scaling when compared to the storage-based algorithm. It is especially important for the NWChem package when the SCF step in a storage-based calculation is up to 4 times faster than a direct SCF step (Tab. 2). We cannot directly compare the performance of the NWChem and GAMESS (US) quantum chemistry packages because they use different setups of exchange-correlation functional integration, different convergence criteria and other hidden

Table 2. The timings of SCF and excitation calculation steps in TD-DFT (B3LYP) benchmark for direct and storage-based SCF algorithms. The benchmark is for 16 nodes, wall clock time is given in seconds

Algorithm	Step	NWChem,s	GAMESS,s
direct SCF	SCF	1175	357
	excitation	1029	431
	total	2204	788
storage-based SCF	SCF	236	135
	excitation	1052	404
	total	1288	539

parameters. However, when using the default configuration in both packages, GAMESS (US) runs faster than NWChem on "Lomonosov-2" supercomputer on he same amount of nodes.

The recent paper [1] compares the results obtained using the TD-DFT and the more accurate *ab initio* XMCQDPT2 method.

Conclusion

TDDFT benchmarks show good scalability for the test system for up to 32 nodes, while GAMESS (US) significantly outperforms NWChem in both single node performance and in scaling. Conventional SCF calculations greatly outperform direct SCF calculations for our test system, and thus conventional SCF procedure should be used whenever a fast-enough storage is available.

Acknowledgements

This work was supported by the Russian Science Foundation (project №17-13-01051). This research is carried out using the equipment of HPC computing resources at Lomonosov Moscow State University [3].The research is carried out using the equipment of the shared research facilities of HPC computing resources at Lomonosov Moscow State University supported by the project RFMEFI62117X0011.

References

1. Polyakov, I.V., Grigorenko, B.L., Mironov, V.A., Nemukhin, A.V.: Modeling structure and excitation of biliverdin-binding domains in infrared fluorescent proteins. Chemical Physics Letters (2018), DOI: 10.1016/j.cplett.2018.08.068

2. Chernov, K.G., Redchuk, T.A., Omelina, E.S., Verkhusha, V.V.: Near-infrared fluorescent proteins, biosensors, and optogenetic tools engineered from phytochromes. Chemical Reviews

117(9), 6423–6446 (2017), DOI: 10.1021/acs.chemrev.6b00700

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

4. Gordon, M.S., Schmidt, M.W.: Advances in electronic structure theory: GAMESS a decade later. In: Dykstra, C., Frenking, G., Kim, K., Scuseria, G. (eds.) Theory and Applications of Computational Chemistry: The First Forty Years, pp. 1167–1189. Elsevier, Amsterdam (2005), DOI: 10.1016/B978-044451719-7/50084-6

5. Valiev, M., Bylaska, E., Govind, N., Kowalski, K., Straatsma, T., Van Dam, H., Wang, D., Nieplocha, J., Apra, E., Windus, T., de Jong, W.: NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. Computer Physics Communications 181(9), 1477–1489 (2010), DOI: 10.1016/j.cpc.2010.04.018

6. Becke, A.D.: Density-functional exchange-energy approximation with correct asymptotic behavior. Physical Review A 38(6), 3098–3100 (1988), DOI: 10.1103/PhysRevA.38.3098

7. Schmidt, M.W., Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.H., Koseki, S., Matsunaga, N., Nguyen, K.A., Su, S., Windus, T.L., Dupuis, M., Montgomery, J.A.: General atomic and molecular electronic structure system. Journal of Computational Chemistry 14(11), 1347–1363 (1993), DOI: 10.1002/jcc.540141112