Available online at www.sciencedirect.com
ScienceDirect
Mendeleev Commun., 2017, **27**, 224–227

Mendeleev Communications

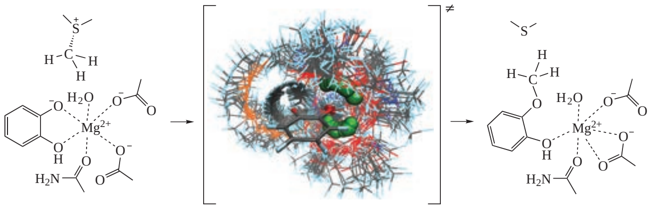
Exhaustive conformational search for transition states: the case of catechol *O*-methyltransferase active site

Michael G. Medvedev,*[a,b] Maria V. Panova,[b] Ghermes G. Chilov,*[b,c] Ivan S. Bushmarinov,[a] Fedor N. Novikov,[b,c] Oleg V. Stroganov,[b,c] Alexey A. Zeifman[b] and Igor V. Svitanko[b]

[a] *A. N. Nesmeyanov Institute of Organoelement Compounds, Russian Academy of Sciences, 119991 Moscow, Russian Federation. E-mail: medvedev.m.g@gmail.com*
[b] *N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation. E-mail: ghermes.chilov@gmail.com*
[c] *MolTech Ltd., 119992 Moscow, Russian Federation*

DOI: 10.1016/j.mencom.2017.05.002

A combination of the common quantum mechanics based transition state theory and exhaustive conformational search for the modeling of difficult reactions with hundreds of competing transition states is proposed. This approach is applied to study all transition state conformations of a reaction occurring in the catechol *O*-methyltransferase (COMT) active site in the absence of a major part of the enzyme, and the results are compared to the recent QM/MM modeling of this reaction within the enzyme. The main points of the method are (i) constraining of forming bonds upon conformer generation and (ii) preliminary constrained optimizations of located conformations to minima using a quantum mechanical method. Importantly, this methodology is applicable to the quantum mechanical part in QM/MM calculations and can reduce demand for large sampling in difficult cases.



The ensemble of possible transition states

In a most general sense, a major task of physical chemistry is to provide macroscopic observables from a sensible averaging of atomic-level events. Historically, the models for molecules considered in that way have been by necessity crude, and the conclusions were derived mostly about highly averaged physical properties of the test systems.[1]

However, in recent years, ever-increasing computational capabilities have allowed researchers to simultaneously study systems in terms of both average observables and finer details. This trend can be easily seen in materials science, where a combination of X-ray diffraction (XRD), neutron diffraction, extended X-ray absorption fine structure (EXAFS) and density functional theory (DFT) molecular dynamics allowed for determination of the structure and even vibrational properties of glassy $Ga_{11}Ge_{11}Te_{78}$.[2] The big-data statistical analysis of crystallographic synthons recently resulted in the discovery of the first example of supramolecular stereoelectronic effects, which leads to the inversion of a preferred carboxylic group conformation in cases of strong hydrogen bonding (*e.g.*, in an insulin hexamer).[3] In another recent study, the full computational assessment of possible catalyst conformations allowed researchers to construct a novel catalyst with a greatly improved viability.[4]

In theoretical chemistry, an ensemble view is often represented by molecular dynamics[5] (MD), which plays a crucial role in theoretical investigations in biochemistry, catalysis and other important areas of modern chemistry. Its power arises from its ability to describe versatile chemical systems in all their diversity. However, to be reliable, molecular dynamics should be long, *i.e.* account for as many states of the system as required for the convergence of macroscopic observables. From this fact arises the major setback of MD – its high computational cost: even using force fields, it requires supercomputers to study large biomolecules; *ab initio* molecular dynamics is even more expensive and rarely if ever has been applied to such systems. Classical molecular force fields are usually not useful when studying chemical reactions since they cannot describe the breakage and formation of bonds (with the exception of the ReaxFF,[6] which requires very rigorous parametrization toward the system under investigation to work well).

The alternative way, which allows one to study reactions adequately and without much time consumption, is the transition state theory (or its modifications such as the variational transition state theory[7] or Kramers' theory[8]) relying upon the postulate that reaction kinetics is determined by the energy of the transition state (TS, saddle point on the minimum gradient path connecting reactants and products) relative to the energy of reactants. While almost any reaction can proceed through different TSs belonging to the same mechanism (further, TS conformers), it is usually assumed that the reaction proceeds through the TS connected to the lowest reagent conformation. However, this is not true if the reagent conformations can interconvert rapidly[9] (at least ten times faster than the main reaction[10,11]). In that case, the reaction is governed by the Curtin–Hammett principle,[9] which states that it proceeds through the lowest transition states, which are not necessarily connected to the lowest reagent conformations (Figure 1). According to the Curtin–Hammett principle, the ratio of products is $[\mathbf{D}]/[\mathbf{C}] = \exp(-\Delta\Delta G^{\neq}/RT)$, notably, not depending on ΔG (Figure 1).

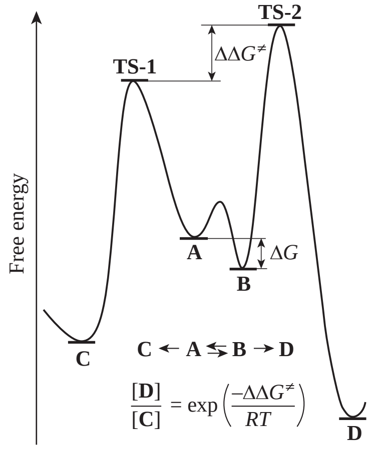


Figure 1 The Curtin–Hammett principle. If conformations **A** and **B** interconvert rapidly, the reaction preferably proceeds through **TS-1**.

As shown by the Curtin–Hammett principle, the accurate modeling of a reaction requires finding and assessing all transition state conformations. However, this is difficult because no currently available program package has this functionality. Meanwhile, any program package for conformational search, which supports geometry constraints on generated conformations, can be used for this task. Such packages range from freeware OpenBabel[12] and RDKit[13] (both require some scripting to apply constraints) to commercial Schrödinger MacroModel[14] (constraints work out-of-the-box). Note that, for small compounds, such as D-glucose, there is a scaled hypersphere search of the anharmonic downward distortion following (SHS-ADDF) method, which allows one to construct full route maps concerning all minima, transition states, and paths between them.[15]

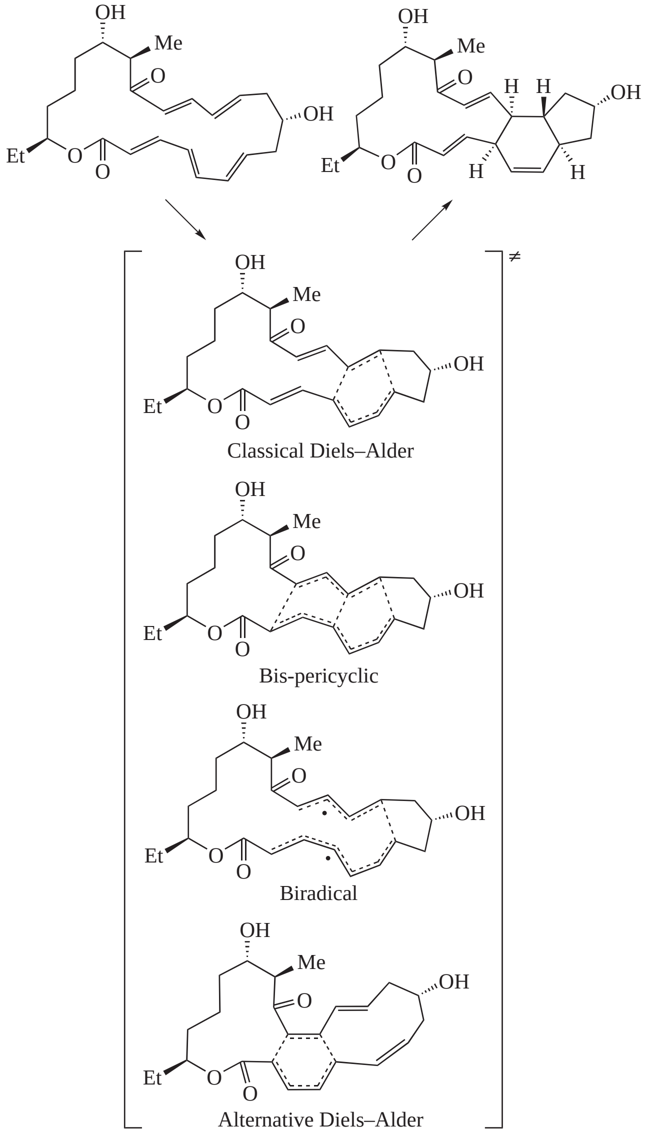
Recently,[16] we have studied all possible transition state conformations for all routes proposed for the SpnF-catalyzed formal Diels–Alder cycloaddition (Scheme 1). To find all Diels–Alder transition state conformations, we have (1) generated all conformations of the product, (2) stretched the two forming bonds closer to their transition-state lengths, (3) optimized structures to the closest minima keeping the two forming bonds constrained and (4) performed unconstrained optimizations toward the closest transition states. The Curtin–Hammett principle applied to the found transition state conformations allowed us to prove that this reaction mostly proceeds through bis-pericyclic transition states.

However, the Diels–Alder reaction is a simple example since its product is isobonded with its transition state (henceforth, isobonded means having the same connectivity graph including the forming bonds), which allowed us to start with the generation of all product conformations. In cases of reactions where the transition state is not isobonded with substrate or product, such an approach is not applicable and one need to search the transition state conformers explicitly. We propose the following algorithm for such cases:

1. Construct a trial transition state guess and locate corresponding TS with a quantum mechanical (QM) method. Trial TS guesses can be constructed automatically using the heuristic-guided method proposed by Reiher and coworkers.[17]

2. Constrain all forming or changing bonds to values found in the trial TS at the previous step and perform a constrained search for all conformers using either (MM- or QM-based) molecular dynamics or torsion sampling with constrained post-optimization or, preferably, a combination of the above.

3. Filter duplicates and TS conformers with too high energies. Note that, if a standard force field was used at the previous step, the chosen cut-off should be sufficiently high to account for possible huge errors on TS structures.



Scheme 1 SpnF-catalyzed reaction, where the exhaustive conformational search allowed us to prove that it mostly proceeds through bis-pericyclic transition states.[16]

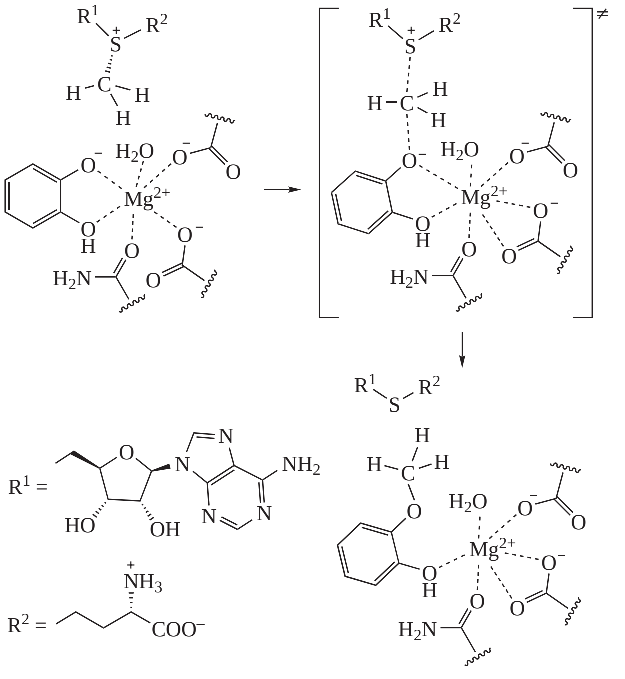
4. Keeping all constraints, optimize each of the remained structures to the closest minimum by a QM method; remove duplicates.

5. For each remained structure, perform an unconstrained optimization to the transition state (or the variational transition state,[7] if they do not coincide) using a QM method; remove duplicates.

6. If necessary, account for the curvature of potential energy surface by means of zero-point and thermal vibrational corrections.

Here, we will apply this methodology to the S_N2 reaction, which takes place in the active site of the catechol *O*-methyltransferase (COMT). COMT is a methyltransferase that regulates neurotransmitters through methylation. It involves the *S*-adenosyl-L-methionine (SAM) and Mg^{2+} in its active site. The catalyzed reaction is shown in Scheme 2.

This reaction in the active site was intensively studied by both molecular dynamics[18] and QM/MM[19] approaches. We have modeled it in the absence of a major part of the enzyme, replacing R^1, R^2 and all attachment points in Scheme 2 with methyl groups. Lack of the enzyme makes the ensemble of possible transition states very diverse and allows us to find TSs, which are geometrically prohibited inside the enzyme.



Scheme 2 Reaction taking place in the active site of the COMT enzyme. Attachment points on the scheme belong to the enzyme.

At the first step, we have generated all possible transition state conformations. For this reason, a trial guess geometry was constructed (Figure S1), and a conformational search was performed using MacroModel[14] with all Mg^{2+}–ligand distances shown in Figure S1 as bonds constrained to 2.0 Å, forming O–C and C–S bonds constrained to 2.02 and 1.95 Å, respectively, and the angle between forming bonds constrained to 180°. An OPLS-aa[20] force field was used to optimize generated conformations preserving constraints, only conformations with energies within 30 kcal mol^{-1} from the lowest one were retained. This resulted in 205 diverse conformations (RMSD > 0.3 Å, excluding hydrogens), which are shown in Figures 2, S2 and S3.

According to Figure 2, generated conformations include both mono- and bidentate coordinations of catechol to Mg^{2+} and the transferring methyl group is always disposed appropriately for the title S_N2 reaction.

QM calculations were performed using Gaussian09[21] implementation of PM6[22] semi-empirical method, which was used for QM part of QM/MM calculations of COMT by Jindal and Warshel.[19] To obtain more reliable results, one should use theoretically grounded DFT methods,[23,24] but the issue here is the proof of the proposed concept, for which accuracy of semi-empirical PM6 method should be sufficient. Polarizable surrounding was modeled with SMD(EtOH)[25] to mimic an averaged enzyme active site, whose dielectric constants are often close to that of ethanol.[26,27]

Since geometries are frequently significantly different between QM and MM methods, an unconstrained optimization toward TS starting from the located MM geometry is likely to lead to wrong TS, which is unacceptable given the number of conformers at hand. To overcome this, initial constrained geometry optimizations to minima were performed using PM6 and SMD(EtOH) keeping the forming bond lengths fixed. After that, unconstrained optimizations toward the closest transition states were performed at the same level of theory. Harmonic frequencies were computed to ensure that located stationary points are proper transition states. A total of 162 transition states were located; their superpositions are shown in Figures 3, S4 and S5. The other 43 cases have not converged to transition states of the requested type. Duplicates removal (by RMSD > 0.2 Å, excluding hydrogens) left 141 unique transition states. Cartesian coordinates of all located transition states are available in Online Supplementary Materials.

As can be seen from a comparison of Figures 2 and 3, while force field predicts a random position of the methyl group toward catecholate, the PM6 method refines it to a ring with the $C_{catechol}$–$O^{-}\cdots C_{Me}$ angle in a range of 116–127° with a median at 122°. The $S_{SAM}\cdots C_{Me}\cdots O^{-}$ angle is in a range of 172–179° with a median at 175°. The distance between the catechol oxygen atom and the SAM sulfur atom is remarkably constant within a range of 3.995–4.035 Å in very good agreement with the results of Jindal and Warshel.[19]

All located conformations have close energies, not more than 8 kcal mol^{-1} from the lowest one. The histogram of relative energies is given in Figure S6; their list is available as Table S1. Conformations close to those at the active site of the real enzyme[19] are 85, 111, 113, 117 and 129 (Figure S7); they have low relative energies ranging from 2.1 to 3.5 kcal mol^{-1}. However, interestingly, the lowest energy TS conformations are significantly different from that in the active site (Figure S8). Note that substrates leading to the located TSs cannot interconvert rapidly, so a fair description of the modeled reaction should be based on the generalized equations[10] (not Curtin–Hammett ones) and include energies of TSs corresponding to substrate interconversions. They can be located using the described approach, as well.

Thus, we propose an algorithm for investigating difficult reactions by means of transition state conformations interplay in cases where *ab initio* molecular dynamics is too expensive to be applied. The important stages are (1) the generation of TS guesses by constrained conformational search approaches, (2) the constrained optimization of generated TS guesses to minima using a

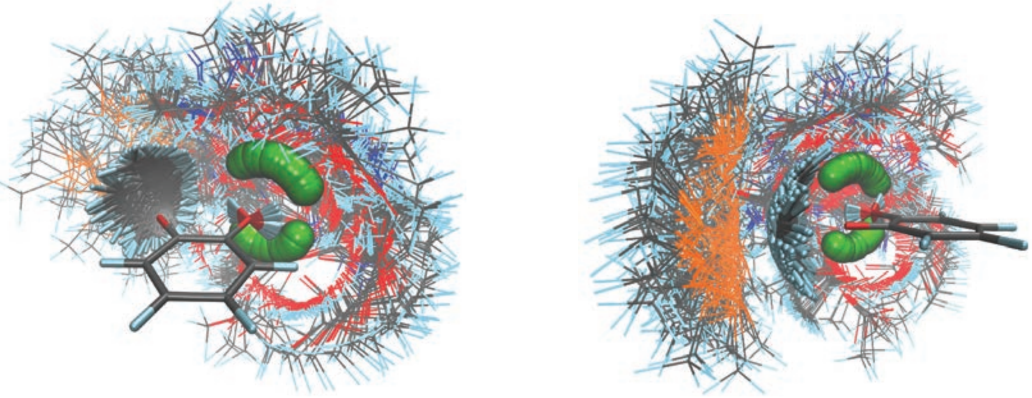


Figure 2 The two views of all TS guess conformations generated by our search, aligned by catechol carbon and oxygen atoms. Catechols and transferring methyl groups are shown as capped sticks; Mg^{2+} ions are shown as spheres, and all other molecules are shown as lines. Coloring: magnesium – green, sulfur – orange, oxygen – red, nitrogen – blue, carbon – grey, hydrogen – light cyan. Full-width images are available in Online Supplementary Materials as Figures S2 and S3.

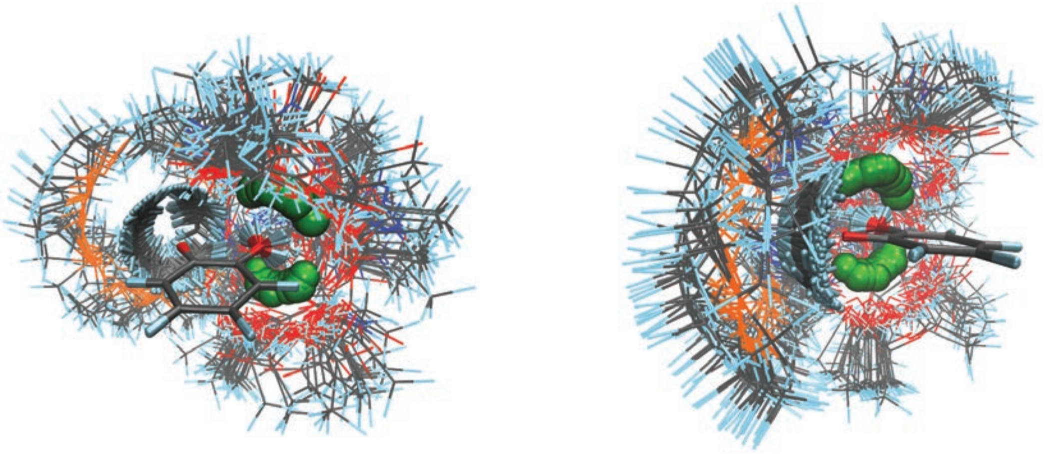


Figure 3 The two views of all TS conformations located at the PM6 SMD(EtOH) level of theory, aligned by catechol carbon and oxygen atoms. Catechols and transferring methyl groups are shown as capped sticks, Mg^{2+} ions are shown as spheres and all other molecules are shown as lines. Coloring: magnesium – green, sulfur – orange, oxygen – red, nitrogen – blue, carbon – grey, and hydrogen – light cyan. Full-width images are available in Online Supplementary Materials as Figures S4 and S5.

QM method and (3) the unconstrained optimization of resulted structures toward the closest TSs using the same or a different QM method. This approach can also be used together with the QM/MM methodology for the simulations of enzymes. Clearly, our approach cannot account for non-Boltzmann coherent fluctuations, but their role in enzymatic catalysis and surface reactions is shown to be negligible in most cases.[28–30] On the other hand, our approach should allow one to apply more accurate *ab initio* methods to bigger QM (sub)systems and to account better for electrostatic effects, which are known to play a crucial role in enzymatic catalysis.[31]

M.G.M., F.N.N., O.V.S., G.G.C., A.A.Z. and I.V.S. acknowledge the support of the Russian Science Foundation (grant no. 17-13-01526); I.S.B. acknowledges the support of the Russian Foundation for Basic Research (grant no. 17-03-00907); and M.G.M. acknowledges the computational resources provided by the Department of Computational Mathematics and Cybernetics of M. V. Lomonosov Moscow State University: IBM Blue Gene/P supercomputer; by the Supercomputing Center of M. V. Lomonosov Moscow State University: the Lomonosov[32] supercomputer; and by the Federal Center for Collective Usage at NRC ‘Kurchatov Institute’, http://computing.kiae.ru/.

Online Supplementary Materials

Supplementary data associated with this article (figures, table of TSs relative energies and an XYZ file with Cartesian coordinates of all located TSs) can be found in the online version at doi: 10.1016/j.mencom.2017.05.002.

References

1 F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York, 1965.

2 I. Voleská, J. Akola, P. Jóvári, J. Gutwirth, T. Wágner, Th. Vasileiadis, S. N. Yannopoulos and R. O. Jones, *Phys. Rev. B*, 2012, **86**, 094108.

3 M. G. Medvedev, I. S. Bushmarinov and K. A. Lyssenko, *Chem. Commun.*, 2016, **52**, 6593.

4 V. V. Gerasimchuk, A. S. Kucherenko, A. N. Fakhrutdinov, M. G. Medvedev, Y. V. Nelyubina and S. G. Zlotin, *Eur. J. Org. Chem.*, 2017, DOI: 10.1002/ejoc.201700166.

5 S. A. Harris and V. M. Kendon, *Phil. Trans. R. Soc. A*, 2010, **368**, 3581.

6 A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard III, *J. Phys. Chem. A*, 2001, **105**, 9396.

7 D. G. Truhlar and B. C. Garrett, *Annu. Rev. Phys. Chem.*, 1984, **35**, 159.

8 *New Trends in Kramers’ Reaction Rate Theory*, eds. P. Talkner and P. Hänggi, Springer, Dordrecht, 1995.

9 J. I. Seeman, *J. Chem. Educ.*, 1986, **63**, 42.

10 N. S. Zefirov, *Tetrahedron*, 1977, **33**, 2719.

11 J. I. Seeman and W. A. Farone, *J. Org. Chem.*, 1978, **43**, 1854.

12 N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminform.*, 2011, **3**, 33.

13 *RDKit: Open-source cheminformatics*, http://www.rdkit.org/.

14 *Schrödinger: Maestro*, Schrödinger, LLC, New York, NY, 2016.

15 H. Satoh, T. Oda, K. Nakakoji, T. Uno, H. Tanaka, S. Iwata and K. Ohno, *J. Chem. Theory Comput*., 2016, **12**, 5293.

16 M. G. Medvedev, A. A. Zeifman, F. N. Novikov, I. S. Bushmarinov, O. V. Stroganov, I. Yu. Titov, G. G. Chilov and I. V. Svitanko, *J. Am. Chem. Soc.*, 2017, **139**, 3942.

17 M. Bergeler, G. N. Simm, J. Proppe and M. Reiher, *J. Chem. Theory Comput.*, 2015, **11**, 5712.

18 N. Patra, E. I. Ioannidis and H. J. Kulik, *PLoS ONE*, 2016, **11**, e0161868.

19 G. Jindal and A. Warshel, *J. Phys. Chem. B*, 2016, **120**, 9913.

20 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225.

21 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 09, Revision D.01*, Gaussian, Inc., Wallingford, CT, 2009.

22 J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173.

23 M. G. Medvedev, I. S. Bushmarinov, J. Sun, J. P. Perdew and K. A. Lyssenko, *Science*, 2017, **355**, 49.

24 T. Vasilevskaya, M. G. Khrenova, A. V. Nemukhin and W. Thiel, *Mendeleev Commun*., 2016, **26**, 209.

25 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378.

26 P. E. Smith, R. M. Brunne, A. E. Mark and W. F. van Gunsteren, *J. Phys. Chem.*, 1993, **97**, 2009.

27 T. Simonson and D. Perahia, *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 1082.

28 A. Warshel and R. P. Bora, *J. Chem. Phys.*, 2016, **144**, 180901.

29 M. H. M. Olsson, J. Mavri and A. Warshel, *Phil. Trans. R. Soc. B*, 2006, **361**, 1417.

30 K. Song, P. de Sainte Claire, W. L. Hase and K. C. Hass, *Phys. Rev. B*, 1995, **52**, 2949.

31 A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu and M. H. M. Olsson, *Chem. Rev.*, 2006, **106**, 3210.

32 V. Sadovnichy, A. Tikhonravov, Vl. Voevodin and V. Opanasenko, in *Contemporary High Performance Computing: From Petascale toward Exascale*, ed. J. S. Vetter, CRC Press, Boca Raton, 2013, pp. 283–307.

Received: 15th December 2016; Com. 16/5110