Budapest University of Technology and Economics

Faculty of Chemical Technology and Biotechnology

George A. Oláh Doctoral School of Chemistry and Chemical Technology

Development of Multilevel Methods Using Frozen Localized Orbitals

Thesis Statements

Written by: Bence Hégely

Advisor: Prof. Mihály Kállay

Department of Physical Chemistry and Materials Science

MTA-BME Lendület Quantum Chemistry Research Group

2019

1 Introduction

Computational chemistry grants theoretical explanations for experiments to facilitate the understanding of nature and to predict phenomena. However, the power of these forecasts are heavily influenced by the basics of the applied models. Quantum chemistry, a discipline of theoretical chemistry, offers the most accurate predictions but its application is limited because of the complicated form of the wave function. Wave function theory (WFT), a specific field of quantum chemistry, introduces a series of known approximations to the wave function from the beginning, or as more commonly said "*ab initio*", and its high-level variants permit simulations for dozens of atoms within 1 kcal/mol accuracy, the typical accuracy of chemical experiments (shortly referred to as chemical accuracy). The predicting power of *ab initio* approaches is marvelous, however, their range of application is quite sobering for most of the chemists, who are mostly interested in condensed-phase reactions and properties. The fact that the simulation time of these approaches scales as a high-power of the system size limits the use of accurate *ab initio* techniques even with growing computational capacities.

Numerous efforts have been made over the years to expand the application range by sacrificing the accuracy to some extent in exchange for faster calculations. For example, the Kohn–Sham (KS) model of density functional theory (DFT) and the combination of quantum chemistry with molecular mechanics (MM) meant groundbreaking developments, which allowed for the simulation of large systems, even nanoparticles. However, the price of the trade off is often in question. Despite the progress of the traditional and composite methods, the development of better approaches is still a must because the progressing fields, for instance, computer-aided materials design, command more accurate and efficient models. Therefore, our aim was to contribute to the field of quantum chemistry by improving existing techniques and developing new ones. The dissertation was written in order to summarize our efforts.

2 Theoretical background

The accurate description of large systems is difficult for traditional quantum chemistry methods because the computational expenses steeply grow with the size of the studied systems (N). The computation time of the Hartree–Fock (HF) and KS DFT calculations using a hybrid functional scales formally as the fourth-power of the systems size. The more advanced *ab initio* correlation methods, such as the second-order Møller–Plesset (MP2) approach or the coupled-cluster (CC) models, e.g., CC with single and double excitations with perturbative triples [CCSD(T)], exhibit a scaling of at least N^5.

One way to decrease the computational cost of the traditional techniques is to split up the large system into smaller subsystems or fragments which are more or less equivalent in size. The conventional methods are only utilized in the fragments, while the interactions among the fragments are described in an approximate manner or completely omitted (see Fig. 1 for illustration). With this idea, the computational time of the costly computational steps can be decreased, and the scaling can be reduced to linear.

Other approaches exploit the fact that many molecular processes exhibit local electronic structure changes, hence it is a logical to partition the system into two (or more) subsystems with different sizes: the active site (A) and its environment (E). Due to the large difference in their size and chemical relevance, it is appealing to apply different methods to the subsystems: the active site is the most relevant part, hence it should be handled by highly accurate but costly techniques, while the environment can be described by a more approximate and computationally cheaper method. Theoretical frameworks exploiting this partitioning strategy are called focused models.

Some focused models are called multilevel approach because different levels of theory are used on different regions of the system, but there is no coupling among the methods. In the case of coupled methods, e.g., the electron density of the active subsystem is polarized by its environment, the approach is considered as an embedding

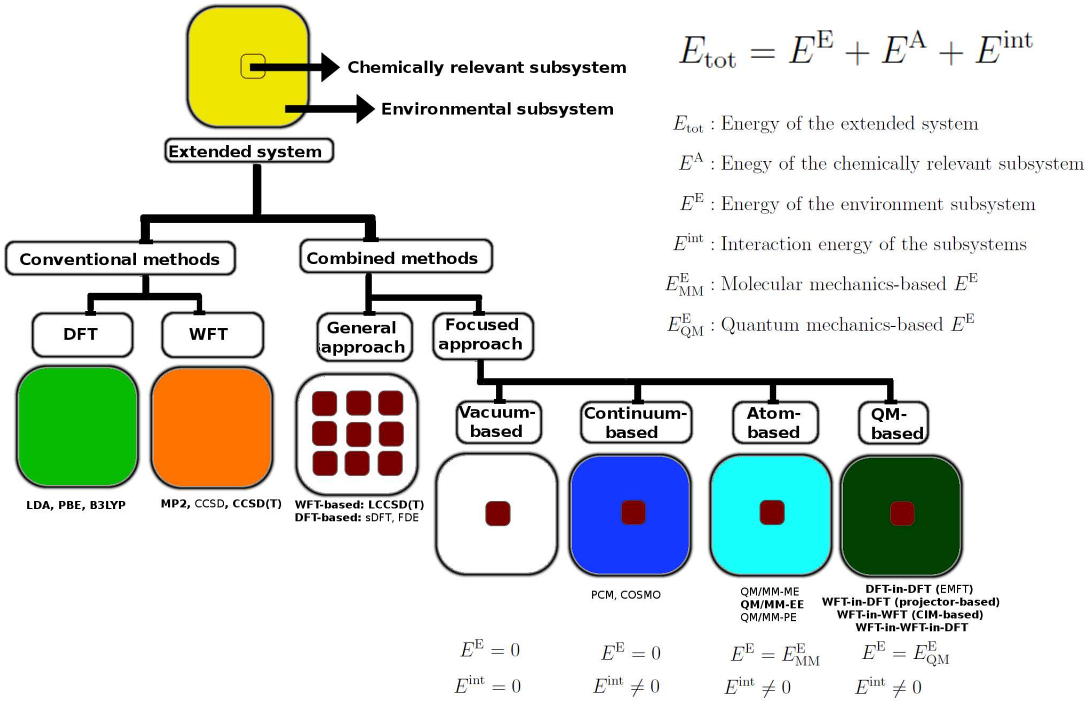

Figure 1: Illustration of the methods that are considered in the thesis. The different techniques are denoted with different colors, whereas the (relatively) high-level methods are highlighted by red. The used and mentioned schemes are set as bold and normal, respectively.

method. The focused models can be further grouped, regardless of their multilevel or embedding character, based on the approximation level of the environment and the interaction of subsystems. The classification is illustrated in Fig. 1. In general, the overall computational cost increases with the approximation level of the environment, but the accuracy does not improve automatically. Nevertheless, one can at least expect that more accurate results can be obtained as the active subsystem enlarges. In the dissertation, the atom-based and quantum-based strategies of the focused models were improved.

The fundamentals of the atom-based strategy stands on MM, which is also called the force field model of molecules. The MM method is based on classical mechanics, electrostatics, and on the empirical inclusion of van der Waals (vdW) interactions. It uses atoms as building blocks, which are bounded by springs and have point-like mass and electric charge. For a practical MM calculation, atom types have to be assigned to each atom, and parameters necessary for the evaluation of the MM energy functional have to be determined. Obviously, the application of MM is problematic when the electronic structure changes, because the predefined atom types are no longer valid. The combined quantum mechanics/molecular mechanics (QM/MM) method offers a solution to this problem by using the MM and quantum chemistry methods on the environment and the active site, respectively. Out of the three QM/MM schemes, we only considered the most widely used electronic embedding (EE) in the dissertation, which allows the electron density of the active system to be polarized by the MM point charges. Despite the success of EE, it still has important issues regarding the coupling of the QM and MM techniques, especially when the system is separated across covalent bonds. In this case, dangling bonds emerge in the QM region, which are commonly handled with capping atoms, known as link atoms. However, this heuristic approach has its limitations, for example, it handles polarized bonds rather poorly. A theoretically more founded approach is the local self-consistent field (LSCF) method, which uses "frozen" localized orbitals in the subsystem boundary, although its computational complexity prevents the development and the extensive use of the approach. Recently, a simpler LSCF method was proposed by Ferenczy [1] which uses the Huzinaga-equation [2]. In the dissertation, the implementation of the Huzinaga-equation-based LSCF (HLSCF) approach is presented with an intent to improve the QM/MM methodology.

The quantum-based strategies face different coupling problems compared with the QM/MM schemes. The linear-scaling subsystem DFT (sDFT) approach, which is a fragmentation-based reformulation of the conventional KS DFT method, is a typical example for the issues regarding quantum-based strategies. In sDFT, the KS equations are solved for each subsystem separately in the embedding potential of other subsystems. With this approach, however, the obtained subsystem electron densities are not

[1] G. G. Ferenczy, *J. Comput. Chem.*, **2013**, 34:862

[2] S. Huzinaga, A. A. Cantu, *J. Chem. Phys.*, **1971**, 55:5543

additive, hence a nonadditive kinetic energy term arises which has to be approximated. Manby *et al.* [3] elaborated a simple, exact, and projector-based embedding scheme which circumvents the above problem by enforcing orbital orthogonality. The approach allows for the embedding of DFT and WFT methods into lower-level DFT approaches utilizing orbital localization. We contributed to this technique by proposing simple modifications to the projector-based embedding scheme. The field of WFT-in-WFT methods were also investigated, where several linear-scaling WFT techniques were extended in the focused models sense.

Combining different levels of theory is not the only way to accelerate the calculations. Since the computation time for any electronic structure calculation strongly depends on the size of the atomic orbital (AO) basis set applied, the most plausible strategy to decrease the computational costs is the reduction of the size of the AO basis. In the case of extended systems with local chemical changes, a trivial way to reduce the AO basis is the use of the mixed-basis set approach. In this simplistic approximation the atoms of the chemically relevant part of the system are described using a large AO basis set, whereas a smaller basis set is employed for the remaining atoms. Although this approach is conceptually simple and trivial to implement, the use of basis sets of different sizes gives rise to the artificial polarization of the electronic structure. In addition, the use of mixed-basis sets might cause convergence problems for SCF algorithms and also increases the probability that the SCF procedure converges to an unstable solution. The so-called dual basis (DB) set approach is a good alternative, which is free from the aforementioned problems. The idea of the DB approach dates back to 1980s, but recent progress is made in this field by Head-Gordon *et al.* [4] In their DB SCF approach an SCF calculation is converged in a small basis set, the density matrix is projected into the large basis set, a single Roothaan step is taken, and an energy correction is evaluated. A similar scheme is proposed by us, which also allows the use of the DB scheme in embedding calculations.

[3] F. R. Manby, M. Stella, J. D. Goodpaster, T. F. Miller, *J. Chem. Theory Comput.* **2012**, 8:2564

[4] Z. Liang, M. Head-Gordon, *J. Phys. Chem. A*, **2004**, 108:3206

3 Technical details

The performance of the different QM/MM boundary schemes, namely the HLSCF and the link atom techniques, was investigated by calculating reaction energies and potential energy surfaces using HF/6-31G* for the QM subsystem. Test systems were chosen in order to represent typical QM/MM applications: the deprotonation of hexanoic acid, the rotation energy profile of histidine, and the potential energy surface of a proton transport process between aspartic acid and histidine were computed. The results were compared with QM-only calculations, while the effect of QM subsystem enlargement was also explored.

Test calculation were carried out for the DFT-in-DFT, WFT-in-DFT, and WFT-in-WFT methods by computing reaction energies of the following systems: the substitution reaction 1-chlorodecane to yield 1-decanol, the deprotonation of decanoic acid, the Diels–Alder reaction between the conjugated octadecanonaene and 1,3-butadiene, and the hydrogenation of pentacene. These systems were designed to model different environments of the subsystem separations. The hybrid DFT functionals and our local natural orbital-based CCSD(T) approach [LNO-CCSD(T)] were utilized as high-level techniques, while various DFT functionals and our local MP2 (LMP2) were used as low-level methods. Several double- and triple-ζ basis sets were employed. The active subsystem was enlarged systematically to investigate the effect of active subsystem size on the accuracy of the combined methods. The results of the combined approaches were compared to vacuum embedding, QM/MM, and calculations utilizing high-level methods only. Most of these test settings were also adopted for the benchmarking of the dual-basis approach. In the case of test calculations with multiple basis sets, the mixed-basis scheme was also utilized. Finally, the best performing techniques were tested on a catalyzed reaction involving the catechol-*O*-methyltransferase enzyme.

4 Results

For the first time, we presented the implementation of the Huzinaga-equation-based LSCF approach in the Amber and Mrcc codes, which makes it possible to apply WFT and DFT methods without orthogonalizing the basis set to the frozen orbitals separating the QM and MM subsystems. A significant novelty of the program is the automatic generation of the frozen localized orbitals obtained from calculations performed for model molecules cut out of the system. The Amber-Mrcc code also allows the use of the link atom approach, thus the benchmarking of the different couplings is now possible. Test calculations utilizing the electronic embedding scheme were performed to check the performance of both the HLSCF and the link atom approaches. It was found that the conceptually appealing HLSCF method is a competitive alternative to the link atom method, however, further developments are necessary for practical applications.

To enhance the efficiency of the QM calculations, we contributed to the development of quantum-based multilevel approaches. The nearly exact, projector-based embedding scheme of Manby *et al.*, which allows for the embedding of DFT and WFT methods into lower-level DFT approaches, was improved. Instead of augmenting the Fock-matrix with the somewhat arbitrary level-shift operator, we proposed the use of the Huzinaga-equation, which strictly enforces the Pauli exclusion principle. Our local correlation method was also extended in the spirit of the focused models: since its kernel splits up the system into local domains, simple multilevel theories can be defined if the domains of the active subsystem and the environment are treated at a different level. Several orbital selection algorithms were also designed to enable the use of the aforementioned techniques as black-box approaches. The considered projector-based Huzinaga-embedding and WFT-in-WFT schemes were benchmarked on simplistic, quasi-one dimensional test systems for reaction energies and compared with QM/MM and vacuum-embedding. We concluded that, in the case of DFT-in-DFT, the Huzinaga-embedding scheme is more efficient than the other approaches, but QM/MM or even simple vacuum embedding is still competitive in particular tests. Concerning the embedding of WFT methods, the clear winner is the LNO-CCSD(T)-in-LMP2 approach, and the Huzinaga-embedding can only be more advantageous if a non-hybrid density functional is employed.

Based on their success, the embedding and multilevel schemes were further developed to increase their efficiency. A DB approach, similar to the DB approximation of Head-Gordon, was designed for embedding calculations with the intent to reduce the computational expenses of the the DFT- and WFT-based methods. The new approach was applied to the Huzinaga-embedding, our WFT-in-WFT method, the combined Huzinaga multilevel local correlation approximations, and it was also used in an application. The performance of the resulting DB DFT and WFT embedding methods was evaluated in extensive benchmark calculations on the aforementioned test set and also compared to that of the corresponding embedding schemes exploiting the mixed-basis approximation. Our results show that, with an appropriate combination of basis sets, the DB approach significantly speeds up the embedding calculations, and, for chemical processes where the electronic structure considerably changes, it is clearly superior to the mixed-basis approximation. The results also demonstrate that the DB approach, if integrated with the mixed-basis approximation, efficiently eliminates the major weakness of the latter, and the combination of the DB and mixed-basis schemes is the most efficient strategy to accelerate embedding calculations. The above findings were also confirmed by the test calculations on a realistic three-dimensional system with biological importance.

5 Applicability

The presented approaches are available in the Amber and Mrcc programs, and they are already useful for the efficient validation of DFT-based QM/MM results and the highly accurate estimation of various properties, such as reaction energies and reaction barrier heights. The implementation of further molecular properties is in progress to widen the applicability of the methods.

6 New scientific results

I The Huzinaga-equation-based local SCF approach using automatically generated frozen localized orbitals has been implemented. It was shown that the theoretically more established method is not yet competitive with the widely used link atom scheme (1).

II The use of the Huzinaga-equation improves the projector-based embedding by making it strictly exact. The projector-based WFT-in-DFT can only be competitive with WFT-in-WFT if a non-hybrid density functional is employed (2).

III A new dual basis scheme has been proposed based on the work of Head-Gordon *et al.* The new scheme is suitable for embedding calculations (3) because the density matrix in the large basis is N-representable. For the first time, the new technique was also used in an application where antioxidant complexes were identified with density functional theory.

IV The combination of the dual basis and mixed basis schemes, the dual-mixed basis approach is the most efficient for embedding and multilevel techniques. The above findings have also been confirmed on the realistic three-dimensional system of the catechol-*O*-methyltransferase enzyme, which has biological importance (3).

7 Publications and presentations

7.1 Publications related to the thesis

1 A QM/MM program using frozen localized orbitals and the Huzinaga equation, B. Hégely, F. Bogár, G. G. Ferenczy, and M. Kállay,
Theor. Chem. Acc., **2015**, 134:132
IF: 1.806, citations: 3

2 Exact density functional and wave function embedding schemes based on orbital localization, B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay,
J. Chem. Phys., **2016**, 145:064107
IF: 2.965, citations: 18

3 Dual basis set approach for density functional and wave function embedding schemes, B. Hégely, P. R. Nagy, and M. Kállay,
J. Chem. Theory Comput., **2018**, 14:4600
IF: 5.399 (2017), citations: 2

4 Melt stabilization of PE with natural antioxidants: Comparison of rutin and quercetin, B. Kirschweng, D. M. Tilinger, B. Hégely, Gy. Samu, D. Tátraaljai, E. Földes, and B. Pukánszky,
Eur. Polym. J., **2018**, 103:228
IF: 3.741, citations: 2

7.2 Other publications

- Solvation and protonation of coumarin 102 in aqueous media: A fluorescence spectroscopic and theoretical study, D. Hessz, B. Hégely, M. Kállay, T. Vidóczy, and M. Kubinyi,
  J. Phys. Chem. A, **2014**, 118:5238
  IF: 2.693, citations: 7

- Melt stabilization of polyethylene with dihydromyricetin, a natural antioxidant, B. Kirschweng, K. Bencze, M. Sárközi, B. Hégely, Gy. Samu, J. Hári, D. Tátraaljai, E. Földes, M. Kállay, and B. Pukánszky,
  Polym. Degrad. Stabil., **2016**, 133:192
  IF: 3.386, citations: 4

7.3 Presentations

- Poster presentation: **The 14th Central European Symposium on Theoretical Chemistry**, 2015, Banská Bystrica, Slovakia
  A QM/MM program using frozen localized orbitals and the Huzinaga equation, B. Hégely, F. Bogár, G. G. Ferenczy, and M. Kállay

- Poster presentation: **The 9th Congress of the International Society for Theoretical Chemical Physics**, 2016, Grand Forks, USA
  Exact density functional and wave function embedding schemes based on orbital localization, B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay

- Oral presentation: **The 14th Conference of the George Oláh Doctoral School**, 2017, Budapest, Hungary
  Egzakt, lokális pályákon alapuló sűrűségfunkcionál és hullámfüggvény beágyazási sémák, B. Hégely, P. R. Nagy, G. G. Ferenczy, and M. Kállay

- Poster presentation: **The 16th International Congress of Quantum Chemistry**, 2018, Menton, France
  Dual basis set approach for density functional and wave function embedding schemes, B. Hégely, P. R. Nagy, and M. Kállay
  Wiley "Computers in Chemistry" Poster award