Final report for NKFIH project KKP126451
“Accurate properties for large molecules”

Introduction

The major goal of this project was to develop quantum chemical methods for the calculation of accurate properties of large molecular systems. The problem was approached from two different directions. On one hand, we developed reduced-cost and local correlation methods, in particular, local coupled-cluster (CC) and second-order Møller–Plesset (MP2) approaches. On the other hand, we also focused on the improvement density functional theory (DFT) methods and combined wave function-DFT approaches such as double-hybrid (DH) functionals. The wave function and DFT methods were also combined utilizing embedding approximations. As the self-consistent field (SCF) calculation became the rate-limiting step in local correlation calculations using our improved methods, we also paid attention to the acceleration of the SCF step. Finally, in collaboration with theoretical and experimental research groups, the methods developed were applied to practical problems in various fields of chemistry.

Local correlation approaches

An optimized implementation of our local natural orbital (LNO) CC singles, doubles, and perturbative triples [CCSD(T)] method [LNO-CCSD(T)] was presented [1]. The integral-direct, in-core, highly efficient domain construction technique of our local MP2 (LMP2) scheme was extended to the CC level. The resulting scheme inherits the beneficial properties of the LMP2 approach, such as the asymptotically linear-scaling operation count, the asymptotically constant data storage requirement, and the completely independent domain calculations. In addition to integrating our recent redundancy-free LMP2 and Laplace-transformed (T) algorithms with the LNO-CCSD(T) code, the memory demand, the domain and LNO construction, the auxiliary basis compression, and the previously rate-determining two-external integral transformation were significantly improved. Our benchmark calculations showed that the optimized method provides average correlation and reaction energy errors of less than 0.07 % and 0.34 kcal/mol, respectively, compared to the canonical CCSD(T) reference. The efficiency of the present LNO-CCSD(T) implementation was demonstrated on realistic, three-dimensional examples. Using the new code an LNO-CCSD(T) correlation energy calculation is feasible on a single processor for a protein molecule including 2380 atoms and more than 44000 atomic orbitals. Our results were highlighted on the cover page of issue 14/8 of the Journal of Chemical Theory and Computation.

Our linear-scaling LNO-CCSD(T) method was further optimized, and extensive benchmark applications were presented [2]. A composite threshold combination hierarchy (Loose, Normal, Tight, etc.) was introduced, which enables black box convergence tests and is useful to estimate the accuracy of the LNO-CCSD(T) energies with respect to CCSD(T). We also demonstrated that the complete basis set limit (CBS) of LNO-CCSD(T) energies can be reliably approached via basis set extrapolation using large basis sets including diffuse functions. Where reference CCSD(T) results are available, the mean absolute errors of LNO-CCSD(T) reaction and intermolecular interaction energies with the default Normal threshold combination are below 0.2-0.3 kcal/mol, while the same measures with the Tight setting are 0.1 kcal/mol for all the tested systems including highly-complicated cases. The performance of LNO-CCSD(T) was also compared with that of other popular local CCSD(T) schemes. The exceptionally low hardware requirements of the present scheme enable the routine calculation of benchmark-quality energy differences within chemical accuracy of CCSD(T)/CBS for systems including a few hundred atoms. LNO-CCSD(T)/CBS calculations can also be performed for more than 1000 atoms with 45000 atomic orbitals using a single, 6-core CPU, about 100 GB memory, and comparable disk space.

Our linear-scaling local MP2 approach and our LNO-CCSD(T) method were further optimized, and extensive benchmark applications were presented. A linear scaling LMP2 method was presented for high-spin open-shell molecules based on restricted open-shell (RO) reference functions [3]. The new LMP2 approach inherits the iteration- and redundancy-free formulation and the completely integral-direct, OpenMP-parallel, and memory- and disk use economic algorithms of our closed-shell LMP2 implementation. By utilizing restricted local molecular orbitals for the demanding integral transformation step and by introducing a novel long-range spin polarization approximation, the computational cost of RO-LMP2 approaches that of closed-shell LMP2. Extensive benchmarks were performed for reactions of radicals, ionization potentials, as well as spin-state splittings of carbenes and transition-metal complexes. Compared to the conventional MP2 reference for systems of up to 175 atoms, local errors of at most 0.1 kcal/mol were found, which are well below the intrinsic accuracy of MP2. RO-LMP2 computations were presented for challenging protein models of up to 601 atoms and 11 000 basis functions which involve either spin-states of a complexed iron ion or a highly delocalized singly occupied orbital. These are the largest correlated computations ever presented for open-shell systems extending the reach of open-shell MP2, as well as its application in the contexts of spin-scaled MP2 and DH DFT. The corresponding runtimes of 9-15 hours obtained with a single, many-core CPU demonstrate that MP2, as well as spin-scaled MP2 and DH density functional computations become widely accessible for open-shell systems of unprecedented size and complexity.

The extension of the highly-optimized LNO-CCSD(T) method was presented for high-spin open-shell molecules based on RO references [4]. The techniques enabling the outstanding efficiency of the closed-shell LNO-CCSD(T) variant were adopted, including the iteration- and redundancy-free MP2 and (T) formulations, as well as the integral-direct, memory- and disk use economic, and OpenMP-parallel algorithms. For large molecules, the efficiency of our open-shell LNO-CCSD(T) method approaches that of its closed-shell parent method due to the application of restricted orbital sets for demanding integral transformations and a novel approximation for higher-order long-range spin-polarization effects. The accuracy of the open-shell LNO-CCSD(T) was extensively tested for radicals and reactions thereof, ionization processes, as well as spin-state splittings and transition-metal compounds against the canonical CCSD(T) reference. The average open-shell LNO-CCSD(T) correlation energies were at least 99.95% accurate, which translates into average absolute deviations of 0.1–0.2 kcal/mol in energy differences already with the default settings. This enables the modeling of large systems with complex electronic structure as illustrated on open-shell organic radicals and transition metal complexes of up to 179 atoms, as well as on challenging biochemical systems, including up to 601 atoms and 11,000 basis functions. While the protein models involve difficulties for local approximations, such as the spin states of a bounded iron ion or an extremely delocalized singly occupied orbital, the corresponding single-node LNO-CCSD(T) computations were feasible in a matter of days with tens to a 100 GB of memory use. Therefore, the new LNO-CCSD(T) implementation enables highly-accurate computations for open-shell systems of unprecedented size and complexity with widely accessible hardware.

In a joint quest with our collaborators, a large-scale benchmark calculations were carried out to assess the accuracy of reference quantum mechanical methods [5]. We compared the performance of quantum diffusion Monte Carlo (DMC) and CCSD(T), which have been categorically shown to yield accurate interaction energies for small molecules. For a small set of large molecular complexes, we showed that DMC and our LNO-CCSD(T) unequivocally disagree by up to 8 kcal/mol in the interaction energies. This leads to differences of up to 6 orders of magnitude in the corresponding binding association constant at room temperature. The disagreement is presumably caused by the inaccuracy of the DMC results due to the poor nodal structure of the Kohn–Sham (KS) density used as the starting point of DMC calculations.

A framework for the reduced-scaling implementation of excited-state correlation methods was developed [6]. An algorithm was introduced to construct excitation-specific local domains, which include all the important molecular orbitals for the excitation as well as for the electron correlation. The orbital space dimensions of the resulting compact domains were further decreased utilizing our reduced-cost techniques developed previously based on the natural auxiliary function (NAF) and LNO approaches. Additional methodological improvements for the evaluation of density matrices were also made. Benchmark calculations were performed at the second-order algebraic-diagrammatic construction [ADC(2)] level. Compared to our reduced-cost algorithm significant, up to 3-9-fold speedups are achieved even for systems of smaller than 100 atoms. At the same time the additional errors introduced by the domain approximations are highly acceptable being about 2-4 meV on the average. The new reduced-scaling algorithm allows us to carry out correlated excited-state calculations using triple-zeta basis sets with diffuse functions for systems of up to 400 atoms or 13000 atomic orbitals in a matter of days using an 8-core processor.

Reduced-cost correlation approaches

A completely integral-direct, disk I/O and network traffic economic CCSD(T) implementation was developed relying on the density-fitting (DF) approximation [7]. By fully exploiting the permutational symmetry the presented algorithm is highly operation-count and memory efficient, while also exhibiting excellent strong scaling via hybrid MPI/OpenMP parallelization. Additionally, our measurements demonstrate a highly competitive, 60-70% utilization of the theoretical peak performance on up to hundreds of cores. The terms whose evaluation time becomes significant only for small- to medium-sized examples were also extensively optimized. Consequently, high performance is also expected for systems appearing in extensive data sets used, e.g., for density functional or machine learning parametrizations, and in calculations required for certain local or reduced-cost approximations of CCSD(T). The efficiency of our implementation allowed us to preform some of the largest CCSD(T) calculations ever presented for systems of 31-43 atoms and 1037-1569 orbitals using only a single CPU and 1-3 days of wall time. The utilization of this well-parallelized CCSD(T) code in combination with our LNO scheme also significantly extended the reach of the present LNO-CCSD(T) implementation.

The accurate and systematically improvable frozen natural orbital (FNO) and NAF cost-reducing approaches were combined with our recent CCSD(T) implementations [8]. Both of the closed- and open-shell FNO-CCSD(T) codes benefit from OpenMP parallelism, completely or partially integral-direct DF algorithms, checkpointing, and hand-optimized, memory- and operation count effective implementations exploiting all permutational symmetries. The closed-shell CCSD(T) code requires negligible disk I/O and network bandwidth, is MPI/OpenMP parallel, and exhibits outstanding peak performance utilization of 50-70% up to hundreds of cores. Conservative FNO and NAF truncation thresholds benchmarked for challenging reaction, atomization, and ionization energies of both closed- and open-shell species are shown to maintain 1 kJ/mol accuracy against canonical CCSD(T) for systems of 31-43 atoms even with large basis sets. The cost reduction of up to an order of magnitude achieved extends the reach of FNO-CCSD(T) to systems of 50-75 atoms (up to 2124 atomic orbitals) with triple- and quadruple-zeta basis sets, which is unprecedented without local approximations. Consequently, a considerably larger portion of the chemical compound space can now be covered by the practically “gold standard” quality FNO-CCSD(T) method using affordable resources and about a week of wall time. Large-scale applications were presented for organocatalytic and transition-metal reactions as well as noncovalent interactions. Possible applications for benchmarking local CCSD(T) methods, as well as for the accuracy assessment or parametrization of less complete models, for example, density functional approximations or machine learning potentials, were also outlined.

A number of approaches were proposed and assessed to reduce the FNO truncation error of CCSD(T) energies [9]. The diagrammatic energy decomposition method of Irmler and Grüneis [*J. Chem. Phys.* **151**, 104107 (2019)] was extended to the FNO truncation correction of the particle-particle ladder (PPL) term in the case of closed- and open-shell molecular systems. The approach was tested for reaction, interaction, and atomization energies, and it was found the most robust for a wider range of FNO truncation thresholds outperforming the commonly employed additive MP2 correction. We also showed that the linear extrapolation (LE) of FNO-CCSD(T) energies as a function of MP2 energies provides the best correlation energies and most balanced energy differences with tighter FNO thresholds, but it lacks systematic error compensation that would be required for the better performance with looser FNO thresholds. Further insight was gained from a diagrammatic and spin-component decomposition based analysis. Moreover, orbital (pair) specific energy decompositions were utilized to introduce size-consistent variants of the promising PPL and LE FNO corrections and their analogues for (T), which are also readily applicable in the context of popular local correlation methods.

A new approach was proposed to reduce the basis set incompleteness error of the triple excitation correction in explicitly correlated CCSD(T) calculations [10]. Our method, termed (T+), is similar to the intuitive triples correction approach of Knizia et al. [*J. Chem. Phys.* **130**, 054104 (2009)] but, in contrast to the latter, is size-consistent. The new approximation is easy to implement, and its overhead is negligible with respect to the conventional (T) correction. The performance of the approach was assessed for atomization, reaction, and interaction energies as well as for bond lengths and harmonic vibrational frequencies. The advantages of its size consistency were also demonstrated.

Our conventional, explicitly correlated, and LNO-based CCSD(T) methods were utilized to closely approach the basis set limit of interaction energies between medium-sized molecular complexes [11]. The study was partly motivated by our previous finding that state-of-the-art DMC and CCSD(T) interaction energies are not in complete agreement even for some of the dimers in the popular S66 compilation [5]. We improved upon the previous S66 CCSD(T) reference by using up to the quadruple-zeta (QZ) level CCSD(T) contributions, utilizing our recent (T+) approach and suggesting a more robust explicitly correlated CC singles and doubles (CCSD) reference. Our efficient LNO-CCSD(T) method allowed us to also compute tightly converged, quintuple-zeta level interaction energies for the S66 set. Moreover, the new CCSD(T)/QZ references enabled the so far most rigorous benchmark of the LNO and other popular local correlation approaches on the S66 set. The accuracy of local approximations was found in the few 10 to few 100 cal/mol interaction energy error range, assuming sufficiently tight settings, such as the Tight and the Normal settings of LNO-CCSD(T) threshold combinations, respectively.

We demonstrated on a representative example that full-dimensional machine-learned potentials can be effectively constructed with CCSD(T) for molecules of 15 atoms using off-the-shelf hardware [12]. This was achieved by accelerating the CCSD(T) computations via the accurate and cost-effective FNO approach. The Δ-machine learning (Δ-ML) approach was employed with the use of permutationally invariant polynomials to fit a full-dimensional potential energy surface (PES) of the acetylacetone molecule, but any other effective descriptor and ML approach can similarly benefit from the accelerated data generation proposed. Our benchmarks for the global minima, H-transfer

TS, and many high-lying configurations show the excellent agreement of FNO-CCSD(T) results with conventional CCSD(T) while achieving a significant time advantage of about a factor of 30–40. The obtained Δ-ML PES shows high fidelity from multiple perspectives including energetic, structural, and vibrational properties. The methodology introduced represents a significant advancement in the efficient and precise construction of potentials at the CCSD(T) level for molecules above the current limit of 15 atoms.

We introduced and tested several approximations to reduce the computational expenses of the explicitly correlated CCSD(T) method for both closed and open-shell species [13]. First, the well-established FNO technique was adapted to explicitly correlated CC approaches. Second, our NAF scheme was employed to reduce the size of the auxiliary basis required for the DF approximation regularly used in explicitly correlated calculations. Third, a new approach, termed the natural auxiliary basis (NAB) approximation, was proposed to decrease the size of the auxiliary basis needed for the expansion of the explicitly correlated geminals. The performance of the above approximations and that of the combined FNO-NAF-NAB approach were tested for atomization and reaction energies. Our results show that overall speedups of 7-, 5-, and 3-times can be achieved with double-, triple-, and quadruple-ζ basis sets, respectively, without any loss in accuracy. The new method can provide, e.g., reaction energies and barrier heights well within chemical accuracy for molecules with more than 40 atoms within a few days using a few dozen processor cores, and calculations with 50+ atoms were still feasible. These routinely affordable computations considerably extend the reach of explicitly correlated CCSD(T).

Based on the reduced-cost explicitly correlated CC method developed, a reduced-cost explicitly correlated MP2 (MP2-F12) approach was presented [14]. The FNO, NAB, and NAF approximations were utilized to decrease the size of the molecular orbital basis and the auxiliary basis sets required for the DF and for the expansion of the explicitly correlated geminals. The necessary modifications of the above approximations for MP2-F12 were elaborated. The performance of the new method was tested for atomization and reaction energies. Our results demonstrate that speedups of 1.5-4 can be achieved in the calculation of MP2-F12 correlation energies with a moderate loss of accuracy.

In collaboration with our Italian colleagues, our newly developed explicitly correlated CCSD(T) ansatz was applied to thermochemistry [15]. The "junChS-F12" composite method was improved by means of our reduced-cost method and validated for the thermochemistry of molecules containing second- and third-row atoms. A thorough benchmark showed that this model, in conjunction with cost-effective revDSD-PBEP86-D3(BJ) reference geometries, offers an optimal compromise between accuracy and computational cost without employing any empirical parameter. If improved geometries are sought, the most effective option is to add MP2-F12 core-valence correlation corrections to explicitly correlated CCSD(T) geometries without the need of performing any extrapolation to the CBS limit. In the same vein, explicitly correlated CCSD(T) harmonic frequencies are remarkably accurate without any additional contribution. Pilot applications to non-covalent intermolecular interactions, conformational landscapes, and tautomeric equilibria confirm the effectiveness and reliability of the model.

Utilizing the DF approximation, we presented an efficient implementation of the density-based basis-set correction scheme, which were primarily developed to mitigate the basis set incompleteness error of wave function methods [16]. A thorough comparison was also carried out for explicitly correlated and density-based approaches. The performance of these approaches was comprehensively tested for the MP2, CCSD, and CCSD(T) methods with respect to the corresponding complete basis set references. It was demonstrated that the density-based correction together with complementary auxiliary basis set (CABS)-corrected Hartree–Fock (HF) energies is highly robust and effectively reduces the error of the standard approaches; however, it does not outperform the corresponding explicitly correlated methods. Nevertheless, what still makes the density-corrected CCSD and CCSD(T) methods competitive is that their computational costs are roughly half of that of the corresponding explicitly correlated variants. Additionally, an incremental approach for standard CCSD and CCSD(T) was introduced. In this simple scheme, the total energies are corrected with the CABS correction and explicitly correlated MP2 contributions. As demonstrated, the resulting methods yield surprisingly good results, below 1 kcal/mol for thermochemical properties even with a double-ζ basis, while their computational expenses are practically identical to those of the density-based basis-set correction approaches.

A reduced-cost implementation of the ADC(2) method was presented [17]. We introduced approximations by restricting virtual natural orbitals and NAFs, which results, on the average, in more than an order of magnitude speedup compared to conventional, DF algorithms for singlet and triplet excitation energies and transition moments. The various approximations were carefully benchmarked, and conservative truncation thresholds were selected which guarantee errors much smaller than the intrinsic error of the ADC(2) method. Using the canonical values as reference, we found that the mean absolute error for both singlet and triplet excitation energies was 0.02 eV, while that for oscillator strengths was 0.001 a.u. The rigorous cutoff parameters together with the significantly reduced operation count and storage requirements allow us to obtain accurate excitation energies and transition properties using triple-zeta basis sets for systems of up to one hundred atoms.

The aforementioned reduced-cost scheme was extended to core excitations [18]. The efficiency of the approximation was demonstrated for the ADC(2) method invoking the core-valence separation (CVS) and DF approaches. The errors introduced by the scheme were comprehensively analyzed for more than 200 excitation energies and 80 oscillator strengths, including C, N, and O K-edge excitations, as well as $1s \rightarrow \pi^*$ and Rydberg transitions. Our results show that significant savings can be gained in computational requirements at the expense of a moderate error. That is, the mean absolute error for the excitation energies, being lower than 0.20 eV, is an order of magnitude smaller than the intrinsic error of CVS-ADC(2), while the mean relative error for the oscillator strengths is between 0.06 and 0.08, which is still acceptable. As significant differences for different types of excitations cannot be observed, the robustness of the approximation was also demonstrated. The improvements in the computational requirements were measured for extended molecules. In this case, an overall 7-fold speedup was obtained in the wall-clock times, while dramatic reductions in the memory requirements were also achieved. In addition, it was also proved that the new approach enables us to perform CVS-ADC(2) calculations within reasonable runtime for systems of 100 atoms using reliable basis sets.

Density functional approximations

We analyzed various possibilities to improve upon the SCAN meta-generalized gradient approximation density functional obeying all know properties of the exact functional which can be satisfied at this level of approximation [19]. We examined the necessity of locally satisfying a strongly tightened lower bound for the exchange energy density in single-orbital regions, the nature of the error cancellation between the exchange and correlation parts in two-electron regions, and the deviation from the second-order gradient expansion of the correlation energy density. We observed a general improvement in the single-orbital electron densities by revising the correlation functional form to follow the second-order gradient expansion in a wider range. Our new revSCAN functional provides more accurate atomization energies, non-covalent interaction energies, and reaction barrier heights than the original SCAN functional. New direct random phase approximation (dRPA) based DH functionals derived from the adiabatic connection formalism were implemented. Their parameters were optimized for both thermochemical properties and densities. The new approaches were benchmarked for thermochemistry, weak interactions, and barrier heights. Our results revealed that the performance of the new models, either with or without spin-scaling, falls behind that of our previous, more empirical dRPA75 method.

Preserving the beneficial properties of the second-order screened exchange (SOSEX) method, such as its freedom from one-body self-correlation error, and its seamless treatment of long-range dispersion interaction, we constructed new non-local functionals by down-scaling the higher-order terms in the SOSEX series to reduce the many-body self-correlation error in molecular systems [20]. Our down-scaled SOSEX (dsSOSEX) and scaled equation SOSEX (seSOSEX) approaches deliver considerably more accurate non-covalent interaction energies, reaction energies, and barrier heights than the original SOSEX method. The dsSOSEX approach improves on the description of medium and long range correlation, while the seSOSEX approach improves on the description of processes with short- and medium-range rearrangement in the electronic structure. As our dsSOSEX method provides a better description of medium- and long-range correlation, it generally leads to more accurate energetics than the seSOSEX method, hence it should be preferred over the latter. We also compared our results to other empirical and non-empirical dRPA and beyond-dRPA methods.

Blending the good performance of the Perdew, Burke, and Ernzerhof (PBE) global hybrid functional (PBE0) at short-range and the dual-hybrid dRPA75 functional at long-range, we proposed a new range-separated (RS) direct random phase approximation (dRPA75rs), which considerably improves on the accuracy of the calculated reaction energies and barrier heights compared to the parent approaches and provides a good description of non-covalent interactions without any dispersion correction [21]. We also combined the new scheme with spin-component scaling, which enables the accurate calculation of energy differences for processes involving electron pair breaking. The new method scaling as the fourth power of the system size shows a balanced performance on a broad test set involving radicals, transition metal atoms, and heavy atoms, which makes it competitive with the best DH functionals based on the second-order perturbation theory. According to the results for the homogeneous electron gas, our dRPA75rs method expectedly gives errors for metallic systems similar to the dRPA approach with an additional error cancellation in the case of partial spin-polarization.

A composite of the time-dependent DFT (TDDFT) and the ADC(2) approach was presented for the efficient calculation of spectral properties of molecules [22]. Our method can be regarded as a new excited-state DH approach or a dressed TDDFT scheme, but it can also be interpreted as an empirically tuned ADC(2) model. Several combinations of exchange-correlation functionals and spin-scaling schemes were explored. Our best-performing method includes the PBE exchange and Perdew's 1986 correlation functional and employs the scaled-opposite-spin approximation for the higher-order terms. The computation time of the new method scales as the fourth power of the system size, and an efficient cost-reduction approach was also presented, which further speeds up the calculations. Our benchmark calculations showed that the proposed model outperforms not only the existing DH approaches and ADC(2) variants but also the considerably more expensive CC methods.

A simple and robust RS DH time-dependent density functional approach was developed for the accurate calculation of excitation energies of molecules within the Tamm–Dancoff approximation [23]. The scheme can be considered as an excited-state extension of the ansatz proposed by Toulouse and co-workers [*J. Chem. Phys.* **148**, 164105 (2018)], which is based on the two-parameter decomposition of the Coulomb potential, for which both the exchange and correlation contributions are range-separated. A flexible and efficient implementation of the new scheme was also presented, which facilitates its extension to any combination of exchange and correlation functionals. The performance of the new approximation was tested for singlet excitations on several benchmark compilations and thoroughly compared to that of representative DH, RS hybrid, and RS DH functionals. The one-electron basis set dependence and computation times were also assessed. Our results show that the new approach improves on standard DHs in most cases, and it can provide a more robust and accurate alternative. In addition, on average, it noticeably surpasses the existing RS hybrid and RS DH functionals.

The aforementioned time-dependent density functional approach was combined with spin-scaling techniques [24]. The proposed spin-component-scaled (SCS) and scaled-opposite-spin (SOS) variants were thoroughly tested for almost 500 excitations including the most challenging types. This comprehensive study provides useful information not only about the new approaches but also about the most prominent methods in the DH class. The benchmark calculations confirm the robustness of the RS-DH ansatz, while several tendencies and deficiencies were pointed out for the existing functionals. Our results show that the SCS variant consistently improves the results, while the SOS variant preserves the benefits of the original RS-DH method reducing its computational expenses. It was also demonstrated that, besides our approaches, only the nonempirical functionals provide balanced performance for general applications, while particular methods are only suggested for certain types of excitations.

Our ADC(2)-based DH ansatz [22] was combined with range-separation techniques [25]. In the presented scheme, both the exchange and the correlation contributions were range-separated, while spin-scaling approaches were also applied. The new methods were thoroughly tested for the most popular benchmark sets including 250 singlet and 156 triplet excitations, as well as 80 oscillator strengths. It was demonstrated that the range separation for the correlation contributions is highly recommended for both the genuine and the ADC(2)-based DH approaches. Our results show that the latter scheme slightly but consistently outperforms the former one for single excitation dominated transitions. Furthermore, states with larger fractions of double excitations were assessed as well, and challenging charge-transfer (CT) excitations were also discussed, where the recently proposed spin-scaled long-range corrected DHs fail. The suggested iterative fourth-power scaling RS-PBE-P86/SOS-ADC(2) method, using only three adjustable parameters, provides the most robust and accurate excitation energies within the DH theory. In addition, the relative error of the oscillator strengths was reduced by 65% compared to the best genuine DH functionals.

The performance of the most recent density functionals was assessed for CT excitations using comprehensive intra- and intermolecular CT benchmark sets with high-quality reference values [26]. For this comparison, the state-of-the-art RS and long-range corrected (LC) DH approaches were selected, as well as global DH and LC hybrid functionals were also inspected. The correct long-range behavior of the exchange-correlation (XC) energy was extensively studied, and various CT descriptors were compared as well. Our results show that the most robust performance is attained by RS-PBE-P86/SOS-ADC(2) as it was suitable to describe both types of CT excitations with outstanding accuracy. Furthermore, concerning the intramolecular transitions, unexpectedly excellent results were obtained for most of the global DHs, however, their limitations were also demonstrated for bimolecular complexes. Despite the outstanding performance of the LC-DH methods for common intramolecular excitations, serious deficiencies were pointed out for intermolecular CT transitions, as well as the wrong long-range behavior of the XC energy was revealed. The application of LC hybrids to such transitions is not recommended in any respect.

The DH TDDFT was extended to vertical ionization potentials (VIPs) and electron affinities (VEAs) [27]. Utilizing the DF approximation, efficient implementations were presented for the genuine DH ansatz relying on the perturbative second-order correction, while an iterative analogue was also elaborated using our ADC(2)-based DH approach. The favorable computational requirements of these schemes were demonstrated. The performance of the recently proposed SCS and SOS RS and LC DH functionals was comprehensively assessed, while popular hybrid and global DH approaches were also considered. For the benchmark calculations, up-to-date test sets were selected with high-level CC references. Our results show that our ADC(2)-based SOS-RS-PBE-P86 approach is the most accurate and robust functional. This method consistently outperforms the excellent SOS-ADC(2) approach for VIPs, although the results were somewhat less satisfactory for VEAs. Among the genuine DH functionals, the SOS-ωPBEPP86 approach is also recommended for describing ionization processes, but its performance is even less reliable for electron-attached states. In addition, surprisingly good results were attained by the LC hybrid ωB97X-D functional, where the corresponding occupied (unoccupied) orbital energies are retrieved as VIPs (VEAs) within the present formalism.

The DH TDDFT was extended to core excitations [28]. Two different DH formalisms were presented utilizing the CVS approximation. First, a CVS-DH variant was introduced relying on the genuine perturbative second-order correction, while an iterative analogue was also presented using our ADC(2)-based DH ansatz. The performance of the new approaches was tested for the most popular DH functionals using the recently proposed XABOOM [*J. Chem. Theory Comput.* **17**, 1618 (2021)] benchmark set. In order to make a careful comparison, the accuracy and precision of the methods were also inspected. Our results show that the genuine approaches are highly competitive with the more advanced CVS-ADC(2)-based methods if only excitation energies are required. In contrast, as expected, significant differences were observed in oscillator strengths; however, the precision was acceptable for the genuine functionals as well. Concerning the performance of the CVS-DH approaches, the PBE0-2/CVS-ADC(2) functional was superior, while its SOS variant was also recommended as a cost-effective alternative. For these approaches, significant improvements were realized in the error measures compared with the popular CVS-ADC(2) method.

Embedding approaches

A dual-basis (DB) approach was proposed which is suitable for the reduction of the computational expenses of the HF, KS, and wave-function-based correlation methods, and is specifically designed for embedding calculations [29]. The new approach was applied to our Huzinaga embedding theory, multi-level local correlation methods, and combined density functional-multi-level local correlation approximations. The performance of the resulting DB density functional and wave function embedding methods was evaluated in extensive benchmark calculations 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.

To enable correlated calculations in realistic solvent environments, we implemented the embedded cluster reference interaction site model (EC-RISM) originally developed by Kloss, Heil, and Kast [*J. Phys. Chem. B* **112**, 4337 (2008)] [30]. This method combines quantum mechanical calculations with the 3D reference interaction site model (3D-RISM). Numerous options, such as buffer, grid space, basis set, charge model, water model, closure relation, etc., were investigated to find the best settings. Additionally, the small point charges, which were derived from the solvent distribution from the 3D-RISM solution to represent the solvent in the QM calculation, were neglected to reduce the overhead without the loss of accuracy. On the MNSOL[a], MNSOL, and FreeSolv databases, our implemented and optimized method provides solvation free energies in water with 5.70, 6.32, and 6.44 kJ/mol root mean square deviations, respectively, but with different settings, 5.22, 6.08 and 6.63 kJ/mol can also be achieved. Only solvent models containing fitting parameters, like COSMO-RS and EC-RISM with universal correction and directly used electrostatic potential, perform better than our EC-RISM implementation with atomic charges.

The performance of multilevel quantum chemical approaches, which utilize an atom-based system partitioning scheme to model various electronic excited states, was studied [31]. In addition, an improved variant of our Huzinaga-embedding scheme was proposed. The considered techniques include the mechanical-embedding (ME) of "our own N-layered integrated molecular orbital and molecular mechanics" (ONIOM) method, the point charge embedding (PCE), the electronic-embedding (EE) of ONIOM, the frozen density-embedding (FDE), the projector-based embedding (PbE), and our local domain-based correlation method. For the investigated multilevel approaches, the ADC(2) approach was utilized as the high-level method, which was embedded in either HF or a density functional environment. The XH-27 test set of Zech et al. was used for the assessment, where organic dyes interact with several solvent molecules. With the selection of the chromophores as active subsystems, we concluded that the most reliable approach is local domain-based ADC(2) [L-ADC(2)], and the least robust schemes are ONIOM-ME and ONIOM-EE. The PbE, FDE, and PCE techniques often approach the accuracy of the L-ADC(2) scheme, but their precision is far behind. The results suggest that a more conservative subsystem selection algorithm or the inclusion of subsystem charge-transfers is required for the atom-based cost-efficient methods to produce high-accuracy excitation energies.

The theory of analytic gradients was presented for the projector based DFT embedding approach utilizing the Huzinaga-equation [32]. The advantages of the Huzinaga-equation-based formulation were demonstrated. In particular, it was shown that the projector employed does not appear in the Lagrangian, and the potential risk of numerical problems is avoided at the evaluation of the gradients. The implementation of the analytic gradient theory was presented for the approaches where hybrid DFT, MP2, or double hybrid DFT is embedded in lower-level DFT environments. To demonstrate the applicability of the method and to gain insight into its accuracy, it was applied to equilibrium geometry optimizations, transition state searches, and PES scans. Our results show that bond lengths and angles converge rapidly with the size of the embedded system, and the embedding approach has the potential to relax the coordinates of the environment as well. These observations together with the timings presented suggest that the Huzinaga-equation-based embedding is a viable alternative for the geometry optimizations of large molecular systems.

In collaboration with our theoretical chemist colleagues, the projected atomic orbitals (PAO) technique was utilized for the construction of virtual orbital spaces in PbE applications [33]. The proposed straightforward procedure produces a set of orthogonal virtual orbitals, which is used in the final, high-level calculation of the embedded active subsystem. The PAO scheme was tested on intermolecular potentials of bimolecular complexes, in ground and excited states, including Rydberg excitations. The results show the outstanding performance of the PbE embedding method with PAO virtual orbitals compared to those produced using common orbital localization techniques. The good agreement of the resulting PbE potential curves with those from high-level ab initio dimer calculations, also in diffuse basis sets, confirms that the PAO technique can be suggested for future applications of this type. The use of D3 dispersion corrections in such calculations was also proposed, supported by a superior accuracy over dispersion terms sourced from the effective fragment potential model.

Improved HF, DFT, and integral calculation algorithms

The calculation of the geometrical derivatives of three-center electron repulsion integrals (ERIs) over contracted spherical harmonic Gaussians has been optimized [34]. We compared various methods based on the Obara–Saika (OS), McMurchie–Davidson, Gill–Head-Gordon–Pople (GHP), and Rys polynomial algorithms using Cartesian, Hermite, and mixed Gaussian integrals. Furthermore, prescreening of the ERI derivatives was discussed, and an efficient approximation of the Cauchy–Schwartz bound for first derivatives was presented. Based on the estimated operation counts the most promising schemes were implemented by automated code generation, and their relative performances were evaluated. In general the OS method with Cartesian Gaussians proved to be the best approach, but when derivatives for one degree of freedom are required at a time the mixed Gaussian OS and GHP algorithms were predicted to be the best performing ones.

An approximation was presented which can efficiently decrease the computational expenses of configuration interaction singles (CIS) and TDDFT methods employing hybrid functionals [35]. The approach is the adaptation of the local DF scheme developed for HF calculations for excited states and reduces the quartic scaling of the methods to cubic. It can also be applied to related methods, such as the time-dependent HF and Tamm–Dancoff approximation TDDFT approaches. Our benchmark calculations showed that, for molecules of 50 to 100 atoms, average speedups of 2 to 4 can be achieved for CIS and TDDFT calculations at the expense of negligible errors in the calculated excitation energies and oscillator strengths, but for bigger systems or molecules of localized electronic structure significantly larger speedups can be gained. We also demonstrated that the approximation enables excited-state calculations on a single processor even for molecules of 1000 atoms using basis sets augmented with diffuse functions including more than 17000 atomic orbitals.

The multipole approximation was utilized for reducing the computational expenses of the exchange contribution in DF Hartree–Fock (DF-HF) calculations [36]. Strategies for approximating the relevant three-center Coulomb integrals with the multipole expansion were developed. Based on the factorized form of the integrals, an algorithm was proposed for the evaluation of the exchange term for both conventional and local DF-HF methods. The accuracy of the resulting energies, the numerical stability of the algorithm, and the speedups achieved were benchmarked with respect to the order of the multipole expansion for various molecular systems. Our results suggest that computation times for a conventional DF-HF calculation can be reduced roughly by a factor of 1.5 for molecules of a couple of hundreds of atoms without any loss of accuracy, while the speedups are somewhat more moderate if local DF approximations are also deployed.

Several further approaches were developed to improve the efficiency of HF and KS SCF calculations relying on a simple first-order energy correction reminiscent of the scheme used in DB SCF methods [37]. The basic idea is to perform an initial SCF calculation computing approximate Fock-matrices, and in the final iteration step, to use a more complete Fock-matrix builder together with the energy correction to diminish the error. The approximation was tested for conventional and local DF SCF approaches combining various auxiliary basis sets, fitting metrics, and Fock-matrix construction algorithms in the initial and final iterations, as well as for seminumerical SCF methods combining integration grids of different quality. We also reported the implementation of the occupied orbital resolution of identity exchange construction (occ-RI-K) algorithm with local DF approximations. Benchmark calculations were presented for total energies, reaction energies, and molecular geometries. Our results show that speedups of up to 80% can be expected utilizing the new approaches without significant loss of accuracy.

A new multilevel approach was presented to the initial guess for SCF calculations, which combines the superposition of atomic densities (SAD) procedure and the density matrix of a semi-empirical quantum mechanics (SQM) calculation through projection [38]. The proposed initial guess method produces a polarized, spin-specific initial density, while its computational costs are a few orders of magnitude lower than the expenses of a scheme that projects the density matrix of a standard quantum chemical calculation utilizing a minimal AO basis set (pMIN). The projected SQM density-based (pSQM) technique was thoroughly tested using the second version of the extended tight binding method (GFN2-xTB), and its efficiency and reliability were compared with those of the standard SAD and density matrix projection techniques. The results indicate that the calculations using the pSQM scheme require somewhat fewer SCF iteration steps compared with the SAD method, however, the SAD, the pSQM, and the pMIN techniques have similar performance.

We developed analytic gradients for local DF HF and hybrid KS density functional methods [39]. Due to the non-variational nature of the local DF algorithm, the method of Lagrange multipliers was used to avoid the solution of the coupled-perturbed HF and KS equations. We proposed efficient algorithms for the solution of the arising Z-vector equations and the gradient calculation, which preserve the third-order scaling and low memory requirement of the original local fitting algorithm. In order to demonstrate the speed and accuracy of our implementation, gradient calculations and geometry optimizations were carried out for various molecular systems. Our results show that significant speedups can be achieved compared to conventional DF calculations without sacrificing accuracy.

Applications to real-life problems

Reliable heats of formation were reported for numerous fluorinated and chlorinated methane and ethane derivatives by means of an accurate thermochemical protocol, which involves explicitly correlated CC calculations augmented with anharmonic, scalar relativistic, and diagonal Born–Oppenheimer corrections [40]. The theoretical results, along with additional experimental data, were further enhanced with the help of the thermochemical network approach. For 28 species, out of 50, our study presented the best estimates, and discrepancies with previous reports were also highlighted. Furthermore, the effects of the less accurate theoretical data on the results yielded by thermochemical networks were discussed.

Accurate thermochemical data were also presented for uracil, thymine, cytosine, and adenine [41]. The gas phase heats of formation of these compounds were determined with the aforementioned thermochemical protocol. The thermochemically relevant tautomers of cytosine were also investigated. To derive heats of formation in condensed phases as well as sublimation enthalpies, the thermochemical network approach was utilized, i.e., the available literature data were collected, reviewed, and combined with our theoretical calculations. Solvation free energies were also determined with various methods and compared to experimental data.

In a joint research project with molecular biologists, we studied the reliability of molecular mechanics (MM) simulations in describing biomolecular ion-driven processes [42]. Using our high-accuracy LNO-CCSD(T) benchmark data, it was demonstrated for a representative MM model that when the polarization descriptors of its ligands are improved to respond to both low and high electric fields, ligand interactions with ions also improve, and transferability errors of MM force fields reduce substantially. The observations can be used to design more reliable MM force fields.

In a similar collaborative research project, we studied the relative interactions of Li^+ and Mg^{2+} ions with water and other biochemical groups [43]. The interactions of these types are important because the therapeutic effect of Li^+, in many cases, stems from its ability to inhibit certain Mg^{2+}-dependent enzymes, where it interacts with or substitutes for Mg^{2+}. Reference data were taken from our LNO-CCSD(T) calculations, followed by a benchmarked van der Waals-inclusive DFT. Improvements were recommended to two simulation approaches: the one that employs an all-atom polarizable

MM model and the other that uses a hybrid quantum and MM implementation of the quasi-chemical theory. Our results show that accurate descriptions of local interactions are essential, and the recalibration of local interactions in these models reduced their errors by almost an order of magnitude. This work also sets up a systematic approach to evaluate and improve molecular models of ions binding to proteins utilizing the accurate local correlation methods developed by us.

In a joint quest with our collaborators, the methylation of lysine was studied [44]. The process can modify non-covalent interactions by altering lysine's hydrophobicity as well as its electronic structure. While ramifications of the former were documented, the effects of the latter remain largely unknown. Understanding electronic structure is important for determining how biological methylation modulates protein-protein binding, and how artificial methylation impacts experiments in which methylated lysines are used as spectroscopic probes and protein crystallization facilitators. The benchmarked first principles calculations undertaken by us revealed that methyl-induced polarization weakens electrostatic attraction of amines with protein functional groups-salt bridges, hydrogen bonds, and cation-pi interactions weaken by as much as 10.3, 7.9 and 3.5 kT, respectively. Multipole analysis shows that weakened electrostatics is due to altered inductive effects that overcome increased attraction from methyl-enhanced polarizability and dispersion. These effects, due to their fundamental nature, are expected to be present in many cases. Survey of methylated lysines in protein structures reveals several cases where methyl-induced polarization is the primary driver of altered non-covalent interactions, and in these cases, destabilizations were found to be in the 0.6-4.7 kT range. The clearest case of where methyl-induced polarization plays a dominant role in regulating biological function is that of the binding domain that recognizes lysine methylated-states on histones. These results broaden our understanding of how methylation modulates non-covalent interactions.

We studied the stability of cyclic phosphinyl radicals with an aim for a systematic assessment of stabilization effects [45]. The radical stabilization energies (RSEs) were calculated using isodesmic reactions for a large number of carbocyclic radicals possessing different ring sizes and grades of unsaturation. In general, the RSE values range from -1.2 to -14.0 kcal/mol, and they show practically no correlation with the spin populations at the P-centers. The RSE values correlate with the reaction Gibbs free energies calculated for the dimerization of the studied simple radicals. Therefore, the more easily accessible RSE values offer a cost-effective estimation of global stability in a straightforward manner. To explore the effect of unsaturation on the RSE values, delocalization energies were determined using appropriate isodesmic reactions. Introducing unsaturations beside the P-center into the backbone of the rings leads to an additive increase in the magnitude of the delocalization energy. Radicals containing exocyclic C=C π-bonds were also investigated, and all of these radicals have rather similar stabilities independently of the ring size, outlining the primary importance of the two exocyclic π-bonds in the conjugation.

In a joint quest with or theoretical chemist collaborators, the prototypical molecule–surface system, the adsorption of CO on MgO was studied [46]. CCSD(T) calculations were carried out to evaluate the adsorption energy of CO. It was shown that our recently developed cluster-based CCSD(T) approach provides a low-cost route toward achieving accurate adsorption energies. Our computed adsorption energies were compared to DMC and experimental results. A consensus could be reached between multiple theoretical calculations and experiments for the first time.

In collaboration with American colleagues, a systematic assessment of the polarizable AMOEBA protein model was provided and critical changes were recommended that substantially improve its predictive performance [47]. High-level CCSD(T) calculation were performed to benchmark the van der Waals-inclusive density functionals that were used to parameterize the nonbonded-fix corrections. A new model of ATP with revised polarization parameters was also presented. It was demonstrated that these improvements lead to a better agreement between theory and experiment for Mg^2+·kinase·ATP complexes.

In a joint quest with our polymer chemist collaborators, the stabilization effect of a flavonoid type natural antioxidant, rutin, was compared to that of quercetin in polyethylene [48]. The experiments showed that the efficiency of the antioxidizing system reaches a minimum at a certain concentration of stabilizers, and molecular modeling was used to investigate the interactions responsible for this effect. The calculations utilized semi empirical methods for the structure optimization step followed by DFT calculations with our DB approach. The calculations predicted the development of strong interactions between quercetin/rutin and the secondary stabilizer, however, the interactions do not involve the hydroxyl groups with antioxidant function. Thus, the complex formation between the antioxidants is not the sole reason for the efficiency decrease.

In a similar study, the aminolysis of poly(ethylene terephthalate) was studied [49]. The potential application of the products of the reaction as cross-linkers and catalysts in epoxy curing was inspected. Combined experimental and theoretical reaction kinetics studies were used to characterize their accelerative effects. Our studies showed that terephthalamides are active catalyst and their efficiency can be tuned with their molecular weight. Our LNO-CCSD(T) calculations suggested that the terephthalamides are in the same order of magnitude in effectiveness as phenolic accelerators. Consequently, terephthalamides are valued materials that can serve as double-purpose components in epoxy curing.

In collaboration with organic chemists and spectroscopists, new types of potential fluorescent labels were tested [50]. The complexation of phenylpyridinium derivatives with cucurbiturils was investigated. The complexation processes were modeled by density functional methods, and the LNO-CCSD(T) approach was employed to calculate accurate binding energies. The experimental observations were interpreted with the aid of our theoretical calculations. The results suggest that one of the cucurbituril complexes considered can be used as a fluorescent indicator displacement assay for the detection of positively charged analytes.

In a similar project, other types of new fluorescent compounds were studied [51]. The photochemical processes of conjugates consisting of an N,N'-dimethyl-bipyridinium unit and naphthalimide derivative fluorophores were investigated. The dyads show only weak fluorescence but form radical pairs with various Lewis-base anions, which results in a dramatic enhancement of the fluorescence intensity. The processes were modeled and the experimental observations were interpreted by theoretical methods.

In collaboration with organic chemists, a new type of potential fluorescent labels was developed [52]. The prototype of the "conditionally activatable photocages" is a new vinyltetrazine-derivatized coumarin. The tetrazine form is disabled in terms of light-induced release, however, bioorthogonal transformation of the modulating tetrazine moiety results in fully restored photoresponsivity. The design of the new compound was carried out and the experimental observations were interpreted with the aid of our theoretical calculations.

In collaboration with our experimental colleagues, we studied the host-guest complex of the anionic macrocycle, carboxylato-pillar[6]arene (WP6) and the cationic dye, oxazine 1 (OX), functioning as a fluorescent turn-on indicator displacement assay (IDA) for the cationic biomolecule, vitamin B1 (thiamine) [53]. The structures of the OX-WP6 and B1-WP6 complexes were characterized theoretical calculations.

In collaboration with our organic chemist colleagues, we participated in the development of xanthenium-derived, water-soluble, low-molecular-weight photocages allowing light-controlled cargo release in the green to red region [54]. We provided theoretical evidence for the experimental results by mapping the possible photolysis pathways. Our results suggest that the photolysis can occur in the first singlet excited state via a heterolytic dissociation resulting in an ion-pair. Another possibility is a fast intersystem crossing to the first triplet state followed by homolytic dissociation. Our theoretical results will also help in the design of further photocages with even more increased activity.

In collaboration with our spectroscopist colleagues, we participated in the development of a novel water-soluble stilbene-type molecular photoswitch synthesized by the reaction of an N-methylpyridinium unit to 7-diethylamino-4-methylcoumarin via an ethylene bridge [55]. We studied the photoisomerization cycle of the photoswitch and its spectroscopic properties using theoretical methods. The PES of the photoisomerization reaction in the ground and the first singlet excited states was computed. It was confirmed that both the cis and the trans isomers are thermally stable in the ground state, but the activation barrier of the reaction is high. Our calculations also proved that the isomerization may occur in the excited state via conical intersections.

Deliverables

In this project, new quantum chemical methods have been developed and implemented which allow for accurate calculation of molecular properties for molecules of several hundreds of atoms. These approaches may open up new opportunities not only for theoretical chemistry but also for other fields of science. In particular, various disciplines that study medium-sized or large molecules composed of first- and second-row atoms may benefit from the results of the project. These disciplines include organic chemistry, biochemistry, polymer science, and molecular spectroscopy. In these fields, the new methods enable accurate calculations for significantly larger molecules than previously possible and thus facilitate the understanding of various phenomena and the design of new compounds for particular purposes. These applications also open the way for the wider use of our results, for example, in drug design, materials science, and chemical technology.

The quantitative list of the deliverables of our project is as follows. The results of our studies have been published in international, peer-reviewed journals. Altogether, 56 papers (listed below) have been published or are under publication, 28 of them in D1-quality journals, and 16 of the remaining ones in Q1-quality journals. 12 invited lectures were delivered at international conferences. In addition, the results of the project were also disseminated in several contributed talks and poster presentations. Four junior scientists of the group have successfully defended their PhD theses, and another thesis has just been submitted. Probably the most tangible deliverable of the project is the computer software including the new developments. The new methods have been implemented in the MRCC package maintained by our group and are already available to the scientific community. A review paper has also been written for the interested researchers summarizing our recent developments, the features of the suite, and our recommendations on how to use the various methods in chemical applications [56].

Publications

1. Péter R. Nagy, Gyula Samu, and Mihály Kállay: Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications, *J. Chem. Theory Comput.* **14**, 4193 (2018).

2. Péter R. Nagy and Mihály Kállay: Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods, *J. Chem. Theory Comput.* **15**, 5275 (2019).

3. P. Bernát Szabó, József Csóka, Mihály Kállay, and Péter R. Nagy: Linear scaling open-shell MP2 approach: algorithm, benchmarks, and large-scale applications, *J. Chem. Theory Comput.* **17**, 2886 (2021).

4. P. Bernát Szabó, József Csóka, Mihály Kállay, and Péter R. Nagy: Linear-scaling local natural orbital CCSD(T) approach for open-shell systems: algorithm, benchmarks, and large-scale applications, *J. Chem. Theory Comput.* **19**, 8166 (2023).

5. Yasmine S. Al-Hamdani, Péter R. Nagy, Dennis Barton, Mihály Kállay, Jan Gerit Brandenburg, and Alexandre Tkatchenko: Interactions between large molecules pose a puzzle for reference quantum mechanical methods, *Nat. Commun.* **12**, 3927 (2021).

6. Dávid Mester, Péter R. Nagy, and Mihály Kállay: Reduced-scaling correlation methods for the excited states of large molecules: Implementation and benchmarks for the second-order algebraic-diagrammatic construction approach, *J. Chem. Theory Comput.* **15**, 6111 (2019).

7. László Gyevi-Nagy, Mihály Kállay, and Péter R. Nagy: Integral-direct and parallel implementation of the CCSD(T) method: Algorithmic developments and large-scale applications, *J. Chem. Theory Comput.* **16**, 366 (2020).

8. László Gyevi-Nagy, Mihály Kállay, and Péter R. Nagy: Accurate reduced-cost CCSD(T) energies: parallel implementation, benchmarks, and large-scale applications, *J. Chem. Theory Comput.* **17**, 860 (2021).

9. Péter R. Nagy, László Gyevi-Nagy, and Mihály Kállay: Basis set truncation corrections for improved frozen natural orbital CCSD(T) energies, *Mol. Phys.* **119**, e1963495 (2021).

10. Mihály Kállay, Réka A. Horváth, László Gyevi-Nagy, and Péter R. Nagy: Size-consistent explicitly correlated triple excitation correction, *J. Chem. Phys.* **155**, 034107 (2021).

11. Péter R. Nagy, László Gyevi-Nagy, Balázs D. Lőrincz, and Mihály Kállay: Pursuing the basis set limit of CCSD(T) non-covalent interaction energies for medium-sized complexes: case study on the S66 compilation, *Mol. Phys.* **121**, e2109526 (2023).

12. Apurba Nandi and Péter R. Nagy: Combining state-of-the-art quantum chemistry and machine learning make gold standard potential energy surfaces accessible for medium-sized molecules, *Artificial Intelligence Chemistry* **2**, 100036 (2023).

13. Mihály Kállay, Réka A. Horváth, László Gyevi-Nagy, and Péter R. Nagy: Basis set limit CCSD(T) energies for extended molecules via a reduced-cost explicitly correlated approach, *J. Chem. Theory Comput.* **19**, 174 (2023).

14. Réka A. Horváth and Mihály Kállay: Basis set limit MP2 energies for extended molecules via a reduced-cost explicitly correlated approach, *Mol. Phys.* **122**, e2304103 (2024).

15. Silvia Di Grande, Mihály Kállay, and Vincenzo Barone: Accurate thermochemistry at affordable cost by means of an improved version of the JunChS-F12 model chemistry, *J. Comput. Chem.* **44**, 2149 (2023).

16. Dávid Mester and Mihály Kállay: Basis set limit of CCSD(T) energies: Explicit correlation versus density-based basis-set correction, *J. Chem. Theory Comput.* **19**, 8210 (2023).

17. Dávid Mester, Péter R. Nagy, and Mihály Kállay: Reduced-cost second-order algebraic-diagrammatic construction method for excitation energies and transition moments, *J. Chem. Phys.* **148**, 094111 (2018).

18. Dávid Mester and Mihály Kállay: Reduced-cost second-order algebraic-diagrammatic construction method for core excitations, *J. Chem. Theory Comput.* **19**, 2850 (2023).

19. Pál D. Mezei, Gábor I. Csonka, and Mihály Kállay: Simple modifications of the SCAN meta-generalized gradient approximation functional, *J. Chem. Theory Comput.* **14**, 2469 (2018).

20. Pál D. Mezei, Adrienn Ruzsinszky, and Mihály Kállay: Reducing the many-electron self-interaction error in the second-order screened exchange method, *J. Chem. Theory Comput.* **15**, 6607 (2019).

21. Pál D. Mezei and Mihály Kállay: Construction of a range-separated dual-hybrid direct random phase approximation, *J. Chem. Theory Comput.* **15**, 6678 (2019).

22. Dávid Mester and Mihály Kállay: Combined density functional and algebraic-diagrammatic construction approach for accurate excitation energies and transition moments, *J. Chem. Theory Comput.* **15**, 4440 (2019).

23. Dávid Mester and Mihály Kállay: A simple range-separated double-hybrid density functional theory for excited states, *J. Chem. Theory Comput.* **17**, 927 (2021).

24. Dávid Mester and Mihály Kállay: Spin-scaled range-separated double-hybrid density functional theory for excited states, *J. Chem. Theory Comput.* **17**, 4211 (2021).

25. Dávid Mester and Mihály Kállay: Accurate spectral properties within double-hybrid density functional theory: A spin-scaled range-separated second-order algebraic-diagrammatic construction-based approach, *J. Chem. Theory Comput.* **18**, 865 (2022).

26. Dávid Mester and Mihály Kállay: Charge-transfer excitations within density functional theory: How accurate are the most recommended approaches?, *J. Chem. Theory Comput.* **18**, 1646 (2022).

27. Dávid Mester and Mihály Kállay: Vertical ionization potentials and electron affinities at the double-hybrid density functional level, *J. Chem. Theory Comput.* **19**, 3982 (2023).

28. Dávid Mester and Mihály Kállay: Double-hybrid density functional theory for core excitations: Theory and benchmark calculations, *J. Chem. Theory Comput.* **19**, 1310 (2023).

29. Bence Hégely, Péter R. Nagy, and Mihály Kállay: Dual basis set approach for density functional and wave function embedding schemes, *J. Chem. Theory Comput.* **14**, 4600 (2018).

30. Ádám Ganyecz and Mihály Kállay: Implementation and optimization of the embedded cluster reference interaction site model, *J. Phys. Chem. A* **126**, 2417 (2022).

31. Bence Hégely, Ádám B. Szirmai, Dávid Mester, Attila Tajti, Péter G. Szalay, and Mihály Kállay: Performance of multilevel methods for excited states, *J. Phys. Chem. A* **126**, 6548 (2022).

32. József Csóka, Bence Hégely, Péter R. Nagy, and Mihály Kállay: Development of analytic gradients for the Huzinaga quantum embedding method and its applications to large-scale hybrid and double hybrid DFT forces, *J. Chem. Phys.* **160**, submitted (2024).

33. Ádám B. Szirmai, Bence Hégely, Attila Tajti, Mihály Kállay, and Péter G. Szalay: Projection based embedding with projected atomic orbitals, *J. Chem. Theory Comput.* **20**, submitted (2024).

34. Gyula Samu and Mihály Kállay: Efficient evaluation of the geometrical first derivatives of three-center Coulomb integrals, *J. Chem. Phys.* **149**, 124101 (2018).

35. Dávid Mester and Mihály Kállay: Reduced-scaling approach for configuration interaction singles and time-dependent density functional theory calculations using hybrid functionals, *J. Chem. Theory Comput.* **15**, 1690 (2019).

36. József Csóka and Mihály Kállay: Speeding up density fitting Hartree–Fock calculations with multipole approximations, *Mol. Phys.* **118**, e1769213 (2020).

37. József Csóka and Mihály Kállay: Speeding up Hartree–Fock and Kohn–Sham calculations with first-order corrections, *J. Chem. Phys.* **154**, 164114 (2021).

38. Bence Hégely and Mihály Kállay: Multilevel approach to the initial guess for self-consistent field calculations, *Int. J. Quantum Chem.* **122**, e26782 (2021).

39. József Csóka and Mihály Kállay: Analytic gradients for local density fitting Hartree–Fock and Kohn–Sham methods, *J. Chem. Phys.* **158**, 024110 (2023).

40. Ádám Ganyecz, Mihály Kállay, and József Csontos: High accuracy quantum chemical and thermochemical network data for the heats of formation of fluorinated and chlorinated methanes and ethanes, *J. Phys. Chem. A* **122**, 5993 (2018).

41. Ádám Ganyecz, Mihály Kállay, and József Csontos: Thermochemistry of uracil, thymine, cytosine, and adenine, *J. Phys. Chem. A* **123**, 4057 (2019).

42. Vered Wineman-Fisher, Yasmine Al-Hamdani, Péter R. Nagy, Alexandre Tkatchenko, and Sameer Varma: Improved description of ligand polarization enhances transferability of ion–ligand interactions, *J. Chem. Phys.* **153**, 094115 (2020).

43. Vered Wineman-Fisher, Julián Meléndez Delgado, Péter R. Nagy, Eric Jakobsson, Sagar A. Pandit, and Sameer Varma: Transferable interactions of Li^+ and Mg^{2+} ions in polarizable models, *J. Chem. Phys.* **153**, 104113 (2020).

44. Sanim Rahman, Vered Wineman-Fisher, Péter R. Nagy, Yasmine Al-Hamdani, Alexandre Tkatchenko, and Sameer Varma: Methyl-induced polarization destabilizes the noncovalent interactions of N-methylated lysines, *Chem. Eur. J.* **27**, 11005 (2021).

45. Anna Ott, Péter R. Nagy, and Zoltán Benkő: Stability of carbocyclic phosphinyl radicals: Effect of ring size, delocalization, and sterics, *Inorg. Chem.* **61**, 16266 (2022).

46. Benjamin Shi, Andrea Zen, Venkat Kapil, Péter R. Nagy, Andreas Grüneis, and Angelos Michaelides: A coming of age for many-body methods: Achieving consensus with experiments for CO on MgO, *J. Am. Chem. Soc.* **145**, 25372 (2023).

47. Julian M. Delgado, Péter R. Nagy, and Sameer Varma: Polarizable AMOEBA model for simulating Mg^{2+}·protein·nucleotide complexes, *J. Chem. Inf. Model.* **64**, 378 (2024).

48. Balázs Kirschweng, Dániel M. Tilinger, Bence Hégely, Gyula Samu, Dóra Tátraaljai, Enikő Földes, and Béla Pukánszky: Melt stabilization of PE with natural antioxidants: Comparison of rutin and quercetin, *Eur. Polym. J.* **103**, 228 (2018).

49. Levente Kárpáti, Ádám Ganyecz, Tibor Nagy, Gergely Hamar, Eszter Banka, Mihály Kállay, and Viktória Vargha: Synthesis and characterisation of isphorondiamine based oligoamides: Catalytic effect of amides during the curing of epoxy resins, *Polym. Bull.* **77**, 4655 (2019).

50. Adrien Paudics, Dóra Hessz, Márton Bojtár, Benjámin Gyarmati, András Szilágyi, Mihály Kállay, István Bitter, and Miklós Kubinyi: Binding modes of a phenylpyridinium styryl fluorescent dye with cucurbiturils, *Molecules* **25**, 5111 (2020).

51. Zoltán Szakács, Márton Bojtár, Dóra Hessz, Sylvia Rousseva, István Bitter, László Drahos, Michiel Hilbers, Hong Zhang, Mihály Kállay, and Miklós Kubinyi: Strong ion pair charge transfer interaction of 1,8-naphthalimide-bipyridinium conjugates with basic anions—Towards the development of a new type of turn-on fluorescent anion sensor, *New J. Chem.* **43**, 6666 (2019).

52. Márton Bojtár, Krisztina Németh, Farkas Domahidy, Gergely Knorr, András Verkman, Mihály Kállay, and Péter Kele: Conditionally activatable visible-light photocages, *J. Am. Chem. Soc.* **142**, 15164 (2020).

53. Adrien Paudics, Dóra Hessz, Márton Bojtár, István Bitter, Viola Horváth, Mihály Kállay, and Miklós Kubinyi: A pillararene-based indicator displacement assay for the fluorescence detection of vitamin B1, *Sens. Actuators B Chem.* **369**, 132364 (2022).

54. Alexandra Egyed, Krisztina Németh, Tibor Á. Molnár, Mihály Kállay, Péter Kele, and Márton Bojtár: Turning red without feeling embarrassed—Xanthenium-based photocages for red-light activated phototherapeutics, *J. Am. Chem. Soc.* **145**, 4026 (2023).

55. Dóra Hessz, Etelka Kiss, Márton Bojtár, Attila Kunfi, Dávid Mester, Mihály Kállay, and Miklós Kubinyi: Photochemistry of a water-soluble coumarin-based photoswitch, *Dyes Pigm.* **221**, 111821 (2024).

56. Mihály Kállay, Péter R. Nagy, Dávid Mester, Zoltán Rolik, Gyula Samu, József Csontos, József Csóka, P. Bernát Szabó, László Gyevi-Nagy, Bence Hégely, István Ladjánszki, Lóránt Szegedy, Bence Ladóczki, Klára Petrov, Máté Farkas, Pál D. Mezei, and Ádám Ganyecz: The MRCC program system: Accurate quantum chemistry from water to proteins, *J. Chem. Phys.* **152**, 074107 (2020).