



SOFTWARE FOCUS

VeloxChem: A Python-driven density-functional theory program for spectroscopy simulations in high-performance computing environments

Zilvinas Rinkevicius^{1,2} | Xin Li¹ | Olav Vahtras¹ | Karan Ahmadzadeh¹ |
Manuel Brand¹ | Magnus Ringholm¹ | Nanna Holmgaard List^{3,4} |
Maximilian Scheurer⁵ | Mikael Scott⁵ | Andreas Dreuw⁵ | Patrick Norman¹

¹Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden

²Department of Physics, Kaunas University of Technology, Kaunas, Lithuania

³Department of Chemistry and the PULSE Institute, Stanford University, Stanford, California

⁴SLAC National Accelerator Laboratory, Menlo Park, California

⁵Interdisciplinary Center for Scientific Computing, Ruprecht-Karls University, Heidelberg, Germany

Correspondence

Patrick Norman, Department of Theoretical Chemistry and Biology, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, SE-106 91 Stockholm, Sweden.
Email: panor@kth.se

Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: COSINE; Norges Forskningsråd, Grant/Award Number: 274918; Vetenskapsrådet, Grant/Award Number: 2018-4343; Villum Fonden, Grant/Award Number: VKR023371; Swedish National Infrastructure for Computing (SNIC); Swedish e-Science Research Centre (SeRC)

Abstract

An open-source program named VeloxChem has been developed for the calculation of electronic real and complex linear response functions at the levels of Hartree–Fock and Kohn–Sham density functional theories. With an object-oriented program structure written in a Python/C++ layered fashion, VeloxChem enables time-efficient prototyping of novel scientific approaches without sacrificing computational efficiency, so that molecular systems involving up to and beyond 500 second-row atoms (or some 10,000 contracted and in part diffuse Gaussian basis functions) can be routinely addressed. In addition, VeloxChem is equipped with a polarizable embedding scheme for the treatment of the classical electrostatic interactions with an environment that in turn is modeled by atomic site charges and polarizabilities. The underlying hybrid message passing interface (MPI)/open multiprocessing (OpenMP) parallelization scheme makes VeloxChem suitable for execution in high-performance computing cluster environments, showing even slightly beyond linear scaling for the Fock matrix construction with use of up to 16,384 central processing unit (CPU) cores. An efficient—with respect to convergence rate and overall computational cost—multifrequency/gradient complex linear response equation solver enables calculations not only of conventional spectra, such as visible/ultraviolet/X-ray electronic absorption and circular dichroism spectra, but also time-resolved linear response signals as due to ultra-short weak laser pulses. VeloxChem distributed under the GNU Lesser General Public License version 2.1 (LGPLv2.1) license and made available for download from the homepage <https://veloxchem.org>.

This article is categorized under:

Software > Quantum Chemistry
Electronic Structure Theory > Density Functional Theory
Theoretical and Physical Chemistry > Spectroscopy

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *WIREs Computational Molecular Science* published by Wiley Periodicals, Inc.

KEYWORDS

circular dichroism, density functional theory (DFT), ECD, high-performance computing (HPC), MPI, OpenMP, response theory, UV/vis

1 | INTRODUCTION

Theoretical chemistry has made tremendous progress in the past few decades and is today an indispensable tool in all fields of molecular science, exemplified by biochemistry and nanotechnology, where it can be employed to reveal the microscopic origins of functionality and interactions as well as to tune performance and guide synthesis. The interplay between simulation and experiment is a key enabling factor for advancements in life and material sciences and spectroscopy represents one of the most important meeting points in between the two disciplines.¹ For instance, electronic circular dichroism spectroscopy is a sensitive probe of both molecular and electronic structures, and reaching spectral agreement between theory and experiment for this spectroscopy may therefore be taken as strong evidence for having correctly elucidated complex intermolecular interactions.² It is imperative to base spectroscopic modeling on first principles as the underlying microscopic events are quantum mechanical (QM) in origin—although³ typically in combination with classical force-field molecular dynamics for the sampling of configuration space—and in order to study complex molecular systems and delocalized electronic transitions, it is therefore important that such first-principles approaches are developed into forms where they may be tractably applied to large systems. It is therefore not surprising that a vast amount of work in computational chemistry has been carried out to extend the computationally tractable system domain boundaries with respect to time and length scales.⁴ For example, the field of quantum chemistry has witnessed several efforts to design algorithms that demonstrate linear scaling with respect to the number of basis functions.⁵

The attainment of linear scaling for methods that inherently scale as N^4 (Hartree–Fock, HF, and Kohn–Sham density functional theory, DFT) or worse (post-HF) does not come without sacrifices and requires spatial locality in some form. This locality can be found naturally in sparse systems, for example, linear polymers, or accomplished with orbital localization schemes and the zeroing out of small numerical integral values. Basis sets based solely on valence-occupied atomic orbitals (AOs) and the use of pure functionals without exact exchange are also enabling factors to reach linear scaling for reasonable-sized systems. This is, however, a poor fit for our goal, which is to model spectroscopic responses for three-dimensional dense systems where diffuse basis functions and the inclusion of exact HF exchange are necessities. Furthermore, as the key quantities in the simulation of spectroscopies are response functions that correspond to sensitive derivatives of the energy with respect to the field strengths of the weak perturbative external electromagnetic fields, there is little room for discarding integrals as it will lead to numerical instabilities—particularly so for higher order interactions with the external fields. A comprehensive study of the parallelization efficiency of the Fock matrix construction has been performed by Chow et al.⁶ and, in line with our interest, they focus on sizable systems below the regime of linear scaling and introduce density-matrix purification as alternative to conventional eigendecomposition approaches. Valuing numerical stability strongly, we implement for the most part conventional algorithms formulated in the molecular orbital (MO) basis, which means that, in the self-consistent field (SCF) optimization, we diagonalize the Fock matrix with an accompanying (low pre-factor) N^3 scaling. In the present version, this diagonalization is performed on a single node with use of a standard threaded math kernel library.

All response properties are expressed in terms of auxiliary Fock matrices, which centralizes the computational cost around the contraction of two-electron integrals with the corresponding auxiliary density matrices. A signum of the VeloxChem program is that it processes efficiently and in parallel a very large number of Fock matrices in the sense that they are *all* created from a *single* evaluation of the set of two-electron integrals. The default energy-gradient and integral-screening thresholds are set to 10^{-6} and 10^{-12} a.u., respectively, which, in our experience, allows for accurate determination of at least up to third-order (i.e., cubic) response functions.

The intended functionality of VeloxChem is that a user with access to moderately powerful contemporary high-performance computing (HPC) cluster resources should be able to carry out spectroscopic simulations in a routine manner for systems with up to and beyond 500 second-row atoms (or some 10,000 Gaussian basis functions with in part diffuse character) in the quantum region and a large polarizable embedding (PE) region (described by distributed

multipole moments and polarizabilities). The parallel processing of Fock matrices is enabled by a very small memory footprint of the integral code, leaving the predominant part of the node memory available for storing auxiliary density and Fock matrices as well as one-particle–one-hole (1p1h) vectors generated in the various iterative reduced-space response equation solvers. The rest of this article is organized as follows: After a brief presentation of the program structure, implementation, and performance, we show three spectroscopic simulation examples to demonstrate aspects of the program capabilities.

2 | PROGRAM STRUCTURE AND MODULARITY

The VeloxChem program is developed with a two-layer structure as illustrated in Figure 1: The upper layer is written in Python and fully handles HPC resource management, workflow control, and input/output streams. The communication of data across HPC cluster nodes is performed with the message passing interface (MPI) standard in the form of the MPI4Py package,⁷ and MPI split communicators are used to achieve cluster resource subdivision. On individual cluster nodes, parallelization is achieved by (a) use of the NumPy⁸ and SciPy⁹ packages in the Python layer linked to a threaded math kernel library and (b) task-based open multiprocessing (OpenMP) programming in the lower C++ layer as to overall result in a hybrid parallel programming model. Similar program language structuring efforts have recently been undertaken and served as inspiration for the present work, most notably (a) the PySCF program platform¹⁰ for Python-based SCF and post-Hartree–Fock (post-HF) electronic structure theory calculations of finite and periodic systems and (b) the Psi4NumPy¹¹ a Python programming environment for facilitating use of the SCF and post-HF kernels from the Psi4 program is obtained through the NumPy library.

The Python layer also contains high-level implementations of the quantum chemical methods that perform the SCF optimization of the electron density and the associated response function calculations needed for the simulation of spectroscopies. In each SCF iteration, a density matrix is formed in the AO basis and exposed to the C++ layer by means of the header-only library Pybind11.¹² In this lower level and based on this density matrix, the corresponding Fock (or Kohn–Sham) matrix is evaluated with contributions from Coulomb and exchange interactions and also, if requested, the PE potential. The numerical DFT-kernel integration is based on a set of spatial grid points and associated weights that are generated once in the beginning of the calculation and stored in memory for repeated use.

The secular and generalized eigenvalue equations arising in linear response theory are solved iteratively and numerically based on reduced-space algorithms residing in the Python layer. The key element in these algorithms consists of the formation of multiple auxiliary Fock matrices based on the contraction of two-electron integrals with the associated density matrices. From a computational point of view, this step is closely analogous to the formation of a single Fock matrix in any given SCF iteration. The only modification that is made is to allow for a treatment of multiple Fock matrices in parallel by having an entire batch of Fock matrices be constructed from a single evaluation of the set of two-electron integrals.

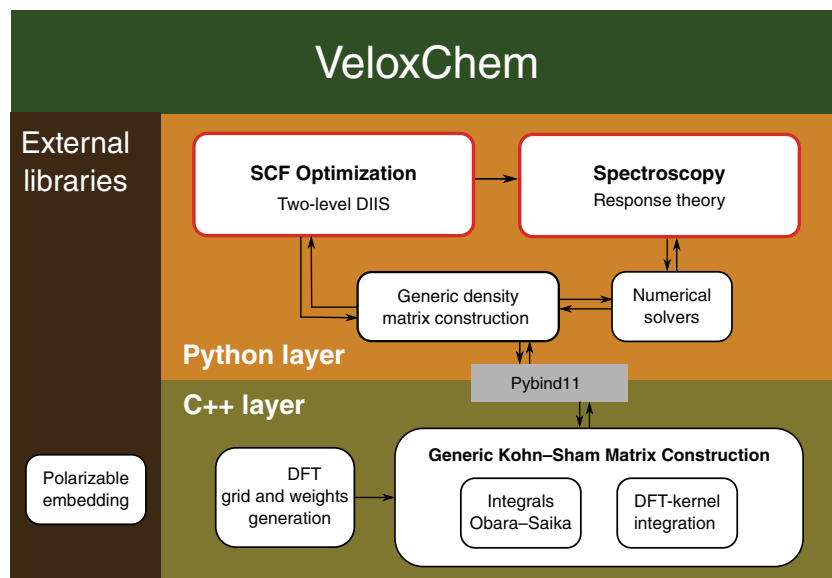


FIGURE 1 VeloxChem program workflow and module overview

The VeloxChem program design is strictly modular with enforced unit testing. This enables easy interaction with external libraries as exemplified by the PE model to incorporate environment effects into QM systems. This PE functionality is provided by the external CPPE library^{13,14} as schematically illustrated in Figure 1. Moreover, the situation can also be reversed, that is, VeloxChem can just as easily serve as an external library in another software package. This aspect is demonstrated in the Gator program, where VeloxChem is the provider of Hartree–Fock reference states and integrals.¹⁵

3 | IMPLEMENTATION AND EFFICIENCY

3.1 | Evaluation of integrals for general-contracted Gaussian basis sets

The evaluation of electron-repulsion integrals (as well as one-electron integrals) in VeloxChem is based on a modified version of the Obara–Saika scheme^{16,17} in which vertical and horizontal recursion steps are combined to increase the efficiency of the scheme on the target processing architectures. Our integral program module implements a workflow as follows:

1. Generation of basic auxiliary primitive integrals $[00|00]^m$ up to m -th order as determined by the sum of angular momenta on all four centers, that is, $m = a + b + c + d$.
2. Build up of primitive integrals $[00|0(c+d)]$ and $[0(a+b)|0(c+d)]$ using vertical Obara–Saika recursion.
3. Generation of contracted Cartesian integrals $(0(a+b)|0(c+d))$ by contracting primitive integrals $[0(a+b)|0(c+d)]$.
4. Build up of contracted integrals $(0(a+b)|cd)$ by applying horizontal angular momentum transfer recursion to primitive integrals $(0(a+b)|0(c+d))$ on the ket-side.
5. Generation of half-transformed contracted integrals $(0(a+b)|(cd)^{sp})$ from contracted integrals $(0(a+b)|cd)$ by applying Cartesian to spherical transformation on the ket-side.
6. Build up of contracted integrals $(ab|(cd)^{sp})$ by applying horizontal angular momentum transfer recursion to contracted integrals $(0(a+b)|(cd)^{sp})$ on the bra-side.
7. Generation of contracted spherical integrals $((ab)^{sp}|(cd)^{sp})$ from half-transformed contracted integrals $(ab|(cd)^{sp})$ by applying Cartesian to spherical transformation on the bra-side.

The VeloxChem implementation of this workflow is based on automated code generation with a high degree of loop structure optimization in order to achieve efficient vectorization on CPUs that implement single instruction multiple data (SIMD) instructions, and the obvious goal of this vectorization is to achieve efficient usage of the vector registers on modern CPUs. In Figure 2, we present a performance profile of the vectorization for the vertical recursion step corresponding to Step 2 in the above integral workflow scheme. For basis set primitives up to angular momentum $l = 5$ (or h -functions), the Intel-compiler reported efficiencies depicted in Figure 2 exceed a speedup of 4.75 for the vectorized code as compared to the scalar code.

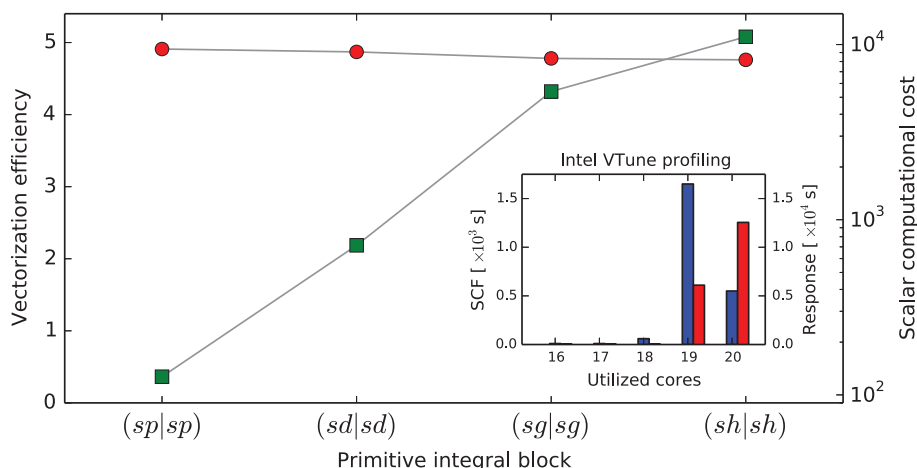


FIGURE 2 Single instruction multiple data (SIMD) vectorization for primitive integral blocks with different angular momenta (red circles) and the corresponding scalar computational costs (green squares) to be understood in a relative sense. Performance profiling results in the inset are obtained for the C_{60} fullerene at the level of Hartree–Fock/def2-SVP and recorded with the performance profiler Intel VTune Amplifier running on a dual-socket (20-core) Intel Xeon node that implements AVX2

The OpenMP parallelization on a single node is benchmarked with use of the Intel VTune Amplifier tool that measures the amount of time spent at different levels of CPU core activity. Thus, the range of activity spans from zero (fully idling CPUs) to the maximum number of cores on the node (in our case 20). The inset in Figure 2 depicts the reported results for the *complete* SCF optimization of the Hartree–Fock ground state of the fullerene C₆₀ as well as the subsequent *complete* response calculation to obtain the five lowest roots of the generalized eigenvalue equation, that is, the five lowest singlet states in the ultraviolet (UV)/visible (vis) spectrum. The overall total OpenMP threading efficiency is seen to reach levels of some 96 and 98% for the SCF (blue bars) and response parts (red bars), respectively.

3.2 | DFT grid/weights generation and kernel integration

Before the SCF optimization begins, a set of grid points and the associated grid-point weights connected with the AO basis set are constructed using modified versions of the Log3-quadrature¹⁸ and SSF-partitioning schemes.¹⁹ This initialization needed for DFT calculations is performed on a single node using OpenMP parallelization. The size of the grid determines the accuracy reached in the numerical integration and it is set by the user in terms of a *grid level* number between 1 and 6 (the program default level is 4). The grid-point data is divided equally in between the nodes that are set aside with an MPI split communicator to handle the kernel integration—meaning that the DFT-kernel integration uses its dedicated part of the cluster resources and is carried out aside and in parallel with the more resource intensive calculation of Fock matrix contributions due to Coulomb and exact exchange interactions (ERIs). The distribution of grid-point data segments is carried out with the operation of MPI scattering.

In Figure 3, we provide an illustration of the prototypical scaling of the kernel integration with respect to the number of employed cores. This example refers to the kernel integration of the ground state density of a medium-sized molecule, zinc porphyrin, with use of a basis set involving 1,352 contracted and 1,986 primitive functions. We do, however, not expect the principal behavior of this scaling to depend strongly on the system size.

Using the default grid level, the Zn-porphyrin derivative spawns a bit more than 1,100,000 grid points. The required wall time in the calculation to generate these grid points and weights is 2 s and the time needed for their scattering across nodes is in comparison insignificant. Employing 32 cores (1 node) the actual kernel integration takes 594 s on this particular CPU that dates back 5 years in time—a result that refers to the reference data point in Figure 3. Up to 256 cores (eight nodes), the scaling is seen to be nearly linear, pushing down the wall time to 81 s at the point of using of 256 cores (or about 1 core per 4,000 grid points). Doubling the number of cores to reach 256 (16 nodes) results in a reduced scaling and the corresponding wall time becomes equal to 47 s. It is in practice not very important to further push down the wall time of the kernel integration as this part of the calculations should be balanced against that of the analytic calculation of ERIs.

Our implementation of the kernel integrator takes advantage of SIMD instructions for vectorization, allowing for rapid evaluation of exchange–correlation functional derivatives and electron densities on the molecular grid. To enable SIMD instructions in the kernel integration, the electron density evaluated on the molecular grid is partitioned into three sets namely (1) $\rho_{\alpha}(\mathbf{r}) > 0$

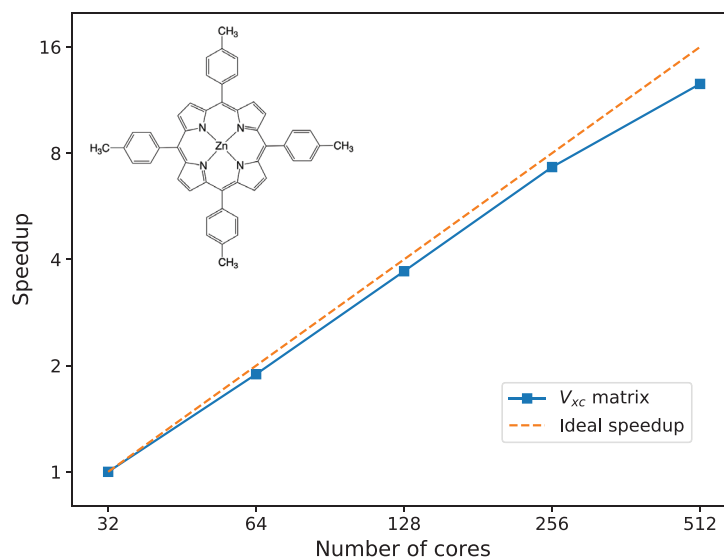


FIGURE 3 Speedup of the density functional theory (DFT)-kernel integration with respect to number of cores and with reference to the 32-core (1 node) case. Results are obtained for a zinc porphyrin derivative at the level of B3LYP/def2-SVPD (without zinc *f*-functions) and using the program default grid level 4 to spawn 1,111,721 grid points

and $\rho_\beta(\mathbf{r}) > 0$; (2) $\rho_\alpha(\mathbf{r}) > 0$ and $\rho_\beta(\mathbf{r}) = 0$; (3) $\rho_\alpha(\mathbf{r}) = 0$ and $\rho_\beta(\mathbf{r}) > 0$, and different forms of the functional derivative code are developed and used for each of these sets. This approach to kernel integration not only enables efficient usage of SIMD instructions but also ensures numerical stability in the integration process and provides a user-defined control of accuracy.

3.3 | Two-level direct inversion of the iterative subspace (DIIS)

VeloxChem implements the C2DIIS algorithm²⁰ with an initial guess based on a superposition of atomic densities followed by a small number (around five) of SCF iterations. The formation of the initial guess is performed in a minimal basis set derived from the user-specified basis set by removal of polarization and diffuse functions. As the adopted minimal basis set is a subset of the user-specified one, we relieve the need for a projection of the initial density from one basis set to the other before starting the *real* SCF optimization. The complete formation of this initial density has a computational cost that is close to one regular SCF iteration.

To illustrate the performance of the SCF scheme in VeloxChem, we use a large organic system with 834 atoms forming a three-armed linked π -conjugated core with extended saturated alkyl chains. In Figure 4, we present electronic energies and gradients during the SCF optimization. The most noteworthy aspect is the fact that the Hartree–Fock energy of the initial guess is a mere 0.3 Hartree higher than the final converged value. To place the initial density in such close proximity to the final one means that there are no observed oscillations in the electronic energy, which is otherwise a typical and undesired behavior in DIIS-based electronic structure optimizations. The illustrated representative SCF convergence is monotonic and takes place in 26 iterations with an employment of a tight gradient threshold norm of 10^{-6} a.u. that in turn enables accurate determination of linear and nonlinear response functions up to at least third order, or cubic response functions.

3.4 | Multifrequency/gradient linear response equation solver

At the single-determinant Hartree–Fock and Kohn–Sham levels of theory, the linear response equation takes the form³

$$\left[\mathbf{E}^{[2]} - (\omega + i\gamma)\mathbf{S}^{[2]} \right] (\mathbf{X}^R + i\mathbf{X}^I) = \mathbf{G}^R + i\mathbf{G}^I, \quad (1)$$

where $\mathbf{E}^{[2]}$ and $\mathbf{S}^{[2]}$ are the real-valued electronic Hessian and overlap matrices, respectively, \mathbf{G} is the complex property gradient, ω is the optical angular frequency of the perturbing electromagnetic field, and γ is a relaxation parameter associated with the inverse lifetime of the excited states. The equation is solved for the unknown associated complex response vector, $\mathbf{X} = \mathbf{X}^R + i\mathbf{X}^I$, that determines the first-order responses in the reference state and from which linear

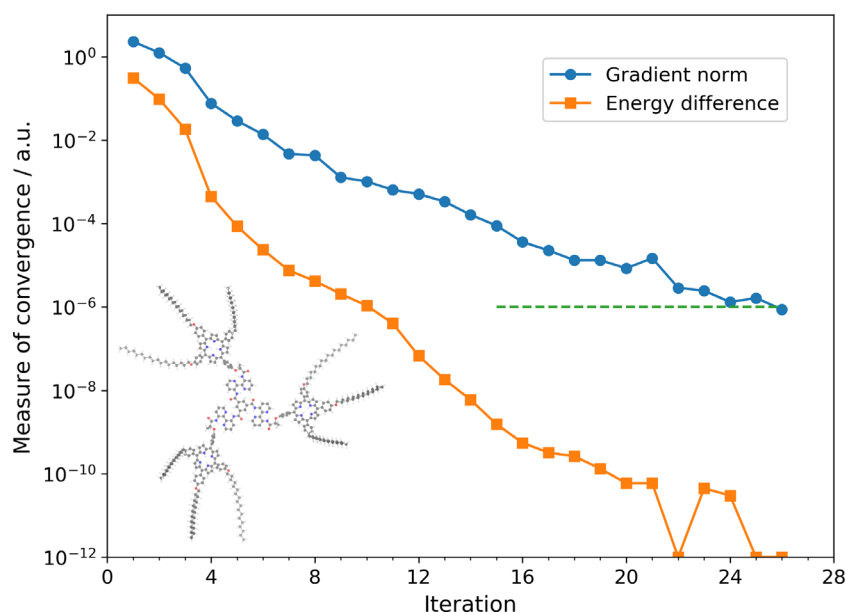


FIGURE 4 Representative SCF convergence characteristics of the two-level direct inversion of the iterative subspace (DIIS) scheme implemented in VeloxChem. Results are presented in atomic units and obtained at the Hartree–Fock/def2-SVP level of theory for a three-armed organic system with 834 atoms, resulting in 7,626 contracted and 12,366 primitive basis functions

and first-order nonlinear response properties can be determined. It has been demonstrated that Equation (1) can be effectively solved by means of an iterative reduced-space algorithm for a large set of optical frequencies discretizing a continuous spectral region of interest.²¹ An extension of this algorithm has been implemented in VeloxChem as to handle not only multiple frequencies in parallel but also multiple right-hand sides (or gradients) in the linear response equation. Parallel here refers to solving all equations with use of a common subspace and a single evaluation of the entire set of two-electron integrals per iteration in the subspace algorithm. If the perturbing electromagnetic fields are treated in the dipole approximation, there are, respectively, three Cartesian electric- and magnetic-dipole operators to consider in Equation (1). The benefits of a parallel handling of gradients becomes even more striking in going beyond the electric-dipole approximation when the perturbation operators are frequency dependent.²²

The convergence characteristics of the multifrequency/gradient linear response equation solver is illustrated in Figure 5 for the case of a calculation of the UV/vis and electronic circular dichroism spectra of noradrenaline. We choose to solve Equation (1) for the set of electric-dipole operators and determine the absorptive parts of the electric and mixed electric–magnetic linear response functions, that is, $\text{Im}\langle\langle\hat{\mu};\hat{\mu}\rangle\rangle_{\omega}$ to obtain the absorption cross section $\sigma(\omega)$ ³ and $\text{Re}\langle\langle\hat{m};\hat{\mu}\rangle\rangle_{\omega}$ to obtain the anisotropy of the decadic molar extinction coefficient $\Delta\epsilon(\omega)$.²³ The former type of response functions is here evaluated in the length gauge whereas the latter is evaluated in the velocity gauge as to guarantee that the trace of the mixed electric–magnetic polarizability tensor is gauge-origin invariant. In figure panel a, we show the convergence behavior when solving Equation (1) for the three electric-dipole operators (length gauge) and a single frequency in resonance with the absorption band at $\omega = 7.3749$ eV. The three equations are solved in parallel and converge smoothly in 20 iterations. To obtain the electronic absorption and circular dichroism (ECD) spectra, the frequency region from 0 to 13.5 eV is uniformly divided by 200 frequency points with a resulting 600 linear response equations to solve. As shown in figure panel b, all 600 equations are converged in, at the most, 11 iterations. Equations involving frequencies up 9.3 eV are all converged in, at the most, seven iterations (results for the remaining equations are shown in rainbow colors) and the three equations depicted in figure panel a are converged in as few as six iterations (shown in red, green, and blue). This clearly demonstrates the convergence benefit of solving linear response equations in parallel in a single common reduced space rather than in serial and, in addition, there is an obvious computational cost benefit.

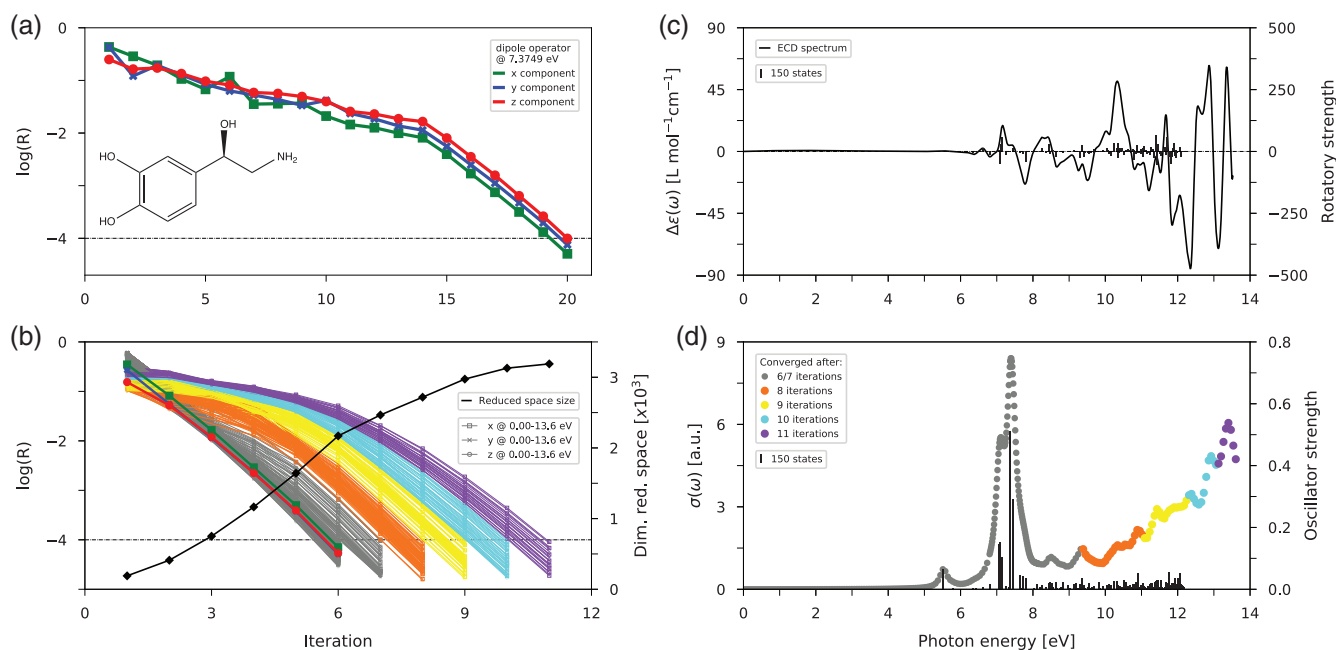


FIGURE 5 Convergence characteristics of linear complex response equation solver. (a) Residuals of convergence (a.u., logarithmic scale) for the three electric-dipole operators for a single resonant frequency, $\hbar\omega = 7.3749$ eV; (b) same as (a) but with parallel handling of 200 optical frequencies and, in addition, a shown dimension of the reduced space; (c) resulting electronic circular dichroism spectrum with rotatory strengths depicted by bars and reported in units of 10^{-40} esu²cm², and (d) spectrum with oscillator strengths depicted by bars. Results are obtained for noradrenaline at the level of Hartree–Fock/aug-cc-pVDZ with a damping term $\gamma = 1,000$ cm⁻¹ and adopting a molecular structure optimized at the level of B3LYP/cc-pVTZ

3.5 | Multicore MPI/OpenMP scaling

A key VeloxChem design objective is to target execution on contemporary and future HPC hardware platforms and to make efficient use of parallelism and (less obvious) the aggregated and distributed memory available in multinode execution runs. As we believe it to be beneficial and productive from a user's perspective, we have decided to implement conventional SCF and response schemes in the MO basis with an emphasis on numerical stability and we aim to push the program performance only up to the point of where the strong N^3 -scaling of the Fock matrix diagonalization takes over and dominates the overall cost.

In Figure 6, we address a titanium oxide nanoparticle that illustrates the size (but not necessarily type) of system that we routinely wish to encompass in the quantum region of a given physical model. With generous and exclusive access to a large part of the main cluster at PDC Center for High Performance Computing in Stockholm, we had the opportunity to test the MPI/OpenMP scaling of VeloxChem up to 16,384 cores (or 512 dual socket 32-core Intel Xeon nodes) and we use the otherwise identical calculation performed on 32 cores (or 1 nodes) as benchmark reference. A *complete* SCF cycle during the optimization of the ground state density requires a wall time of 70,565 s in the reference calculation and out of which 70,493 s and 72 s are spent on the construction of the Fock matrix and other parts (including the Fock matrix diagonalization), respectively. Apart from the Fock matrix construction, the SCF module is written in Python (Figure 1) and executes on a single node and the corresponding timing is therefore nearly constant along the series of calculations presented in Figure 6.

Owing to favorable differences in the dynamic workload balancing between the available MPI processes and OpenMP threads, the Fock matrix construction scales slightly beyond linear with respect to the number of cores (up to a point of 384 nodes) and reaches an efficiency of 162.9% at the point of using 64 nodes and compared to the reference calculation. This example thus demonstrates almost perfect scaling up to a point of allocating almost one CPU core per primitive basis function. The wall time for the Fock matrix construction using 512 nodes is reduced to 149 s, so while it may be possible to observe a reasonable scaling beyond 512 nodes in the present example, it makes little sense in practice as the SCF/DIIS part becomes comparable in duration.

As a final remark in the context of program scaling, we note that the reported wall time of 72 s for the SCF cycle excluding the Fock matrix construction is still not representing a practical program bottleneck per se. With cubic scaling of the Fock matrix diagonalization, it is realistic to remain within 10 min of wall time for this part of the SCF cycle with twice the number of basis functions (or double the system size) as compared to the present example.

4 | CAPABILITIES

Release version 1.0 of VeloxChem encompasses response theory simulations of general linear (one-photon) electronic spectroscopies at the level of Hartree–Fock and Kohn–Sham DFT employing standard localized Gaussian basis sets ranging up to angular momentum $l = 4$ (g-functions). In the case of DFT, a set of standard exchange–correlation

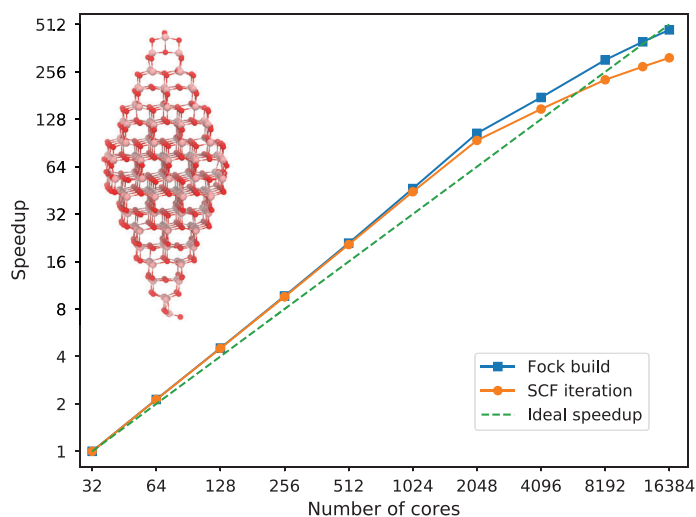


FIGURE 6 Scaling with respect to number of cores for the optimization of the ground state wave function. Speedups for one complete SCF cycle (orange circle) and the isolated associated Fock matrix construction (blue square) are reported separately and with reference to the calculation using 32 cores. Results are obtained for a titanium oxide nanoparticle, $\text{Ti}_{165}\text{O}_{330}$, at the level of Hartree–Fock/def2-SV(P), resulting in 8,580 contracted and 18,810 primitive basis functions

functionals are provided in the categories of pure, generalized gradient approximation, hybrid, long-range corrected, and short-range corrected.

Real and complex linear response functions are implemented to describe dispersive and absorptive electronic spectroscopies in the entire region of photon frequencies—in practical applications, however, the regions of main interest are UV/vis and soft X-ray as neither geometric derivatives nor relativistic effects are considered. The complex response function technique is also known as the complex polarization propagator (CPP) approach.²⁴

For real linear response functions, calculations of residues and poles are offered to enable the calculation of ground-to-excited state transition moments and the corresponding vertical transition energies. These properties are provided based on solutions to the secular equation in the full random phase approximation (RPA) as well as the more limited Tamm–Dancoff approximation (TDA). The former of the two approaches is also known under the names of time-dependent Hartree–Fock (TDHF) and time-dependent DFT (TD-DFT), depending on the adopted level of electronic structure theory.

For complex linear response functions, calculations are enabled for “spectral windows” so that the requested response functions are determined for a dense set of optical real-valued frequencies in an interval covering the user-specified spectral region. In this type of spectrum calculation, a uniform imaginary damping parameter is assumed. In addition, it is also possible to request the calculation of response functions for a set of imaginary frequencies without further inclusion of damping terms as to determine dispersion interaction coefficients such as the C_6 dipole–dipole dispersion coefficients. The program evaluates the Casimir–Polder integral by means of the Gauss–Legendre quadrature scheme and a user-specified number of pretabulated weights and frequency points.²⁵

Description of a molecular environment is enabled by the self-consistent PE approach,^{26–28} which uses distributed electric multipole moments to represent the permanent electrostatic effects of the environment on the QM subsystem and distributed localized electric-dipole polarizability tensors to describe anisotropic polarization effects both internally in the environment and due to interactions with the QM subsystem. PE parameters can be obtained with use of an integrated implementation of the LoProp scheme^{29,30} and the corresponding potential is consistently integrated in the response theory module under the assumption that dispersion effects in the polarizabilities of the environment can be ignored.³¹ The PE functionality in VeloxChem is handled through the external CPPE library.^{13,14}

5 | NUMERICAL EXAMPLES

5.1 | Electronic circular dichroism of helicenes

It has been demonstrated that ECD in linear and nucleosomal forms of DNA can be simulated by means of summing spectra from base-pair dimers and thereby effectively transform the spectrum calculation of a large-scale system into a large number of calculations for moderately sized systems.³² Similarly, ECD spectra of very short peptides and unfolded oligo-peptides/proteins are in close agreement but, at the same time, the involvement of dynamic explicit water molecules in the electronic structures of accessible states makes the situation more complex and renders classical excitonic models unsuitable.³³ In the general case, however, systems cannot be subdivided while maintaining quantitative accuracy in ECD spectrum calculations. One extreme is represented by extended π -conjugated systems with delocalized excitons, in which case there is a need for a treatment of large QM regions. This is offered by VeloxChem and we illustrate this capability by studying the near-UV spectral convergence of helicenes with respect to increasing polymer chain length.

A comprehensive combined experimental and theoretical study has been performed on the series of carbo[n] helicenes, or CH[n], with the number of phenyl rings, n , ranging from 6 to 10.³⁴ Characterized by symmetry, the two lowest bands showing large Cotton effects were denoted as 1B_b and 1B_a , respectively. For (*P*)-(+)-CH[6], the 1B_b and 1B_a bands were observed at 320 nm (positive Cotton effect) and 250 nm (negative Cotton effect), respectively, or 3.87 and 4.96 eV. The reported theoretical spectra were obtained at the level of coupled cluster (CC2) and showed an excellent agreement with experiment for the whole series of systems.

In Figure 7, we report the ECD spectra at the level of Hartree–Fock for the series ranging from CH[6] to CH[30]. For the shortest member, CH[6], corresponding to one complete helix turn, the 1B_b and 1B_a bands are found at 4.6 and 5.4 eV, respectively. The signs of the Cotton bands are correctly predicted but the transition energies are strongly overestimated as may be expected at the level of Hartree–Fock. Completing a second helix turn to reach CH[12], the 1B_b and 1B_a bands are found red shifted and now appear at transition energies of 3.8 and 4.5 eV, respectively.

Going beyond CH[12], we see band progressions that are less predictable. ECD spectra for the series CH[12]–CH[19] are shown in the inset of Figure 7. It is seen that the 1B_b band does neither grow stronger nor becomes further red

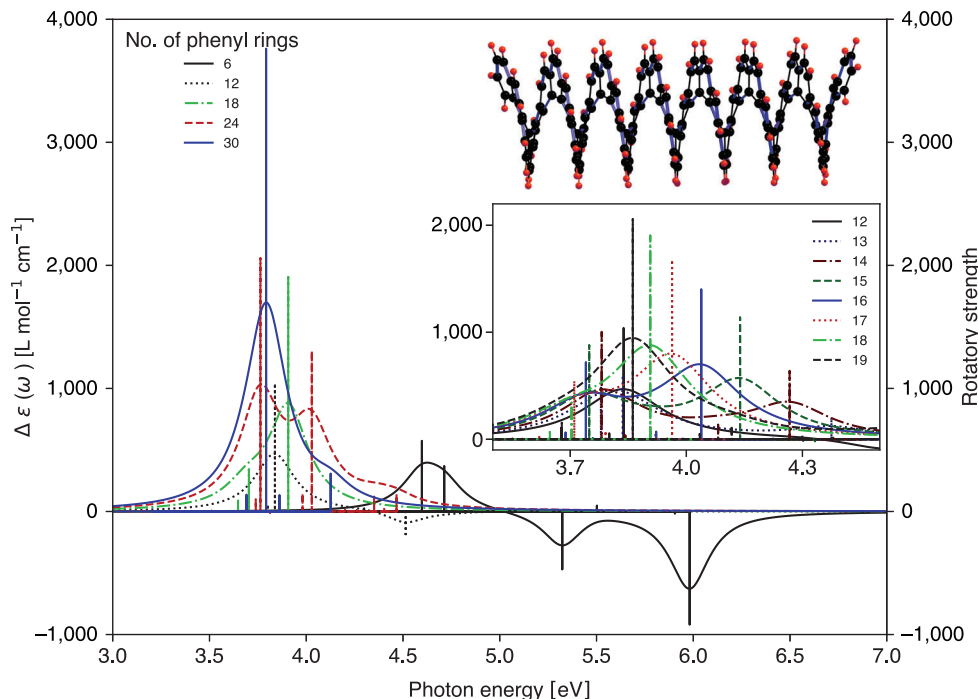


FIGURE 7 ECD spectra for the series of helicenes built from 6 to 30 phenyl rings. Results refer to the calculation of rotatory strengths at the level of Hartree–Fock/def2-SVPD for up to the 20 lowest excited states, depending on system size, as to include the first bright state. These rotatory strengths are broadened by a Lorentzian line profile with a half-width at half-maximum (HWHM) of 0.124 eV to produce the displayed ECD spectra. Structures are optimized with the density functional based tight binding DFTB approach

shifted with helicene length. Instead there appears a new band at 4.3 eV for CH[14] that rapidly gains in intensity and red shifts with helicene length. For CH[19], this new band is found at 3.9 eV and completely dominates the near-UV spectrum due to its strong intensity. This observation of irregular band progressions demonstrates that spectral extrapolations as made from CH[10] to the polymer limit should be done with a high degree of caution.³⁴ We are inclined to conclude that explicit calculations for longer systems as opposed to extrapolations become a necessary requirement to make accurate predictions of ECD spectra for the family of longer helicenes.

In spectral studies of large molecular systems, it is always difficult to guess beforehand the number of states that must be converged in the generalized eigenvalue equation in order to fully cover the spectral region of interest. The VeloxChem program addresses this issue by implementing a reduced-space eigenvalue solver with an efficient restart capability. The reduced-space information from a calculation made to converge a first set of n excited states can be fully retrieved in a second run to quickly converge an additional set of m excited states (in a bottom-up Davidson-solver sense).

5.2 | UV/Vis absorption and dichroism of a probe DAPI in an aqueous DNA PE

Quantitative simulations of spectroscopic response properties of chromophores in the condensed phase require the modeling not only of short-range quantum but also long-range electrostatic solute–solvent interactions. For the latter, VeloxChem implements a PE scheme that discretizes the environment into atomic sites with associated electric multipole moments and polarizabilities. We illustrate this program capability by the simulation of the absorption and circular dichroism spectra of the 4',6-diamidino-2-phenylindole (DAPI) biomarker inside the minor groove of DNA.

Our example calculation refers to a molecular dynamics snapshot configuration with embedding potential parameters (point charges and isotropic polarizabilities) taken from previous work.² The model system illustrated in Figure 8 is set up with (i) a single DAPI in the QM region, (ii) a PE region containing the DNA sequence, the three additional bound DAPIs, and a 15 Å shell of water molecules and a small number of sodium and chlorine ions (the entire system is charge neutral), and (iii) the remaining water molecules and ions in the surrounding point-charge (PC) region. In total, the model includes some 85,000 and 23,000 point-charge and polarizable sites, respectively.

The linear absorption and circular dichroism spectra of the model system are shown in Figure 8. In the original work, it was demonstrated that the experimental circular dichroism spectrum is due to an intricate interplay between several mechanisms namely (i) excitonic coupling between the four bound DAPIs, (ii) off-resonant excitonic coupling between DAPIs and nucleobases, (iii) charge transfer between probe and host, and (iv) chiral imprint of DNA onto the

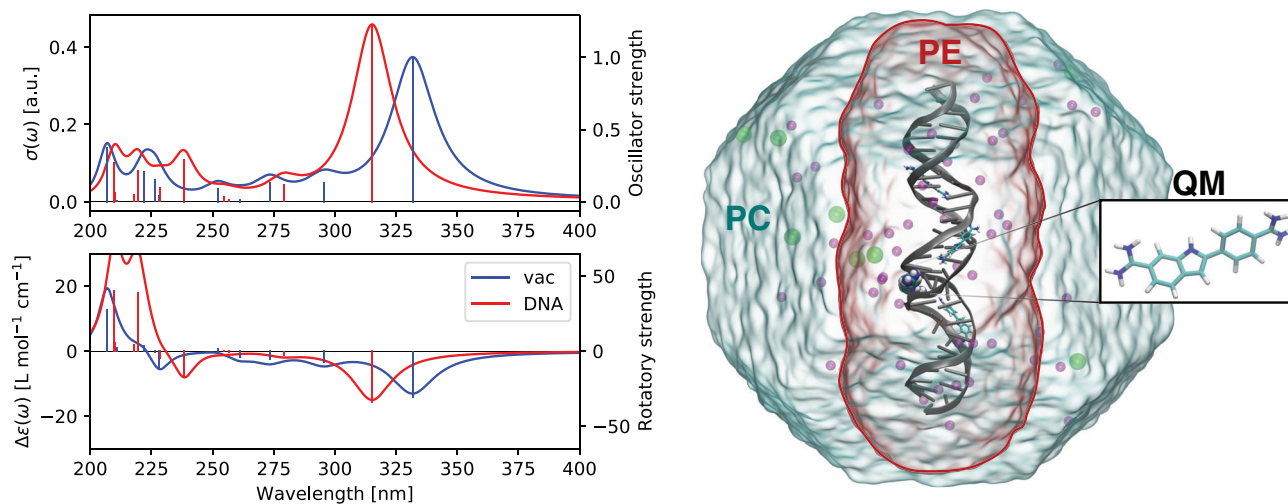


FIGURE 8 Absorption and ECD spectra for DAPI in vacuum and bound to DNA in aqueous solution. Results are obtained at the level of TDA/BHandHLYP/6-31+G(p,d). Oscillator strengths and rotatory strengths (in units of 10^{-40} esu²cm²) are broadened by a Lorentzian line profile with a half-width at half-maximum (HWHM) of 0.124 eV to produce the displayed spectra

electronic and molecular structures of DAPI. The ECD spectrum in Figure 8 is the result of contribution (iv). Modeling of contributions (ii) and (iii) to the ECD spectrum requires extended QM regions with inclusion made of nearby nucleobases and coordinating water molecules. Modeling of contribution (i) can be achieved by either extended QM calculations including all four probes or by means of the excitonic coupling model (ECM). In the present work, it has been amply demonstrated that VeloxChem is well suited for spectral calculations of extended systems and it also implements the ECM as described previously in the literature,³⁵ so all four contributions to the ECD spectrum of this complex biochemical system can readily be simulated with the current version of the program.

5.3 | Linear responses to pulsed electric fields

The present version of the VeloxChem program can be used to calculate the first-order time-domain response resulting from the interaction of a system with a pulsed electric field, such as a Gaussian pulse. The spectral envelope of a field of this form will also be Gaussian and centered around the carrier frequency of the pulse, confining the spectral components of the resulting time-domain signal to the regions where the spectral envelope of the field is appreciably strong. We demonstrate this functionality in Figure 9 (see caption for details), which shows the calculated first-order correction to the dipole moment of noradrenaline in the time domain upon interaction with a Gaussian-shaped pulse. This is accomplished by the use of Fourier transformation of the frequency-domain response function weighted by the spectral profile of the pulse (Figure 9b).

Observing the time-domain signal, we note an initial increase in the signal strength in synchronicity with the strength of the applied field. The signal later resolves into a weaker signal pattern originating from interference between induced oscillations of differing frequencies, where the resonant feature around 7.5 eV is assumed to contribute the strongest.

A more detailed description of this functionality, including the incorporation of pulse-shape effects for nonlinear response, will be presented in a future work and is a natural complement to the planned development of nonlinear response functionality in VeloxChem.

6 | SUMMARY AND OUTLOOK

The VeloxChem 1.0 program release offers a versatile tool for the simulation of spectroscopies described by linear response theory. It efficiently implements SCF theory, that is, Hartree–Fock and Kohn–Sham DFT, for targeted execution on HPC cluster resources and by means of the standard hybrid OpenMP/MPI programming model. At the same time, it also

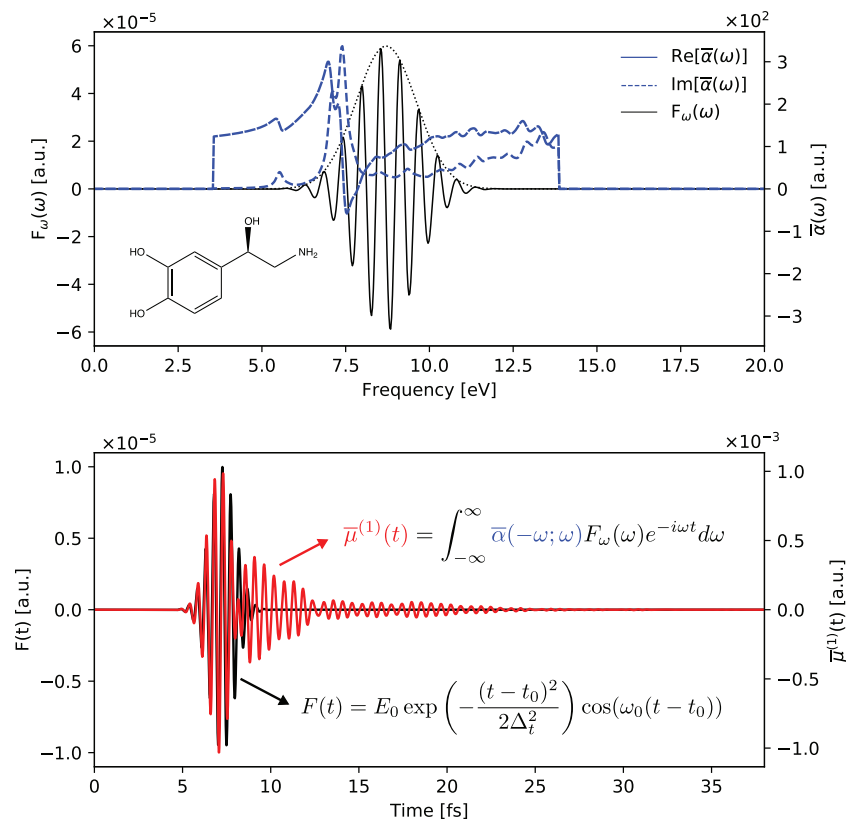


FIGURE 9 First-order interaction of noradrenaline with a Gaussian-shaped optical pulse ($\omega_0 = 0.32$ a.u.; $\Delta_t = 30$ a.u.; $E_0 = 10^{-5}$ a.u.; $t_0 = 300$ a.u.). (a) Section of real and imaginary parts of the frequency-domain isotropic electric-dipole polarizability, $\bar{\alpha}(-\omega; \omega)$, truncated by the region of appreciable field spectral amplitude, $F_\omega(\omega)$. (b) Time-domain representation of the electric field, $F(t)$, and the resulting first-order-induced dipole moment, $\bar{\mu}^{(1)}(t)$. Computational details are identical to those in Figure 5

executes well on single-node desktops and laptops, and the strictly modular implementation in a Python/C++ layered fashion enables interactive usage in a notebook environment suitable for educational purposes. Such training aspects are described in some further detail in the release paper of Gator 1.0¹⁵ that represents a twinning project of ours aimed at spectroscopy simulations with the use of electron-correlated wave functions.

All computationally intensive parts of the program are formulated in terms of the construction of Fock matrices and it is a program hallmark to efficiently handle and construct a very large number of such matrices in parallel. This core functionality enables the efficient implementation of a multifrequency and multigradient complex (and real) linear response equation solver. As thousands of Fock matrices can be processed in parallel, “windows” of spectral regions that are up to and beyond 10 eV wide are conveniently addressed and the problematic resolution of eigenstates in dense spectral regions can thereby be circumvented. Furthermore, the atomic-to-molecular orbital transformation that is a cornerstone in conventional implementations of correlated electronic structure theory methods can also be Fock-matrix driven and thereby carried out on HPC cluster platforms. This has been described in detail in the program release paper of Gator 1.0.¹⁵

The low memory-footprint integral code is based on a modified version of the Obara–Saika scheme with caution exercised not to introduce hardware-specific solutions while yet remaining performance competitive. To reach the reported high degree of SIMD vectorization comes as the result of C++ source code solutions that are not suitable to be implemented and, more importantly, maintained by a human programmer. We manage to circumvent this issue by means of meta-coding, that is, we have developed a C++ code of manageable size and complexity that writes the extensive compile-stage source code. All things considered, we believe we have put ourselves in a position to have an integral code platform flexible enough to navigate the landscape of HPC hardware solutions, at least for the foreseeable future. This belief will be put to a first test as we will next begin the endeavor of addressing heterogeneous hardware—in particular, execution on general-purpose graphical processing units.

In terms of program capability, VeloxChem 1.0 provides a potent and flexible platform for future developments and expansions of the spectroscopy toolbox. We expect that the strictly modular and layered programming model will serve us well in this respect, and future extensions to other spectroscopies involving linear and also nonlinear response functions are expected to be quite straightforward, constituting suitable tasks for students in the field. Specifically, the next program release will include an efficient implementation of complex response theory up to third order, or cubic response functions, and, among other things, it will encompass the important property of two-photon absorption.³⁶

Refraining from listing all of the functionality not included in this first release version of the program, we nevertheless conclude this discussion by pointing out the lack of analytical molecular gradients and Hessians. We regard this as another crucial development for the next release version as it enables molecular structure optimizations and simulation of vibrational spectroscopies.

ACKNOWLEDGMENTS

Financial support is acknowledged from the European Commission in the form of the ITN titled “*Computational Spectroscopy in Natural Sciences and Engineering (COSINE) (Grant No. 765739)*,” the Swedish Research Council (Grant No. 2018-4343), the Research Council of Norway (Grant No. 274918), the Villum Foundation (Grant No. VKR023371), and the Swedish e-Science Research Centre (SeRC) as well as computational resources provided by the Swedish National Infrastructure for Computing (SNIC).

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Zilvinas Rinkevicius: Conceptualization; methodology; project administration; software. **Xin Li:** Conceptualization; methodology; project administration; software. **Olav Vahtras:** Methodology; software. **Karan Ahmadzadeh:** Software. **Manuel Brand:** Software; visualization. **Magnus Ringholm:** Software; visualization; writing-original draft. **Nanna List:** Software; visualization; writing-original draft. **Maximilian Scheurer:** Software. **Mikael Scott:** Software; visualization; writing-original draft. **Andreas Dreuw:** Conceptualization; funding acquisition; project administration. **Patrick Norman:** Conceptualization; funding acquisition; methodology; project administration; writing-original draft.

ORCID

Manuel Brand  <https://orcid.org/0000-0003-3992-043X>

Patrick Norman  <https://orcid.org/0000-0002-1191-4954>

RELATED WIRES ARTICLE

[The Dalton quantum chemistry program system](#)

REFERENCES

1. Lindon JC, Tranter GE, Holmes JL, editors. *Encyclopedia of spectroscopy and spectrometry*. San Diego, CA: Academic Press, 2010.
2. List NH, Knoops J, Rubio-Magnieto J, et al. Origin of DNA-induced circular dichroism in a minor-groove binder. *J Am Chem Soc*. 2017a; *139*:14,947–14,953.
3. Norman P, Ruud K, Saue T. *Principles and practices of molecular properties*. Chichester: John Wiley & Sons, Ltd, 2018.
4. Akimov AV, Prezhdo OV. Large-scale computations in chemistry: A bird's eye view of a vibrant field. *Chem Rev*. 2015; *115*:5797–5890.
5. Ochsenfeld C, Kussmann J, Lambrecht DS. Chapter 1: Linear-scaling methods in quantum chemistry. In: Lipkowitz BK, Cundari TR, editors. *Reviews in computational chemistry*. Volume 23. Hoboken, NJ: Wiley, 2007; p. 1–82.
6. Chow E, Liu X, Smelyanskiy M, Hammond JR. Parallel scalability of Hartree-Fock calculations. *J Chem Phys*. 2015; *142*(10):104,103. <https://doi.org/10.1063/1.4913961>.
7. Dalcín L, Paz R, Storti M, D'Elía J. MPI for Python: Performance improvements and MPI-2 extensions. *J Parallel Distrib Comput*. 2008; *68*:655–662.
8. van der Walt S, Chris Colbert S, Varoquaux G. The NumPy array: A structure for efficient numerical computation. *Comput Sci Eng*. 2011; *13*(2):22–30.
9. Jones E, Oliphant T, Peterson P. SciPy: Open source scientific tools for Python. 2001. [Cited 2019 Jul 26]. Available from: <http://www.scipy.org>.
10. Sun Q, Berkelbach TC, Blunt NS, et al. PySCF: The Python-based simulations of chemistry framework. *WIREs Comput Mol Sci*. 2018; *8*(1):e1340.
11. Smith DGA, Burns LA, Sirianni DA, et al. Psi4NumPy: An interactive quantum chemistry programming environment for reference implementations and rapid development. *J Chem Theory Comput*. 2018; *14*(7):3504–3511.
12. Jakob W, Rhineland J, Moldovan D. pybind11—Seamless operability between C++11 and Python. 2017. [Cited 2019 Jul 26]. Available from: <https://github.com/pybind/pybind11>.
13. Scheurer M, Reinholdt P, Kjellgren ER, Olsen JMH, Dreuw A, Kongsted J. CPPE: An open-source C++ and Python library for polarizable embedding. *J Chem Theory Comput*. 2019; *15*(11):6154–6163.
14. Scheurer, M. (2019), *CPPE: C++ and Python library for polarizable embedding*. <https://doi.org/10.5281/zenodo>

15. Rehn DR, Rinkevicius Z, Herbst M, et al. Gator: A Python-driven program for spectroscopy simulations using correlated wave functions. *WIREs Comput Mol Sci*. 2019, submitted.
16. Obara S, Saika A. Efficient recursive computation of molecular integrals over Cartesian Gaussian functions. *J Chem Phys*. 1986;84:3963–3974.
17. Ahlrichs R. A simple algebraic derivation of the Obara–Saika scheme for general two-electron interaction potentials. *Phys Chem Chem Phys*. 2006;8:3072–3077.
18. Treutler O, Ahlrichs R. Efficient molecular numerical integration schemes. *J Chem Phys*. 1995;102:346–354.
19. Stratmann ER, Scuseria GE, Frisch MJ. Achieving linear scaling in exchange-correlation density functional quadratures. *Chem Phys Lett*. 1996;257:213–223.
20. Sellers H. The C2-DIIS convergence acceleration algorithm. *Int J Quantum Chem*. 1993;45(1):31–41.
21. Kauczor J, Norman P. Efficient calculations of molecular linear response properties for spectral regions. *J Chem Theory Comput*. 2014;10:2449–2455.
22. List NH, Kauczor J, Saue T, Jensen HJA, Norman P. Beyond the electric-dipole approximation: A formulation and implementation of molecular response theory for the description of absorption of electromagnetic field radiation. *J Chem Phys*. 2015;142(24):244,111.
23. Jiemchooraj A, Norman P. Electronic circular dichroism spectra from the complex polarization propagator. *J Chem Phys*. 2007;126:134,102.
24. Norman P. A perspective on nonresonant and resonant electronic response theory for time-dependent molecular properties. *Phys Chem Chem Phys*. 2011;13:20,519–20,535.
25. Jiemchooraj A, Norman P, Sernelius BE. Electric dipole polarizabilities and C6 dipole-dipole dispersion coefficients for sodium clusters and C60. *J Chem Phys*. 2006;125(12):124,306.
26. Olsen JMH, Kongsted J. *Chapter 3 - molecular properties through polarizable embedding, pp. 107–143*. San Diego, CA: Academic Press, 2011.
27. List NH, Norman P, Kongsted J, Jensen HJA. A quantum-mechanical perspective on linear response theory within polarizable embedding. *J Chem Phys*. 2017b;146(23):234,101.
28. Pedersen MN, Hedegård ED, Olsen JMH, Kauczor J, Norman P, Kongsted J. Damped response theory in combination with polarizable environments: The polarizable embedding complex polarization propagator method. *J Chem Theory Comput*. 2014;10:1164–1171.
29. Gagliardi L, Lindh R, Karlström G. Local properties of quantum chemical systems: The LoProp approach. *J Chem Phys*. 2004;121(10):4494–4500.
30. Ågren H, Harczuk I, Vahtras O. Decomposition of molecular properties. *Phys Chem Chem Phys*. 2019;21(5):2251–2270.
31. Nørby MS, Vahtras O, Norman P, Kongsted J. Assessing frequency-dependent site polarisabilities in linear response polarisable embedding. *Mol Phys*. 2017;115(1–2):39–47.
32. Norman P, Parello J, Polavarapu PL, Linares M. Predicting near-UV electronic circular dichroism in nucleosomal DNA by means of DFT response theory. *Phys Chem Chem Phys*. 2015;17:21,866–21,879.
33. Kumar A, Schweitzer-Stenner R, Wong BM. A new interpretation of the structure and solvent dependence of the far UV circular dichroism spectrum of short oligopeptides. *Chem Commun*. 2019;55(40):5701–5704.
34. Nakai Y, Mori T, Inoue Y. Theoretical and experimental studies on circular Dichroism of Carbo[n]helicenes. *J Phys Chem A*. 2012;116:7372–7385.
35. Li X, Parrish RM, Liu F, Kokkila Schumacher SIL, Martínez TJ. An Ab initio exciton model including charge-transfer excited states. *J Chem Theory Comput*. 2017;13:3493–3504.
36. Fahleson T, Norman P. Resonant-convergent second-order nonlinear response functions at the levels of Hartree-Fock and Kohn-Sham density functional theory. *J Chem Phys*. 2017;147:144,109.

How to cite this article: Rinkevicius Z, Li X, Vahtras O, et al. VeloxChem: A Python-driven density-functional theory program for spectroscopy simulations in high-performance computing environments. *WIREs Comput Mol Sci*. 2019;e1457. <https://doi.org/10.1002/wcms.1457>