eChem: A notebook exploration of quantum chemistry

Thomas Fransson,† Mickael G. Delcey,† Iulia Emilia Brumboiu,‡ Manuel Hodecker,† Xin Li,† Zilvinas Rinkevicius,† Andreas Dreuw,¶ Young Min Rhee,§ and Patrick Norman*,†

†Division of Theoretical Chemistry and Biology, KTH Royal Institute of Technology, 100 44 Stockholm, Sweden
‡Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, 87-100 Toruń, Poland
¶Interdisciplinary Center for Scientific Computing, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany
§Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Korea

E-mail: panor@kth.se

Abstract

The eChem project features an e-book published as a web page (https://doi.org/10.30746/978-91-988114-0-7), collecting a repository of Jupyter notebooks developed for the dual purpose of explaining and exploring the theory underlying computational chemistry in a highly interactive manner as well as providing a tutorial-based presentation of the complex workflows needed to simulate embedded molecular systems of real biochemical and/or technical interest. For students ranging from beginners to advanced users, the eChem book is well suited for self-directed learning, but workshops led by experienced instructors and targeting student bodies with specific needs and interests can readily be formed from its components. This has been done by using eChem as the base for a workshop directed towards graduate students learning the theory and practices of quantum chemistry, resulting in very positive assessment of the interactive nature of this framework. The members of the eChem team are engaged in both education and research and as a mirroring activity, we develop the open-source software upon which this e-book is predominantly based. The overarching vision and goal of our work is to provide a science- and education-enabling software platform for quantum molecular modeling on contemporary and future high-performance computing systems, and to document the resulting development and workflows in the eChem book.

Introduction

Theoretical chemistry has made tremendous progress in the past few decades and it is today an indispensable tool in all fields of molecular science. However, due to the complexity of the problems addressed, as well as the multitude of available software packages, the computational modelling of chemical processes may seem as a daunting task, not only to beginners but also to experienced users in need to expand their arsenal of methodologies. Also, software adaptation by students engaged in method development is made difficult by the fact that links between equations and software implementations are often vague. Therefore, the development of user-
friendly software, interactive educational tools, and instructive tutorials is important. From a general educational point of view, we argue that a deeper level of understanding can be reached by introducing a component of student software exploration, where students create their own algorithms to model physical processes in line with the quip:

What I cannot create, I do not understand. —Richard Feynman

By combining interactive notebooks with modern, Python-driven, software packages, the eChem project provides a platform to enable such creation. Dual purposes are here being served namely (i) to concretize the sometimes opaque theory of quantum chemistry by including explicit calculations alongside the derivation of the theory and (ii) to include more advanced tutorials showing various aspects of state-of-the-art computational chemistry. The eChem book benefits a wide readership, but it is perhaps particularly useful for advanced undergraduate and beginner graduate students.

The chosen framework is Jupyter Notebook\(^1\) with which web pages can be built by means of the Jupyter Book project.\(^2\) Jupyter is an increasingly popular interactive tool\(^3,4\) that can combine code, data collection, text, calculations, and more. A notebook consists of an arbitrary number of cells with markdown text or program code that is executed separately. This provides a large flexibility for exploration, explanation, and execution, at a noted risk of suboptimal code structuring.\(^3,5,6\) These features explain the massive growth in the number of scientific Jupyter notebooks, with notable examples in the discovery of gravitational waves,\(^7\) or first imaging of a black hole.\(^8\) Additionally, interactive notebooks offer an educational tool attractive to students\(^9\) and which can provide deep insights into complex topics.\(^10\) Notebooks can be implemented in a problem-based learning protocol,\(^11\) and it is particularly interesting to be able to use research-grade computational chemistry software.\(^11\) Educational notebooks have been realized in a wide range of different fields,\(^9,10,12–22\) and modules like nbgrader have been developed to simplify the use of notebooks for exercises and assignment grading.\(^23\)

While the external framework for the eChem project is provided by Jupyter Notebook, the quantum chemical software packages at the core of the project offer the necessary routines for quantum chemistry calculations. The primary packages are those developed by us, including VeloxChem,\(^24\) MultiPsi,\(^25\) and Gator.\(^26,27\) Some additional chemistry software is also used such as semiempirical methods from xTB,\(^28,29\) structure optimization with geomeTRIC,\(^30\) and molecular dynamics with OpenMM.\(^31\) For analysis and visualization, routines in NumPy,\(^32\) matplotlib,\(^33\) py3Dmol,\(^34,35\) and k3d\(^36\) are used. These software modules are chosen for their performance and ease of use, as well as having a Python layer from which calculations can be run, enabling easy manipulation, exploration, and development. Furthermore, these are free and open-source software packages, which enable wider adoption, improved user friendliness, flexibility, and a democratization of science.\(^37,38\) We note a range of other modules and software packages with similar features,\(^37,39–45\) including some with focus on education.\(^46–49\) Additionally, projects like Psi4Numpy,\(^46\) Dalton Project,\(^45\) and Open Chemistry\(^38\) include interactive notebooks and have some overlap with eChem.

In the following, we present the philosophy of the eChem project. The use of this e-book for education is discussed, including self-directed and workshop situations, as well as our experiences in teaching using eChem. In closing, we provide an outlook on the future expansion of the e-book and some general conclusions.

Our philosophy

The “e” in eChem can refer to electronic as well as educational. The goal of this project and our software development is to be both science- and education-enabling, providing a comprehensive and living repository of theory, workflows, and tutorials. The living descriptor means that the eChem web page is subject to constant changes, being updated with new and improved discus-
sions of underlying theory, as well as workflows and illustrations of our most recent research efforts. Together with the ability of providing numerical examples and tutorials, this approach has major advantages over a more static printed book. This means that the book is never finished, and portions of it are in a rather preliminary state—this article thus focuses on the parts that are sufficiently polished at the time of writing.

The term science-enabling refers in this context to workflows and recommendations for a wide range of methodologies and applied studies, providing a very transparent overview of how we model a range of chemical processes. For this, we require software capable of considering dense 3D systems of large size (with stable and reliable solvers), excited states of diffuse character, time-efficient prototyping of novel methods, providing transparent exposure of data and data-structures to enable in-depth analysis, interaction with other software packages, and programs suitable for both laptops and top high performance computing (HPC) resources. The exploratory and flexible nature of a Jupyter framework here benefits novices and veterans in the field alike, enabling the use of novel computational workflows, interactive computations and analysis, and rapid development cycles. For instance, during the writing of the eChem book the authors appreciated how easily unusual workflows and examples could be implemented and visualized, resulting in many examples illustrated on the web page.

In the context of education-enabling, the notion of deep learning refers to taking each student’s understanding of the subject matter to another (deeper) level. Our experience tells us that the process of implementing methods to solve fundamental equations is supremely efficient as a means to achieve that sort of deeper learning and understanding, but only a small number of students are granted this opportunity as many core modules of scientific software were written a long time ago and have often been made obscure by code optimization. We believe that lectures and textbooks are indispensable to present the foundations of theory but also limited in their ability to develop the intuition needed as a skill set of the (computational) chemist. The eChem book is a complement that allows a hands-on approach to computational chemistry. The black-box characteristic of a program code is dismantled into its algorithmic pieces, which in turn are examined and built back together as to become representatives of the main methods of modern computational chemistry. Discussions of theory are thus intertwined with numerical examples, bringing the educational material into the realm of deep learning. This, coupled with the exploratory nature of notebooks and the ample opportunities for visualization provides a stimulus for self-directed and deep learning.

Finally, an important part of our philosophy is that research and education should, as far as possible, be freely and easily available as well as user-friendly. As such, we favor software packages that are free and open-source, easy to install with a common package managing tool, and offering interoperability at the Python level. We have chosen Jupyter Notebook as our vehicle for interactive computations. We note that, historically, such a choice has suffered from lower performance than running more specialized codes written in, for instance, Fortran through a batch queuing system, but developments made to combine Python modules with C++ routines for compute-intensive parts as well as frameworks such as JupyterHub to access powerful computational resources is constantly diminishing this gap. In this new landscape, we believe that the decrease in cognitive cost reached by adapting computational protocols to the sequentially operating (single-core) brain rather than multi-tasking computers often outweighs the increase in wall time.

The eChem book

The title page of the eChem book is shown in Fig. 1. The book is composed of six parts with underlying chapters and sections, intertwining theory with numerical examples and analyses. The first five chapters are directed towards a discussion of the underlying theory, encompassing the topics electronic ground states, molecu-
lar structure and dynamics, spectra and properties, environment, and visualization. Notebook markdown cells with text and equations presenting the theory are intertwined with code cells implementing the said expressions. These code cells, which also perform insightful calculations with associated figure illustrations, are key to the deeper student learning, not least since they are made available for download and further curiosity-driven exploration. Following this, the final chapter is titled tutorials and workflows, and focuses on practical calculation and analyses of spectra and properties. This demonstrates how one or several software programs (or modules) are used to perform commonly conducted modeling in theoretical chemistry and how the notebook approach to simulations naturally promotes software interoperability. A brief outline of the (current) content of each chapter and section is found in the Supplementary Information.

Examples of theory integrated with numerical results, access to deeper routines, and visualization tools are found in Fig. 2. Panels (a) and (b) use H₂ and CO, respectively, to illustrate the concept of electron correlation. Panel (a) presents a calculation of the correlation energy. As the interpretation of quantum mechanics is probabilistic, the one- and two-electron densities take center stage in quantum chemistry. The eChem book is able to provide the student with visual aids to understand this difficult topic of ubiquitous importance, as shown in panel (b). For visualizing molecular orbitals, panel (c) shows how the OrbitalViewer class provides a light-weight and highly responsive interface. For multi-reference calculations, this module also presents natural occupation numbers and provides an interactive interface for molecular orbital selection and the construction of active spaces. Finally, panel (d) shows how the exposure of grid points from the VeloxChem program enables an exploration of the exchange-correlation kernel integration that
typically is treated in a black-box manner in teaching as well as research.

**eChem for education**

**Self-directed learning**

Chemistry students in general, including undergraduates and experimentalists, will find an opportunity in the eChem book to learn the basics of theoretical chemistry by performing actual calculations that mirror the theory and concepts under study. We argue that, just as a theoretician benefits from knowledge about the basic experimental methods, the experimentalist benefits from knowledge about the basic computational methods. This promotes a mutual understanding and interdisciplinary work.

Self-directed learning with the eChem book can also be applied in the context of flipped learning, where the focus is on active learning from the part of the student, as supported by instructors. While flipped learning comes with its set of challenges, it has been seen to generally provide additional insight and improved critical thinking skills,\(^{51-53}\) with the field of chemistry in particular likely to benefit from such an approach.\(^{51}\)

Additionally, a reoccurring event in a dynamic and active research group is the graduation of developed PhD students followed by the recruitment of new ones. In this situation, the eChem book serves as a means to transfer knowledge. The beginner student may have access to a curriculum of advanced courses, yet there will always be a gap between basic theory and the field-specific knowledge and practical skills that need to be acquired before research activities can be commenced. This training is greatly accelerated with interactive and targeted study materials that are made suitable for self-directed learning.

**Designing a course module or workshop**

There are entire books available discussing the application of computational methods for teaching chemistry and chemical concepts as well as the introduction of programming into the chemistry curriculum.\(^{54,55}\) It has been noted that access to computational chemistry is essential for undergraduate chemistry programs, a viewpoint that is strongly supported by us. However, the integration of computational chemistry in undergraduate teaching is hampered by a lack of user-friendly and easily accessible software/hardware environments and instructors with adequate training.\(^{47,56}\) The eChem project is designed to break these barriers.

The use of notebooks for courses and workshops represents a form of blended learning which promotes active learning and has been noted to fare well when compared to traditional learning activities.\(^{57}\) In active learning, the student is engaged in active, meaningful exercises via technological tools that provide cognitive support.\(^{58}\) It can be noted that assessment practices are about as important as presentation practices,\(^{59}\) and in this context nbgrader\(^\text{23}\) is of particular relevance. This is a tool which facilitates creating and grading assignments (coding or written form) in a Jupyter notebook environment, including auto-grading functionalities as well.

A module in an undergraduate course based on the eChem book would typically take the form of a series of computer exercises dispersed among more conventional lectures covering the topics addressed in the parts titled *Electronic ground states* and *Molecular structure and dynamics*. A workshop is likely to cover the more advanced topics covered in *Spectra and properties* and *Environment* and combine a presentation of the underlying theoretical foundations with hands-on exercises either taken directly or inspired from the part titled *Tutorials and workflows*.

**Our experience: eChem workshop at KAIST**

To probe the value of using the interactive tools provided by the eChem book, we organized a one-week workshop for chemistry students at the Korea Advanced Institute of Science and
Technology (KAIST) in November 2022, which attracted 23 students in total. The workshop was held as part of a semester-long course in Quantum Chemistry and nine students were
from the course. The workshop aimed to illustrate fundamental concepts in quantum chemistry by combining different interactive Jupyter notebooks from the eChem book. As about half of the students were not familiar with Python, or the Jupyter notebook framework, the week started with an introductory hands-on session aimed to familiarize the participants with the necessary Numpy routines, as well as some central VeloxChem class objects. The subsequent two days were dedicated to electronic structure theory, where the students wrote their own Hartree–Fock (HF) and density functional theory (DFT) programs and explored the concept of electron correlation by visualizing two-particle densities at the HF and second-order Møller–Plesset (MP2) levels of theory. The final two days were dedicated to the exploration of potential energy surfaces (PESs), where the students wrote their own geometry optimization routines, determined the vibrational modes of a molecule, and constructed analytical force fields. At the end of the week, we asked the students to choose one or two projects out of a pool of six, which were to be carried out individually or in pairs. These projects served as examination as well as a stimulus for self-directed learning.

Even though the students had very different backgrounds and experience with Python or computational chemistry, the workshop received very positive feedback. The results of a student survey carried out at the end of the workshop are summarized in Fig. 3. Students were asked to what extent they agree or disagree, ranging from 1 (strongly disagree) to 5 (strongly agree), with a series of statements, included as insets in the graphs. The fourth panel contains the average answers to four specific questions, as detailed in the Supporting Information.

![Figure 3: Results of the student survey carried out after the eChem workshop at KAIST. The students were asked to what extent they agree or disagree, ranging from 1 (strongly disagree) to 5 (strongly agree), with a series of statements, included as insets in the graphs. The fourth panel contains the average answers to four specific questions, as detailed in the Supporting Information.](image)

understanding facilitated by the interactive aspects of eChem:

*The interactive course was much better for solving my questions / difficulties in understanding.*

*It’s very helpful to understand the chemical concepts while doing the exercises and computational calculation.*

*Things finally became tangible, practical, when I could see processes and results.*

Additionally, they felt that interactive learning made them more invested in the course compared to a traditional setup:
It helped me to be more invested and active than “normal” classes.

The few negative remarks to be found were concerned with the (too short) time restrictions on the interactive sessions and the level of difficulty and depth at which some of the topics were addressed:

Wish we had [another] week to explore deeper concepts.

One suggestion for improvement that we are considering to incorporate in future workshops is to use pop-quizzes to check that all students have reached the necessary understanding to continue further with the presentation.

Finally, one student provided a very rewarding summarizing statement about the eChem KAIST workshop:

I would highly recommend this workshop to anyone, and encourage you to reach many students in the world.

Summary and Outlook

We have here presented the eChem educational project, featuring an online e-book that enables and stimulates interactive self-directed learning but can also be used to form course modules and workshops. Numerical examples with enlightening figure plots carried out in notebook code cells are intertwined with presentations of basic theory in markdown cells. These self-contained notebooks are easy to download and execute on the reader’s personal computer, which promotes further student exploration driven by author-designed exercises and/or stimulated curiosity.

Different parts of the eChem book are suited for students not only at different stages in their educational training but also with different focus in their own work. Undergraduate and beginner graduate students are likely to benefit mostly from the chapters devoted to basic theory and concepts and, in this regard, the overwhelmingly positive student assessment from the workshop held at KAIST is highly encouraging to us. More senior graduate students and researchers in computational chemistry are likely to benefit mostly from the workflows and tutorials describing methodologies and software interoperability to tackle complex molecular systems of real technical and/or biochemical interest. In addition, the eChem notebooks demonstrate the exposure of C++ data structures in the Python layer and which in turn enables the prototyping of novel scientific methods. This will benefit graduate students and researchers with a focus in their work on theory development and program implementation.

As noted earlier, the eChem book is a living document, designed to be constantly improved and expanded such that its repository of notebooks exemplifies and explains the state-of-the-art methodologies in theoretical chemistry. It will never be complete in coverage and will always be influenced by the research interests of its authors. The most imminent additions to be made include tutorials on catalysis and reactivity, optical activities, and nonlinear field interactions.

Supporting information

Overview and content of eChem (PDF). Archive with material from KAIST workshop (ZIP).

Web page

URL: https://doi.org/10.30746/978-91-988114-0-7

QR code:
Acknowledgements

We are very grateful to Prof. Seogjoo J. Jang (City University of New York) for incorporating the eChem workshop into his course during his sabbatical stay at KAIST. The authors are also deeply indebted to KAIST Institute of Science-X (KAIX) for its financial and administrative supports toward the eChem workshop at KAIST. Further acknowledgements are directed toward financial supports from the Swedish e-Science Research Centre (SeRC), the Swedish Research Council (Grants No. 2018-4343 and 2020-04639), the Air Force Office of Scientific Research (Grant FA8655-20-1-7010), and Samsung Electronics (I0211126-09175-01) as well as computing resources provided by the Swedish National Infrastructure for Computing (SNIC).

Author declarations

Conflicts of interest

The authors have no conflicts of interest to disclose.

Licensing

The eChem book is published under the Creative Commons Attribution-ShareAlike license CC BY-SA.

References


(11) Rodríguez-Becerra, J.; Caceres-Jensen, L.; Diaz, T.; Druker, S.; Padilla, V. B.; Pernaa, J.; Aksela, M. Developing technological pedagogical science knowledge through educational computational chemistry: a


