What drives computational chemistry forward: theory or computational power?

Wenfa Ng

Citizen scientist, Singapore, Email: ngwenfa771@hotmail.com

Abstract

History is often thought to be dull and boring – where large numbers of facts are memorized for passing exams. But the past informs the present and future, especially in delineating the context surrounding specific events that, in turn, help provide a deeper understanding of their causes and implications. Scientific progress (whether incremental or breakthroughs) is built upon prior work. Chronological examination of computational chemistry's evolution reveals the existence of major "epochs" (e.g., transition from semi-empirical methods to first principles calculations), and the centrality of key ideas (e.g., Schrodinger equation and Born Oppenheimer approximation) in potentiating progress in the field. The longstanding question of whether computing power (both capacity and speed) or theoretical insights play a more important role in advancing computational chemistry was examined by taking into account the field's development holistically. Specifically, availability of large amount of computing power at declining cost, and advent of graphics processing unit (GPU) powered parallel computing are enabling tools for solving hitherto intractable problems. On the other hand, this essay argues (using Born Oppenheimer approximation as an example) that theoretical insights' role in unlocking problems through simple (but insightful) assumptions is often overlooked. Collectively, the essay should be useful as a primer for appreciating major development periods in computational chemistry, from which counterfactual questions illuminate the relative importance of theoretical insights and advances in computer science in moving the field forward.

Keywords: theory, simulation, computational power, epochs, science history,

Subject areas: chemistry, theoretical chemistry,

1

Simulation, together with theory and experiment, comprises the triumvirate of science. By tracing computational chemistry's development path, the relative importance of theory and computing power in advancing the field is examined in this essay. Specifically, chronological examination of computational chemistry's development shed light on three major overlapping periods: (i) theory development for explaining observations of spectroscopic emission lines of elements and prediction of their atomic structure, (ii) utilization of simplifying assumptions and experiment data for circumventing problems associated with lack of computing power in solving the Schrodinger equation in the pre-computing era, and (iii) dramatic increase in computing power giving rise to first principles (*ab initio*) methods for solving, with few or no simplifying assumptions, large systems comprising polyatomic and long chain molecules. Besides the three eras mentioned, the increasing popularity of hybrid methods utilizing fine grained approaches for simulating key aspects of a model, and less accurate (but adequate) methods for examining other facets is increasingly recognized as a defining feature of the current period. The essay closes by critically assessing the relative importance of theoretical ingenuity and computational power in seeding new developments and breakthroughs, where simple but elegant insights, such as the Born Oppenheimer approximation, help open paths to previously inaccessible solutions.

Existence of distinct eras in computational chemistry

Electronic structure calculations, a sub-field of computational chemistry, focuses on developing methods for explaining and predicting atom organization and interactions between sub-atomic particles (i.e., neutrons, electrons and protons), and latter, how electron density is distributed across orbitals and their role in mediating bond formation between atoms.[1] As in other research areas, a voluminous body of literature describes myriad methods and tools developed at various junctures in the field's evolution. Though seemingly disparate and not easy to organize, analyzing different methods and tools chronologically lends clarity to their inter-relationships and reveals different phases in the field's development. Specifically, the three overlapping periods are: (i) initial experiment and theoretical studies elucidating the structure of the atom, and the motions of its sub-atomic constituents (a period where theory lagged behind experiment); followed by (ii) use of simplifying assumptions for solving models of single or few atoms and calibration of parameters using experiment data in an era of relatively low computing power (this was a period where approximate and semi-empirical methods dominate); and finally, (iii) the emergence of first principles (*ab initio*) modeling approaches for solving large systems (comprising hundreds to thousands of molecules) with few or no assumptions and the aid of low cost computing power.

Currently, we may be in the midst of a nascent era where a variety of coarse-graining or model reduction approaches that incorporate simplifying assumptions or experiment data (for model parameters) helps researchers tackle problems previously only accessible on supercomputers.[2] [3] Specifically, by using fine-grained methods on aspects of a problem that directly informs the answers sought while allowing some inaccuracies to permeate in other areas, model reduction approaches significantly reduce the computing load. More importantly, these approaches allow large systems comprising complex molecules to be tackled using affordable and accessible computing resources such as a small cluster of graphics processing units (GPU) powered computers, yielding results and answers in a reasonable amount of time compared to brute-force approaches.

Evolution of the field

The discovery of sub-atomic particles such as the electron, proton and neutron sow the seeds of computational chemistry as an independent field of scientific inquiry. Specifically, researchers of the day debated competing theories concerning atom organization, and the mechanistic underpinnings of the forces mediating interactions between sub-atomic particles. Success of the quantum mechanical approach (over classical physics) in explaining the key observation that orbiting negatively-charged electrons do not spiral into the positively-charged nucleus ushered in the emerging field of electronic structure calculations, whose main goal at that time was to explain the emission spectra of various elements. Specifically, peaks present on the emission spectrum of elements (e.g., sodium and hydrogen) arise from the release or absorption of energy during transition of electrons between different energy levels. Realization that electrons, or more accurately, electron densities, are arranged in defined energy levels and spatial regions led to the proposal of the atomic and molecular orbital concepts. From a quantum mechanical perspective, these are regions where electrons of specific energies are located. This era was defined by the promulgation of many of the foundational concepts and tools of computational chemistry, where the theoretical tools of quantum mechanics illuminate spectroscopic observations, highlighting that theory lagged behind experiment. Perhaps the enabling contribution of this era was the formulation, by Erwin Schrodinger, of an equation that describes the total energy of any system. Known simply as the Schrodinger equation, its intractability to solution spawned an entire sub-field seeking to develop methods and strategies for solving it.[4] More specifically, solution of the equation is crucial for understanding the placement of electrons of differing energies in distinct orbitals, which in turn, determines an atom's chemical properties.

Development of various approximate methods (incorporating simplifying assumptions) for solving the Schrodinger equation dominated the second era of electronic structure calculations, of which the Born Oppenheimer approximation is the most iconic.[5] [6] Specifically, the objective was to devise increasingly better and faster techniques for solving the electronic portion of the total

system energy formulation with the help of simplifications such as ignoring the electrostatic repulsions between electrons (also known as electron-electron correlations). One example that exemplifies the utility of approximations and for solving intractable problems (with slight but tolerable inaccuracies) is the Born Oppenheimer approximation, which decouples electronic and nuclear motions encapsulated within the Schrodinger equation. Specifically, coupled motions of the nucleus and electrons, where electrons' movement influences the atom's nucleus and vice versa, accounts for the mathematical intractability of the Schrodinger equation, which more accurately defined, is a many body problem. The key to resolving this conundrum hinges on the observation that, for atoms of sufficiently large atomic mass, the nucleus is essentially fixed in space; thus, allowing the entangled motions of the nucleus and electrons to be solved independently. More importantly, the approximation would increasingly lead to more accurate solution as the atomic mass increases. By applying the approximation, only the electronic component of system energy needs to be solved; thus, significantly reducing the computational load.

Given the inability of mechanical slide rule and rudimentary calculators in computing properties of atoms with sufficient accuracy and precision, the second era of computational chemistry was also characterized by the emergence of many semi-empirical methods,[7] [8] where experiment data (usually from spectroscopy studies) were used in calibrating essential parameters in system models. These parameters describe key characteristics of atoms and could not be calculated from first principles in the pre-computing era. In addition, lack of computing power also constrained the types of systems studied to those involving single or few atoms. In the case of larger systems comprising more atoms, models exist to solve them approximately, but the simplifications used were unrealistic.

Increases in computing speed and capacity, and the availability of user-friendly software packages signal the arrival of the third era of electronic structure calculations. Specifically, greater computing power allows the calculation, from first principles, of most system properties with minimal reliance on simplifying assumptions, in systems comprising large numbers of polyatomic molecules.[9] [10] [11] Although systems of thousands of proteins remain inaccessible to even the fastest supercomputers, significantly larger systems of at least few tens of molecules (more than 10 atoms) have become increasingly accessible to simulation.[9] This allow meaningful answers to questions concerning reactivity, chemical kinetics and evolution of transition states to be obtained.[9] In addition, advances in computing also democratizes computational chemistry; specifically, by allowing non-specialist researchers to perform routine investigations of simple systems using user-friendly software packages on desktop computers – compared to command-line programmes on mainframe or supercomputers in earlier eras. Although not the sole *ab initio* method available, density functional theory (DFT) utilizing Gaussian type orbitals is the dominant technique for tackling a wide range of questions concerning reactivity and molecular recognition between molecules, in fields as diverse as material science, biochemistry and physics.[12]

Finally, desire for simulating ever larger systems of long chain molecules (representative of real world systems) using less computing time, or on desktop computers and small parallel computing clusters, has driven the development of various model reduction strategies, in what is emerging as a nascent fourth era. This development taps on the efficiency and speed of semi-empirical and approximate methods, and the need of tackling large scale systems at spatiotemporal resolutions more closely resembling those of natural systems. Specifically, even though multi-fold speedup is available from high performance computers, simulating the dynamics of excited electrons within a single macromolecular biomolecule is the technical limit; thus, motivating the development of hybrid methods to more effectively simulate a system's most critical aspects for answering a given question. Within the family of model reduction strategies, coarse-graining approaches, which combines first principles methods with simplifying assumptions, is increasingly used.[13] [14] Essentially, coarse-graining seeks to employ the most suitable tool for different components of a problem. *Ab initio* techniques, for example, would be used in simulating the precise atomic movement during the binding and cleavage of a molecule at an enzyme's active site, while the important (but less critical) interactions between enzyme and water solvent would be approximated by a mean field that captures, in aggregate, all the electrostatic and *van der Waals* interactions between water molecules, and those between the enzyme and water molecules. Thus, using a mean field (to substitute for more fine-grained calculations) in simulating the solvent effect on enzyme catalysis, significantly reduces the computational requirement that a full explicit treatment of water molecules' interaction with the enzyme would create. Current trends, where many investigators are employing myriad simulation techniques for solving problems at spatial and temporal scales relevant to real world problems, together with large system size and complexity meant that, in the absence of a significant leap in computing power and reduction in cost, model reduction approaches would likely remain popular. However, future development of algorithms capable of first principles simulation of large systems at a fraction of current computational cost, would revolutionize the field by making obsolete many of the model reduction approaches currently in vogue.

History seldom evolves linearly, but rather, is punctuated by sets of related events that arose due to unique circumstances at certain periods of time. For example, closer examination of the delineated timeline reveals the clustering of different methods into distinct periods, the binning of which depends on the assumptions used and the extent to which experiment data from spectroscopy and other instruments helps model building. One framework for demarcating different periods in the field clusters methods into four overlapping eras: (i) theoretical postulations and experiment elucidation of the structure of atoms and their sub-atomic constituents, (ii) calculations of electron density distribution and understanding the basis of chemical bonding for single or few atoms using simplified models calibrated with experiment data, (iii) simulations of systems comprising large number of polyatomic molecules with few or no assumptions (i.e., first principles calculations), and (iv) combined use of fine-grained and less accurate methods in model reduction approaches for solving large systems (i.e., coarse-graining approaches). Thus, the evolution of electronic

structure calculations can be understood chronologically or through the identification of different developmental periods. However, clustering and binning myriad developments into categories inevitably requires a set of arbitrary criteria – such as the relative importance of simplifying assumptions, experiment data input, theory or computing power. Choice of binning criteria is dependent on the questions asked and solutions sought. Thus, criteria used and perspectives taken in examining a question determines the different eras defined; a point crucial to remember as it indicates the essentiality of choosing relevant criteria to avoid self-fulfilling conjectures.

Theory or computational power?

Since past events cannot be rewound, counterfactual analysis (i.e., asking "what if" questions) is useful for illuminating the likely consequence or implications of alternative trajectories of events at specific periods of time. Similarly, counterfactual analysis could shed new light on the longstanding question concerning the relative importance of computing power and theoretical insights in advancing computational chemistry. Dramatic increases in computing power is generally recognized to have propelled computational chemistry forward; however, this commentary argues that the interplay between computational power and theory may be more nuanced. For instance, closer examination of events following promulgation of the Schrodinger equation reveals that the significant computational challenge posed by the coupled motions of nucleus and electrons might had presented a stumbling block to research if not for the Born Oppenheimer approximation. Specifically, the Born Oppenheimer simplification enabled researchers to solve the simpler problem of electron motion in the context of a fixed nucleus – an approximation which progressively improves with increasing atomic mass. Doing so allows partial solutions to be obtained, which although with caveats attached, helped illuminate paths or strategies to solving real world problems. Thus, theoretical insight's importance in unlocking or initiating developments in computational chemistry is often under appreciated. In addition, theoretical insights also serve as a useful check on your thinking, especially in clarifying the cloud of convoluted equations that might otherwise obfuscate meaning, or act as roadblocks in the smooth flow of logical thought.

This commentary is an initial attempt at providing some thoughts on a longstanding debate – and certainly does not provide the last word on the issue. More detailed analysis of the question would come from science historians. History is the continuous evaluation, from different perspectives, of existing evidence in light of new developments. This, together with successive generations of scholars casting their backward glance on past events from different vintage points, would engender differing interpretations of the same events, which is healthy for promoting intellectual debate. Specifically, future developments in computational chemistry and re-interpretations of old evidence from fresh perspectives may lead to slightly different conclusions on the above debate.

Collectively, analysis of computational chemistry's chronology highlights four overlapping phases in the field's development, i.e., (i) development of theories for explaining experiment observations of emission lines of elements and prediction of electron density distributions in atoms; (ii) semi-empirical and approximate methods utilizing experiment data and simplifying assumptions for calculating electronic structure of single or few atoms; (iii) first principles methods for tackling systems comprising large numbers of polyatomic molecules; and (iv) combination of fine-grained and approximate methods in model reduction approaches for reducing the computational requirement during solution of large systems at meaningful spatiotemporal resolution within reasonable time. Finally, exploration of the relative roles played by computing power and theoretical insights in advancing the field illuminates the importance of theoretical ingenuity in unlocking up-to-now intractable problems, while acknowledging the importance of large amount of inexpensive computing power in potentiating the transition from semi-empirical to first principles methods.

References

(1)   Kratzer, P.; Neugebauer, J. The Basics of Electronic Structure Theory for Periodic Systems. *Front. Chem.* **2019**, *7*, 106. https://doi.org/10.3389/fchem.2019.00106.

(2)   Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42* (1), 73–93. https://doi.org/10.1146/annurev-biophys-083012-130348.

(3)   Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

(4)   Nakatsuji, H. Discovery of a General Method of Solving the Schrödinger and Dirac Equations That Opens a Way to Accurately Predictive Quantum Chemistry. *Acc. Chem. Res.* **2012**, *45* (9), 1480–1490. https://doi.org/10.1021/ar200340j.

(5)   Echenique, P.; Alonso, J. L. A Mathematical and Computational Review of Hartree–Fock SCF Methods in Quantum Chemistry. *Mol. Phys.* **2007**, *105* (23–24), 3057–3098. https://doi.org/10.1080/00268970701757875.

(6)   Born–Oppenheimer and Non-Born–Oppenheimer, Atomic and Molecular Calculations with Explicitly Correlated Gaussians | Chemical Reviews https://pubs-acs-org.libproxy1.nus.edu.sg/doi/10.1021/cr200419d (accessed Dec 29, 2019).

(7)   Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116* (9), 5301–5337. https://doi.org/10.1021/acs.chemrev.5b00584.

(8)   Thiel, W. Semiempirical Quantum–Chemical Methods. *WIREs Comput. Mol. Sci.* **2014**, *4* (2), 145–157. https://doi.org/10.1002/wcms.1161.

(9)   Friesner, R. A. Ab Initio Quantum Chemistry: Methodology and Applications. *Proc. Natl. Acad. Sci.* **2005**, *102* (19), 6648–6653. https://doi.org/10.1073/pnas.0408036102.

(10)   Friesner, R. A.; Guallar, V. Ab Initio Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics (Qm/Mm) Methods for Studying Enzymatic Catalysis. *Annu. Rev. Phys. Chem.* **2005**, *56* (1), 389–427. https://doi.org/10.1146/annurev.physchem.55.091602.094410.

(11)   Friesner, R. A.; Dunietz, B. D. Large-Scale Ab Initio Quantum Chemical Calculations on Biological Systems. *Acc. Chem. Res.* **2001**, *34* (5), 351–358. https://doi.org/10.1021/ar980111r.

(12)   Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115* (19), 2315–2372. https://doi.org/10.1080/00268976.2017.1333644.

(13)   Saunders, M. G.; Voth, G. A. Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* **2013**, *42* (1), 73–93. https://doi.org/10.1146/annurev-biophys-083012-130348.

(14)   Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.

Conflicts of interest

The author declares no conflicts of interest.

Funding

No funding was used in this work.