Quantum chemistry: a complementary supporting tool for crystallography

Jernej Stare

National Institute of Chemistry, Ljubljana, Slovenia

1. Summary. *Quantum chemistry* (QC)[1] encompasses a wide range of computational methods based on the most fundamental physical background governing the structure of matter, namely quantum mechanics. The core part of QC methods is the calculation of the electronic structure based on the solving of the Schrödinger equation, a highly sophisticated numerical protocol. QC provides unique insight into matter, facilitating characterization of structure, reactivity, spectroscopic features and many other properties. As such, quantum chemistry represents a valuable tool for the interpretation of the experiment and evaluation of properties that are beyond the reach of experiment. Together with the more approximate but size-efficient classical methods, quantum chemistry comprises the field of *molecular simulation* (molecular modeling) which has evolved into a standalone discipline and has been routinely used in various fields of research.

With constantly growing capabilities of parallel computing, the interpretative power of quantum chemistry has been increasingly exploited in studies of crystalline solids. A class of QC methods features inherent support for periodicity (e.g., periodic DFT), making them ideal for the modeling of crystals. Applications include validation of structure determination, improved refinement (e.g., hydrogen atoms), stability of polymorphs and isomers, external pressure effects, investigation of phase transitions and analysis of intermolecular interactions. The notable enhancement of research establishes QC methods as indispensable tool in crystallography, stimulates interdisciplinary studies and provides new research perspectives.

2. Background. The key task of any modeling protocol, whether based on classical or quantum physics, is the calculation of the potential energy of a chemical system defined by its atomic positions; therefore, *for a given set of atomic coordinates, the protocol yields the corresponding energy*. That alone facilitates a number of useful applications such as geometry optimization, vibrational analysis and molecular dynamics; it also allows for hierarchical ranking of isomers or conformers of the same chemical system.
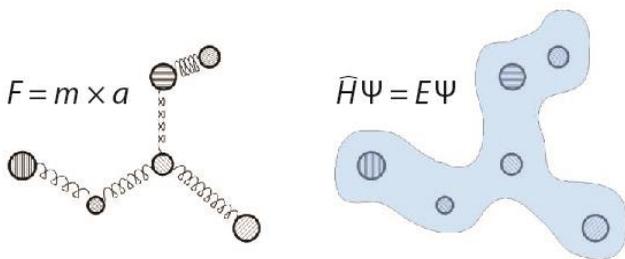
Evidently, the crucial part of any computational method is how the energy is derived from the supplied atomic coordinates. Depending on the underlying physical background, classical or quantum methods are being used. ***The classical methods are based on Newtonian physics*** and the matter is represented as ensemble of particles (atoms) connected by elastic springs (chemical bonds). The energy is formulated as empirical function of intramolecular coordinates (bond lengths, valence angles and torsions) augmented with non-bonding interactions, of which electrostatics is represented by the Coulomb potential using empirical atomic charges, whereas the Lennard-Jones potential is

---

[1] The herein used acronym "QC" is rather non-standard in quantum chemistry communication. Instead, the acronym "QM" (quantum mechanics) is often being used.

normally used for the modeling of Van der Waals interactions. ***The parameters defining the potential energy function are obtained empirically*** by fitting to experimental data or to highly accurate quantum calculations. A generic term ***force field*** is normally used for the set of parameters and equations used in classical simulation of molecular systems.

Importantly, ***classical models do not include explicit treatment of the electronic structure*** of the system; all the information about the electron density is collapsed into force field parameters and atomic charges. This is due to the fact that unlike atomic nuclei electrons are beyond classical treatment, because Newtonian physics breaks down for sufficiently light and fast particles. In order to fully explore the electronic structure (and its role in bonding, reactivity etc.), a first principle treatment of electrons is required, engaging the principles of quantum mechanics. The corresponding quantum

Figure 1. Schematic comparison between classical (left) and quantum-chemical model (right).

treatment is based on the solving of the ***Schrödinger equation*** (SE), a mathematical formulation of the most fundamental physical principles:

$$\hat{H}\Psi = E\Psi.$$

In the case of chemical systems, the Hamiltonian operator $\hat{H}$ and wavefunction Ψ pertain to the electrons of the system of interest. The wavefunction defined in the 3D space is related to the probability density of the particles (i.e., electrons) while the operator $\hat{H}$ extracts the energy from the wavefunction. The Hamiltonian can be divided into two parts, namely kinetic ($\hat{T}$) and potential ($\hat{V}$):

$$\hat{T} = -\sum_i \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2};$$
$$\hat{V} = \sum_{\alpha,\beta} \frac{1}{r_{\alpha\beta}} - \sum_{i,\alpha} \frac{1}{r_{i\alpha}} + \sum_{i,j} \frac{1}{r_{ij}}.$$
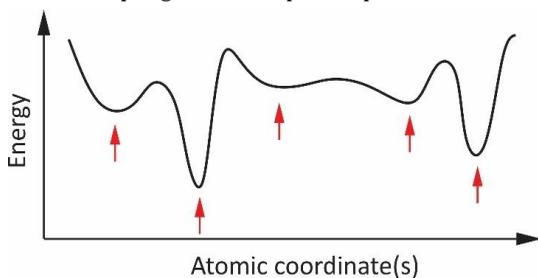
According to postulates of quantum mechanics, the kinetic operator (kinetic energy of electrons) includes double derivatives; therefore, the SE is a differential equation of the second order. The three terms in the potential energy operator are the following pairwise electrostatic interactions: repulsion between nuclei, attraction between nuclei and electrons, and repulsion between electrons. Due to the electron repulsion term ($1/r_{ij}$) the variables in the equation are not separable; consequently, the SE can only be solved analytically for a system of one nucleus and one electron (hydrogen atom) whereas ***for molecular systems it has to be solved approximately by numerical techniques***.

A number of methodologies for the solving of the SE exist, differing in complexity, accuracy and computational cost, but virtually all of them share the formulation of the searched wavefunction in terms of the so called ***basis functions*** (also called ***basis set***). Two major groups of basis sets exist, namely (i) ***atomic basis sets*** and (ii) ***plane wave***

basic sets. The former are functions resembling orbitals of hydrogen-like atoms (1s, 2s, 2p, 3s, 3p, 3d,...) that are centered at nuclear positions; the latter are sinusoidal functions of different periodicity delocalized over the simulation cell. Techniques using plane waves are ideal for the modeling of crystalline solids due to the inherent periodicity of plane waves. Among the approaches for the solving of the SE the most well known are the Hartree-Fock (HF) method, *Density Functional Theory (DFT)* methods and perturbation methods. Perhaps the most popular to date are the DFT methods due to their cost-efficiency and high accuracy. DFT methods support plane wave basis sets and represent vast majority of quantum calculations on crystalline solids (*"periodic DFT"*). Recently, DFT methods previously suffering from underestimated dispersion interactions have been greatly improved by including suitable corrections.

A detailed explanation of QC approaches and the underlying mathematical/numerical formalisms is beyond the scope of this presentation. For more information the reader is referred to the corresponding textbooks and online resources.[2]

3. Basic types of calculations. In addition to the calculation of energy and electron density, most quantum calculations (and all classical ones) also support the calculation of forces (gradients of the energy) along atomic coordinates. This facilitates the search of lowest energy conformation for a given system, a protocol named energy minimization or *geometry optimization*. Optimization of geometry is perhaps the most widely used protocol in molecular simulation, as it is carried out in almost every computational study. In the case of periodic structures it is also possible to optimize unit cell parameters. Another extremely important treatment derived from forces is *molecular dynamics simulation* (MD), i.e. integration of equations of motion, yielding real-time thermal fluctuations of the system indispensable for the sampling of the phase space and for free energy calculations. However, both in the case of optimization and MD one needs to mind that the complexity of the system often prevents from finding the lowest energy minimum or from sampling the entire phase space relevant for the investigated system. Particularly the crossing of barriers can impose a problem that needs to be solved by advanced simulation and sampling techniques.

The fact that quantum chemistry is based on explicit treatment of the electron density facilitates analysis of the density by methods of *population analysis*, yielding atomic charges, contours of orbitals, information on reactivity, etc. In addition, quantum chemistry features inherent support for the modeling of

Figure 2. Simplified potential energy surface with indicated minima. Very often it is not possible to sample the entire surface due to complexity and high barriers.

[2] See for example: W. J. Hehre, L. Radom, P. von R. Schleyer, J. Pople, *Ab Initio Molecular Orbital Theory*, Wiley 1986; A. R. Leach, Molecular Modeling: Principles and Applications, Pearson Education Ltd. 2001; https://en.wikipedia.org/wiki/Quantum_mechanics.

chemical reactions, i.e. bond breaking and making. With classical methods this cannot be trivially done, as the bonds are designated to be 'permanent', and modifications of the force field and advanced sampling methods are required to support reactivity.

The two main approaches used in molecular simulation – classical and quantum – are to a great extent complementary, as outlined in Table 1. While classical methods are appropriate for large systems and complex environment and for the sampling of large number of possible conformations, quantum methods are much more accurate and useful for a detailed insight into the structure and reactivity. A special class of simulation methods combines the advantages of both, treating the central subsystem of interest by a quantum method whereas the surroundings is modeled by a more approximate classical method. Such type of modeling is known as **multiscale modeling** or **QM/MM** approach.
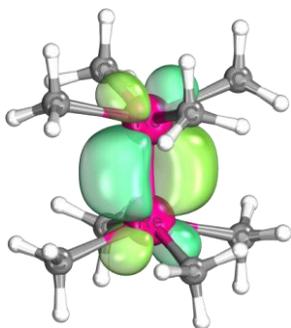
Figure 3. Example of visualization of bond orbitals. Taken from web presentation of G. Knizia's research group, Penn state University, http://sites.psu.edu/knizia/.

Table 1: Comparison of selected features of classical and quantum chemical methods.

Feature	Classical methods	Quantum methods
Physical background	Newtonian	Quantum-mechanical
Calculation of energy	From empirical force field	Directly from first principles
Information on electronic structure	None (implicitly included in empirical parameters)	Fully available
Modeling of reactions	Only by advanced methods ('permanent' bonds)	Inherently supported
Time for energy evaluation	~0.01 sec. for 10,000 atoms	A minute to many days for 50 atoms
Size limit (approx.)	>100,000 atoms	100-1,000 atoms
Overall description	Approximate but suitable for large systems	Accurate but size limited

A number of quantum chemistry program packages are available, and many of them support periodic calculations. Examples of widely used programs are Gaussian, VASP, CPMD and CRYSTAL, but there are many others. The packages differ in the treatment of the electron density (either by atomic orbitals or plane wave basis sets) and in the available protocols such as molecular dynamics and analysis of the density. Licensing and pricing policies are also quite diverse, but for academic research it is generally easy to

acquire license for the appropriate program package at relatively low cost or even free of charge.

While the workflow of quantum calculations is very complex, most of the QC programs allow for fairly simple handling. The input data supplied by the user typically includes atomic positions (together with unit cell details and symmetry code for periodic systems), simple specification of the computational methodology (e.g., which density functional and which basis set), and keywords defining the protocol (e.g., geometry optimization, molecular dynamics, vibrational analysis, etc.). Also, the output information can be fairly easily digested and processed in most cases. Visualization of results is facilitated by a number of molecular graphics programs (Molden, VMD, XCrySDen…).

4. Applications. As computational techniques (quantum in particular) are universally applicable to any molecular system, it is not a surprise that they found a variety of applications in the field of crystallography. One of the most widely used (and at the same time one of the simplest) applications is ***validation of structure determination***. In short, geometry optimization of the structure is performed using the experimental crystallographic data as starting point, and the resulting structure is compared to the experimental one, possibly yielding improved structural model. Such procedure is particularly useful in conjunction with powder diffraction where uncertainties in atomic positions is typically higher than with single crystal diffraction. A recently published in-depth review of validation of powder diffraction crystal structures includes as many as 215 diverse examples.[3] The structures optimized by periodic DFT were critically evaluated against the single crystal structures that served as benchmark. The general agreement between DFT-optimized and experimental structures was very good, and cases when DFT prediction failed were rare. The study confirms that periodic DFT methods represent a valuable independent tool for validation of crystal structures. The power of DFT validation is reflected in the growing number of published studies combining crystallographic and computational techniques.
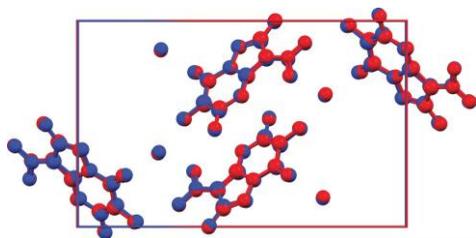
Figure 4. Example of overlaid experimental (red) and DFT-optimized (blue) crystal structure taken from ref. 3.

Figure 5. Structure determined from powder X-ray diffraction (left) and refined structure after DFT

---
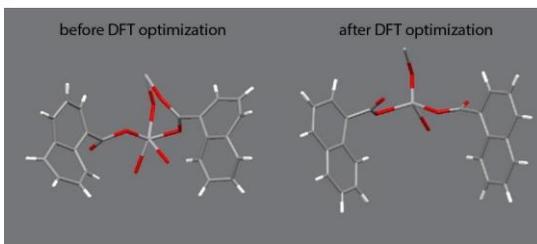
[3] J. van de Streek, M. A. Neuman, *Acta Cryst.* **B70**, 1020-1032 (2014). See also references cited therein.

optimization followed by refinement (right).

Our research experience includes a number of examples of crystal structure validation. Perhaps the most pronounced role of DFT calculations was in the case of structure determination of vanadium oxo arene carboxylates (nanocrystalline hybrid organic-inorganic materials), where DFT optimization noticeably altered the initial structural model and subsequent Rietveld refinement starting from the optimized structure resulted in a significantly reduced R factor.[4] Among the rest, DFT optimization flattened the arene rings, changed coordination number of vanadium from 5 to 4 and the carboxylate biding type from partially bidentate to monodentate. The optimized structure was further validated by comparison of the computed normal modes with the experimental infrared spectrum, confirming that the use of calculations indeed provided great improvement in structure determination.

A trivial standalone application of the structure validation by DFT methods is the **refinement of hydrogen positions** that can hardly be reliably determined by X-ray diffraction due to the low electron density around hydrogen atoms. While neutron diffraction can overcome this problem, the availability of neutron facilities is limited and requirements for the quantity and quality of the sample are much higher than with X-ray methods, rendering neutron methods unpractical. Reportedly, DFT optimization can greatly improve bond distances involving hydrogen, which is particularly valuable in studies of hydrogen bonds.[5] Additionally, DFT methods can elucidate the hydrogen disorder issue, particularly related to tautomerism and rotational isomerism.[3]

Disorder is a common issue accompanying structure determination by crystallographic techniques. Generally, disorder can originate from non-equal content between individual unit cells (static disorder) or from dynamical isomerism within the unit cells (dynamic disorder). In both cases the resulting experimental structure pertains to a long-time average of structures present in the sample, possibly yielding averaged (non-typical) geometry parameters, or non-unity occupancies of atomic sites. **Periodic DFT calculations can in principle elucidate the sources of disorder**; however, such calculations can be rather demanding due to the necessity of using large supercells. For example, one of the most popular titanium-doped silicalite catalysts TS-1
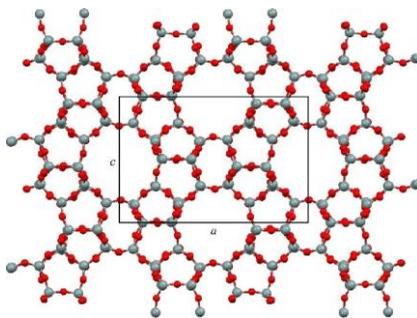
Figure 6. Crystal structure of TS-1 viewed along b axis.

---

[4] I. Djerdj *et al.*, *J. Mater. Chem.* **22**, 10255-10265 (2012).
[5] K. Molčanov *et al.*, *Phys. Chem. Chem. Phys.* **16**, 998-1007 (2014); E. A. Kapustin *et al.*, *Cryst. Growth Des.* **14**, 1851-1864 (2014).

contains small amount of titanium within the parent MFI zeolite crystal structure containing 12 crystallographically inequivalent silicon sites,[6] but not much has been known about the distribution of titanium among these sites. By using periodic DFT and HF calculations the possible structural isomers were optimized, yielding hierarchical energy preference of the substitution sites.[7] However, the results of different studies were rather controversial, suggesting that the structure of TS-1 may be governed by kinetic factors (at the stage of synthesis) rather by thermodynamic stability.

A similar but more consistent elucidation was performed on the structure of a ternary mixture of lanthanum, niobium and titanium oxide ($LaTi_2Nb_3O_{13}$), a class of ceramics with application in microwave technologies, in which two titanium and three niobium atoms are distributed over five available sites, yielding 10 possible isomers. By DFT calculations energy preference of isomers was calculated and the partial occupancies of Ti and Nb were estimated, in fine agreement with experimental observations.[8]

Periodic DFT calculations are also suitable for the modelling of **effects induced by external pressure**. The pressure is included in the DFT formalism by using a stress tensor which, multiplied by the cell volume, represents an additional energy term in the model Hamiltonian. High pressure crystallographic studies can reveal important information about intermolecular interactions governing the structure; they are often accompanied by calculations.[9] We recently studied polymorphism and phase transitions of crystalline L-serine induced by high pressure using periodic DFT and gas phase calculations.[10] The results suggest that at a microscopic level phase transitions are lead by decrease in the volume of the crystal unit cell (contributing to the enthalpy change), but not the internal crystal energy. At a microscopic level we suggest that H-bond overstrain leads to jump-wise phase transition with substantial experimental hysteresis, while no such overstrain was found for "normal type" phase transition. The predicted pressures for the phase transitions deducted by the minimum enthalpy criterion are in a very good agreement with the observed ones. The study suggests a new computational approach to understanding the mechanisms of pressure-induced phase transitions in organic crystals.

In order to explain the observed structural features, the predictive power of periodic DFT can be exploited beyond the treatment of the investigated structure. With modeling it is possible to construct and
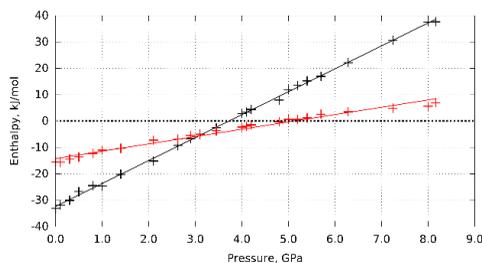

Figure 7. Predicted pressures of the phase transitions L-serine I → L-serine II and L-serine II → L-serine III. Black - ΔH_{II-I}, Red - ΔH_{III-II}, the transitions can be expected when the ΔH values become equal to zero.

rc, J. Stare, *Acta Chim. Slov.* **55**, 951-959 (2008).

[9] N. Casati, P. Macchi, A. Sironi, *Chem. Commun.* 2679-2681 (2009); P. Macchi *et al. Cryst. Eng. Comm.* **12**, 2596-2603 (2010).
[10] D. Rychkov, J. Stare, E. Boldyreva, *in preparation* (2016).

investigate hypothetical structures and acquire helpful information for the interpretation of a similar structure of interest. One such example is our work on cystalline nitranilic acid hexahydrate, a benchmark system containing the Zundel cation ($H_5O_2^+$).[5a] Due to the high acidity of the ring OH groups the proton is transferred to the crystalline water molecules and forms the $H_5O_2^+$ entity.
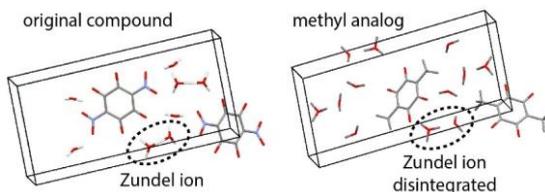
Figure 8. DFT-optimized crystal structure of the original compound (left) and its methyl analog with indicated disintegration of the Zundel cation upon substitution.

We examined the source of the electronegativity of the ring group on the acidity and replaced the ring nitro group by various functional groups of lower electronegativity and re-optimized the crystal structure. Consequently, on reducing electronegativity of the ring substituent, the Zundel cation disintegrates and ultimately leads to a "neutral" structure consisting of protonated acid and non-protonated water molecules. This confirms the decisive role of electronic effects for the existence of a peculiar structural feature. ***It is virtually impossible to elucidate this aspect by experimental methods because the corresponding analogs do not exist.*** In addition, in the present case periodic DFT methods (including MD) were extensively used for the characterization of proton dynamics (probed experimentally by infrared and NMR spectroscopy), synergistically enhancing crystallographic investigation of the selected structural features.

Finally, periodic quantum calculations tackle the issue of ***crystal structure prediction*** (CSP), i.e. predicting how a molecule will crystalize solely from the structure of the molecule. This challenge has a tremendous impact for industries relying on synthesis of organic materials, particularly pharmaceutical industry. CSP has been stimulated by organizing competitions of research groups in the so called blind tests, where the participants are given the molecular structures of various compounds differing in complexity, whereas their crystal structures are kept secret. The participants are required to predict the crystal structures by using computational methods and predictive algorithms of their choice. The project was started in 1999[11] and recently the sixth blind test has been completed,[12] considering systems of increasing complexity and constantly improving the success of predictions. The computational assignment can be divided into two stages, namely (i) ***sampling***, in which a number of plausible crystal structures are generated on the basis of the structure of the molecule of interest, and (ii) ***ranking***, where the so obtained structures are evaluated for their actual existence. The sampling issue is

---

[11] J. P. M. Lommerse *et al., Acta Cryst.* **B56**, 697-714 (2000).
[12] A. M. Reilly *et al., Acta Cryst.* **B72**, 439-459 (2016); A. J. Cruz-Cabeza, *Acta Cryst.* **B72**, 437-438 (2016).

related to the capability of thoroughly sampling the intermolecular phase space, i.e. generating and briefly evaluating as much structures as possible. Because of the enormous complexity of the potential energy landscape, this stage requires fast, albeit approximate sampling methods, therefore classical (empirical) simulation techniques are preferred. However, the empirical parameters of the simulation are often determined by fitting to the results of the more sophisticated quantum calculations. For successful ranking accurate assessment of thermodynamic stability of the selected structures is required, and periodic DFT and other QC methodologies are indeed prevalent at this stage. An increasing importance of quantum methods has been observed in the development of CSP. Despite the clearly advancing power of CSP, many challenges remain open for further research, particularly in the areas of nucleation and growth of crystals and kinetic factors. In the future quantum methods will likely provide important contribution in these compelling issues.

5. Conclusions. Quantum chemistry finds a diverse array of applications in crystallography, providing an independent tool for the characterization of structure, dynamics, reactivity and other properties of crystalline solid materials. The synergism between X-ray crystallographic techniques and theoretical methods is reflected in the fact that at high resolution the experimental density is expressed in fine detail, thus approaching the first principle representation defined by quantum mechanics. Therefore, linking the high resolution XRD and QC calculations appears like a natural evolution of both methodologies, representing a challenge for further development.