Quantum Computing for Molecular Biology

Baiardi, Alberto; Christandl, Matthias; Reiher, Markus

Published in:
ChemBioChem

DOI:
10.1002/cbic.202300120

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Molecular biology and biochemistry interpret microscopic processes in the living world in terms of molecular structures and their interactions, which are quantum mechanical by their very nature. Whereas the theoretical foundations of these interactions are well established, the computational solution of the relevant quantum mechanical equations is very hard. However, much of molecular function in biology can be understood in terms of classical mechanics, where the interactions of electrons and nuclei have been mapped onto effective classical surrogate potentials that model the interaction of atoms or even larger entities. The simple mathematical structure of these potentials offers huge computational advantages; however, this comes at the cost that all quantum correlations and the rigorous many-particle nature of the interactions are omitted. In this work, we discuss how quantum computation may advance the practical usefulness of the quantum foundations of molecular biology by offering computational advantages for simulations of biomolecules. We not only discuss typical quantum mechanical problems of the electronic structure of biomolecules in this context, but also consider the dominating classical problems (such as protein folding and drug design) as well as data-driven approaches of bioinformatics and the degree to which they might become amenable to quantum simulation and quantum computation.

1. Introduction

The elucidation of the molecular foundations of biology has changed the way we understand biology at the cellular level. Examples are discoveries in the molecular machinery of cells such as the structural characterization of proteins and their mode of action at atomistic resolution\[1–4\] as well as the genome analysis of cells of different organisms that allowed for a correlation to the phenomenological theory of evolution.\[5–8\]

Despite the fact that molecular biology addresses the atomistic resolution of biological structures and function and may therefore be expected to be inherently quantum, it is typically not perceived as a field in which quantum mechanical phenomena play a dominant role. This is due to the fact that detailed molecular dynamics simulations rest almost exclusively on classical Newtonian mechanics. Moreover, the resulting qualitative understanding of the functioning of biomolecular machinery, supported by extensive experimental results, often does not require a detailed atomistic structure and certainly not detailed atomistic dynamics. Instead, cartoonish representations of macromolecules are used in order to sketch a sequence of events that they undergo. This knowledge alone is already sufficient in order to create functional networks of biochemical pathways (classical examples are the glycolysis pathway and the Krebs cycle), leading to an understanding of molecular functions that may be viewed as a what-comes-next static picture of molecular events.

Still, only detailed quantitative physical simulations – in parallel to detailed experiments with a high spatial and temporal resolution – will allow extracting such a picture with high confidence. Classical molecular dynamics simulations provide efficient models and can be rigorously based on quantum mechanics (technically speaking, this is possible through the Born-Oppenheimer approximation, which separates electronic and nuclear motions, with the latter being then identified as the atomic motion in classical dynamics). Unfortunately, more detailed simulations of the quantum mechanical equations are very difficult and only possible for a very small number of atoms.

However, if we were to advance molecular simulations by quantum computation fueled by the current hardware and algorithm developments in this field,\[9–13\] we may wonder to what extent biomolecular simulations would actually benefit from such developments and whether quantum computation will become key in computational quantum molecular biology.\[15–18\] Brought to the point, the question is, whether the emerging branches of quantum computation may eventually deliver a significant advance over traditional approaches.

Whereas an intensive search is ongoing for quantum effects on function in biology,\[19–22\] the most important quantum effects are first and foremost rooted in the electronic structure of biomolecules and, to a lesser degree, in their quantum nuclear motion (giving rise, for instance, to tunneling and kinetic isotope effects). The electronic structure of molecules is indeed key for the quantitative theoretical description and prediction of chemical reactions in terms of reaction energies and activation barriers through the Born-Oppenheimer potential energy surface (PES; see Figure 1). In other words, reaction...
mechanisms of chemical processes in living cells can be analyzed and understood by computationally solving the relevant quantum mechanical equations. Such calculations provide insights into covalent bond formation in biomolecules important to biochemical pathways, into proton and electron transfer processes as well as into biological phenomena driven by visible and infrared light.

The fact that reversibility is key to molecular function in biochemistry causes macromolecules to interact predominantly through (many) non-covalent interactions. These weak interactions, such as van der Waals forces and hydrogen bonds, can be easily broken under ambient conditions. They originate from the quantum correlations of electrons. However, these electron correlation effects are very hard to describe in principle in a huge many-electron system such as a protein because the computational effort for the solution of the underlying electronic Schrödinger equation scales exponentially with system size. This curse of dimensionality creates a steep wall that cannot be overcome with traditional computational approaches (and therefore limits such attempts to small systems of comparatively few atoms, which represent only segments of biomolecules).

However, it is well known that these types of interactions can be accurately modeled by classical surrogate potentials for interacting atoms as new entities; typically only two fit parameters per pair potential are required as in Lennard-Jones potentials (note that no parameters whatsoever are required in explicit quantum chemical simulations of electronic structures). Such force-field approximations can be combined with algorithms that, being based on classical mechanics, do not suffer from the curse of dimensionality, but have a low polynomial scaling with the number of atoms. The dramatic gain in efficiency is a result of the mapping of the electron correlations onto pair interactions of atoms and thus mainly...
due to the different algebraic structure of classical vs. quantum mechanics.

Such classical model potentials are the key ingredients of biomolecular force fields in classical mechanics simulations of bio-macromolecules, which approximate the Born-Oppenheimer PES and facilitate the sampling of the high-dimensional configuration space through classical molecular dynamics or Monte-Carlo methods in order to produce free energies from which the thermodynamic behavior of a biochemical system can be deduced. Unfortunately, the underlying model potentials require sophisticated parametrizations, generally omit quantum mechanical effects, lack transferability to other molecule classes, and therefore cannot describe well arbitrary chemical reactions and molecular interactions (if at all). By contrast, none of these deficiencies is present in approaches that involve quantum mechanical effects, which are, however, hampered by the highly unfavorable scaling problem with system size.

In this work, we, therefore, discuss key computational target problems in molecular biology from the perspective of quantum computation and elaborate on where and how to achieve a quantum advantage in this field. The goal of this work is not only to deliver an overview of potential application areas and refer to pioneering results already accomplished but to give a specific roadmap on where to expect breakthroughs that would not be accessible with traditional approaches.

This paper is organized as follows: We first provide a tailored introduction to the foundations of computational physics modelling applied to molecular biology in order to establish some common ground in this highly interdisciplinary field for a better understanding of the specific examples that follow. We begin with some principles of physical modeling of biomolecular processes (Section 2); we then consider specific quantum equations (Section 3) and their traditional solution (Section 4), before we dwell on the basics of quantum computation (Section 5). Equipped with the key concepts and basic notation, we then turn to actual problems in molecular biology, where we first consider the theoretical description of chemical reactions in biology (Section 6), followed by phenomena of quantum biology (Section 7) and molecular function amenable to classical molecular mechanics (Section 8). We finish (Section 9) with remarks on the potential of data-driven methods for bioinformatics applications and eventually provide some general conclusions and outlook (Section 10).

2. Physical modeling of biological processes

The ways we describe, model, and understand processes in biology can be very diverse. If we consider molecular biology as the core of biology, from which biological structures and functions, cells, and eventually the living world as a whole emerge, achievements in chemistry and molecular physics can be exploited to build a quantitative theory of the dase and interactions of biomolecules.

In this ansatz, the dynamical behavior of key entities is described by a mechanical equation of motion, which allows for a redistribution of energy among kinetic and potential contributions that eventually determine biomolecular function. Since molecular biology rests on chemistry, it is also governed by the key theory of physics that quantitatively describes the electromagnetic interactions of electrons and atomic nuclei in and between (bio)molecules, namely quantum electrodynamics. Starting from this basis, a quantum many-particle theory (see Section 3) and several computational solution methods (cf. Section 4) have been devised.[31–34] As a result, energies and molecular structures can be predicted that allow us to describe molecular behavior with breathtaking accuracy and depth.

However, what can be accomplished in terms of actual computations depends on the size of the system, i.e., on the number of electrons and atomic nuclei. Very high accuracy can only be achieved for very small molecules. Increasing molecular size generally is accompanied by a loss in accuracy for feasibility reasons, due to the high computational cost of exact quantum simulations. More and more approximations need to be invoked as the molecular size grows. Ab initio wave-function-based methods aim at solving the many-body problem with approximate wave function parameterizations (such as truncated CI,[35] coupled cluster,[36] and perturbation theory,[37]), which are routinely applicable to molecules with several dozens of atoms. Larger systems can be studied with methods that also approximate the molecular Hamiltonian in terms of parameterized models. A prominent example is approximate density functional theory (DFT)[38–41] and related approaches (for instance, the random phase approximation[39]). The model parameters are often optimized to reproduce either highly accurate quantum chemical or experimental reference data. Although many benchmark studies have set out to quantify the accuracy of different flavors of density functional theory, it remains hard to assess a priori the accuracy of a given density functional approximation for a specific molecule not included in the training data.[40]

For reasons to be discussed later, we may consider the size of a molecular system of a few hundred atomic nuclei to be routinely treatable with approximate traditional quantum chemical approaches. Obviously, such molecular sizes are far smaller than those of functional biological macromolecules embedded in an aqueous phase.

To alleviate this drawback, embedding techniques have been devised, which provide a quantum magnifying glass for a cluster of atomic nuclei and electrons considered to be decisive for the quantum process within a bio-macromolecule (such as the active site of an enzyme). The neglected part of the biomacromolecule (and of the surrounding and penetrating water molecules) is then treated as an environment. If part of the environment does not simply play a “spectator” role, but is reactive – such as (small) part of a protein environment (e.g., a solvent molecule or a few amino acid residues) – it becomes part of the quantum subsystem of the embedding scheme. The remainder of the environment is treated by a classical mechanical model with its (possibly dynamic) partial charges and geometric constraints affecting the quantum region through an additional electrostatic potential. The determination of this quantum region can be automated (see, e.g., Refs. [41]...
and (42)). Numerous approaches have been devised for the embedding of a small active quantum part into an environment (see, e.g., Refs. [24, 43–53]). We note, however, that large-amplitude motions and long-range effects may be difficult to describe in such settings.

In terms of chemical diversity, biomolecules are mostly organic molecules, often with rather regular bonding patterns (such as proteins, polysugars, nucleic acids, and other biopolymers). It is for this reason that electronic quantum correlations turn out to be rather regular in the sense that they can be mapped to simple classical models. These classical models are much easier to handle in actual computations, and therefore, much larger molecular sizes can be accessed (such as large proteins and protein complexes).

In other words, the underlying Coulomb interaction of electrons and nuclei within and between biomolecules can be mapped onto specific forms of potential energy functions, now defined between new dynamical entities: atoms, united atoms, or even larger coarse-grained aggregates. Surprisingly transferable potential energy functions are obtained that capture the fundamental interactions in a molecular mechanics framework, which is now rooted in classical Newtonian mechanics rather than quantum mechanics. Prototypical analytical forms for the distance-dependent potential energies are parametrized electrostatic and Lennard-Jones model potentials. Only if chemical bonds are formed or broken, such descriptions become more involved and require so-called reactive force fields [14] that are, however, highly specialized and tailored toward a given class of molecules. Moreover, their accuracy is often difficult to assess – as is the case for any methods based on classical surrogate potentials. Due to the lack of generality and possibility of reliably controlling the error, only explicit quantum mechanical calculations can deliver reliable insights into any biochemical process that is accompanied by breaking and forming bonds.

All of these approaches refer to the same basic concept, namely that of a potential energy hypersurface, i.e., the PES, which assigns an energy to any atomic configuration of atomic nuclei/atoms in a molecule, and hence, allows one to judge structural changes in terms of a change in potential energy. The (quantum) energies that constitute the PES are the electronic energies defined in the Born-Oppenheimer approximation which assigns energy to a fixed scaffold of atomic nuclei, in which the electrons of the molecule move. Depending on the well depths, the number of wells, and the barrier heights exhibited by this hypersurface, different physical modeling approaches can be activated (see Figure 1).

The key take-home message is that we can describe the quantum mechanical behavior of electrons and nuclei in biomolecular systems in principle, but often not in practice, as such calculation would be too computer-time demanding. In such cases, one resorts to classical models and accepts a reduced accuracy and a loss in transferability. The promise of quantum computations is that a quantum description can be extended to larger atomistic structures and to higher accuracy (compared to approximate traditional quantum chemical approaches).

3. Theoretical foundations: Quantum electronic and vibrational structure

Molecular motion is governed by the laws of quantum mechanics. Defining these laws and the equations they lead to represent the first step toward understanding the potential of quantum computing for molecular simulations. Quantum mechanics describes a molecule through the molecular wave function \( \Psi(r, R, t) \), which is a complex-valued function that depends on the positions of all its elementary particles, i.e. the electrons \( r \) and the nuclei \( R \), and on time \( t \). The Born rule states that the squared norm of the wave function, \( |\Psi(r, R, t)|^2 \), defines the probability density of finding the molecule in a given configuration \( (r, R) \) at a given time \( t \). The second key quantity is the molecular Hamiltonian \( \hat{H}(r, R, t) \), which determines the time evolution of the wave function through the time-dependent Schrödinger equation,

\[
\hat{H}(t)\Psi(r, R, t) = \hat{\hbar}\partial_t \Psi(r, R, t).
\]  

As shown in Figure 2, the Hamiltonian describing a molecule contains contributions associated with the kinetic energy of the atomic nuclei and of the electrons as well as the interaction between these particles. The interactions are very well described by pairwise electrostatic Coulomb potential energies, although they neglect, for instance, magnetic interactions. Since these interactions do not depend on time, we can rigorously eliminate the \( t \) dependence of the Hamiltonian, i.e., \( \hat{H}(t) \rightarrow \hat{H} \). Now, any solution to Eq. (1) can be expressed in terms of specific wave functions, often referred to as stationary states, whose probability amplitudes do not change with time. These wave functions, \( \Psi_n(r, R) \), are the solutions to the time-independent Schrödinger equation,

\[
\hat{H}\Psi_n(r, R) = E_n \Psi_n(r, R).
\]  

i.e. they are the eigenvectors of the Hamiltonian \( \hat{H} \) with eigenvalues \( E_n \) their energy, assigned to them.

Molecular machinery in biology is viewed as intricate systems of numerous molecules, whose actions are governed mostly by reversible interactions of the biomolecules contained in the systems. Examples are hydrated protein complexes embedded in a membrane (such as ATPase) or ribosomes. It is because of the weakness of these reversible interactions that quantum mechanical modeling can efficiently address the quantum states of the individual molecular players (and their interactions) rather than describing all of them as a whole system; i.e., separated biomolecules and their interactions are studied rather than an organelle or a cell as a whole. While this separation into components makes the solution of the Schrödinger equation then feasible (at least for not too large molecular structures), it also allows one to piece together and eventually understand biomolecular machinery in terms of individual energies assigned to the states of the molecular players. For this reason, Eq. (2) is the main cornerstone of the simulation of biochemical processes.
Figure 2. Graphical representation of the key equations for quantum mechanical molecular calculations. Electronic properties are described, for a given configuration of the nuclei \( R \), by the electronic Hamiltonian \( H_{\text{ele}}^R \) defined in the upper left part of the figure. \( H_{\text{ele}}^R \) is obtained from the molecular Hamiltonian by neglecting the nuclear kinetic energy, i.e., by adopting the fixed-nuclei approximation. It contains, therefore, the kinetic energy of the electrons \( T_{\text{ele}}(r) \), the electron-electron Coulomb interaction \( V_{\text{ee}}(r) \), the nuclear-electronic Coulomb attraction \( V_{\text{ne}}^R(r) \), and the nuclear-nuclear Coulomb repulsion \( V_{\text{nuc}}^R(r) \). The corresponding ground and excited states wave functions \( \Psi_{\text{ele},\text{GS}}^R(r) \) and \( \Psi_{\text{ele},\text{ES}}^R(r) \) depend parametrically, on the nuclear ones \( R \) because it does not contain any derivative operator with respect to \( \mathbf{r} \). The corresponding electronic ground and excited states wave functions \( \psi_{\text{ele},\text{GS}}^R(r) \) and \( \psi_{\text{ele},\text{ES}}^R(r) \) will display the same parametric dependence. The variation of the electronic energy of a given electronic state with the nuclear coordinates defines the PES (represented as a green curve for the electronic ground state and as a red curve for the lowest-energy excited state). The PES is usually only calculated point-wise (black crosses) and approximated with fitting algorithms. We approximate the green PES around the left local minimum by a parabolic function (the corresponding curve is displayed with a dashed line), which defines the harmonic approximation. The vibrational Schrödinger equation for the electronic ground state is reported in the lower part of the figure. Its solution yields the vibrational wave functions \( \psi_{\text{vib,GSO}}^{(j)}(R) \). Since the PES changes with the electronic state, a different set of vibrational wave functions will be obtained for the excited state.

\[
\hat{H}_{\text{ele}}^R = T_{\text{ele}}(r) + V_{\text{ee}}(r) + V_{\text{ne}}^R(r) + V_{\text{nuc}}^R(r)
\]

\[
\hat{H}_{\text{ele}}^R(r)\psi_{\text{ele}}^R(r) = E_{\text{quantum}}^R \psi_{\text{ele}}^R(r)
\]

\[
(\hat{T}_n(R) + E_{\text{ele},\text{GS}}^R)\psi_{\text{nuc,GSO}}^{(j)}(R) = E_{\text{GS}}^j \psi_{\text{nuc,GSO}}^{(j)}(R)
\]

Already for a molecule with a few dozen atoms, the wave function depends on hundreds of variables, and Eqs. (1) and (2) become extremely hard to solve on classical computers. The Schrödinger equation can be simplified by invoking the Born-Oppenheimer approximation: since the mass of the nuclei is at least three orders of magnitude larger than that of the electrons, the nuclear dynamics are much slower than the electronic ones. It is, therefore, reasonable to calculate the wave function in two steps. First, the electronic wave function is calculated by neglecting the kinetic energy contribution of the nuclei. This yields the so-called fixed-nuclei Hamiltonian \( H_{\text{ele}}^R \) (see Figure 2), where the nuclear coordinates \( R \) do not enter as variables, but rather as fixed parameters. Therefore, the eigenvalues \( E_{\text{quantum}}^R \) and eigenfunctions \( \psi_{\text{ele}}^R(r) \) depend parametrically on the nuclear coordinates \( R \), which we therefore indicate as a superscript. They describe the allowed states and the corresponding energy of the electrons for a given configuration of the nuclei \( R \). The nuclear part of the molecular wave function is then calculated by assuming that the change in the nuclear coordinates is slow enough to ensure that the electrons are described by the same eigenstate (e.g., the ground state) \( \psi_{\text{ele}}^R(r) \) for all \( R \) values. This approximation can be seen as a result of the adiabatic theorem of quantum mechanics. The eigenvalue \( E_{\text{quantum}}^R \) associated with the eigenfunction describing the electrons defines the effective interaction potential between the electrons and the nuclei. The nuclear wave function \( \psi_{\text{nuc}}^R(R) \) describing this motion is, therefore, obtained by solving the Schrödinger equation for a Hamiltonian containing \( E_{\text{quantum}}^R \) as interaction term (see Figure 2). The overall molecular wave function is then given by \( \Psi_{\text{nuc,GSO}}(R) \times \psi_{\text{ele}}^R(r) \).

Whether one needs to solve either the time-independent or the time-dependent Schrödinger equation, either only for the electrons or also for the nuclei, depends on the biochemical process at hand. In all cases, the wave function is the solution of a differential equation in which all variables are coupled by the potential energy operator. This coupling correlates the motion of the quantum particles and renders the solution of the differential equation a daunting computational task for large molecules.

Why is the solution of many-body quantum problems such a major computational hurdle? We consider the electronic problem, although our conclusions will apply to the nuclear one as well. In order to represent it on hardware, either classical or quantum, the many-electron wave function must be discretized. This discretization is realized by expressing the wave function in terms of a finite set of one-electron functions \( \{ \phi_i(r, a) \}_{i=1}^{N_{\text{ele}}} \), the so-called molecular orbitals. Specifically, the wave function is expressed as a linear combination of antisymmetrized products of molecular orbitals as sketched in Figure 3. An antisymmetrized product is often referred to as Slater determinant or as electronic configuration. Methods to solve the Schrödinger equation within the resulting wave function ansatz without any restriction of the space of configurations are, in chemistry, referred to as full configuration...
interaction (full-CI), because the wave function is expressed as a linear superposition of all electronic configurations. In physics, this is known as exact diagonalization. The molecular orbitals are themselves discretized as a linear combination of a finite set of orbitals – usually referred to as atomic orbitals to distinguish them from the molecular ones. The atomic orbitals are basis functions that are pre-optimized to represent the orbitals of a given atom (and its polarization) accurately. They are expressed, in practice, as a linear combination of Gaussian basis functions. This facilitates the calculation of integrals that define the representation of the molecular Hamiltonian in the atomic orbital basis.

For a molecular system, the union of the atomic orbitals of all atoms defines the atomic basis in which the molecular orbitals are expanded. The expansion coefficients are called molecular orbital coefficients, which can then be optimized based on the variational principle. When the optimization is only carried out for a single Slater determinant, this yields the Hartree-Fock (HF) method.

In the limit of a dense, complete atomic basis set – a limit called the complete basis set limit – the many-electron wave function can be encoded exactly as a linear superposition of Slater determinants. The complete basis set limit cannot be reached in numerical simulations that must necessarily rely on finite, incomplete basis sets. Extrapolation techniques have been designed to tame this problem. These methods assume a finite, incomplete basis sets. Extrapolation techniques have been designed to tame this problem. These methods assume that the electronic energy \( E_{\text{ele}} \) depends on the basis set size \( n \) according to a given, fixed power law. One can then solve the electronic Schrödinger equation for increasingly larger basis sets, determine the power law coefficients by least-squares fitting, and estimate the complete basis set electronic energy \( E_{\text{ele}} \) from the limiting value obtained for \( n \to +\infty \). These methods are, however, inherently heuristic because the form of the power law must be fixed a priori.

Encoding an extremely complex high-dimensional wave function in terms of much simpler entities, the orbitals, comes with a price. Let us consider a molecule with \( N_{\text{ele}} \) electrons described by \( N_{\text{orb}} \) molecular orbitals. Let us also assume that \( N_{\text{orb}} \) is proportional to \( N_{\text{ele}} \). This is a reasonable assumption because, as highlighted above, \( N_{\text{orb}} \) is proportional to the number of atoms of the molecule that is, in turn, proportional to \( N_{\text{ele}} \). Since one can choose any one of the \( N_{\text{ele}} \) molecular orbitals to describe each electron, the number of configurations that can be constructed grows combinatorially with \( N_{\text{ele}} \). This implies that the computational representation of the wave function requires storing a number of coefficients that grows exponentially with the number of electrons. The resulting unfavorable scaling is often referred to as the curse of dimensionality and renders full-CI impractical on classical hardware already for molecular systems with more than three to five atoms.

The curse of dimensionality plagues any full-CI-based method aiming at solving quantum many-body problems in terms of one-particle functions. This hurdle is, therefore, not limited to time-independent electronic calculations. The same challenges will be faced for the solution of the time-dependent Schrödinger equation, both for the nuclear and for the electronic case. The form of the differential equation changes – the time-independent Schrödinger equation is an eigenvalue equation, while the time-dependent one is a partial differential equation that involves both the spatial and the time variables. Still, both equations are defined in terms of the Hamiltonian operator that couples the variables on which the wave function depends. Full-CI is, therefore, needed to describe the resulting quantum correlations completely, leading once more to the emergence of the curse of dimensionality.

Research in quantum chemistry has mostly focused on taming the curse of dimensionality with algorithms that trade the full-CI accuracy and generality for computational efficiency.
Specifically, they rely on wave function approximations which scale polynomially in the number of expansion parameters but are tailored to a specific regime of electron correlation. In quantum chemistry, the term electron correlation denotes an ensemble of effects that are not accounted for by the Hartree–Fock ansatz. This notion corresponds well with notions of correlation and entanglement in the context of quantum information theory, the number of necessary configurations corresponding to the Slater rank[69] – the Fermionic analog of the Schmidt rank quantifying quantum entanglement.[69] More detailed information about the weight components can be cast as eigenvalues of the one-particle reduced density matrix with naturally associated entropic measures and relation to Pauli’s principle.[61–64]

Consider the idealized case in which the Coulomb repulsion between the electrons is vanishing and, therefore, the electrons are not coupled. In this case, the full-CI wave function coincides with the Hartree-Fock determinant. Therefore, electron correlation effects are those that impedes encoding the wave function as a single determinant and require it to be expressed as a linear superposition of configurations.

Qualitatively speaking, the larger the number of configurations with non-negligible expansion coefficients, the stronger the correlation effects will be. Paradigmatic examples of strong correlation effects in molecules are transition metal complexes (as they occur in active sites of metalloenzymes and FeS clusters) and bond-breaking processes.[65–67]

In the former case, the manifold of d-type atomic orbitals of transition-metal active sites, and transition metal complexes in general, often contains near-degenerate orbitals that are partially occupied. The corresponding configurations will, therefore, have a similar weight in the full-CI wave function. Similarly, bond-breaking processes yield electronic configurations where unpaired electrons are localized on different parts of the molecule – a situation that cannot be captured by a single configuration in a singlet state because spin pairing will always produce two degenerate configurations with alpha and beta spin orbitals exchanged on the dissociating parts.

Moreover, weak electron correlation effects – referred to as dynamical correlations in quantum chemistry – must also be taken into account. Although they do not alter the qualitative structure of the full-CI wave function, they yield a non-negligible contribution to the electronic energy because of the sheer number of configurations with very small weights.

Dynamical correlations comprise two classes, short- and long-range correlations. From a physical perspective, long-range correlations are associated with the London dispersion forces, which, qualitatively speaking, describe the induced dipolar interaction between two molecular fragments that result from instantaneous polarization of the corresponding electronic density. Although they require, in principle, a proper quantum description of the electron-electron interaction, dispersion interactions can be well-approximated with effective pairwise classical potentials.[66]

Short-range correlations are, instead, associated with the short-range electron-electron interaction, i.e., in the limit of a vanishing interelectronic distance in which the positions of two electrons coalesce. It can be proven that, at the coalescence of two electrons, the many-body wave function displays a cusp (the so-called Kato cusp), i.e., its derivative is not continuous. Attempting to fulfill the Kato cusp condition with products of cupless Gaussian basis functions yields highly non-compact full-CI wave functions, with many configurations with small but non-negligible coefficients – the signature of dynamical correlations. Accordingly, the basis set expansion of the wave function is often supplemented by factors that depend explicitly on the interelectronic distance[69]

4. Computational foundations: Traditional approaches

The curse of dimensionality limits practical applications of full-CI to molecules with approximately 20 electrons,[65,70] a size that characterizes very small molecules, far smaller than any biomolecule. The impossibility of routinely applying full-CI to large molecules has sparked the development of approximate methods not suffering from the curse of dimensionality. The key idea behind many of these approximations is to identify wave function parametrizations that can approximate full-CI while being defined by a number of parameters that scales polynomially with the system size. These approximate parametrizations have been tailored to two idealized situations: weakly correlated and strongly correlated electronic structures of molecules. Unfortunately, strong and weak correlation effects often coexist in a molecular system and, therefore, these methods must be combined in accurate molecular simulations.

Above, we defined weakly correlated wave functions as those in which the Hartree–Fock configuration has a predominant weight (given as the square of its expansion coefficient in the expansion of configurations). Since the wave function is normalized to a fixed value (usually one), all other determinants will have a comparatively small weight. Moreover, it is reasonable to expect that the weight of the other configurations will decrease as their departure from the Hartree-Fock determinant increases. This departure can be measured in terms of so-called excitations (which are orbital substitutions see Figure 4): a determinant will be defined as a \(N\)-fold excited determinant if it differs from the Hartree-Fock reference one by at most \(N\) orbitals. We note that the term excitation is misleading, despite having become a standard notion, as it implies that an orbital substitution corresponds to an electronically excited state, which is only true for non-interacting electrons and obvious when considering that the ‘excitations’ are still elements of a many-electron determinantal basis set into which the ground-state wave function is expanded in the configuration interaction ansatz. Based on this idea of orbital substitutions, the full-CI wave function can be systematically approximated by including only up to \(N\) \(N\)-tuple substitutions in the wave function expansion. This leads to so-called truncated CI approaches, whose computational complexity scales polynomially with the number of orbitals, with an exponent proportional to \(N\). We note that the emerging
Many-body wave function
\[ \Psi(r_1, \ldots, r_{N_{\text{ele}}}) \]

Superposition of configurations
\[ \Psi = C_0 |\Psi_0\rangle + \sum_i C_i^n |\Psi_i^n\rangle + \sum_{ij} C_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + \cdots \]

Full CI

Occupied
\{ \varphi_1(r), \varphi_2(r), \ldots, \varphi_{N_{\text{orb}}}(r) \} 

Unoccupied

One-particle functions

Reference configuration

| \varphi_1 \cdots \varphi_i \cdots \cdots \varphi_{N_{\text{ele}}} \rangle = |\Psi_{i,j}^{ab}\rangle 

Double substitution

| \varphi_1 \cdots \varphi_i \cdots \cdots \varphi_{N_{\text{ele}}} \rangle = |\Psi_i^n\rangle 

Single substitution

| \varphi_1 \cdots \varphi_j \cdots \cdots \varphi_{N_{\text{ele}}} \rangle = |\Psi_0\rangle 

Figure 4. Graphical representation of configuration interaction-based wave function approximations (illustrated for the electronic wave function of butadiene). Configuration Interaction methods encode the many-body wave function (upper part of the figure) in terms of one-particle functions, known as orbitals in electronic structure theory (bottom left part of the figure). Molecular orbitals may be optimized with Hartree-Fock or Density Functional Theory and are not scale linearly with the number of electrons, that it is not size extensive, included in the truncated expansion of configurations. The key limitation of truncated CI (by contrast to full-CI), however, is that the number of negligible expansion coefficients \( C_{i,j}^{ab} \text{sub} \) which they differ compared to the reference determinant. Configurations differing by a single spinorbital are referred to as ‘single excitations’ (represented in gray), while the ones differing by two spin orbitals are referred to as ‘double excitations’ (represented in yellow). Truncated CI approximates the full-CI wave function by including only all those configurations up to \( N_{\text{orb}} \)-fold excitations. For \( N_{\text{orb}} = 1 \), one obtains the configuration interaction singles (CIS) approximation, whereas for \( N_{\text{orb}} = 2 \) the method is referred to as configuration interaction with singles and doubles excitations (CISD). If \( N_{\text{orb}} \) coincides with the overall number of electrons of the molecule, one retrieves full-CI. Note that the general notation for the expansion coefficients \( C_{i,j}^{ab} \text{sub} \) in Figure 3 is here replaced to follow the partitioning according to excitation classes as \( C_1, C_2, \ldots \)

Hierarchy of approximations is an ad hoc choice that does not guarantee that all truly important configurations are actually included in the truncated expansion of configurations. The key limitation of truncated CI (by contrast to full-CI), however, is that it is not size extensive, i.e. the corresponding energy does not scale linearly with the number of electrons, which makes it impossible to reliably compare the energy of molecules with different numbers of electrons.

The established canonical solution to this problem is the coupled cluster (CC) method. CC relies on an exponential wave function ansatz that contains a number of parameters that scales polynomially with system size. Specifically, CC with singles, double, and perturbative triple excitations, abbreviated CCSD(T), has become the de facto reference method for calculating weakly correlated electronic wave functions in order to reach chemical accuracy of about 1 kcal/mol.

Approaches such as truncated CI or CC – usually referred to as single-reference methods, because they rely on a single configuration, the Hartree–Fock one, as a reference – can become inaccurate as soon as the full-CI wave function is not dominated by a single configuration. For such multi-reference cases, modern full-CI-type methods have made it possible to push the limits of full-CI to systems with approximately 100 electrons. These methods can be qualitatively divided into two classes: selected configuration interaction methods and algorithms based on tensor network states.

Selected configuration interaction methods rely on the assumption that, although not being dominated by a single configuration, the full-CI expansion remains highly sparse so that the number of negligible expansion coefficients \( C_{i,j}^{ab} \text{sub} \) (see Figure 3) is in overwhelming majority. Full-CI Quantum Monte Carlo (FCIQMC) exploits this idea by solving the Schrödinger equation via a stochastic Monte Carlo-based realization of the imaginary-time evolution algorithm. Deterministic variants of FCIQMC are also available, that originate from the “Configuration Interaction using a Perturbative Selection made Iteratively” (CIPSI) method introduced by Malrieu that identifies the non-negligible configurations based on second-order perturbation theory. In practice, selected CI methods can target up to 100 electrons, which is a huge leap forward for chemical applications, but still represents a severe limitation for generally large biomolecules. Surely, other approximate quantum approaches that do not rely on expansions of configurations – most importantly contemporary methods of Density Functional Theory (DFT; see below) – can target systems that are two orders of magnitude larger, but they also come with an uncontrollable error for a specific biomolecule under consideration.
An alternative route is offered by tensor network states (TNSs). In essence, a tensor network approximates the tensor $C = C_{ij...kl}$ of all full-CI expansion coefficients (cf. Figure 3) with tensor compression algorithms. This tensor compression reduces the number of wave function parameters but is inevitably associated with a loss of accuracy. However, TNSs can be designed to yield an optimal balance between compression degree and accuracy. In many numerical studies it turned out that a specific class of TNSs, so-called matrix product states, reliably approximate the electronic Schrödinger equation, whose parameters can be very efficiently optimized with an iterative algorithm known as density matrix renormalization group (DMRG) approach. Although we have focused so far on the key problem of molecular quantum simulations, namely the solution of the time-independent electronic Schrödinger equation (cf. Figure 2), we note that the approaches discussed so far can be generalized to the nuclear problem as well (see, for example, Ref. [93] for the introduction of vibrational DMRG) and to the time domain to solve time-dependent Schrödinger equations.

Even for complex molecules with several dozen atoms, electronic correlation is rarely strong for more than 100 electrons. This has inspired the development of so-called active-space algorithms: a full-CI-based scheme is applied only on the strongly correlated orbitals, while the remaining weak correlation effects are captured \textit{a posteriori}. Translating such an apparently simple idea into an algorithm is a hard task. In fact, designing algorithms to include accurately weak correlation effects on top of a DMRG or FCIQMC simulation remains an open challenge of computational chemistry. One should not forget that the second hurdle of active space-based methods is the need of identifying the strongly-correlated electrons \textit{a priori}, without knowing the exact full-CI wave function. In this respect, various strategies have been proposed to efficiently identify strong correlation effects based either on single-reference calculations or on partially converged full CI-type simulations.

Finally, we emphasize that DFT represents a route to solve the electronic structure problem that has become the workhorse of computational molecular quantum mechanics. The key difference between a superposition of configurations and DFT is that DFT calculations in practice rely on the Kohn-Sham variant, where the density $\rho(r)$ is obtained from a single determinant wave function in such a way that it can be calculated as a sum of absolute squares of the orbitals entering the determinant. The key insight here is that one can obtain the exact ground-state electron density of an interacting quantum system from this determinant, which describes exactly a non-interacting surrogate system of fermions. The orbitals are obtained by solving a Hartree-Fock-like set of equations that is defined in terms of the so-called exact density functional $E[\rho(r)]$. While this procedure is exact, the exact form of $E[\rho]$ is, unfortunately, not known. Practical DFT calculations rely on heuristic forms of $E[\rho]$ that are either data-driven or inspired by the form of the functional obtained for simple model systems. Estimating the accuracy of approximate functionals is often impossible, which prevents quantifying and systematically improving the accuracy of a DFT calculation, which is the major drawback of contemporary DFT.

We pause at this point for the conceptual observation that the ground state energy can be written in variational terms as

$$E_g = \min \langle \psi | H | \psi \rangle$$

(3)

and that methods that optimize over an ansatz (variational methods) such as Hartree-Fock, full-CI, DFT, restrict the minimization and always lead only to an upper bound on the true ground state energy. Whereas efforts can be made to assess the closeness to the true ground state energy (e.g. extrapolation to an infinite basis set in the case of full-CI), there can ultimately not be any rigorous closeness estimate of such methods (as the infinite basis set limit cannot be reached in practice) or the ansatz within a given basis set only samples a subset of the states (e.g. Hartree-Fock, CC, …). It was already noted in 1960 by Coulson\textsuperscript{100} that rigorous lower bounds could be obtained by the observation that molecular Hamiltonians only involve pairwise interactions $H = \sum h_{ij}^2$ and that, therefore, the ground state energy can be cast as an optimization over the two-particle reduced density matrix (2-RDM) $\rho^{(2)}$

$$E_g = \min \frac{N}{2} \text{Tr}(h^{(2)} \rho^{(2)}).$$

(4)

The function to be optimized is linear and the set of 2-RDMs is convex and only of polynomial dimension in the basis set, sparking hope for a solution to the problem. The earlier discussed difficulty of the ground state energy problem is now hidden in the complexity of describing this convex set. The concrete question is: Given a matrix $\rho^{(2)}$, is it the reduced density matrix of an $N$-particle electronic wave function? This problem is known as the N-representability problem of quantum chemistry and solving it as Coulson’s challenge.

Whereas the exact characterization of the set has been shown to be Quantum Merlin Arthur (QMA)-hard (QMA is the quantum analog of the famous class of computational problems known as nondeterministic polynomial time (NP)) and thus unlikely to be solved even on a quantum computer,\textsuperscript{102} we note that any outer approximation to the set will result in a lower bound on the energy, in stark contrast to the situation discussed above for the variational methods.

Concretely, such approximations can be given by semi-definite constraints (rendering the optimization problem a semi-definite program) on the set and a number of nontrivial constraints have been found and, in general, significant efforts have been devoted to this problem.\textsuperscript{59–64,104–106} For chemical problems, unfortunately, the rigorous lower bounds obtained with this method have not been as close from below to the believed true ground state energy as the variational methods from above. The success of such methods has therefore been limited in practice.
Many of the methods outlined so far target the lowest-energy solution of the electronic Schrödinger equation (see Figure 2) which corresponds to the electronic ground state. The ground-state energy is the key quantity for reactions that happen on the electronic ground-state PES (see Figure 1). However, some biological processes are driven by the absorption of electromagnetic radiation. In this case, the reaction involves the surface of an electronically excited state, as shown in Figure 1, which corresponds to higher-energy solutions of Eq. (2). These solutions are not only important to describe a biomolecular system after light excitation, but they also offer information that can be related to spectroscopic techniques and, hence, to experimental spectroscopies that identify specific states of biomolecular machinery. Many of the methods reviewed above can be straightforwardly generalized to target excited states – although these generalizations are often harder than their ground-state counterparts.

Molecules can be subjected to spectroscopic techniques through their response to external electromagnetic perturbations (i.e., light). The spectroscopic response of a molecule does not depend only on its energy, but also on its properties, such as the electric and magnetic dipole moment. Response properties can be calculated based on the methods listed above with time-dependent perturbation theory.[107]

Identifying the method providing the best compromise between cost and accuracy for time-dependent simulations is an especially hard task. When driven out of its ground state, the nature of the wave function and, therefore, the correlation between its particles may change, even dramatically. It is known that the time evolution of a wave function following an external perturbation (usually known as quench in quantum physics, which corresponds most often to an electromagnetic perturbation in chemistry) may lead to a continuous growth of its entanglement. This makes it difficult to accurately represent the time-evolving wave function as a TNS over the whole propagation. For example, the entanglement growth makes the MPS representation of the wave function increasingly inaccurately with increasing time. How this loss in accuracy affects the simulation accuracy depends on the simulation target at hand.

5. Computational foundations: The promise of quantum simulation

It was Feynman[108] who noticed that using a classical system, e.g. a traditional computer, to simulate a quantum system will in general result in an exponential overhead in the time required for the simulation due to the curse of dimensionality, but that building a simulator (or computer) directly from quantum mechanical constituents (henceforth called a quantum simulator or, if very versatile, quantum computer) could circumvent this curse and therefore achieve an efficient simulation. With efficient simulation we mean a simulation that requires time and space resources that scale proportionally (or at most polynomially) with the size of the biomolecule to be simulated. Analogous to the bits of a computer, the constituents in a quantum computer are most naturally cast in terms of quantum bits (qubits), quantum mechanical two-level systems. Qubits in a quantum computer are combined with the tensor product operation. The idea to build a quantum system to simulate another quantum system – that quantum simulates quantum – is as self-evident and simple as it is revolutionary.

The technical reason that hinders classical simulators at the possibility of simulating quantum systems is the different composition rules of its elementary components. In fact, the classical composition rule is weaker than the quantum mechanical. This means, in practice, that only a subset of the possible quantum mechanical states of \( N \) qubits can be represented by \( N \) of their classical counterparts. This is precisely remedied by building a simulator directly from quantum mechanical constituents, since those evidently adhere to the same quantum mechanical composition rule and, therefore, offer a computational platform perfectly suited to the task of simulating a quantum system.

By simulating a quantum system we mean, just as before, simulating relevant properties of quantum systems – the most relevant being, in chemistry and biology – molecular properties. Primary examples of relevant quantities to be simulated are, for static phenomena, ground- and excited-state energies and, for non-equilibrium processes, time-dependent correlation functions. Simulating quantum systems is not the unique application of quantum computing. Other computational problems that are relevant for physics, chemistry or biology and that are hard to solve on a classical computer may be solved more efficiently on quantum computers. A candidate in the biological context is the protein folding problem, which is an optimization problem, for which heuristic quantum optimization algorithms are being investigated.[109,110]

It is important to emphasize that, although quantum simulates quantum is such a self-evident slogan, the existing quantum simulation algorithms are highly non-trivial and problem-dependent. Their future development will therefore require ingenuity and insight both from the general principles of quantum algorithmic design (mathematics, physics, and computer science) and from the application side (physics, chemistry, biology).

Quantum simulators naturally fall into two categories, traditionally called analog and digital quantum simulators. Analog quantum simulators are more akin to purpose-built machines that through their build-up mimic the interactions and, therefore, the behavior of the system to be simulated. When equipped with knobs to modify the behavior of the simulator they can have aspects of universality.[111,112] Yet, they are meant as purpose-specific computing machines and, therefore, are not meant to be universal computing machines. Mimicking the interactions of a quantum system with an analog quantum simulation is often hard due to the fact that the many-body interaction terms of a given Hamiltonian (for instance, the electronic Hamiltonian of Figure 2) may not be directly encoded on the hardware at hand. This limit can be overcome with so-called gadgets that map, via perturbation theory, the desired interactions of the system to be simulated to effective interactions built from those available in the
hardware. The price to be paid is a slight increase, usually a multiplicative constant, of the hardware size needed to simulate a given quantum system.

The amazing experimental progress, especially in the area of optical lattices, where many atoms are trapped and interact in a lattice of standing light waves, has delivered remarkable simulation results of condensed matter systems surpassing the practical simulation capability of traditional computers.\[113,114\] Yet, since the constituents of analog quantum simulators also experience decoherence – a quantum form of errors – which cannot be suppressed arbitrarily (this is true at present, and may remain so even in the far future), formal guaranteed statements about the overall accuracy of the achieved simulation are difficult if not impossible. Control in such a system is continuously being improved, yet, in general, experimental analog quantum simulators are characterized by limited control and accuracy, but have a relatively easily scalable number of constituents. The limited control constitutes a major challenge when simulating (bio)chemical processes with digital simulators. The key difference between Hamiltonians relevant in quantum physical problems and those appearing in chemical and quantum biological contexts is that the (discretized) interaction is drastically more irregular in the latter case than in the former one. Tunability is, therefore, a key requirement of an analog quantum simulator for quantum chemical and quantum biological simulations.

In a nutshell, analog quantum simulators are single-purpose machines with an easily scalable size designed to mimic the behavior of interesting quantum systems. Affected by decoherence on the individual-component level and without complete error-correcting and fault-tolerant mechanisms in place, however, few rigorous computational guarantees are available.

Digital quantum simulators are, by contrast, based on the gate-based model of quantum computation in direct analogy to the gate-based model of standard (classical) computation (see the upper part of Figure 5). Here, a quantum algorithm is cast as a sequence of quantum gates, i.e., operations acting on quantum bits (qubits). Qubits are the analog of the bits in a standard computer, but can, in addition to the states 0 and 1, also assume a quantum superposition of both, in the same way that an arrow cannot only point up or down but also in any other direction. In a physics context, qubits are also known as two-level systems and can often be viewed as a spin-\(\frac{1}{2}\) system. Importantly, the composition rule of quantum particles, and

---

**Figure 5.** Graphical representation of the steps involved in the simulation of a molecular process with digital and analog quantum simulation. For a given biochemical target (such as the LHC-II complex in the upper left part of the figure), one first chooses the Hamiltonian \(H_{\text{mol}}\) that describes the process at hand. For LHC-II, this could be the excitonic Hamiltonian describing the excitation energy transfer process. Assume we aim at solving the time-dependent Schrödinger equation (box in the left part of the figure) for a given initial wave function \(\Psi(0)\). In digital quantum simulations (upper part of the figure), one first maps the target quantum system to a set of \(N\) qubits (in the figure \(N = 5\)), and maps \(H_{\text{mol}}\) onto a corresponding qubit Hamiltonian. A sequence of gates is then defined to solve the time-dependent Schrödinger equation. In the figure, we selected the specific case of the Trotter algorithm for simulating quantum dynamics (see Eq. (5)). In analog quantum simulations (lower part of the figure), the quantum system is first mapped onto a set of interacting quantum entities (such as, for instance, atoms trapped in arrays of optical tweezers). The interaction between these entities is then tuned (for instance, with a laser, as represented in the figure) such that the Hamiltonian describing these interactions \(H_{\text{sim}}\) mimics the original Hamiltonian \(H_{\text{mol}}\). The quantum dynamics is then simulated directly on the resulting simulator, without the need of designing any algorithm for solving the time-dependent Schrödinger equation. The result of a simulation is a time-dependent property \(\langle O \rangle(t)\), as represented in the right part of the figure. The analog quantum simulation does not admit rigorous approximation bounds yet aims to give an accurate qualitatively reliable simulation for a long propagation time. The digital simulation, by contrast, will track the property with high accuracy, yet only for a shorter period of time due to the stringent hardware requirements of fault-tolerant quantum computation.
hence also of quantum bits, posits that the state of \( n \) qubits requires the description with \( 2^n \) probability amplitudes (i.e., expansion coefficients; see Figure 3) describing the precise superposition of the strings of \( n \) bits (thereby making up the ‘state’ or ‘wave function’ of the \( n \) qubits). Simulating or even just storing the amplitudes on a standard computer, therefore, requires more than \( 2^n \) bits and therefore an exponential overhead (the curse of dimensionality eluded to earlier).

Whereas the storage of classical bits is stable, as the ‘0’ or ‘1’ is copied many times over, for instance, in the many spins of a magnetic patch storing a bit on a hard drive, the situation is markedly different in quantum computation. Quantum information cannot be copied and is therefore much more fragile and affected by quantum noise, the previously mentioned decoherence. Consequently, storage and gate operations on quantum bits are prone to errors that need to be corrected. Note that quantum error correction, first put forward by Shor in 1996,\(^{[115]}\) alone is not sufficient, because the quantum circuits carrying out the error correction may themselves introduce new errors. Luckily, techniques have been developed that replace each noisy physical qubit and physical gate in a circuit with an encoded logical qubit and logical gate, thereby allowing an entirely noise-free quantum computation to be simulated on noisy quantum hardware (given that the noise is below a threshold – the ‘threshold theorem’).\(^{[116]}\) In addition to the quantum hardware and quantum algorithms, quantum computers therefore also require an important error correction layer in between. That fault-tolerant quantum computation is possible at all is mathematically highly non-trivial, but allows us to imagine that digital quantum simulation is possible in the future, despite decoherence possibly staying around indefinitely.

There is a price to be paid for the fault-tolerant hardware in terms of the overhead required. Whereas recently, it has been shown that only a constant overhead is necessary (i.e. the number of logical qubits is proportional to the number of physical ones)\(^{[117]}\) and one may therefore generally see no fundamental obstacle to building a perfect digital quantum simulator in the future, the number of physical qubits required for the first viable algorithms to run is extremely large and therefore not expected to be reached in the near future, but only in the medium or long term. We note that one may hope for novel hardware solutions that lead to a lower overhead due to stronger protection of the native qubits to noise, e.g., through topological protection.\(^{[118,119]}\)

Therefore, research on quantum algorithms splits, on the one hand, into algorithms for fault-tolerant quantum computers, which we may think of as a future perfect quantum computer, and, on the other hand, into algorithms for near-term or noisy-intermediate stage quantum (NISQ)\(^{[120]}\) devices of dozens to thousands of physical qubits, that are characterized by the absence of complete fault-tolerance.

Naturally, the algorithm design focuses on rigorous runtime bounds for the fault-tolerant machines and heuristic arguments for NISQ devices. Common to problems from both chemistry and molecular biology is the necessity of mapping the representation of electrons onto the qubit structure. This mapping is often done with the Jordan-Wigner\(^{[121]}\) or the Bravyi-Kitaev transform,\(^{[122,123]}\) but further algorithms-tailored methods are expected to be developed in the future. Common to all simulation algorithm designs is the drive to the most favorable scaling of simulation time with the size of the simulated system in order to make actual quantum computations feasible.

An important class of algorithms for NISQ hardware are so-called hybrid quantum-classical algorithms that partition the computational workload between a classical and a quantum computer.\(^{[124,125]}\) The most prominent example of this class of algorithms relevant for electronic structure calculations is the variational quantum eigensolver (VQE), which presents a variational (and thus heuristic) algorithm to upper bounding the ground state energy problem.\(^{[126–128]}\) In a nutshell, a given wave function ansatz is represented with a parametrized quantum circuit (i.e. a quantum circuit, where the individual gates are dependent on parameters), and its energy is estimated with a term-by-term measurement.

As we have discussed, there exist wave function ansätze inspired by the ones described in the previous section for which the cost of measurement would grow exponentially on classical hardware and polynomially on quantum hardware, therefore allowing to circumvent the exponential overhead. However, VQE yields the exact solution to the full-CI problem only if the final wave function can be represented with that ansatz. The parameters of the wave function ansatz are then optimized in an outer loop on a classical computer taking the quantum energy measurements as input in order to obtain a variational upper bound to the true ground state energy.

Other examples of hybrid classical-quantum algorithms are the variational quantum time evolution for quantum dynamics simulations\(^{[129,130]}\) and the variational imaginary-time evolution,\(^{[131]}\) which represents an alternative to VQE for ground state energy optimizations.

On a fault-tolerant computer, one may directly measure the full-CI energy by first preparing a wave function ansatz on the quantum hardware (as in VQE) and then applying the Quantum Phase Estimation algorithm, which projects a prepared state into an energy eigenstate of the Hamiltonian.\(^{[131,132–136]}\) The ground state will be found with high probability after \( O(\frac{1}{\epsilon}) \) repetitions, where \( \epsilon \) is the overlap between the prepared state and the exact ground-state wave function. Note, however, that the ground state energy problem is known to be QMA-hard and it might, therefore, be expected that no general exponential speedup will be obtained, rendering this again a heuristic algorithm. In practice, however, for those electronic structure problems that are dominated by dynamic correlation (see above), the Hartree-Fock state will be a reliable initial state with sufficient overlap with the final state, as this has been the key insight into the development of single-reference approaches such as coupled cluster for decades. In the presence of strong static correlation, the Hartree-Fock state may not have a sufficiently large overlap with the final wave function (see Ref. [137]). However, in these cases, an amazing qualitative and quantitative understanding of good candidates for functions (such as the FCI-type wave functions optimized by DMRG or FCIQMC; see above) has been assembled that approximate such
strongly correlated systems well so that there is hope that the overlap × indeed can be made substantial.

For Quantum Phase Estimation, one is required to perform a time evolution with a Hamiltonian that describes the (bio)molecule under study with parameters that have been calculated as integrals in a so-called four-index transformation from the atomic orbital basis set to the molecular orbital set. Since the Hamiltonian consists of a sum of terms, the basic idea is to use a Lie-Trotter formula in order to be able to decompose the generally complicated time evolution into gates that only touch two particles at a time. For a Hamiltonian with two terms \( H = H_1 + H_2 \), time-evolved for time \( t \), this yields the following approximation

\[
e^{-iHt} = e^{-iH_1t}e^{-iH_2t}
\]

\[
\approx \left( e^{-iH_1 t/N} e^{-iH_2 t/N} \right) \ldots \left( e^{-iH_1 t/N} e^{-iH_2 t/N} \right),
\]

where \( N \) is the total number of terms. Note that it is a sequence of shorter propagations each time only involving one of the terms in the Hamiltonian. By now, significant improvements on this initial decomposition idea have been devised and an excellent, indeed asymptotically optimal, understanding has been obtained \(^{[138,139]} \). Yet, significant further development is needed, especially in regard to concrete and tailored approaches to simulation of quantum processes.

Finally, we emphasize that whether analog or digital, fault-tolerant or NISQ, it is important to realize that asymptotic runtime bounds will only be a guide to the performance and practical resource estimates will have to be obtained. These bounds will additionally depend strongly on the available hardware, e.g., on the number, the connectivity, and the control of the qubits, on the gate set, and the type of noise. To make practical use of quantum computation for molecular biology (or molecular simulations in general), all of these aspects of the available hardware need to be combined with an in-depth knowledge of the molecular system(s) at hand and, therefore, might require highly specialized knowledge and expertise ranging from molecular biology to the understanding of the state of the quantum hardware to eventually the quantum algorithm design.

In the remaining parts of the manuscript, we will detail the prospects of quantum computation in several distinct areas of molecular biology.

6. Local quantum effects in biochemical processes

We will now turn our attention to (local) quantum mechanical phenomena in molecular biology, which mostly relate to chemical reactions of biomolecules in the most general sense (not only covering stoichiometric or enzymatic reactions, but any kind of chemistry of biomolecules, even involving solvation and association processes). Chemical processes have been recognized as a potential high-impact application of future fault-tolerant quantum computers \(^{[140,141]} \). Accordingly, chemical reactions in biomolecular systems are an obvious target for quantum computations with significant advantages under certain conditions. In fact, such a scenario has been first elaborated in detail for the mode of action of the enzyme nitrogenase \(^{[140]} \) and was further expanded on for other catalytic processes with improved quantum algorithms \(^{[142]} \).

The key feature exploited in this context is the spatial locality of chemical reactions; i.e., they involve comparatively few atoms that are involved in dissociation and association (bond-breaking and bond-formation) processes. The huge rest of the biomolecular machinery is treated as a spectator environment in a given chemical reaction step. This environment can be efficiently described in a classical embedding scheme, an idea that won the Nobel Prize in Chemistry in 2013 \(^{[14] \}. \) As a result, the complex environment can be efficiently modeled and only the local quantum description of the reaction event remains the major challenge: the biochemical problem has been reformulated as a purely chemical problem, and all developed solution approaches in the chemical context automatically transfer to the biochemical application.

In reaction chemistry, these processes are all related to changing the connectivity of atoms in molecules, i.e., to changing covalent bonding in some way. Examples are enzymatic reactions, but also any kind of chemical reaction that carries biological significance. Hence, it starts already with the creation of the first biomolecules (abiogenesis) and the origin of life. Accordingly, quantum computation promises to advance not only the accuracy of quantum predictions for enzyme kinetics, especially for strong correlation cases as they are found in metalloenzymes with 3d metal-atom active sites such as copper and iron-sulfur enzymes (including electron and proton transfer processes), but also for stoichiometric reactions and biochemical pathways in living cells.

Naturally, any chemical reaction network of relevance to molecular biology can benefit from quantum computation. Intellectually appealing examples may even be taken from prebiotic chemistry (see, for example, the formose reaction network \(^{[142,143]} \)) where (traditional and quantum) simulations may provide insights into prebiotic potentialities that have not been scrutinized by experiment yet. Quantum computation might even deliver accurate electronic energy estimates for reactive biomolecular precursors on surfaces that act as catalysts for the production of the first (functional) biomolecules (cf. the Waechtershaeuser hypothesis \(^{[140,143]} \)).

Even more relevant would be the application of quantum computing to simulate reactions that are difficult to reproduce in a laboratory setting. An example are astrochemical reaction networks emerging under the extreme conditions of extraterrestrial environments. In the context of molecule formation in interstellar space, often the lack of an environment (producing conditions far away from thermal equilibrium) will require high-accuracy rovibronic quantum-state resolved formation theories and corresponding calculations. The high accuracy needed to reliably predict the outcome of these reactions, combined with the relatively small size of the
intermediates entering astrochemical reaction networks, makes astrochemistry\textsuperscript{[142–146]} an appealing target for quantum computing. In the context of astrobiology, biomolecules (such as amino acids) on surfaces of interstellar objects (exoplanets or asteroids) subjected to the light of stars may well fall into the category of approaches applicable to pre-biotic scenarios on Earth.\textsuperscript{[150]} Moreover, computer simulation can even adjust the conditions to account for the macroscopic constraints such as increased temperature and pressure. Hence, within this context, potential routes leading to the emergence of life in outer space could also be scrutinized based on quantum computing simulations.

In the physical description, chemical processes are mapped to a path across the Born-Oppenheimer PES (cf. Figure 1) so that the calculation of the electronic energy as a solution of the electronic Schrödinger equation is the key first step (cf. Figure 2). Accordingly, the accurate quantum mechanical calculation of the electronic energy for a given arrangement of the atomic nuclei in the biomolecules under consideration suffers from the curse of dimensionality through the large number of orbitals required for accurate results.

In the past decades, a zoo of practical solution methods has been devised for traditional computing hardware (see Section 4). However, every method strikes a compromise between feasibility and accuracy: the more accurate the energies (and, in particular, the energy differences) need to be, the smaller the systems (in terms of the number of atoms or electrons) are for which an accurate method is feasible on some traditional computer hardware. Still, these traditional approaches have reached a remarkable degree of sophistication, accuracy, usefulness, and acceptance.\textsuperscript{[151–153]} Detailed resource estimates\textsuperscript{[13,140]} showed that algorithms based on quantum phase estimation are likely to compete and beat the best traditional approaches for cases of strong electron correlation as they occur in active sites of metalloenzymes (and in electronically excited states; see below). Their main advantage though is the explicit control over the error bar on the result (provided that the initial state prepared features sufficient overlap with the final state; see discussion in the previous section), which is not available from any traditional approach and also not from the currently most favored algorithm for NISQ devices, namely, VQE with a unitary coupled cluster ansatz because of limitations on the underlying parametrization.

7. Global quantum effects in functional biomolecules

We highlighted in the previous section that a key feature of quantum effects associated with the rearrangement of nuclei and electrons that drive biochemical reactions is their locality in space – i.e. they are relevant only for the reacting nuclei and their electrons (and for a comparatively tiny fraction of their environment). Other biomolecular processes are, instead, influenced by quantum effects acting on a larger spatial scale. For instance, large amplitude motions of solvated proteins are, at their very heart, governed by dynamic electron correlation effects. While these can often be well modelled by classical surrogate potentials, this is not always the case. Two prominent examples are light harvesting and magnetoreception (see Figure 6). Light harvesting represents the initial step of photosynthesis, i.e. the process exploited by plants to convert light into chemical energy. Photosynthesis proceeds via a chain of complex chemical processes, the first of which is light harvesting, which corresponds to the absorption of visible light by so-called pigment-protein complexes.\textsuperscript{[154–156]} Magnetoreception\textsuperscript{[157]} is the process that enables birds to sense weak magnetic fields and, in particular, the Earth magnetic field that can be used as a guide then. It has been argued that quantum effects play a key role in both phenomena.

In contrast to the chemical reactions described in the previous section, quantum effects span, in this case, the whole molecule. These phenomena are the subject of study of so-called quantum biology.\textsuperscript{[159–165]} We prefer not to use this term because the effects described in Section 6 have also a quantum nature, but are usually not referred to in the realm of quantum biology.

Pigment-protein complexes encapsulate several chromophores, usually porphyrin derivatives. The light is first absorbed by the chromophores. The resulting electronic excitation is then transferred, via a chain of excitation energy transfer (EET) processes, to the reactive center that converts it into chemical energy.\textsuperscript{[159–160]} The overall efficiency of photosynthesis depends critically on whether the electromagnetic energy of the absorbed light can be transferred to the reaction center without losses. Multidimensional electronic spectroscopy measurements of light-harvesting systems have revealed the presence of quantum coherences (i.e., that the electronic wave function is a coherent superposition of excited states centered on each chromophore) in the early phase of the EET process.\textsuperscript{[161–163]} This observation, combined with the fact that quantum mechanical models predict a higher efficiency compared to semiclassical calculations, has led to the hypothesis that light harvesting is driven by quantum effects. Although it was later observed that the lifetime of electronic coherences may not be long enough to have an effect on the EET rate,\textsuperscript{[164–166]} it has been argued that molecular vibrations, through their coupling with the electronic excitation (usually referred to as excitonic coupling), may enhance the coherence lifetime by quantum coherences.\textsuperscript{[166–168]}

Similar observations have been made for magnetoreception. According to the currently most favored hypothesis, magnetoreception relies on the generation of a radical pair in so-called cryptochromes.\textsuperscript{[169,210]} The latter are proteins located, in particular, in the bird’s eyes, where they can absorb visible light. Light absorption triggers a chain of electron transfer reactions that generates two unpaired electrons (the radical pair) localized on a flavin adenine dinucleotide and a tryptophan residue of the protein. The spin state of the radical pair can be either singlet or triplet, and an external magnetic field can affect the relative amount of these two states. Since the rate at which the radical pair decays to the electronic ground state via spin recombin-
is believed to be at the origin of magnetoreception.\cite{169,170} Semiclassical simulations predict a radical pair lifetime that is much shorter than that obtained with quantum-mechanical models.\cite{171}

As for light harvesting, this has been taken as an indication that quantum mechanical effects play a key role in magnetoreception. For both phenomena, quantum mechanical effects affect the interaction between molecules – the chromophores, for light harvesting, and the unpaired electrons, for magnetoreception. Simulating quantum effects happening on such large spatial scale is much harder than for chemical reactions. The electronic wave functions of these systems depend on thousands of electrons, and this makes applying the theory described in Section 3 extremely challenging for both classical and quantum computations. As a way out, effective Hamiltonians (usually referred to as excitonic Hamiltonians) have been designed that largely reduce the problem complexity but still encode the key quantum mechanical effects that drive these biological phenomena.

In these Hamiltonians, the chromophores are approximated as quasiparticles, the excitons, with two quantum levels each, one corresponding to the electronic ground state and the other one to the excited state. The inter-chromophore interaction is then approximated with effective potential energy terms. For the radical pair dynamics, each electron of the radical pair is described by two spin operators, $\hat{S}_1$ and $\hat{S}_2$, that interact with the Earth magnetic field $\mathbf{B}$, among themselves via a dipolar spin-spin interaction ($\mathbf{D}$ is the dipolar interaction tensor, whereas $\hat{I}_k$ is the unit matrix), and with the nuclear spin (referred to as $\hat{I}_k$ for the $k$-th nucleus). Note that a system as large as the cryptochrome contains several dozens of nuclei with non-zero spin.

Figure 6. Key steps involved in the modelling of photosynthesis (left part) and magnetoreception (right part; “Biological function”). First, the biomolecule that tunes the target biological function is identified (“Biochemical system”): a light-harvesting complex for photosynthesis and a cryptochrome protein for magnetoreception. Second, the “Physical process” linked to the biological function of the molecules is identified. For photosynthesis, this is the excitation energy transfer (EET) that enables to funnel the energy captured by the chromophores through light absorption to the reaction center. For magnetoreception, the key physical process is the generation of the radical pair in the FAD-tryptophan complex inside the cryptochrome. Finally, one defines a “Model Hamiltonian” that describes the quantum process which tunes the biochemical function. For the EET in a light-harvesting complex, the shown excitonic Hamiltonian can describe the chromophores as two-level quantum systems (with energies $\epsilon_{i,0}$ and $\epsilon_{i,1}$ for the $i$-th chromophore, both of which depend on the chromophore nuclear coordinates $Q_i$), that interact via a nearest-neighbor potential. For the radical pair dynamics, each electron of the radical pair is described by two spin operators, $\hat{S}_1$ and $\hat{S}_2$, that interact with the Earth magnetic field $\mathbf{B}$, among themselves via a dipolar spin-spin interaction ($\mathbf{D}$ is the dipolar interaction tensor, whereas $\hat{I}_k$ is the unit matrix), and with the nuclear spin (referred to as $\hat{I}_k$ for the $k$-th nucleus). Note that a system as large as the cryptochrome contains several dozens of nuclei with non-zero spin.
account for the excitation energy transfer phenomenon. Condensing all the electrons of a chromophore into a two-level quantum system drastically reduces the Hamiltonian dimensionality and, consequently, the complexity of solving the corresponding quantum equations, but at a huge loss in terms of detail and resolution. Excitonic Hamiltonians can be further extended to include also the nuclear degrees of freedom, which leads to so-called (phenomenological) vibronic Hamiltonians.

Even within the excitonic approximation, solving exactly the Schrödinger equation for a faithful Hamiltonians describing a pigment-protein complex or a radical pair is extremely hard with classical computing algorithms. In the case of light harvesting, a pigment-protein complex may be composed of more than 10 chromophores, each one with several dozens of degrees of freedom. For magnetoreception, the radical pair interacts with the nuclear spins. The number of nuclei with non-zero spin may be large, which makes simulating the radical pair computationally hard. In both cases, this results in many-body Hamiltonians depending on more than 100 particles, which is out of the reach of the classical simulation strategies described in Section 3. For this reason, the claim that global quantum effects are relevant in photosynthesis and magnetoreception relies on simulations of simplified problems that further approximate the biochemical process. For instance, quantum mechanical simulations of magnetoreception have been limited to radical pairs that are much smaller and simpler than the cryptochromes.\cite{172-174} Quantum computing might have the potential to overcome these limits, provided that either a sufficiently large number of logical qubits can be provided in huge fault-tolerant quantum computers or a large analog quantum simulator can be built that can represent Hamiltonians with so many quasiparticles.

In fact, both the efficiency of the EET process in pigment-protein complexes and the lifetime of the radical pair can be predicted by solving the time-dependent Schrödinger equation for the corresponding Hamiltonian. As we discussed in Section (5), this is one of the few examples of a problem for which theoretical guarantees of an exponential speedup can be derived. In practice, the major challenge will be ensuring quantum hardware stability over the whole propagation time. As we display Figure 5, achieving such stability is particularly hard in digital quantum computing. Conversely, quantum simulators may be better suited for this class of simulations. As a matter of fact, some pilot experimental realizations of quantum simulators for light-harvesting systems have been proposed in recent years,\cite{173,176} although they are limited to very simple excitonic Hamiltonians.

8. Quantum computing for structural biology

Processes driven by quantum effects induced either by the breaking or forming of covalent bonds (see Section 6) or by light excitation (see Section 7) represent only a very small portion of the realm of biological processes. In fact, we already highlighted in the introduction that molecular biology is governed by reversible, non-covalent interactions between biomolecules. Although these interactions can be described with the quantum approaches presented in Section 5, such a route is both impractical and superfluous. It is impractical because, due to the curse of dimensionality, full atomistic quantum mechanical simulations of large biomolecules will most likely remain out of reach for any simulation method, both with classical and quantum computation. Furthermore, it is superfluous, because weak interactions are effectively approximated by classical surrogate potentials, i.e., by the so-called force fields.\cite{177-180} A force field approximates the PES of a molecule through a classical parametric interaction potential between its atoms, as shown in Figure 1. The interaction energy between atoms that are linked by a covalent bond is expressed as a power series in terms of bond lengths $r$, and angles $\theta_i$, and as a Fourier series in terms of the dihedral angles $\gamma_m$ (see Figure 1). The non-covalent dispersion-repulsion interaction between non-bonded atoms is, instead, expressed as a sum of pairwise Lennard-Jones-type interactions. The parameters entering the force field (i.e., $k_i^b$, $k_i^{\theta_i}$, $C_i^b$, and $C_i^{\theta_i}$ in Figure 1) are obtained by least-squares fitting either to results of quantum chemical calculations or to experimental data. The force field then defines a potential that can be used in classical atomistic simulations based on the Newton equation of motion of the atoms.

At this point, it is important to highlight that any biological process cannot be considered an isolated process. In fact, it happens at about room temperature and involves a continuous energy exchange with the environment (for instance, the cell in which a protein is located). A biomolecule is then not to be represented by a single molecular structure, but rather as a thermal ensemble of many configurations, each one weighted by the appropriate thermodynamic weight. The key quantity is, therefore, not the energy of the microstate of a single configuration, but the free energy calculated from the partition function of all microstates. The partition function also gives access to all thermodynamic properties. The PES is not the appropriate quantity to interpret the reactivity within such a thermodynamic ensemble as the driving force of a reaction is, in fact, the free energy (see Figure 7). Typically, the potential of mean forces\cite{181;182} is then considered, which generalizes the concept of PES to thermal systems. For a given reaction profile (e.g., one of the paths of Figure 1), the potential of mean forces is calculated, at each point of the PES, by averaging out all degrees of freedom of the molecule but the reactive one. Despite this conceptual difference, the electronic energies remain the cornerstone also for calculating the potential of mean forces because they are a key contribution to the partition function and, therefore, to any thermodynamic quantity. One may wonder whether quantum computing may be applied to speed-up also this step. While algorithms to calculate the partition function of quantum systems on quantum hardware have been proposed,\cite{182} which is not surprising in view of the quantum simulates quantum principle introduced above, extensions of these schemes to compute classical thermal averages are still scarce. Quantum computing realizations of the Monte-Carlo algorithm exist,\cite{183} and so it may be argued that they can be applied to calculate thermal
averages via sampling techniques. However, these algorithms assume that the distribution function can be encoded as the probability distribution of a many-body wave function (see Section 3). Hence, applying them to calculate classical thermodynamic averages requires mapping a classical probability distribution onto the qubit states, a task that is hard itself, as we will discuss in the following section.

Once an appropriate force field is available, simulating the classical dynamics of a biomolecule represents the cornerstone of several types of biological simulations. The most prominent example is the protein folding problem, i.e., the task of predicting the three-dimensional structure of a protein based on its amino acid sequence. The protein folding problem can be solved by simulating the dynamics of a protein under thermal equilibrium conditions. For sufficiently long propagation times, the protein will evolve into its most stable conformation, thereby yielding the solution to the protein folding problem. The three-dimensional structure of a protein can be reliably predicted with force fields because it is mostly determined by the non-covalent interactions between the amino acid residues of the protein. Not being affected by the curse of dimensionality, classical simulations can target much larger molecules compared to quantum mechanical simulations so that folding simulations of small fast-folding proteins are possible.

Since classical simulations are not impacted by the curse of dimensionality, one may argue whether quantum computing can be beneficial also to this field. The first, natural combination of these two fields is to parametrize a force field based on energies obtained from quantum computations. It would be possible, for instance, to parametrize or refine a force field that describes molecular interactions in proteins based on exact quantum chemical simulations on small polypeptides. A small polypeptide may contain a few dozen atoms and a few hundred electrons and is, therefore, too large to be targeted by classical quantum chemical calculations. Such a molecular size may be, in the near future, within the application range of quantum computing, in which case such a route would provide very accurate data for parametrizing force fields (provided that the quantum machine can represent the discretized problem well). This idea has been recently explored in Refs. [190] and [191], where VQE-based methods produced data for parametrizing classical force fields. In the future, the very same problems can be explored with algorithms for fault-tolerant quantum computers.

Quantum computing can also be applied to classical simulations that rely on a given force field. For instance, protein folding can be rephrased as an optimization problem. The stable structure of a protein corresponds, in fact, to the structure yielding the lowest free energy. However, this optimization problem is classically complex (NP-hard) to solve. For this reason, the folding problem can be solved based on this idea only for simplified lattice models that drastically reduce the conformational freedom of a molecule.

Note that protein folding is the only molecular biology problem that can be rephrased as an optimization problem. Another example is molecular docking, i.e., the prediction of the most stable binding configuration of a ligand to a given target, which represents the first step of in silico drug design. However, also in this case, the optimal configuration may be obtained by
minimizing a classical force field describing the interaction between the ligand and the target. As for the protein folding, the resulting optimization problem is classically hard.\cite{196,197,198}

Since optimization problems may be one of the first and primary application fields of quantum computing, the question arises of whether quantum computing may boost the efficiency of optimization-based molecular biology algorithms. Whereas polynomial-time quantum algorithms are not expected for NP-hard problems,\cite{199} like general optimization problems, more modest speedups, still significant in practice, might be achievable. An example is the quadratic speedup that can be achieved based on an exhaustive search for the optimal solution with Grover’s search algorithm\cite{200} (leaving aside questions of efficient data transfer and storage in the quantum machine; see below).

Moreover, quantum optimization algorithms for which a gain in performance cannot be guaranteed, but has been observed heuristically for small test cases, have been designed. In the latter category, the quantum approximate optimization algorithm (QAOA)\cite{201} is noteworthy and is currently under in-depth investigation.\cite{202,203} Another example of heuristic algorithms is quantum annealing, which solves a quadratic binary optimization by mapping it onto the equivalent problem of calculating the ground state configuration of an Ising Hamiltonian. The latter is then calculated on a specialized simulator. The idea is that quantum tunneling will explore the complicated energy landscape of this model more efficiently and, therefore, avoid converging into local minima.\cite{204,205}

All these classes of optimization algorithms can be, in principle, applied to biochemical problems such as those mentioned above. The first attempts to solve the protein folding optimization problem relied on quantum annealing\cite{206,207} while, more recently, hybrid classical-variational algorithms based on QAOA,\cite{208} VQE,\cite{209} and on quantum random walks\cite{210} have been proposed. The efficiency of all these methods has only been explored so far for simplified lattice models of small polymers with a few dozen of amino acids. On the one hand, this may be taken as an indication that these quantum algorithms may be promising candidates for molecular biology applications. At the same time, it is impossible to assess, based only on this observation, if and which one of them may yield a practical quantum advantage for systems that are out of the reach of classical computers. Furthermore, going beyond simplified lattice models and applying them to a real force field is a crucial step in order to truly compete with methods based on classical dynamics. We note that finding the lowest-energy conformation of a polypeptide is not the unique relevant optimization problem in protein science. For instance, it was proposed in Ref. \cite{204} to reformulate the problem finding the most favourable reaction pathways as an optimization problem, and to solve it with quantum-annealing algorithms. A key conceptual difference between simulation- and optimization-based solutions to the protein folding problem will remain for both classical and quantum computing: Solving folding as an optimization problem yields only the final structure, corresponding to the lowest (free) energy, but this procedure does not provide any insights into the physical dynamics of the folding process itself. As illustrated by the Levinthal paradox,\cite{211} a protein cannot fold simply by randomly sampling all its possible conformations because, in this way, the folding time would be much higher than what is observed experimentally. The local interactions between amino acids drive, while the protein folds, the protein toward the lowest-energy conformation, as highlighted by the famous “folding funnel” metaphor. While it is obvious that a quantum mechanical description of these interactions to yield accurate electronic energies that then enter sampling procedures of structural configurations would be desirable, it is highly uncertain whether a quantum machine can accomplish this for such large atomistic structures (let alone the fact that also the role of chaperone proteins and other mechanisms that can assist or steer a folding process may need to be considered). However, embedding approaches that dissect a quantum system into smaller quantum subsystems, which are then amenable to actual computations, can deliver key concepts to meet this challenge.\cite{212,213}

Finally, we note that a third strategy to solve the protein folding problem is offered by data-driven approaches. The three-dimensional structure of many proteins is often available from nuclear magnetic resonance or X-ray crystallography. It is, therefore, possible to use this enormous amount of experimental data as a training set for a machine-learning model. A prime example of such an application is the deep neural network AlphaFold,\cite{214} which can predict the three-dimensional structure of a protein using, as input, its amino acid sequence. Data-driven approaches can overcome the steep scaling of explicit classical and quantum simulations. However, since the former make predictions based on previously learned data, the accuracy of machine learning for genuinely new instances depends on its transferability to these unseen input data – a property that is usually referred to as generalization power, which is extremely hard to estimate \textit{a priori}.

It is therefore not surprising that it was recently shown\cite{215} that data-driven approaches can hardly predict the structure of so-called intrinsically disordered proteins,\cite{216} in which only a small fraction of the amino acid sequence is organized into a well-defined three-dimensional structure, while the rest of the protein is inherently dynamical and disordered. Inferring the structure of these proteins based on data-driven approaches is extremely challenging because their conformation is inherently different from any other protein and inference from data lacking this information breaks down. Physical modeling such as first-principles simulations does not suffer from these limitations by construction, but many practical challenges of simulating intrinsically disordered proteins with classical algorithms remain.\cite{217} We will discuss in the next section whether quantum computing may have the potential to also advance machine learning applications.
9. Quantum computing for data-driven approaches to molecular biology

One may argue whether quantum computing has the potential to boost machine-learning algorithms in molecular biology.\(^\text{(217,218)}\) Answering this question is the overarching goal of quantum machine learning.\(^\text{(219,220)}\) In essence, machine learning is a mathematical framework to learn a target function by knowing its value at a given set of points (usually known as training set). The target function is approximated with a parameterized function, and the parameters are optimized to reproduce its value on the training set.

Quantum computing may boost machine learning in two respects. First, it can speed up the training process. Second, it can make the algorithm more expressive in order to be able to predict more complex functions. The first advantage has been leveraged in deriving the quantum realization of the principal components analysis algorithm\(^\text{(221)}\) and of a support vector machine.\(^\text{(222)}\) Formal guarantees of their quantum speedup over their classical counterpart might be available for these algorithms, but they require, as the initial step, preparing the qubits in a linear superposition with coefficients that depend on the input data set. In principle, such a step can be realized with the quantum realization of the random access memory (RAM) architecture.\(^\text{(223)}\) However, depending on the nature of the input data, this loading phase may become the bottleneck of the overall simulation, therefore eliminating any practical quantum advantage.\(^\text{(224)}\)

Methods that leverage the quantum computer only for boosting the representation power of the machine-learning model do not suffer, in principle, from this limitation because they can carry out the training process classically. For instance, a neural network approximates the target functions in terms of elementary, non-linear functions – the so-called neurons. Increasing the number of neurons increases the representation power of the network but, at the same time, makes the training phase more complex. A quantum circuit (see Figure 5) can be interpreted as a quantum neural network, where the gates are the quantum counterpart of the neurons. For a given number of neurons, quantum neural networks can be more expressive – even exponentially, for specific network architectures – than their classical counterpart.\(^\text{(225–227)}\) Note that this expressivity boost is not limited to neural networks and has been observed, for instance, for support vector machines.\(^\text{(222,228,229)}\) However, the training phase, when carried out classically, may itself become hard due to the emergence of so-called Barren plateaus, especially for large data sets.\(^\text{(230,231)}\) This makes the training process itself a computationally hard problem and therefore eradicates a quantum speedup.

Quantum algorithms for solving data-driven problems are not limited to machine learning. These algorithms can also be used to identify correlations within data that are collected experimentally. For instance, genome sequencing requires reconstructing the nucleobase sequence of an organism based on small portions of the full DNA sequence.\(^\text{(232)}\) By interpreting the DNA segments as vertices of a graph, which are connected if the terminal segments of their nucleobase sequence overlap, genome sequencing becomes equivalent to the problem of finding a Hamiltonian cycle of the graph. This problem is known to be NP-hard classically,\(^\text{(233,234)}\) although efficient heuristic classical solutions to it have been developed over the years.\(^\text{(235,236)}\) The heuristic quantum algorithms introduced in Section 8 for solving optimization problems can be straightforwardly applied also to this context.

A key take-home message of the discussion so far is that quantum machine learning may become applicable, in the future, only for data-driven quantum algorithms in which a quantum speedup can be observed already for small data sets. Otherwise, either loading the classical data on quantum hardware or carrying out the classical optimization will become the main bottleneck, nullifying any quantum advantage. For instance, quantum machine learning algorithms for identifying common patterns between two DNA sequences – a task known as DNA alignment – have been proposed in recent years.\(^\text{(237,238)}\) As already mentioned above, the Grover search yields only a quadratic speedup over its classical realization. With such a limited speedup, a practical advantage will only be observed for problems with very large data sets.\(^\text{(239)}\) However, these will be difficult to deal with on currently designed hardware with rather few physical qubits (to be compared to the vast number of bits in a classical computer). Moreover, it remains to be seen how to experimentally realize a so-called quantum random access memory,\(^\text{(223)}\) which would represent a crucial component for loading a large data set on a quantum computer. For these reasons, one may not expect quantum algorithms based on the Grover search to represent – at least in the foreseeable future – suitable candidates for demonstrating a practical quantum advantage.

Other applications of quantum machine learning algorithms may, in principle, yield a practical quantum advantage in the future. This is the case of the recently proposed quantum algorithm for DNA sequencing\(^\text{(240,241)}\) and for identifying the most probable binding site of a transcription factor to DNA.\(^\text{(242)}\) In both cases, the quantum hardware is leveraged to solve a classically hard optimization problem with heuristic algorithms. However, being heuristic, the efficiency of these algorithms must be assessed at a case-by-case level. At this stage of hardware development, with quantum computers composed of at most hundreds of noisy qubits (see Section 5), these algorithms can be validated only on very small data sets. It is, therefore, currently extremely hard to predict whether they will yield an advantage when applied to problems that are out of the reach of classical hardware. This limitation does not only hold for machine learning algorithms, but for any problem where classical data are mapped onto the qubit space, including, among others, the task of finding the lowest-energy configuration of a biological macromolecule given a force field describing the interaction between its atoms.
10. Conclusions

The living world is built from atoms and molecules and, therefore, it is clearly governed by quantum physics. Due to the large size of most biomolecules and the relatively high temperatures of living organisms, classicality emerges via the process of decoherence. This raises the question in which situations accurate quantum mechanical modeling is required and where effective classical description suffices. In an in-silico descriptive and predictive context, this question directly affects the computational resources required to simulate biological processes, because full quantum mechanical simulations are affected by the curse of dimensionality when simulated with classical computers.

This article has laid out the prospects of quantum computing coming to rescue, moving from smaller to larger biomolecular structures. Quantum mechanical simulations are essential when simulating biochemical processes that are accompanied by the creation of new chemical bonds. Simulating these phenomena has, therefore, the highest prospect of impact by quantum computers. However, we also discussed how quantum effects may be crucial also in larger functional biomolecules, where quantum effects shape classical effective descriptions and require a more targeted use of quantum computing at this quantum-classical barrier.

Moving to even larger structures with an effectively cured curse of dimensionality, biocomputations are now faced with solving a still hard, but now entirely classical, computational problem, often in the form of an optimization problem, e.g. in protein folding. Also here, quantum computing might come to supplement classical computers. From a proven quadratic speedup for exhaustive search strategies to tailored algorithms to be run on fault-tolerant computers, to the heuristic approaches for near-term quantum devices, there is a long list of approaches for this class of problems, which appears not only in biomolecular sciences but in all sciences and industries, ranging from logistics to finance.

Nevertheless, the specifics of the optimization problems in biomolecular sciences require a targeted and in-depth understanding of the type of optimization problem as well as a tailored solution. The same is true more generally for important classical computational problems in the life sciences, ranging from genomics to patient data from, e.g., imaging or sensory data such as from wearable devices. Here, it seems possible that quantum computers might become a supplement and accelerator of machine learning methods. More generally, quantum technologies, such as quantum sensing, might enhance the quality of imaging data of biomolecular structures (see e.g. Ref. [244]) and, therefore, improve the quality of the data on which further classical or quantum computations would be based. In a future scenario, not only the quantum computers but also the quantum sensors might be connected with help of entanglement, thereby harnessing a further increase in resolution (cf. Ref. [245]).

We emphasize that the first-principles nature of the quantum mechanical basis of the physical modelling approach makes the methods discussed in this work also applicable to other branches of molecular and materials science. Accordingly, approaches for modelling biochemical processes will also allow for modelling of general chemical reactions. However, this is true also from a more general perspective of physical modelling or data-driven science. For instance, approaches for protein simulation will also be applicable in the context of polymer physics.

Acknowledgments

The authors are grateful for the generous funding through the ‘Quantum for Life Center’ funded by the Nova Nordisk Foundation (grant NNF20OC0059939). Open Access funding provided by Eidgenössische Technische Hochschule Zürich.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords: Quantum Computing · Molecular Biology · Quantum Chemistry · Quantum Biology · Molecular Simulations

(231) C. Ortiz Marrero, M. Kieferová, N. Wiebe, PRX Quantum 2021, 2, 040316.
(232) James M. Heather, Benjamin Chain, Genomics 2016, 107, 1.
(233) E. W. Myers, Bioinformatics 2005, 21, i79.

Manuscript received: February 16, 2023
Revised manuscript received: April 27, 2023
Accepted manuscript online: May 7, 2023
Version of record online: June 1, 2023