

The landscape of computational approaches for artificial photosynthesis

Received: 14 October 2022

Accepted: 11 April 2023

Published online: 26 June 2023

 Check for updates

Ke R. Yang ^{1,2}, Gregory W. Kyro ¹ & Victor S. Batista ^{1,2} 

Artificial photosynthesis is an attractive strategy for converting solar energy into fuels, largely because the Earth receives enough solar energy in one hour to meet humanity's energy needs for an entire year. However, developing devices for artificial photosynthesis remains difficult and requires computational approaches to guide and assist the interpretation of experiments. In this Perspective, we discuss current and future computational approaches, as well as the challenges of designing and characterizing molecular assemblies that absorb solar light, transfer electrons between interfaces, and catalyze water-splitting and fuel-forming reactions.

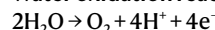
We rely on finite fuels and chemicals whose utility is exhausted once consumed in nearly all aspects of life. Technological advances and improvements in overall standard of living are likely to require more energy and lead to rising energy costs. The potential for artificial photosynthesis to provide sustainable energy on a global scale is an exciting prospect for combatting the increase in energy consumption and cost. However, achieving that goal requires the development of photocatalytic devices from Earth-abundant materials that are able to withstand solar radiation and aqueous, oxidative conditions. Although progress has been made in recent years, the intricate mechanisms of charge separation and photocatalysis pose many challenges^{1–4}.

Designing and characterizing components of artificial photosynthesis devices requires close collaboration between experimental work and computational modeling. This is because the efficiency of the photocatalytic interface depends on the performance and stability of its multiple components including semiconductor electrodes covalently linked to molecular dyes for visible light absorption and catalysts for water oxidation. Additionally, many requirements and limitations must be considered when designing these components, making the analysis more challenging than other areas of materials or chemical research. The main outstanding challenge is the design of materials to create chemical fuels from water and protons or carbon dioxide (CO₂) using visible and infrared photons from sunlight. Computational approaches are central to this effort because they can provide guidelines for materials design and fundamental understanding of the underlying mechanisms of light absorption, charge separation and catalysis at the molecular level. However, many computational challenges persist due to the balance between cost and accuracy, as is the

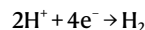
case for determining relevant energy levels and reduction potentials, representing solvent interactions and designing catalysts, thus urging the development of cheaper and more efficient methods.

Artificial photosynthesis mimics natural photosynthesis by absorbing solar light and splitting water into O₂, protons and electrons (Fig. 1). The electrons extracted from water can reduce protons or CO₂ to produce energy carrier fuels such as H₂ or hydrocarbons.

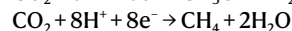
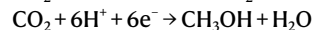
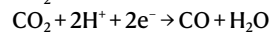
Water oxidation reaction:



Proton reduction reaction:



CO₂ reduction reactions:



To develop an artificial photosynthesis device to efficiently convert solar energy into chemical fuels, it is therefore necessary to tune photoabsorption to the solar spectrum (mostly visible and infrared) with materials that are resilient and do not degrade under oxidative aqueous conditions⁵.

Computational design has been combined with synthesis and spectroscopic characterization to create artificial photosynthetic systems with semiconductor electrodes sensitized to visible light absorption by covalently attaching molecular dyes⁶. Computational modeling has been useful for designing sensitizers and anchoring groups that can survive in aqueous and oxidative conditions while facilitating rapid electron transfer at the interface⁷. Moreover, quantum dynamics simulations have revealed that photoexcitation of the

¹Department of Chemistry, Yale University, New Haven, CT, USA. ²Energy Sciences Institute, Yale University, West Haven, CT, USA.

 e-mail: victor.batista@yale.edu

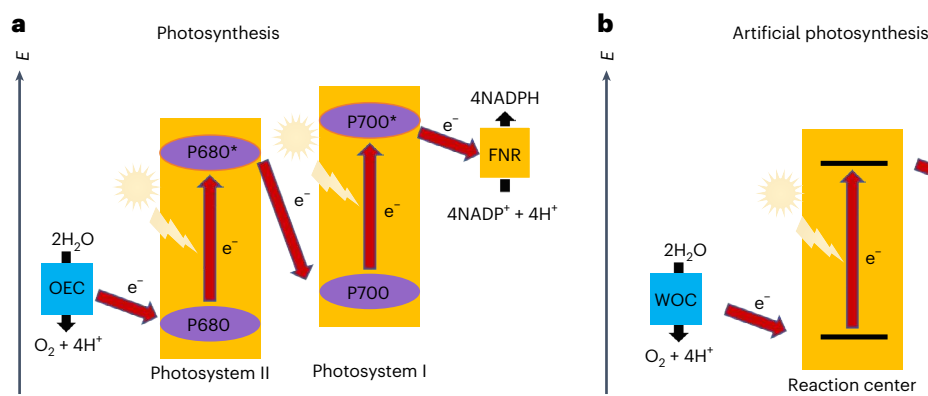


Fig. 1 | Schematic representation of natural photosynthesis and artificial photosynthesis. a, b. In both natural (a) and artificial (b) photosynthesis, light-induced charge separation occurs at reaction centers by absorption of solar light, which drives the water oxidation and proton or CO₂ reduction reactions. In natural photosynthesis, the oxidized form of NADP⁺ is reduced to NADPH, while in artificial photosynthesis, protons or CO₂ are reduced to H₂ or CO, respectively, or hydrocarbons. To achieve efficiency in artificial photosynthesis, the light absorbers must cover the solar spectrum as much as possible while still having enough of a band gap to straddle the potentials for water oxidation and proton reduction, therefore driving both reactions. In **b**, this is shown by the water oxidation catalyst transferring electrons to the lower energy state of the reaction

center, and the high energy state of the reaction center transferring electrons to the proton reduction catalyst. Additionally, robust and efficient water oxidation and proton or CO₂ reduction catalysts are needed to lower overpotentials. Effective interfacial structures need to be formed between each component of an artificial photosynthetic system to enable effective interfacial charge transfer. Red arrows indicate transfer pathways of electrons, e⁻ denotes electrons and *E* represents energy. FNR, ferredoxin–NADP(+) reductase; NADP⁺, nicotinamide adenine dinucleotide phosphate; NADPH, reduced form of nicotinamide adenine dinucleotide phosphate; OEC, oxygen evolving complex; P680, pigment absorbed at ~680 nm in photosystem II; P700, pigment absorbed at ~700 nm in photosystem I; PRC, proton reduction catalyst; WOC, water oxidation catalyst.

chromophore triggers ultrafast electron transfer on a sub-picosecond timescale, creating a charge-separated state that quickly separates further and injects the photoelectron into the semiconductor electrode.

Balancing the rates of electron transfer and charge recombination in robust materials is essential for achieving high photoconversion efficiency. After electron injection into the semiconductor electrode, photogenerated electrons are then used to reduce protons or CO₂ to create fuels, while the holes left behind are used to catalytically oxidize water, effectively storing the absorbed solar energy in the form of chemical bonds (Fig. 2a). One of the challenges addressed by computational studies has been the elucidation of the catalytically active species responsible for water oxidation. A complication in understanding catalytic performance is the propensity of catalysts to change form in solution. In this scenario, the originally prepared catalysts turn out to be pre-catalysts, while the actual catalytic species are formed in situ and remain elusive to identification. Resolving the identity of water oxidation catalysts has therefore especially benefited from computational modeling and close collaboration with spectroscopic analyses^{8,9}.

Considering the central role of computational studies in the development and characterization of artificial photosynthetic systems, it is natural to anticipate that more advanced computational approaches, in conjunction with advances in machine learning (ML)^{10,11} and quantum computing^{12–14}, will continue to make significant contributions essential for exploration and innovation in this field.

Synergistic computational and experimental approaches

Computational approaches can be most effective in the field of artificial photosynthesis when applied to assess structural or mechanistic hypotheses formulated for the interpretation of experiments. Building and validating computational models thus typically requires experimental information on the nature of the specific material investigated, as well as data from experiments addressing well-defined questions about the behavior of the system.

Benchmark model systems allow for validation of the computational approaches and level of theory. After validation, computations provide insights at the molecular level that can support or otherwise disfavor the formulated hypotheses. Moreover, simulations allow for

rigorous interpretations of experiments and for predictions of observables that could be checked by a new round of experiments, establishing a synergistic iterative loop between theory and experiments.

The iterative approach has proven particularly effective for the development of molecular assemblies on semiconductor surfaces, anchoring sensitizers and catalysts that are stable under aqueous oxidative conditions, while providing understanding of the timescales and mechanisms for charge separation and catalysis in artificial photosynthesis devices^{8,9}. For example, quantum simulations of interfacial electron transfer (IET) initially addressed the feasibility of sensitizing a TiO₂ semiconductor surface to visible light by functionalization of catechol adsorbates. The simulations suggested sensitization to visible light and a sub-picosecond timescale for IET, promoting the idea that photoabsorption and IET would not be rate-limiting compared with the microsecond timescale at which solar photons arrive to the photosensitizer under normal conditions (1 sun = 1 kW m⁻² is roughly 2,000 photons nm⁻², with a typical surface coverage of 1 adsorbate nm⁻²)⁸. Experiments subsequently confirmed those predictions by ultraviolet–visible (UV-vis) spectroscopy and terahertz pump–probe spectroscopy, showing injection times on the sub-picosecond timescale. Subsequent experimental studies showed that catechol anchoring groups were not sufficiently robust under aqueous oxidative conditions, motivating the development of a new generation of anchoring groups more suitable for high-potential photoanodes.

Computational modeling has also been essential for advancing our ability to design efficient and stable catalysts, as proven by successful collaborations between computational and experimental studies on water oxidation, hydrogen evolution and CO₂ reduction reactions. For example, the actual molecular structure of the active form of the water oxidation catalysts Cp*Ir^{III} (Cp*, pentamethyl-cyclopentadienyl) in solution was computationally resolved⁹. It was shown that the Cp* ligand undergoes oxidative degradation and forms an Ir^{III} dimer that turns out to be the catalytically active species in solution. The proposed structural model, obtained at the density functional theory (DFT) level, was subsequently validated by simulations of high-energy X-ray scattering (HEXS), extended X-ray absorption fine structure (EXAFS) spectroscopy and direct comparisons with experiments. Another example where computations were essential to resolve ambiguities

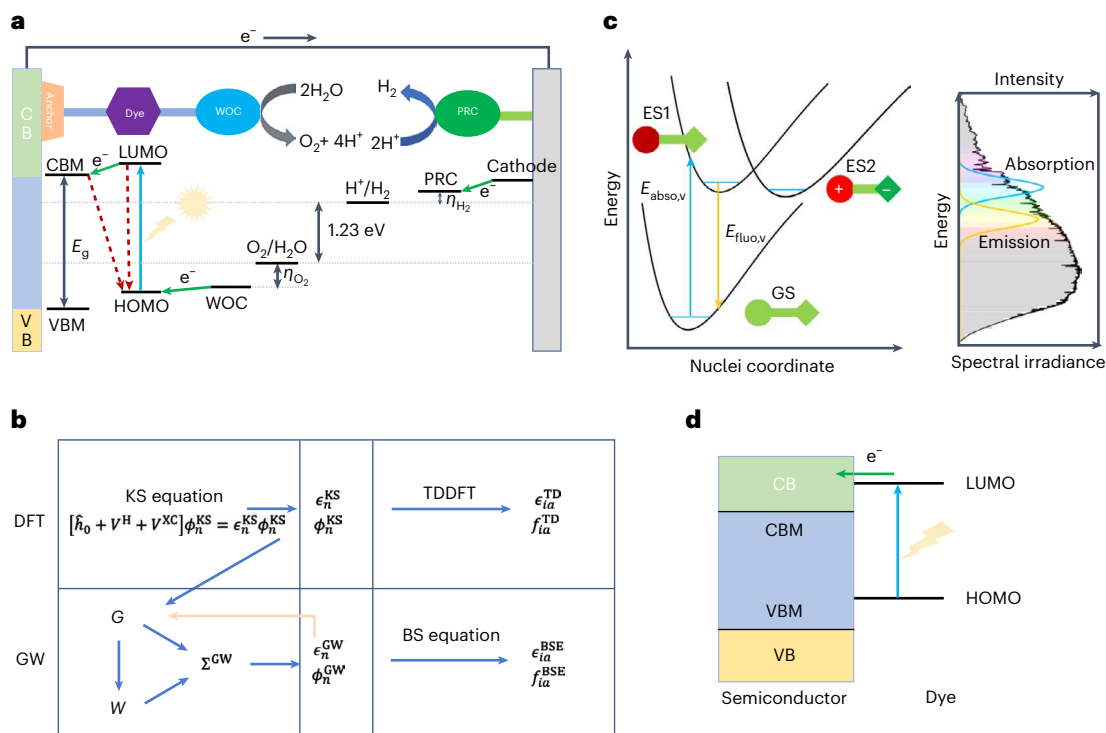


Fig. 2 | Water splitting driven by photoexcitation. a, Energy level alignment of the normal hydrogen electrode, reduction potential of O₂ or H₂O, valence band maximum (VBM) and conduction band minimum (CBM) of semiconductors, and HOMO and LUMO of molecular dyes. **b**, Procedures to perform a DFT calculation to obtain Kohn–Sham (KS) orbitals (ϕ) and orbital energies (ϵ); a TDDFT calculation to obtain excitation energies (ϵ) and oscillation strengths (f); the *GW* method to obtain more accurate orbital energies (ϵ) and orbitals (ϕ) using the KS orbitals and orbital energies as input; and the Bethe–Salpeter (BS) equation to obtain excitation energies

(ϵ) and oscillation strengths (f). **c**, Local excitation (ES1) and charge-transfer excitation states (ES2) and an overlay of absorption/fluorescence spectra and solar spectrum. GS refers to a ground-state transition, E_{abs} is the absorption energy from the electronic ground state to an excited state, and E_{flu} is the fluorescence energy from the lowest energy excited state back to the ground state. **d**, IET between a molecular dye and semiconductor. Valence band (VB), VBM, conduction band (CB) and CBM are shown. Molecular dye injects electrons into the conduction band of semiconductor after photoexcitation.

regarding the nature of the catalytically active species in solution has been the demonstration of in situ generation of self-improved water oxidation catalysts with the general formula [(trpy)(5,5'-X₂-bpy)Ru^{IV}(μ-O)Ru^{IV}(trpy)(O)(H₂O)]⁴⁺ (trpy, tripyridine; bpy, 2,2'-bipyridine; X = H or F) found to form in solution during the catalytic cycle from their corresponding pre-catalyst mononuclear counterparts [Ru(trpy)(5,5'-X₂-bpy)(H₂O)]²⁺ (ref. 15). DFT computational analysis established the coexistence of two interconnected catalytic cycles where the mononuclear catalytic system is slowly and irreversibly converted to the more stable dinuclear catalytic system.

Computational approaches have enabled us to make great strides in understanding light absorption¹⁶, charge separation¹⁷ and catalytic water oxidation processes^{18–20} of artificial photosynthesis. In the following sections, we discuss the limitations of current computational methods and identify potential future steps to progress the field. We look forward to seeing these advances come to fruition, ultimately enabling us to create efficient solar energy conversion devices from Earth-abundant materials, paving the way for a sustainable-energy society.

Light absorption

Light absorption is the first step in artificial photosynthesis. Light is typically absorbed by a semiconductor material or molecular dye adsorbed to a semiconductor, generating electron–hole pairs that subsequently separate and drive oxidation and reduction reactions. The seminal work of Fujishima and Honda demonstrated that TiO₂ can be used as a photoanode to absorb light and catalyze water oxidation when applying a small bias potential²¹. However, TiO₂ is only capable of absorbing UV light, which is just a small fraction of the solar

spectrum. In prolonged attempts to increase the efficiency of photo-absorption, many other materials have since been analyzed, including perovskites, Fe₂O₃, WO₃ and BiVO₄, as well as semiconductor materials sensitized with molecular adsorbates⁷. Efforts in materials design to maximize absorption of the solar spectrum while enabling efficient photoconversion by charge separation have greatly benefitted from computational approaches discussed in subsequent sections, including DFT calculations of electronic structure, the analysis of electronic-level alignment, inverse design and simulations of IET.

Computational tools provide a fast and inexpensive approach for in silico screening and design of semiconductors for optimal light absorption. In many cases, light absorption has been improved by heterogenization of molecular dye sensitizers inspired by dye-sensitized solar cells used in dye-sensitized photoelectrosynthesis cells⁵ for solar fuel production (Fig. 2a). The energies of photogenerated electrons and holes can be estimated by computing electron affinities (EAs) and ionization potentials (IPs), often approximated according to Koopmans' theorem²² in terms of the energies of frontier orbitals such as the lowest-unoccupied molecular orbital (LUMO) and highest-occupied molecular orbital (HOMO) energies of a molecule, respectively. The difference between the EA and IP defines the 'quasiparticle gap' (or also the 'fundamental gap' or 'physical gap') of a molecule or material¹. For semiconductor materials, this corresponds to the band gap, defined as the energy difference between the edge of the conduction band minimum and the edge of the valence band maximum.

Solar light absorption must power water oxidation and proton or CO₂ reduction reactions to generate chemical fuels. Therefore, computational design and modeling of energy level alignment must

ensure that the photogenerated holes have enough energy to oxidize water to dioxygen, producing electrons (that is, reducing equivalents) that could be used to reduce protons to H_2 or CO_2 to hydrocarbons (Fig. 2a). Therefore, the quasiparticle gap (often approximately the HOMO–LUMO gap of a dye molecule, or the band gap of a semiconductor) must straddle the redox potentials of water splitting into dioxygen and hydrogen, while the chromophore maintains the ability to span as much of the solar spectrum as possible without degrading through oxidation, or detaching from the surface in aqueous solution. Computational approaches for determining the quasiparticle gap are crucial to the design and development of dye-sensitized photoelectrosynthesis cells for their ability to assess the viability of different materials at high speed and low cost. Quasiparticle gaps are often determined by DFT calculations. However, the predicted HOMO–LUMO gaps for molecules and band gaps for solids tend to be smaller than the actual values¹ due to self-interaction errors (SIEs)²³, which arise from spurious self-interactions of electrons, particularly when using local spin density approximation (LSDA) and generalized gradient approximation (GGA) functionals. There are two common approaches to partially correcting SIEs, including hybrid functionals and the DFT+*U* methodology. DFT+*U* was introduced to properly describe Mott insulators such as 3*d* transition-metal oxides²⁴. Since then, rotationally invariant implementations^{25,26} have been widely used. The DFT+*U* approach can properly describe the electronic structure and band gap of hematite, and has been used to determine the band gaps and band edge positions of pristine²⁷ and doped²⁸ hematite. However, it is insufficient when applied to semiconductors with *d*⁰ or *d*¹⁰ metal ions such as TiO_2 (ref. 2) and Cu_2O (ref. 3), which require treatments based on hybrid functionals²⁹ or *GW* methods (where *G* stands for the one-body Green's function and *W* stands for the dynamically screened Coulomb interaction)³⁰. While computationally more expensive, hybrid functionals such as B3LYP²⁹ are generally more accurate than non-hybrid functionals for consideration of the exact exchange from Hartree–Fock calculations, and have been extensively applied to calculating band gaps of a wide range of semiconductor materials with a mean absolute error of only 0.3 eV for TiO_2 , WO_3 , $BiVO_4$, Cu_2O and GaN (ref. 16) for solar water splitting. Nevertheless, accurate yet efficient methods for calculating the electronic structure of semiconductor materials have yet to be developed, a challenge that is expected to benefit from recent progress in ML methods³¹.

Figure 2c shows a schematic diagram of the energy levels associated with the primary event of artificial photosynthesis involving light absorption, a process that promotes the absorber to an electronically excited state (ES1). Usually, the system relaxes back to the ground state by dissipation of the absorbed energy in the form of heat (vibrations), emitting a photon by fluorescence, or generating a charge-separated state (ES2) that can be used to drive redox chemistry to produce fuels^{6–8}. Computational modeling of the electronic structure can provide insights of the relaxation mechanisms as well as guidelines for the design of materials that promote charge-separated states. Absorption energies can be approximated as quasiparticle gaps, although calculations of absorption intensities and energy dissipation require more rigorous treatments of electronic excitations and vibronic transitions.

For molecular systems, excited states can be described quite rigorously using wavefunction-based quantum chemical methods such as the equation-of-motion coupled-cluster single–double (EOM-CCSD) method, complete active space self-consistent field (CASSCF) theory, complete active space perturbation theory second-order (CASPT2), *n*-electron valence state perturbation theory second-order (NEVPT2) and multireference configuration interaction (MRCI). However, the computational cost of these methods scales unfavorably with system size, limiting the applicability to small benchmark molecular models¹⁸. Time-dependent DFT (TDDFT)³² provides a cost-effective method for studying electronic excitations, and is therefore widely used to calculate absorption spectra of dye molecules and small nanoclusters. TDDFT has the advantage of producing reliable oscillator strengths,

which are directly related to absorption intensities. However, current implementations of TDDFT are not sufficiently reliable. For example, the adiabatic approximation is almost universally made in TDDFT implementations, and its impact on accuracy has not been fully investigated⁴. Moreover, the magnitudes of the particle–hole SIE associated with the exchange functional and the orbital relaxation error associated with using fixed ground state orbitals (both within the adiabatic approximation) have only been estimated in terms of their impacts on EA (EA-TDDFT)³³. The orbital-optimized DFT (OO-DFT) method³⁴ has been developed to avoid these errors.

Sophisticated functionals have been developed to address charge-transfer excitation, including the range-separated hybrid functionals, which are quite promising³⁵. Besides TDDFT, the Bethe–Salpeter equation³⁶ at the *GW* level (BSE-*GW*) has recently attracted significant interest in the chemistry community since it can properly describe charge-transfer excitations in both molecular and extended systems, essential for the computational modeling of complex interfaces involved in artificial photosynthesis. In fact, the BSE-*GW* approach has been used to calculate absorption spectra of photoelectrode materials including TiO_2 , WO_3 , Fe_2O_3 , $BiVO_4$, $SrTiO_3$ and GaN, all of which show good agreement with experimental spectra. Moreover, given the recent successes of ML methods for predicting optical spectroscopic properties, we envision that emerging techniques will be very useful for addressing the current challenges of excited-state computational methods³⁷.

Charge separation

Electron–hole pairs are generated upon photoabsorption, storing the absorbed solar energy in the form of electrochemical energy. The charge separation process is particularly difficult because there is only a small window of time for kinetically favored processes to predominate over the otherwise thermodynamically preferred charge recombination. In sensitized semiconductors, the process of IET from the dye molecule to the surface is controlled by the electronic level alignment and electronic coupling between the donor states mostly localized on the adsorbate dye and the acceptor energy levels in the semiconductor (Fig. 2d). Thus, the design of efficient artificial photosynthesis devices must solve the challenge of promoting charge separation via IET using anchoring groups that enable efficient electron transfer and are structurally and functionally robust even after photo-oxidation of the anchored molecular assembly in aqueous solutions. DFT calculations of molecular assemblies anchored to semiconductor surfaces can provide characterization of binding modes, while calculations of molecular conductance⁶ and simulations of IET from adsorbate dyes into semiconductor surfaces through anchoring groups allow for estimations of the timescales and mechanisms of IET.

Quantum dynamics simulations of electron transfer have been reported for a variety of model systems, going back at least to the initial study of IET from catechol into TiO_2 (Fig. 2d)⁸. In that work, ab initio molecular dynamics (MD) was used to obtain representative nuclear configurations, and electronic wavefunctions were then propagated by solving the time-dependent Schrödinger equation with an extended Hückel Hamiltonian. Although approximate, models based on tight-binding Hamiltonians can make IET simulations possible even for thousands of atoms. That methodology has also been widely used to study the effects of thermal nuclear dynamics³⁸, exposed facets³⁹ and solvation⁴⁰ on IET for several dye– TiO_2 complexes⁴¹, and has allowed for the discovery of a two-step mechanism for ultrafast IET⁴², highlighting its applicability to screening dye–semiconductor complexes for efficient IET in artificial photosynthesis. However, a tight-binding Hamiltonian, such as extended Hückel, requires parameterization and is not suitable to describe the effects of changes in nuclear configurations. Therefore, it is often coupled to classical MD simulations of the nuclei such as in the DynEMol software package, as demonstrated in studies of vibronic effects in the ultrafast IET of perylene-sensitized TiO_2 surfaces⁴³.

Ab initio nonadiabatic MD (NAMD) simulations using DFT orbitals is also typically applied to study IET¹⁷. Due to the computational cost, current ab initio NAMD simulations are limited to the analysis of short simulations. Exploring longer time dynamics, including carrier-trapping mechanisms that limit the overall efficiency of charge transport through the semiconductor material, is still beyond the capabilities of current computational methods and would require development of cost-effective methodologies.

A challenging aspect of modeling artificial photosynthetic assemblies is the proper description of electronic energy level alignment, which is critical for efficient photoelectrocatalysis. The performance of photoelectrocatalysis is dependent on the energy alignment between electron donor and electron acceptor states, such as the alignment of the photoelectrode band edges and the water oxidation and proton or CO₂ reduction potentials (Fig. 2a). Essentially, the alignment of electronic energy levels is critical since it determines the overall efficiency of charge transfer and thus whether photo-generated holes and electrons are able to oxidize water and reduce protons or CO₂. The computational analysis of level alignment is typically based on DFT calculations of electronic energy levels relative to vacuum, and free energy calculations of redox potentials. For a bulk semiconductor, the absolute band edge position can be obtained through slab calculations. LSDA/GGA functionals are often used due to their low computational cost, although they typically underestimate the band gaps. More reliable approaches such as hybrid functionals⁴⁴ and *GW* approximations⁴⁵ can be used to determine the band edge positions through slab calculations, but they are computationally expensive. An approximate approach combines LSDA/GGA slab calculations with band gap corrections from hybrid functionals and *GW* methods. Recent work has combined ML and DFT and has shown success in rapidly advancing electronic structure applications such as in silico materials discovery and the search for new chemical reaction pathways⁴⁶.

It is important to note that determining band edges under realistic conditions would require modeling photoelectrode–water interfaces rather than semiconductor–vacuum interfaces. This could be done with implicit solvation models using joint DFT approaches⁴⁷, or with explicit solvation models using DFT-based ab initio MD⁴⁸. Implicit solvation models can provide a computationally efficient way to describe solid–liquid interfaces. However, important aspects of solvation, including the description of critical hydrogen-bonding interactions, require fully atomistic models. Unfortunately, explicit solvation models based on ab initio calculations are quite expensive, so they are restricted to short simulations and rather limited sampling of configurations. Therefore, there is an urgent need for development of accurate and efficient computational methods that can properly determine the effects of solvation on the electronic bands of photoelectrodes in aqueous solution.

Catalysis

Modeling catalytic reactions

After light absorption and the subsequent creation of electron–hole pairs, electrons are employed to reduce protons or CO₂, thus creating chemical fuels, while holes drive the extraction of electrons from water through catalytic water oxidation. As such, the development of computational models for artificial photosynthesis has been largely directed at understanding the relevant electrochemical reactions. These are multi-electron reactions that tend to be slow, so catalysts are needed to speed up their rate. For these reactions, the catalyst must be stable in water and not degrade when exposed to light. As a result, it remains rather challenging to create efficient water oxidation catalysts that are affordable and abundant enough to be used worldwide.

Computational modeling of the key chemical reactions in artificial photosynthesis must elucidate the mechanisms driving the electron or proton transfer steps, bond cleavage and bond formation. This analysis is essential to map reaction pathways and understand how

catalysts enable these mechanisms, considering both kinetic and thermodynamic data. In general, the precise mechanism of water oxidation depends on the reaction conditions and the nature of the catalyst. However, it is commonly assumed that the minimum free energy pathway enables water oxidation by formation of a series of intermediate states, including *OH₂, *OH, *O, *OOH and *OO generated by proton-coupled electron transfer (Fig. 3d).

Theoretical overpotentials (the difference between the minimal applied potential and the equilibrium potential of an electrochemical reaction) can then be estimated at the DFT level by computing the standard reduction potentials for each one-electron oxidation step (Fig. 2d). In the case of water oxidation, the equilibrium potential relative to the standard hydrogen electrode is 1.23 V at pH = 0, and the minimum applied potential is the maximum reduction potential of all intermediates. The computational hydrogen electrode has been used to calculate the theoretical overpotentials of water oxidation over a wide range of metal oxide surfaces¹⁹, finding that the adsorption energy difference between *O and *OH can serve as a good descriptor for oxygen evolution reactions. Clearly, the free energy change for each reduction reaction should be close to 1.23 eV for an efficient water oxidation catalyst²⁰.

Optimizing CO₂ reduction is very challenging since large overpotentials are often required to activate the rather inert CO₂ molecules, and heterogeneous catalysts suffer from poor product selectivity. Computational approaches are therefore essential to elucidate the underlying reaction mechanisms to enable the rational design of robust catalysts with low overpotentials, small activation free energies and high product selectivity. As shown in Fig. 3a, the standard free energy changes for an oxidation–reduction pair ($\Delta G_{\text{gas}}^{\circ}(\text{Ox}|\text{Red})$) are typically determined with the Born–Haber cycle from the IP, the thermal correction to free energy changes in the gas phase, and the solvation energies of reactants and products. To calculate $\Delta G_{\text{gas}}^{\circ}(\text{Ox}|\text{Red})$, we need potential energy surfaces for both the oxidized and reduced forms of a redox pair Ox|Red, which correspond to the ground states of $N - 1$ and N electron systems, respectively. Therefore, both forms can be well described by ground-state quantum chemical methods, which generally include single-reference and multireference methods (Fig. 3b). The common single-reference methods include the Hartree–Fock method, Møller–Plesset perturbation theory, coupled-cluster methods and DFT⁴⁹. Multireference methods are generally more complicated but are necessary when the energy separations between electronic states are small, which is often the case for transition-metal catalysts, excitation processes, and chemical bond cleavage and formation in most artificial photosynthesis systems. Common multireference methods include CASSCF, CASPT2, CASPT3 and so on, NEVPT2, configuration interaction (CI), multireference CI (MRCI) and multireference coupled-cluster (MRCC) theory¹⁸. It is worth noting that active-space approaches are highly dependent on the choice of active space, and although they are potentially powerful, they can also be unreliable. Additionally, modern local correlation approaches to coupled-cluster methods such as domain-based local pair natural orbital-coupled-cluster singles and doubles have enabled highly accurate redox and spin-state energetic calculations in large molecular systems that are of direct relevance to artificial photosynthesis research. Among these quantum chemical methods for calculating the gas-phase reaction energies, the cost of DFT scales most favorably with respect to system size. However, the performance of DFT depends on the choice of exchange–correlation functionals and, unlike wavefunction-based methods, there is no systematic way to improve its performance. As has been mentioned previously to address many of the computational concerns related to modeling both light absorption and charge separation, ML can be combined with DFT to improve electronic structure-based techniques for modeling catalytic reactions⁴⁶.

Solvation effects can contribute significantly to the changes in reaction free energy in solution, and therefore need to be accounted for

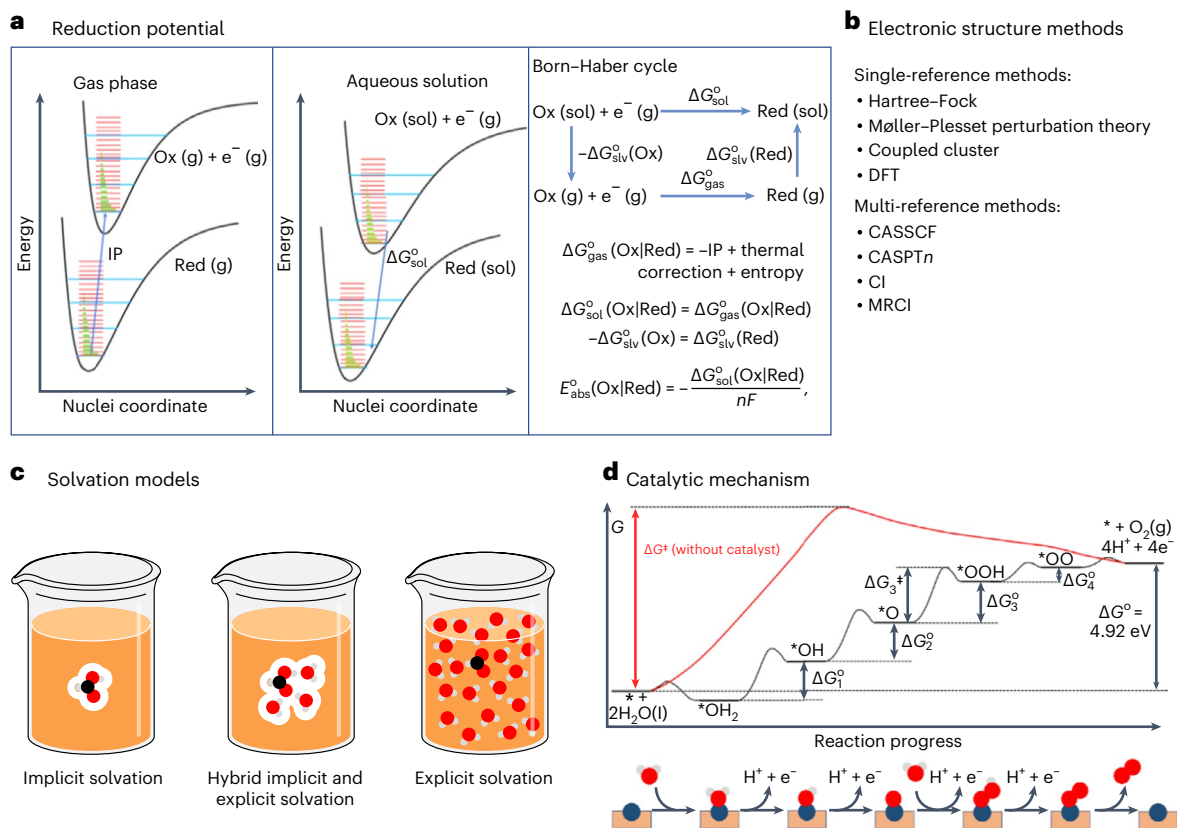


Fig. 3 | Computational modeling of reduction potentials and catalytic mechanism in artificial photosynthesis. **a**, Schematic representation of slices of potential energy surfaces (in black) of the oxidized (Ox) and reduced (Red) forms of a generic oxidation–reduction (Ox|Red) redox couple in gas phase and in aqueous solution. Blue and red lines represent vibrational and rotational levels, respectively. The green boxes show the Boltzmann population of vibrational and rotational states at a finite temperature, T . Ionization potential (IP) and the gas phase free energy change ($\Delta G_{\text{gas}}^{\text{o}} (\text{Ox}|\text{Red})$) are shown. **b**, A list of popular quantum chemical methods that can be used to calculate $\Delta G_{\text{gas}}^{\text{o}} (\text{Ox}|\text{Red})$.

by implicit⁵⁰, explicit⁵¹ or hybrid implicit and explicit solvation models (Fig. 3c)⁵². Implicit solvation models treat the solvent as a polarizable dielectric medium, without representing any solvation molecules explicitly. Through careful parameterization, implicit solvation models can properly account for the average bulk electrostatic contribution to solvation effects. However, they cannot properly describe hydrogen-bonding interactions. By contrast, explicit models treat the solvent as explicit molecules and consider explicit interactions between solute and solvent molecules. To obtain reliable solvation free energies with explicit models, it is necessary to describe the interactions between solute and solvent molecules as accurately as possible and sample the configuration space exhaustively through either MD or Monte Carlo techniques. Hybrid implicit and explicit models treat a small number of solvent molecules explicitly, such as those that form hydrogen bonds with solute molecules, and represent the rest of the solvent molecules implicitly. With current solvation models, the error in solvation energy calculations can be as low as 0.2 eV (ref. 18). However, it remains the dominant error in free energy calculations of redox potentials. Therefore, it is imperative to develop more accurate methods for estimating solvation energies that could be used in free energy calculations based on traditional methods as well as new approaches based on ML⁵³.

Inverse design of catalysts

The discovery and development of new catalysts can be substantially accelerated with ML through inverse design (such as a data-driven

method that explores chemical-wide space)^{54,55}. ML requires a data set to train the model, and therefore much of the model's predictive capability is dependent on both the quality and quantity of the data at hand. However, mature databases of catalysts and materials for artificial photosynthesis have yet to be developed and made widely available. Some databases that are currently available include AtomWork⁵⁶, the Automatic Flow of Materials Discovery Library⁵⁷, the Inorganic Crystal Structure Database⁵⁸, the Materials Project⁵⁹ and the Open Quantum Materials Database⁶⁰.

Both supervised and unsupervised ML approaches have been utilized to assist catalyst design. Supervised learning involves training on data with known labels and learning the proper parameters of the ML model to predict with high accuracy the output label for a given input. Deep learning models have been shown to perform significantly better than other methods on supervised learning tasks⁶¹. Convolutional neural networks (CNNs) first developed for image detection have been tuned for chemical prediction, often incorporating methods such as those presented in ResNet and VGG. In the case of CNNs, molecules are embedded into a spatial grid, either 2D or 3D, and a filter scans over the data to generate an output, learning its weight to properly fit the data relative to the corresponding labels. A sufficiently trained ML model can make accurate predictions and provide information on the relationship between the input and output data. Therefore, it is possible to extract information regarding how strongly each feature used to represent the input data contributes to the prediction.

This approach has already been successfully applied to identify potential catalysts to be used in artificial photosynthesis devices, including water oxidation catalysts⁶², CO₂ reduction catalysts⁶³ and hydrogen evolution catalysts⁶⁴.

ML has also been used to predict DFT-derived properties of catalytic materials^{10,11}. For example, there have been recent studies using ML to predict the surface adsorption energies of reactants and reaction intermediates⁵⁴. Additionally, ML has been used to predict *d*-band centers⁶⁵, which are widely used to describe a variety of reactions occurring over transition-metal catalysts. There are also instances of ML being used to predict band gaps and surface phase diagrams⁶⁶.

ML models trained with data from various types of spectroscopic and microscopic techniques for a large number of catalysts would be extremely valuable. It is natural to anticipate those models would be able to predict and generate the structures of active sites, intermediates and product distributions. Also, reactions could be predicted with an ML model trained on experimentally generated data for reactions promoted by different catalysts. While still at an early stage, this approach has already been shown for CH₄ conversion, CO oxidation, CO₂ reduction, the oxygen evolution reaction as well as the photocatalytic splitting of water⁵⁴. With a proper understanding of the catalytic mechanisms involved in artificial photosynthesis reactions, one could enable the *in silico* design of better catalysts with lower activation free energies and lower overpotentials.

A major challenge in ML-based catalyst design arises from the fact that most work reported to date has focused on a very narrow subset of catalysts and reactions. Additionally, ML studies have been limited to establishing correlations between human-selected reaction descriptors such as reaction rates and energies, adsorption energies and turnover numbers. It is, therefore, crucial to bridge the gap between current efforts, mostly focused on simple descriptor-based screenings of catalysts, and the capabilities of pattern recognition that could be provided by deep learning models. As has already been shown in other areas of artificial intelligence applications, ML models will become increasingly powerful computational tools in the field as soon as databases become more available and better organized for the class of materials and catalysts relevant to artificial photosynthesis. However, it is worth noting that ML models are not self-aware, and therefore will require human tuning to achieve optimal performance.

Looking forward

Artificial photosynthesis is clearly an intrinsically complex process with multiple chemical events spanning different timescales and involving materials and molecular components of different sizes (Fig. 4), as discussed herein. Multiscale computational approaches are available for studying artificial photosynthesis, including a variety of methods with a wide range of costs and accuracies. Atomic modeling of processes in artificial photosynthesis is enabled for a reasonable computational cost by *ab initio* wavefunction and DFT-based methods, semi-empirical quantum mechanics, density functional tight binding methods⁶⁷ and reactive force field methods⁶⁸. Coarse-grained⁶⁹ and continuum models⁷⁰, on the other hand, sacrifice accuracy of atomic details to describe larger systems with longer dynamics. The integration of atomistic and coarse-grained models involving different lengths and timescales remains to be developed, an effort that could provide a complete mechanistic understanding of the catalytic transport processes involved in artificial photosynthesis.

Computational modeling of light absorption on the femtosecond timescale requires accurate computational methods to describe both the ground state and excited states, as well as the influence of the surrounding environment on the oscillator strengths and vibronic couplings. TDDFT and BSE-*GW* methods are widely used for calculations of absorption spectra. However, the adiabatic linear-response implementation of TDDFT underestimates charge-transfer excitation energies, and therefore BSE-*GW* methods are often preferred.

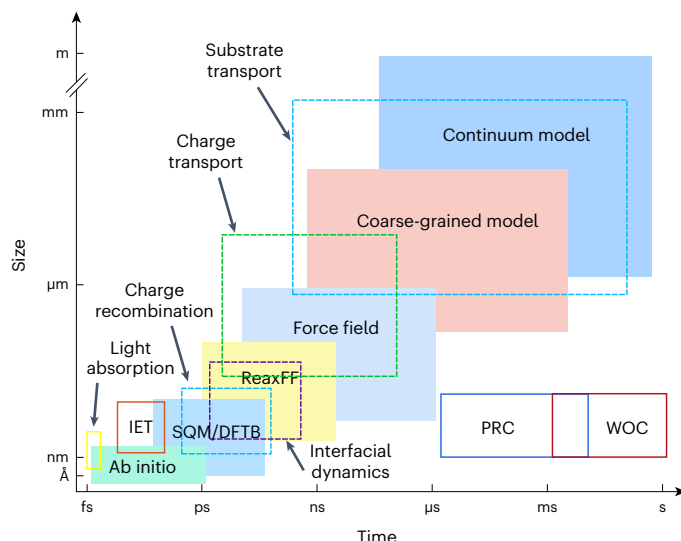


Fig. 4 | Sizes and timescales of computational modeling methods and chemical events in artificial photosynthesis. Filled squares depict approximate time and size scales feasible for specific computational methods. *Ab initio* methods include wavefunction-based methods, DFT and their variant for excited states; semi-empirical quantum mechanics (SQM) and density functional tight binding (DFTB)⁶⁷; reactive force field method (ReaxFF)⁶⁸, which uses force fields with flexible functional forms to describe bond breaking and formation; force field methods for describing interactions between atoms; coarse-grained models⁶⁹, which use coarse-grained representations rather than atomic representation to describe complex systems; and continuum models⁷⁰, which use continuous representations to describe systems. Unfilled boxes represent approximate time and size scales of chemical events involved in artificial photosynthesis: light absorption, IET between an excited dye molecule and semiconductor photoelectrode or between catalysts and dye molecules; charge recombination of photo-generated electrons and holes; interfacial dynamics describe the dynamics of photoelectrode–electrolyte interfaces, which are important for stabilizing surface charges; charge and substrate transport refer to the transport of charges in photoelectrodes and substrates in the solution; PRC and WOC are catalytic reactions for solar water splitting.

To account for solvation and environmental effects on light absorption, quantum or molecular mechanics methods are commonly used, incorporating either implicit, explicit or hybrid solvation models. Similar approaches with BSE-*GW* methods are currently under development to account for solvation and environment effects⁷¹. Aside from solvation and environment effects, vibronic couplings⁴³ and nuclear dynamics can also affect light absorption, as demonstrated in the study of BiVO₄, showing that modeling nuclear dynamics with classical MD lowers the bandgap by 0.70 eV, while modeling with path-integral MD methods further lowers the gap by 0.22 eV (ref. 72). It is, therefore, important to account for both the complex environment and the dynamical nature of dye molecules and photoelectrodes when modeling light absorption under realistic conditions.

Quantum dynamics methods based on tight-binding model Hamiltonians have provided an effective way to study IET between the photo-excited dye molecule and the semiconductor surface⁸. However, new parameterizations from reliable reference data are needed to study charge injection into other semiconductors, such as Fe₂O₃, WO₃ and BiVO₄, that may have important applications in artificial photosynthesis. Advancements have led to implementation of explicit non-adiabatic effects, allowing for the description of vibronic effects, and the influence of the electron–hole interaction and the changing electron density on the nuclear dynamics⁷³. Further developments of computational approaches are required to enable non-adiabatic quantum simulations of both nuclear and electronic dynamics, as demonstrated for small molecular systems with tensor network methods⁷⁴. The NAMM

approach combined with TDDFT can be used to study IET without parametrization, but the computational cost currently limits the size and timescale of the simulations that can feasibly be performed. Charge recombination dynamics, which are rarely modeled computationally due to the long associated timescale, could in principle be described using more cost-effective approaches such as time-dependent density functional tight binding⁶⁷. In addition, ML methods have been recently applied to study excited state dynamics of molecules⁷⁵, and could be extended to study excited-state dynamics in extended bulk materials and interfacial systems with larger sizes and longer timescales. Further outstanding challenges regarding simulations of IET dynamics also concern the role of spin transitions ubiquitous to electronic relaxation at photoelectrode interfaces.

Computational modeling enables researchers to examine the key chemical reactions of artificial photosynthesis at a molecular level, such as water oxidation and proton or CO₂ conversion into fuels. Most of these catalysts involve transition metals that have significant static correlations and therefore, in principle, require multireference treatments with methods such as CASPT2 or MRCI. Owing to the computational cost of these methods, however, many computational studies of catalytic mechanisms are currently limited to approximate approaches such as broken-symmetry DFT. It is also worth mentioning that it is necessary to adequately consider relativistic effects and spin-orbit coupling in the case of heavy-element systems. The recent developments of quantum Monte Carlo⁷⁶, multi-configurational DFT⁷⁷, and tensor network methods⁷⁸ such as density matrix renormalization group theory provide accurate treatment of larger systems with transition metals. Although these methods are still too expensive for routine application, they can be used to benchmark broken-symmetry DFT approaches with approximate functionals and even guide the development of better functionals. In fact, a functional has been developed by training a neural network, and has demonstrated good performance for main-group atoms and molecules with significant static correlations⁷⁹, providing a viable approach to develop reliable functionals for transition-metal catalysts in artificial photosynthesis. Additionally, there are multiple ongoing developments in ML that are expected to be particularly valuable for identification of catalytic intermediates and calculations of molecular properties that can be directly compared to X-ray absorption and emission spectroscopies⁸⁰, electron paramagnetic resonance spectroscopy, Mössbauer spectroscopy and magnetic circular dichroism.

ML approaches are promising in the field of artificial photosynthesis and have already been explored in a variety of applications, including the description of interatomic interactions from MD simulations⁸¹. Aside from catalyst design and interatomic interactions, ML methods have been used for photocatalyst screening⁸², UV-vis spectra prediction for molecules in solution⁸³, and for correcting redox potentials in implicit and explicit solvation models⁸⁴. However, as a data-driven approach, ML requires sufficient data to populate the selected feature space, which needs to be generated either from reliable experiments or highly accurate quantum mechanics calculations. Further developments are required to make ML models flexible to ensure faithful representations of input data, transferable across chemical space and scalable to systems of different sizes. Even considering these major challenges, we envision that ML will soon be a truly transformative tool in computational studies of artificial photosynthesis. In addition, emerging quantum computing methods^{12–14} have tremendous potential for applications to artificial photosynthetic systems due to their exponential speed compared with classical computation.

Powerful computational approaches have been developed and applied to study the components of cells for artificial photosynthesis. However, a full atomistic understanding of the overall process remains elusive. Such treatments would be particularly valuable to model Z-schemes and rectification mechanisms that could be essential for development of highly efficient photocatalytic cells. Advancements

in hardware such as graphics processing unit acceleration and massive parallelization, as well as algorithm development such as linear scaling and fragment methods, have already facilitated large-scale calculations in artificial photosynthesis⁸⁵. However, many aspects remain difficult to model and require further development, including modeling dynamical and environmental effects at complex interfaces, photo-induced charge separation and recombination dynamics including non-adiabatic effects and spin crossover transitions, charge and substrate transport mechanisms that could be rate limiting, changes and structural evolution of catalysts under working conditions, and redox properties of intermediates in catalytic cycles. Substantial advances are thus required to develop a rigorous understanding of the components involved in artificial photosynthesis, and guidelines for the design of efficient solar energy conversion devices that could provide humanity with a means of sustainable energy.

References

1. Cramer, C. J. & Truhlar, D. G. Density functional theory for transition metals and transition metal chemistry. *Phys. Chem. Chem. Phys.* **11**, 10757–10816 (2009).
2. Pacchioni, G. Modeling doped and defective oxides in catalysis with density functional theory methods: room for improvements. *J. Chem. Phys.* **128**, 182505 (2008).
3. Scanlon, D. O., Morgan, B. J. & Watson, G. W. Modeling the polaronic nature of p-type defects in Cu₂O: The failure of GGA and GGA+U. *J. Chem. Phys.* **131**, 124703 (2009).
4. Casida, M. E. & Huix-Rotllant, M. in *Density-Functional Methods for Excited States* (eds Ferré, N. et al.) 1–60 (Springer, 2016).
5. Young, K. J. et al. Light-driven water oxidation for solar fuels. *Coord. Chem. Rev.* **256**, 2503–2520 (2012).
6. Negre, C. F. A. et al. Efficiency of interfacial electron transfer from Zn-porphyrin dyes into TiO₂ correlated to the linker single molecule conductance. *J. Phys. Chem. C* **117**, 24462–24470 (2013).
7. Sivula, K. & van de Krol, R. Semiconducting materials for photoelectrochemical energy conversion. *Nat. Rev. Mater.* **1**, 15010 (2016).
8. Rego, L. G. C. & Batista, V. S. Quantum dynamics simulations of interfacial electron transfer in sensitized TiO₂ semiconductors. *J. Am. Chem. Soc.* **125**, 7989–7997 (2003).
9. Yang, K. R. et al. Solution structures of highly active molecular Ir water-oxidation catalysts from density functional theory combined with high-energy X-ray scattering and EXAFS spectroscopy. *J. Am. Chem. Soc.* **138**, 5511–5514 (2016).
10. Doan, H. A. et al. Quantum chemistry-informed active learning to accelerate the design and discovery of sustainable energy storage materials. *Chem. Mater.* **32**, 6338–6346 (2020).
11. Yao, Z. et al. Machine learning for a sustainable energy future. *Nat. Rev. Mater.* **8**, 202–215 (2022).
12. Liu, J., Wan, L., Li, Z. & Yang, J. Simulating periodic systems on a quantum computer using molecular orbitals. *J. Chem. Theory Comput.* **16**, 6904–6914 (2020).
13. Mizuta, K. et al. Deep variational quantum eigensolver for excited states and its application to quantum chemistry calculation of periodic materials. *Phys. Rev. Res.* **3**, 043121 (2021).
14. Cao, Y. et al. Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019).
15. López, I. et al. A self-improved water-oxidation catalyst: is one site really enough? *Angew. Chem. Int. Ed.* **53**, 205–209 (2014).
16. Heyd, J., Peralta, J. E., Scuseria, G. E. & Martin, R. L. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Chem. Phys.* **123**, 174101 (2005).
17. Wang, L., Long, R. & Prezhdo, O. V. Time-domain ab initio modeling of photoinduced dynamics at nanoscale interfaces. *Annu. Rev. Phys. Chem.* **66**, 549–579 (2015).

18. González, L. & Lindh, R. *Quantum Chemistry and Dynamics of Excited States: Methods and Applications* (John Wiley and Sons, 2020).
19. Man, I. C. et al. Universality in oxygen evolution electrocatalysis on oxide surfaces. *ChemCatChem* **3**, 1159–1165 (2011).
20. Seh, Z. W. et al. Combining theory and experiment in electrocatalysis: insights into materials design. *Science* **355**, eaad4998 (2017).
21. Fujishima, A. & Honda, K. Electrochemical photolysis of water at a semiconductor electrode. *Nature* **238**, 37–38 (1972).
22. Stowasser, R. & Hoffmann, R. What do the Kohn–Sham orbitals and eigenvalues mean? *J. Am. Chem. Soc.* **121**, 3414–3420 (1999).
23. Cohen, A. J., Mori-Sánchez, P. & Yang, W. Insights into current limitations of density functional theory. *Science* **321**, 792–794 (2008).
24. Anisimov, V. I., Zaanen, J. & Andersen, O. K. Band theory and Mott insulators: Hubbard U instead of Stoner I. *Phys. Rev. B* **44**, 943–954 (1991).
25. Liechtenstein, A. I., Anisimov, V. I. & Zaanen, J. Density-functional theory and strong interactions: orbital ordering in Mott-Hubbard insulators. *Phys. Rev. B* **52**, R5467–R5470 (1995).
26. Dudarev, S. L., Botton, G. A., Savrasov, S. Y., Humphreys, C. J. & Sutton, A. P. Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study. *Phys. Rev. B* **57**, 1505–1509 (1998).
27. Liao, P. & Carter, E. A. New concepts and modeling strategies to design and evaluate photo-electro-catalysts based on transition metal oxides. *Chem. Soc. Rev.* **42**, 2401–2422 (2013).
28. Meng, X. Y. et al. Enhanced photoelectrochemical activity for Cu and Ti doped hematite: the first principles calculations. *Appl. Phys. Lett.* **98**, 112104 (2011).
29. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
30. Hedin, L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys. Rev.* **139**, A796–A823 (1965).
31. Kulik, H. et al. Roadmap on machine learning in electronic structure. *Electron. Struct.* **4**, 023004 (2022).
32. Runge, E. & Gross, E. K. U. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **52**, 997–1000 (1984).
33. Carter-Fenk, K., Cunha, L. A., Arias-Martinez, J. E. & Head-Gordon, M. Electron-affinity time-dependent density functional theory: formalism and applications to core-excited states. *J. Phys. Chem. Lett.* **13**, 9664–9672 (2022).
34. Hait, D. & Head-Gordon, M. Orbital optimized density functional theory for electronic excited states. *J. Phys. Chem. Lett.* **12**, 4517–4529 (2021).
35. Kümmel, S. Charge-transfer excitations: a challenge for time-dependent density functional theory that has been met. *Adv. Energy Mater.* **7**, 1700440 (2017).
36. Salpeter, E. E. & Bethe, H. A. A relativistic equation for bound-state problems. *Phys. Rev.* **84**, 1232–1242 (1951).
37. Joung, J. F. et al. Deep learning optical spectroscopy based on experimental database: potential applications to molecular design. *JACS Au* **1**, 427–438 (2021).
38. Abuabara, S. G., Rego, L. G. C. & Batista, V. S. Influence of thermal fluctuations on interfacial electron transfer in functionalized TiO₂ semiconductors. *J. Am. Chem. Soc.* **127**, 18234–18242 (2005).
39. Li, C. et al. Facet-dependent photoelectrochemical performance of TiO₂ nanostructures: an experimental and computational study. *J. Am. Chem. Soc.* **137**, 1520–1529 (2015).
40. Menzel, J. P. et al. Photoinduced electron injection in a fully solvated dye-sensitized photoanode: a dynamical semiempirical study. *J. Phys. Chem. C* **124**, 27965–27976 (2020).
41. Jiang, J. et al. Molecular design of light-harvesting photosensitizers: effect of varied linker conjugation on interfacial electron transfer. *Phys. Chem. Chem. Phys.* **18**, 18678–18682 (2016).
42. Liu, C. & Jakubikova, E. Two-step model for ultrafast interfacial electron transfer: limitations of Fermi's golden rule revealed by quantum dynamics simulations. *Chem. Sci.* **8**, 5979–5991 (2017).
43. Oliboni, R. S. et al. Vibronic effects in the ultrafast interfacial electron transfer of perylene-sensitized TiO₂ surfaces. *J. Phys. Chem. C* **123**, 12599–12607 (2019).
44. Komsa, H.-P., Broqvist, P. & Pasquarello, A. Alignment of defect levels and band edges through hybrid functionals: effect of screening in the exchange term. *Phys. Rev. B* **81**, 205118 (2010).
45. Toroker, M. C. et al. First principles scheme to evaluate band edge positions in potential transition metal oxide photocatalysts and photoelectrodes. *Phys. Chem. Chem. Phys.* **13**, 16644–16654 (2011).
46. Snyder, J. C., Rupp, M., Hansen, K., Müller, K.-R. & Burke, K. Finding density functionals with machine learning. *Phys. Rev. Lett.* **108**, 253002 (2012).
47. Letchworth-Weaver, K. & Arias, T. A. Joint density functional theory of the electrode-electrolyte interface: application to fixed electrode potentials, interfacial capacitances, and potentials of zero charge. *Phys. Rev. B* **86**, 075140 (2012).
48. Cheng, J. & Sprik, M. Aligning electronic energy levels at the TiO₂/H₂O interface. *Phys. Rev. B* **82**, 081406 (2010).
49. Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models* (Wiley, 2004).
50. Tomasi, J., Mennucci, B. & Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **105**, 2999–3094 (2005).
51. Basdogan, Y., Maldonado, A. M. & Keith, J. A. Advances and challenges in modeling solvated reaction mechanisms for renewable fuels and chemicals. *WIREs Comput. Mol. Sci.* **10**, e1446 (2020).
52. Pliego, J. R. & Riveros, J. M. The cluster–continuum model for the calculation of the solvation free energy of ionic species. *J. Phys. Chem. A* **105**, 7241–7247 (2001).
53. Jinich, A., Sanchez-Lengeling, B., Ren, H., Harman, R. & Aspuru-Guzik, A. A mixed quantum chemistry/machine learning approach for the fast and accurate prediction of biochemical redox potentials and its large-scale application to 315 000 redox reactions. *ACS Cent. Sci.* **5**, 1199–1210 (2019).
54. Toyao, T. et al. Machine learning for catalysis informatics: recent applications and prospects. *ACS Catal.* **10**, 2260–2297 (2020).
55. Freeze, J. G., Kelly, H. R. & Batista, V. S. Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. *Chem. Rev.* **119**, 6595–6612 (2019).
56. Xu, Y., Yamazaki, M. & Villars, P. Inorganic materials database for exploring the nature of material. *Jpn. J. Appl. Phys.* **50**, 11RH02 (2011).
57. Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
58. Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B* **58**, 364–369 (2002).
59. Jain, A. et al. The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
60. Saal, J. E., Kirklin, S., Aykol, M., Meredig, B. & Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **65**, 1501–1509 (2013).
61. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

62. Palkovits, R. & Palkovits, S. Using artificial intelligence to forecast water oxidation catalysts. *ACS Catal.* **9**, 8383–8387 (2019).
63. Guo, Y. et al. Machine-learning-guided discovery and optimization of additives in preparing Cu catalysts for CO₂ reduction. *J. Am. Chem. Soc.* **143**, 5755–5762 (2021).
64. Zheng, J. et al. Symbolic transformer accelerating machine learning screening of hydrogen and deuterium evolution reaction catalysts in MA2Z4 materials. *ACS Appl. Mater. Interfaces* **13**, 50878–50891 (2021).
65. Takigawa, I., Shimizu, K.-i, Tsuda, K. & Takakusagi, S. Machine-learning prediction of the d-band center for metals and bimetals. *RSC Adv.* **6**, 52587–52595 (2016).
66. Zhang, Y. & Xu, X. Machine learning band gaps of doped-TiO₂ photocatalysts from structural and morphological parameters. *ACS Omega* **5**, 15344–15352 (2020).
67. Spiegelman, F. et al. Density-functional tight-binding: basic concepts and applications to molecules and clusters. *Adv. Phys. X* **5**, 1710252 (2020).
68. Senftle, T. P. et al. The ReaxFF reactive force-field: development, applications and future directions. *NPJ Comput. Mater.* **2**, 15011 (2016).
69. Kmiecik, S. et al. Coarse-grained protein models and their applications. *Chem. Rev.* **116**, 7898–7936 (2016).
70. Samanta, B. et al. Challenges of modeling nanostructured materials for photocatalytic water splitting. *Chem. Soc. Rev.* **51**, 3794–3818 (2022).
71. Blase, X., Duchemin, I. & Jacquemin, D. The Bethe–Salpeter equation in chemistry: relations with TD-DFT, applications and challenges. *Chem. Soc. Rev.* **47**, 1022–1043 (2018).
72. Wiktor, J., Reshetnyak, I., Ambrosio, F. & Pasquarello, A. Comprehensive modeling of the band gap and absorption spectrum of BiVO₄. *Phys. Rev. Mater.* **1**, 022401 (2017).
73. da Silva Oliboni, R., Bortolini, G., Torres, A. & Rego, L. G. C. A nonadiabatic excited state molecular Mechanics/Extended Hückel Ehrenfest method. *J. Phys. Chem. C* **120**, 27688–27698 (2016).
74. Greene, S. M. & Batista, V. S. Tensor-train split-operator fourier transform (TT-SOFT) method: multidimensional nonadiabatic quantum dynamics. *J. Chem. Theory Comput.* **13**, 4034–4042 (2017).
75. Westermayr, J. & Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **121**, 9873–9926 (2021).
76. Rudshteyn, B. et al. Calculation of metallocene ionization potentials via auxiliary field quantum Monte Carlo: toward benchmark quantum chemistry for transition metals. *J. Chem. Theory Comput.* **18**, 2845–2862 (2022).
77. Gaggioli, C. A., Stoneburner, S. J., Cramer, C. J. & Gagliardi, L. Beyond density functional theory: the multiconfigurational approach to model heterogeneous catalysis. *ACS Catal.* **9**, 8481–8502 (2019).
78. Szalay, S. et al. Tensor product methods and entanglement optimization for ab initio quantum chemistry. *Int. J. Quantum Chem.* **115**, 1342–1391 (2015).
79. Kirkpatrick, J. et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* **374**, 1385–1389 (2021).
80. Guda, A. A. et al. Understanding X-ray absorption spectra by means of descriptors and machine learning algorithms. *NPJ Comput. Mater.* **7**, 203 (2021).
81. Behler, J. Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **145**, 170901 (2016).
82. Masood, H., Toe, C. Y., Teoh, W. Y., Sethu, V. & Amal, R. Machine learning for accelerated discovery of solar photocatalysts. *ACS Catal.* **9**, 11774–11787 (2019).
83. Chen, Z., Bononi, F. C., Sievers, C. A., Kong, W.-Y. & Donadio, D. UV-visible absorption spectra of solvated molecules by quantum chemical machine learning. *J. Chem. Theory Comput.* **18**, 4891–4902 (2022).
84. Hruska, E., Gale, A. & Liu, F. Bridging the experiment-calculation divide: machine learning corrections to redox potential calculations in implicit and explicit solvent models. *J. Chem. Theory Comput.* **18**, 1096–1108 (2022).
85. Seritan, S. & Martinez, T. OMSC 2019 nonadiabatic dynamics study of the B850-B800 complex in LH2 using TeraChem Cloud. *VIRT&L-COMM* **19**, VIRT&L-COMM.19.2019.21 (2019).

Acknowledgements

The authors acknowledge support from the Photosynthetic Systems Program, Office of Basic Energy Sciences, United States Department of Energy, under contract no. DESC0001423 (V.S.B.). We thank the National Energy Research Scientific Computing Center and the Yale Center for Research Computing for computing time. We also acknowledge support from the National Science Foundation under grant 2124511 (Centers for Chemical Innovation Phase I: National Science Foundation Center for Quantum Dynamics on Modular Quantum Devices) for the development of quantum computing algorithms applied to photosynthetic systems.

Author contributions

K.R.Y., G.W.K. and V.S.B. reviewed the relevant material, brainstormed solutions and wrote the paper.

Competing interests

The authors declare no competing interest.

Additional information

Correspondence should be addressed to Victor S. Batista.

Peer review information *Nature Computational Science* thanks Dimitrios Pantazis and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Kaitlin McCardle, in collaboration with the *Nature Computational Science* team.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© Springer Nature America, Inc. 2023