

ARTICLE OPEN



Quantum simulation of preferred tautomeric state prediction

Yu Shee^{1,2}, Tzu-Lan Yeh³, Jen-Yueh Hsiao⁴, Ann Yang³, Yen-Chu Lin^{3,5} and Min-Hsiu Hsieh⁴

Prediction of tautomers plays an essential role in computer-aided drug discovery. However, it remains a challenging task nowadays to accurately predict the canonical tautomeric form of a given drug-like molecule. Lack of extensive tautomer databases, most likely due to the difficulty in experimental studies, hampers the development of effective empirical methods for tautomer predictions. A more accurate estimation of the stable tautomeric form can be achieved by quantum chemistry calculations. Yet, the computational cost required prevents quantum chemistry calculation as a standard tool for tautomer prediction in computer-aided drug discovery. In this paper we propose a hybrid quantum chemistry-quantum computation workflow to efficiently predict the dominant tautomeric form. Specifically, we select active-space molecular orbitals based on quantum chemistry methods. Then we utilize efficient encoding methods to map the Hamiltonian onto quantum devices to reduce the qubit resources and circuit depth. Finally, variational quantum eigensolver (VQE) algorithms are employed for ground state estimation where hardware-efficient ansatz circuits are used. To demonstrate the applicability of our methodology, we perform experiments on two tautomeric systems: acetone and Edaravone, each having 52 and 150 spin-orbitals in the Slater Type Orbital - 3 Gaussian (STO-3G) basis set, respectively. Our numerical results show that their tautomeric state prediction agrees with the Coupled Cluster Singles and Doubles (CCSD) benchmarks. Moreover, the required quantum resources are efficient: in the example of Edaravone, we could achieve chemical accuracy with only eight qubits and 80 two-qubit gates.

npj Quantum Information (2023)9:102; <https://doi.org/10.1038/s41534-023-00767-9>

INTRODUCTION

Tautomers are constitutional isomers that spontaneously convert to one another in dynamic equilibrium. The process of this interconversion is called tautomerization. Typical tautomerization involves the movement of a proton from one position to another and rearrangement of a double bond within the molecule. Other types of tautomerisms include annular, ring-chain, and valence tautomerisms^{1–3}. One well-known example of tautomerization, and quite often involved in the field of drug development, is the keto-enol tautomerism, in which the carbonyl double bond (keto form) is interconverted to an alkene double bond (enol form). This is accompanied by the shift of the alpha proton in the keto form to the hydroxyl group in the enol form, as illustrated in Fig. 1.

Tautomerization plays an important role in biological systems. Non-Watson-Crick base pairing can occur due to tautomerization of nucleic acid base pairs. Such nucleic acid mismatches induced by tautomerization result in spontaneous mutagenesis and hence genetic instability^{4,5}. In the case of drug molecules, the movement of proton within the molecules will lead to the conversion from a hydrogen bond donor to a hydrogen bond acceptor or vice versa, which is essential for the analysis of structure-activity relationship (SAR). Moreover, it has been estimated that more than a quarter of marketed drugs can exhibit tautomerism⁶ and analysis of chemical databases showed that 10–30% of potential drug molecules have potential tautomers^{7–11}. Prediction of tautomeric states of compounds of interest is therefore an important subject in the field of computer-aided drug design.

State-of-the-art algorithms for tautomeric prediction usually involve enumeration of possible tautomers, followed by prediction of the dominant form or estimation of the tautomer populations^{12–16}. The estimation of tautomeric ratios in aqueous medium can be achieved by pKa calculations. Such algorithms usually aim

to provide a list of possible tautomeric forms with ratios of the corresponding species, since it is important to estimate the population of the tautomers due to the small differences in their free energies that could be easily compensated by the interaction with proteins. Previous efforts on pKa estimation rely on empirical pKa prediction models. For example, Epik Classic utilizes Hammett-Taft linear free energy relationship (LFER)¹⁷. However, such empirical methods are known to have limited accuracy, partly due to the under-coverage of chemical space in the database^{18,19}. This could be problematic as the lack of experimental parameterization for novel chemical scaffolds prevents accurate prediction against novel compounds. Moreover, experimental challenges in obtaining the relative tautomer stability as well as isolating the distinct tautomeric forms could also be the cause of prediction uncertainty¹⁵. One potential rational solution to this would be prediction methods based on electronic structures. Indeed, more recent focuses have shifted to quantum-mechanics-based (QM) pKa prediction^{20–22}. Such QM-based methods, however, are computationally demanding and may not be practical for applications like large-scale virtual screening. It should also be noted that solvent is another important factor in determining tautomeric ratios in solution.

Computational quantum chemistry provides a more accurate approach to describing molecular structures in drug design than classical force field-based molecular mechanics methods. QM calculations play a critical role for certain problems in the field, such as metal binding in metalloenzymes, reaction mechanism study, covalent inhibitor design²³. Other applications of QM methods include calculation of molecular properties, QM-based descriptor for QSAR, parameterization of forcefields, and estimation of protein-ligand binding affinities^{23–26}. Major QM methods used in the field of computer-aided drug discovery (CADD) can be

¹Department of Chemistry, Yale University, New Haven, CT 06511, USA. ²Bleximo Corp., Berkeley, CA 94720, USA. ³Insilico Medicine Taiwan Ltd., Taipei City 110208, Taiwan. ⁴Hon Hai (Foxconn) Research Institute, Taipei City 114, Taiwan. ⁵Department of Pharmacy, National Yang Ming Chiao Tung University, Taipei City 112304, Taiwan.

✉email: jimmy.lin@insilicomedicine.com; min-hsiu.hsieh@foxconn.com

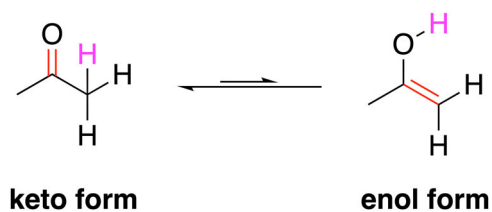


Fig. 1 Keto-enol tautomerization in acetone. This schematic represents keto-enol tautomerization in acetone, highlighting the transfer of a proton (depicted in magenta) and the movement of a double bond (in red). The staggered equilibrium arrows indicate a preference for the keto form.

classified to semi-empirical QM methods, Hartree-Fock (HF), post-Hartree-Fock (pHF) methods, and density functional theory (DFT). While quantum chemistry approaches have brought accuracy to the calculations, their computational cost makes them unsuitable as a daily tool in CADD except for the semi-empirical methods. Depending on the level of theory, the integral transformation could scale as high as $O(M^5)$, where M is the number of atomic basis functions (AOs)²⁷.

Therefore, even though the aforementioned classical quantum chemistry algorithms are well-designed, they still have unreasonable time and space requirements for medium and large molecules²⁸. More often, applications of quantum chemistry methods are used in specific drug design scenarios and the implementation of the QM methods is hard to generalize. This poses severe limitations on the general use of quantum chemistry methods in drug discovery, e.g., prediction of dominant product in tautomerism, understanding the dynamics of protein folding, and ligand binding free energy calculations.

Quantum simulation is considered one of the most promising applications of quantum technology since it has the potential of overcoming the exponential barrier of solving electronic structure problems^{29,30}. Jordan-Wigner (JW) transformation³¹ was found to give mappings from fermionic creation and annihilation operators to qubit operators³². This result paved the road to generically simulate physical systems on well-controlled quantum computers. Moreover, the quantum phase estimation (QPE) algorithm^{33–42} provides accurate spectral calculations for molecular Hamiltonian with the potential of exponential speedup. However, the QPE method is not applicable on current quantum devices because it requires long coherence time and high gate fidelity. Instead, several variational methods^{37,43–53} suitable for noisy intermediate-scale quantum (NISQ) devices were proposed. One of the most practical approaches is the variational quantum eigensolver (VQE) algorithm^{43,44}. VQE is a hybrid quantum-classical method that approximates ground states of molecular Hamiltonian by variationally tuning the ansatz parameters and is expected to give better accuracy than CCSD results with polynomial costs.

Quantum simulation of small molecules has since then been widely implemented. Researchers have worked towards molecules (BeH_2 , H_2O , and H_{12}) with improved algorithms and hardware design^{45,54,55}. Others also provide strategies to simulate slightly larger systems, such as CO_2 , C_2H_4 , C_{18} , and the nitrogenase iron-sulfur molecular clusters^{56–58}, using symmetries, fragmentation of molecules, spin-model simplifications, or qubit encoding methods^{59–64}. However, these quantum simulation results are still quite limited in the problem size and often lack real world applications.

In this work, we aim to design a general methodology for a pharmaceutical application, namely, the prediction of preferred tautomeric states. We expect that our scheme could work for both current and future quantum devices. Specifically, we presented a hybrid quantum chemistry-quantum computation approach to predict the dominant form of tautomers, where quantum chemistry methods are used to construct the system of interest

and reduce the size of the Hamiltonian to meet current quantum hardware requirements, and the reduced Hamiltonian was mapped to qubits using qubit-efficient encoding (QEE)⁶⁴ and subsequently simulated with variational quantum eigensolver (VQE). We focus on predicting the preferred tautomeric states by considering the energetics of the states using quantum chemistry approaches to make the system applicable on current quantum computing schemes. This approach was practiced on two illustrative examples, acetone and Edaravone. Our quantum simulation results showed reasonable agreement with CCSD calculations.

RESULTS

To compare the relative stability of tautomers, one would need to calculate the energies of the systems. Before jumping into detailed calculation of the electronic structures, the most straightforward and hand-wavy way to estimate tautomers' relative stability is by considering the difference in bond dissociation energy for a simple system like the case of acetone tautomerism. The bond dissociation enthalpy⁶⁵ difference between $\text{C}=\text{O}$, $\text{C}-\text{C}$, $\text{C}-\text{H}$ (keto form, $745 + 347 + 413 = 1505$ kcal/mol) and $\text{O}-\text{H}$, $\text{C}-\text{O}$, $\text{C}=\text{C}$ (enol form, $467 + 358 + 614 = 1439$) is 66 kJ/mol (15.8 kcal/mol), favoring the keto form. However, as the complexity of the molecules increases, the relative stability could not be attributed so easily. Factors like electrostatic effect, steric hindrance from the other parts of the molecule and intra-molecular hydrogen bonds need careful consideration as well. Commercial or publicly available CADD tools deal with the tautomerism problem using empirical or rule-based cheminformatics methodologies, e.g., scoring tautomers based on the prediction of microstate and microstate pKa values. There is, however, still room for improvement with such an approach due to the lack of extensive databases and over-parameterization of the prediction models.

In principle, two important factors decide the tautomeric equilibrium. The relative potential energy difference, which determines the equilibrium direction, accounts for the stability between the isomers. On the other hand, the rate of isomer interconversion depends on the activation energy. As tautomerism involves bond breaking and bond formation, the best way to handle this problem should be quantum mechanics-based approaches by considering the energetics of the individual tautomeric forms. Indeed, methods involving implicit solvent density functional theory calculations have been implemented in the tautomer enumeration and scoring workflow⁶⁶. However, such first-principle calculations are impractical for some of the CADD tasks even with the computation resources nowadays. For example, the enumeration of the tautomers and prediction of dominant species of the compound library is an essential step at the very beginning of a virtual screening campaign. A computational bottleneck on DFT calculations prevents practical applications of such algorithms from processing large databases. Here, we suggest an alternative method by a quantum mechanics-based computational scheme with current quantum computing simulators to predict the relative stability of the tautomeric forms. With the advent of quantum computing technologies, quantum chemical approaches could play an important and applicable role for tautomer predictions and computer-aided drug discovery when implemented on future large-scale quantum devices.

In the following, we first present the quantum chemistry overview of the two tautomeric systems: (1) acetone and its enol form, and (2) Edaravone's keto, enol, and amine forms in the STO-3G basis set. Next, we introduce current challenges of simulating medium-to-large molecules and our workflow that could contribute to resolving the problems. Besides, we provide the background of the qubit-efficient encoding methods for quantum simulation. Lastly, we show the numerical results for the implementation of our workflow with quantum simulations of the two tautomeric systems.

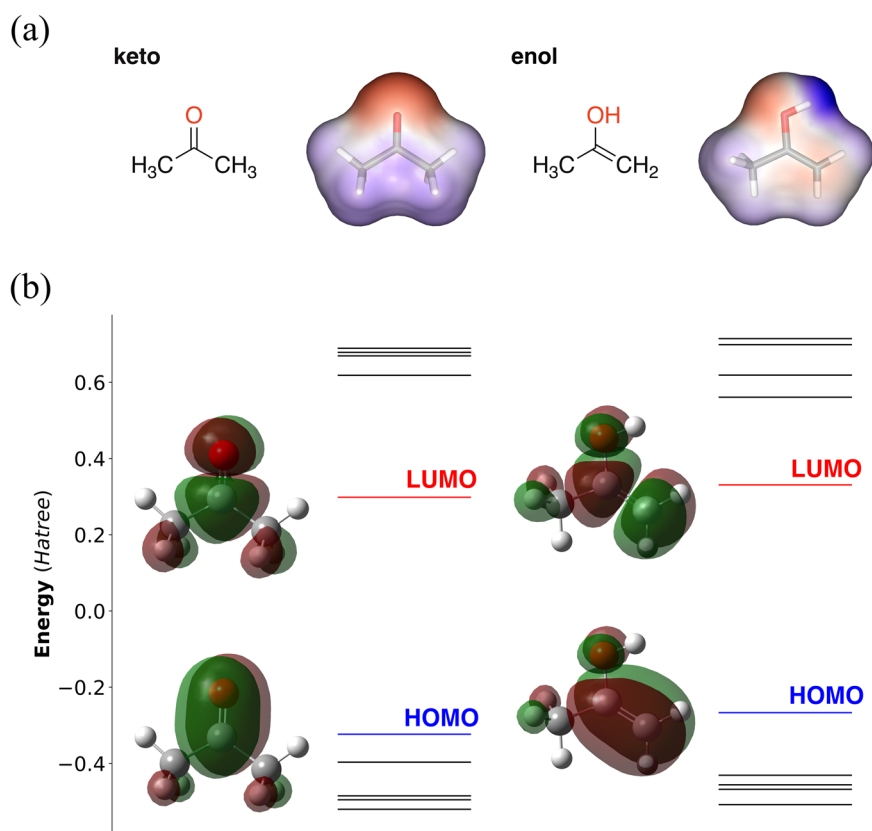


Fig. 2 Acetone tautomers and orbital energies. **a** Tautomeric isomers of acetone. The geometries were optimized at the B3LYP/6-311++G(d,p) level and the electrostatic potential surface generated by ESP_DNN⁹³. **b** Hartree-Fock orbital energy levels around HOMO/LUMO as calculated using the STO-3G basis set. The full list of molecular orbital eigenvalues can be found in Supplementary Information.

Overview of systems

Acetone. We start with a relatively simple example of tautomerization. Acetone exhibits keto-enol tautomerization, in which the acetone and its enol form, propen-2-ol, interconvert to each other (Fig. 2a). The relative stability of these two isomers can be easily assessed by comparing their bond energy, and the keto form is more stable than the enol form by about 16 kcal/mol. Figure 2a shows the optimized geometries of the keto and enol forms of the acetone at the level of 6-311++G(d,p) using B3LYP in the Gaussian 16 program⁶⁷. To enable quantum computing of the system at a reasonable computational cost, we reconstituted the molecules with the minimal STO-3G at the second-order Møller-Plesset perturbation theory (MP2) level with 52 spin-orbitals. The corresponding energy diagrams of the molecular orbitals near HOMO/LUMO are shown in Fig. 2b for the keto and enol form, respectively. The keto form is more stable than the enol form in most cases because the C=O double bond is stronger than the C=C bond. Polarization of the C=O double bond, as depicted in electrostatic potential surface of the acetone, gives it a relatively higher bond energy. The rearrangement of the double bond between the keto and enol form is apparent from the shift of the π bonding patterns of the HOMOs for the corresponding tautomers.

Edaravone. Edaravone is an FDA-approved drug for the treatment of amyotrophic lateral sclerosis (ALS). The mechanism of the action of Edaravone involves its radical scavenging activity. The idea originated from a research program in Mitsubishi Yuka Pharmaceutical Corporation, in which the scientists try to utilize the radical-scavenging activity of phenol and avoid related toxicity. The researchers believed that an aromatic heterocyclic system with the potential of keto-enol tautomerization could exist in the form with a hydroxyl group⁶⁸. Edaravone exists in three

forms of tautomeric isomers - keto, enol, and amine (Fig. 3a), with varying physicochemical properties. Moreover, the three forms of Edaravone exhibit different antioxidant activity. Therefore, it is important to know which form among the three dominates to better understand the pharmacological effect of Edaravone. Estimation of relative stability among the three tautomers for Edaravone is not as trivial as for acetone, since the molecular structure of Edaravone is much more complicated. For example, the aromaticity of the *N*-substituted pyrazolone core in the three isomers varies. Moreover, the surface charge distribution of the three tautomers differs quite a lot and the twisted angles between the pyrazolone and the *N*-substituted benzene are distinct among the three according to the optimized geometries of the forms (at the level of B3LYP/6-311++G(d,p) theory) as depicted in Fig. 3a. Reconstitution of the electronic configuration using the minimal STO-3G basis at MP2 level yields 150 molecular orbitals. Energy diagrams of the molecular orbitals near HOMO/LUMO and the HOMOs and LUMOs of the three tautomeric species do not provide clear pictures for the relative stabilities of the three species as in the case of acetone (Fig. 3b).

Notice that it is already computationally expensive to construct the electronic structures of Edaravone at the MP2/STO-3G level of theory. However, in the field of medicinal chemistry, there are more compounds with larger sizes. It is therefore not an efficient way to predict the thermodynamic stability of different tautomer states by comparing the energies of the isomers using such quantum chemistry calculation directly. Calculations based on wave function approaches for pharmaceutical molecules remains impractical even if the sizes of the system can be recused using methods such as complete active space (CAS) approaches because CI expansion scales exponentially and the affordable active size is generally considered to be limited to 18 electrons in 18 orbitals⁶⁹,

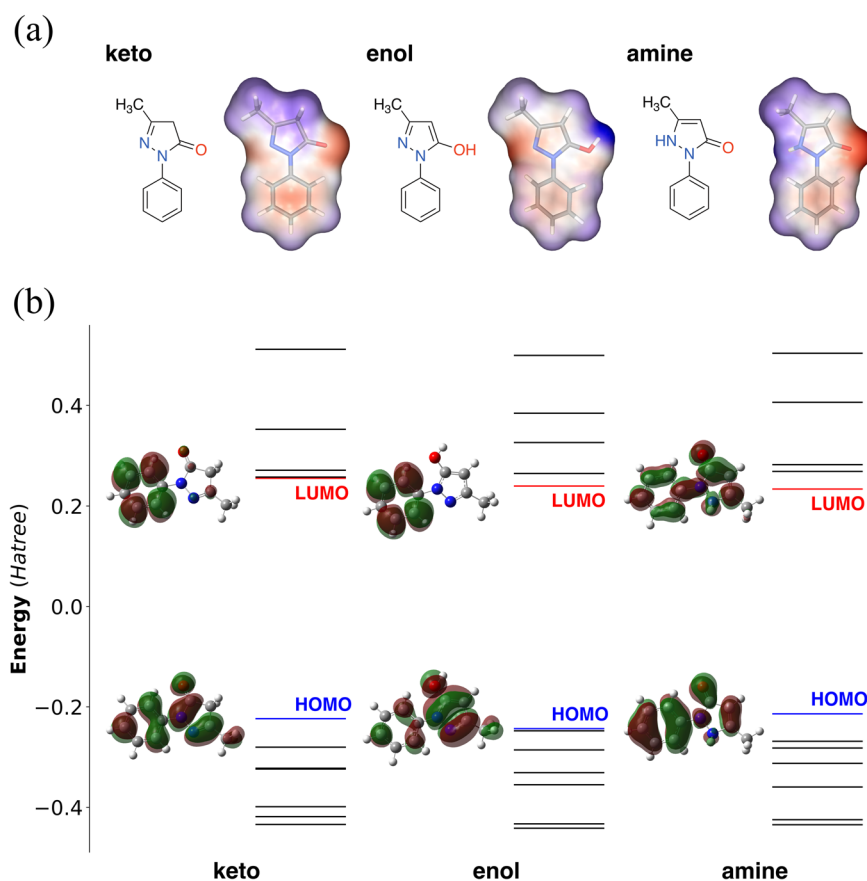


Fig. 3 Edaravone tautomers and orbital energies. **a** Tautomeric isomers of Edaravone. The geometries were optimized at the B3LYP/6-311++G(d,p) level and the electrostatic potential surface generated by ESP_DNN⁹³. **b** Hartree-Fock orbital energy levels around HOMO/LUMO as calculated using the STO-3G basis set. The full list of molecular orbital eigenvalues can be found in Supplementary Information.

although the limit had been pushed to 22 electrons in 22 orbitals with parallel multiconfigurational SCF implementation⁷⁰.

In order to leverage the power of quantum computation, we proceed as follows. Firstly, the second-quantized electronic Hamiltonian can be written as

$$H = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{2} \sum_{pqrs} h_{pqrs} a_p^\dagger a_q^\dagger a_r a_s, \quad (1)$$

where h_{pq} and h_{pqrs} are the overlap and exchange integrals. The indices of the spin-orbitals are represented by p , q , r , and s in the summation of Eq. (1). For example, the indices run from 0 to 51 for acetone and from 0 to 149 for Edaravone because the molecules in the STO-3G basis set have 52 and 150 spin-orbitals respectively. To simulate electronic structure problems on quantum computers, the creation and annihilation operators and the electronic states have to be mapped to qubits.

Common qubit encoding methods are Jordan-Wigner, parity, and Bravyi-Kitaev encoding where the qubit requirements are $\mathcal{O}(N)$ where N is the number of spin-orbitals. For example, Jordan-Wigner (JW) encoding uses N qubits to store the N spin-orbital occupation number. In this encoding scheme, the $|0\rangle$ qubit state represents that the spin-orbital is not occupied while the $|1\rangle$ qubit state represents that the spin-orbital is occupied. The mappings of the creation and annihilation operators for JW are

$$a_p^\dagger = \frac{1}{2}(X_p - iY_p) \otimes Z_{p-1} \otimes \cdots \otimes Z_0, \quad (2)$$

$$a_p = \frac{1}{2}(X_p + iY_p) \otimes Z_{p-1} \otimes \cdots \otimes Z_0, \quad (3)$$

where the Z Pauli strings address the fact that electronic wavefunctions are anti-symmetric. For acetone and Edaravone in the STO-3G basis set, JW encoding will map the electronic structure problems to 52 qubits and 150 qubits. No meaningful data would be collected for VQE using that many qubits on NISQ devices because of decoherence. It is, therefore, important to reduce the problem size.

Workflow

Here we propose a hybrid computational scheme for the prediction of the preferred tautomeric state, comprising classical quantum chemistry (using classical computers) and quantum computation. The workflow involves (1) geometry optimization of the isomers using density functional theory, (2) reconstruction of the molecular orbital space in a minimal basis set, (3) reduction of the MO space, (4) single point energy calculations of active MO sets, (5) selection of the most representative active MO set, and finally (6) energy calculation in Quantum computing scheme using QEE, as illustrated in Fig. 4. We use the CCSD energies of the full systems as the benchmark of our workflow performance.

Specifically, to reduce the problem size, we use quantum chemistry methods to select active molecular orbitals (MOs) and map the Hamiltonian onto fewer qubits using a qubit-efficient encoding method. The geometries of the tautomeric isomers are prepared and initially optimized using Avogadro⁷¹ with the MMFF94 force field^{72–76} and further optimized at the level of 6-311++G(d,p) using B3LYP^{77,78} in the Gaussian 16 program. Due to limited resources for quantum computing, the molecules are reconstituted with the minimal STO-3G at the MP2 level⁷⁹,

followed by the calculation of occupancies of natural orbitals (NOs) using Gaussian 16. The STO-3G basis set has been widely used in several pioneering works of computation chemistry with quantum computing scheme^{45,54,55}. Although the computational cost has been greatly reduced with the minimal basis set, the number of required qubits still exceeds the affordable amount of current QC simulators or real quantum devices.

The reduction of the MO space is then achieved by considering the occupation numbers and quantum hardware limitations. The selection of active space was achieved mostly by excluding the inner core and the outer virtual orbitals. In practice, the natural orbitals with occupancy closest to 0 (completely virtual) or 2 (fully occupied) were frozen or removed. Besides, quantum hardware constraints and the efficiencies of the qubit-efficient encoding method are also considered in the selection process. The reduced second-quantized Hamiltonian is then encoded onto qubits.

The natural orbitals for Acetone and Edaravone tautomers calculated at the MP2/STO-3G level can be found in Supplementary Information. Selected combinations of active MOs are listed in Supplementary Information for acetone and Edaravone, respectively. After the active space selection from quantum chemistry methods, we have 12 spin-orbitals (6 molecular orbitals) for acetone, propen-2-ol, and the three tautomers of Edaravone. With Jordan-Wigner (JW) encoding, the qubit count to simulate the quantum systems is 12 qubits which is still a difficult task for noisy intermediate-scale quantum devices without error mitigation. Therefore, we use qubit-efficient encoding (QEE) to encode the systems onto 8 qubits shown in Table 1 because it provides fewer qubit counts and renders the systems to be suitable for the use of hardware-efficient ansatzes. We then use a hardware-efficient ansatz to simulate the systems with variational quantum eigensolver (VQE).

Even though the qubit requirement is reduced by classical active space selection methods, it is still hard to be implemented on a quantum device as the circuit depth would not permit. For JW encoding, some unphysical (e.g. not particle conserving or violating other symmetries) electronic configurations are also encoded. This necessitates the usage of chemical-inspired ansatzes (e.g. the unitary couple cluster (UCC) ansatz) that often have larger circuit depth so that the trial wavefunction only represents the states in the chemical subspace. It is still possible to use a hardware-efficient ansatz for JW encoded problems, but the number of entangling layers has to be large. With more entangling layers, the ansatz circuit would have better

expressibility⁸⁰ and entanglement to include the solution space⁸¹, but it may suffer from the vanishing gradient and the barren plateau problem^{82–84}. Thus, we resort to a qubit-efficient encoding (QEE) method proposed in⁶⁴ which provides an logarithmic saving of qubit counts (8 qubits for QEE). Additionally, QEE only encodes the physical (significant) electronic configurations such that low depth hardware-efficient ansatzes are suitable for QEE encoded problems.

Variational quantum circuits and parameter initialization. After obtaining the qubit Hamiltonian from QEE, we employ two different hardware-efficient ansatz, as illustrated in Fig. 5, for the VQE algorithm. The first hardware-efficient ansatz consists of four two-qubit entangling building blocks, and the layers are arranged in a staggered form where Fig. 5a gives an example of two hardware-efficient layers. The first ansatz used in this work consists of at most 20 layers and eight parameterized R_y rotations at each layer in the end. Therefore, the largest circuit in this form has 80 CNOT gates and 168 parameters. While for the second/alternative ansatz, each layer consists of seven CNOT gates and eight parameterized R_y rotations shown in Fig. 5b. The largest circuit in this form consists of 10 layers and has 70 CNOT gates and 80 parameters. We have also run this alternative ansatz circuit on a noisy simulator where the circuit consists of four layers and has 28 CNOT gates and 32 parameters.

For the initial state, we have compared two initialization

Molecule	Active Space Orbitals	(Electrons, MO)	JW Qubit Count	QEE Qubit Count
acetone	14-19	(4e, 6o)	12	8
propen-2-ol	14-19	(4e, 6o)	12	8
Edaravone (keto)	44-49	(4e, 6o)	12	8
Edaravone (enol)	44-49	(4e, 6o)	12	8
Edaravone (amine)	44-49	(4e, 6o)	12	8

Calculations were done in the STO-3G basis sets and equilibrium bond distances with some molecular orbitals being frozen/removed. Note that the molecular orbital indices are ordered from the lowest to the highest energies and the indices start from 0.

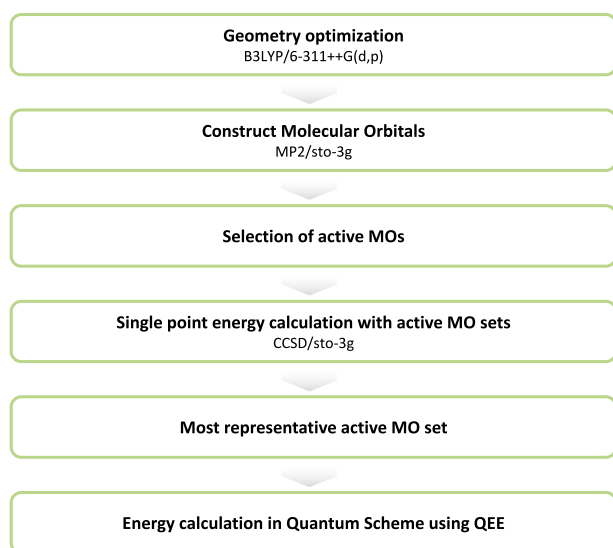


Fig. 4 Proposed workflow of preferred tautomeric state prediction.

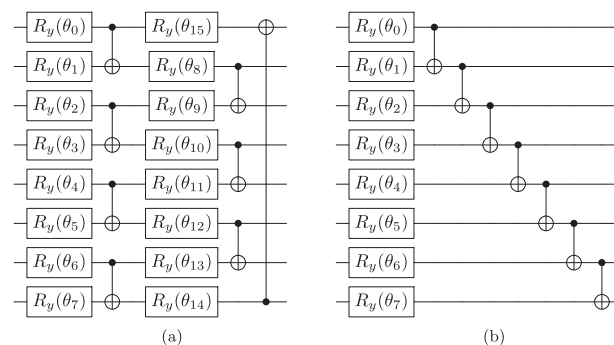


Fig. 5 The hardware-efficient ansatz circuit layers used in this work. **a** Two hardware-efficient layers of the 8-qubit ansatz circuit. Note that there are single-qubit parametrized R_y rotations on all the qubits at the end of this ansatz circuit. **b** One hardware-efficient layer of the alternative 8-qubit ansatz circuit. Note that there is no single-qubit rotation at the end of this ansatz circuit.

strategies. First, we have used the Hartree-Fock (HF) state as the initial state where all the parameters are set to zero initially⁶⁴. We have also compared HF initialization with the Gaussian initialization strategy proposed in ref. ⁸⁵. Zhang et al.⁸⁵ proposed this Gaussian initialization strategy to escape from barren plateau where the initial parameters are sampled from a Gaussian distribution. The Gaussian distribution used in this work has a mean of 0 and a variance of 0.3.

Numerical results

We investigate the performance of our workflow with several numerical experiments. First, the CCSD energies for the full systems of the tautomers in the STO-3G basis set are calculated and are considered our benchmark for energy estimation performance. Second, energy calculations of VQE with hardware-efficient ansatz on a noiseless quantum simulator are performed for both the acetone and the Edaravone tautomers. Third, VQE energies on a noiseless quantum simulator for the acetone system using an alternative ansatz circuit are obtained. Lastly, a noisy simulation (with the noise model detailed in Supplementary Information) of the acetone system has been done using four layers of the alternative ansatz circuit.

The CCSD energies of the full systems and the active-space systems in the STO-3G basis set are computed. For the acetone system (Table 2), the full system CCSD energy of the enol form relative to the keto form is 24.070 kcal/mol, and the active-space system CCSD energy of the enol form relative to the keto form is 23.766 kcal/mol. For the Edaravone system (Table 3), the full

Table 2. Relative ground state energies of acetone and propen-2-ol in kcal/mol calculated by CCSD and with the VQE results calculated on a noisy simulator.

	Acetone	Propen-2-ol
Full system CCSD energies relative to acetone	0	24.070
Active-space CCSD energies relative to acetone	0	23.766
Exact active-space system energies relative to acetone	0	23.773
Active-space system VQE energies relative to acetone	0	22.801
Active-space system VQE Error	2.787	1.805

Relative energies for both the full MO and the active space (HOMO-1, HOMO, LUMO, LUMO+1, LUMO+2, and LUMO+3) were shown. Various combinations of orbital reduction are provided in Supplementary Information. Exact energies of the active-space systems were calculated by the diagonalization of the Hamiltonian. VQE energies were calculated using a 4-layer hardware-efficient ansatz (Fig. 5b) with Gaussian initialization for the parameters on a noisy simulator (see Supplementary Information). VQE errors were calculated from the energy difference between VQE energies and the exact energies of the active-space systems.

Table 3. Relative ground state energy of the keto, enol, and amine form of Edaravone in kcal/mol.

	Keto	Enol	Amine
Full system CCSD energies relative to keto form	0.000	13.726	25.947
Active-space CCSD energies relative to keto form	0.000	9.197	25.398

Relative energies for both the full MO and the active space (HOMO-1, HOMO, LUMO, LUMO+1, LUMO+2, and LUMO+3) were shown. Various combinations of orbital reduction are provided in Supplementary Information.

system CCSD energy of the enol form relative to the keto form is 13.726 kcal/mol and the CCSD energy of the amine form relative to the keto form is 25.947 kcal/mol. While for active-space system of Edaravone, the CCSD energy of the enol form relative to the keto form is 9.197 kcal/mol and the CCSD energy of the amine form relative to the keto form is 25.398 kcal/mol.

For the VQE results of acetone shown in Fig. 6, we use two different parameter initialization strategies (Gaussian and HF) and investigate how different number of ansatz circuit layers affect VQE calculations. For acetone in Fig. 6a with 20 circuit layers using Gaussian initialization, the keto form has an error of 0.135 kcal/mol and the enol form has an error of 0.258 kcal/mol which are well below chemical accuracy (1 kcal/mol). It can be seen that more entangling ansatz layers provides lower VQE error (see Fig. 6a) and better estimations of the relative energy gaps between the keto and enol forms of acetone (see Fig. 6b). Besides, VQE calculations of active-space systems have the same trends with CCSD calculations of the full systems for the predictions of stability between the tautomers (see the dotted lines in Fig. 6b). However, for the case where initial states are HF states, the VQE results are often found to be trapped in local minimum or barren plateau. This can be clearly observed from the keto form of acetone in Fig. 6c where the final states are almost the same as the initial HF state. While for the case where Gaussian initialized parameters are used, the VQE shows faster convergence to chemical accuracy (require lower circuit depth) where only 12 circuit layers are needed to reach chemical accuracy. Also, the relative energy gaps between the keto and enol forms of acetone are not accurate when using HF initialization (see Fig. 6d).

Similar VQE results can be found in Fig. 7 for the Edaravone system. For Edaravone in Fig. 7a with 20 circuit layers, the keto form has an error of 0.730 kcal/mol, the enol form has an error of 1.006 kcal/mol, the amine form has an error of 1.168 kcal/mol which are fairly close to chemical accuracy. Figure 7a also shows that more entangling ansatz layers often provides lower VQE error and better estimations of the relative energy (Fig. 7b). For the case of HF initialization for Edaravone, the VQE errors and calculations of relative energy gaps also reach similar accuracy in Fig. 7c and Fig. 7d. However, it can still be observed that Gaussian initialization has a better convergence with respect to the number of circuit layers.

Additionally, we tested the performance of an alternative ansatz (see Fig. 5b for a single layer) on the acetone system. The results in Fig. 8 show that the energy gap estimation has reached chemical accuracy with just 4 entangling layers (28 two-qubit gates). The circuit depth and the two-qubit gate count are reasonable for near-term quantum devices. Thus, we have run the circuit with the parameters being set at the optimal point on a noisy simulator (the noise model is based on the thermal relaxation channel with specific values listed in Supplementary Information). The results for noisy simulation can be found in Table 2 where the VQE errors and the energy gaps are found to be close to chemical accuracy.

DISCUSSION

In this work, we have proposed a hybrid quantum-classical workflow for the prediction of the relative stability of tautomers. It is of the pharmaceutical industry's interest to accurately predict the preferred tautomeric state of a given molecules and its computation methodology. However, the computational requirements using classical approaches for quantum chemistry are too expensive, so we have made use of quantum simulation for electronic structure calculations.

We are aware of some studies that are making the effort to realize quantum simulation as a practical application for quantum chemistry. Recently, Tazhigulov et al.⁵⁸ simulated strongly-correlated molecules that are more relevant to real world problems on Google's Sycamore quantum processor. They

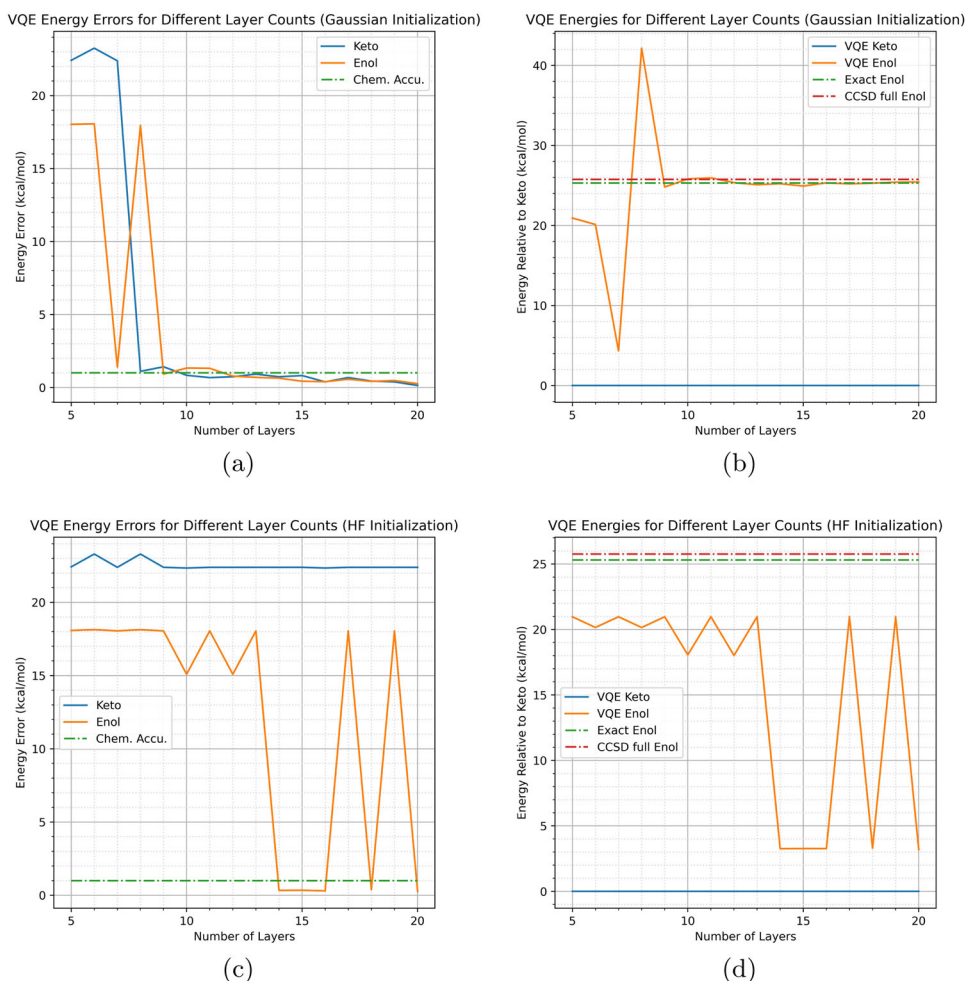


Fig. 6 The relationship between acetone VQE results and the number of hardware-efficient ansatz layers. **a**, **b** are the results when Gaussian-initialized parameters are used while **c** and **d** are the results when initial states are the HF states. All the x-axes represent the number of hardware-efficient circuit layers. For **a** and **c**, the y-axes represent VQE energy errors of the keto or enol form of acetone, and the dashed lines represent chemical accuracy (Chem. Accu.). For **b** and **d**, the y-axes represent the energies relative to the keto form (so VQE energies, CCSD energies, and exact energies of the keto form are set to zero). Note that the exact energies (Exact in the legend) are the results from exact diagonalization of the active-space Hamiltonian for each tautomer, and CCSD full in the legend represents the CCSD energy of the full system relative to the keto form.

mapped the electronic structures of iron-sulfur molecular clusters and α -RuCl₃ into low-energy spin models using the results from theoretical and spectroscopic studies. This simplification reduced the qubit requirements and rendered the possibility of simulating larger molecules. They then used the finite-temperature version of quantum imaginary time evolution to capture physical observables. However, the circuit depth for the evolution were too large for the quantum device. Thus, they recompiled the circuit on a classical noiseless emulator, and numerous error mitigation techniques were employed. One of their experiments that used 9 qubits and 82 two-qubit gates achieved acceptable accuracy, while the data for the experiment that used 11 qubits and 310 two-qubit gates were not quantitatively meaningful. This research benchmarked the performance and limitation of quantum simulation on quantum hardware. It is, therefore, important to reduce hardware requirements of physical simulation to this limitation (e.g. using similar qubit and two-qubit gate counts).

Instead of utilizing empirical information, we have used ab initio quantum chemistry methods to reduce quantum hardware requirements. Besides, the problem of interest is to find the relative ground state energy differences rather than absolute energies or other observables, so there are more rooms for orbital reduction. First, we optimize the geometry using B3LYP. Next, we

use the natural orbitals from MP2 to select active orbitals. The selection criteria includes the occupation numbers and hardware limitation. To exploit as much quantum resources, we map the reduced Hamiltonian onto fewer qubits using QEE. Finally, we simulate the molecules/tautomers with 12 active spin-orbitals using VQE with 8 qubits and 80 two-qubit gates, which is fairly close to the quantum hardware limitation from⁵⁸ (9 qubits and 82 two-qubit gates). In this work, not much efforts were done on exploring the best ansatz circuit for the VQE simulation, but we have found that an alternative ansatz circuit with 28 two-qubit gates can sufficiently provide accurate VQE calculations for the acetone system. This two-qubit gate count is much fewer than the quantum hardware limitation from⁵⁸ which shows the applicability of our methodology.

The VQE results for predicted stability agrees with those from full system CCSD results. For the prediction of acetone, its keto form is more stable than its enol form. For the case of Edaravone, its keto form is the most stable and its amine form is the most unstable. Note that these results did not incorporate solvent effects, which is an important factor in consideration of tautomerism in biological systems. Continuum solvation models, such as Polarizable Continuum Model (PCM)⁸⁶, the Solvation Model based on Density (SMD)⁸⁷ or the Conductor-like Screening

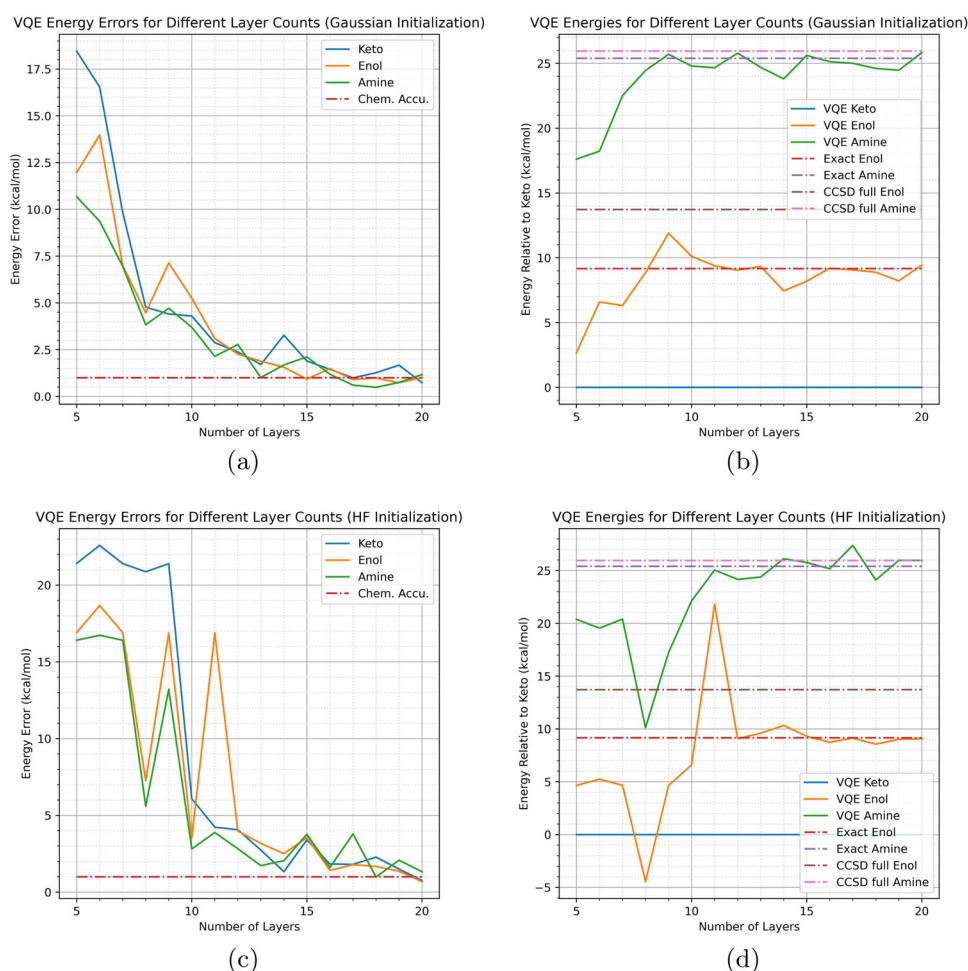


Fig. 7 The relationship between Edaravone VQE results and the number of hardware-efficient ansatz layers. **a, b** are the results when Gaussian-initialized parameters are used while **c** and **d** are the results when initial states are the HF states. All the x-axes represent the number of hardware-efficient circuit layers. For **a** and **c**, the y-axes represent VQE energy errors of the keto, enol, or amine form of Edaravone, and the dashed lines represent chemical accuracy (Chem. Accu.). For **b** and **d**, the y-axes represent the energies relative to the keto form (so VQE energies, CCSD energies, and exact energies of the keto form are set to zero). Note that the exact energies (Exact in the legend) are the results from exact diagonalization of the active-space Hamiltonian for each tautomer, and CCSD full in the legend represents the CCSD energies of the full systems relative to the keto form.

Model (COSMO)⁸⁸, can be used to take into account the solvent effects. For example the PCM-VQE algorithm provides a self-consistent way to include the solvent effect which does not require extra quantum resources.

The CCSD results are considered to be the solution for the comparison with VQE results since FCI is too expensive for such large systems. Nevertheless, there could be even larger systems where CCSD starts becoming intractable. This is where our hybrid quantum-classical workflow could come into play because MP2 (which provides natural orbital occupancy) is computationally cheaper than CCSD and quantum simulation has the potential of becoming a more advantageous approach.

Note that even though the VQE data are from noiseless simulator, there are still some VQE errors. The reasons could be that either the ansatz circuit could not capture the numerically true ground states or the parameter optimization stopped at local minimum or got stuck on barren plateau. Nevertheless, the errors are within chemical accuracy and does not affect the stability prediction.

The inclusion of active space selection process in this work alleviates the burden on quantum resources. Not too much has been emphasized on the selection of active space in this work. The selection of CAS is itself a sophisticated work and remains a

challenge in the field⁸⁹. Future work will focus on generalization of active space selection for the tautomerization process. Potential focuses could be placed on the π -orbitals since the process involves the migration of double bonds or on the atom types that are heavily involved in the tautomerization process such as N or O. Similar complete active space concept has been used for generation of tautomers by micro-pKa predictions⁹⁰. Of course a quantum device with more qubits that are less error-prone can bring value into quantum chemistry, but we are not close to achieving large-scale fault-tolerant in any architecture. As quantum computers will certainly be improved, the active orbital selection criteria can be adjusted such that more orbitals can be included. Such versatility stems from the fact that the limitation of quantum hardware is included in the selection criteria.

Given the importance of tautomerism in pharmaceutical research, our work aims to extend the applications of quantum computing in drug discovery and predictions of chemical events. To realize useful applications for quantum computers in near term, hybrid methods and reasonable classical preprocessing are necessary. On the other hand, the classical methods should be adaptable to the advancement of future quantum computers. Certain modifications on the quantum subroutines can also be done to improve the whole picture. We have employed QEE to

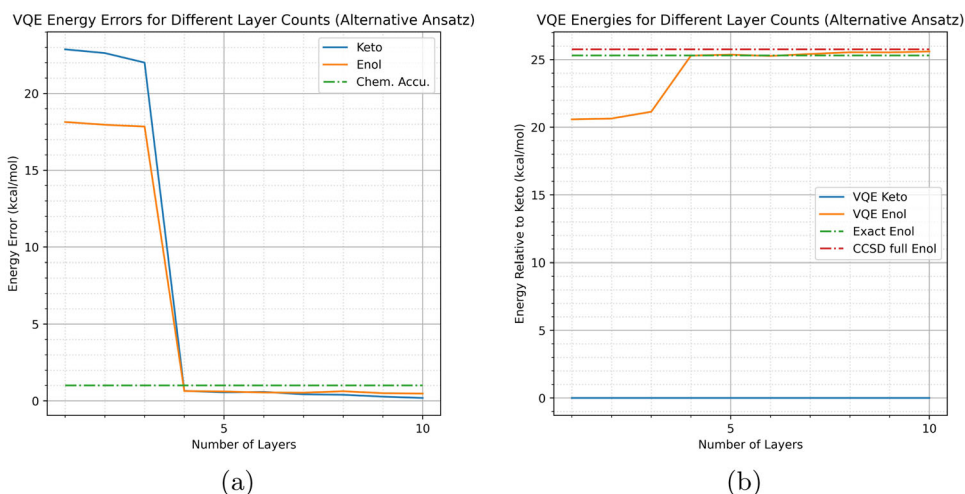


Fig. 8 The relationship between acetone VQE results and the number of the alternative ansatz layers from the main text. Gaussian-initialized parameters are used. All the x-axes represent the number of hardware-efficient circuit layers. For **a**, the y-axis represents VQE energy errors of the keto or enol form of acetone, and the dashed line represents chemical accuracy (Chem. Accu.). For **b**, the y-axis represents the energies relative to the keto form (so VQE energies, CCSD energies, and exact energies of the keto form are set to zero). Note that the exact energies (Exact in the legend) are the results from exact diagonalization of the active-space Hamiltonian for each tautomer, and CCSD full in the legend represents the CCSD energy of the full system relative to the keto form.

increase the maximal number of spin-orbitals that can be simulated on limited quantum resources. Besides, to avoid the local minimum/barren plateau problem seen in this work, one could employ some parameter optimization strategies such as the initialization strategy proposed in refs. ⁸⁵ and ⁹¹. One could also change the VQE subroutine to methods such as imaginary time evolution⁹² for better convergence or quantum algorithms with better scaling when future hardware admits.

METHODS

Qubit efficient encoding

As we want to work on a larger electronic system, an encoding scheme using $\mathcal{O}(N)$ qubits (like JW encoding) would not be practical on current and near-future devices. Therefore, it would be suitable to resort to encoding methods using fewer qubits. Here, we choose Qubit Efficient Encoding (QEE) from ref. ⁶⁴ where the qubit scales logarithmically with respect to N . This is done by only mapping particle conserving or other electronic configurations with certain symmetries to the qubit basis states which exploits the exponential growth of qubit Hilbert space.

Since the encoding scheme works in a subspace of the space spanned by the occupation basis, we will not be able to map a single creation or annihilation operator as the operators change the number of electrons in the system. Nevertheless, the second-quantized Hamiltonian can be written as a linear combination of excitation operators, which conserve the number of particles, $E_{pq} \equiv a_p^\dagger a_q$ yielding

$$H = \sum_{pq} h_{pq} E_{pq} + \frac{1}{2} \sum_{pqrs} h_{pqrs} (\delta_{qr} E_{ps} - E_{pr} E_{qs}) \quad (4)$$

by using the anti-commutation relations of fermionic operators. Any excitation operator E_{pq} can be written as $E_{pq} = \sum_{k,k'=0}^{|\mathcal{F}_m|-1} c_{k'k}^{pq} |\mathbf{f}_{k'}\rangle_f \langle \mathbf{f}_k|_f$, where \mathcal{F}_m is the set of all particle conserving (or with other symmetries) electronic configurations $|\mathbf{f}_k\rangle_f$ and $c_{k'k}^{pq} = \langle \mathbf{f}_{k'}|_f E_{pq} |\mathbf{f}_k\rangle_f$. $c_{k'k}^{pq}$ is zero if the transition from $|\mathbf{f}_k\rangle_f$ to $|\mathbf{f}_{k'}\rangle_f$ via excitation operator E_{pq} is impossible. The coefficient $c_{k'k}^{pq}$ can be ± 1 according to the antisymmetric property of fermions where $c_{k'k}^{pq} = \prod_{i=\min(p,q)+1}^{\max(p,q)-1} (-1)^{f_i}$. With a transformation

\mathcal{E} that maps the selected electronic configurations one-to-one to qubit basis states in an arbitrary order, the corresponding qubit operator of E_{pq} can be written as

$$\tilde{E}_{pq} = \mathcal{E} \circ E_{pq} \circ \mathcal{E}^{-1} = \sum_{k,k'=0}^{|\mathcal{F}_m|-1} c_{k'k}^{pq} |\mathbf{q}_{k'}\rangle_q \langle \mathbf{q}_k|_q, \quad (5)$$

where $|\mathbf{q}_k\rangle_q = \mathcal{E}|\mathbf{f}_k\rangle_f$ is the encoded qubit state of $|\mathbf{f}_k\rangle_f$ under the map \mathcal{E} . The transition between the two basis states $|\mathbf{q}\rangle_q = |q_{Q-1}, \dots, q_0\rangle_q$ and $|\mathbf{q}'\rangle_q = |q'_{Q-1}, \dots, q'_0\rangle_q$ can be factorized as $|\mathbf{q}'\rangle_q \langle \mathbf{q}|_q = \bigotimes_{k=0}^{Q-1} |q'_k\rangle_q \langle q_k|_q$. Therefore, the encoded excitation operator can be expressed as

$$\tilde{E}_{pq} = \sum_{k,k'=0}^{|\mathcal{F}_m|-1} c_{k'k}^{pq} |\mathbf{q}_{k'}\rangle_q \langle \mathbf{q}_k|_q = \sum_{k,k'=0}^{|\mathcal{F}_m|-1} \bigotimes_{w=0}^{Q-1} c_{k'k}^{pq} T_{k'k,w}, \quad (6)$$

where $T_{k'k,w}$ could be one of the following operators

$$Q^+ = |1\rangle_q \langle 0|_q = \frac{1}{2}(X - iY), \quad (7)$$

$$Q^- = |0\rangle_q \langle 1|_q = \frac{1}{2}(X + iY), \quad (8)$$

$$N^{(0)} = |0\rangle_q \langle 0|_q = \frac{1}{2}(I + Z), \quad (9)$$

$$N^{(1)} = |1\rangle_q \langle 1|_q = \frac{1}{2}(I - Z) \quad (10)$$

according to the qubit basis state transitions.

As each operator $T_{k'k,w}$ is a sum of two Pauli operators, this allows us to express the qubit Hamiltonian

$$H_q = \sum_{pq} h_{pq} \tilde{E}_{pq} + \frac{1}{2} \sum_{pqrs} h_{pqrs} (\delta_{qr} \tilde{E}_{ps} - \tilde{E}_{pr} \tilde{E}_{qs}) \quad (11)$$

as a sum of Pauli operator strings. With the Hamiltonian in Pauli strings, the energy expectation value can be obtained using a quantum computer. In this work, we select electronic configurations that are particle-conserving and singlet so we reduce the qubit requirement from 12 qubits to 8 qubits. In the original QEE paper, the authors discovered that the number of qubit

Hamiltonian terms is bounded as follows: when $m \leq \frac{N}{2}$, the upper bound is $\mathcal{O}\left(\frac{N^{2m+1}}{(m-1)!m!}\right)$ and when $m > \frac{N}{2}$, the upper bound is $\mathcal{O}\left(\frac{N^{2(N-m)+1}}{(N-m-1)!(N-m)!}\right)$. Here, N represents the number of spin-orbitals, and m represents the number of electrons. It is worth noting that this scaling is generally higher than $\mathcal{O}(N^4)$ for JW encoding. Nevertheless, the authors present several scenarios in which QEE yields significant reductions in qubit counts while maintaining similar or even fewer numbers of Hamiltonian terms. The QEE paper provides use cases where 4 to 8 qubits are utilized after obtaining the QEE Hamiltonian, which is also the case for this work.

Variational parameter optimization

For noiseless VQE simulation, the statevector simulator from Qiskit was used, and the classical optimization of the variational parameters was done using the Sequential Least Squares Programming (SLSQP) method with 50k iterations. For noisy VQE simulation, the statevector simulator from Qiskit was used with 100k measurement shots per circuit. There is no classical optimization of the variational parameters for noisy simulation as we use the optimal circuit parameters from noiseless simulation.

DATA AVAILABILITY

The optimized structures of the tautomers are available as GitHub repository at <https://github.com/randyshee/tautomer-quantum-simulation>.

CODE AVAILABILITY

The implementation of the quantum simulation experiments is available as GitHub repository at <https://github.com/randyshee/tautomer-quantum-simulation>.

Received: 28 November 2022; Accepted: 27 September 2023;

Published online: 16 October 2023

REFERENCES

- Antonov, L. *Tautomerism: Concepts and Applications in Science and Technology* (Wiley Online Library, 2016).
- Muller, P. Glossary of terms used in physical organic chemistry (IUPAC recommendations 1994). *J. Macromol. Sci. Part A Pure Appl. Chem.* **66**, 1077–1184 (1994).
- Alkorta, I., Goya, P., Elguero, J. & Singh, S. P. A simple approach to the tautomerism of aromatic heterocycles. *Natl Acad. Sci. Lett.* **30**, 139 (2007).
- Wang, W., Hellinga, H. W. & Beese, L. S. Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl Acad. Sci. USA* **108**, 17644–17648 (2011).
- Bebenek, K., Pedersen, L. C. & Kunkel, T. A. Replication infidelity via a mismatch with Watson–Crick geometry. *Proc. Natl Acad. Sci. USA* **108**, 1862–1867 (2011).
- Martin, Y. C. Let's not forget tautomers. *J. Comput. Aided Mol. Des.* **23**, 693–704 (2009).
- Dhaked, D. K., Ihlenfeldt, W.-D., Patel, H., Delannée, V. & Nicklaus, M. C. Toward a comprehensive treatment of tautomerism in cheminformatics including in InChI V2. *J. Chem. Inf. Model.* **60**, 1253–1275 (2020).
- Cruz-Cabeza, A. J. & Groom, C. R. Identification, classification and relative stability of tautomers in the Cambridge structural database. *CrystEngComm* **13**, 93–98 (2010).
- Guasch, L. et al. Experimental and cheminformatics study of tautomerism in a database of commercially available screening samples. *J. Chem. Inf. Model.* **56**, 2149–2161 (2016).
- Milletti, F., Storch, L., Sforza, G., Cross, S. & Cruciani, G. Tautomer enumeration and stability prediction for virtual screening on large chemical databases. *J. Chem. Inf. Model.* **49**, 68–75 (2009).
- Sitzmann, M., Ihlenfeldt, W.-D. & Nicklaus, M. C. Tautomerism in large databases. *J. Comput. Aided Mol. Des.* **24**, 521–551 (2010).
- Warr, W. A. Tautomerism in chemical information management systems. *J. Comput. Aided Mol. Des.* **24**, 497–520 (2010).
- Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **24**, 591–604 (2010).
- Rupp, M., Korner, R. & V. Tetko, I. Predicting the pKa of small molecules. *Comb. Chem. High Throughput Screen.* **14**, 307–327 (2011).
- Martin, Y. C. Experimental and pKa prediction aspects of tautomerism of drug-like molecules. *Drug Discov. Today Technol.* **27**, 59–64 (2018).
- Navo, C. D. & Jiménez-Osés, G. Computer prediction of pKa values in small molecules and proteins. *ACS Med. Chem. Lett.* **12**, 1624–1628 (2021).
- Shelley, J. C. et al. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21**, 681–691 (2007).
- Settimo, L., Bellman, K. & Knegtel, R. M. A. Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds. *Pharm. Res.* **31**, 1082–1095 (2014).
- Balogh, G. T., Tarcsay, A. & Keserü, G. M. Comparative evaluation of pKa prediction tools on a drug discovery dataset. *J. Pharm. Biomed. Anal.* **67–68**, 63–70 (2012).
- Bochevarov, A. D., Watson, M. A., Greenwood, J. R. & Philipp, D. M. Multi-conformation, density functional Theory-Based pKa prediction in application to large, flexible organic molecules with diverse functional groups. *J. Chem. Theory Comput.* **12**, 6001–6019 (2016).
- Pracht, P., Wilcken, R., Udvarhelyi, A., Rodde, S. & Grimme, S. High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pKa values in the context of the SAMPL6 challenge. *J. Comput. Aided Mol. Des.* **32**, 1139–1149 (2018).
- Geballe, M. T., Skillman, A. G., Nicholls, A., Guthrie, J. P. & Taylor, P. J. The SAMPL2 blind prediction challenge: introduction and overview. *J. Comput. Aided Mol. Des.* **24**, 259–279 (2010).
- Kotev, M., Sarraf, L. & Gonzalez, C. D. *Quantum Mechanics in Drug Discovery* (ed. Heifetz, A. p. 231–255 (Springer US, 2020).
- Manathunga, M., Götz, A. W. & Merz, Jr, K. M. Computer-aided drug design, quantum-mechanical methods for biological problems. *Curr. Opin. Struct. Biol.* **75**, 102417 (2022).
- Zhou, T., Huang, D. & Caflich, A. Quantum mechanical methods for drug design. *Curr. Top. Med. Chem.* **10**, 33–45 (2010).
- Arodola, O. A. & Soliman, M. E. Quantum mechanics implementation in drug-design workflows: does it really help? *Drug Des. Dev. Ther.* **11**, 2551–2564 (2017).
- Simons, J. Why is quantum chemistry so complicated? *J. Am. Chem. Soc.* **145**, 4343–4354 (2023).
- Dykstra, C., Frenking, G. & Kim, K. *Theory and Applications of Computational Chemistry: the First Forty Years* (Elsevier Science, 2011).
- Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- Blunt, N. S. et al. A perspective on the current state-of-the-art of quantum computing for drug discovery applications. *arXiv* <https://doi.org/10.48550/arXiv.2206.00551> (2022).
- Jordan, P. & Wigner, E. ber das Paulische Äquivalenzverbot. *Z. Physik* **47**, 631–651 (1928).
- Somma, R., Ortiz, G., Gubernatis, J. E., Knill, E. & Laflamme, R. Simulating physical phenomena by quantum networks. *Phys. Rev. A* **65**, 042323 (2002).
- Kitaev, A. Y. Quantum measurements and the Abelian Stabilizer Problem. *arXiv* <https://doi.org/10.48550/arXiv.quant-ph/9511026> (1995).
- Du, J. et al. NMR implementation of a molecular hydrogen quantum simulation with adiabatic state preparation. *Phys. Rev. Lett.* **104**, 030502 (2010).
- Lanyon, B. P. et al. Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2**, 106–111 (2010).
- Li, Z. et al. Solving quantum ground-state problems with nuclear magnetic resonance. *Sci. Rep.* **1**, 88 (2011).
- O'Malley, P. J. et al. Scalable quantum simulation of molecular energies. *Phys. Rev. X* **6**, 031007 (2016).
- Paesani, S. et al. Experimental Bayesian quantum phase estimation on a silicon photonic chip. *Phys. Rev. Lett.* **118**, 100503 (2017).
- Santagati, R. et al. Witnessing eigenstates for quantum simulation of Hamiltonian spectra. *Sci. Adv.* **4**, eaap9646 (2018).
- Wang, Y. et al. Quantum simulation of helium hydride cation in a solid-state spin register. *ACS Nano* **9**, 7769–7774 (2015).
- Abrams, D. S. & Lloyd, S. Quantum algorithm providing exponential speed increase for finding eigenvalues and eigenvectors. *Phys. Rev. Lett.* **83**, 5162 (1999).
- Aspuru-Guzik, A., Dutoi, A. D., Love, P. J. & Head-Gordon, M. Simulated quantum computation of molecular energies. *Science* **309**, 1704–1707 (2005).
- Peruzzo, A. et al. A variational eigenvalue solver on a photonic quantum processor. *Nat. Commun.* **5**, 4213 (2014).
- McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New J. Phys.* **18**, 023023 (2016).
- Kandala, A. et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549**, 242–246 (2017).
- Kandala, A. et al. Error mitigation extends the computational reach of a noisy quantum processor. *Nature* **567**, 491–495 (2019).

47. Jiang, Z., Sung, K. J., Kechedzhi, K., Smelyanskiy, V. N. & Boixo, S. Quantum algorithms to simulate many-body physics of correlated fermions. *Phys. Rev. Appl.* **9**, 044036 (2018).
48. Kivlichan, I. D. et al. Quantum simulation of electronic structure with linear depth and connectivity. *Phys. Rev. Lett.* **120**, 110501 (2018).
49. Wecker, D. et al. Solving strongly correlated electron models on a quantum computer. *Phys. Rev. A* **92**, 062318 (2015).
50. Babbush, R., McClean, J., Wecker, D., Aspuru-Guzik, A. & Wiebe, N. Chemical basis of Trotter-Suzuki errors in quantum chemistry simulation. *Phys. Rev. A* **91**, 022311 (2015).
51. Sugisaki, K. et al. Quantum chemistry on quantum computers: a polynomial-time quantum algorithm for constructing the wave functions of open-shell molecules. *J. Phys. Chem. A* **120**, 6459–6466 (2016).
52. Sugisaki, K. et al. Quantum chemistry on quantum computers: a method for preparation of multiconfigurational wave functions on quantum computers without performing post-hartree-fock calculations. *ACS Cent. Sci.* **5**, 167–175 (2019).
53. Du, Y., Hsieh, M.-H., Liu, T., You, S. & Tao, D. Erratum: learnability of quantum neural networks. *PRX Quant.* **3**, 030901 (2022).
54. Google AI Quantum and Collaborators*†. et al. Hartree-Fock on a superconducting qubit quantum computer. *Science* **369**, 1084–1089 (2020).
55. Nam, Y. et al. Ground-state energy estimation of the water molecule on a trapped-ion quantum computer. *npj Quant. Inf.* **6**, 33 (2020).
56. Progress toward larger molecular simulation on a quantum computer: Simulating a system with up to 28 qubits accelerated by point-group symmetry. *Phys. Rev. A* **105**, 062452 (2022).
57. Toward practical quantum embedding simulation of realistic chemical systems on near-term quantum computers. *Chem. Sci.* **13**, 8953 (2022).
58. Tazhigulov, R. N. et al. Simulating models of challenging correlated molecules and materials on the sycamore quantum processor. *PRX Quant.* **3**, 040318 (2022).
59. Bravyi, S., Gambetta, J. M., Mezzacapo, A. & Temme, K. Tapering off qubits to simulate fermionic Hamiltonians. *arXiv* <https://doi.org/10.48550/arXiv.1701.08213> (2017).
60. Moll, N., Fuhrer, A., Staar, P. & Tavernelli, I. Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer. *J. Phys. A: Math. Theor.* **49**, 295301 (2016).
61. Babbush, R. et al. Exponentially more precise quantum simulation of fermions in the configuration interaction representation. *Quant. Sci. Technol.* **3**, 015006 (2018).
62. Steudtner, M. & Wehner, S. Fermion-to-qubit mappings with varying resource requirements for quantum simulation. *New J. Phys.* **20**, 063010 (2018).
63. Kirby, W., Fuller, B., Hadfield, C. & Mezzacapo, A. Second-quantized fermionic hamiltonians for quantum simulation with polylogarithmic qubit and gate complexity. *arXiv* <https://doi.org/10.48550/arXiv.2109.14465> (2021).
64. Shee, Y., Tsai, P.-K., Hong, C.-L., Cheng, H.-C. & Goan, H.-S. Qubit-efficient encoding scheme for quantum simulations of electronic structure. *Phys. Rev. Res.* **4**, 023154 (2022).
65. Silberg, M. S. *Chemistry: The Molecular Nature Of Matter And Change/Martin S. Silberg*, chap. 9 (McGraw-Hill, Boston, 2009), 5th ed. edn.
66. Levine, D. S. et al. Pattern-free generation and quantum mechanical scoring of ring-chain tautomers. *J. Comput. Aided Mol. Des.* **35**, 417–431 (2021).
67. Frisch, M. J. et al. *Gaussian 16, Revision C.01* (Gaussian Inc. Wallingford CT, 2016).
68. Watanabe, K., Tanaka, M., Yuki, S., Hirai, M. & Yamamoto, Y. How is edaravone effective against acute ischemic stroke and amyotrophic lateral sclerosis? *J. Clin. Biochem. Nutr.* **62**, 20–38 (2018).
69. Aquilante, F. et al. Molcas 8: new capabilities for multiconfigurational quantum chemical calculations across the periodic table. *J. Comput. Chem.* **37**, 506–541 (2016).
70. Vogiatzis, K. D., Ma, D., Olsen, J., Gagliardi, L. & de Jong, W. A. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *J. Chem. Phys.* **147**, 184111 (2017).
71. Hanwell, M. D. et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17 (2012).
72. Halgren, T. A. Merck molecular force field. v. extension of mmff94 using experimental data, additional computational data, and empirical rules. *J. Comput. Chem.* **17**, 616–641 (1996).
73. Halgren, T. A. Merck molecular force field. II. MMFF94 van der waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **17**, 520–552 (1996).
74. Halgren, T. A. Merck molecular force field. iii. molecular geometries and vibrational frequencies for mmff94. *J. Comput. Chem.* **17**, 553–586 (1996).
75. Halgren, T. A. Merck molecular force field. i. basis, form, scope, parameterization, and performance of mmff94. *J. Comput. Chem.* **17**, 490–519 (1996).
76. Halgren, T. A. & Nachbar, R. B. Merck molecular force field. iv. conformational energies and geometries for mmff94. *J. Comput. Chem.* **17**, 587–615 (1996).
77. Becke, A. D. Becke's three parameter hybrid method using the lyp correlation functional. *J. Chem. Phys.* **98**, 5648–5652 (1993).
78. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The J. Phys. Chem.* **98**, 11623–11627 (1994).
79. Head-Gordon, M., Pople, J. A. & Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **153**, 503–506 (1988).
80. Du, Y., Hsieh, M.-H., Liu, T. & Tao, D. Expressive power of parametrized quantum circuits. *Phys. Rev. Res.* **2**, 033125 (2020).
81. Sim, S., Johnson, P. D. & Aspuru-Guzik, A. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quant. Technol.* **2**, 1900070 (2019).
82. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R. & Neven, H. Barren plateaus in quantum neural network training landscapes. *Nat. Commun.* **9**, 4812 (2018).
83. Zhang, K., Hsieh, M.-H., Liu, L. & Tao, D. Toward trainability of quantum neural networks. *arXiv* <https://doi.org/10.48550/arXiv.2011.06258> (2020).
84. Zhang, K., Hsieh, M.-H., Liu, L. & Tao, D. Toward trainability of deep quantum neural networks. *arXiv* <https://doi.org/10.48550/arXiv.2112.15002> (2021).
85. Zhang, K., Liu, L., Hsieh, M.-H. & Tao, D. *Advances in Neural Information Processing Systems*. (eds. Oh, A. H., Agarwal, A., Belgrave, D. & Cho, K.) (2022).
86. Amovilli, C. et al. Recent advances in the description of solvent effects with the polarizable continuum model. In Löwdin, P.-O. (ed.) *Advances in Quantum Chemistry*, vol. 32, 227–261 (Academic Press, 1998).
87. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
88. Klamt, A. & Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2* 799–805 (1993).
89. Shao, Y. et al. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **8**, 3172–3191 (2006).
90. Watson, M. A., Yu, H. S. & Bochevarov, A. D. Generation of tautomers using micro-pka's. *J. Chem. Inf. Model.* **59**, 2672–2689 (2019).
91. Grant, E., Wossnig, L., Ostaszewski, M. & Benedetti, M. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum* **3**, 214 (2019).
92. McArdle, S. et al. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quant. Inf.* **5**, 75 (2019).
93. Rathi, P. C., Ludlow, R. F. & Verdonk, M. L. Practical High-Quality electrostatic potential surfaces for drug discovery using a Graph-Convolutional deep neural network. *J. Med. Chem.* **63**, 8778–8790 (2020).

ACKNOWLEDGEMENTS

The authors thank Prof. Yuan-Chung Cheng from National Taiwan University for many useful suggestions.

AUTHOR CONTRIBUTIONS

Y.S. and T.L.Y. contributed equally to this work. Y.C. Lin and M.H. Hsieh are the team leaders from Insilico Medicine Taiwan and Hon Hai (Foxconn) Research Institute, respectively. They supervise the research team during the project. Y.S. and T.L.Y. designed the workflow with the help of all the other authors. T.L.Y., A.Y. and Y.S. conducted quantum chemistry simulations. Y.S. and J.Y.H. perform noiseless and noisy quantum simulations. The first draft of the manuscript was written by Y.S. and T.L.Y. All authors contributed to the interpretation of the results and the final draft.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41534-023-00767-9>.

Correspondence and requests for materials should be addressed to Yen-Chu Lin or Min-Hsiu Hsieh.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023