Practical protein-pocket hydration-site prediction for drug discovery on a quantum computer

Daniele Loco [*]
Qubit Pharmaceuticals, Advanced Research Team, 75014 Paris, France

Kisa Barkemeyer [*] and Andre R. R. Carvalho
Q-CTRL, Berlin, Germany
and Q-CTRL, Sydney, New South Wales, Australia

Jean-Philip Piquemal [†]
Advanced Research Department, Qubit Pharmaceuticals, 24 rue du Faubourg Saint-Jacques, 75014 Paris, France
and Sorbonne Université, LCT, UMR 7616 CNRS, 75005 Paris, France

(Received 31 January 2026; accepted 29 July 2026; published 21 September 2026; corrected 16 October 2026)

Demonstrating the practical utility of noisy intermediate-scale quantum hardware for recurrent tasks in computer-aided drug discovery is of paramount importance. We tackle this challenge by performing three-dimensional hydration-site prediction in protein pockets on a quantum computer. Formulating the water placement problem as a quadratic unconstrained binary optimization, we use a hybrid approach coupling a classical three-dimensional reference-interaction site model to an efficient quantum optimization solver to run various hardware experiments up to 123 qubits. Matching the precision of classical approaches, our results reproduced experimental predictions on real-life protein-ligand complexes. Furthermore, through a detailed resource estimation analysis, we show that accuracy can be systematically improved with increasing number of qubits, indicating that full quantum utility is within reach as devices scale in the near term. The method has potential for assisting simulations of protein-ligand complexes for drug lead optimization and setup of docking calculations.

DOI: 10.1103/gyqr-mv1h

I. INTRODUCTION

In the past 20 years, computer-aided drug discovery (CADD) has become an important tool for the pharmaceutical industry to enhance drug discovery pipelines [1–3]. Accurate numerical methods capable of modeling chemical processes, physical interactions between complex molecular structures, and dynamical effects occurring in biological systems are of great value, improving experimental data interpretability, while potentially reducing the overall costs thanks to improved theoretical predictions.

Between the many challenges faced in the long and expensive process of developing a drug, there is the need to understand the role of water. Water, as a primary component of biological environments, profoundly influences the overall structural stability of proteins, as well as the interactions of proteins with small molecules, often referred to as ligands, making the understanding of its behavior fundamental for estimating drug binding and optimizing drug candidates [4–6]. Water molecules can act as an intermediate between a protein and a ligand. These can be called interfacial waters, and they play a fundamental role in the binding process [4,7]. However, predicting the configuration of water molecules presents a formidable computational challenge due to the high dimensionality of the problem and the necessity to sample a vast conformational space [5,6,8–10].

Classically, molecular dynamics and Monte Carlo strategies are widely used to address the problem of configurational sampling in an effective way [8,11–15]. These types of algorithms are general enough to handle the configurational sampling of molecules larger than water, such as biomolecules and large aggregates. However, sampling water will always remain difficult due to the complexity of the hydrogen-bond networks that water can form, while either exchanging rapidly with bulk solvent or, conversely, becoming structurally stabilized in binding sites [16–19]. Modern enhanced sampling techniques enable to precisely capture the water-protein and water-ligand interactions in complex biological systems [20–22] to provide predictive drug binding free-energy estimation [5,23,24]. This, however, comes with a significant computational cost and state-of-the-art simulation codes heavily rely on massive parallelism and GPU acceleration [25–29]. In parallel, promising machine learning approaches [30,31] have been steadily improving and becoming able to sample molecular equilibrium structures. However, to date, these approaches do not yet reach sufficient accuracy for CADD applications [3,32,33].

[*]These authors contributed equally to this work.
[†]Contact author: jean-philip.piquemal@sorbonne-universite.fr

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Alternatively, quantum computing holds the promise of addressing these classically hard computational tasks. Indeed, quantum speedups are expected to be substantial for implementations of statistical physics problems related to configuration sampling [34,35]. A number of studies have presented applications of different types of quantum algorithms to biomolecular systems [33,36–38] and, more generally, full drug discovery pipelines that incorporate quantum-based computations have been proposed and reviewed in the literature [39,40].

In this context, some of us recently presented an approach for the quantum prediction of protein pocket hydration [41]. Using a quadratic unconstrained binary optimization (QUBO) cost function embodying the water placement problem, we mapped the problem of identifying hydration sites onto an Ising Hamiltonian. This mapping enabled the implementation of an adiabatic protocol on an analog quantum device based on Rydberg cold atom physics [41]. The method used two-dimensional (2D) slices of 3D-RISM (three-dimensional reference interaction site model) continuous water density functions to define a discrete optimization problem, which was then mapped onto the Hamiltonian of the corresponding 2D Rydberg atom layer [41]. Although this work relied on 2D slices only, which poses challenges for applicability to 3D systems, it represented a key step toward the use of quantum computing in CADD.

In this paper, taking further steps toward developing practical quantum approaches for CADD, we present an extension of the previous work [41], leveraging digital quantum computing to predict the 3D hydration sites in protein pockets in real-life conditions using quantum hardware. Our methodology integrates a classical formulation of the hydration-site prediction problem with quantum optimization techniques, simultaneously leveraging advanced software [42,43] and hardware to address the problem on a quantum computer. The presented end-to-end workflow includes the problem encoding, the quantum algorithm implementation, and hardware execution using an efficient quantum solver [43], enabling us to explore the practical utility for large-scale instances of the problem on current noisy intermediate-scale quantum (NISQ) devices. This strategy allowed us to describe realistic 3D systems without additional approximations [41]. We applied our approach to a selected list of proteins relevant to the pharmaceutical industry, encompassing FDA (U.S. food and drug administration) approved drugs in complex with their target protein. Our results were obtained from experiments performed on the latest Heron 156-qubit IBM devices and compared to existing classical solvers. Finally, we discuss the pathway to quantum advantage by presenting a resource estimation analysis to forecast the device specifications needed to address protein complexes at scales where classical methods struggle to find solutions.

II. RESULTS

Our hydration-site prediction workflow consists of multiple steps that are described in detail in Sec. IV. It starts from the calculation of 3D-RISM densities [$g(\mathbf{r})$] from the 3D molecular structure of the target protein for the hydration-site prediction. This 3D-RISM density contains the essential information about the water distribution around the protein. We then describe this density as a sum of Gaussian distributions whose centers are associated with the location of the water molecules and with variances corresponding to the uncertainty in the water position. Therefore, the problem of locating the positions of water molecules is mapped onto the problem of finding the best Gaussian mixture that approximates the 3D-RISM density. In Sec. IV, we give more details about the 3D-RISM and how to interpret high-density regions, which are the most relevant for our task. A key element put forward in a previous work [41] and further explored here is framing this as a QUBO problem amenable to a quantum optimization. Once this optimization is performed, the outputs are the 3D spatial coordinates of the most likely sites where water molecules can be found.

TABLE I. Selection of protein structures considered in this study. The columns represent the protein PDB ID, the number of CW molecules present in the crystallographic structure inside the binding pocket, the number of protein residues (amino acids) that identify the binding pockets, the protein MW, and the resolution of the x-ray diffraction experiment that determined the molecular structure. PDB IDs 3b7e and 3beq are, respectively, the IDs used to register a holo and an apo structure of the same protein.

PDB ID	No. of CW	No. of residues	MW (kDa)	Resolution (Å)
1f9g	6	25	83.7	2.0
1x70	16	18	174.9	2.1
2f9w	15	10	59.9	1.9
3b7e	28	15	86.8	1.45
4h2f	13	27	60.8	1.85
3beq	10	15	87.0	1.64

This section is divided into five subsections reflecting the workflow described above. We first present the set of protein complexes selected for this study, including the relevant discretization and Gaussian parameters used to define our test cases. We then report on the results of the crucial optimization step executed on quantum computers, up to the current hardware limits. To understand the performance of the method beyond this limit, we also present results for large-scale systems using classical approximate optimization approaches. Finally, we connect these optimization results to the other components of the workflow to obtain our final hydration-site prediction. These results are on par with, or sometimes better than, other hydration prediction methods available in the literature. At the end of the section, we forecast the resources needed on a quantum device to achieve such problem scales.

A. Test systems: realistic drug-protein complexes

We start by choosing the protein 3D structures required in the first step of our workflow, shown in Fig. 9. We tested our approach on a set of proteins of interest for health care applications, all of them being related to human diseases for which a drug has already been developed and later approved by the FDA agency [44]. Table I shows the crystal structures used in this work, each identified by their Protein Data Bank [45] (PDB) ID. We focus our attention only on the binding

site of each protein and show in Table I the number of crystal water (CW) molecules in the crystallographic structure inside the binding pocket, the number of protein residues identifying the pockets, the protein molecular weight (MW), and the resolution of the x-ray diffraction experiments used to determine the molecular structure.

The first five structures were selected out of the 108 analyzed by Samways *et al.* in Ref. [44], and they all include protein structures with a ligand bound to it, which are commonly denoted as protein-ligand complexes or holo structures. We kept our selection small to remain compatible with quantum computing resource constraints, while ensuring sufficient diversity across protein sizes and families. For these structures, we can decide whether to include or remove the explicit ligand molecular structure for the hydration-site prediction. These will correspond to two different predictions, as we will show later. Additionally, we included in this list the example of the apo structure of PDB ID 3b7e (PDB ID = 3beq). The term apo refers to the protein 3D conformation where no ligand is bound to it. Knowing the differences in the protein hydration between holo and apo structures is an important information for drug discovery modelers. In practice, knowing the hydration in the apo state can help the design of new efficient binders, mimicking the protein-water interactions observed in the corresponding holo structure. This additional example complements our test set to further show the potential of our approach for applications to real use cases in drug discovery.

Dimension reduction

Mapping the full 3D-RISM densities of the large selected systems would result in QUBO instances of ∼1 000 000 binary variables, which would make the corresponding optimization problem unattainable for both classical methods and current quantum computers. Therefore, to simplify the problem, one needs to define the problem on a grid coarser than that of the initial high-resolution 3D-RISM density. For this, we resample $g(\mathbf{r})$ using a grid spacing δ larger than the original 3D-RISM one ($\delta = 0.5$ Å) and remove low-density values according to an arbitrary threshold value τ_g (see Sec. IV). Such a strategy impacts the hydration-site prediction accuracy, as discussed later in the text, but enables us to address the problem within a reasonable time using modest classical computing resources, while at the same time making the problem suitable for execution on quantum hardware by choosing parameters compatible with current limitations on qubit number and two-qubit gates. The QUBO instances resulting from this resampling procedure are shown in Table II, together with the relevant parameters used to preprocess the computed 3D-RISM density and to map them onto each QUBO problem instance. The structures considered are a holo structure with ligand (L), a holo structure without ligand (NL), or the apo structure (APO) from Table I.

B. Quantum hardware demonstration of 3D hydration-site prediction

To assess the performance of the quantum optimization step for our 3D hydration-site prediction workflow, we solved the underlying QUBO problem for the instances described in Table II on IBM devices using Q-CTRL's quantum optimization solver (see Sec. IV for details). The instances reported in Table II represent reduced-size problems selected to fit the QPU resources currently available. They do not represent yet a full scale application, which would require more advanced QPUs. The scope of these simulations is to demonstrate the feasibility of our workflow on quantum hardware. In Secs. II D and II F, we discuss in more detail problems relevant for drug discovery applications and the resources needed to address them with our quantum-based approach. Exemplary results for the largest instance considered within the test set, for instance, "d" in Table II, which requires 116 variables corresponding to 116 qubits, are shown in Fig. 1.

TABLE II. Test cases listed by PDB ID and labeled according to the inclusion (L) or exclusion (NL) of the ligand. We show the number of binary variables of the associated QUBO problem, the grid parameters (δ and τ_g), and the variance of the Gaussian distributions used for each problem instance. We label each of them with the characters from the last column for easy of reference throughout the text. Additional details about the parameter choice can be found in the Supplemental Material [46].

Instances	No. of variables	δ	τ_g	σ^2	Label
1f9g (L)	98	1.05	0.17	0.8	a
1f9g (NL)	91	1.05	0.17	0.8	b
1×70 (L)	69	0.95	0.10	0.8	c
1×70 (NL)	116	0.95	0.10	0.8	d
2f9w (L)	41	1.05	0.10	0.8	e
2f9w (NL)	92	1.05	0.10	0.8	f
3b7e (L)	71	1.15	0.10	1.0	g
3b7e (NL)	90	1.15	0.10	1.0	h
4h2f (L)	82	0.95	0.12	0.8	i
4h2f (NL)	79	0.95	0.12	0.8	l
3beq (APO)	72	1.25	0.10	1.0	m

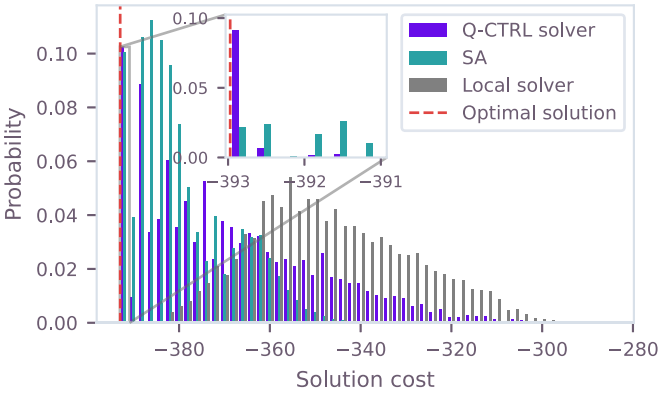


FIG. 1. Cost distributions for the hydration-site prediction QUBO problem instance d as specified in Table II, requiring 116 qubits. The Q-CTRL solver output obtained using IBM Kingston (purple bars) is compared to the optimal solution calculated classically with CPLEX (dashed red line). For reference, the results of SA (teal bars) as well as the results of a greedy local solver (gray bars) are displayed. Inset: Closed-up view of the near-optimal region using finer bins.

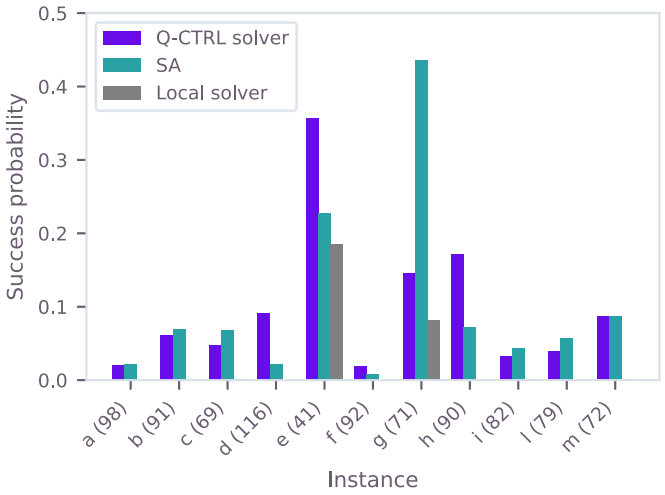


FIG. 2. Probability to sample the optimal solution using the Q-CTRL solver on IBM Kingston (purple bars), SA (teal bars), and a greedy local solver (gray bars) for the test-set instances labeled according to Table II, with variable numbers indicated in parentheses.

It displays the cost distribution of sampled bitstrings together with the optimal solution calculated classically using CPLEX. For the considered problem instance size, finding provably optimal solutions is still possible, while it becomes increasingly difficult for larger-scale instances, as we show below. The Q-CTRL solver finds the optimal solution with a sampling probability of around 9%, a significantly higher success probability than that of simulated annealing (SA), used as a classical reference heuristic. SA results were obtained with the Neal library [47] using a geometric temperature schedule and 10 000 samples to evaluate the final cost distribution. A detailed description of the annealing schedule and the option used for all our simulation is reported in Table S1 of the Supplemental Material [46]. Although providing an overall distribution favoring lower-cost solutions, SA identifies the optimal solution with a probability of only about 2%. In contrast, a local solver that applies greedy optimization to randomly sampled candidate solutions is significantly outperformed and fails to identify the optimal solution. Details on the implementation of the local solver that can be interpreted as a zero-temperature version of SA can be found in Ref. [43].

The same process of analyzing the QPU output and comparing it to the exact CPLEX solution was repeated for the whole test set to extract the probability of obtaining optimal solution. These probabilities are shown in Fig. 2 for all instances, demonstrating that the quantum solver executed on real hardware is able to successfully solve the hydration-site prediction QUBO problem for a variety of systems with a performance that is overall comparable to SA. Notably, the quantum solver identifies the optimal solution in all cases. In contrast, the greedy local solver only identifies the optimal solution with a non-negligible probability for a few smaller instances and deviates from optimality as instance size grows (see Sec. S5 in the Supplemental Material for additional details on the cost gaps [46]).

Since for the instances considered in this work (Table II) exact classical solvers find provably exact solutions quickly, in practice, we would not deploy quantum resources for these cases. Nevertheless, as the problem size grows, exact methods typically scale poorly, and beyond a certain point, high-quality heuristics, classical or quantum, become preferable. We, thus, aim to identify a point where the performance of exact classical approaches for solving these problems deteriorates, leaving room for heuristic approaches, such as the Q-CTRL quantum optimization solver, to yield better performance.

To determine the hardness of the hydration-site prediction QUBO problems, we benchmark how the computational complexity increases with problem scale for the exemplary case of the 3b7e-with-ligand structure. Using CPLEX on a standard laptop, we consider problem instances at a range of scales, obtained by varying the grid spacing. We observe that for 123 variables, corresponding to a grid spacing of 1.9 Å, it is no longer possible to find the optimal solution within a cutoff time of 3 h. As shown in Fig. 3(a), the incumbent solution only improves in the first few steps before plateauing. The optimality gap indicates how far the incumbent solution may be from the true optimum and thus acts as a proxy for solution quality. As shown in the inset of Fig. 3(a), the optimality gap stalls around 40%, indicating that, in principle, there is still significant room for improvement of the solution. In fact, the incumbent CPLEX solution after the cutoff time ($C_{\text{best, CPLEX}} = -319.90$) is not the optimal solution: Running the same instance with the Q-CTRL solver yields a lower cost ($C_{\text{best, Q-CTRL}} = -321.32$) after a total run-time of only 25 min on IBM Pittsburgh.

Note that the results presented in Fig. 3(a) indicate that the instance is challenging to solve for an exact solver, and one needs to resort to approximate methods. As we see in the output cost distributions in Fig. 3(b), both the heuristic SA and the Q-CTRL solver yield the same best solution cost, with the former achieving higher probability. This can be attributed to the fact that the quantum resources required to run this instance reach current device limits in terms of gate fidelities, circuit depth, and T1 times. To reliably tackle even larger problems, quantum algorithm development and further quantum hardware improvements toward larger and better devices must go hand in hand. The effect of the progress on the hardware side was evident during our study: While we were producing our results, IBM's newest Heron r3 processor, the ibm_pittsburgh backend, came online. We ran a subset of test-set instances on both Heron r2 and r3 devices and observed significant performance differences, as shown in Fig. 4. This highlights the potential for substantial improvements, including the possibility of outperforming SA, as next-generation devices arrive in line with IBM's quantum development road map [48]. Expected future developments are discussed in more detail in Sec. II F.

C. Performance evaluation of 3D hydration-site prediction

Now that we have shown that, for the systems considered so far, the optimization step of our workflow can be successfully solved using Q-CTRL's solver on a quantum computer, we can look at how these results impact the ultimate goal of predicting the hydration sites.

The first step is to consider how to interpret the output bitstrings corresponding to the optimal solutions as spatial water

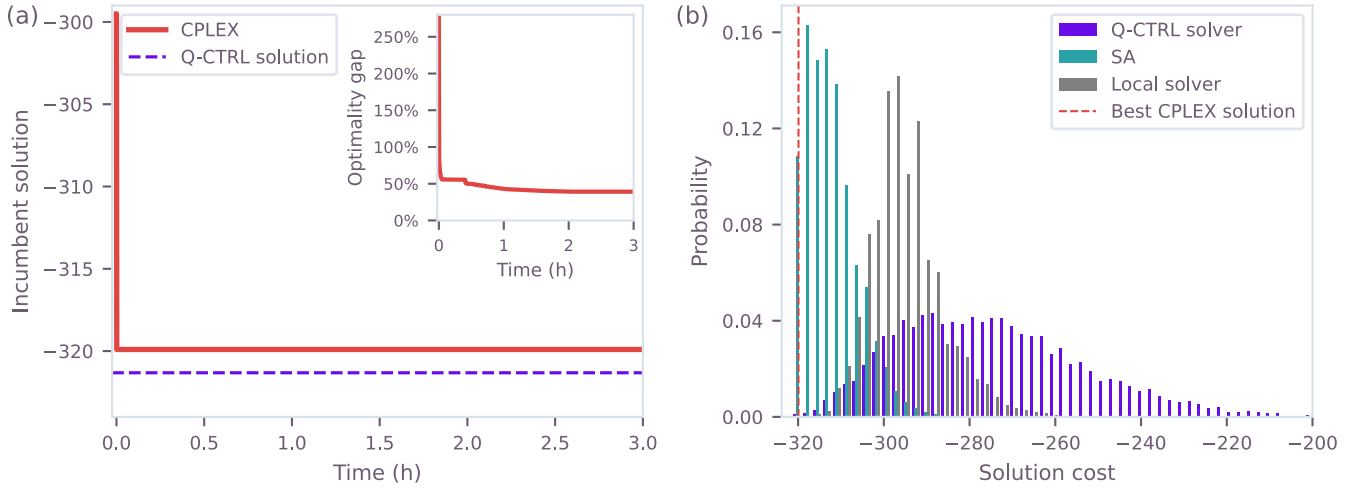


FIG. 3. Results for the 3b7e-with-ligand instance with a grid spacing of 1.9 Å corresponding to 123 variables. (a) Exact classical (CPLEX) results. Main: Incumbent solution (red line) over time compared to the best solution identified using the Q-CTRL solver on IBM Pittsburgh (dashed purple line). Inset: Optimality gap over time. (b) Cost distributions for the Q-CTRL solver on IBM Pittsburgh (purple bars), simulated annealing (teal bars), and a greedy local solver (gray bars) are shown together with the best CPLEX solution (dashed red line).

configurations. In our QUBO formulation (see Sec. IV), each qubit has been mapped onto a spatial location, and the values 0 and 1 correspond, respectively, to the absence or presence of a water molecule at the point in space associated with that register. Thus, we read the corresponding 3D coordinates associated with qubits in state 1 and place a water at that point in space. In practice, we only place the oxygen atom of the water molecule, which is sufficient to identify its location. Repeating this operation for each element of a bitstring, we obtain the predicted water (PW) molecules configuration.

The next step is to assess the accuracy of these predictions. It is common practice in the field to use the crystallographic water (CW) positions to assess the quality of a given numerical method aiming at identifying hydration sites [6,8–10,13,49]. Here, we will use the CWs available in the experimental data as a reference. One metric we use to assess the quality of our predictions is the percentage of CWs that we are able to identify. We use the character C to refer to this metric, which is computed by searching for the presence of at least one PW in a radius of 3 Å around each CW. If more than one PW is found, we also record how many additional PWs are associated with each CW. We consider such a set of PWs as belonging to the same cluster. Performing this analysis for each CW, we use as an additional evaluation metric the average number of PWs found per CW and call it the average cluster size ($\langle CS \rangle$). We are also interested in determining how spatially close the PWs are from the CWs. We denote this third metric as the precision (P), which we measure both for the closest PW associated with each CW, namely, P*, and for the ensemble of the cluster associated with each CW ($\langle P \rangle$). The definitions of these metrics are reported in Sec. IV.

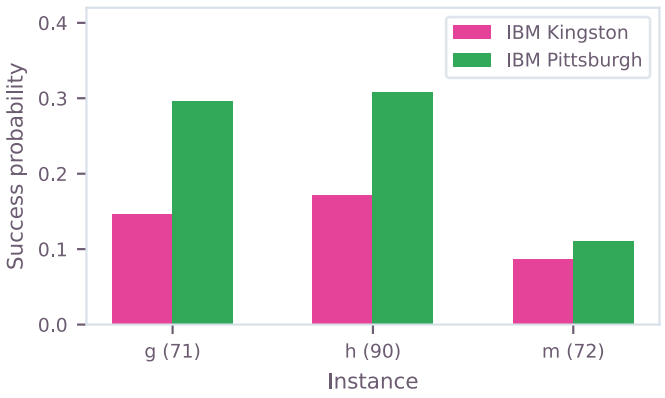


FIG. 4. Probability to sample the optimal solution using the Q-CTRL solver on the Heron r2 backend ibm_kingston (red bars) and the Heron r3 backend ibm_pittsburgh (blue bars) for a selection of the test-set instances labeled according to Table II, with variable numbers indicated in parentheses.

For this performance evaluation, different aspects need to be considered. In the first place, it is important to note that the crystal structures used in this work have been determined with a reasonably good experimental resolution (see Table I). This ensures that the comparison between PW and CW is meaningful. Furthermore, the accuracy of our QUBO-based approach also depends on the underlying 3D-RISM densities and on how well they actually represent the water molecule distribution in the protein. Available benchmarks from the literature assess this factor, proving a generally good cost-balanced 3D-RISM's quality for hydration free-energy prediction [50], in line with the quality requirements of fast virtual screening tasks and systems preparation in CADD. On top of these factors, the discretization step and dimensionality reduction procedure we perform to map the 3D-RISM density into QUBO (see Fig. 9) also impact the method's accuracy. In particular, we can expect to obtain the best mapping between the continuous density and the QUBO when δ is equal to or smaller than the spacing of the $g(\mathbf{r})$ grid (0.5 Å) and if the τ_g threshold remains negligibly small.

Based on these considerations, we explored different mappings between QUBO problem instances and the underlying $g(\mathbf{r})$ density. We varied the parameters used to generate system "g" in Table II to give a fair estimation of the impact of these approximations on the method's performance. To do so,

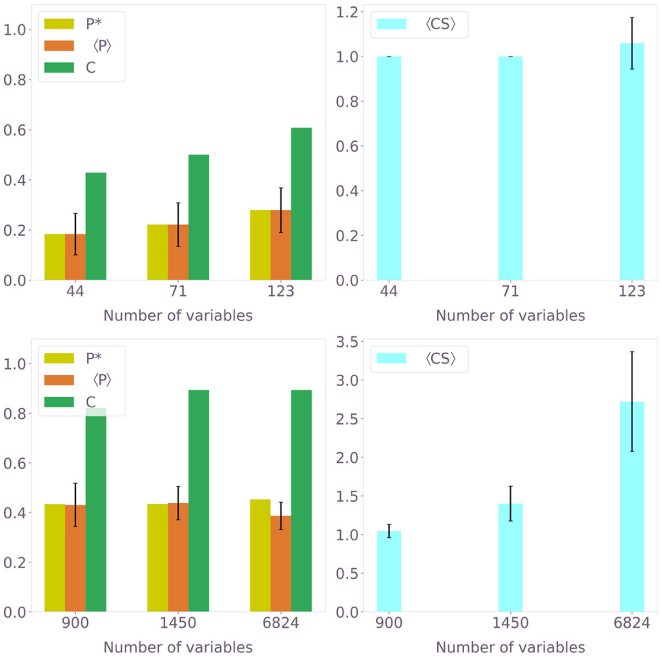


FIG. 5. Analysis of the hydration-site prediction performance across different QUBO instance sizes for PDB ID = 3b7e from Table I, including the ligand in the 3D-RISM calculation. Top panel shows the performance analysis for the quantum optimization run using the Q-CTRL solver on the IBM devices, on the QUBO instances obtained from the following parameters: $\sigma^2 = 1.0$ Å^2 and $\tau_g = 0.1$ are used uniformly, and, from smaller to larger instances, $\delta = 1.35, 1.15, 0.95$ Å, to discretize the 3D-RISM density for the mapping onto the QUBO problem. Bottom panel shows the same analysis performed on the classical SA results on QUBO instances obtained from the following parameters: $\sigma^2 = 1.0$ Å^2 and, from smaller to larger instances, $\tau_g = 0.1, 0.05, 0.002$ and $\delta = 0.5, 0.5, 0.35$ Å. For each panel, we report as follows: on the left-side plot, P* (closest water placement precision), ⟨P⟩ (cluster-averaged precision), and C (fraction of crystal waters identified), and on the right side of the plot ⟨CS⟩ (average cluster size); error bars show the 95% confidence interval for ⟨P⟩ and ⟨CS⟩. Each metric is computed as described in Sec. IV, extracting the PWs corresponding to the best solution obtained from the corresponding optimization.

we reduced the δ and τ_g parameters used in our original set (see Table II) to increase the size and the overall complexity of the optimization problem. We ensure that the sizes of these problems are still tractable on quantum hardware. The results are displayed in the top panel of Fig. 5, where the performance metrics are computed by extracting the best solutions obtained for each new QUBO instance using Q-CTRL's optimization solver. As with the cases presented in Table II, the quantum optimizer was also able to find the optimal solution.

Both precision metrics (P*, ⟨P⟩) and the fraction of successfully identified CWs (C) consistently increase with the size and complexity of the QUBO. The larger QUBO instance (123 variables) is able to recover up to ∼60% of CWs, showing a notable increase with respect to the 71-variable instance. The metric ⟨CS⟩ exceeds the value of 1.0, which could be related to diffused regions of high enough density that can be mapped by more than one variable. This is not possible, or less likely to happen, for smaller QUBO instances.

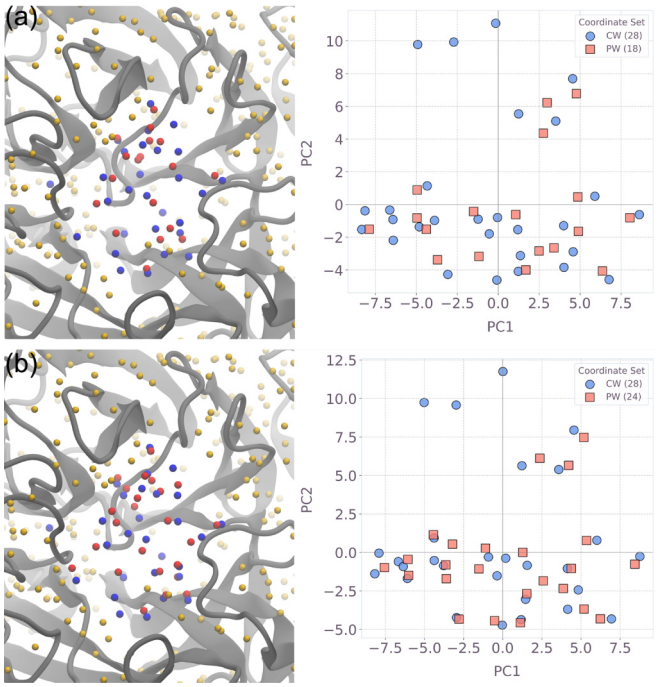


FIG. 6. Visualization of hydration-site location corresponding to the best solutions obtained from the 123-variable instance [panel (a)] and 900-variable instance [panel (b)] analyzed in Fig. 5. On the left side of each panel, we show a 3D visualization of the molecular system, using VMD software; red spheres represent the oxygen atoms placed as a solution of the QUBO, blue the reference CWs, while in yellow are represented all the crystal waters available in the PDB original file. On the right side, we have a 2D space representation of the hydration-site location obtained through a principal component analysis (PCA) performed for an easier visualization of the corresponding 3D space placement. Light-red squares represent the PW location on the principal components (PC), while the light-blue circles represent those of the CWs, with their respective numbers reported in the legend.

To give a better idea of the results obtained from these predictions, we report in Fig. 6(a), the 3D geometry of the identified hydration site from the best solution obtained on the 123-variable QUBO displayed on Fig. 5. On the left side of the figure, the 3D structure of the binding pocket is represented in gray, together with the oxygens of CWs present in the original PDB (yellow spheres). We highlight the set of water we targeted for the hydration-site prediction task in blue, while the red spheres represent the PW locations.

The 3D visualization being hard to interpret, we performed a PCA as a dimensionality reduction technique to better apprehend the variance (difference and spread) between the two sets of 3D coordinates, the CW coordinates and the PW coordinates, in a simple 2D plot. The PCA model is fitted to the combined 3D data (i.e., the CW and PW coordinate sets combined) to find the two directions (principal components PC1 and PC2) in the 3D space that capture the maximum amount of variance between the two sets of coordinates. Both data points are then projected onto the new 2D space defined by PC1 and PC2, as shown in Fig. 6. As observed from the metrics analysis, not all CWs (light-blue dots) have been identified by at least one PW (light-red squares). The QUBO-based

approach places roughly 60% of the CWs and the closest PWs are placed nearby the corresponding identified CW.

D. Problem scale-up

In the previous section, we showed that the discretization step and dimensionality reduction procedure we performed to map the 3D-RISM density into QUBO (see Fig. 9) have an impact on method's prediction accuracy. For instance, the problem instances we solved on the IBM QPU were generated to fit the machine's requirements, meaning that we considered a reduced portion of the binding pockets for the systems from Table I. Therefore, to assess the full potential of our strategy and to showcase the performance of our approach on problems with size and complexity relevant for drug discovery applications, we further scaled the QUBO instances complexity and size up to better map the underlying 3D-RISM density $g(\mathbf{r})$, and we evaluated their performances. Problems of this size cannot be solved directly on available QPUs; therefore, we rely only on classical heuristic methods to find the hydration sites corresponding to these scaled-up scenarios.

We kept using the 3D-RISM density computed on PDB ID = 3b7e with ligand as example while further reducing δ and τ_g. We generated QUBO instances exceeding the hardware limitation of 156 binary variables (each binary variable being mapped onto one qubit). The results are reported in Fig. 5, bottom panel, together with all the parameters used to generate the QUBO instances.

Compared to the 123-variable instance, P*, ⟨P⟩, and C all show a notable increase starting from the 900-variable instance. No remarkable improvements are observed beyond 900 variables on P* and ⟨P⟩, while C reaches a plateau value ∼0.9 starting at 1450 variables. ⟨CS⟩ shows a significant increase for the 6824-variables instance. Nevertheless, the 55-7 fold increase in the number of variables with respect to the 123-variable instance only corresponds to a ∼2.5 increase in ⟨CS⟩.

⟨P⟩ and ⟨CS⟩ both exhibit an associated statistical variance, which we report as the 95% confidence interval. We see how this attribute consistently decreases with the system size. We observe the same general trend in the additional results reported in Sec. S4 and Figs. S5 and S6 of the Supplemental Material [46], reinforcing this evidence. This trend is not that clearly followed by ⟨CS⟩. To support this observation, we also report their coefficients of variation in Table S4 of the Supplemental Material [46], computed as the ratio between the standard deviation and the mean value of the computed metrics. We can conclude that the PW location with respect to the CW reference becomes gradually more robust for larger instances, since the variability of ⟨P⟩ tends to become lower while its average value increases.

Before moving to the next section, we stress that the increase in the prediction accuracy and statistical consistency with the increase of the QUBO size is also observed in other examples taken from Table I, varying again δ and τ_g to obtain QUBO instances of increasing complexity, spanning up to ∼6000 binary variables. The evaluations of the resulting hydration-site predictions are reported in Sec. S4 and Figs. S3–S6 of the Supplemental Material [46].

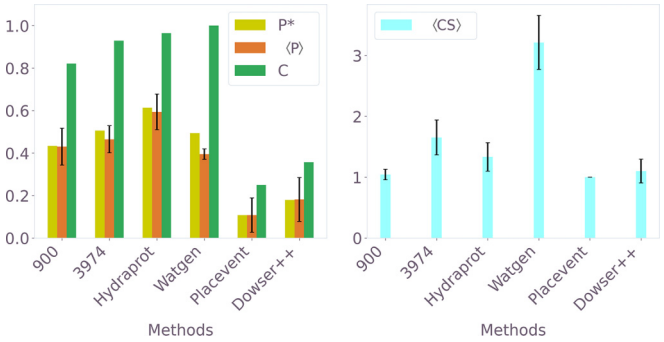


FIG. 7. Comparative performance analysis of different classical methods: the two best-performing QUBO models for PDB ID = 3b7e from Table I, with ligand included in the 3D-RISM density, characterized by 900 variables, $\sigma^2 = 1$ Å$_2$, $\tau_g = 0.10$, $\delta = 0.5$ Å, and 3974 variables with $\sigma^2 = 0.8$ Å$_2$, $\tau_g = 0.05$, and $\delta = 0.35$ Å, Hydraprot, Watgen, Placevent, and Dowser++. The figure presents four performance metrics as a function of the number of binary variables in the QUBO formulation. P* (closest water placement precision), ⟨P⟩ (cluster-averaged precision), C (fraction of crystal waters identified) are presented on the left plot; ⟨CS⟩ (average cluster size) are shown on the right plot. Error bars show the 95% confidence interval for ⟨P⟩ and ⟨CS⟩.

E. Comparison to other methods for hydration-site prediction

The task of predicting hydration sites in proteins is the focus of a variety of methods [4,6,10,49]. In this section, we compare our QUBO-based approach to a selected set of methods from the literature. Some of them are designed to be computationally affordable and can be used on modest-size hardware [4,9,10], at the cost of a lower accuracy in the prediction. Molecular dynamics and Monte Carlo-based simulations are in comparison more demanding in computational resources, reaching higher precision in the predictions while offering a potentially complete view on the thermodynamics of the system [5,8]. In Fig. 7, we compare our method's performances to a limited selection of popular, fast classical methods including: Hydraprot [51], a neural network crafted for this task, Placevent [52], Watgen [53], and Dowser++ [54]. These methods are based on different algorithms and therefore perform different types of computations with respect to each other. A short description of these methods is available in Sec. S3 of the Supplemental Material [46]. They have been classified in the literature and range from the so-called knowledge-based to deep-learning-based approaches [10,49]. They all exhibit low computational costs, only requiring, on average, just a few minutes of computing on a regular CPU. The selection we propose does not represent a full benchmark of the available classical methods, which is out of the scope of the present work. Among these methods, only Placevent uses 3D-RISM densities to identify the hydration sites, but is important to note that Placevent does not use a QUBO formulation, as the one proposed in this work.

To perform this comparison, we selected the best-performing QUBO instances for the test case of PDB ID = 3b7e with ligand, discussed in more detail in Fig. S5 of the Supplemental Material [46]. The solutions for these instances are obtained from simulations using SA heuristic, as

already explained above. All parameters used to generate the QUBO instances are reported in the caption of Fig. 7, where we included the results from both the 900-variable and the 3974-variable instances. Again, we observe an increase in the performance of the QUBO-based method passing from 900 to 3974 variables. Most notably, P and C considerably increase; the variance on ⟨P⟩ is reduced even if ⟨CS⟩ is on average ∼50% larger for the 3974 instance. In other words, with a ∼4.5-fold increase in the number of variables, only a ∼1.5-fold average increase of PW is observed.

To produce the hydration-site predictions of the selected methods from the literature, we used publicly available codes to run simulations on the same PDB ID and performed comparable performance analysis. We were not able to generate a set of hydration sites using WaterKit on this system, due to atom types definition while converting the PDB file to perform the ensemble structure generation, which is used by the software to estimate the best water sites. We were able to do so for another test system, reported in Fig. S7 of the Supplemental Material [46], showing quite a good performance of the method. From Fig. 7, we observe that our QUBO approach and Hydraprot show the best performances for this test case: The precision metrics are similar between these two methods, both of which substantially outperform the others. See Sec. S9 of the Supplemental Material [46] for more details on the metrics. Watgen identifies all the CW positions, but the average cluster size ⟨CS⟩ is more than double and presents a larger variance compared to those of QUBO and Hydraprot. This means that Watgen has the tendency to place a large number of water beside those corresponding to the CW sites.

Most notably, Placevent, which also uses 3D-RISM as input, clearly underperformed with respect to the QUBO approach and also with respect to all other methods. These observations are reinforced by a similar analysis performed on PDB ID = 4h2F without ligand, as reported in Fig. S7 of the Supplemental Material [46].

We conclude that the QUBO-based approach exhibits performances on par with, and in some cases superior to, other classical methods available in the literature. We can expect from our previous results that a quantum optimization will deliver results in at least the same range of precision when quantum hardware at the required scale is available.

As a final remark, it is worth noting that our QUBO simulations are based on a rather simple modeling of the intermolecular interactions governing protein-water and water-water interactions. All 3D-RISM simulations do not take into account the polarization contribution to such interactions. This missing physical effect is particularly relevant for charged protein residues or highly polarizable species, such as the water molecule itself. Including polarization would increase the cost of the classical simulations required to generate the QUBO cost function, but could potentially improve its performance.

F. Advantage forecasting

Projecting when quantum computing will have the potential to outperform classical approaches is nuanced. To forecast quantum advantage for hydration-site prediction, we first need to determine when our QUBO-based hydration-site prediction method achieves performance that is competitive with alternative approaches, reliably predicting crystal water positions. Second, we need to assess when quantum computing can be expected to outperform classical methods in solving the corresponding QUBO problem instances. As a next step, we have to estimate the quantum resources required to reach the potential threshold. Comparing these resource requirements with quantum hardware provider road maps, we can then get an idea of when the required resources are likely to be available. In this work, we focused specifically on IBM hardware, as our benchmarks were run on these devices and IBM provides a detailed public road map [48].

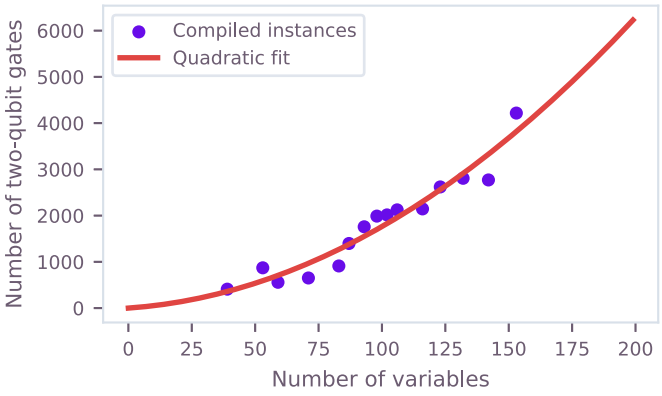


FIG. 8. Two-qubit gate scaling with problem size. For the 3b7e-with-ligand structure, purple dots show two-qubit gate counts after compilation for varying numbers of variables determined by the grid spacing. A quadratic fit (red line) is used for extrapolation.

For the QUBO-based approach to hydration-site prediction, we have seen that it becomes practically relevant at scales on the order of 1000 variables. For the 3b7e-with-ligand case within the specified grid extent, a grid spacing of 0.5 Å, corresponding to 900 variables (see Fig. 5, bottom panel), can be taken as an indicative threshold for the onset of competitiveness with alternative approaches (see Sec. II C). Consequently, the ability to reliably solve instances at this scale establishes a lower bound for the definition of a potential quantum advantage threshold.

Achieving quantum advantage requires outperforming any classical algorithm, whether exact or heuristic. An important milestone on the path toward full quantum advantage is the ability to outperform exact classical solvers, a regime referred to as quantum utility by IBM. As we have seen in our QPU benchmarks, an exact classical solver already struggles to solve the QUBO problem instances at a size of 123 variables, so the 900-variable instance lies well within this regime and can therefore be retained as a reasonable potential quantum utility threshold.

To estimate quantum resource requirements at this potential quantum utility threshold, we compiled instances up to 156 variables to a 156-qubit Heron device and extrapolated the scaling (Fig. 8). The two-qubit gate count scales approximately quadratically with the number of variables; under this scaling, a 900-variable instance would require on the order of 100 000 two-qubit gates. Comparing with IBM's quantum development road map [48], qubit counts around 1000 will likely be available by 2027. The projected two-qubit gate

requirements, however, remain out of near-term reach and will require early forms of error correction. According to IBM's road map [48], error-corrected devices capable of running 100 M gates are expected by 2029. Notably, the higher (square) connectivity in next-generation hardware, along with improved compilation techniques, could substantially reduce the number of gates required [55]. Note that if the gate count can be pushed down to 15k, then the target problem scales could be achieved as early as 2028 using IBM's Nighthawk devices.

To achieve full quantum advantage, the quantum approach must ultimately outperform not only exact but also heuristic classical solvers. However, a quantitative comparison with classical heuristics, such as SA, at the target scale is not yet possible because the required quantum resources are not available. Algorithmic considerations and empirical studies suggest that SA will increasingly struggle to find high-quality solutions within practical run times as problem size grows [56–58], whereas the quantum solver may benefit from a more structured exploration of the search space, albeit subject to potential expressibility limitations in the low-depth limit [59,60]. In line with the benchmarking results presented in the QPU benchmarks section, we therefore regard it as plausible that the Q-CTRL solver will achieve performance at least comparable to SA once hardware is sufficiently stable to support reliable circuit execution at the target scale of 900 variables. However, ultimately, any claim of quantum advantage must be supported by empirical evidence and comparisons with state-of-the-art classical approximate solvers; based on the above resource estimation and IBM's hardware road map, we anticipate that the necessary tests may become feasible by 2028.

III. DISCUSSION

We tested our QUBO-based method to identify 3D hydration sites in protein pockets, using a quantum solver to find the corresponding optimal solution. Hardware calculations have been run on two Heron IBM's devices (r2 and r3) using up to 123 qubits and have been complemented by larger classical QUBO simulations to estimate the accuracy of the approach. Various protein-ligand complexes of interest for health care have been used as test cases, demonstrating the feasibility of the simulations on NISQ devices, using Q-CTRL's optimization solver, including Fire Opal automated error suppression [42,43].

Comparing the small instances attainable on quantum hardware (∼100 variables) to larger instances that can be approximately solved through classical QUBO simulations (∼1000 variables), we showed how the method's accuracy scales toward improved solution qualities. For most cases, a good correspondence between the predicted water placement and the reference experimental measures is achieved. We also compared our method to a representative set of other classical approaches known in the literature, which are not using the QUBO formulation we proposed. From this comparison, we observed that our QUBO approach performs on par with the best fast hydration-site prediction methods we could test.

Moreover, we showed that, even around ∼100 variables, classical optimization solvers fail to provide a provably optimal solution within a reasonable computing time, whereas quantum optimizations or classical heuristics outperform their best-found approximate solution. To further estimate the quantum resource required at the utility threshold, we performed a detailed resource analysis based on number of optimization variables and on two-qubit gate counts. A 900-variable instance (900 qubits), which would allow us to tackle most of the industry test cases, requires on the order of 100 000 two-qubit gates and some early forms of error correction. This should make the target problem reachable within 5 years.

This opens to the possibility that quantum optimization might deliver a quantum advantage by providing better solutions, or equally good solutions but at a lower computational cost. It also demonstrates how practical quantum computing could rapidly integrate CADD pipelines to perform meaningful computations, as in the step needed to prepare protein-ligand complexes for MD simulations required for drug lead optimization or docking calculations.

IV. METHODS

In this section, we explain step-by-step our hydration-site prediction workflow: from approximating the input 3D-RISM density as a Gaussian mixture model (GMM), to optimizing the GMM using a QUBO formalism, to the implementation of the method on the quantum hardware using Q-CTRL's optimization solver on real hardware.

A. Step-by-step workflow for the QUBO-based 3D hydration-site prediction

In Fig. 9, we show a diagram with the workflow we followed to integrate the QUBO formulation into a computational pipeline to perform the hydration-site prediction simulations. It represents the sequence of computational tasks we apply to perform the simulations described in Sec. II, and to evaluate their accuracy.

The overall procedure requires to first select the 3D molecular structure that constitutes the target of the hydration-site prediction. This structure is needed to compute the 3D-RISM density, which bears the essential information about the water distribution around the solvated molecule, a protein in our case. A restricted area of the molecular structure is selected, where our attention is focused to perform the hydration-site prediction. A set of discrete variables is used to map the localized area's 3D-RISM density onto a QUBO problem, resampling such underlying 3D-RISM density on a coarse grid to reduce the problem size and complexity. Using this variable grid, the problem is cast into a GMM optimization problem, and the corresponding QUBO matrix is computed. Q-CTRL's Fire Opal solver is used to solve the optimization problem on a quantum hardware, and the best solution obtained is interpreted as the set of positions of the predicted water molecules.

1. 3D-RISM density computations

The very first step consists in selecting from Table I an x-ray structure and extracting the atomic coordinates of the system to compute the 3D-RISM continuous water density, as illustrated in Fig. 9. To compute the 3D-RISM continuous water density $g(\mathbf{r})$, we used the popular AMBERTOOLS

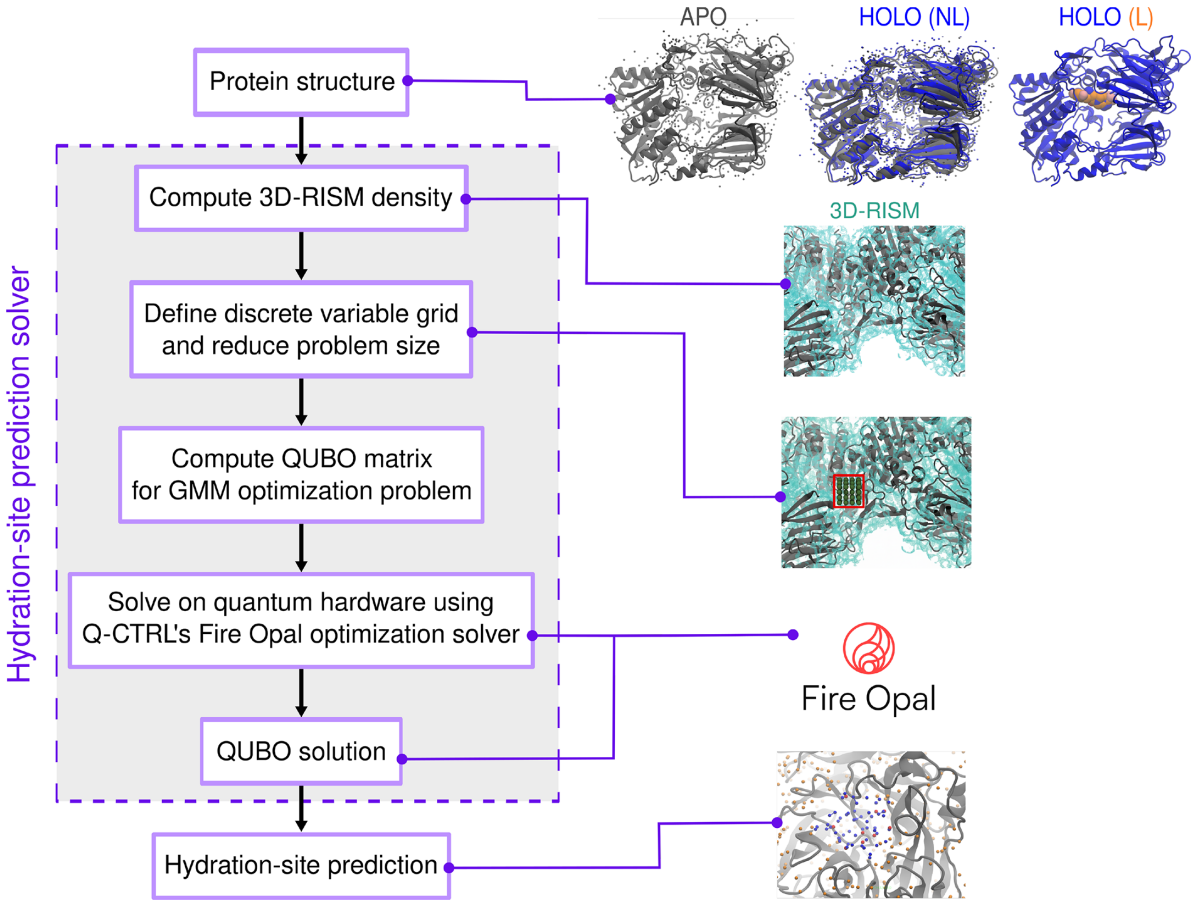


FIG. 9. Workflow illustrating the full-stack conversion of the hydration prediction problem into its QUBO form, followed by the translation to a QAOA-like hybrid quantum algorithm, from which we predict the hydration sites in a protein. An initial 3D-RISM continuous water density $g(\mathbf{r})$ within the protein of choice is converted into a binary variable placement grid. This grid is used to evaluate the QUBO matrix elements that map the information contained in $g(\mathbf{r})$ into the combinatorial optimization form that we use to find the GMM that best approximates the target density $g(\mathbf{r})$. The QUBO matrix can now be injected into the digital quantum optimization solver, including the error suppression pipeline. The output directly reads as coordinates in 3D space of the most likely sites for stable water molecules. Example structures of an apo (gray), holo without ligand (blue), and holo with ligand (ligand in orange) are added to the diagram. A 3D-RISM density isosurface is shown in cyan, as well as a grid of variable generated inside a focused region (green spheres inside a red square)

package, a widely used package in the community of MD simulations [50,61,62]. More details on the software and the parameters used to compute the 3D-RISM densities are reported in Sec. S7 of the Supplemental Material [46].

The 3D-RISM theory provides probability density of all the solvent molecules around a solute on a set of given positions in 3D space $\mathbf{r}$. It gives as output one density function for each atom type of the solvent molecule [50,61]. In the case of water, this means that two different sets of density can be computed, one describing the distribution of oxygen atoms, and one of the hydrogen atoms. High values of $g(\mathbf{r})$ correspond to regions with high solvent concentration compared to the pure bulk solvent. One can interpret those high-density regions as locations where water molecules are more likely to be found. Such locations can correspond to thermodynamically favorable locations, thanks to the balance between enthalpic and entropic contributions. Such locations can also correspond to metastable states where water molecules are kinetically trapped due to the surrounding environment. In our work, the solute is always represented by proteins or proteins in a complex with a ligand. The structure of the solute remains fixed, without any relaxation effect accounted for in the computation of $g(\mathbf{r})$, which, as a consequence, is biased by the imposed crystal structure.

3D-RISM requires as input the 3D coordinates of each atom of the molecular structure of the solute that one needs to solvate. In our case, the proteins molecular structures are stored in the PDB files we selected. In addition to the molecular structures, 3D-RISM also requires the force field parameters that are needed to solve the integral equations that are the very core of the method [50,61]. For the protein, we used the ff14SB [63] Amber force field, while the popular SPC/E water model was used to describe the solvent [64]. The default closure, namely, the Kovalenko-Hirata (KH) closure, was employed in all 3D-RISM calculations. The main advantage of KH compared to the hypernetted-chain closure is the superior numerical convergence, especially for systems such as those we study. Both closures handle Coulombic interactions well, while overestimate non-Coulombic interactions [65].

For the five holo structures presented in Table I, we provided two sets of 3D-RISM densities, one computed with

the effect of the ligand, and one without it. In such cases, we used the same PDB ID, the protein structure itself being unchanged, while the electrostatic and steric effects of the ligand in the binding pocket were taken into account (L) or ignored (NL), resulting in a different 3D-RISM density. This is a pure model-related information, which can be used to estimate the ligand effect on the computed 3D-RISM density and its impact on the hydration-site prediction. Ligand parameters for 3D-RISM calculations were derived using the General Amber Force Field. Partial charges were assigned via the AM1-BCC semiempirical method using the Antechamber module in AMBERTOOLS. Missing bonded parameters were generated using parmchk2 to ensure a complete force field description of the ligands. PDB ID = 3beq is the only case for which we provide only one 3D-RISM density, since it represents the apo structure of PDB ID = 3b7e. For each PDB ID, and each L and NL structure, the density function $g(\mathbf{r})$ was computed on a grid with 120 × 120 × 120 points with spacing $\delta = 0.5$ Å. The only exception is represented by PDB ID = 1 × 70, for which larger grids have been used, with 200 × 200 × 200 points for case L and 180 × 180 × 180 points for case NL, respectively. The computed densities were saved in a *.dx* format, which contains the information on the geometry of the grid and on the point-by-point values of the computed functions. We ran the 3D-RISM simulations on standard CPUs, with an average simulation time of ∼30 min. For the larger boxes presented in our tests, computing times were considerably larger, on the order of ∼3 h. Machine learning models that predict 3D-RISM densities have been proposed and might improve both accuracy and computational efficiency [66,67]. Moreover, it is always possible to rely on high-performance computing to reduce execution time, which was not the intention in the present work.

2. Discrete variable grid and problem size reduction

The original grids used for the computation of the 3D-RISM densities described above encompass a number of points on the order of ∼1 000 000. The quantum hardware resources presently available are insufficient for this problem sizes, while classical computational strategies to solve QUBO problems of such dimensionality become extremely expensive. Therefore, following the workflow presented in Fig. 9, we operated a dimension reduction of the problem. Discretizing the $g(\mathbf{r})$ function on coarse grids, we generated QUBO problems of adequate sizes for the quantum calculations performed above. First, we focused only on the binding site of the protein. This choice is motivated by the fact that the binding site is often the most relevant portion of the protein for a drug discovery project. We further reduce the size of the problem by selecting the density to be considered in the QUBO formulation. We define a density threshold τ_g to ignore locations where the density is lower than a chosen value. This step is not necessary to model the hydration-site prediction problem as a QUBO, but is only required to handle the practical constraints imposed by the quantum hardware on the one hand and to reduce the classical simulation costs on the other.

In practice, to operate the dimensionality reduction described above, after having computed $g(\mathbf{r})$ given a protein.pdb, we proceeded as follows:

(1) We select a cubic subdomain of the density including the protein binding pocket.

(2) The crystal waters present in the binding pockets are selected and their coordinates are saved, since they will be later used to evaluate the performance of the hydration-site prediction [44].

(3) We define a QUBO problem inside the selected region, where the variables are arranged in a regular cubic grid $\mathbb{G}$ with defined grid spacing δ.

(4) We keep only the grid points located in correspondence to, or close to, a $g(\mathbf{r})$ value $> \tau_g$.

(5) The QUBO matrix is computed on the resulting collection of grid points, using a uniform Gaussian variance at each point, so that $\sigma_i^2 \equiv \sigma^2 \quad \forall i \in \mathbb{G}$.

To prepare for hardware tests, we consider cubic subdomains of 15 Å sides and use a coarser grid spacing compared to the original 3D-RISM grid setup, with QUBO grid points spacing larger than the original 0.5 Å in the 3D-RISM density. The main parameters of the grids used are reported in Table II. More information about the effects of the choice of δ and σ^2 is discussed in Secs. S2 and S8 of the Supplemental Material [46] . The grid generated in this way is used to formulate the hydration-site prediction problem as a GMM optimization, cast as a QUBO, as shown in Fig. 9. The QUBO matrix elements are computed on this grid and the details of the QUBO formulation of the hydration-site prediction problem are described in the next section.

The choice of a density threshold $\tau_g > 0$ excludes dehydrated regions [$g(\mathbf{r}) < 1$]. We acknowledge that while density-depleted zones are often relevant for drug-designing purposes, their explicit characterization remains outside the scope of the present approach, which is tailored to resolve high-probability water localization. This choice is motivated by the formulation of the problem and the structure of the QUBO cost function. As described in more detail in the Supplemental Material [46] (Sec. S1), we subtract the bulk water value from the 3D-RISM density so that $g(\mathbf{r}) \leftarrow g(\mathbf{r}) - 1$, and later we scale the amplitude of each normalized Gaussian to match the local 3D-RISM shifted density. Therefore, when considering negative density regions, the diagonal terms in the QUBO matrix could still be negative, incorrectly indicating a favorable location for placing a water molecule where the negative density actually implies no water should be present. Furthermore, nondiagonal terms could also be potentially negative: if the two "interacting" Gaussian functions are placed one in a negative density area and the other in a nearby positive density region. This is in contradiction with the fact that the nondiagonal (quadratic) terms are meant to introduce a penalty (increase the value of the QUBO cost function) and avoid two water molecules to be placed too close to each other. The potential impact of failing to distinguish between positive and negative density areas is to place water in locations with negative density values, overpopulating nearby areas due to the negative off-diagonal terms.

B. QUBO formulation of the 3D hydration-site prediction problem

While methods like the 3D-RISM provide a continuous density map of water molecule distribution on a spatial grid,

the explicit location of stable water molecules and their thermodynamics is complementary, valuable information. Approaches such as Placevent [52] and GaSol [68] work in this direction, extracting an optimal water network coerced by the continuous 3D-RISM distribution.

Some of the authors proposed an alternative approach to transition from the continuous density to a discrete optimization problem, approximating the 3D-RISM density distribution with a GMM [41]. Each Gaussian component of the GMM represents a potential, localized hydration site, transforming the problem from continuous sampling to the selection of an optimal subset of discrete sites [41]. The problem then becomes to find the GMM that best fits the given 3D-RISM density.

To do so, we consider a set of N potential hydration sites: Our aim is to associate with each site a Gaussian center in the optimal GMM, which we need to determine. We introduce a set of binary variables, $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, where each $x_i \in \{0, 1\}$ is associated with the ith Gaussian component of the GMM [41], with

(1) $x_i = 1$ if a water molecule is placed at the ith potential site;

(2) $x_i = 0$ if no water molecule is placed at the ith site.

The central challenge is to select a configuration of water sites, $\mathbf{x}$, that minimizes the total system energy, while satisfying physical constraints, such as the minimum repulsive distance between two water molecules. In our approach, all the thermodynamics information is contained in the 3D-RISM density function $g(\mathbf{r}) : \mathbb{R}^3 \to \mathbb{R}$, with $\mathbf{r}$ being a spatial coordinate. This function, which can be computed at a reduced computational cost on a predefined grid of points, has large positive values in correspondence of locations where the water-water and water-protein interactions are favorable [41,61]. Therefore, we aim at finding the GMM that minimizes the following L^2 norm [41]:

$$I^2 := \int_{\mathcal{C}} \left(g(\mathbf{r}) - \sum_{i=1}^{M} \mathcal{G}(\mathbf{q}_i, \sigma^2)(\mathbf{r})\, x_i \right)^2 d\mathbf{r}. \tag{1}$$

Expanding the square and with minimal manipulation of the resulting terms, it is easy to see how this can be written as a QUBO cost function. We refer to the original work for more details on the derivation [41]. In Eq. (1), we denote by $\mathcal{G}(\mathbf{q}_i, \sigma^2)$ an isotropic Gaussian with mean $\mathbf{q}_i \in \mathbb{R}^3$ and variance σ^2. The set of $\mathbf{q}_i$ defines the grid in 3D space where the binary variables are placed. Note that this grid can correspond to the 3D-RISM one.

The hydration-site prediction problem is solved by minimizing a quadratic objective function $C(\mathbf{x})$, as defined by the standard QUBO equation

$$\min_{\mathbf{x}\in\{0,1\}^N} C(\mathbf{x}) = \sum_{i=1}^{N} Q_{ii} x_i + 2 \sum_{i<j}^{N} Q_{ij} x_i x_j. \tag{2}$$

In this context, the coefficients of the symmetric QUBO matrix $\mathbf{Q} \in \mathbb{R}^{N,N}$ are specifically engineered to encode the energy landscape of water placement:

(1) The linear terms Q_{ii} correspond to the linear bias applied to the binary variable x_i. This term can be seen as a self-energy term, connected to the solvation free energy of placing a single water molecule at site i, reflecting its interaction with the static protein environment (e.g., protein residues and bulk solvent). A negative Q_{ii} value indicates an energetically favorable site.

(2) Quadratic terms Q_{ij} explicitly encode the pairwise interaction energy between two Gaussians, placed at sites i and j. This term is critical for introducing constraints. For two sites i and j that are too close, for example, Q_{ij} becomes a large positive penalty value, pushing the solver toward configurations that avoid too close water molecules.

By minimizing $C(\mathbf{x})$, a QUBO solver identifies the optimal configuration of water sites $\mathbf{x}^*$ that achieves the lowest overall energy, while simultaneously satisfying the necessary spatial separation constraints via the penalty terms [41]. More details on the parameters needed to evaluate the QUBO matrix elements are discussed in Secs. S1 and S2 of the Supplemental Material [46].

In the next section, we describe how we use the quantum hardware, in combination to Q-CTRL's Fire Opal, to perform the optimization needed to perform the hydration-site prediction.

C. Hardware deployment of variational quantum optimization via Q-CTRL's fire opal

Solving binary combinatorial optimization problems, such as the QUBO instances defined for the hydration-site prediction, is a challenging task for classical computers. Exact methods scale poorly, motivating heuristics that seek high-quality solutions within practical run times. While fully fault-tolerant quantum computing remains out of reach on today's noisy hardware, variational quantum algorithms have emerged as practical quantum heuristics with potential benefits over classical alternatives. In particular, variational quantum optimization partitions computation between quantum and classical resources: A quantum computer is used to sample from a parametrized quantum circuit, while a classical routine iteratively updates the circuit parameters to minimize an objective function encoding the target problem.

The prototypical variational quantum optimization algorithm is the quantum approximate optimization algorithm (QAOA), which applies layers of alternating cost and mixer unitaries, implementing the problem's cost function and facilitating transitions between states, respectively, to an equal superposition initial state [69]. While QAOA admits theoretical performance guarantees in the large-depth limit of many layers for certain problem types [70–72], naive, small-depth implementations on noisy hardware often fail to deliver high-quality solutions at practically relevant scales due to compilation overhead, device noise, and challenging optimization landscapes [73–75]. Realizing benefits therefore requires an end-to-end pipeline with system-level optimization across all stages.

Thus, to be able to reliably tackle the hydration-site prediction QUBO problem on IBM quantum devices, we employ Q-CTRL's Fire Opal Optimization Solver. As detailed in Ref. [43], the workflow comprises the following:

(1) *Ansatz setup.* The input problem is specified as a cost polynomial

$$C(\mathbf{x}) = \mathbf{x}^T \mathbf{Q}\mathbf{x} \quad \mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{N,N}, \quad \mathbf{x} \in \{0,1\}^N, \tag{3}$$

generated from the QUBO problem matrix $\mathbf{Q}$ as introduced in Eq. (2). This objective is used to set up an enhanced variational ansatz circuit compared to standard QAOA. Note that while for the hydration-site prediction we focus on quadratic cost functions, the solver also natively supports higher-order cost polynomials.

(2) *Circuit compilation.* To minimize overhead in the hybrid loop, the ansatz circuit is subjected to efficient parametric precompilation, making use of an enriched gate set with precalibrated, pulse-level instructions.

(3) *Error-suppressed hardware execution.* Within the solver workflow, Q-CTRL's performance management software applies an automated error-suppression pipeline that deterministically reduces the impact of noise and realizes maximum hardware performance without additional sampling overhead [42].

(4) *Classical parameter optimization.* The circuit parameters are updated by a classical optimizer configured to maximize the efficiency of quantum-resource utilization.

(5) *Classical postprocessing.* Measured bitstrings are refined via efficient local search correcting for uncorrelated bitflip errors.

Leveraging this combination of techniques, the solver has demonstrated that nontrivial binary combinatorial optimization problems can be solved at full device scale, outperforming competing approaches on published benchmarking instances [43], and enabling record-scale solutions to real-world optimization problems [76].

D. Performance metrics

Using the combination of the technologies described above, we were able to extract from the optimal solution obtained on the quantum hardware the discrete positions of the optimal network of water molecules. We assess the quality of the PW positions against experimentally determined CWs, which are treated as true hydration sites [6,9,10,44].

Let n be the total number of CWs and the set $\{a_i\}_{i=1}^n$, with $a_i \in \mathbb{R}^3$, their positions in 3D space. We consider that the QUBO-based method is able to place m PWs, with $\{b_j\}_{j=1}^m$, with $b_j \in \mathbb{R}^3$ the set of their positions. m can be greater, smaller, or equal to n. We evaluate whether at least one of the PWs matches one of the reference CWs. Each $a_i \in \mathbb{R}^3$ is considered as a center of a sphere of radius $R_s = 3$ Å and if at least one of the $b_j \in \mathbb{R}^3$ is found, the CW is considered as successfully identified.

The number $n*$ of PWs that successfully identified CWs is used to compute the fraction of identified hydration sites

$$C = \frac{n*}{n} \in [0,1], \tag{4}$$

with C close to 1 meaning that the PWs successfully cover the set of reference crystal waters. The threshold of 3 Å has been chosen to mimic a medium-low crystal structure resolution, being 1.5 Å or below generally considered a good resolution.

We complement this information by quantifying how many PWs have been found in the 3 Å radius spheres around each CW and their position. Let $\mathcal{C}_i$ be the PWs cluster around the successfully identified CW i, and $N_i = |\mathcal{C}_i|$ be its size. We use the union of all the $\mathcal{C}_i$ clusters to evaluate

$$\langle \mathrm{P} \rangle = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{b_j \in \mathcal{C}_i} \frac{S_{i,j}}{N_i} \right), \quad \text{with}$$
$$S_{i,j} = \frac{1}{1 + \|a_i - b_j\|} \in \left[\frac{1}{1+R_s}, 1 \right], \tag{5}$$

and

$$\langle \mathrm{CS} \rangle = \frac{\sum_{i=1}^{m} N_i}{m}, \tag{6}$$

the averaged precision [Eq. (5)] and the average cluster size [Eq. (6)], respectively. $\langle \mathrm{P} \rangle$ scores the quality of identified hydration sites based on both the number of PWs that identify a CW and the inverse of the Euclidean distance between them and the CWs. $S_{i,j}$ by itself accounts for the distance effect, rewarding closer placements. $\langle \mathrm{P} \rangle$ penalizes placement methods that provide large, diffuse clusters, overestimating the number of PWs around a single CW position. At the same time, it penalizes methods that are not able to successfully identify all CWs positions. $\langle \mathrm{CS} \rangle$ gives a quantitative information on average size of the identified clusters $\mathcal{C}_i$. Both metrics are designed so that a value close to 1 represent a good scoring for a hydration-site identification method. For both $\langle \mathrm{P} \rangle$ and $\langle \mathrm{CS} \rangle$, we evaluate the 95% confidence interval (CI_{95}), computed using the standard error of the mean from their standard deviation.

In addition to their average and CI_{95}, we compute the coefficient of variation (CV) that measures relative variability of the metrics and is defined as the ratio of standard deviation and the mean. A robust prediction is associated with a low CV value.

At last, we complement these metrics with the single closest water precision (P^*), defined as

$$\mathrm{P}^* = \frac{1}{n} \sum_{i=1}^{n} S_i^*, \quad \text{with} \quad S_i^* = \max_{b_j \in \mathcal{C}_i} (S_{i,j}). \tag{7}$$

This metric quantifies how precise is the identification of each hydration-sites CW based on the single closest PW position, b_i^*, within the cluster $\mathcal{C}_i$, ignoring the cluster size and the other points of the cluster.

In Sec. S9 of the Supplemental Material [46], we show the effect of smaller thresholds on the computed metrics, considering the PDB 3b7e, which has the best x-ray experiment resolution between our example (1.45 Å) and has been used as test case for a number of comparisons in this work.

ACKNOWLEDGMENTS

This work was funded in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 810367), project EMC2 (J.-P.P.). We acknowledge the use of IBM Quantum Credits via the IBM Quantum Startups Program for this work.

The views expressed are those of the authors, and do not reflect the official policy or position of IBM or the IBM Quantum Platform team.

J.-P.P. is shareholder and cofounder of Qubit Pharmaceuticals. All the remaining authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available because they contain commercially sensitive information. The data are available from the authors upon reasonable request. No software or code has been included, and all generated data are included as results in the document and in the Supplemental Material [46]. A Python notebook with an example code to run the optimization of hydration sites on IBM quantum hardware is available in Q-CTRL's Fire Opal documentation [77]. The QUBO matrix data required for running this code are available from the corresponding author upon reasonable request [78]. Execution also requires access to Fire Opal software and IBM Quantum Platform [79].

[1] A. B. Gurung, M. A. Ali, J. Lee, M. A. Farah, and K. M. Al-Anazi, An updated review of computer-aided drug design and its application to COVID-19, Biomed Res. Int. **2021**, 8853056 (2021).

[2] D. Vemula, P. Jayasurya, V. Sushmitha, Y. N. Kumar, and V. Bhandari, CADD, AI and ML in drug discovery: A comprehensive review, Eur. J. Pharm. Sci. **181**, 106324 (2023).

[3] Y. Chang, B. A. Hawkins, J. J. Du, P. W. Groundwater, D. E. Hibbs, and F. Lai, A guide to in silico drug design, Pharmaceutics **15**, 49 (2023).

[4] D. Bucher, P. Stouten, and N. Triballeau, Shedding light on important waters for drug design: Simulations versus grid-based methods, J. Chem. Inf. Model. **58**, 692 (2018).

[5] M. Blazhynska, N. Ansari, L. Lagardère, and J.-P. Piquemal, From water networks to binding affinities: Resolving solvation dynamics for accurate protein-ligand predictions, bioRxiv (2026), doi:10.1101/2025.10.17.683050.

[6] G. A. Ross, G. M. Morris, and P. C. Biggin, Rapid and accurate prediction and scoring of water molecules in protein binding sites, PLoS One **7**, e32036 (2012).

[7] Y. Lu, R. Wang, C.-Y. Yang, and S. Wang, Analysis of ligand-bound water molecules in high-resolution crystal structures of protein–ligand complexes, J. Chem. Inf. Model. **47**, 668 (2007).

[8] J. Michel, J. Tirado-Rives, and W. L. Jorgensen, Prediction of the water content in protein binding sites, J. Phys. Chem. B **113**, 13337 (2009).

[9] S. M. Ivanov, Calculated hydration free energies become less accurate with increases in molecular weight, PLoS One **19**, e0309996 (2024).

[10] M. L. Samways, R. D. Taylor, H. E. B. Macdonald, and J. W. Essex, Water molecules at protein–drug interfaces: Computational prediction and analysis methods, Chem. Soc. Rev. **50**, 9104 (2021).

[11] L. Lagardère, L. Maurin, O. Adjoua, K. El Hage, P. Monmarché, J.-P. Piquemal, and J. Hénin, Lambda-ABF: Simplified, portable, accurate, and cost-effective alchemical free-energy computation, J. Chem. Theory Comput. **20**, 4481 (2024).

[12] N. Ansari, Z. F. Jing, A. Gagelin, F. Hédin, F. Aviat, J. Hénin, J.-P. Piquemal, and L. Lagardère, Lambda-ABF-OPES: Faster convergence with high accuracy in alchemical free energy calculations, J. Phys. Chem. Lett. **16**, 4626 (2025).

[13] W. L. Jorgensen, Efficient drug lead discovery and optimization, Acc. Chem. Res. **42**, 724 (2009).

[14] S. A. Adcock and J. A. McCammon, Molecular dynamics: Survey of methods for simulating the activity of proteins, Chem. Rev. **106**, 1589 (2006).

[15] J. Hénin, T. Lelièvre, M. R. Shirts, O. Valsson, and L. Delemotte, Enhanced sampling methods for molecular dynamics simulations [article v1.0], Living J. Comput. Mol. Sci. **4**, 1583 (2022).

[16] M. S. Bodnarchuk, R. Viner, J. Michel, and J. W. Essex, Strategies to calculate water binding free energies in protein–ligand complexes, J. Chem. Inf. Model. **54**, 1623 (2014).

[17] S. de Beer, N. P. Vermeulen, and C. Oostenbrink, The role of water molecules in computational drug design, CTMC **10**, 55 (2010).

[18] J. Setiadi, F. Biedermann, W. M. Nau, and M. K. Gilson, Thermodynamics of water displacement from binding sites and its contributions to supramolecular and biomolecular affinity, Angew. Chem. Int. Ed. **64**, e202505713 (2025).

[19] Y. Ge, O. J. Melling, W. Dong, J. W. Essex, and D. L. Mobley, Enhancing sampling of water rehydration upon ligand binding using variants of grand canonical Monte Carlo, J. Comput.-Aided Mol. Des. **36**, 767 (2022).

[20] T. J. Inizan, F. Célerse, O. Adjoua, D. El Ahdab, L.-H. Jolly, C. Liu, P. Ren, M. Montes, N. Lagarde, L. Lagardère, P. Monmarché, and J.-P. Piquemal, High-resolution mining of the SARS-CoV-2 main protease conformational space: Supercomputer-driven unsupervised adaptive sampling, Chem. Sci. **12**, 4889 (2021).

[21] D. El Ahdab, L. Lagardère, T. J. Inizan, F. Célerse, C. Liu, O. Adjoua, L.-H. Jolly, N. Gresh, Z. Hobaika, P. Ren, R. G. Maroun, and J.-P. Piquemal, Interfacial water many-body effects drive structural dynamics and allosteric interactions in SARS-CoV-2 main protease dimerization interface, J. Phys. Chem. Lett. **12**, 6218 (2021).

[22] A. Schahl, L. Lagardère, B. Walker, P. Ren, H. Wioland, M. Ballet, A. Jégou, M. Chavent, and J.-P. Piquemal, Histidine 73 methylation coordinates β-actin plasticity in response to key environmental factors, Nat. Commun. **16**, 2304 (2025).

[23] L. El Khoury, Z. Jing, A. Cuzzolin, A. Deplano, D. Loco, B. Sattarov, F. Hédin, S. Wendeborn, C. Ho, D. El Ahdab, *et al.*, Computationally driven discovery of SARS-CoV-2 M_{pro} inhibitors: From design to experimental validation, Chem. Sci. **13**, 3674 (2022).

[24] N. Ansari, C. Liu, F. Hédin, J. Hénin, J. W. Ponder, P. Ren, J.-P. Piquemal, L. Lagardère, and K. El Hage, Targeting RNA with small molecules using state-of-the-art methods provides highly

predictive affinities of riboswitch inhibitors, Commun. Biol. **8**, 1405 (2025).

[25] L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Y. Ren, J. W. Ponder, and J.-P. Piquemal, Tinker-HP: A massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields, Chem. Sci. **9**, 956 (2018).

[26] O. Adjoua, L. Lagardère, L.-H. Jolly, A. Durocher, T. Very, I. Dupays, Z. Wang, T. J. Inizan, F. Célerse, P. Ren, J. W. Ponder, and J.-P. Piquemal, Tinker-HP: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using GPUs and multi-GPU systems, J. Chem. Theory Comput. **17**, 2034 (2021).

[27] B. Kohnke, C. Kutzner, and H. Grubmüller, A GPU-accelerated fast multipole method for GROMACS: Performance and accuracy, J. Chem. Theory Comput. **16**, 6938 (2020).

[28] J. C. Phillips, D. J. Hardy, J. D. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, *et al.*, Scalable molecular dynamics on CPU and GPU architectures with NAMD, J. Chem. Phys. **153**, 044130 (2020).

[29] D. A. Case, D. S. Cerutti, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, M. Ghazimirsaeed, G. M. Giambaşu, T. J. Giese, A. W. Götz, J. A. Harris, *et al.*, Recent developments in Amber biomolecular simulations, J. Chem. Inf. Model. **65**, 7835 (2025).

[30] F. Bigi, S. Chong, A. Kristiadi, and M. Ceriotti, FlashMD: Long-stride, universal prediction of molecular dynamics, arXiv:2505.19350.

[31] M. Plainer, H. Wu, L. Klein, S. Günnemann, and F. Noé, Consistent sampling and simulation: Molecular dynamics with energy-based diffusion models, arXiv:2506.17139.

[32] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, Machine learning for molecular simulation, Annu. Rev. Phys. Chem. **71**, 361 (2020).

[33] A. Agathangelou, D. Manawadu, and I. Tavernelli, Quantum algorithm for protein side-chain optimisation: Comparing quantum to classical methods, arXiv:2507.19383.

[34] B. Claudon, J.-P. Piquemal, and P. Monmarché, Quantum speedup for nonreversible Markov chains, Nat. Commun. **16**, 10732 (2025).

[35] B. Claudon, P. Rodenas-Ruiz, J.-P. Piquemal, and P. Monmarché, Quantum circuits for the Metropolis-Hastings algorithm, J. Phys. A: Math. Theor. **59**, 305304 (2026).

[36] A. Robert, P. K. Barkoutsos, S. Woerner, and I. Tavernelli, Resource-efficient quantum algorithm for protein folding, npj Quantum Inf. **7**, 38 (2021).

[37] I. Liliopoulos, G. D. Varsamis, T. Karamanidou, C. Papalitsas, G. Koulouras, V. Pantazopoulos, T. G. Stavropoulos, and I. G. Karafyllidis, Quantum algorithm for protein-ligand docking sites identification in the interaction space, J. Comput.-Aided Mol. Des. **39**, 40 (2025).

[38] Q.-M. Ding, Y.-M. Huang, and X. Yuan, Molecular docking via quantum approximate optimization algorithm, Phys. Rev. Appl. **21**, 034036 (2024).

[39] W. Li, Z. Yin, X. Li, D. Ma, S. Yi, Z. Zhang, C. Zou, K. Bu, M. Dai, J. Yue, Y. Chen, X. Zhang, and S. Zhang, A hybrid quantum computing pipeline for real world drug discovery, Sci. Rep. **14**, 16942 (2024).

[40] P. Das, A. Ray, S. Bhattacharyya, J. Platos, V. Snasel, L. Mrsic, T. Huang, and I. Zelinka, A brief review on quantum computing based drug design, WIREs Data Min. Knowl. Discov. **14**, e1553 (2024).

[41] M. D'Arcangelo, L.-P. Henry, L. Henriet, D. Loco, N. Gouraud, S. Angebault, J. Sueiro, J. Forêt, P. Monmarché, and J.-P. Piquemal, Leveraging analog quantum computing with neutral atoms for solvent configuration prediction in drug discovery, Phys. Rev. Res. **6**, 043020 (2024).

[42] P. S. Mundada, A. Barbosa, S. Maity, Y. Wang, T. Merkh, T. Stace, F. Nielson, A. R. R. Carvalho, M. Hush, M. J. Biercuk, and Y. Baum, Experimental benchmarking of an automated deterministic error-suppression workflow for quantum algorithms, Phys. Rev. Appl. **20**, 024034 (2023).

[43] N. Sachdeva, G. S. Hartnett, S. Maity, S. Marsh, Y. Wang, A. Winick, R. Dougherty, D. Canuto, Y. Q. Chong, M. Hush, P. S. Mundada, C. D. B. Bentley, M. J. Biercuk, and Y. Baum, Quantum optimization using a 127-qubit gate-model IBM Quantum computer can outperform quantum annealers for nontrivial binary optimization problems, New J. Phys. **28**, 064506 (2026).

[44] M. L. Samways, H. E. B. Macdonald, R. D. Taylor, and J. W. Essex, Water networks in complexes between proteins and FDA-approved drugs, J. Chem. Inf. Model. **63**, 387 (2023).

[45] https://www.rcsb.org/.

[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/gyqr-mv1h for additional details on parameter choices in Sec. S2, cost gaps, metrics, and method descriptions, which include Refs. [10,49,80].

[47] https://github.com/dwavesystems/dwave-neal.

[48] IBM Quantum computing|Technology and roadmap, https://www.ibm.com/quantum/technology.

[49] E. Nittinger, F. Flachsenberg, S. Bietz, G. Lange, R. Klein, and M. Rarey, Placement of water molecules in protein structures: From large-scale evaluations to single-case examples, J. Chem. Inf. Model. **58**, 1625 (2018).

[50] D. Roy and A. Kovalenko, Performance of 3D-RISM-KH in predicting hydration free energy: Effect of solute parameters, J. Phys. Chem. A **123**, 4087 (2019).

[51] A. Zamanos, G. Ioannakis, and I. Z. Emiris, Hydraprot: A new deep learning tool for fast and accurate prediction of water molecule positions for protein structures, J. Chem. Inf. Model. **64**, 2594 (2024).

[52] D. J. Sindhikara, N. Yoshida, and F. Hirata, Placevent: An algorithm for prediction of explicit solvent atom distribution-application to HIV-1 protease and F-ATP synthase, J. Comput. Chem. **33**, 1536 (2012).

[53] H.-H. Bui, A. J. Schiewe, and I. S. Haworth, Watgen: An algorithm for modeling water networks at protein–protein interfaces, J. Comput. Chem. **28**, 2241 (2007).

[54] A. Morozenko and A. A. Stuchebrukhov, Dowser++, a new method of hydrating protein structures, Proteins: Struct. Funct. Bioinf. **84**, 1347 (2016).

[55] Avoiding an unexpected roadblock in quantum computing—Compilation|Q-CTRL, https://q-ctrl.com/topics/avoiding-an-unexpected-roadblock-in-quantum-computing-compilation.

[56] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, Optimization by simulated annealing: An experimental evaluation; Part I, graph partitioning, Oper. Res. **37**, 865 (1989).

[57] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, Optimization by simulated annealing: An experimental evaluation; Part II, graph coloring and number partitioning, Oper. Res. **39**, 378 (1991).

[58] D. Henderson, S. H. Jacobson, and A. W. Johnson, The theory and practice of simulated annealing, in *Handbook of Metaheuristics*, edited by F. Glover and G. A. Kochenberger (Springer US, Boston, MA, 2003), pp. 287–319.

[59] M. Streif and M. Leib, Comparison of QAOA with quantum and simulated annealing, arXiv:1901.01903.

[60] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices, Phys. Rev. X **10**, 021067 (2020).

[61] T. Luchko, S. Gusarov, D. R. Roe, C. Simmerling, D. A. Case, J. Tuszynski, and A. Kovalenko, Three-dimensional molecular theory of solvation coupled with molecular dynamics in Amber, J. Chem. Theory Comput. **6**, 607 (2010).

[62] D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, *et al.*, AmberTools, J. Chem. Inf. Model. **63**, 6183 (2023).

[63] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB, J. Chem. Theory Comput. **11**, 3696 (2015).

[64] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, The missing term in effective pair potentials, J. Phys. Chem. **91**, 6269 (1987).

[65] https://ambermd.org/doc12/Amber24.pdf.

[66] K. Osaki, T. Ekimoto, T. Yamane, and M. Ikeguchi, 3D-RISM-AI: A machine learning approach to predict protein–ligand binding affinity using 3D-RISM, J. Phys. Chem. B **126**, 6148 (2022).

[67] D. J. Fowles and D. S. Palmer, Solvation entropy, enthalpy and free energy prediction using a multi-task deep learning functional in 1D-RISM, Phys. Chem. Chem. Phys. **25**, 6944 (2023).

[68] L. Fusani, I. Wall, D. Palmer, and A. Cortes, Optimal water networks in protein cavities with GAsol and 3D-RISM, Bioinformatics **34**, 1947 (2018).

[69] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[70] R. Shaydulin, C. Li, S. Chakrabarti, M. DeCross, D. Herman, N. Kumar, J. Larson, D. Lykov, P. Minssen, Y. Sun, *et al.*, Evidence of scaling advantage for the quantum approximate optimization algorithm on a classically intractable problem, Sci. Adv. **10**, eadm6761 (2024).

[71] S. Boulebnane and A. Montanaro, Solving Boolean satisfiability problems with the quantum approximate optimization algorithm, PRX Quantum **5**, 030348 (2024).

[72] S. Omanakuttan, Z. He, Z. Zhang, T. Hao, A. Babakhani, S. Boulebnane, S. Chakrabarti, D. Herman, J. Sullivan, M. A. Perlin, R. Shaydulin, and M. Pistoia, Threshold for fault-tolerant quantum advantage with the quantum approximate optimization algorithm, arXiv:2504.01897.

[73] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Noise-induced barren plateaus in variational quantum algorithms, Nat. Commun. **12**, 6961 (2021).

[74] J. Weidenfeller, L. C. Valor, J. Gacon, C. Tornow, L. Bello, S. Woerner, and D. J. Egger, Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware, Quantum **6**, 870 (2022).

[75] E. Pelofske, A. Bärtschi, and S. Eidenbenz, Short-depth QAOA circuits and quantum annealing on higher-order Ising models, npj Quantum Inf. **10**, 30 (2024).

[76] Accelerating the schedule for quantum-enhanced rail|Q-CTRL, Section: Case Studies, 2025, https://q-ctrl.com/case-study/accelerating-the-schedule-for-quantum-enhanced-rail.

[77] Predicting protein-pocket hydration sites using Fire Opal|Q-CTRL, accessed 13 April 2026, https://docs.q-ctrl.com/fire-opal/apply/predicting-protein-pocket-hydration-sites-using-fire-opal.

[78] QUbit Pharmaceuticals, Protein pocket hydration QUBO data. In Practical protein-pocket hydration-site prediction for drug discovery on a quantum computer, Zenodo, 2026, https://doi.org/10.5281/zenodo.19912753.

[79] Both Q-CTRL and IBM offer limited free access to their platforms.

[80] J. Eberhardt and S. Forli, WaterKit: Thermodynamic profiling of protein hydration sites, J. Chem. Theory Comput. **19**, 2535 (2023).

Correction: A support statement in the Acknowledgments has been corrected.