

# A Perspective on Protein Structure Prediction Using Quantum Computers

Hakan Doga,\* Bryan Raubenolt,\* Fabio Cumbo, Jayadev Joshi, Frank P. DiFilippo, Jun Qin, Daniel Blankenberg, and Omar Shehab



Cite This: *J. Chem. Theory Comput.* 2024, 20, 3359–3378



Read Online

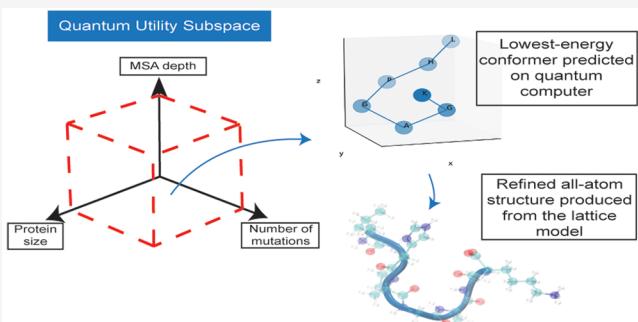
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Despite the recent advancements by deep learning methods such as AlphaFold2, *in silico* protein structure prediction remains a challenging problem in biomedical research. With the rapid evolution of quantum computing, it is natural to ask whether quantum computers can offer some meaningful benefits for approaching this problem. Yet, identifying specific problem instances amenable to quantum advantage and estimating the quantum resources required are equally challenging tasks. Here, we share our perspective on how to create a framework for systematically selecting protein structure prediction problems that are amenable for quantum advantage, and estimate quantum resources for such problems on a utility-scale quantum computer. As a proof-of-concept, we validate our problem selection framework by accurately predicting the structure of a catalytic loop of the Zika Virus NS3 Helicase, on quantum hardware.



As a proof-of-concept, we validate our problem selection framework by accurately predicting the structure of a catalytic loop of the Zika Virus NS3 Helicase, on quantum hardware.

## 1. INTRODUCTION

The intricate dance of life at the molecular level is orchestrated by proteins, with virtually all biological activity tied to the three-dimensional conformations they adopt. The phenomenon by which these structures are predetermined from their primary amino acid sequence is known as the “protein folding problem”, and it is inherently central to all life and its myriad of diseases. In nature, protein folding is a path-dependent process, meaning that the optimal path is usually taken.<sup>1,2</sup> For most biomedical research applications, predicting the optimal structure itself, without necessarily reproducing the optimal path, is arguably most important and more attainable. This is known as protein structure prediction (PSP) (see Figure 1). Decades of scientific inquiry have sought to unravel the mysteries of how and why proteins assume their complex structures, often driven by the quest for understanding diseases at the molecular level. To comprehend the function of a protein and devise effective drugs targeting them, an accurate depiction of their physiologically active structure is indispensable. Traditionally, this has been achieved through laborious wet lab experiments involving genetic modifications, protein isolation, and purification. Techniques like X-ray crystallography, NMR, and CryoEM have been instrumental in solving protein structures, revolutionizing our grasp of diseases. However, these methods are time-consuming, expensive, and not without limitations. Recognizing the need for alternatives, researchers turned to machine learning, exemplified by AlphaFold2,<sup>3</sup> RoseTTAFold,<sup>4</sup> and iTASSER,<sup>5</sup> which leverage

experimentally determined structures. While transformative, these methods may lack a nuanced understanding of the underlying physics, potentially hindering predictions of novel protein structures.<sup>6</sup> Physics-based methods (also known as template-free modeling) on the other hand, such as molecular dynamics (MD) simulations, face challenges in scalability and practicality. Interested readers may refer to<sup>7,8</sup> for comparative analyses of the above-mentioned methods.

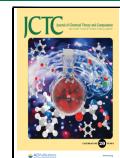
Fundamentally, protein folding is a physics problem. A computer algorithm armed with a realistic thermodynamic description and mathematical framework for modeling interactions between amino acids (residues) could theoretically navigate the vast conformational space, ultimately arriving at an optimal solution.<sup>9</sup> Yet, the computational complexity of this challenge has made it a persistent hurdle in the life sciences. In this landscape, quantum computing holds the potential to provide meaningful utility for this problem. While there have been major advances in quantum hardware and algorithm development,<sup>10</sup> finding the appropriate class of problems amenable to quantum advantage is still an open question for all areas of practical applications. Estimating the resources needed

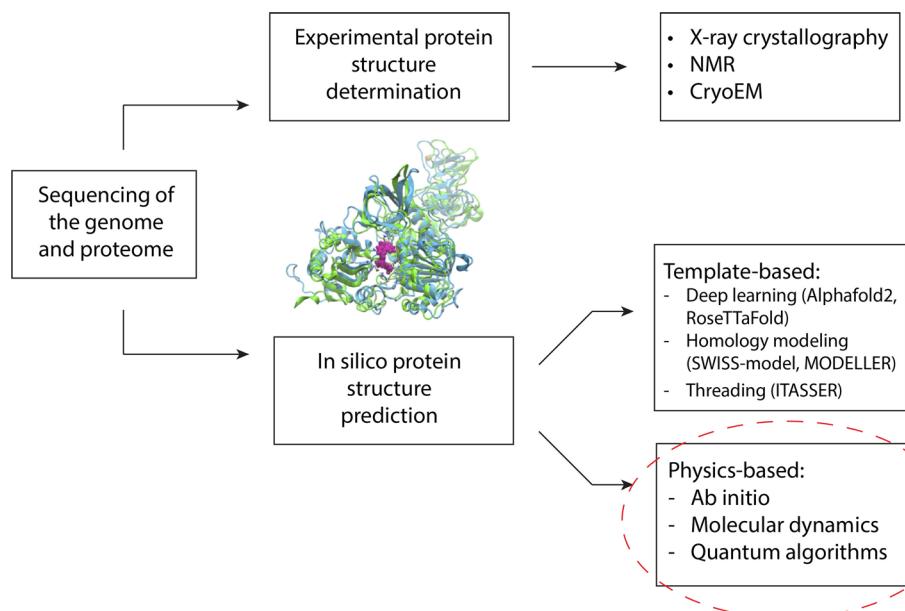
Received: January 22, 2024

Revised: April 19, 2024

Accepted: April 22, 2024

Published: May 4, 2024





**Figure 1.** Overview of the PSP pipeline. Following genomic sequencing, the primary amino acid sequence is determined. The experimental method then starts with expressing this protein by genetically modifying another organism with this new sequence. This organism will then translate these proteins, and the new protein of interest can be isolated, purified, and then solved using X-ray crystallography, NMR, or CryoEM. The in silico methods on the other hand, simply take the primary amino acid sequence as input and the structure is predicted by either a physics-based method (where the underlying biophysics is somehow simulated) or a template-based method (where machine learning algorithms predict structures based on patterns found in a training set of experimental templates). The method we adopt in this work falls under the category of physics-based algorithms. As an illustrated example, an in silico model and X-ray crystal structure of the SARS-CoV2 NSP13 helicase (PDB: 7NN0) are superimposed, along with a docked known inhibitor (colored in magenta).

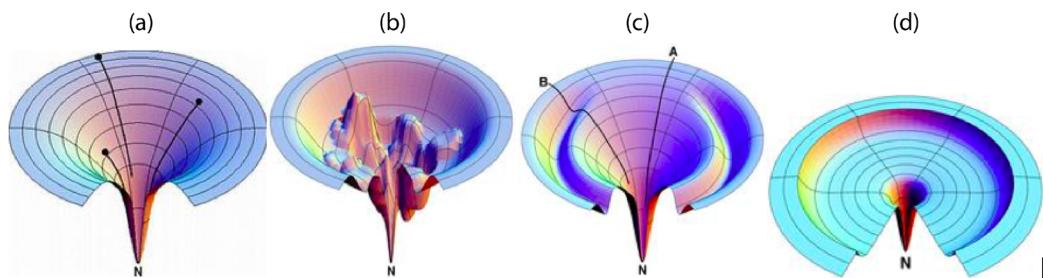
to enable quantum advantage is equally challenging, with sporadic progresses. In this work, we provide our perspective, a scalable framework to identify PSP problem instances that may be amenable to quantum advantage. Narrowing the problem space through a combination of protein sequence length, intrinsic disorder, number of mutations and multiple sequence alignment (MSA) depth, we opine on how to select problems at different scales that have been found challenging for state-of-the-art classical methods. In particular, MSA depth, which is characterized by the number of effective sequences ( $N_{eff}$ ) in the alignment and computed using a similarity metric as described in,<sup>3</sup> impacts the accuracy of the prediction. One can interpret the MSA depth as the availability of training data related to the target protein from which the relevant statistical summaries are calculated. Furthermore, the approach we take here makes our problem selection framework explainable, which is a challenge in classical machine learning.<sup>11</sup>

**Section 2** introduces quantum computing fundamentals, laying the groundwork for understanding its application in solving the protein folding problem, first demonstrated in.<sup>12</sup> **Section 3** describes our perspective. It delves into the complexities of PSP, highlighting challenges related to protein size, intrinsic disorder and mutations, and the role of MSAs. The focus is on identifying the problem subspace where quantum computing might outperform deep learning methods reliant on MSAs. Experimental results in **Section 4** showcase our quantum-classical hybrid workflow predicting the structure of the seven amino acid catalytic P-loop from a vital Zika virus protein (NS3 Helicase) using the IBM\_Cleveland quantum computer, validating our perspective with a concrete example. **Section 5** discusses the scalability of the quantum algorithm and provides high level quantum resource estimation for PSP problems. The conclusion reflects on the significance of the

work, opening avenues for future exploration in leveraging quantum computing for protein folding challenges.

## 2. QUANTUM COMPUTING: A BRIEF INTRODUCTION

Quantum Computing is a new model of high performance computing where the traditional foundation of computing, i.e. binary logic, has been replaced by theories of quantum mechanics.<sup>13,14</sup> This section aims to provide a very concise introduction to the topic, while the readers are encouraged to review<sup>15</sup> for a comprehensive overview, and **Section 2** of<sup>16</sup> for a primer from the perspective of healthcare and life sciences. The power of quantum computing comes from quantum mechanical effects such as superposition, entanglement, negative state probability i.e. interference, and probabilistic measurement. These phenomena sometimes allow a quantum algorithm to naturally map the degrees of freedom of quantum hardware to those of a target quantum system and simulate in an efficient manner. For some suitably structured problems, it is possible to exploit these phenomena to design quantum algorithms capable of traversing a search space or optimization cost landscape in a more efficient probabilistic manner. Over the last few decades, quantum advantages have been theoretically demonstrated for prime factorization,<sup>17</sup> unstructured search,<sup>18</sup> network flows,<sup>19</sup> quantum simulation,<sup>20</sup> topological invariants,<sup>21</sup> partition functions,<sup>22</sup> semidefinite programming,<sup>23</sup> linear systems,<sup>24</sup> differential equations,<sup>25</sup> dynamic programming,<sup>26</sup> bilinear functions,<sup>27</sup> etc. A few recent works have also reported quantum utility<sup>28</sup> or empirical quantum advantage,<sup>27,29–34</sup> as defined in.<sup>33</sup> Several architectures have been proposed for scalable quantum computers including neutral atom qubits,<sup>35</sup> spin qubits,<sup>36</sup> topological qubits,<sup>37</sup> trapped-ion qubits,<sup>38</sup> and superconducting qubits.<sup>39</sup> With 30+ quantum computers, IBM Quantum Experience is



**Figure 2.** This graphic was originally created and released in the public domain by Ken A. Dill. Original caption: The illustrations of proposed energy landscape that each demonstrate the degree of freedom a protein possesses in terms of configurations and the multidimensional routes that a protein can take to achieve its final configuration. From left to right for proposed funnel-shaped energy landscapes: the idealized smooth funnel, the rugged funnel, the Moat funnel, and the Champagne Glass funnel.<sup>73</sup>

the first ever and largest cloud-based quantum computing service. Quantum software development kits like Qiskit,<sup>40</sup> CUDA Quantum,<sup>41</sup> Forest,<sup>42</sup> PennyLane,<sup>43</sup> CirQ<sup>44</sup> and Braket<sup>45</sup> are also available with increasingly more application verticals in each release. Several vendors have also announced their developmental road maps<sup>46–49</sup> with IBM leading the way to achieve one hundred thousand qubits with increased gate quality and speed within a decade.<sup>10,50</sup> In parallel with the progress of technologies, the research communities have also presented their point of views for using quantum computers in several areas of applications including scientific discovery,<sup>51</sup> biological sciences,<sup>52</sup> nuclear physics,<sup>53</sup> high-energy physics,<sup>54</sup> cell-centric therapeutics,<sup>16</sup> financial engineering,<sup>55</sup> climate science,<sup>56</sup> etc. Finally, the responsible usage of quantum computing is also emerging as an area of research.<sup>57</sup>

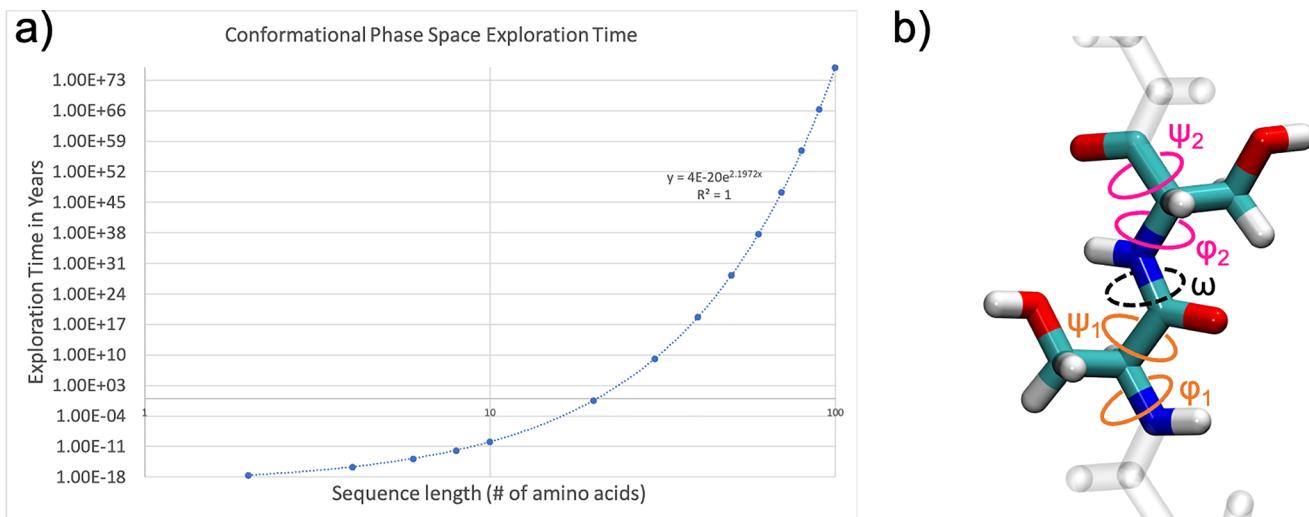
### 2.1. Quantum Search and Optimization Algorithms.

While finding the global minima of the energy landscape of protein conformation is an optimization problem, one could also imagine it as a search problem. In that case, the database consists of all possible conformers with the lowest energy conformer being the marked item to search for. The quantum algorithm for unstructured database search is Grover's algorithm which provides quadratic speedup over state-of-the-art classical search algorithms.<sup>18</sup> The algorithm initializes a quantum state in an equal superposition over all database entries. It then iterates a step which amplifies the amplitude of the target entry. After  $O(\sqrt{N})$  iterations, a measurement will find the entry with high probability, and it has been shown as optimal.<sup>58</sup> The optimality of Grover's algorithm is also supported by the observation that it defines a geodesic in quantum Hilbert space.<sup>59,60</sup> Therefore, any other quantum search or optimization algorithm may be considered as a parametrized approximation of Grover's algorithm. One such algorithm is quantum approximate optimization algorithm (QAOA)<sup>61</sup> (or the equivalent variational quantum eigensolver (VQE) algorithm<sup>62</sup>). It works by setting up a cost Hamiltonian whose ground state encodes the search result. A quantum system is initialized in an easy-to-prepare state. A series of unitary operations alternately apply a mixer Hamiltonian and the cost Hamiltonian. At the end, the state of the system is measured, giving a candidate solution. While low-depth QAOA is not expected to outperform the state-of-the-art classical algorithm<sup>63,64</sup> (see<sup>65,66</sup> for a more formal study of the limitations of variational quantum algorithms), at a higher depth the algorithm increasingly becomes an approximation of the Grover's algorithm and has the potential to maximize fractional Grover speedup.<sup>61</sup> See Section 4 of the Supporting

Information for a detailed discussion of algorithmic structures needed for quantum speedup.

**2.2. The Protein Folding Funnel and the Prospect of Quantum Advantage.** The theory of protein folding considers how proteins' primary amino acid sequences dictate how they fold rapidly and specifically into their native 3D structures. The classical view was that folding occurred through discrete intermediates along a linear pathway. In contrast, the energy landscape theory views folding as a progressive organization of an ensemble of partially folded structures funneled toward the native state.<sup>67–69</sup> Evolution has shaped proteins to have a rugged, funnel-like landscape biased towards the functional native structure. The funnel shape implies folding robustness, with different routes down the funnel possible for the same protein. To discuss this robustness relation,<sup>68</sup> uses the analogy of an ensemble of skiers distributed over a mountain representing the energy landscape, following their trajectories to reach the bottom of the valley, the global minimum. Figure 2a) shows an idealized version of this funnel without any bumps. However, in reality, as characterized first in<sup>70,71</sup> using spin-glass based models and statistical mechanics and later further analyzed in<sup>72</sup> viewing folding funnels as convergent kinetic pathways, the folding funnel can be more rugged (See Figure 2b), leading to multistate kinetics. In these more rugged energy landscapes, as discussed in-depth in,<sup>68</sup> the Brownian motion of convergence to the native state can be slowed down by kinetic traps, local minima of intermediate states, energy barriers, forcing the destruction of favorable contacts between residues and narrow pathways to reach the native state. This kinetic view suggests that the overall complexity of the funnel impacts the robustness of the folding process. Just like most other spontaneous thermodynamic phenomena, protein folding is largely path dependent—nature takes the optimal path to the optimal solution. The native state's stability increases moving down the funnel through local folding events. Common patterns emerge like topology determining mechanisms, but details depend on subtle sequence variations. The funnel concept explains how folding can be fast despite the vast number of possible non-native conformations. The landscape guides the protein through a directed search. Figure 2 shows four proposed folding funnels.

Finding the global minimum energy in the folding funnel can be understood as a search problem where the database entries are conformational energies. The funnel shape indicates that despite the overall "easy-to-follow" macro structure, the ruggedness on the funnel wall and the bottom induces hardness to conformation prediction.



**Figure 3.** A graphical representation of the rapid growth in the search space as a function of  $n$  amino acids. (a) The data considers that for a protein with  $n$  amino acids, there are  $n - 1$  peptide bonds. For each peptide bond, there are also 2 other bond angles on either side of the  $\alpha$  carbon,  $\phi$  and  $\psi$ , and assumes that each peptide in the sequence can adopt up to 3 conformations (3 combinations of these bond angles). (b) So, for a protein with  $n$  amino acids, there are a total of  $3^{2(n-1)}$  possible conformations. Assuming 1 ps is spent to sample each conformation, the y-axis represents the total exploration time in years to sample all possible conformations for a protein with  $n$  amino acids. For a detailed discussion on Levinthal's paradox and growth in the conformational search space, please see Section 2 of the Supporting Information.

The ruggedness of the energy landscape at the bottom of the folding funnel correlates with the complexity and diversity of conformations a protein can adopt.<sup>69,74–83</sup> In fact, in<sup>84</sup> the authors construct a mathematical model to analyze the relation between this complexity and the protein folding kinetics and thermodynamics (See eq 2 in<sup>84</sup>). This overlap function  $\chi$  is a measure of existence of non-native structures for the protein with respect to the native state, and built in relation to the structure of the energy landscape. A smooth, single-minimum funnel bottom indicates a rigid protein with few accessible conformers. A rugged funnel bottom with many minima reflects a flexible protein existing as an ensemble of diverse conformers interconverting between substates. The more rugged the bottom, the more conformational flexibility the protein has to sample alternate structures. Proteins with rugged, multimimima landscapes can bind ligands in a nonspecific way by selecting conformers complementary to each ligand from their ensemble. In contrast, proteins with smooth funnels have less conformational heterogeneity and exhibit more specific binding to preorganized native structures. Increased ruggedness also enables phenomena like induced fit binding, crystal packing effects, domain swapping, and misfolding aggregation by allowing access to alternate conformers. Molecular chaperones may smooth rugged landscapes to promote proper folding over misfolding. In summary, the complexity and interconversion of a protein's conformational substates is directly related to the ruggedness of the energy landscape at the bottom of its folding funnel.<sup>79</sup> Rugged landscapes confer flexibility and multispecificity, while smooth funnels restrict accessible conformers and interactions.

The ruggedness in the folding funnel has a curious parallel in unstructured search problems when one interprets the problem as traversing the geodesic of a search space (Grover's quantum search algorithm) or optimizing a rugged cost function using a variational algorithm with a parametrized quantum circuit (QAOA algorithm). Besides, the space of the folding pathways has a natural tree structure (See Figure 5 of<sup>85</sup>) similar to the search tree structure in Grover solution space. These

algorithms have been briefly introduced in Section 2.1. It has been observed that the ruggedness of the search or optimization landscape creates amenability for quantum advantage<sup>86–95</sup> (see Section 4 of the Supporting Information for a detailed discussion). The quantum advantage appears when there is a least amount of information or a highest amount of uncertainty about the search space which is also known as the “worst case” in algorithmic complexity analysis. In the case of Grover's algorithm one can attribute this quantum speed up to the fact that quantum mechanics allow us to associate negative probabilities to search paths. Therefore, a subset of the search paths will cancel each other, rendering the search problem smaller and easier. If one is using the QAOA algorithm, the same phenomena may be viewed as an effect of quantum tunneling where a quantum search process may cross energy barriers in a probabilistic manner that does not have any classical analog. Under these circumstances, we can conjecture that the more rugged a folding funnel is, the easier it will be for a quantum algorithm to find the lowest energy conformation relative to the performance of its classical competitors.

### 3. WHAT MAKES PROTEIN STRUCTURE PREDICTION HARD?

**3.1. Performance of Physics-Based Methods against Increasing Sequence Length.** Regardless of the physics-based computational method used, sequence length quickly becomes a major limitation. As a protein sequence becomes larger, there is an exponential increase in the search space (number of possible conformations), and a corresponding exponential growth in the required run time for an exhaustive search (see Figure 3a). Most nonmolecular dynamics (non-MD) PSP methods (also known as *ab initio* or free modeling methods) have generally been limited to structures of only a few dozen amino acids in length. One study<sup>96</sup> compared the performance of four of these methods. While most of these methods appeared to produce backbone Root Mean Squared Deviations (RMSDs) within 3.0 Å when compared to the

experimental results, the investigated structures were all 31 amino acids or less in length. PEP-FOLD3<sup>97</sup> is designed to model peptides between 5 and 50 amino acids, and has been successful in doing so in different studies. Quark,<sup>98</sup> an *ab initio* program developed by the same group who created ITASSER,<sup>99</sup> can predict structures accurately, but was mainly designed for fragments up to 20 amino acids in length. For small peptides at this scale, these programs perform well (run times of less than 14 h for all test cases reported in<sup>96</sup>) and can offer viable solutions with the right compute resources. For larger, biologically relevant proteins, physics-based classical algorithms like these may not be as suitable due to the rapid increase in the conformational space.

Although there has been remarkable success in some studies where all-atom MD simulations are applied to predict folded structures, with the most impressive results arguably being those of D.E. Shaw's research group,<sup>100,101</sup> sequence length still remains a limitation. These efforts led to the simulation of possible folding paths, as well as an accurate prediction of the folded state, for human ubiquitin<sup>102</sup> and a set of 12 "fast folding" proteins.<sup>103</sup> The main issue is the amount of sampling time (the length of the simulation in time, as defined by the product of the total number of discrete time steps and step size) as well as compute resources it took to achieve these results, since most researchers may have limited access to either. Although the majority of the structures in<sup>103</sup> folded to near perfect accuracy, much of this occurred in the order of hundreds of microseconds to milliseconds of sampling time (depending on the system size, this generally requires significant run times ranging from weeks to months even when using conventional HPC resources) and the investigated proteins were relatively small, ranging in size from roughly 10–80 residues. As an example, in their study,<sup>103</sup> the TC10b variant of the Trp-cage mini protein (sequence length of 20 amino acid residues, PDB: 2JOF) was simulated for approximately 208  $\mu$ s, with folding and unfolding events being observed for 14  $\mu$ s. This was done on the Anton supercomputer, a highly sophisticated cluster designed to accelerate all-atom MD simulations. To highlight the discrepancy between generated sampling time and the corresponding computer run time, we can refer to some of our own simulations of Trp-cage (discussed in the following section). Using 16 cores and 4 NVIDIA A100 GPUs, our team was able to generate 725 ns of sampling time per every 24 h of computer wall time, for a system of Trp-cage immersed in 2200 water molecules (a total of 9092 atoms in the system). At this rate, it would have taken us 287 days to generate the 208  $\mu$ s of sampling time observed in.<sup>103</sup> That is almost 10 months just to fold a small 20 amino acid protein using conventional MD simulations. Thus, much like the other physics-based methods, standard (unbiased) all-atom molecular simulations still present a threshold in sequence size for which they may no longer be a practical solution for predicting larger protein structures.

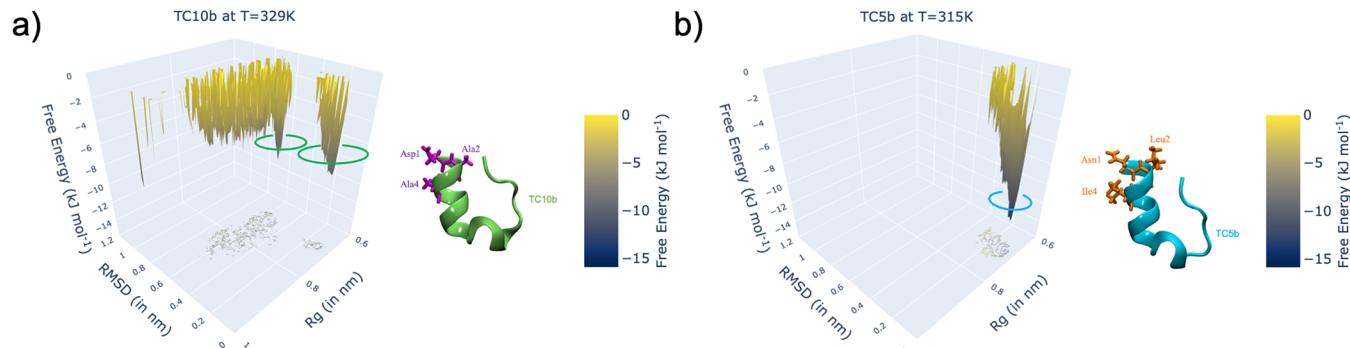
This gap between practical sampling times in simulation and the real folding times in experiments has been known for quite some time, and it is one of the main reasons enhanced sampling methods (such as replica exchange molecular dynamics REMD) were developed. REMD is an excellent tool for thermodynamic sampling of large conformational spaces, with relatively short time scales. Rather than having a single, long simulation of the system, several "replicas" of the same system are simulated at different temperatures (usually

room temperature, as well as above and below the melting points) and different random seeds for the initial velocities. In most implementations, the idea is that these temperature-dependent simulations can then undergo Monte Carlo swaps between neighboring replicas, preventing the lower temperature conformations from becoming "trapped" in a local minima. While the sampling time can be orders of magnitude less than running a single conventional simulation at fixed temperature, that is compensated by the number of simulations one has to run, which can regularly exceed a few dozen replicas even for a small 20 amino acid protein like Trp-Cage.<sup>104,105</sup> Because of this burden, new methods of REMD have been developed which can significantly reduce the number of replicas required to sample the same conformational space.<sup>106</sup>

A separate emerging field also leverages machine learning techniques to accelerate physics-based methods like MD. While a popular alternative for speed up of MD simulations has historically involved the use of coarse grain potentials, most recently those derived from neural networks<sup>107</sup> seemed to perform quite well, achieving similar results with the same set of proteins from,<sup>108</sup> much faster however. Among the limitations is the fact that long time scale all-atom simulations still need to be run beforehand for the neural network to train on, and even then these potentials are still not generalizable (they cannot be readily used for proteins largely different from the training set). On the other hand, there also appears to be much promise in using statistical information derived from machine learning to accelerate all-atom MD. An example of this is the novel approach been taken in,<sup>109</sup> where their all-atom MD method "ML x MELD x MD" leverages residue-residue contact information from machine learning algorithms to create spatial restraints and reparameterize the force field. The method appeared to not only dramatically accelerate the run time to fold proteins from an unfolded linear chain when, it also appears to be quite scalable as it successfully predicted structures up to 200 residues in length (much larger than the structures in refs 107,108).

**3.2. Mutations and Intrinsic Disorder.** One of the main advantages of the template-based deep learning methods like AlphaFold2 and RoseTTAFold is the sheer size of the structures they are able to produce. They are not limited to a few dozen residues as we've seen with the physics-based methods. Both AlphaFold2 and RoseTTAFold can readily produce models up to a couple thousand residues in length, in part due to the fact that their databases (PDB: <https://www.rcsb.org/>) include experimentally determined structures which span across this size range. Obviously the larger the sequence, the more MSAs and fragments are needed, and thus computing power. The upperbound is roughly between 1280 and 2700 amino acids, according to the European Bioinformatics Institute (<https://alphafold.ebi.ac.uk/faq>). Despite this impressive capability, success still depends on how accurate the models are. Specifically, the models are known to produce discrepancies when dealing with proteins with a) mutated sequences and b) intrinsically disordered regions. Here we discuss the effect of these two factors.

**3.2.1. Mutations.** While the ability for these programs to predict wildtype proteins containing mostly ordered secondary structures, i.e. localized structures that form based on interactions within the protein backbone, is unquestionable, significant errors have been reported when attempting to predict mutated variants. The discrepancies can be observed in both the 3D coordinates of the produced structure as well as



**Figure 4.** Free energy landscapes of (a) TC10b and (b) TC5b derived from molecular dynamics simulations at their melting points of 329 and 315 K, respectively. The folded NMR structures are quite similar (0.8 Å RMSD between them), but the free energy landscape is dramatically different, with TC10b demonstrating two possible folding pathways and their minima, as opposed to the more well-defined funnel and obvious global minimum in the case of TC5b. The three point mutations between these structures give rise to different physicochemical properties such as their melting temperatures, and accordingly different free energy landscapes.

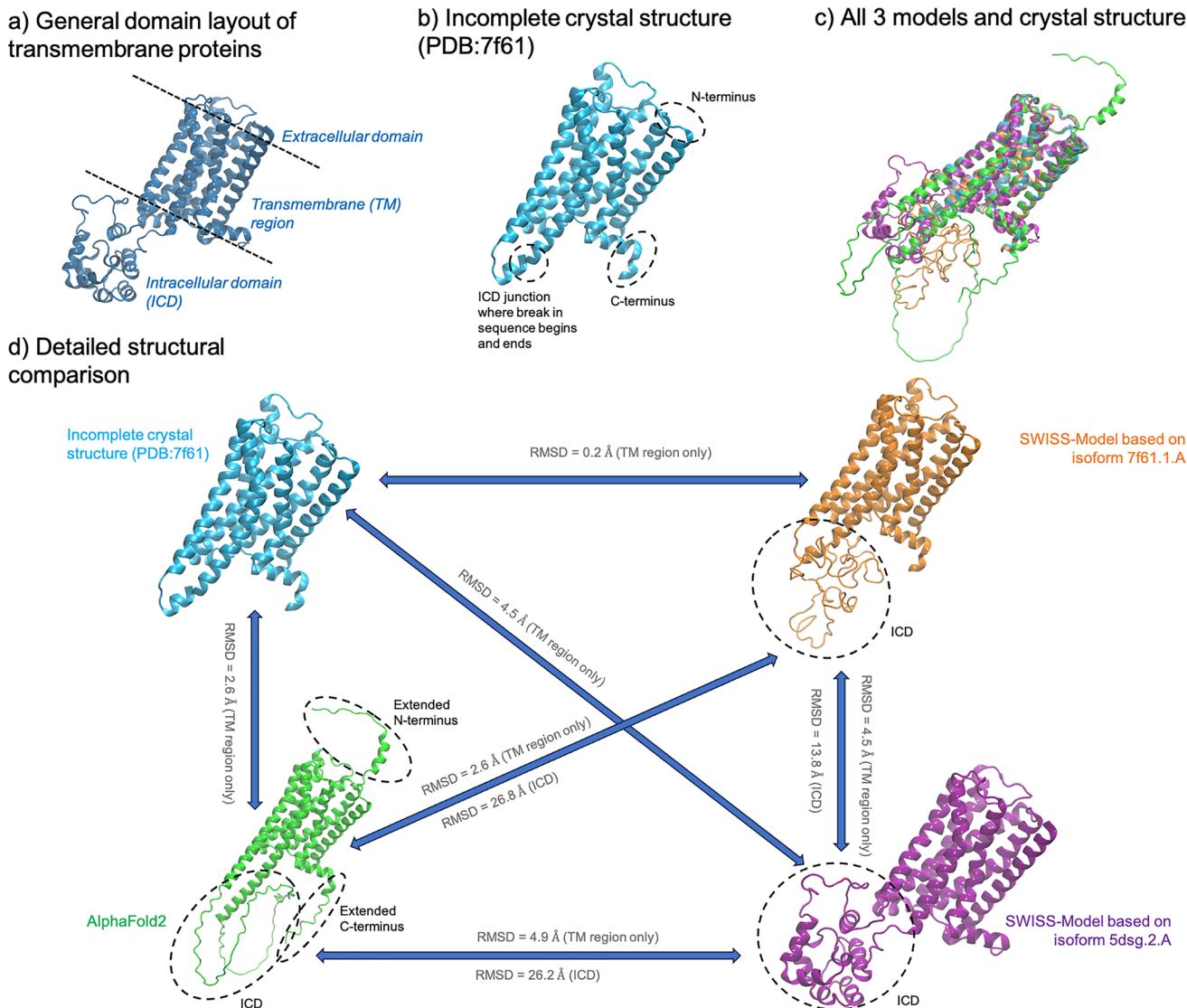
the predicted thermodynamic stability of the mutated residues relative to their confidence scores.<sup>110,111</sup> Mutations change the local chemical space, and thus the physics of the interacting residues. Even the slightest mutations between similar residues (for example valine to leucine) can have notable impacts on the local structure. Some mutations can be drastically different, such as going from a slender, flexible, charged residue like glutamic acid to a bulky hydrophobic side chain like tryptophan. Such cases can have profound effects, such as disrupting charge distributions beyond the local structure and often breaking relevant hydrogen bond networks vital to the protein's activity. In order for PSP methods to be successful, the algorithm needs to capture the consequences of mutations. From a global perspective, mutations can change the free energy landscape of a protein. This can sometimes lead to a more rugged landscape (or more smooth), perhaps introducing new local minima (see section 2.2). This can be problematic, as an algorithm's optimization protocol could get "stuck" in these regions, producing a conformation it thinks corresponds to the lowest energy, particularly if heuristic optimizers are involved. Of course exhaustive search algorithms could naturally avoid this, but they are not scalable beyond a certain size.

The Trp-cage proteins are 20 amino acid engineered "miniproteins" that have long been studied as ideal benchmarking candidates for physics-based PSP methods.<sup>103–105,112,113</sup> One of the original variants of this protein is TC10b (PDB: 2JOF), with sequence DAYAQWLKDGGPSSGRPPPS. A closely related variant is TC5b (PDB: 1L2Y), with sequence NLYIQWLKDGGPSSGRPPPS. The structures share a sequence identity of 85%, with mutations occurring at residues 1, 2, and 4. The mutations at 2 and 4 still preserve the hydrophobic chemical space, but result in more bulky side chains (leucine and isoleucine instead of alanines). The mutation at 1 is more pronounced, however, with a charged aspartate residue being substituted by an uncharged, but still polar, asparagine residue. Although the NMR structures are quite similar, with an RMSD of 0.8 Å between the two, the folding pathways and free energy landscapes can be noticeably different.

Proteins at their melting point are in a state of unfolding and refolding, much like a physical system transitioning between two states. One way to model what this looks like is through molecular dynamics simulations. As a proof of concept, we can explore the free energy landscape of the two variants of Trp-

Cage, TC10b and TC5b,<sup>114</sup> derived from conventional molecular dynamics simulations at their theoretical melting points (329 and 315 K, respectively). These simulations were performed with GROMACS2022<sup>115</sup> (as implemented in the Galaxy platform<sup>116</sup>) using the FF99SB force field<sup>117</sup> and the TIP3P water model,<sup>118</sup> for approximately 500 ns in the isothermal–isobaric (NPT) ensemble. Time-dependent RMSDs and radii of gyration ( $R_g$ ) of the  $\alpha$  carbon backbone were extracted to measure conformational changes over time, and then posteriorly analyzed using MD DaVis.<sup>119</sup> Figure 4 illustrates the resulting data, where the free energy (z-axis) is plotted as a function of RMSD (x-axis) and  $R_g$  (y-axis). The difference between both plots is quite clear. In Figure 4a, TC10b presents a much more rugged landscape, so much so that there appears to be almost two separate folding paths, leading to two distinct optimal conformations (circled in green). Both minima are relatively close, with the smaller funnel reaching  $-9.01$  kJ/mol and the larger funnel arriving at  $-9.4$  kJ/mol. Both optimal conformations appear to share nearly the same  $R_g$  (6.8 vs 6.9 Å), while having distinct RMSDs (0.8 vs 5 Å). TC5b on the other hand, presents a much more conservative free energy landscape (Figure 4b). There is a single global minima (circled in cyan) corresponding to an energy of  $-13.5$  kJ/mol. This optimal conformation shares the same radius of gyration as TC10b (6.9 Å) with a slightly different RMSD (0.7 Å). Thus, assuming the force fields are reproducing the experimental melting points accurately (which is beyond the scope of this work) this brief qualitative exercise highlights a central thesis of this perspective: even when mutated sequences still lead to similar experimentally solved structures, the free energy landscape and dynamic folding pathways can be quite different. Computational methods must therefore be aware and capable of capturing these unique interactions, while enumerating and sampling the entire conformational space. This is precisely what a quantum algorithm could be capable of doing, at scale.

**3.2.2. Intrinsically Disordered Regions.** While most proteins have ordered domains (with  $\alpha$ -helices,  $\beta$ -sheets, and  $\beta$ -barrels), many intrinsically disordered regions (IDRs) can exist throughout the structure. In fact, there are many proteins that are significantly disordered (containing at least 30% disordered residues), appropriately regarded as intrinsically disordered proteins (IDPs). The human proteome is estimated to be approximately 32% intrinsically disordered.<sup>120</sup> These regions serve a purpose as they are naturally flexible, allowing

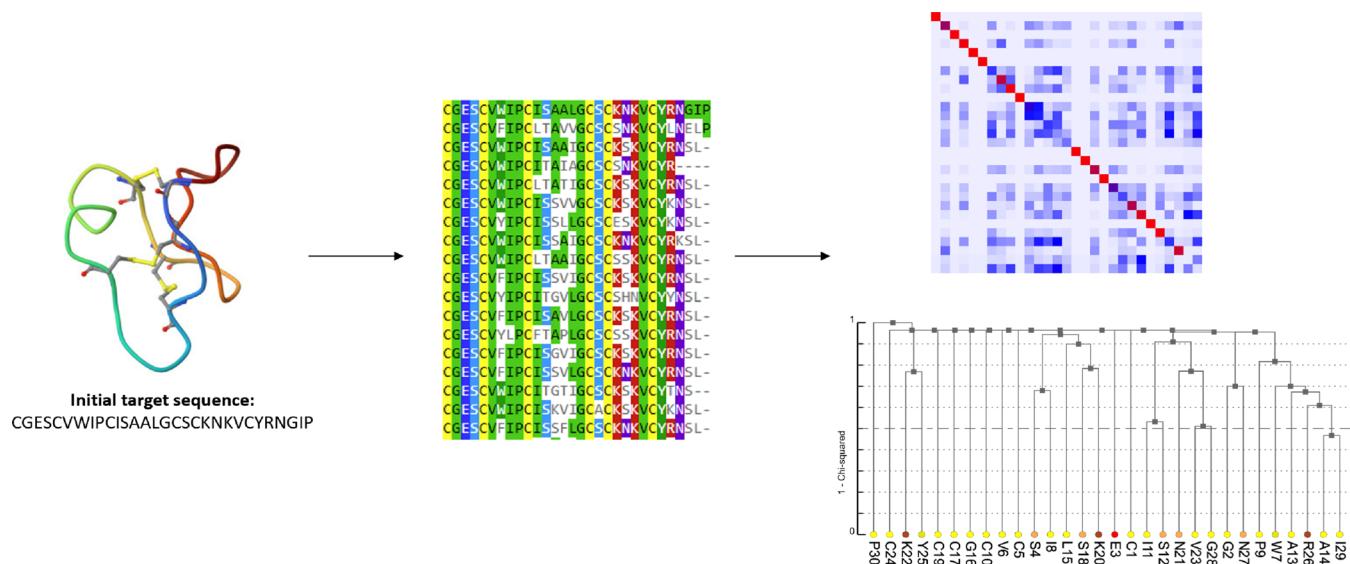


**Figure 5.** (a) The common anatomy of transmembrane proteins like the G-protein coupled receptor (GPCR) superfamily. (b) The incomplete experimentally determined crystal structure of H3HR, which is essentially missing the entire ICD. (c) Alignment and comparison between the experimental structure and computational models of H3HR, highlighting the excellent agreement in the TM region, but the significant discrepancies in the predicted ICD. (d) A detailed comparison of the structures reveals that the TM region is closely modeled with RMSDs of 0.2 to 4.9 Å, while the ICD varies drastically with RMSDs of 13.8 to 26.8 Å. The least accurate model is arguably AlphaFold2's, with a highly disordered ICD and an extended C- and N-terminus.

them to often serve as binding domains for other proteins or ligands. From a thermodynamic perspective, one can think of these regions as occupying a series of local minima in the free energy landscape, which are close in magnitude and thus equiprobable, with the global minimum often induced by a binding process. Despite their disorder, there is likely an ensemble average of structures, and it is important for PSP methods (and subsequent drug design efforts) to capture this accurately. The deep learning methods often fail at properly predicting these conformations.<sup>121</sup> What appears to be consistent though, at least in the case of AlphaFold2, is that these regions are usually predicted with low confidence scores, and for those IDRs which conditionally fold (when bound to a substrate), a recent study determined that it usually predicts a conformation closely related to that folded state.<sup>122</sup> This perhaps presents an unfavorable bias in the learning model, which could mislead researchers who are assuming the model is predicting the unbound protein. Another study found

interesting correlations between the AlphaFold2 model's confidence scores and the observed flexibility throughout the protein in molecular dynamics simulations.<sup>123</sup> Regions with lower confidence scores in the model were found to have higher flexibilities in the simulations, while those with higher scores usually corresponded to more stable secondary structures. This was quantified by measuring root-mean-square fluctuations (RMSFs), a common metric to assess flexibility in MD simulations and NMR ensembles. While this is an interesting correlation, this inference does not indicate that the predicted conformation of these IDRs is accurate, and researchers should be aware of this. To get a proper representation of these disordered regions, an ensemble average must be sampled. In order to do this, the physics at play must somehow be simulated, which is something most template-based methods remain arguably agnostic to.<sup>6</sup>

A perfect example of the importance of representing the disordered regions correctly occurs with transmembrane



**Figure 6.** A typical MSA generation pipeline to extract statistical summaries. After the sequences are aligned, the covariance matrix and the phylogenetic tree are constructed to capture coevolutionary information.

proteins. As the name implies, these proteins are embedded in cell membranes. They are key players in promoting cellular homeostasis, by maintaining membrane integrity, aiding in cell binding and adhesion, active and passive transport of substrates in and out of the cell, and many other vital functions. They come in many shapes and forms. One important superfamily of transmembrane proteins is the G-protein coupled receptors (GPCRs). The basic anatomy of these proteins is outlined in Figure 5a. There are three main domains: the extracellular domain, the transmembrane (TM) region, and the intracellular domain (ICD). The intracellular and extracellular domains interact with molecules in and out of the cell, respectively. As such, flexible loop regions tend to be found here, in addition to some segments of ordered secondary structures. The transmembrane region on the other hand is almost entirely made up of a bundle (seven in this case) of  $\alpha$ -helices forming a channel through the lipid bilayer of the cell membrane, allowing substrates to enter and leave the cell. The structures in Figure 5 include the incomplete X-ray crystal structure (b) and three computational models (c-d, obtained from the SWISS-MODEL repository; <https://swissmodel.expasy.org/repository/uniprot/Q9Y5N1>) of the H3 human histamine receptor (H3HR). Its primary function is the binding and release of the neurotransmitter histamine. As such, it is vital for normal neurological function and its impairment is associated with a number neurodegenerative disorders, including Multiple Sclerosis (MS), Alzheimer's disease (AD) and Parkinson's disease (PD).<sup>124</sup> The crystal structure (PDB: 7f61) is missing the entire intracellular domain, approximately 105 residues. Transmembrane proteins like H3HR and other GPCRs are generally difficult to determine experimentally for several reasons. Part of this is due to the fact that these proteins exist in different media—the intracellular and extracellular domains exist in aqueous hydrophilic media (cytosol and plasma) while the transmembrane region is trapped in a hydrophobic environment surrounded by lipids. Experimentally, it is very difficult to provide a solvent for proteins that can emulate both environments. Thus, this class of proteins relies heavily on some sort of homology modeling for researchers to further investigate them. The ability of the deep learning methods to

accurately fill in the gaps here has been investigated before.<sup>125,126</sup> The work<sup>126</sup> found that AlphaFold2 performed well, but most of their structures appeared to mainly involve the ordered, transmembrane alpha helical bundles. Similarly, the work of<sup>125</sup> found that AlphaFold2 and RoseTTAFold performed well in predicting the TM regions, but there were significant discrepancies when modeling the surrounding loop regions. This is precisely what we observe in our analysis of the H3HR models. The AlphaFold2 model is colored in green, a SWISS-MODEL based on isoform 7f61.1.A (essentially modeled without a template for the ICD) colored in orange, and another SWISS-Model based on isoform 5dsg.2.A (using the M4 muscarinic receptor as a template for all domains) colored in purple. Protein isoforms are similar proteins (or variants) arising from a common gene (or family of genes). The variations in the protein sequence, and thus corresponding 3D structure, are due to differences in how the corresponding genes are transcribed and translated. In other words, different codons may be selected to encode the protein sequences, leading to unique isoforms. When generating homology models using templates based on difference isoforms, some of these structural differences may influence the resulting model. The SWISS-MODEL server scores models based on a composite scoring function, QMEANDisCo. During our studies, these 2 isoforms yielded models with the highest QMEANDisCo scores (0.70 for 7f61.1.A and 0.63 for 5dsg.2.A), among the models for H3HR in the repository. The TM regions align quite well in all 4 structures as seen in Figure 5c, and in all cases the RMSDs for this region is under 5.0 Å (Figure 5d). The differences in the ICD structures were much more pronounced between the models. The largest RMSD is observed between AlphaFold2 and both SWISS-MODEL (averaging 26.5 Å). The RMSD in the ICD between the SWISS-MODEL is still significant (13.8 Å), although about half as much as AlphaFold2's, where there appears to be virtually no order in the structure. In reality, the ICDs of these proteins are neither fully ordered nor fully intrinsically disordered, but rather a combination of segments of both, much like what is observed in the case of the purple isoform in Figure 5d. Another interesting fact worth pointing out is that,

for reasons which remain unclear, the AlphaFold2 model (<https://alphafold.ebi.ac.uk/entry/Q9YSN1>) is noticeably larger. The model includes an additional 26 residues at the N-terminus and another 14 residues at the C-terminus. With this, the model has 445 residues total, while the FASTA sequence for the H3HR crystal structure (PDB:7f61) contains only 407 residues, including those missing from the ICD.

The ability to accurately generate full length *in silico* models for transmembrane proteins is both challenging and crucial. They remain among the most experimentally difficult to determine and are accordingly the most expensive human proteins to solve via crystallographic methods (with an average cost of \$2.5 M per protein, at a 10% success rate<sup>127</sup>). Yet, solving these structures holds tremendous therapeutic value. As a clear example, approximately 35% of FDA approved drugs target GPCRs.<sup>128</sup>

**3.3. Multiple Sequence Alignment Availability and Diversity.** Multiple Sequence Alignment (MSA) is a fundamental technique in bioinformatics, crucial for understanding protein sequence relationships and evolutionary patterns. Essentially, an MSA is a collection of amino acid sequences derived from a single query sequence, illustrating how similar sequences are to the original one. These alignments encode significant evolutionary insights utilized across various domains like Protein Structure Prediction (PSP), protein classification, and protein design (See Figure 6). As a result, in the absence or the limited availability of MSAs, deep learning methods struggle to predict the protein structures with high accuracy.

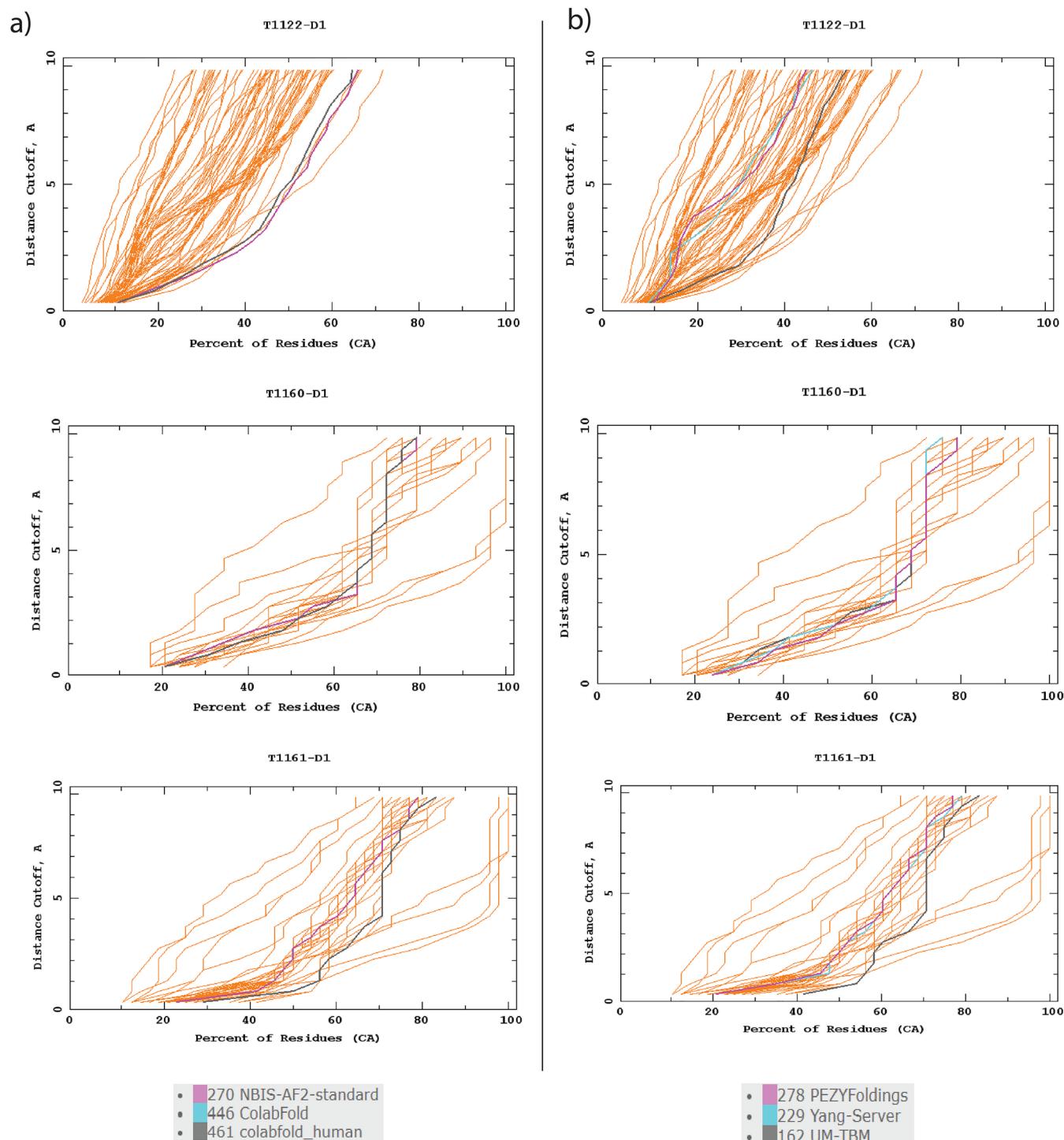
The utilization of MSA-based information has a long-standing history in PSP. Early methodologies<sup>129,130</sup> laid the groundwork for employing covariance methods to identify coevolving residue pairs. By analyzing substitution patterns among sequences from different organisms, these methods investigated the correlations essential for understanding protein–protein interactions and spatial proximity in structure prediction.<sup>131,132</sup> The underlying concept hinges on the observation that coevolving residues tend to maintain close proximity in the final native 3D structure. Recent advancements, in particular in machine learning methods, have seen statistical analyses extracting coevolutionary insights, which are then integrated into deep learning architectures for 3D conformation prediction.<sup>3,4,133,134</sup> Notably, AlphaFold2's introduction of the "Evoformer," a transformer block capable of capturing long-range dependencies among residues, marked a significant breakthrough, enabling highly accurate structure predictions.<sup>3</sup> In tandem with the rise of large language models, Protein Language Models (PLMs) are also emerging as significant tools in PSP<sup>135–137</sup> that utilize MSA information. MSA Transformer, for instance, leverages a data set of 26 million MSAs and employs row and column attention mechanisms in its transformer-based architecture.<sup>138</sup> These unsupervised models harness MSA patterns to predict protein conformations.

Predicting protein structure from a single amino acid sequence remains challenging, particularly for deep learning methods. The availability of Multiple Sequence Alignments (MSAs) significantly influences the performance and applicability of these approaches. Studies show a substantial decrease in the prediction accuracy for shallow MSAs with median depths below 30, as evidenced by low predicted local distance difference test (pLDDT) scores.<sup>3,123,139</sup> Orphan proteins, which lack homologous sequences and comprise about 20%

of all metagenomic protein sequences,<sup>140</sup> exemplify cases where limited MSAs hinder the utilization of coevolutionary information. Recent advancements in Protein Language Models (PLMs) demonstrate improved accuracy in predicting orphan protein structures compared to AlphaFold2 and RosettaFold, leveraging linguistic processing of MSA summaries that are less reliant on alignment.<sup>140–142</sup> Furthermore, the computational challenge of accessing diverse and deep MSAs underscores the need for alternative MSA-free computational methods capable of reliably predicting protein structures with high accuracy, given the scarcity of publicly available precomputed MSA databases like those used in AlphaFold2.<sup>143</sup>

**3.4. A Multidimensional Perspective on Computational Hardness.** Protein sequence length does not singularly dictate prediction accuracy<sup>144</sup> but rather, a combination of factors such as mutations, disorder, and MSA depth collectively influences the ruggedness of the energy landscape. This complexity necessitates a broader scope of information beyond sequence length for a quantum solution to protein conformation problems. The fitness landscape in protein sequences, a critical metric for mutation analysis, directly influences biomolecular evolution by determining the Colony-Forming Unit (CFU) and approximating free energy.<sup>145</sup> As mutations occur, local peaks emerge within the fitness landscape, leading to changes in the energy landscape, as depicted in Figure 4, introducing uncertainty regarding its structure and potentially enabling quantum advantage. In comparison to structured proteins, Intrinsically Disordered Proteins (IDPs) possess shallow, rugged energy landscapes housing multiple local minima, especially in hypervariable regions.<sup>146</sup> This stands in contrast to ordered regions like the guanosine triphosphate (GTP) binding domains, which exhibit a deep energy minimum. IDPs' sensitivity to environmental changes, including post-translational modifications, can significantly alter their energy landscapes by manipulating local minima and barriers, thereby creating uncertainty and ruggedness conducive to leveraging quantum advantages for conformation predictions. Additionally, studies on kinase families have shown that unregularized mutational statistics inference struggles to converge for shallow MSAs, indicating an increasingly rugged and ill-conditioned energy landscape.<sup>147</sup> The results consistently highlight that reduced MSA depths result in Potts models reflecting more rugged energy landscapes, affirming the potential for quantum advantage in solving protein prediction problems affected by such increased ruggedness and uncertainty.

In this section, we present our view on the hardness of a PSP task in terms of three main parameters: the protein size (sequence length), the number of point mutations and the MSA depth. We discuss a general framework for identifying proteins for a structure prediction task, in particular using physics-based quantum algorithms, which have a high potential for yielding competitive results compared to state-of-the-art deep learning, MSA-based methods. We expect that this will be beneficial for both quantum researchers who plan to benchmark their novel quantum algorithms against state-of-the-art results, and also for biomedical researchers who wish to explore quantum capabilities to enhance their structure prediction results. Our perspective relies on a careful analysis of the current landscape of the PSP research and the most recent Critical Assessment of Structure Prediction (CASP) outcomes, a biennial event where the latest PSP methods are



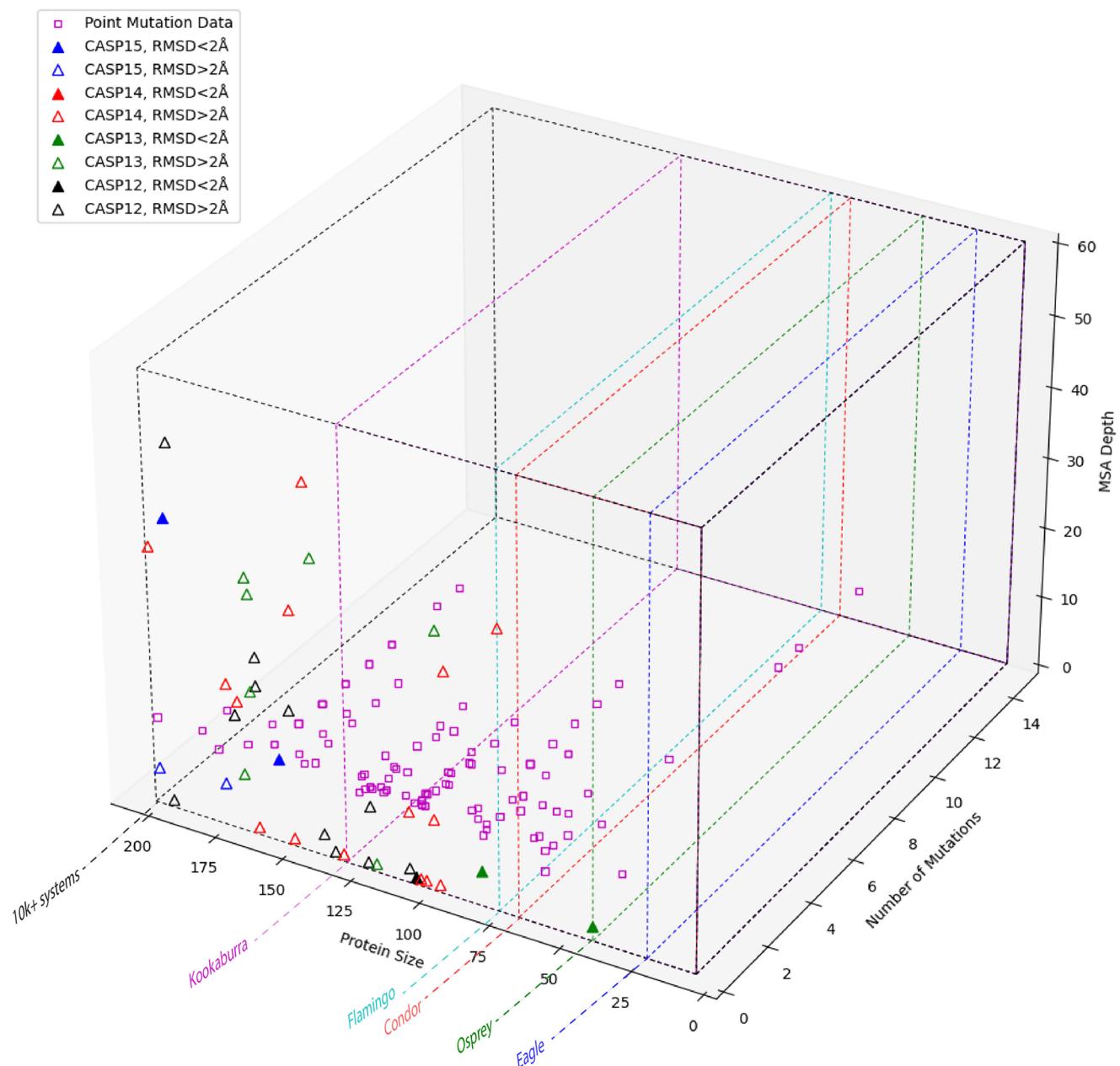
**Figure 7.** The plots highlight the performance of all groups for the target proteins, however we further distinguish: (a) Performance of versions of AlphaFold2 that participated in CASP15 for the three targets. (b) Performance of top three ranked groups with respect to overall z-scores in CASP15 for the three targets. The graphs show the percentage of residues in the target protein that are under a certain distance cutoff. It is easy to see that for the target protein T1122, no group was able to predict all residues under 10 Å. For T1160 and T1161, which exhibit point mutations, deep learning-based methods used by the groups listed has 80% of residues under 10 Å. In general, sharper increasing curves indicate poorer predictions. Images generated from CASP15 official Web site from the results.

showcased and their performance is benchmarked.<sup>148</sup> Our perspective is meant to serve as a general guideline for maximizing the best possible outcome in the regime where the best deep-learning based methods are known to perform poorly. While we identify this subspace of the problem space based on a particular quantum algorithm from<sup>149</sup> and the

actual boundaries of the subspace may change depending on the method adopted, we believe that the general framework provided persists if one wants to use an *ab initio* method for PSP.

In the previous sections, we discussed in detail three sources of hardness for the PSP problem and how they relate to the

## Problem Space for Quantum Utility



**Figure 8.** Subspace of PSP problems partitioned with different color rectangular boxes. This is in parallel with the IBM Quantum hardware roadmap as system size increases in terms of qubits. The nested rectangular boxes represent the subset of proteins where deep learning based methods are known to perform poorly (data point markers with no facecolors), hence other ab initio methods, including quantum algorithms, can potentially yield better predictions. CASP data is obtained from,<sup>151</sup> however sequence length data is added from CASP website directly. For each protein on the plot, average RMSD from the top 10% of the groups is calculated and added to the data. Point mutation data is obtained from,<sup>152</sup> and the MSA depth ( $N_{\text{eff}}$ ) values are calculated using HHblits and HHpred tools.<sup>153</sup> Average RMSD values are not calculated for point mutation data set since they are not CASP targets. Clearly, there are many more proteins within these boxes. Our goal is to show that there are nontrivial, high-value targets for each of these regimes. The alignment with the IBM Quantum Roadmap is more of a symbolic representation. While we estimated qubit costs for the problem instances, one needs to perform a rigorous resource estimation for a concrete representation. This is beyond the scope of this work. We assume that for any range of qubit number, the quantum computer is able to perform a reasonable number of gate operations under reasonable time frame.

energy landscape, in particular increasing complexity and ruggedness. Some of the target proteins from the most recent CASP15 provide further evidence for these hardness metrics. As discussed earlier, it is known that in the absence of deep and variable MSAs, DL methods fail to produce accurate predictions. In particular, it is reported in<sup>3</sup> that if the MSA depth is less than 30 sequences, AlphaFold2 prediction

accuracy decreases significantly. Moreover, it is also mentioned that if the MSA depth exceeds 100 sequences, the algorithm has diminishing returns. Hence, one can conclude that for any protein sequence for which the generated MSA has depth less than 30, these deep learning based methods will yield low accuracy predictions. When measuring the performance of the predictions, there are two common metrics CASP utilizes;

Global Distance Test - Total Score (GDT\_TS score) and Template Modeling score (TM score) (For detailed description, see <https://predictioncenter.org/casp15/doc/help.html>). Two target proteins T1122 and T1131 in CASP15 are examples of orphan proteins where a reliable MSA cannot be generated. In particular, none of the participants were able to produce a good prediction in the absence of MSAs. The best GDT\_TS score is less than 40 and TM scores are around 0.5. A GDT\_TS score around 50 is interpreted to roughly predict the overall topology of the structure correctly, and similar interpretations for TM score as well. One can see for most predictions of T1122, only 60% of residues are under 10 Å. The other target orphan protein T1131 has similar results for all participants with the best GDT\_TS score around 25 (this is usually deemed as good as a random guess) and the best TM score in 0.3s (See Figure 7).

Similarly, predicting the impact of the mutation is a difficult task. As detailed in,<sup>150</sup> detecting how the mutations alter the conformation was a big challenge in CASP15. In particular, two target proteins T1160 and T1161 differ by mutations at five residues. Both proteins were categorized as “easy” since templates were provided and they are relatively short proteins with 48 amino acids. The top performing groups reached TM scores in 0.3 and 0.5s for the two target proteins, respectively. Best GDT\_TS scores are 72.41 and 72.4 for the two proteins, while most groups achieved GDT\_TS scores in 50s. This is another example of how disruptive point mutations can be for a structure prediction task, in particular using a technique that requires reliable MSA-based information.

Based on these observations, if one wants to consider the parameters that can make a protein structure task difficult, we propose that along with the protein size, MSA depth and the number of (point) mutations play a significant role. A detailed analysis is presented in the previous sections in terms of how they increase hardness and introduce more complexity in the energy landscape. In particular, our perspective is helpful if one chooses to use a method that does not rely on MSA-based information, such as physics-based or free modeling methods. In this work, we use the *ab initio* quantum algorithm from<sup>149</sup> to identify the subspace of the problem space where one can expect to produce predictions that are potentially more accurate than state-of-the-art, deep learning based methods.

The algorithm from,<sup>149</sup> to our knowledge, is the best scaling quantum algorithm that constructs a 3D  $\text{Ca}$ -only backbone of the protein from a single amino acid sequence. In this coarse-grained tetrahedral lattice model, a variational quantum algorithm is adopted to find the lowest energy conformation of the protein based on precalculated residue–residue interaction energies. We present a sequence of nested subspaces based on these three sources of hardness and the resource estimates for this particular quantum algorithm (See Section 5 for further details).

Using the dense encoding of an amino acid sequence in,<sup>149</sup> we observe that we can target proteins up to 22 amino acids in length using IBM’s 127-qubit devices and extend this to 41 amino acids on IBM Osprey, a 433-qubit device. We have also estimated how this subspace gets larger as we have access to larger devices, aligned with IBM Quantum’s hardware roadmap (For further information see <https://www.ibm.com/quantum/roadmap>). Since we have established that proteins with shallow MSAs are good candidates for this algorithm,<sup>149</sup> we target the proteins with MSA depth less than 60 as shown in Figure 8. MSA depth is measured by the number of effective sequences

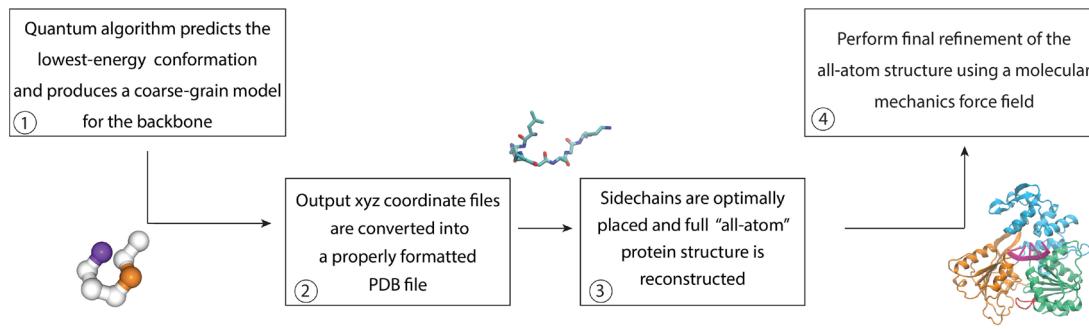
$N_{\text{eff}}$  similar to,<sup>3</sup> and the previous CASP data in Figure 8 is obtained from the public repository<sup>151</sup>, with the inclusion of some extra data curated by us. Since the target protein can exhibit any number of (point) mutations not exceeding the overall residue count, the protein size potentially determines the boundaries for the other axis representing the number of point mutations. However, the data we curated contains maximum 15 point mutations, so the axis boundaries are adjusted accordingly for better visualization. As a result, we argue that any target protein contained in the rectangular box regions in Figure 8 is potentially a suitable candidate for a physics-based quantum algorithm to benchmark the performance and yield competitive results compared to state-of-the-art, deep learning template-based algorithms.

We make a few important remarks and observations before concluding this section. Most importantly, while the actual numerical boundaries of the subsets provided in Figure 8 depend on the quantum algorithm we adopt, the general axes that describe computational hardness and the underlying idea of “going wide instead of going deep” provide a systematic way of benchmarking *ab initio* quantum algorithms to maximize the outcome. As quantum hardware becomes more capable, we expect that this subspace will scale accordingly and include more potential target proteins. Moreover, with the current boundaries provided, this is a nontrivial subset of all proteins. There are 1321 protein structures in PDB that contain 22 or less residues. Similarly, almost 3000 protein sequences contain 41 or less amino acids in PDB database and if we query the condition that the sequences exhibit at least one mutation, we see that there are 509 protein sequences. (All data collected from PDB Web site; note that some of these sequences can be part of a larger protein sequence. Since the proteins from PDB are already being utilized for training deep learning based algorithms, we note that these numerical values from PDB are meant to represent the variety of proteins that are available at these smaller scales.)

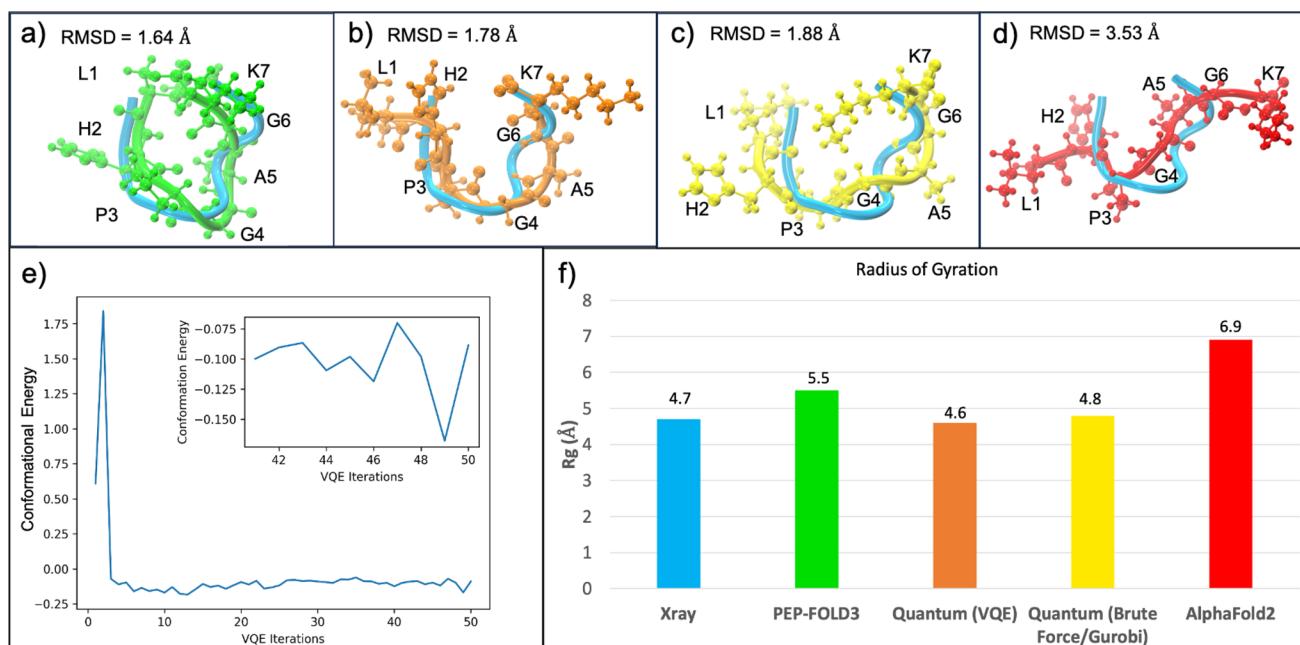
Benchmarking quantum algorithms and techniques is an important research area within the quantum community. Our perspective provides a novel and systematic approach that aims to maximize the utility of adopting a physics-based quantum algorithm and benchmarking the outcome. As a result, we believe that whether developing a quantum algorithm or exploring existing quantum algorithms for PSP tasks, selecting a target protein in the regions provided in Figure 8 maximizes the utility and creates the best possible outcome.

#### 4. A QUANTUM-CLASSICAL HYBRID WORKFLOW FOR PROTEIN STRUCTURE PREDICTION

While the promise of general quantum advantage comes with strong theoretical foundation, quantum computers are not expected to replace classical computers in every task. A scientific workflow may have a component that is computationally intensive and could be handled by a quantum algorithm while the remaining steps are treated on classical hardware. For example, simulating a catalytic reaction in the active site of an enzyme with a quantum Hamiltonian while the rest of the domains are modeled using classical algorithms (such as using conventional molecular mechanics force fields).<sup>154</sup> Initially, we envision quantum algorithms for PSP to be designed as part of hybrid quantum-classical workflows. Lattice models can help simplify the PSP problem by representing proteins as a chain of beads on a lattice. Generally speaking, these models employ an energy function to score and



**Figure 9.** A schematic representation of the workflow used in this study. The most computationally demanding part, finding a coarse-grain representation of the lowest energy conformation of the protein structure, is performed on a quantum computer. The following steps are handled classically to convert the output into a desired format and postprocess to construct the full structure, while preserving the quantum algorithm's originally predicted backbone geometry from the coarse grain model. A final refinement of the all-atom structure is then performed through further energy minimization using a molecular mechanics force field. This last step allows the protein to potentially reach an even more optimal configuration as the atoms and bonds are no longer constrained to the four turns of the original lattice structure.



**Figure 10.** An initial validation of the workflow with the Zika virus helicase P-loop (LHPGAGK). In all cases, the coordinates from the experimental crystal structure are colored in cyan. The lowest RMSD relative to the experimental structure is achieved by (a) PEP-FOLD3, followed very closely by (b) the quantum algorithm executed on IBM\_Cleveland, and then by (c) the problem Hamiltonian solved classically by brute force and the QUBO Hamiltonian solved by a classical mixed-integer linear solver (both approaches yielded the same solution). The least accurate model was produced by (d) AlphaFold2. This is observed by the relative RMSD values in each case, as well as (f) the measured radius of gyration  $R_g$ . The conformational energy plot in (e) appears to demonstrate that VQE begins to continuously sample conformations around the basin after a handful of VQE iterations. Note: while these residues are numbered 1–7 here, they correspond to 194–200 in the crystal structure, PDB: 5gj.

rank the enumerated conformers, with a search algorithm used afterward to identify the lowest energy conformation. The energy functions may range in sophistication, from a basic hydrophobic-polar model (where the alphabet is reduced to a binary representation, allowing for just three possible interaction types) to more detailed models considering various physical forces including the hydrophobic effect, electrostatic interactions,  $\pi-\pi$  stacking, hydrogen bonding, induced dipole moments, and side-chain packing. Statistical or knowledge-based potentials are effective scoring functions derived from a statistical analysis of pairing frequencies between amino acids in a database of protein structures. These empirically derived potentials have been widely used in protein–ligand and

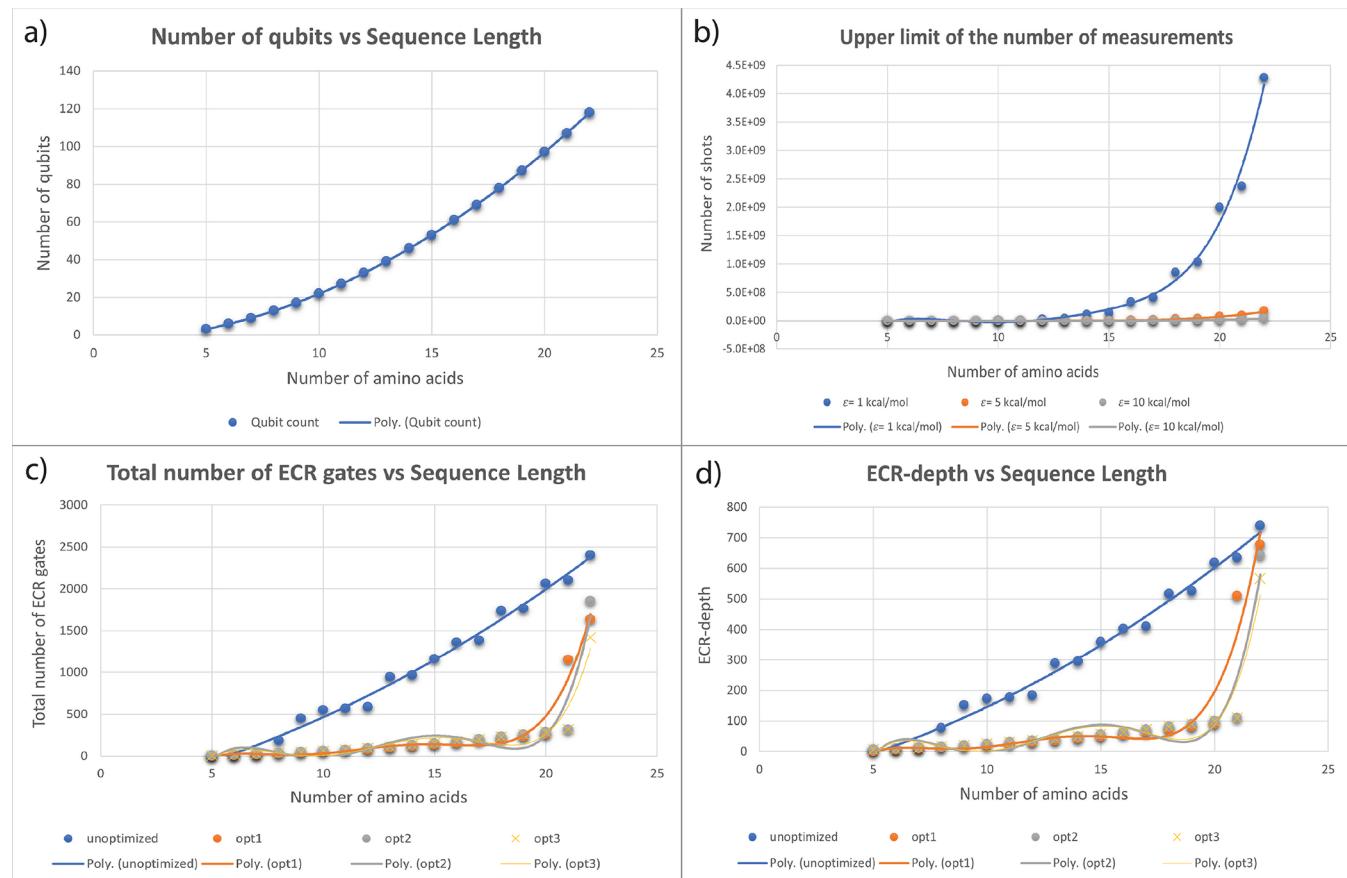
protein–protein docking, protein structure prediction, comparing native folds vs decoys, and modeling protein folding processes. Within the young field of quantum algorithms for PSP, the Miyazawa-Jernigan<sup>155</sup> statistical potentials have become the gold standard for determining interaction energies and scoring the conformers. They have the advantage of providing unique pairwise interaction energies for the full alphabet of amino acids, as opposed to a simple binary representation of all 20 amino acids being either hydrophobic or polar (as in the HP model). They can be readily used for coarse grain modeling of proteins, making them computationally efficient solutions, which is necessary as we remain limited to the available quantum resources of the near-term quantum

computers of today. In any case, the objective should always be to effectively reproduce the free energy associated with every possible type of interaction. The more detailed the energy function is and the more granular the lattice model, the greater the number of physically reasonable enumerated conformations. The challenge for the search algorithm then becomes the ability to effectively traverse this enormous conformational space containing all possible conformations, which remains NP-hard even at a coarse grain resolution on a 2D lattice.

The series and order of steps involved in this proposed workflow (see Figure 9, and section 5.1 of the Supporting Information for a more detailed discussion of the workflow) are not uncommon in *ab initio* PSP methods. For example, the Rosetta algorithm<sup>156</sup> first creates small coarse grain fragments using a knowledge-based potential (the Rosetta energy function), much like what is done in the first step of this workflow (using the Mizayawa-Jernigan potentials instead<sup>155</sup>). These fragments are then assembled using a set of reference PDB structures until the full size of the sequence is built. In the later stages, the Rosetta “all-atom” energy function<sup>157</sup> is then employed which refines the all-atom models. A clustering analysis is performed on the resulting models from each iteration traversing the search space, with the average structure from the largest cluster becoming the chosen solution. The fact that templates are used in this process is part of the reason why some members of the community have debated as to whether or not Rosetta is in fact a true *ab initio* method, since it is mainly the fragments that are constructed exclusively with an energy function. By now, we are of course aware that efficiency plummets when trying to do this beyond a few dozen amino acids. Nonetheless, incorporating the physics this way has clearly allowed the method to preserve accuracy while handling large sequences, an advantage over most other template-based methods. With the possible advantage quantum computing brings in handling large search spaces, fragment assembly may not be required, allowing for the enumeration and sampling of large conformers in a single pass of the algorithm. This is what our proposed workflow attempts to do—build the entire sequence in one step, sample the energies of all possible conformers, and finally refine the optimal solution classically.

We tested steps 1–3 of this workflow on a small, but highly relevant seven amino acid fragment (LHPGAGK)—the “P-loop” of the Zika virus NS3 helicase protein. See section 3 of the Supplementary Information for more discussion about the relevance of this protein in drug discovery. We chose the P-loop as the “toy model” for this workflow and present our results as an initial proof of concept. A comparison is made in Figure 10 between the results from a) PEP-FOLD3, b) the quantum algorithm executed on IBM\_Cleveland and solved by VQE, c) the quantum algorithm’s Ising Hamiltonian (a classical cost function originally defined in<sup>149</sup>) solved classically (by an exhaustive brute force search as well as Gurobi<sup>158</sup>), and d) AlphaFold2. In every case, the loop’s experimental structure (extracted from the full crystal structure, PDB: 5gjb) is colored in cyan, while the models are colored in green, orange, yellow, and red, respectively. The backbones were aligned to the experimental structure and RMSDs were measured in each case using VMD.<sup>159</sup> In terms of RMSD, PEP-FOLD3 produced the best model, with an RMSD of 1.64 Å. However, this was only a marginal improvement over the quantum algorithm’s results from the IBM\_Cleveland quantum device, with an RMSD of 1.78 Å, with respect to the crystal structure. During sampling, the

conformational energy initially rises sharply but begins to continuously hover around the basin shortly after a handful of VQE iterations (Figure 10e). The Ising Hamiltonian produced by the quantum algorithm contains higher order polynomial terms (cubic and above), which can often increase the complexity of optimization in general, and more so on classical computers. We have also solved the same Hamiltonian classically to compare the performance. While the brute force search was performed on the native higher order Hamiltonian, it was also quadratized and converted into a QUBO (quadratic unoptimized binary optimization) Hamiltonian using a method developed by Fujitsu,<sup>160</sup> and posteriorly solved with Gurobi. The Hamiltonian term corresponding to the optimal solution, as determined by both Gurobi and the brute force search, yielded an RMSD of 1.88 Å, slightly higher than VQE’s solution from the execution on quantum hardware (Figure 10c). AlphaFold2 produced the least accurate model (Figure 10d), with an RMSD of 3.53 Å, nearly twice as much as that predicted by the quantum algorithm. The radius of gyration was also measured in each case, and the same trend can be observed (Figure 10e). In this regard, the best results came from the quantum algorithm. The average radius of gyration between the solution from IBM\_Cleveland (4.6 Å) and the classical solvers (4.8 Å) is the same as the experimental structure (4.7 Å). PEP-FOLD3’s solution presents a larger radius of gyration (5.5 Å). AlphaFold2’s prediction is noticeably greater at 6.9 Å, indicating that the model is much more extended than the experimental structure and the quantum algorithm’s solutions. There are a few interactions in the models that lead to the observed differences in structure. The PEP-FOLD3 model places residues L1 and K7 close to each other, which is primarily driven by the electrostatic attraction of the backbone N-terminal amine group (−NH<sub>3</sub><sup>+</sup>) of L1 and the C-terminal carboxylate (−COO<sup>−</sup>) of K7 (Figure 10a.) The experimental structure also presents a close interaction between these two residues, which contributes to the PEP-FOLD3 model’s accuracy, although in the crystal structure it is due to hydrophobic contacts between L1’s isopropyl side chain and the aliphatic carbons of K7’s side chain. In the case of the quantum algorithm’s results from IBM\_Cleveland (Figure 10b), we observe that there is a rotation at the  $\alpha$  carbon of H2, which orients the adjacent L1 residue perpendicularly with respect to its experimental conformation. This simultaneously results in an attractive electrostatic interaction, a hydrogen bond between the protonated Histidine side chain and the backbone carboxylate group of K7. It is possible that this rotation at H2 could be influential to the subsequent dihedrals of the backbone as well, which leads to a near perfect alignment of P3 and G4. Relative to the experimental structure, we observe that K7 is also rotated, with the side chain pointing away towards the right, away from the loop. This contributes to the aforementioned hydrogen bond with H2, but could also influence the preceding backbone dihedrals, which result in an excellent alignment of G6. Although there is a difference in the conformation of A5, it is not severe, with a displacement of 2.62 Å in the  $\alpha$  carbon, with respect to the experimental structure. In the case of the classical solver’s results (Figure 10c), the sequence of turns yields a rotation of the alpha carbons of H2 and K7. The H2 side chain adopts a conformation aiming outward, away from the loop, while the K7 side chain is rotated inward, closely resembling the experimental conformation. This results in a hydrogen bond



**Figure 11.** (a) The total number of qubits scale quadratically as the protein size increases. We consider both configuration and interaction qubits to encode a given amino acid sequence, and analyze the scaling. (b) We estimate an upper bound for the number of measurements needed to predict a protein structure within a fixed energy error margin using the work from.<sup>162</sup> The energy unit is converted to kcal/mol from Hartree for consistency with our workflow. See Section 5.4 of the Supporting Information for mathematical formula to calculate the upper bound. The three plots show different upper limits to predict a protein structure within  $\epsilon = 1, 5, 10$  kcal/mol range of the lowest energy conformation. The upper bound is not known to be tight. We further expect the empirically sufficient number of measurements to be significantly lower than this upper bound. (c) Shows the total number of ECR gates in the circuit as the protein size increases for different optimization levels. Qiskit transpiler allows four levels of optimization which are defined in.<sup>163</sup> (d) Shows the ECR-depth as a function of protein size.

between the positively charged side chain amine group of K7 and the backbone carbonyl group of L1. Although this helps preserve the overall “U” shape of the loop, it displaces the alignment of P3, G4, and A5 by 2.62, 2.04, and 3.03 Å, respectively. In the case of the AlphaFold2 structure (Figure 10d), there appears to be no obvious intramolecular interaction that would stabilize the loop, as observed in the previous 3 models. This is consistent with the fact that the structure is largely elongated, alternatively curving along a single axis. A reasonable alignment is found mostly at P3, with large displacements in the other 6 residues, which contribute to the structure having both the largest global RMSD and radius of gyration. This brief analysis informs that the models’ accuracy is dependent on the rotations (or turns along a lattice, in the case of a quantum algorithm) of residues L1, H2, and K7. The backbone alignment of the remaining residues seems to be influenced by the intramolecular interactions formed based on the conformations of these 3 residues.

These results highlight a few important things: (1) Both the tetrahedral lattice model and the Miyazawa-Jernigan potentials<sup>155</sup> employed in the algorithm<sup>149</sup> could be sophisticated enough for applications in *ab initio* PSP. (2) Even though VQE may not always find the absolute lower bound in the energy, it could be well suited for sampling near-optimal low energy

conformations that may even produce more accurate results than an exhaustive search when the measure of success is based on structural similarity to a ground truth experimental solution (VQE’s heuristic solution was structurally more accurate than the ground truth solution found by the exhaustive search). It is well-known that there could be several local minima that are quite close in magnitude and sign, yet be structurally distinct. An example of this can be seen in Figure 4a, where the bottom of both funnels correspond to two distinct conformations similar in energy, but only one of them is closest to the experimental structure. (3) Classically rebuilding the all-atom structure can still preserve the originally predicted geometry by the quantum algorithm. (4) A quantum-classical hybrid workflow like this could potentially be used as a tool in tandem with template-based programs to help predict binding domains characterized by loops.

## 5. RESOURCE ESTIMATION

Similar to determining which PSP problems may benefit from quantum computational advantages, accurately estimating the resources required for quantum algorithms that solve PSP problems presents its own difficulties. It is important to note that a rigorous resource estimation of a quantum algorithm for a given PSP problem is beyond the scope of a perspective

paper. However, we contextualize our perspective with a simplified resource estimation framework. We only consider a utility-scale quantum algorithm, however a similar framework can also be developed for a fault-tolerant quantum algorithm.

For the discussion presented in this section, we operate under the following set of assumptions:

- We use IBM quantum computers as our target devices. The diversity in QPU qubit connectivity and topology can have an impact on the resource estimation. Our estimates are based on the 127-qubit Eagle (R3) quantum chip on IBM\_Washington.
- The quantum algorithm employed changes the various resources needed to execute the experiment. We implement the algorithm from,<sup>149</sup> which is a variational quantum algorithm. As a result, qubit count depends on the choice of lattice for the coarse-grain model and the interaction model between residues. Moreover, the ansatz we have used is the RealAmplitudes ansatz from Qiskit's circuit library for only one repetition. We point the readers to<sup>161</sup> for comparative overview of different types of ansatze.
- A more comprehensive resource estimation requires analyzing the entire workflow in Figure 9, whereas we only focus on the first step of the workflow. Furthermore, while we analyze some commonly used metrics such as qubit count, circuit depth, ECR depth, and number of measurements, there are other parameters one can consider such as overall and gate-level execution time, overhead from error mitigation, etc.
- In a future work, our goal is to provide a robust resource estimation for PSP, where we consider the other parameters mentioned above.

For simplicity, we have adopted a simple regression model for extrapolation. Our approach involves mapping the problem to a quantum circuit, characterizing the circuit in terms of the number of gates and qubits required, applying different levels of circuit optimization and analyzing the number of ECR gates required for execution. We derived our estimates using the sequence of an experimentally verified protein structure (membrane-proximal cytoplasmic domain tail of platelet integrin  $\alpha$ IIb- $\beta$ 3) (The PDB entry can be found here <https://www.rcsb.org/structure/1M8O>). The estimates provided in the Figure 11 were calculated for up to 22 amino acids. Using the regression model we built, we have established the resource cost in different categories such as number of qubits, number of ECR gates that can be executed in parallel, total number of gates, total circuit depth for both optimized and unoptimized circuits, and upper limit on the number of measurements.

Our estimates show that there is a perfect quadratic relationship between the protein sequence length and the number of qubits required. The primary reason we have based our regression model up to 22 amino acids is the current capabilities of the quantum hardware. Our estimates show that for a protein sequence with 22 amino acids, the number of qubits needed is 118 (see Figure 11a). Extrapolated according to these estimates, we expect that we can go up to 41 amino acids with the release of IBM's Osprey, a 433 qubit quantum computer, and up to 67 amino acids on IBM's Condor, a 1121 qubit machine. As a comparison, similar estimates were conducted when converting the Hamiltonian into a QUBO format and this can be found in section 5.3 of the Supporting

Information. While qubit count is an important constraint for hardware experiments, we are also assuming that we will be able to run the number of gates that we project in these estimates. As we have access to more qubits and higher gate fidelity, we believe that it will be possible to predict larger and larger protein structures using quantum algorithms. While in terms of protein size, current deep learning based methods can target much larger structures, as we discussed in Section 3, we believe that there are still regimes in smaller proteins where physics-based quantum algorithms can yield better results.

In conclusion, we have established an initial analysis of the quantum computational resource cost for protein conformation predictions using the quantum algorithm from,<sup>149</sup> based on the full protein structure of platelet integrin  $\alpha$ IIb- $\beta$ 3 cytoplasmic domain.<sup>164</sup> Our estimates are crucial for understanding the limits of the subset of the problem space we want to target. There are various other techniques such as optimizing logical-to-physical qubit mapping, tensor-based circuit transpilation methods, circuit cutting and dynamic circuits that can enable us to predict larger proteins. In our future work, our goal is to explore some of these methods to enhance the performance of the quantum algorithm.

## 6. GENERAL DISCUSSION

The protein folding problem and protein structure prediction remain challenging tasks despite advances. Template-based methods, particularly those propelled by machine learning, have enabled researchers to obtain realistic models faster than wet lab experiments, establishing structure–function relationships that would have otherwise taken a very long time to discover. Growth of databases like the Protein Data Bank, now exceeding 200,000 experimental structures, underpins the success of these methods. However, the number of known protein sequences is outpacing the number of solved structures, around 300 million versus 200,000.<sup>165</sup> This highlights expansive unseen biodiversity. Advancing physics-based PSP methods is crucial to help bridge this gap.

Some argue that most natural folds have already been discovered, and by extension the template-based methods should be capable of predicting most structures. Remarkable performances like those of AlphaFold2 support this claim, given enough homology. However, improvement is still needed when dealing with mutations (even in conserved regions), orphan proteins, shallow MSAs, and novel sequences. AlphaMissense may address mutations.<sup>166</sup> Still, there are many missing pieces in this puzzle according to many in the biophysics community. Overcoming these limitations fundamentally requires simulating physics. Molecular dynamics could be a solution, but is constrained by time scales, system sizes, and the chosen force fields. *Ab initio* methods also scale poorly on classical hardware as the search space grows unmanageably large.

Quantum computing shows promise in both accuracy and scalability, as seen by the results of our quantum-classical workflow in predicting the Zika virus NS3 helicase P-loop. It produced an accurate model with a backbone RMSD of less than 2.0 Å compared to the experimental structure, and with superior scalability versus classical optimization.<sup>149</sup> With 10,000 qubits and sufficient circuit fidelity, prediction of biomedically relevant proteins and mutants could be attainable. For instance, our estimates indicate hemoglobin's 141 residues may be predictable using 4,967 entangled qubits—an extremely valuable prospect since its structure enables

fundamental life functions. Although this structure has long been determined experimentally, it is at this size and scale where some of the most biologically relevant proteins exist.

Early quantum algorithms will likely mirror classical physics methods. Eventually quantum chemical calculations could replace knowledge-based energy functions, perhaps initially with DFT or Hartree–Fock methods, for larger peptides than what is currently possible on classical hardware. The ultimate goal could be simulating exact electron dynamics, deriving molecular and electronic structures in real time, capturing nature in its most absolute form. This could bring about a new era for molecular force fields as well, yielding the ability to not only predict optimal protein structures irrespective of path, but truly help solve the path-dependent protein *folding* problem itself.

The future of structural biology depends profoundly on developing and scaling up efficient physics-based PSP methods. Quantum computing's potential to simulate nature's fundamental mechanics may help overcome the historical barriers. This will not have to be mutually exclusive, but could be rather synergistic, with quantum PSP methods helping complement the strengths of classical template-based methods, and vice versa. Multidisciplinary collaboration among biophysics, chemistry, computer science (both classical and quantum), and structural biology can help unravel the remaining mysteries. For the first time in decades, elucidating how and why proteins fold appears within reach, presenting a milestone that could deeply advance our comprehension of the subtle intricacies enabling life.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00067>.

A brief introduction to Protein Structure Prediction; The Levinthal paradox and the exponential growth of the search space; The “P-loop” of the Zika virus NS3 helicase protein; Structures for quantum speedup; The quantum algorithm ([PDF](#))

## ■ AUTHOR INFORMATION

### Corresponding Authors

Hakan Doga – IBM Quantum, Almaden Research Center, San Jose, California 95120, United States; [orcid.org/0000-0003-2574-2038](https://orcid.org/0000-0003-2574-2038); Email: [hakandoga@ibm.com](mailto:hakandoga@ibm.com)

Bryan Raubenolt – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States; Email: [raubebn@ccf.org](mailto:raubebn@ccf.org)

### Authors

Fabio Cumbo – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States

Jayadev Joshi – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States

Frank P. DiFilippo – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States

Jun Qin – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States

Daniel Blankenberg – Center for Computational Life Sciences, Lerner Research Institute, The Cleveland Clinic, Cleveland, Ohio 44106, United States; [orcid.org/0000-0002-6833-9049](https://orcid.org/0000-0002-6833-9049)

Omar Shehab – IBM Quantum, IBM Thomas J Watson Research Center, Yorktown Heights, New York 10598, United States

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.4c00067>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Portions of this material are based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001122C0102 to O.S. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency. The authors thank Vikram Khipple Mulligan from the Flatiron Institute, Ken A. Dill from Stony Brook University, Alán Aspuru-Guzik from University of Toronto, Kenneth Merz, Iris Smith, and William Martin from the Cleveland Clinic, and Emily Pritchett, Saif Rayyan, Antonio Mezzacapo, Courtney Long, Wiktor Mazin, Niall Robertson, Michal Rosen-Zvi, Jeannette Garcia, Filippo Utro, Kahn Rhrissorakrai, Ryan Sweke, Joseph Morrone, Yoel Shoshan, Ivano Tavernelli, Sarah Mostame, Sara Capponi, Jannis Born and Cynthia McKinney from IBM, and Robert Jernigan from Iowa State University of Science and Technology for their expert opinions and insightful feedback on the manuscript or during the project. The authors also acknowledge DARPA, NASA, and the MIT Lincoln Lab team involved in the Phase 1 of the DARPA Quantum Benchmarking program led by Joseph Altepeter, for their rigorous feedback.

## ■ REFERENCES

- (1) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.
- (2) Dill, K. A.; MacCallum, J. L. *Science* **2012**, *338*, 1042–1046.
- (3) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. *Nature* **2021**, *596*, 583–589.
- (4) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; et al. *Science* **2021**, *373*, 871–876.
- (5) Zhou, X.; Zheng, W.; Li, Y.; Pearce, R.; Zhang, C.; Bell, E. W.; Zhang, G.; Zhang, Y. *Nat. Protoc.* **2022**, *17*, 2326–2353.
- (6) Outeiral, C.; Nissley, D. A.; Deane, C. M. *Bioinformatics* **2022**, *38*, 1881–1887.
- (7) Noé, F.; De Fabritiis, G.; Clementi, C. *Curr. Opin. Struct. Biol.* **2020**, *60*, 77–84.
- (8) Kuhlman, B.; Bradley, P. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 681–697.
- (9) Ghosh, K.; Ozkan, S. B.; Dill, K. A. *J. Am. Chem. Soc.* **2007**, *129*, 11920–11927.
- (10) Bravyi, S.; Dial, O.; Gambetta, J. M.; Gil, D.; Nazario, Z. *J. Appl. Phys.* **2022**.
- (11) Burkart, N.; Huber, M. F. *Journal of Artificial Intelligence Research* **2021**, *70*, 245–317.
- (12) Perdomo-Ortiz, A.; Dickson, N.; Drew-Brook, M.; Rose, G.; Aspuru-Guzik, A. *Sci. Rep.* **2012**, *2*, 571.
- (13) Dirac, P. A. M. *The Principles of Quantum Mechanics*; Clarendon Press, 1981.

- (14) Heisenberg, W. In *The Physicist's Conception of Nature*; Springer: 1973; pp 264–275.
- (15) Nielsen, M. A.; Chuang, I. L. *Quantum Computation and Quantum Information*; Cambridge University Press: 2010.
- (16) Basu, S.; Born, J.; Bose, A.; Capponi, S.; Chalkia, D.; Chan, T. A.; Doga, H.; Goldsmith, M.; Gujarati, T.; Guzman-Saenz, A., et al. *arXiv*, July 11, 2023.
- (17) Shor, P. W. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*; 1994; pp 124–134.
- (18) Grover, L. K. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*; 1996; pp 212–219.
- (19) Ambainis, A.; Spalek, R. In *Annual Symposium on Theoretical Aspects of Computer Science*; 2006; pp 172–183.
- (20) Lloyd, S. *Science* **1996**, *273*, 1073–1078.
- (21) Aharonov, D.; Jones, V.; Landau, Z. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing*, 2006; pp 427–436.
- (22) Wocjan, P.; Chiang, C.-F.; Abeyesinghe, A.; Nagaj, D. *arXiv*, November 4, 2008.
- (23) Kalev, A.; Li, T.; Lin, C. Y.-Y.; Svore, K. M.; Wu, X., et al. *Leibniz International Proceedings in Informatics*; 2019.
- (24) Harrow, A. W.; Hassidim, A.; Lloyd, S. *Physical review letters* **2009**, *103*, No. 150502.
- (25) Engel, A.; Smith, G.; Parker, S. E. *Phys. Rev. A* **2019**, *100*, No. 062315.
- (26) Ambainis, A.; Balodis, K.; Iraids, J.; Kokainis, M.; Prumacris, K.; Vihrovs, J. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*; 2019; pp 1783–1793.
- (27) O'Meara, C.; Yogaraj, K.; Ghosh, K.; Sabino, P.; Fernández-Campoamor, M.; Cortiana, G.; Bernabé-Moreno, J.; Tacchino, F.; Mezzacapo, A.; Shehab, O., et al. *arXiv*, April 20, 2023.
- (28) Kim, Y.; Eddins, A.; Anand, S.; Wei, K. X.; Van Den Berg, E.; Rosenblatt, S.; Nayfeh, H.; Wu, Y.; Zaletel, M.; Temme, K., et al. *Nature* **2023**, *618*, 500–505.
- (29) Daley, A. J.; Bloch, I.; Kokail, C.; Flannigan, S.; Pearson, N.; Troyer, M.; Zoller, P. *Nature* **2022**, *607*, 667–676.
- (30) Ristè, D.; Da Silva, M. P.; Ryan, C. A.; Cross, A. W.; Córcoles, A. D.; Smolin, J. A.; Gambetta, J. M.; Chow, J. M.; Johnson, B. R. *npj Quantum Information* **2017**, *3*, 16.
- (31) Dale, H.; Jennings, D.; Rudolph, T. *Nat. Commun.* **2015**, *6*, 8203.
- (32) Stamatopoulos, N.; Mazzola, G.; Woerner, S.; Zeng, W. J. *Quantum* **2022**, *6*, 770.
- (33) Krunic, Z.; Flother, F.; Seegan, G.; Earnest-Noble, N.; Omar, S. *IEEE Trans. Quantum Eng.* **2022**, *3*, 1–11.
- (34) Wu, S. L.; Sun, S.; Guan, W.; Zhou, C.; Chan, J.; Cheng, C. L.; Pham, T.; Qian, Y.; Wang, A. Z.; Zhang, R., et al. *Physical Review Research* **2021**, *3*, No. 033221.
- (35) Levine, H.; Keesling, A.; Semeghini, G.; Omran, A.; Wang, T. T.; Ebadi, S.; Bernien, H.; Greiner, M.; Vuletić, V.; Pichler, H., et al. *Physical review letters* **2019**, *123*, No. 170503.
- (36) Burkard, G.; Ladd, T. D.; Nichol, J. M.; Pan, A.; Petta, J. R. *arXiv*, December 16, 2021.
- (37) Conlon, A.; Pellegrino, D.; Slingerland, J. K.; Dooley, S.; Kells, G. *Phys. Rev. B* **2019**, *100*, No. 134307.
- (38) Kozhanov, A.; Yu, Y.; Zhukas, L.; Feng, L.; Biswas, D.; Harraz, B.; Yan, K.; Zhang, V.; Noel, C.; Monroe, C. *Quantum* **2.0**; 2023; QM3A-2.
- (39) Steffen, M.; DiVincenzo, D. P.; Chow, J. M.; Theis, T. N.; Ketchen, M. B. *IBM J. Res. Dev.* **2011**, *55*, 13:1–13:11.
- (40) Cross, A. *APS March Meet. Abstr.* **2018**, *2018*, L58–003.
- (41) TCQ Development team. *CUDA Quantum*, version 0.4.0, 2023.
- (42) R Computing. *Pyquil documentation*, <http://pyquil.readthedocs.io/en/latest/webs/>, 2019.
- (43) Bergholm, V.; Izaac, J.; Schuld, M.; Gogolin, C.; Ahmed, S.; Ajith, V.; Alam, M. S.; Alonso-Linaje, G.; AkashNarayanan, B.; Asadi, A., et al. *arXiv*, November 12, 2018.
- (44) Developers, C. *Cirq, version v1.1.0. See full list of authors on GitHub*: <https://github.com/quantumlib/Cirq/graphs/contributorwebs>, 2022.
- (45) Gonzalez, C. *Digitale Welt* **2021**, *5*, 14–17.
- (46) *IonQ Technology Roadmap*, 2023.
- (47) *Quantinuum Technology Roadmap*, 2023.
- (48) Bourassa, J. E.; Alexander, R. N.; Vasmer, M.; Patil, A.; Tzitrin, I.; Matsuurra, T.; Su, D.; Baragiola, B. Q.; Guha, S.; Dauphinais, G., et al. *Quantum* **2021**, *5*, 392.
- (49) Bartolucci, S.; Birchall, P.; Bombin, H.; Cable, H.; Dawson, C.; Gimeno-Segovia, M.; Johnston, E.; Kieling, K.; Nickerson, N.; Pant, M., et al. *Nat. Commun.* **2023**, *14*, 912.
- (50) *IBM Quantum Roadmap*, 2015.
- (51) Alexeev, Y.; Bacon, D.; Brown, K. R.; Calderbank, R.; Carr, L. D.; Chong, F. T.; DeMarco, B.; Englund, D.; Farhi, E.; Fefferman, B. *PRX Quantum* **2021**, *2*, No. 017001.
- (52) Emani, P. S.; Warrell, J.; Anticevic, A.; Bekiranov, S.; Gandal, M.; McConnell, M. J.; Sapiro, G.; Aspuru-Guzik, A.; Baker, J. T.; Bastiani, M., et al. *Nat. Methods* **2021**, *18*, 701–709.
- (53) Beck, D.; Carlson, J.; Davoudi, Z.; Formaggio, J.; Quaglioni, S.; Savage, M.; Barata, J.; Bhattacharya, T.; Bishop, M.; Cloet, I., et al. *arXiv*, February 28, 2023.
- (54) Di Meglio, A.; Jansen, K.; Tavernelli, I.; Alexandrou, C.; Arunachalam, S.; Bauer, C. W.; Borras, K.; Carrazza, S.; Crippa, A.; Croft, V., et al. *arXiv*, July 6, 2023.
- (55) Egger, D. J.; Gambella, C.; Marecek, J.; McFaddin, S.; Mevissen, M.; Raymond, R.; Simonetto, A.; Woerner, S.; Yndurain, E. *IEEE Transactions on Quantum Engineering* **2020**, *1*, 1–24.
- (56) Berger, C.; Di Paolo, A.; Forrest, T.; Hadfield, S.; Sawaya, N.; Stechly, M.; Thibault, K. *arXiv*, June 23, 2021.
- (57) Kop, M.; Aboy, M.; De Jong, E.; Gasser, U.; Minssen, T.; Cohen, I. G.; Brongersma, M.; Quintel, T.; Floridi, L.; Laflamme, R. Available at SSRN 4393248, 2023.
- (58) Bennett, C. H.; Bernstein, E.; Brassard, G.; Vazirani, U. *SIAM journal on Computing* **1997**, *26*, 1510–1523.
- (59) Alvarez, J.; Gomez, C. *arXiv*, October 27, 1999.
- (60) Cafaro, C.; Mancini, S. *Physica A: Statistical Mechanics and its Applications* **2012**, *391*, 1610–1625.
- (61) Farhi, E.; Goldstone, J.; Gutmann, S. *arXiv*, November 14, 2014.
- (62) Peruzzo, A.; McClean, J.; Shadbolt, P.; Yung, M.-H.; Zhou, X.-Q.; Love, P. J.; Aspuru-Guzik, A.; O'Brien, J. L. *Nat. Commun.* **2014**, *5*, 4213.
- (63) Bravyi, S.; Kliesch, A.; Koenig, R.; Tang, E. *Physical review letters* **2020**, *125*, No. 260505.
- (64) Hastings, M. B. *arXiv*, May 16, 2019.
- (65) De Palma, G.; Marvian, M.; Rouzé, C.; França, D. S. *PRX Quantum* **2023**, *4*, No. 010309.
- (66) Clary, J. M.; Jones, E. B.; Vigil-Fowler, D.; Chang, C.; Graf, P. *Int. J. Quantum Chem.* **2023**, *123*, No. e27097.
- (67) Dill, K. A.; Alonso, D. O.; Hutchinson, K. *Biochemistry* **1989**, *28*, 5439–5449.
- (68) Dill, K. A.; Chan, H. S. *Nature structural biology* **1997**, *4*, 10–19.
- (69) Onuchic, J. N.; Wolynes, P. G. *Curr. Opin. Struct. Biol.* **2004**, *14*, 70–75.
- (70) Bryngelson, J. D.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 7524–7528.
- (71) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902–6915.
- (72) Leopold, P. E.; Montal, M.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 8721–8725.
- (73) *Folding funnel*, [https://en.wikipedia.org/wiki/Folding\\_funnel](https://en.wikipedia.org/wiki/Folding_funnel), 2023 (accessed on October 6, 2023).
- (74) Ma, B.; Kumar, S.; Tsai, C.-J.; Nussinov, R. *Protein engineering* **1999**, *12*, 713–720.
- (75) Clark, P. L. *Trends in biochemical sciences* **2004**, *29*, 527–534.
- (76) Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267*, 1619–1620.

- (77) Karplus, M. *Nat. Chem. Biol.* **2011**, *7*, 401–404.
- (78) Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Soccia, N. D. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 3626–3630.
- (79) Tsai, C.-J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181–1190.
- (80) Chikenji, G.; Fujitsuka, Y.; Takada, S. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 3141–3146.
- (81) Jin, W.; Kambara, O.; Sasakawa, H.; Tamura, A.; Takada, S. *Structure* **2003**, *11*, 581–590.
- (82) Hansmann, U. H.; Okamoto, Y.; Onuchic, J. N. *Proteins: Struct., Funct., Bioinf.* **1999**, *34*, 472–483.
- (83) Soccia, N. D.; Nymeyer, H.; Onuchic, J. *Physica D: Nonlinear Phenomena* **1997**, *107*, 366–382.
- (84) Camacho, C. J.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6369–6372.
- (85) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- (86) Bapat, A.; Jordan, S. *arXiv*, December 6, 2018.
- (87) Wu, R.-B.; Hsieh, M. A.; Rabitz, H. *Phys. Rev. A* **2011**, *83*, No. 062306.
- (88) Baldwin, C.; Laumann, C. *Phys. Rev. B* **2018**, *97*, No. 224201.
- (89) Leier, A.; Banzhaf, W. In *The 2003 Congress on Evolutionary Computation, 2003; CEC'03*, 2003; Vol. 1, pp 170–177.
- (90) Stepney, S.; Clark, J. A. *J. Comput. Theor. Nanosci.* **2008**, *5*, 942–969.
- (91) Streif, M.; Leib, M. *Quantum Science and Technology* **2020**, *5*, No. 034008.
- (92) McClean, J. R.; Harrigan, M. P.; Mohseni, M.; Rubin, N. C.; Jiang, Z.; Boixo, S.; Smelyanskiy, V. N.; Babbush, R.; Neven, H. *PRX Quantum* **2021**, *2*, No. 030312.
- (93) Wang, Z.; Zheng, P.-L.; Wu, B.; Zhang, Y. *Physical Review Research* **2023**, *5*, No. 023171.
- (94) Wang, Z.; Rubin, N. C.; Dominy, J. M.; Rieffel, E. G. *Phys. Rev. A* **2020**, *101*, No. 012320.
- (95) Shaydulin, R.; Lotshaw, P. C.; Larson, J.; Ostrowski, J.; Humble, T. S. *ACM Transactions on Quantum Computing* **2023**, *4*, 1–15.
- (96) Yousef, M.; Abdelkader, T.; El-Bahnasy, K. *Ain Shams Engineering Journal* **2019**, *10*, 713–719.
- (97) Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. *Nucleic Acids Res.* **2016**, *44*, W449–W454.
- (98) Xu, D.; Zhang, Y. *Proteins: Struct., Funct., Bioinf.* **2012**, *80*, 1715–1735.
- (99) Zhang, Y. *BMC Bioinformatics* **2008**.
- (100) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T., et al. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis; 2021*; pp 1–11.
- (101) Robustelli, P.; Piana, S.; Shaw, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **2018**.
- (102) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Phys. Chem. B* **2016**, *120*, 8313–8320.
- (103) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, S17–S20.
- (104) Paschek, D.; Nymeyer, H.; García, A. E. *J. Struct. Biol.* **2007**, *157*, S24–S33.
- (105) Day, R.; Paschek, D.; García, A. E. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1889–1899.
- (106) Rick, S. W.; Schwing, G. J.; Summa, C. M. *J. Chem. Inf. Model.* **2021**, *61*, 810–818.
- (107) Majewski, M.; Pérez, A.; Thölke, P.; Doerr, S.; Charron, N. E.; Giorgino, T.; Husic, B. E.; Clementi, C.; Noé, F.; De Fabritiis, G. *Nat. Commun.* **2023**.
- (108) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, S17–S20.
- (109) Nassar, R.; Brini, E.; Parui, S.; Liu, C.; Dignon, G. L.; Dill, K. A. *J. Chem. Theory Comput.* **2022**, *18*, 1929–1935.
- (110) Buel, G. R.; Walters, K. J. *Nature Structural and Molecular Biology* **2022**, *29*, 1–2.
- (111) Pak, M. A.; Markhieva, K. A.; Novikova, M. S.; Petrov, D. S.; Vorobyev, I. S.; Maksimova, E. S.; Kondrashov, F. A.; Ivankov, D. N. *PLoS One* **2023**, *18*, No. e0282689.
- (112) Pitera, J. W.; Swope, W. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 7587–7592.
- (113) Geng, H.; Chen, F.; Ye, J.; Jiang, F. *Computational and Structural Biotechnology Journal* **2019**, *17*, 1162–1170.
- (114) Barua, B.; Andersen, N. *The Trp-cage: Optimizing the Stability of a Globular Miniprotein*; 2008.
- (115) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. *SoftwareX* **2015**, *1*–*2*, 19–25.
- (116) Afgan, E.; et al. *Nucleic Acids Res.* **2022**, *50*, W345–W351.
- (117) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (118) Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
- (119) Maity, D. *djmaity/md-davis: MD DaVis v0.4.0*, version v0.4.0, 2022.
- (120) Deiana, A.; Forcelloni, S.; Porrello, A.; Giansanti, A. *PLoS One* **2019**, *14*, No. e0217889.
- (121) Ruff, K. M.; Pappu, R. V. *J. Mol. Biol.* **2021**, *433*, No. 167208.
- (122) Alderson, T. R.; Pritišanac, I.; Kolarić, D.; Moses, A. M.; Forman-Kay, J. D. *BioRxiv* **2022**.
- (123) Guo, H.-B.; Perminov, A.; Bekele, S.; Kedziora, G.; Farajollahi, S.; Varaljay, V.; Hinkle, K.; Molinero, V.; Meister, K.; Hung, C.; Dennis, P.; Kelley-Loughnane, N.; Berry, R. *Sci. Rep.* **2022**.
- (124) Rocha, S. M.; Pires, J.; Esteves, M.; Baltazar, G.; Bernardino, L. *Front. Cell. Neurosci.* **2014**.
- (125) Lee, C.; Su, B.-H.; Tseng, Y. J. *Brief. Bioinformatics* **2022**, DOI: [10.1093/bib/bbac308](https://doi.org/10.1093/bib/bbac308).
- (126) Hegedűs, T.; Geisler, M.; Lukács, G. L.; Farkas, B. *Cell. Mol. Life Sci.* **2022**.
- (127) Stevens, R. C. *The Cost and Value of Three-Dimensional Protein Structure*; 2003.
- (128) Sriram, K.; Insel, P. A. *Mol. Pharmacol.* **2018**, *93*, 251–258.
- (129) White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. *Methods Enzymol.* **2007**, *422*, 75–101.
- (130) Burger, L.; Van Nimwegen, E. *Mol. Syst. Biol.* **2008**, *4*, 165.
- (131) Weigt, M.; White, R. A.; Szurmant, H.; Hoch, J. A.; Hwa, T. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 67–72.
- (132) Jones, D. T.; Buchan, D. W.; Cozzetto, D.; Pontil, M. *Bioinformatics* **2012**, *28*, 184–190.
- (133) Liu, Y.; Palmedo, P.; Ye, Q.; Berger, B.; Peng, J. *Cell systems* **2018**, *6*, 65–74.
- (134) Xu, J.; Mcpartlon, M.; Li, J. *Nature Machine Intelligence* **2021**, *3*, 601–609.
- (135) Alley, E. C.; Khimulya, G.; Biswas, S.; AlQuraishi, M.; Church, G. M. *Nat. Methods* **2019**, *16*, 1315–1322.
- (136) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; et al. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118*, No. e2016239118.
- (137) Fang, X.; Wang, F.; Liu, L.; He, J.; Lin, D.; Xiang, Y.; Zhang, X.; Wu, H.; Li, H.; Song, L. *HelixFold-Single: MSA-free Protein Structure Prediction by Using Protein Language Model as an Alternative*; 2023.
- (138) Rao, R. M.; Liu, J.; Verkuil, R.; Meier, J.; Canny, J.; Abbeel, P.; Sercu, T.; Rives, A. In *Proceedings of the 38th International Conference on Machine Learning*; Meila, M., Zhang, T., Eds.; PMLR, 2021; Vol. 139, pp 8844–8856.
- (139) Mariani, V.; Biasini, M.; Barbato, A.; Schwede, T. *Bioinformatics* **2013**, *29*, 2722–2728.
- (140) Michaud, J. M.; Madani, A.; Fraser, J. S. *Nature biotechnology* **2022**, *40*, 1576–1577.
- (141) Meng, Q.; Guo, F.; Tang, J. *Brief. Bioinformatics* **2023**, *24*, No. bbad217.
- (142) Chowdhury, R.; Bouatta, N.; Biswas, S.; Floristean, C.; Kharkar, A.; Roy, K.; Rochereau, C.; Ahdriz, G.; Zhang, J.; Church, G. M.; et al. *Nat. Biotechnol.* **2022**, *40*, 1617–1623.

- (143) Ahdritz, G.; Bouatta, N.; Kadyan, S.; Jarosch, L.; Berenberg, D.; Fisk, I.; Watkins, A. M.; Ra, S.; Bonneau, R.; AlQuraishi, M. *OpenProteinSet: Training Data for Structural Biology at Scale*; 2023.
- (144) Stevens, A. O.; He, Y. *Biomolecules* **2022**, *12*, 985.
- (145) Hayashi, Y.; Aita, T.; Toyota, H.; Husimi, Y.; Urabe, I.; Yomo, T. *PLoS One* **2006**, *1*, No. e96.
- (146) Cornish, J.; Chamberlain, S. G.; Owen, D.; Mott, H. R. *Biochem. Soc. Trans.* **2020**, *48*, 2669–2689.
- (147) Haldane, A.; Levy, R. M. *Phys. Rev. E* **2019**, *99*, No. 032405.
- (148) Kryshtafovych, A.; Schwede, T.; Topf, M.; Fidelis, K.; Moult, J. *Proteins: Struct., Funct., Bioinf.* **2021**, *89*, 1607–1617.
- (149) Robert, A.; Barkoutsos, P. K.; Woerner, S.; Tavernelli, I. *npj Quantum Inf.* **2021**.
- (150) Elofsson, A. *Curr. Opin. Struct. Biol.* **2023**, *80*, No. 102594.
- (151) Pereira, M. *CASP15 High Accuracy*; 2023.
- (152) Xavier, J. S.; Rezende, P. M.; Velloso, J. P. L.; Nguyen, T.-B.; Karmarkar, M.; Portelli, S.; Ascher, D. B.; Pires, D. E. V. *ThermoMutDB*, 2021.
- (153) Zimmermann, L.; Stephens, A.; Nam, S.-Z.; Rau, D.; Kübler, J.; Lozajic, M.; Gabler, F.; Söding, J.; Lupas, A. N.; Alva, V. *Computation Resources for Molecular Biology*. *J. Mol. Biol.* **2018**, *430*, 2237–2243.
- (154) Marx, V. *Nat. Methods* **2021**, *18*, 715–719.
- (155) Miyazawa, S.; Jernigan, R. L. *Journal of molecular biology* **1996**, *256*, 623–644.
- (156) Leman, J. K.; Weitzner, B. D.; Lewis, S. M.; Adolf-Bryfogle, J.; Alam, N.; Alford, R. F.; Aprahamian, M.; Baker, D.; Barlow, K. A.; Barth, P.; et al. *Nat. Methods* **2020**, *17*, 665–680.
- (157) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; et al. *J. Chem. Theory Comput.* **2017**, *13*, 3031–3048.
- (158) Gurobi Optimization, LLC *Gurobi Optimizer Reference Manual*; 2023.
- (159) Spivak, M.; Stone, J. E.; Ribeiro, J.; Saam, J.; Freddolino, P. L.; Bernardi, R. C.; Tajkhorshid, E. *J. Chem. Inf. Model.* **2023**, *63*, 4664–4678.
- (160) Mandal, A.; Roy, A.; Upadhyay, S.; Ushijima-Mwesigwa, H. *Compressed Quadratization of Higher Order Binary Optimization Problems*; 2020.
- (161) Du, Y.; Tu, Z.; Yuan, X.; Tao, D. *Phys. Rev. Lett.* **2022**, *128*, No. 080506.
- (162) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. *nature* **2017**, *549*, 242–246.
- (163) IBM *Quantum Documentation*; 2023.
- (164) Vinogradova, O.; Velyvis, A.; Velyviene, A.; Hu, B.; Haas, T. A.; Plow, E. F.; Qin, J. *Cell* **2002**, *110*, 587–597.
- (165) Jisna, V. A.; Jayaraj, P. B. *Protein Journal* **2021**, *40*, 522–544.
- (166) Cheng, J.; Novati, G.; Pan, J.; Bycroft, C.; Žemgulytedot, A.; Applebaum, T.; Pritzel, A.; Wong, L. H.; Zielinski, M.; Sargeant, T.; et al. *Science* **2023**, *381*, No. eadg7492.