Enhancing computational peptide therapeutic design with quantum chemistry and quantum computing

Vikram K. Mulligan[1], Tristan Zaborniak[2], P. Douglas Renfrew[1], and Benjamin Brown[3].

1. Biomolecular Design Group, Center for Computational Biology, Flatiron Institute, New York, NY 10010, United States of America.
2. Department of Computer Science, University of Victoria, Victoria, BC V8W 2Y2, Canada.
3. Department of Chemistry, Vanderbilt University, Nashville, TN 37240, United States of America.

https://doi.org/10.17952/37EPS.2024.P1286

Introduction

Advances in synthetic chemistry have greatly increased the number of amino acid and amino acid-like monomers that can be incorporated into a synthetic heteropolymer, creating exciting opportunities for the rational design of peptide and peptide-like drugs, nanomaterials, and catalysts. In recent years, we have worked to generalize the Rosetta software suite to permit its application to the design of synthetic peptides and heteropolymers built from non-canonical amino acids and other artificial building-blocks, and to allow validation of designs by simulation of molecules' folding properties and propensities [1–7]. Examples of synthetic peptides designed with our pipeline and validated by X-ray crystallography or NMR spectroscopy are shown in **Fig. 1**.

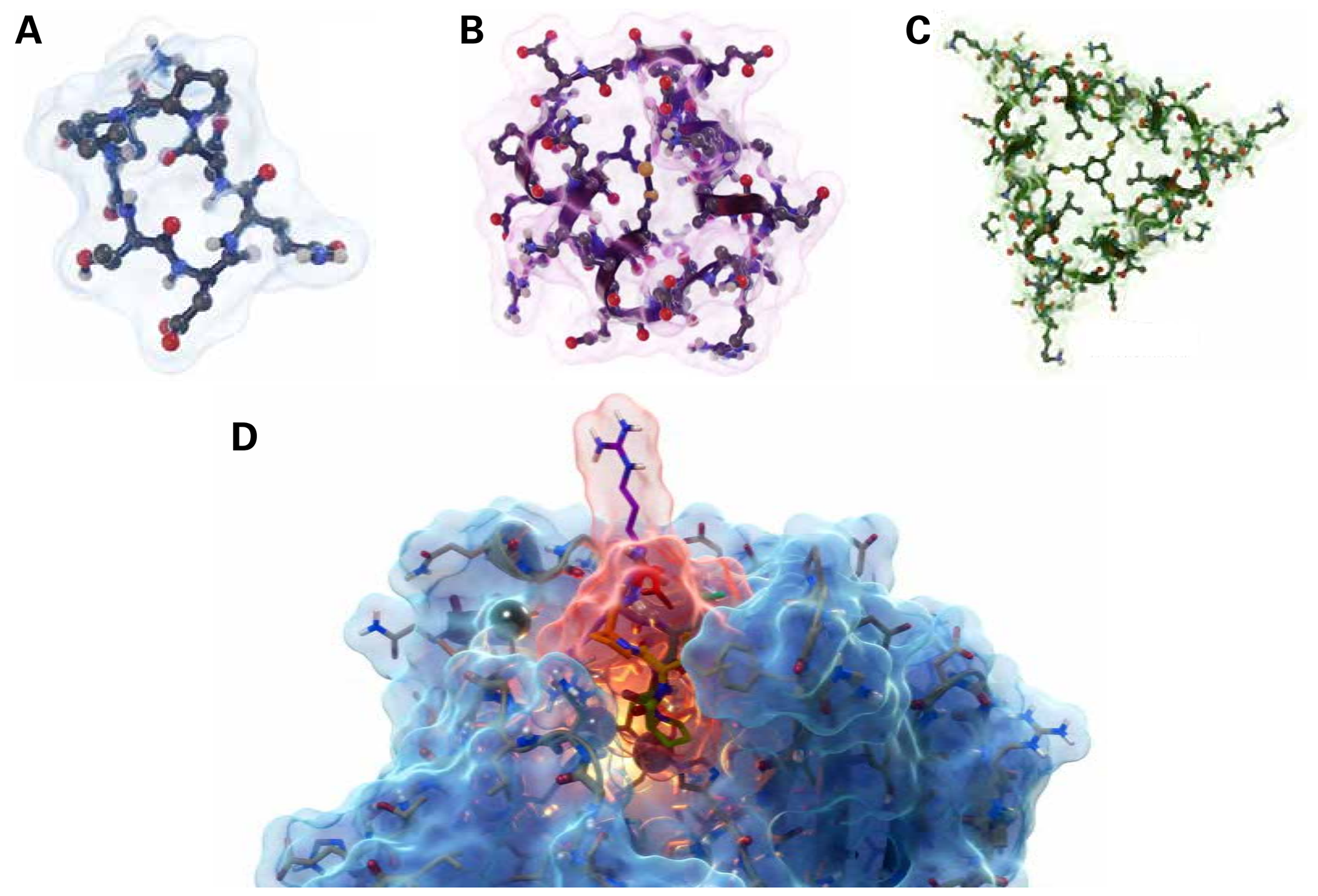

Figure 1: Synthetic peptides designed using the Rosetta software suite. (**A**) A 7-residue cyclic peptide, made from a mixture of D- and L-amino acids, that folds into a rigid, designed fold [2,4,7]. (**B**) A 26-residue cyclic peptide made from D- and L-amino acids that folds into a tertiary fold with helices of opposite handedness packing against each other, a fold impossible for any natural protein made from canonical amino acids alone [1]. (**C**) A 60-residue cyclic polypeptide with a fold stabilized by 1,3,5-tris(bromomethyl) benzene that serves both as a covalent three-way crosslink and as a component of a small hydrophobic core [3]. (**D**) An 8-residue cyclic peptide inhibitor (pink) of the New Delhi metallo-β-lactamase 1 (blue), an enzyme involved in antibiotic resistance. This peptide, also built from a mixture of D- and L-amino acids, folds rigidly into its binding-competent conformation, largely eliminating the entropic cost associated with ordering a disordered peptide on binding to a target [5].

However, this chemical diversity creates two challenges for rational design. First, force field-based design methods generalize poorly, and must be reparameterized as new chemical building blocks are added. More exotic building-blocks often have few experimental data available for empirically calibrating force fields. Second, the expansion in the number of possible building blocks vastly increases the number of combinatorial possibilities, creating an intractably large search space for computational optimization-based design approaches. Here, we describe recent work to address both of these challenges. To achieve greater accuracy and generality, we have developed *RosettaQM*, a set of software tools allowing quantum chemistry calculations to be incorporated into any existing Rosetta protocol. To tackle the combinatorial problem, we have developed a new open-source software library that extends Rosetta or other peptide modelling packages, which we call *Masala*. Masala provides diverse high-efficiency solvers for combinatorial optimization problems like the sequence design problem. Beyond traditional Monte Carlo based solvers for classical computing hardware, we have mapped these problems to current-generation quantum annealers and gate-based quantum computers, which may offer scaling advantages as the combinatorial space continues to balloon. We believe that these computational tools will be an important complement to experimental screening and other approaches.

Introducing Quantum Chemistry in Peptide Design and Validation Pipelines

Historically, the Rosetta software (used for macromolecular design and structure prediction) has been limited by its reliance on classical force fields, which are of limited accuracy and which are difficult to extend to support synthetic chemical building-blocks. Quantum mechanics-based (QM) calculations permit far greater accuracy and generality, albeit at greater computational cost depending on the level of approximation used to solve the multi-body Schrödinger equation. Historically, these have been laborious to set up, however. We have developed the RosettaQM communications bridge to allow Rosetta to communicate with GAMESS, Orca, or other quantum chemistry software, permitting QM energy calculations in the context of any Rosetta protocol. **Fig. 2A** illustrates how a user may use low-cost, low-accuracy force fields, medium-cost, medium accuracy semi-empirical methods, and high-cost, high-accuracy DFT calculations in the context of a single protocol, reserving the more expensive calculations for samples that pass intermediate filters. **Fig. 2B** shows RosettaQM features that facilitate its application to large peptides and proteins, including fragment molecular orbital (FMO) calculations that use a "divide and conquer" approach to get around the scaling of full DFT calculations, as well as the RosettaQM MultiScoreFunction, which allows different levels of QM theory to be applied to different regions of a structure.

We demonstrate some of RosettaQM's many applications by showing the computation of mainchain potentials (Ramachandran maps) for different canonical and non-canonical amino acids (**Fig. 2C**). Precomputed lookup tables for the mainchain potential are essential for the force fields used in peptide design and validation, so this provides facile means of generating these for new non-canonical amino acids [8]. In addition, RosettaQM permits QM-based energy calculations in the context of existing protocols, including peptide conformational sampling and structure prediction protocols. **Fig. 2D** shows how even approximate semi-empirical methods, such as DFTB, offer accuracy and generality well beyond those of force fields, permitting the accurate prediction of the folded structure of cyclosporine A in organic solvent for the first time. RosettaQM also generalizes well to more exotic chemical building blocks. We are currently applying it to the design and validation of peptoid polymers (N-functionalized poly-glycine) with secondary and tertiary folds (**Fig. 2E**).

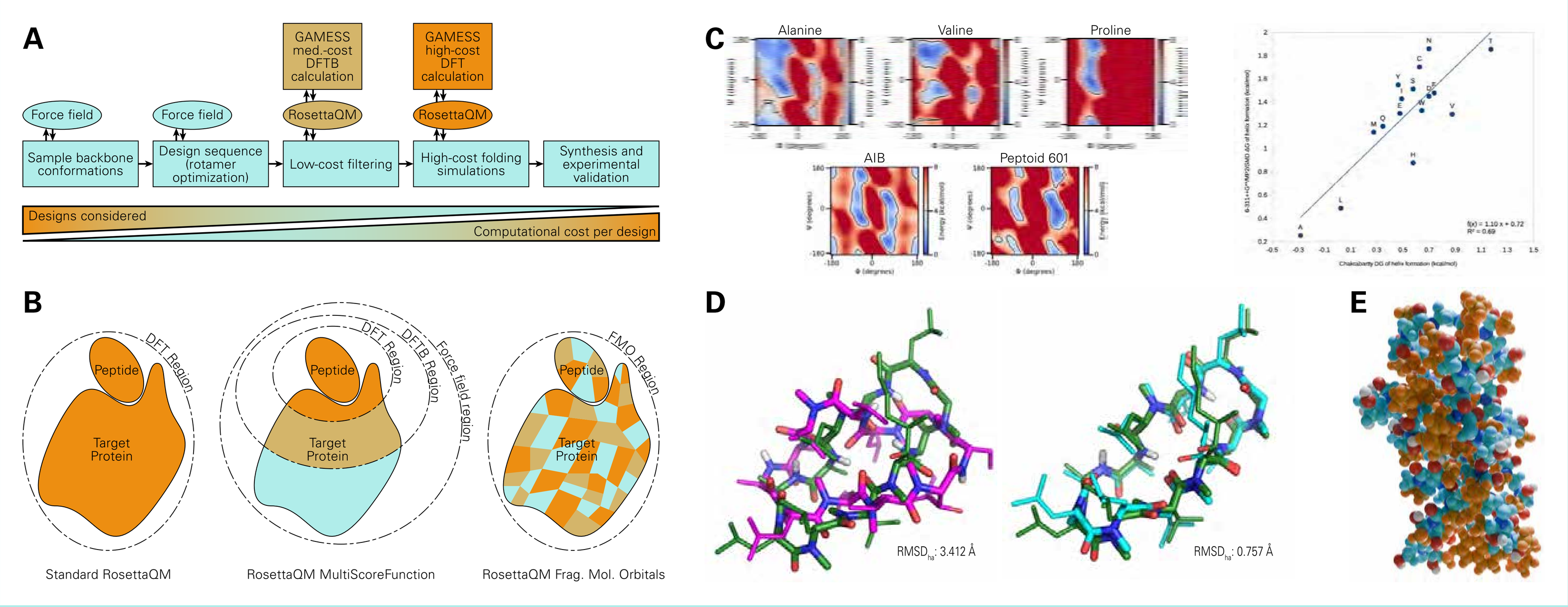

Figure 2: Quantum mechanics-based (QM) energy calculations in peptide design and validation pipelines using RosettaQM. (**A**) Existing design and validation pipelines enhanced with RosettaQM. While force fields can continue to be used for steps that must be fast, where accuracy is of finite importance, semi-empirical calculations or full DFT calculations can be performed at later stages. RosettaQM permits a QM calculation to be configured from within Rosetta, sent to GAMESS or Orca, executed, and the result imported into Rosetta, all without user intervention. (**B**) Modes of operation of RosettaQM. Left: a basic RosettaQM calculation treats a molecular structure as a quantum-mechanical system. Middle: the RosettaQM "MultiScoreFunction" permits regions of a structure to be scored with force fields or with different levels of QM theory, using mechanical or electrostatic embedding schemes. Right: RosettaQM supports fragment molecular orbital (FMO) calculations in GAMESS [12], in which separate QM calculations performed on individual amino acids or on pairs of interacting amino acids are combined to produce an estimate of the overall energy of the system. (**C**) Computation of Ramachandran potentials with RosettaQM (using the 6-311++G** level of theory and MP2). Left: Ramachandran potentials for alanine, proline, and valine, plus non-canonical amino acid 2-aminoisobutyric acid and peptoid monomer (S)-N-(1-phenylethyl)-glycine. Right: Correlation between computed and experimentally-determined helical propensity of amino acids [11]. (**D**) Prediction of the folded state of cyclosporine A in organic solvent. Left: overlay of prediction using Rosetta ref2015 energy function (magenta) on the X-ray crystal structure (green). The predicted structure has no resemblance to the reality. Right: overlay of prediction using RosettaQM DFTB semi-empirical method (cyan) on the X-ray crystal structure (green), showing excellent agreement. (**E**) Designed peptoid helical bundle fold using RosettaQM for validation. Polar groups are shown in cyan, and apolar groups in orange. Work to make peptoids that fold into intricate tertiary folds as proteins do is ongoing.

Mapping the Sequence Design Problem to Quantum Computers

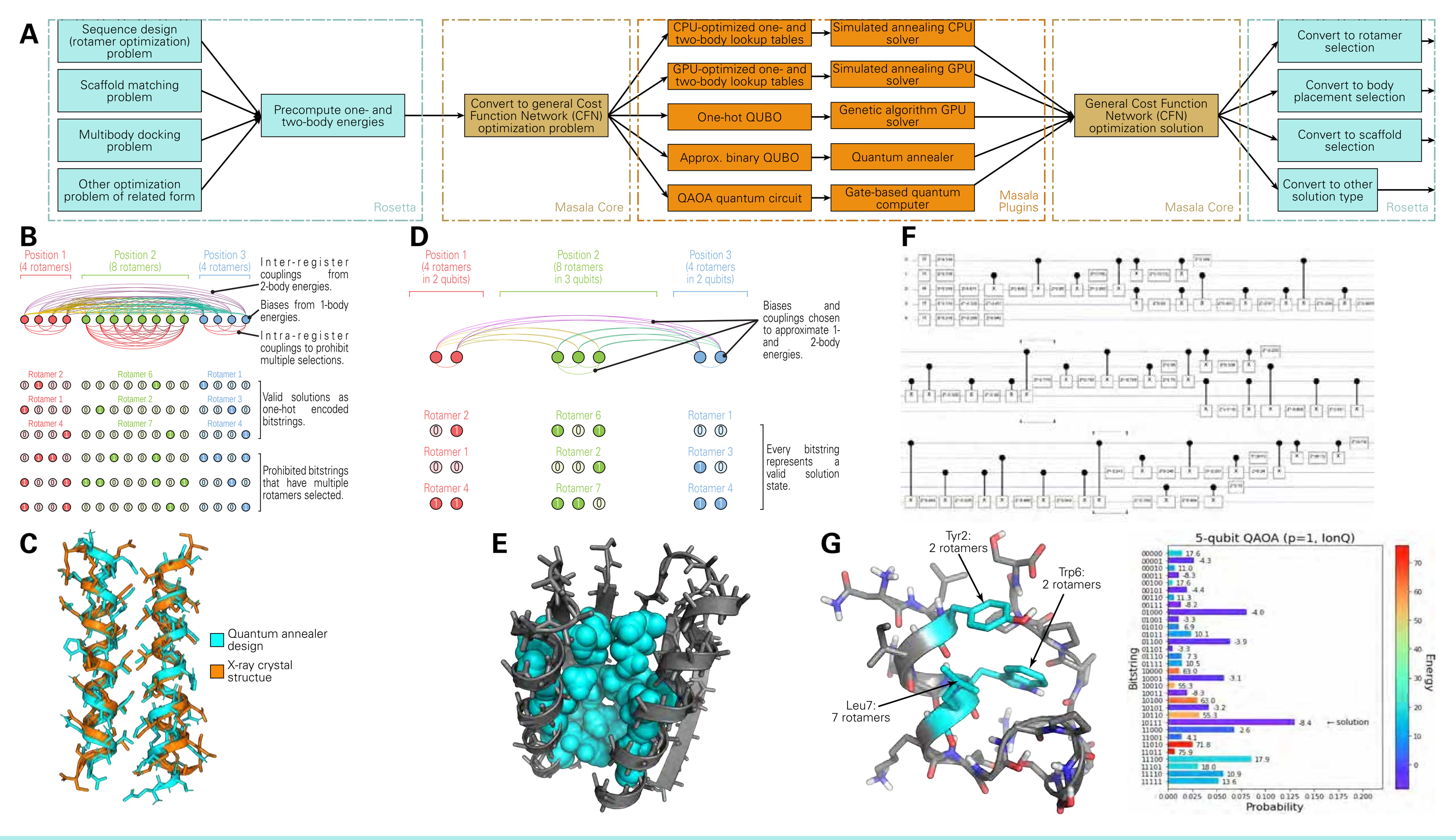

Figure 3: Implementing support for external combinatorial optimizers, including quantum optimizers, for use in Rosetta protocols. (**A**) Schematic of Rosetta's communication with the Masala libraries. Rosetta precomputes one- and two-body energies for rotamer optimization or related problems, then sends the problem description to Masala expressed as a generic cost function network optimization (CFN) problem. Masala's core detects available plug-in optimizer (CPU, GPU, or QPU-based) at runtime, sends the CFN problem to a user-selected optimizer. The CFN solution is transmitted back to Rosetta and converted to a rotamer optimization solution. (**B**) Mapping of the rotamer optimization problem to the quantum annealer, using *ND* qubits to represent *D* candidate rotamers at each of *N* possible positions. Final solutions are expressed as a bitstring with one-hot encoding of selected rotamers [9,10]. (**C**) A self-assembling heterochiral helical bundle peptide designed on the quantum annealer using the hybrid classical-quantum QBSolv algorithm. The design (cyan) is shown superimposed on the X-ray crystal structure (orange). (**D**) An approximate binary encoding of the rotamer optimization problem for the quantum annealer, using $N \log_2(D)$ qubits. This encoding uses far fewer qubits, but introduces some error in states' scores. (**E**) A well-packed protein core (cyan) designed entirely on the quantum annealer with the approximate binary encoding. (**F**) Mapping of the rotamer optimization problem to a gate-based quantum computer, using the quantum approximate optimization algorithm (QAOA). (**G**) A small rotamer optimization problem solved on the IonQ quantum computer. Left: three side-chains in the core of the Top7 protein were optimized with 7, 2, and 2 rotamers (28 total combinations). Right: with repeated sampling, the lowest-energy solution was found most frequently and showed considerable enrichment.

Given *D* possible synthetic building blocks in an N-residue peptide, how is one to choose which of the D^N possible sequences to synthesize? Even supercomputers cannot exhaustively enumerate possible sequences as *D* or *N* grow large. Historically, Rosetta has used simulated annealing as a heuristic for finding good sequences likely to produce a desired fold and function, at the expense of any guarantees of finding the best. However, Rosetta's simulated annealer lacks support for modern hardware (multi-core CPUs, GPUs), extensibility (support for alternative search algorithms), or versatility (support for optimization problems other than sequence design). We have built the Masala software library as an extension library against which software like Rosetta can be linked. Masala has versatile plug-in architecture, allowing easy development of plug-in Masala optimizers that can be detected at runtime and used by Rosetta without recompilation. **Fig. 3A** shows Rosetta's interaction with Masala's solvers. The sequence design problem is separable into a energy calculation step, in which the energies of candidate rotamers (amino acid identities with fixed side-chain conformations) and pairs of rotamers are pre-computed and stored, and a combinatorial optimization step, in which the precomputed one- and two-body energies are used to try to find a selection of one rotamer per position that minimizes energy. The first step can be performed by Rosetta, and the second by a general-purpose solver provided by Masala.

Quantum computers offer a possible means of surmounting the scaling obstacles of difficult combinatorial optimization problems. We have implemented solvers in Masala that allow Rosetta design using quantum computers. Our initial implementation (described in [9]) used an exact one-hot encoding of rotamer selections, requiring *ND* qubits to represent *D* rotamers at *N* positions (**Fig. 3B**). As a proof of principle, we used this to design a self-assembling mixed-chirality peptide helical bundle on the D-Wave quantum annealer, which we synthesized by solid-phase methods and validated by x-ray crystallography (**Fig. 3C**). However, this required more qubits than were available on the then-current D-Wave 2000Q quantum annealer, so that it was necessary to use a hybrid classical-quantum algorithm called QBSolv to divide the problem into pieces that could be solved on the annealer. We subsequently developed an approximate binary encoding (**Fig. 3D**) that compresses the problem to use $N \log_2(D)$ qubits at the expense of some accuracy in energy values. This has permitted the direct design of much larger polypeptides (*e.g.* **Fig. 3E**) entirely on the QPU. We have also explored a mapping to gate-based quantum computers using the QAOA algorithm for optimization (**Fig. 3F**). On current devices, such as systems built by IBM or IonQ, this allows only very small rotamer optimization problems to be solved; nevertheless, it is able to rapidly find lowest-energy solutions (**Fig. 3G**). Since we have previously shown that the multibody docking problem, another NP-hard problem, maps to the same functional form [10], we anticipate that these solvers will be broadly useful.

Conclusions

Robust computational peptide design faces challenges of accuracy, generality, and tractability. By introducing quantum mechanical energy calculations into design and validation pipelines through the RosettaQM bridge, we have addressed the major sources of inaccuracy in energy calculations, and have greatly improved our ability to generalize the methods to new chemical building-blocks. By mapping design problems to quantum annealers and gate-based quantum computers, we have met tractability challenges head-on, and opened the door to tackling design problems involving far more candidate building-blocks as quantum computers improve.

References

[1] G. Bhardwaj*, V.K. Mulligan*, C.D. Bahl*, J.M. Gilmore, P.J. Harvey, O. Cheneval, G.W. Buchko, S.V.S.R.K. Pulavarti, Q. Kaas, A. Eletsky, P.-S. Huang, W.A. Johnsen, P.J. Greisen, G.J. Rocklin, Y. Song, T.W. Linsky, A. Watkins, S.A. Rettie, X. Xu, L.P. Carter, R. Bonneau, J.M. Olson, E. Coutsias, C.E. Correnti, T. Szyperski, D.J. Craik, D. Baker, Accurate de novo design of hyperstable constrained peptides, *Nature* 538 (2016) 329–335. (*Co-primary authors.)
[2] P. Hosseinzadeh*, G. Bhardwaj*, V.K. Mulligan*, M.D. Shortridge, T.W. Craven, F. Pardo-Avila, S.A. Rettie, D.E. Kim, D.-A. Silva, Y.M. Ibrahim, I.K. Webb, J.R. Cort, J.N. Adkins, G. Varani, D. Baker, Comprehensive computational design of ordered peptide macrocycles, *Science* 358 (2017) 1461–1466. https://doi.org/10.1126/science.aap7577. (Co-primary authors.)
[3] B. Dang*, H. Wu*, V.K. Mulligan*, M. Mravic, Y. Wu, T. Lemmin, A. Ford, D.-A. Silva, D. Baker, W.F. DeGrado, De novo design of covalently constrained mesosize protein scaffolds with unique tertiary structures, *Proc. Natl. Acad. Sci. U.S.A.* 114 (2017) 10852–10857. https://doi.org/10.1073/pnas.1710695114. (Co-primary authors.)
[4] V.K. Mulligan, C.S. Kang, M.R. Sawaya, S. Rettie, X. Li, I. Antselovich, T.W. Craven, A.M. Watkins, J.W. Labonte, F. DiMaio, T.O. Yeates, D. Baker, Computational design of mixed chirality peptide macrocycles with internal symmetry, *Protein Sci.* 29 (2020) 2433–2445. https://doi.org/10.1002/pro.3974.
[5] V.K. Mulligan, S. Workman, T. Sun, S. Rettie, X. Li, L.J. Worrall, T.W. Craven, D.T. King, P. Hosseinzadeh, A.M. Watkins, P.D. Renfrew, S. Guffy, J.W. Labonte, R. Moretti, R. Bonneau, N.C.J. Strynadka, D. Baker, Computationally designed peptide macrocycle inhibitors of New Delhi metallo-β-lactamase 1, *Proc. Natl. Acad. Sci. U.S.A.* 118 (2021) e2012800118. https://doi.org/10.1073/pnas.2012800118.
[6] P. Hosseinzadeh, P.R. Watson, T.W. Craven, X. Li, S. Rettie, F. Pardo-Avila, A.K. Bera, V.K. Mulligan, P. Lu, A.S. Ford, B.D. Weitzner, L.J. Stewart, A.P. Moyer, M. Di Piazza, J.G. Whalen, P.J. Greisen, D.W. Christianson, D. Baker, Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites, *Nat. Commun.* 12 (2021) 3384. https://doi.org/10.1038/s41467-021-23609-8.
[7] G. Bhardwaj, J. O'Connor, S. Rettie, Y.-H. Huang, T.A. Ramelot, V.K. Mulligan, G.G. Alpkilic, J. Palmer, A.K. Bera, M.J. Bick, M. Di Piazza, X. Li, P. Hosseinzadeh, T.W. Craven, R. Tejero, A. Lauko, R. Choi, C. Glynn, L. Dong, R. Griffin, W.C. van Voorhis, J. Rodriguez, L. Stewart, G.T. Montelione, D. Craik, D. Baker, Accurate de novo design of membrane-traversing macrocycles, *Cell* 185 (2022) 3520-3532.e26. https://doi.org/10.1016/j.cell.2022.07.019.
[8] V. Mironov, Y. Alexeev, V.K. Mulligan, D.G. Fedorov, A systematic study of minima in alanine dipeptide, *J. Comput. Chem.* 40 (2019) 297–309. https://doi.org/10.1002/jcc.25589.
[9] V.K. Mulligan, H. Melo, H.I. Merritt, S. Slocum, B.D. Weitzner, A.M. Watkins, P.D. Renfrew, C. Pelissier, P.S. Arora, R. Bonneau, Designing Peptides on a Quantum Computer, *bioRxiv* (2019) 752485. https://doi.org/10.1101/752485.
[10] M. Pandey, T. Zaborniak, H. Melo, A. Galda, V.K. Mulligan, Multibody molecular docking on a quantum annealer, *arXiv* (2022). https://doi.org/10.48550/arXiv.2210.11401.
[11] A. Chakrabartty, T. Kortemme, R.L. Baldwin, Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions, *Protein Sci. Publ. Protein Soc. 3* (1994) 843–852. https://doi.org/10.1002/pro.5560030514.
[12] D.G. Fedorov, The fragment molecular orbital method: theoretical development, implementation in GAMESS, and applications, *WIREs Comput. Mol. Sci.* 7 (2017) e1322. https://doi.org/10.1002/wcms.1322.

Collaboration Opportunity

The Biomolecular Design Group seeks collaborators interested in working together to develop macromolecules for binding to targets of therapeutic interest, for functional nanomaterial applications, or for catalysis of reactions of interest. In particular, we seek experimental collaborators interested in recombinant expression of proteins, chemical synthesis of peptides, peptoids, or other heteropolymers, or experimental screening of designed macromolecules in in vitro or in vivo assays. Please contact vmulligan@flatironinstitute.org if you are interested.

Biomolecular Design Group
Center for Computational Biology,
Flatiron Institute,
162 5th Ave, New York, NY 10010

simonsfoundation.org/flatiron/