

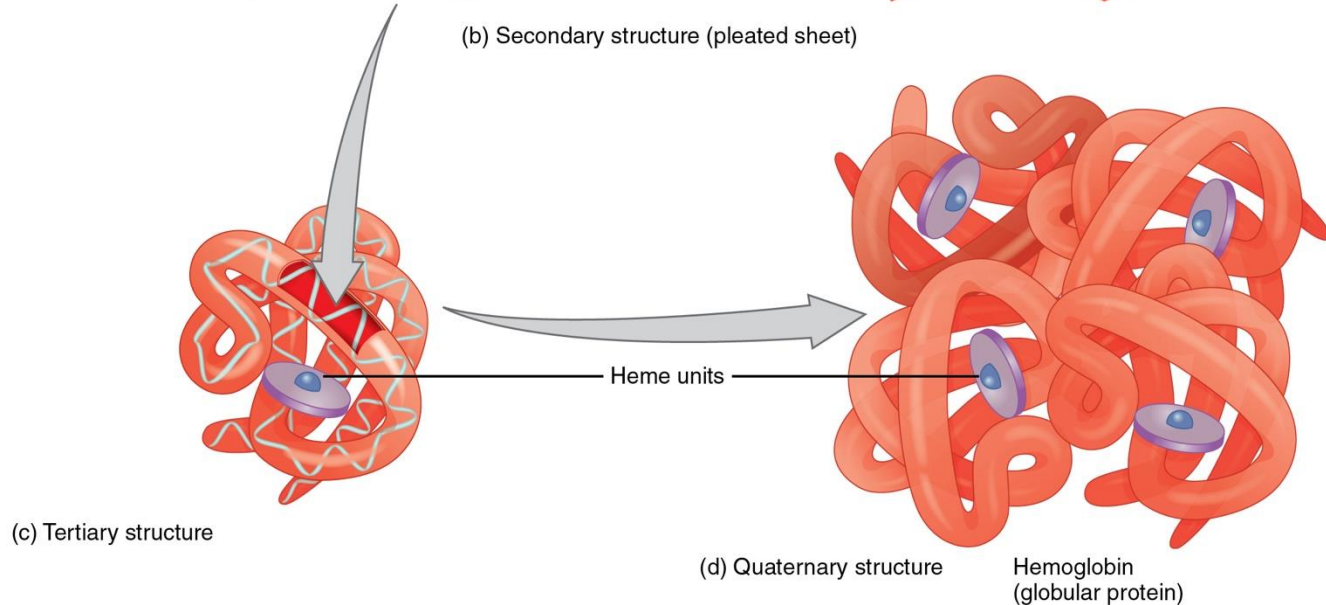
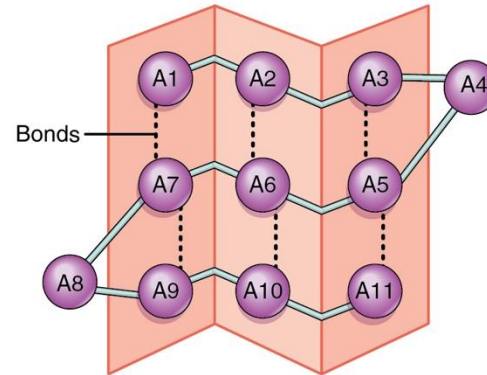
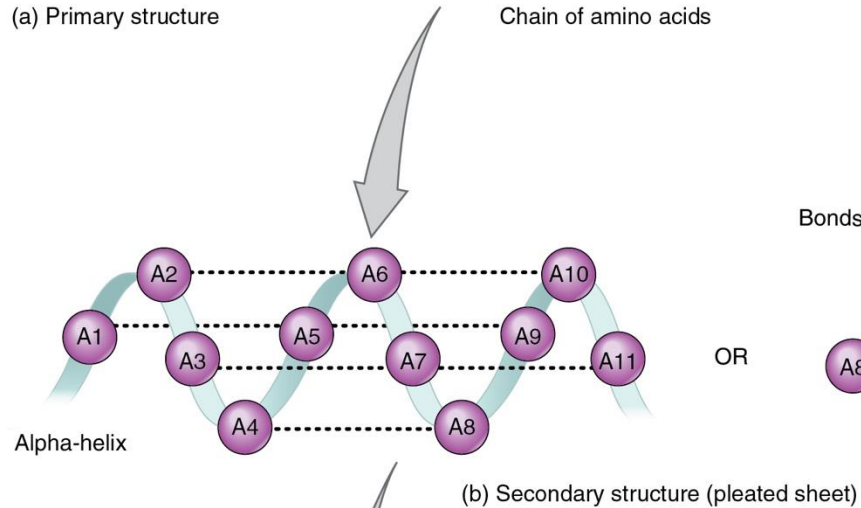
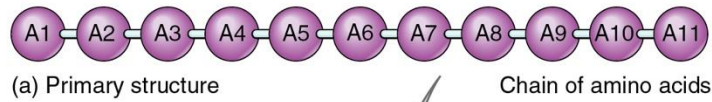
A robust and versatile computational peptide design pipeline to inform wet-lab experiments

Vikram K. Mulligan, Ph.D.
Research Scientist, Biomolecular Design Group
Center for Computational Biology
Flatiron Institute
Monday, 16 December 2024

Outline

1. Peptide and protein drug design on classical computers.
2. Enhancing peptide and protein drug design with quantum chemistry calculations (on classical computers).
3. Enhancing peptide and protein drug design with quantum computers

What do we mean by biomolecule “design”?



Sequence

Find a sequence of chemical building-blocks...

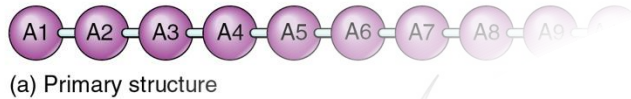
Fold

...that folds into a desired structure or fold...

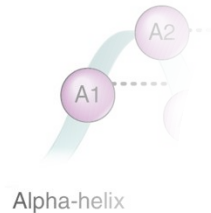
Function

...that produces a desired function.

What do we mean by biomolecule “design”?

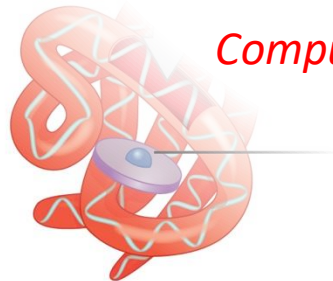


But an N -residue peptide made from B different building blocks has B^N possible sequences.



For a 10-residue peptide made from the 20 canonical amino acids, there are 20^{10} , or about **10^{13} (ten trillion), possible sequences** – already too many to screen using the largest libraries that we can plausibly make.

If we increase the number of possible building-blocks to 1,000, then there are 1000^{10} , or **10^{30} , possible sequences**. If you made one molecule of each, your library would weigh roughly 30 tonnes.



(c) Tertiary structure

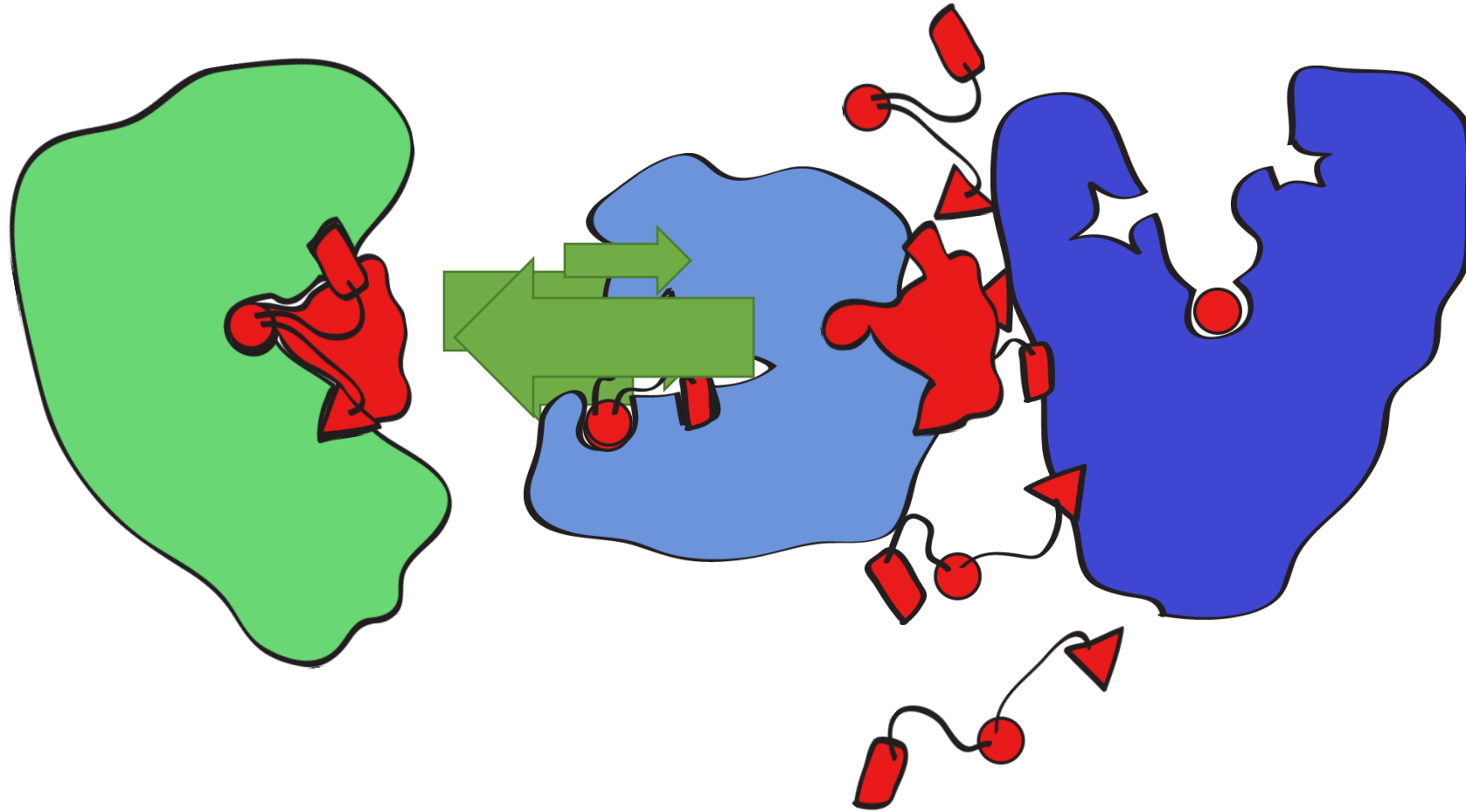
(d) Quaternary structure

Hemoglobin
(globular protein)

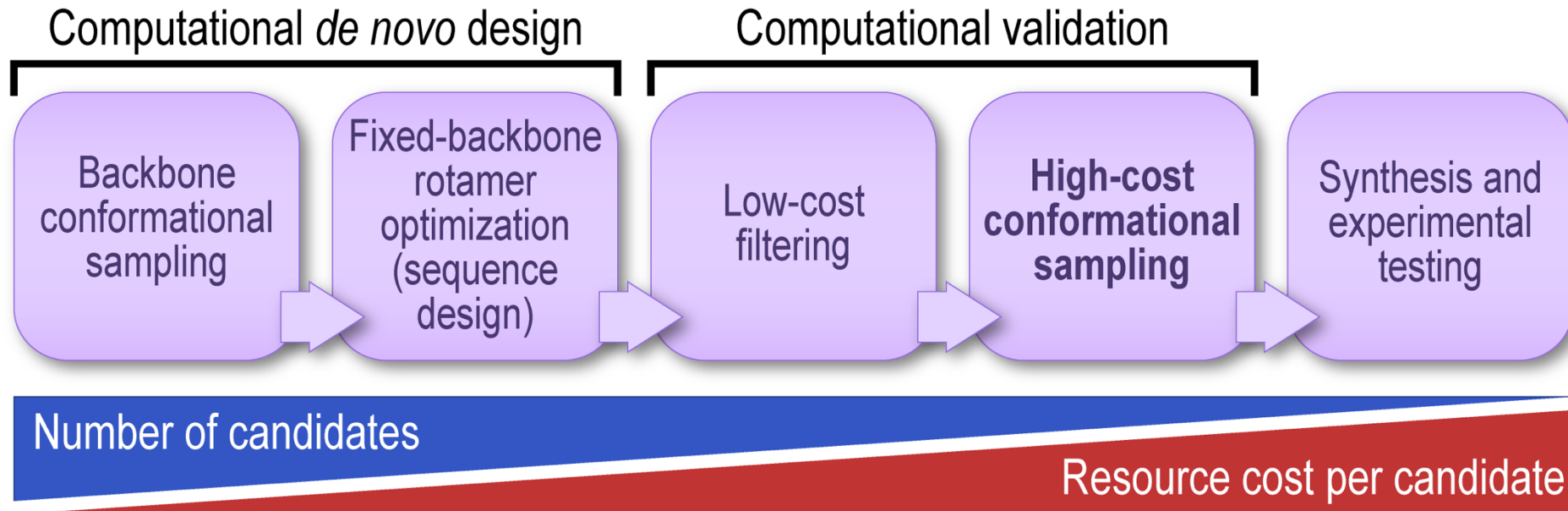
Computational design is therefore an essential complement to library screening.

... produces a desired function.

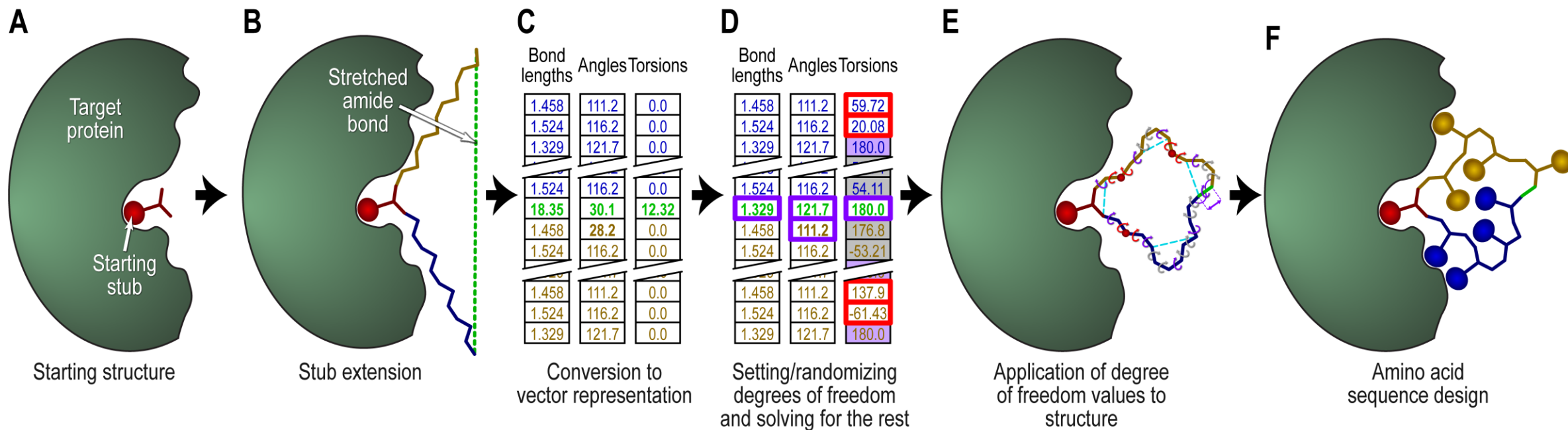
Conformational entropy: The Achilles heel of macrocycles?



Pipeline for designing rigidly-folded peptide drugs



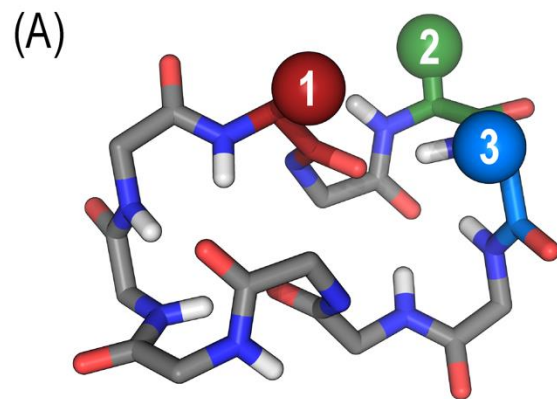
Classical peptide macrocycle design with the Rosetta software suite



From Mulligan VK. (2022) "Computational methods for peptide macrocycle drug design." Chapter in *Peptide Therapeutics: Fundamentals of Design, Development, and Delivery*, Jois S., ed. Berlin, Germany: Springer. 2022.

DOI: 10.1007/978-3-031-04544-8_3.

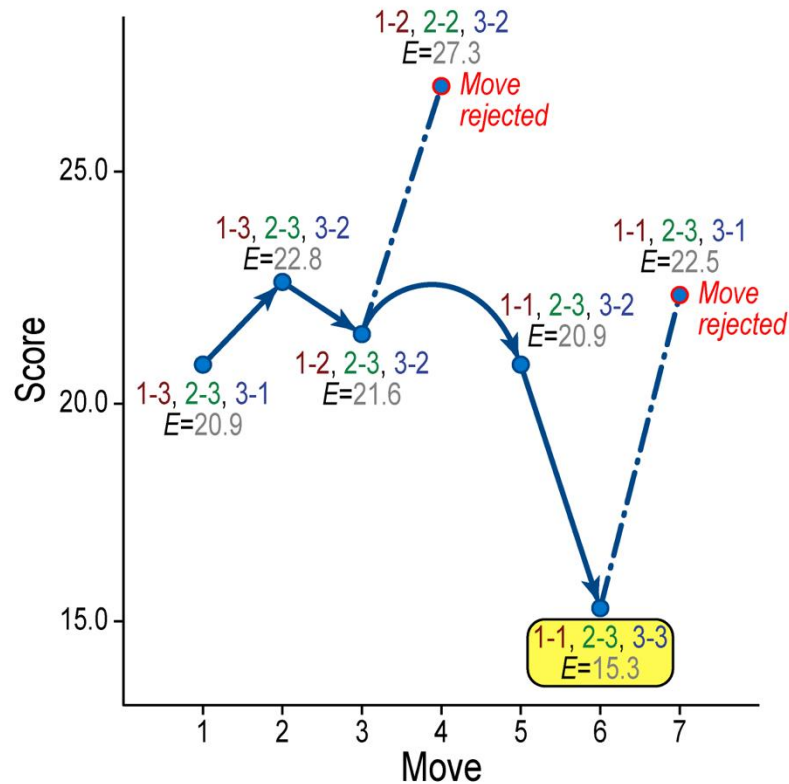
The rotamer optimization (design) problem



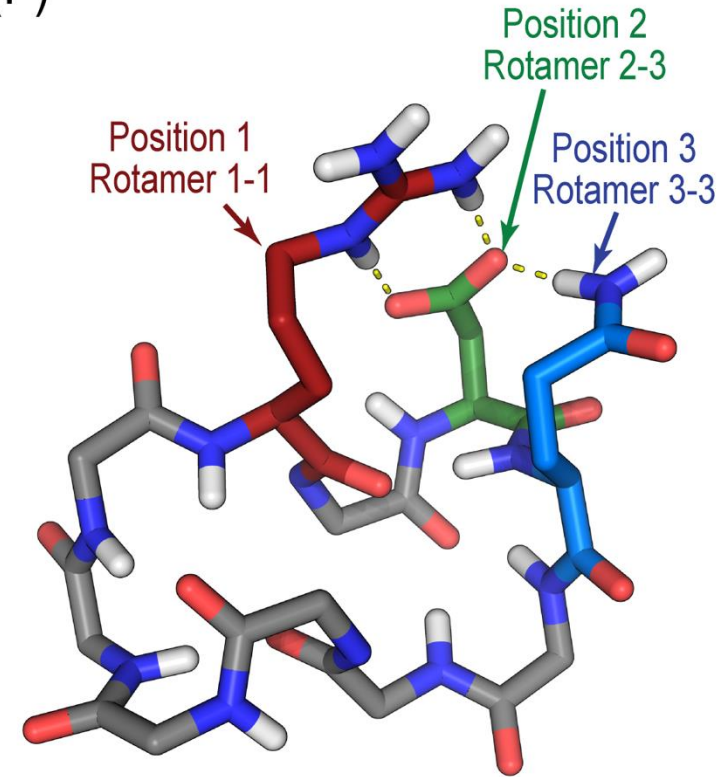
From Mulligan VK and Hosseinzadeh P. (2022) “Computational Design of Peptide-Based Binders to Therapeutic Targets.”
Chapter in *Approaching the Next Inflection in Peptide Therapeutics: Attaining Cell Permeability and Oral Bioavailability*,
Ghodge S. V. *et al.*, eds. Washington DC: American Chemical Society.

The rotamer optimization (design) problem

(E)

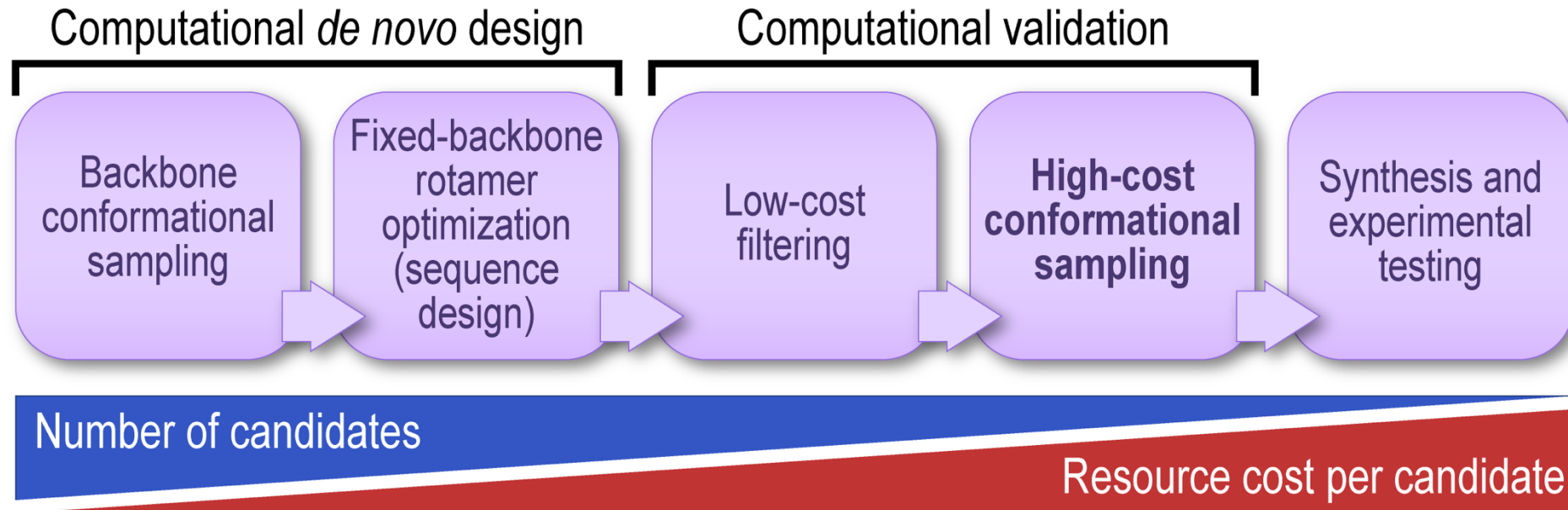


(F)

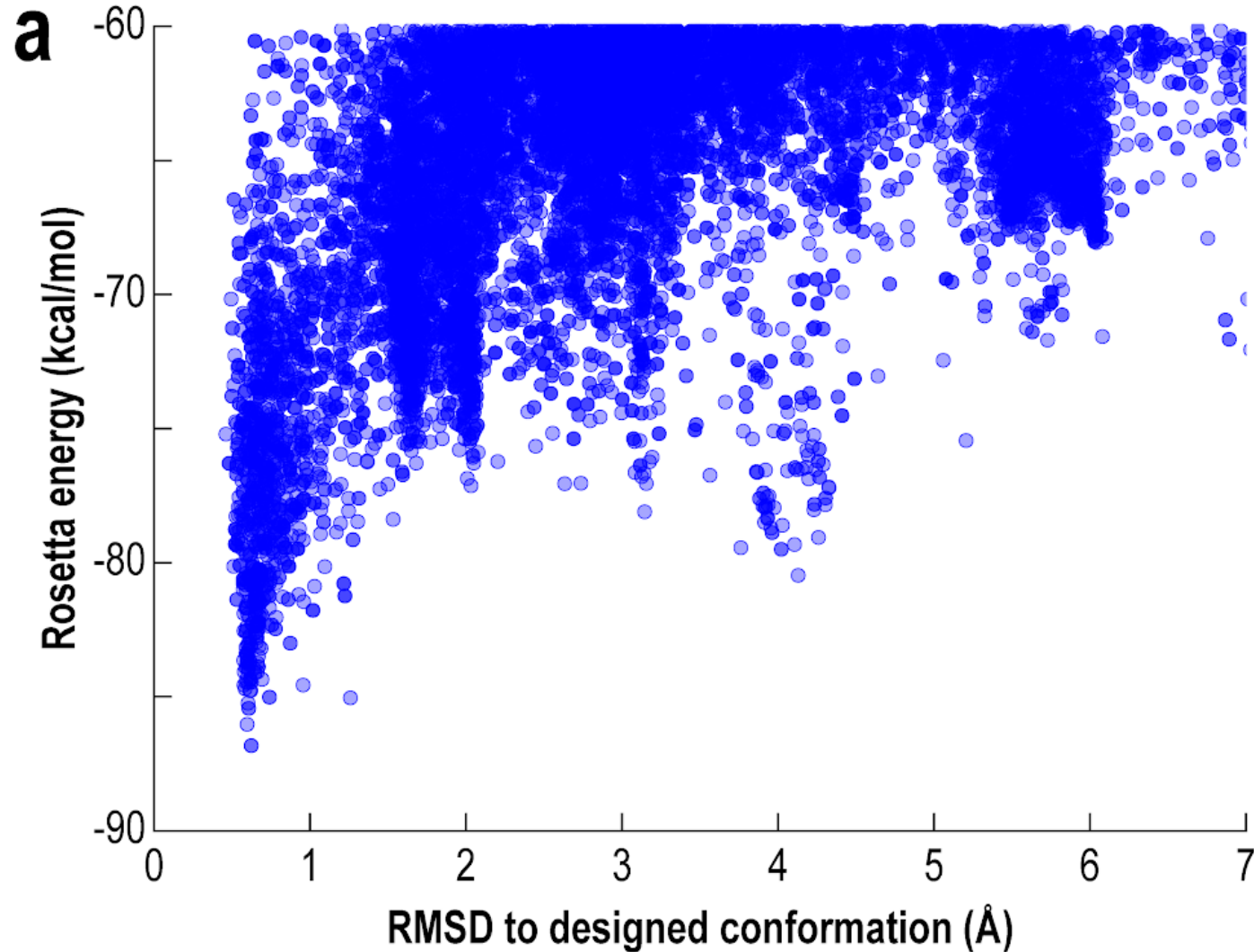


From Mulligan VK and Hosseinzadeh P. (2022) “Computational Design of Peptide-Based Binders to Therapeutic Targets.” Chapter in *Approaching the Next Inflection in Peptide Therapeutics: Attaining Cell Permeability and Oral Bioavailability*, Ghodse S. V. *et al.*, eds. Washington DC: American Chemical Society.

Pipeline for designing rigidly-folded peptide drugs



Validating peptide designs with large-scale conformational sampling (Rosetta's simple_cycpep_predict application)



$$P_{near} = \frac{\sum_{i=1}^N e^{-\frac{RMSD_i^2}{\lambda^2}} e^{-\frac{E_i}{k_B T}}}{\sum_{j=1}^N e^{-\frac{E_j}{k_B T}}}$$

The toolkits: The Rosetta software suite



Home | **Software** | Documentation & Support | Developer Resources | About | Blog

Search

```
density.dimension(grid[0],grid[1],grid[2]);
for (int i=0; i<density.u1()*density.u2()*density.u3(); ++i) density[i]=0.0;
numeric::xyzVector< core::Real > cartX, fracX;
numeric::xyzVector< core::Real > atm_i, atm_j, del_ij;
const core::Real ATOM_MASK_PADDING = 1.5;
for (Size n=1; n<=nposes; ++n) {
  core::pose::Pose &pose = *(poses[n]);
  int nres = pose.total_residue();
  for (int i=1 ; i<=nres; ++i) {
    conformation::Residue const &rsd_i (pose.residue(i));
```

A unique partnership between universities, government laboratories, institutes, research centers, and partner corporations

Software

License and Download

Ways to Use

Documentation

Release Notes

Related Projects

Servers

Home

Overview

The Rosetta software suite includes algorithms for computational modeling and analysis of protein structures. It has enabled notable scientific advances in computational biology, including de novo protein design, enzyme design, ligand docking, and structure prediction of biological macromolecules and macromolecular complexes.

Rosetta is available to all non-commercial users for free and to commercial users for a fee. [License Rosetta](#) to get started.

Rosetta development began in the laboratory of Dr. David Baker at the University of Washington as a structure prediction tool but since then has been adapted to solve common computational macromolecular problems.

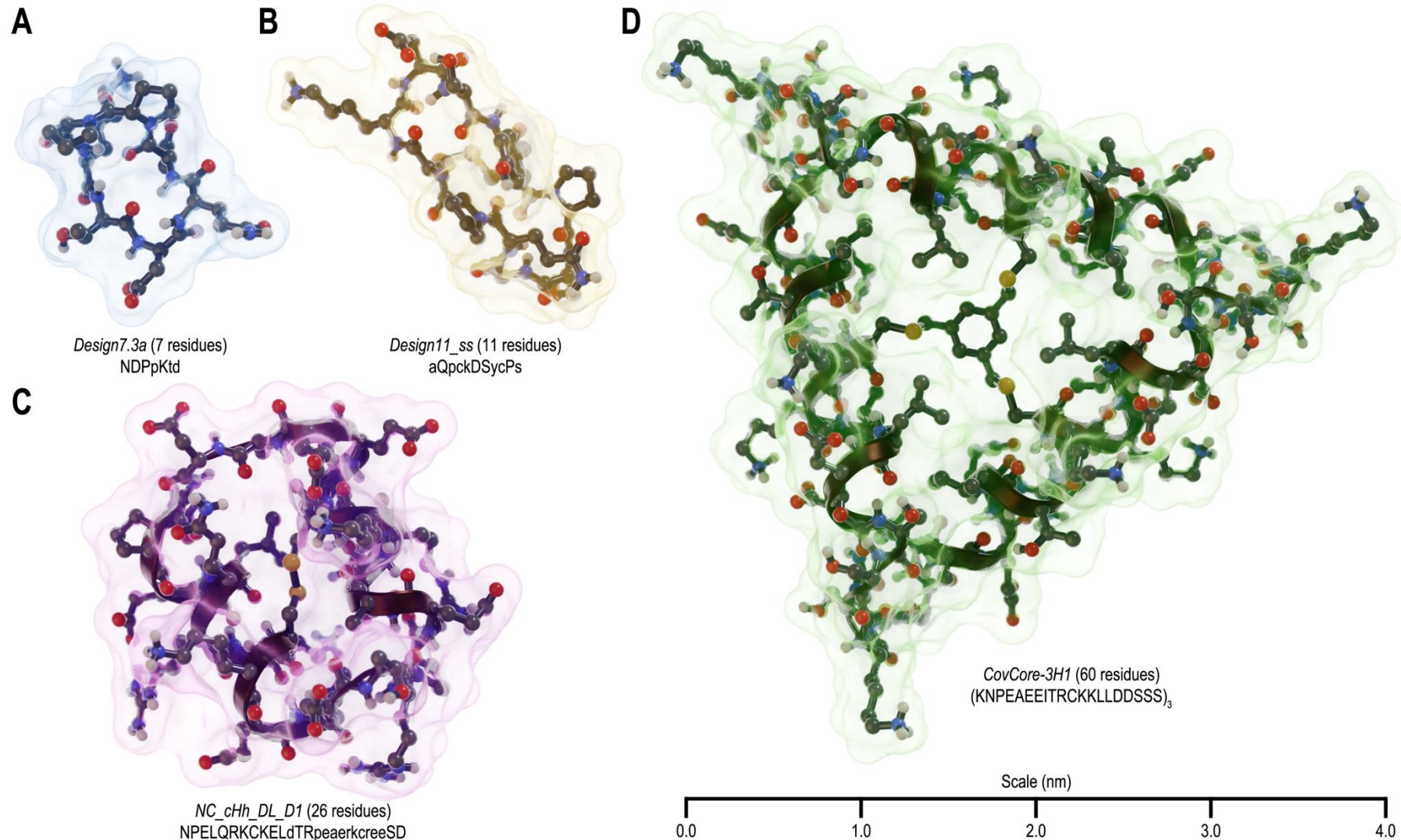
Development of Rosetta has moved beyond the University of Washington into the [members of RosettaCommons](#), which include government laboratories, institutes, research centers, and partner corporations.

The Rosetta community has many goals for the software, such as:

- Understanding macromolecular interactions
- Designing custom molecules
- Developing efficient ways to search conformation and sequence space
- Finding a broadly useful energy functions for various biomolecular representations

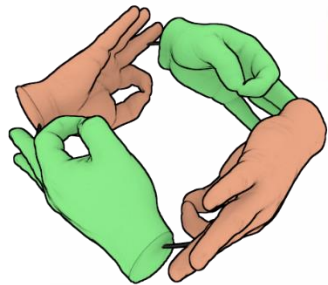
- Rosetta is protein modelling software that has been generalized for more exotic macromolecules.
- The software is free for academics, nonprofits, and governments, and is licenced for a fee for commercial use.
- Originally started in David Baker's lab, Rosetta is now developed and maintained by more than 70 labs in many countries.

Synthetic peptides designed to fold into rigid structures with the Rosetta software suite

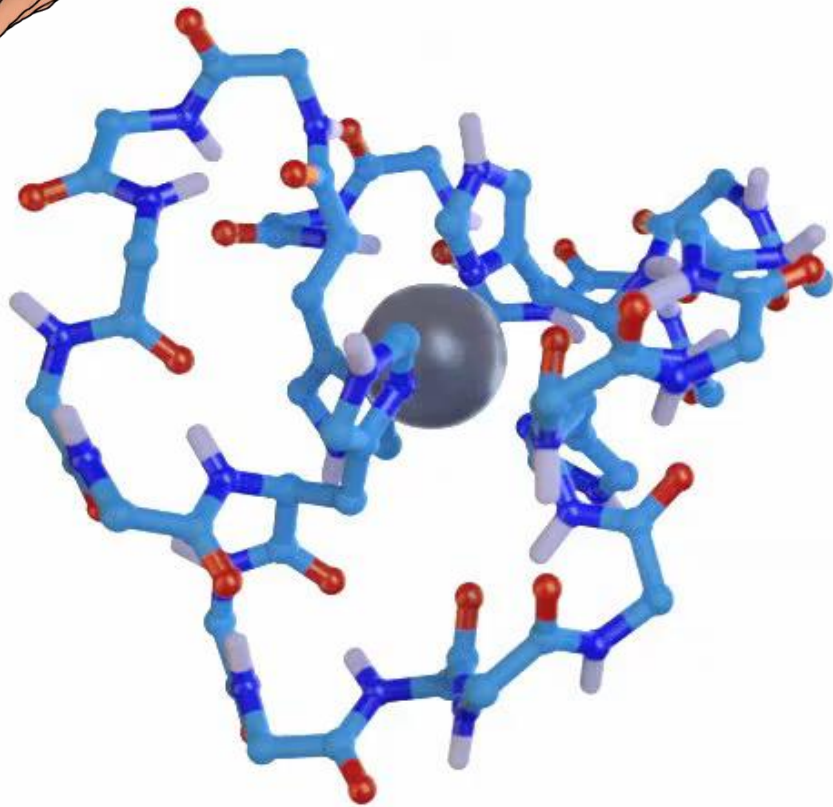


From Bhardwaj, Mulligan, Bahl *et al.* (2016) *Nature* 538(7625):329-35; Hosseinzadeh, Bhardwaj, Mulligan *et al.* (2017). *Science* 358(6369):1461-6; Dang, Wu, Mulligan *et al.* (2017). *Proc Natl Acad Sci USA* 114(41):10852-7.

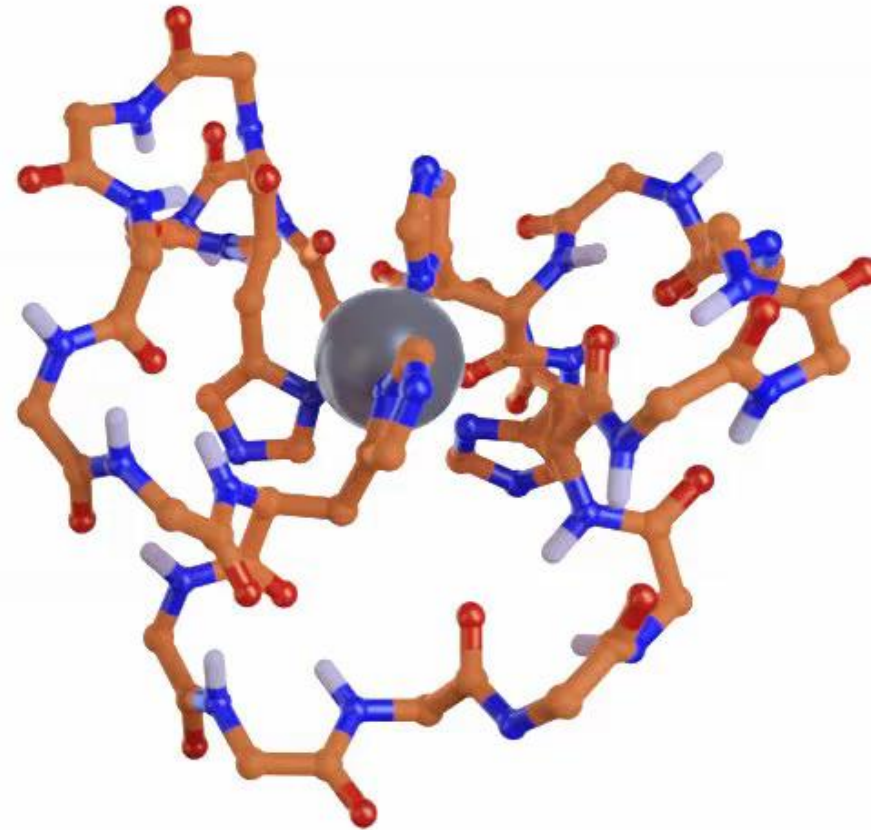
Crystal structure of a designed S₄-symmetric mixed-chirality Zn-binding peptide



Design model

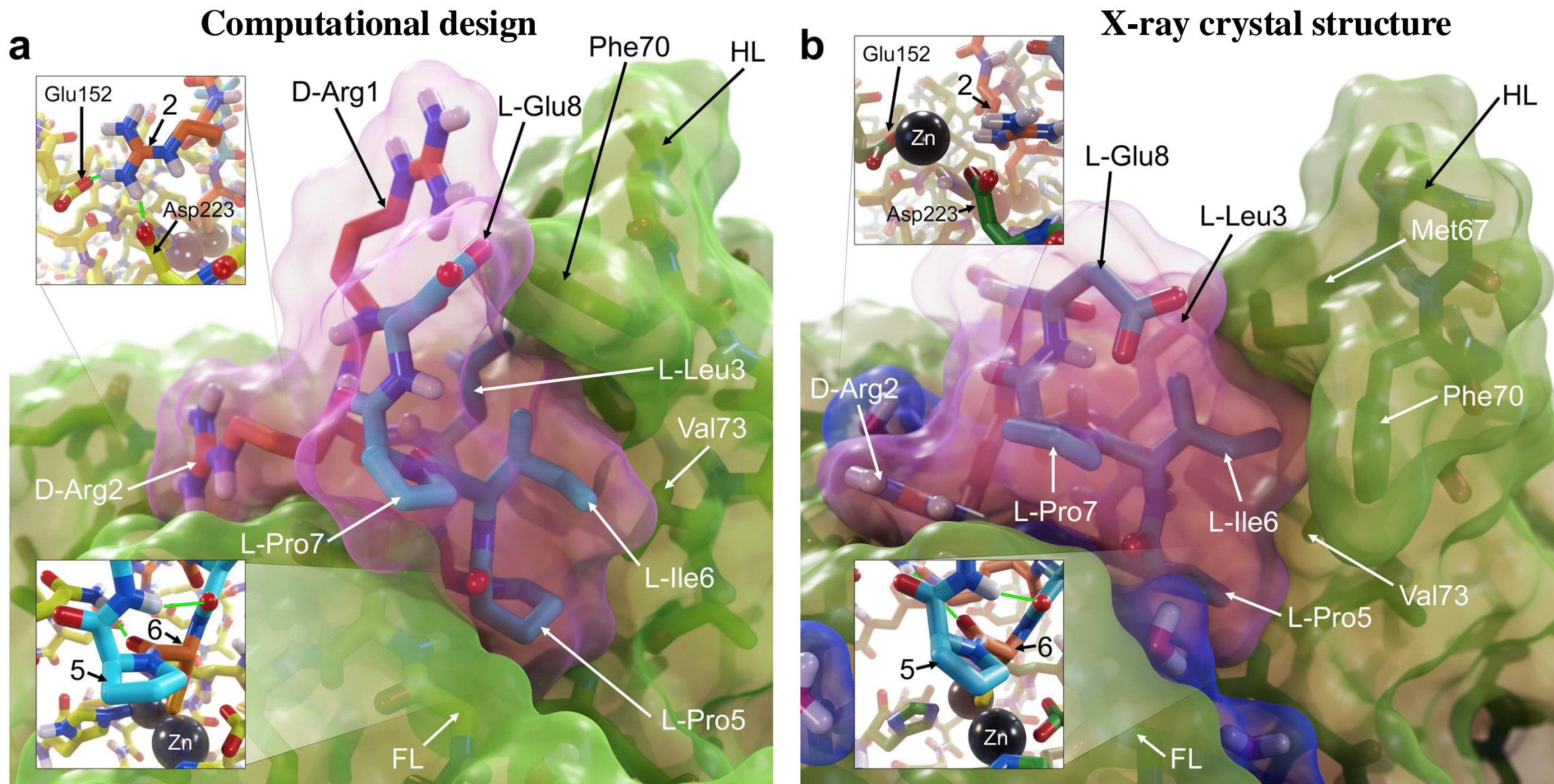


Crystal structure



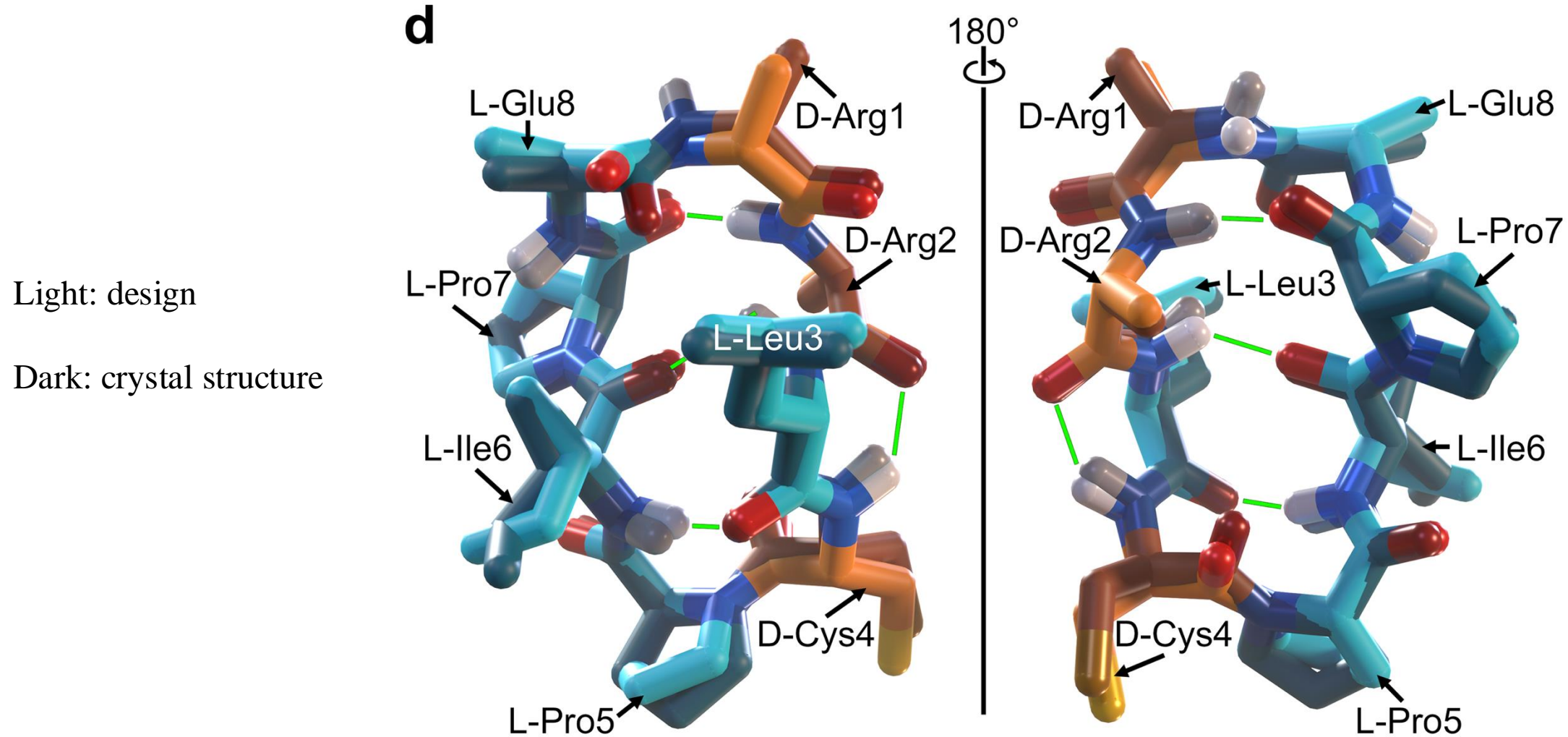
From Mulligan, Kang, Sawaya *et al.* (2021.) Computational design of mixed chirality peptide macrocycles with internal symmetry. *Protein Sci.* 29(12):2433-45. DOI: 10.1002/pro.3974.

A designed inhibitor of the New Delhi metallo- β -lactamase 1 (NDM-1)



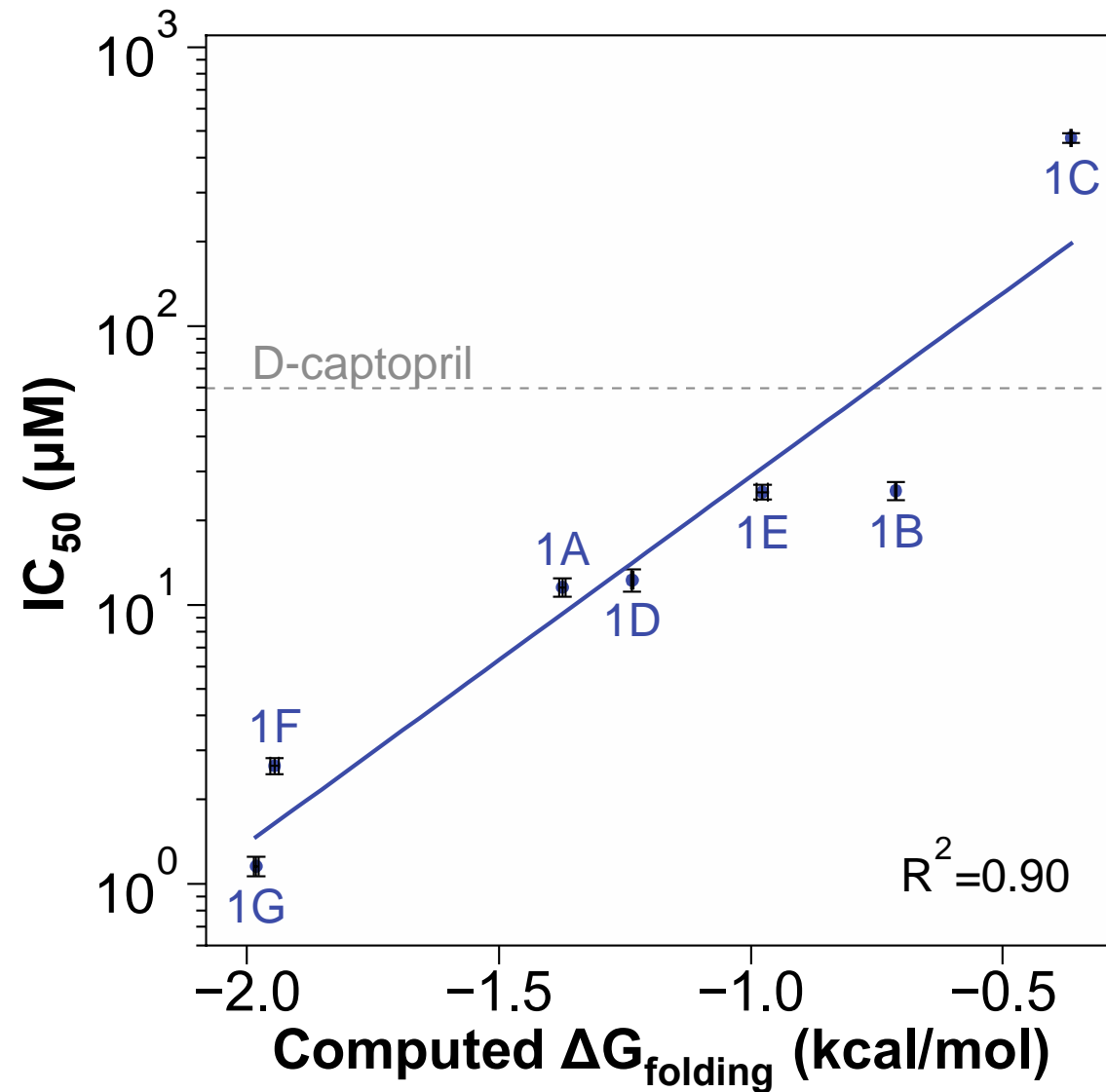
From Mulligan *et al.* (2021). Computationally-designed peptide macrocycle inhibitors of New Delhi metallo- β -lactamase 1. *Proc Natl Acad Sci USA* 118(12):e2012800118. DOI: 10.1073/pnas.2012800118.

A designed inhibitor of the New Delhi metallo- β -lactamase 1 (NDM-1)



From Mulligan *et al.* (2021). Computationally-designed peptide macrocycle inhibitors of New Delhi metallo- β -lactamase 1. *Proc Natl Acad Sci USA* 118(12):e2012800118. DOI: 10.1073/pnas.2012800118.

Simulations predict success in experiments



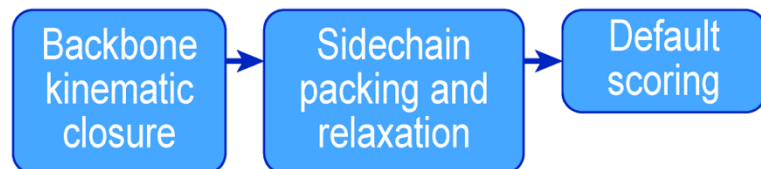
From Mulligan *et al.* (2021). Computationally-designed peptide macrocycle inhibitors of New Delhi metallo- β -lactamase 1. *Proc Natl Acad Sci USA* 118(12):e2012800118. DOI: 10.1073/pnas.2012800118.

Outline

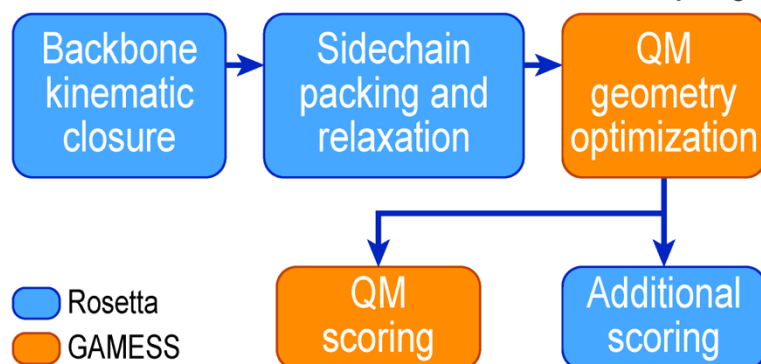
1. Peptide and protein drug design on classical computers.
2. Enhancing peptide and protein drug design with quantum chemistry calculations (on classical computers).
3. Enhancing peptide and protein drug design with quantum computers

RosettaQM-based prediction of the structure of cyclosporine A in organic solvent

A Traditional macrocycle conformational sampling:

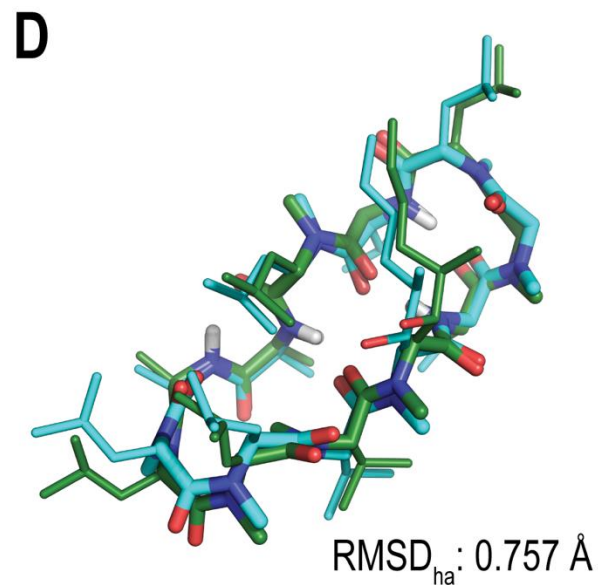
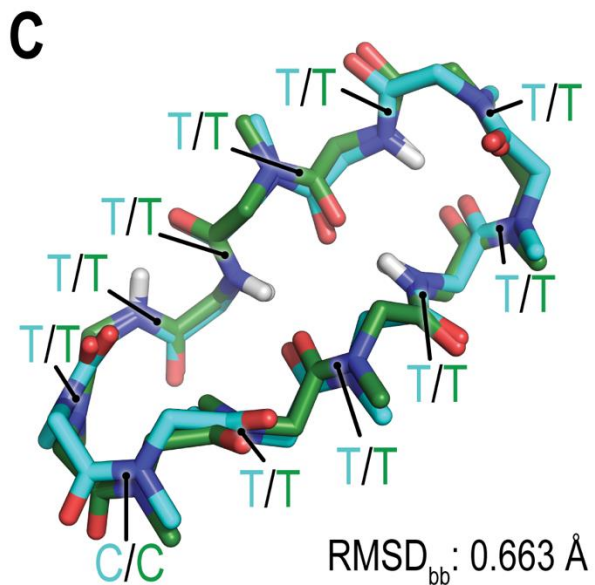
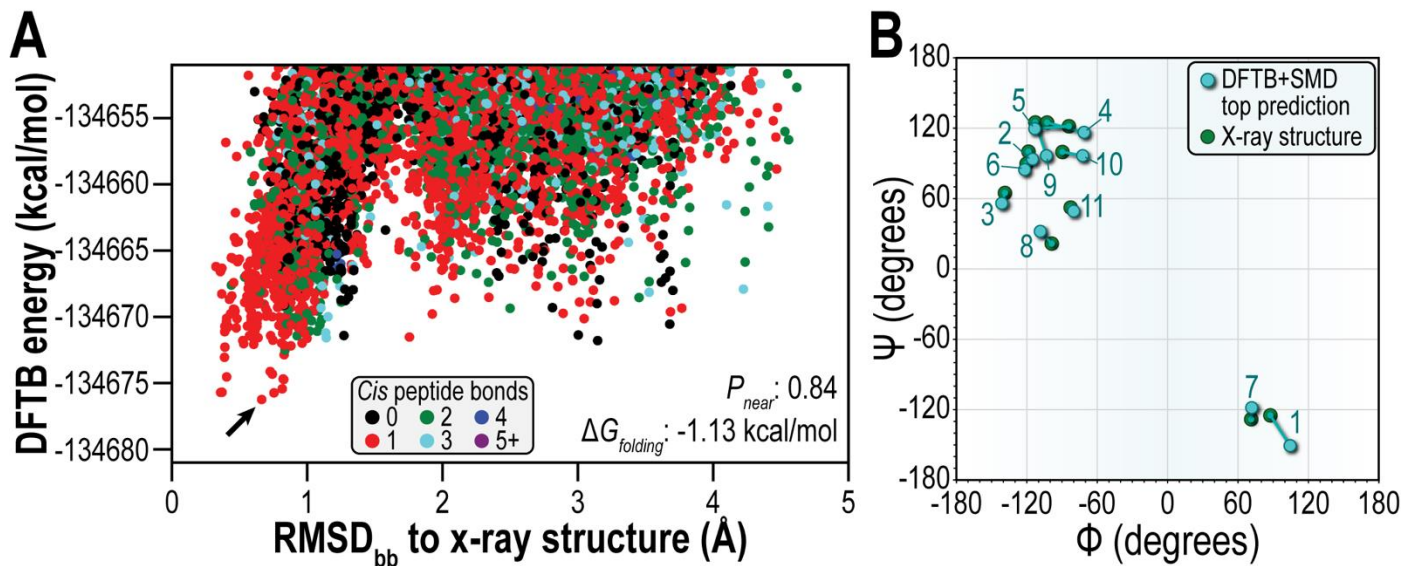


RosettaQM-enhanced conformational sampling:



With Benjamin Brown, P. Douglas Renfrew, Chris Jurich, Nancy Hernandez, and Bargeen Turzo.

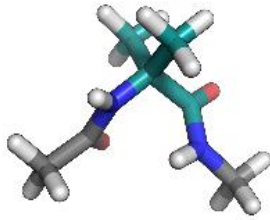
Improving Accuracy and Generality with QM Energy Calculations



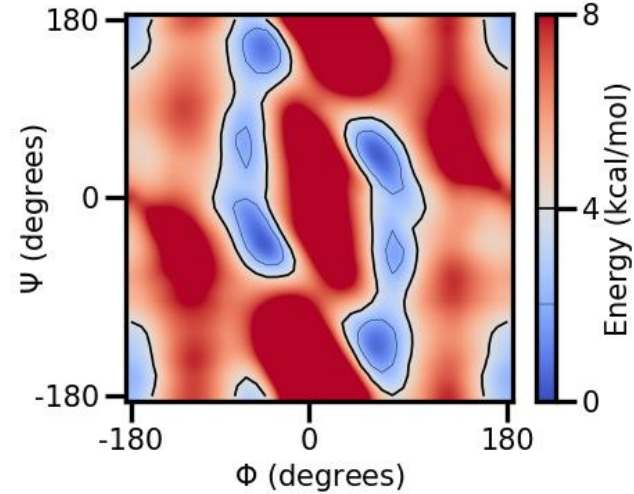
With Benjamin Brown, P. Douglas Renfrew, Chris Jurich, Nancy Hernandez, and Bargeen Turzo.

Prediction of the Ramachandran map: RosettaQM-based parameterization of noncanonical force fields

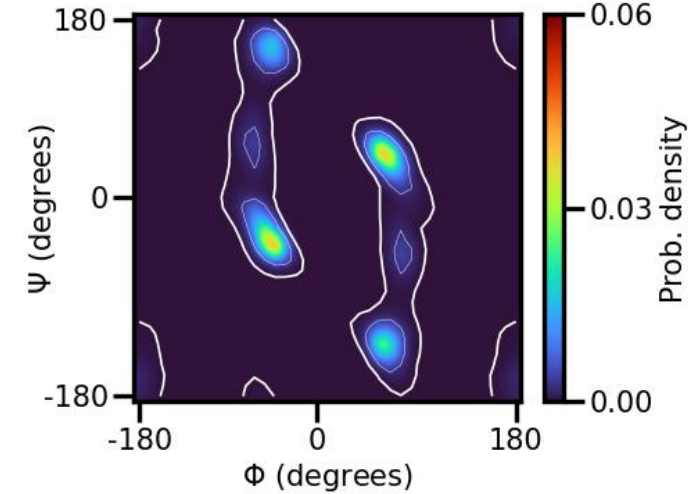
**2-Amino
isobutyric acid**



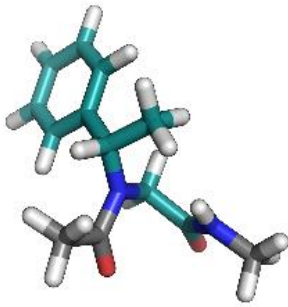
**Energy
(6-311++G**/MP2/SMD)**



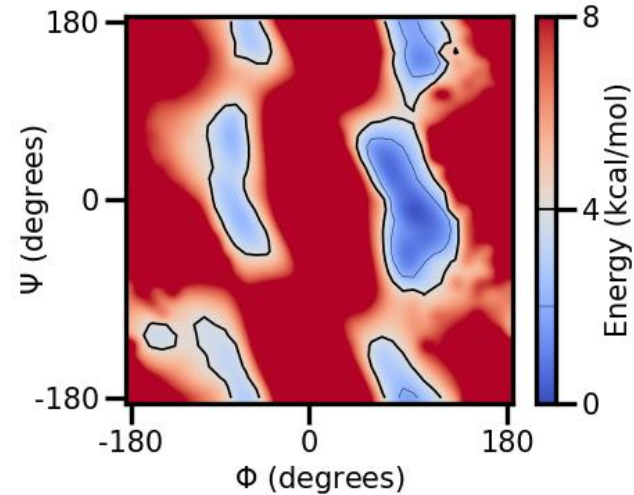
**Probability
(6-311++G**/MP2/SMD)**



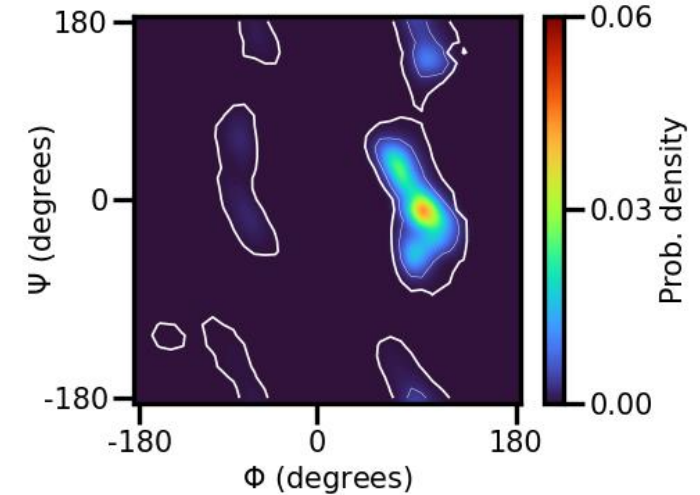
**Peptoid
Monomer 601**



**Energy
(6-311++G**/MP2/SMD)**



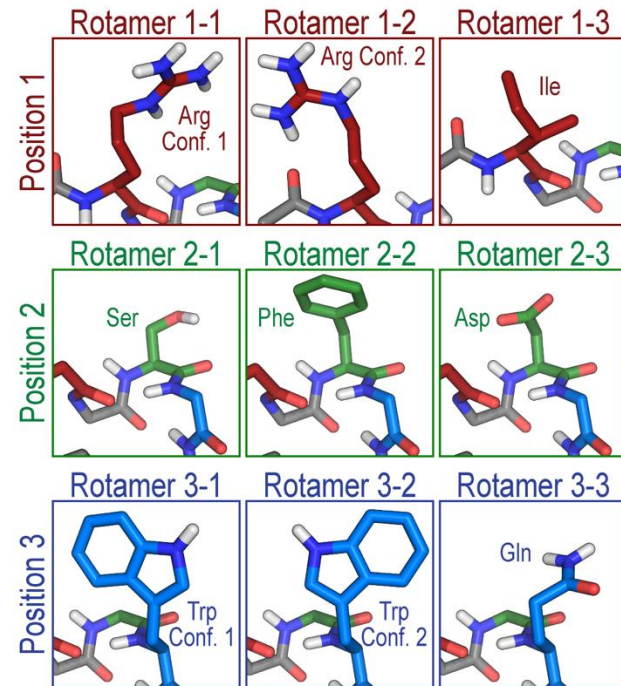
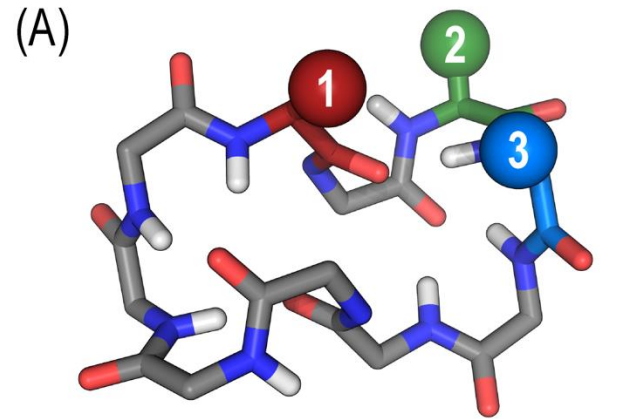
**Probability
(6-311++G**/MP2/SMD)**



Outline

1. Peptide and protein drug design on classical computers.
2. Enhancing peptide and protein drug design with quantum chemistry calculations (on classical computers).
3. Enhancing peptide and protein drug design with quantum computers

Revisiting the rotamer optimization (design) problem



(B)

$$E = \sum_{i=1}^N \alpha_i + \sum_{j=1}^{N-1} \sum_{k=2}^N \beta_{jk}$$

N - Number of designable positions

α_i - Internal energy of selected rotamer at position i

β_{jk} - Interaction energy of selected rotamers at positions j and k

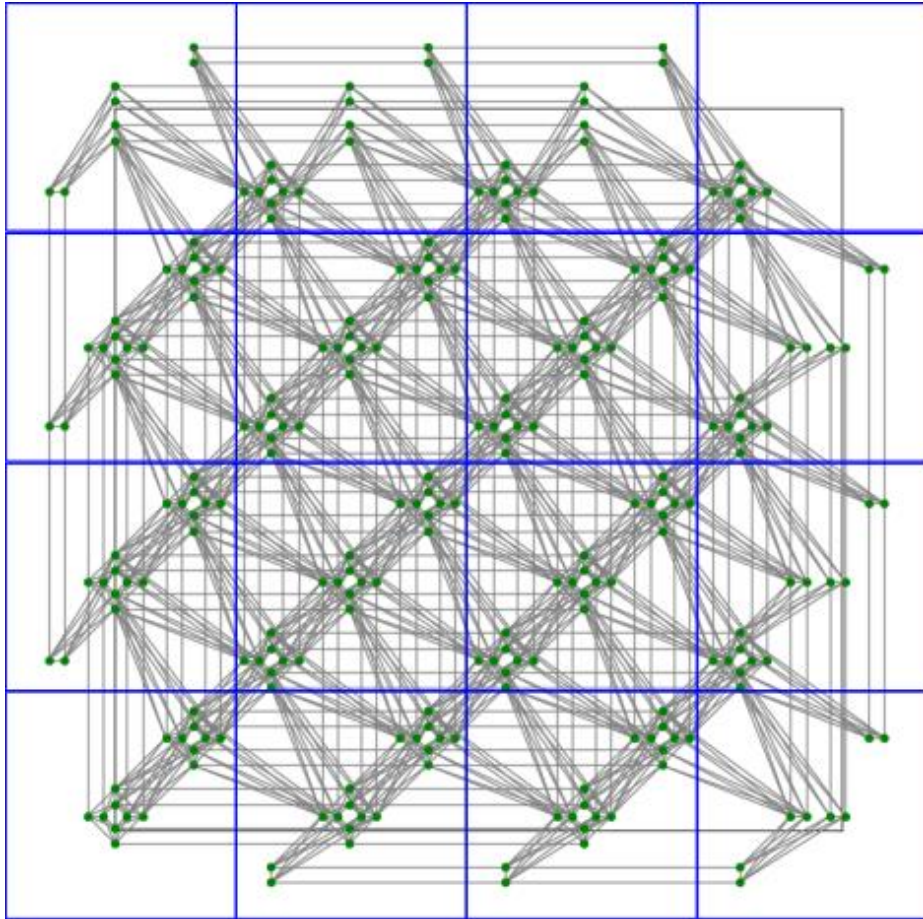
Single rotamer internal energy

1-1	5.2
1-2	3.3
1-3	7.1
2-1	2.1
2-2	6.8
2-3	3.1
3-1	3.1
3-2	3.5
3-3	1.5

Rotamer pair interaction energy

		Second rotamer					
		2-1	2-2	2-3	3-1	3-2	3-3
First rotamer	2-1	5.3	1.6	0.7	7.9	4.3	4.1
	1-2	5.8	4.8	2.3	1.3	5.3	5.8
	1-3	3.1	3.5	3.7	1.4	1.3	1.1
2-3	2-1				3.0	3.1	2.9
	2-2				3.0	3.6	1.7
	1-1				2.5	4.1	0.7

The D-Wave Advantage adiabatic quantum annealer



- The D-Wave Advantage offers about 5,000 sparsely-connected physical qubits. Each is connected to 15 others. This can emulate 177 fully-connected virtual qubits.
- The user provides inputs by setting single-qubit biases (h_i) for each qubit and two-qubit couplings ($J_{i,j}$) for each pair of qubits.
- The total energy of a given state of the computer is:

$$E = \sum_{i=1}^Q q_i h_i + \sum_{i=2}^Q \sum_{j=1}^{i-1} q_i q_j J_{i,j}$$

- In the above, Q is the number of qubits, and q_i and q_j are the value (0 or 1) of the i^{th} and j^{th} qubit (defining the *state*). The annealing process returns as output values q_i for all qubits such that E is a minimum.

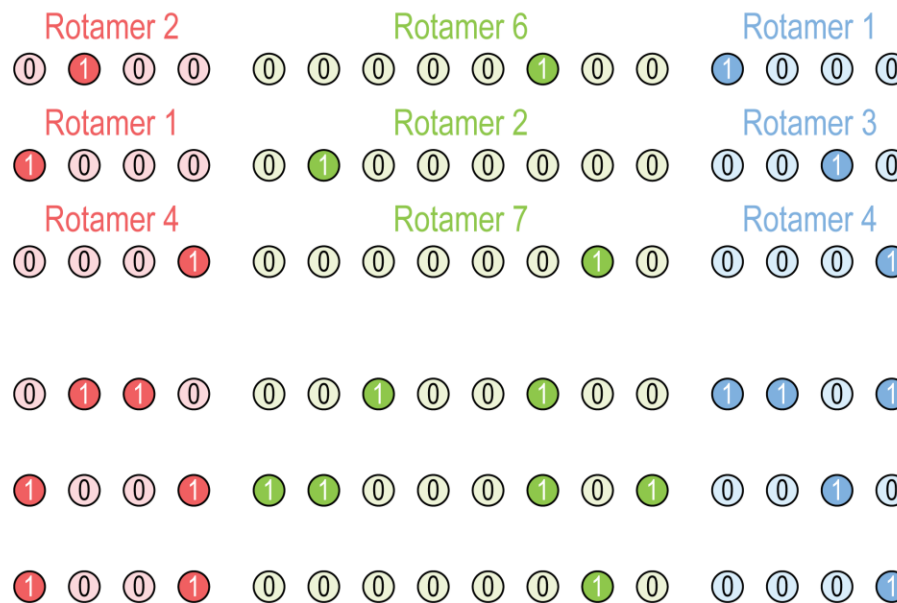
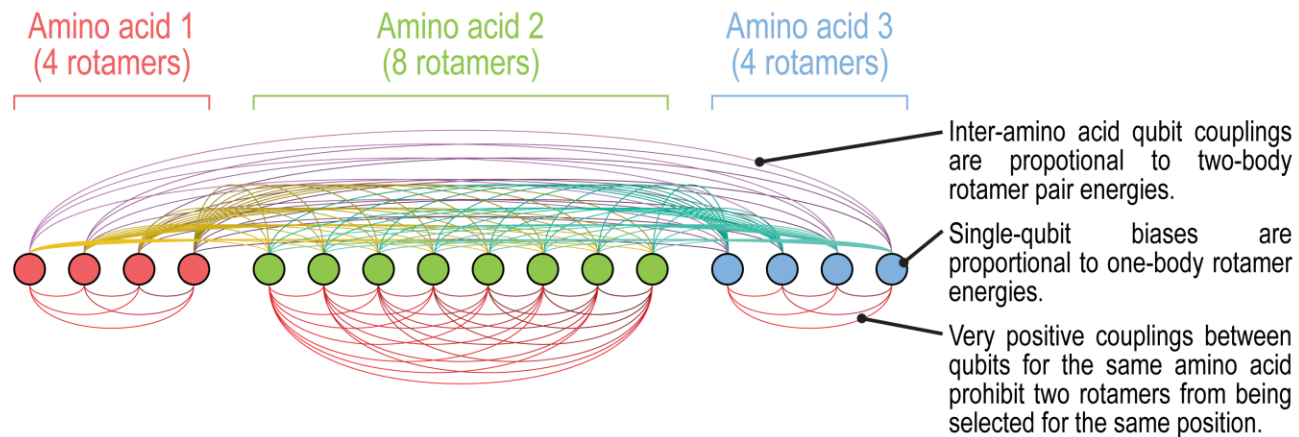
Designing peptides using a quantum annealer: QPacker



In collaboration with
Hans Melo, CEO,
Menten AI



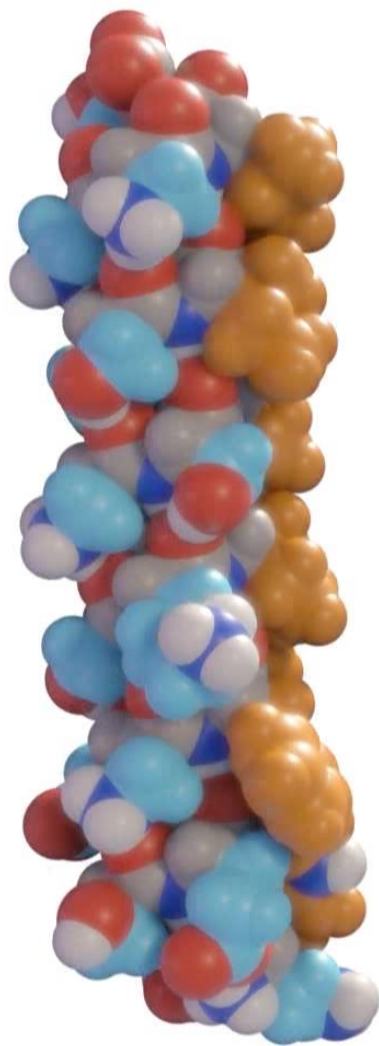
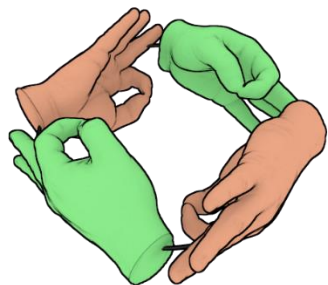
In collaboration with
Brian Weitzner,
Principal Scientist,
Outpace Bio



Examples of valid states that correspond to possible solutions to the packing problem, each with a single selected rotamer.

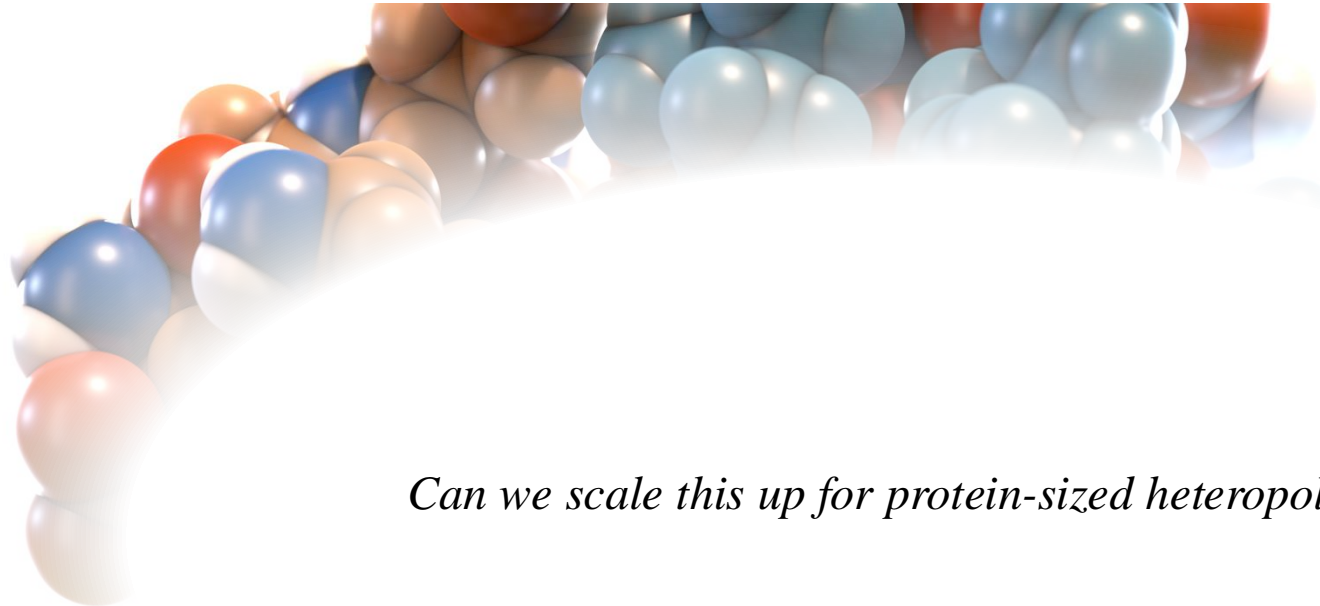
Examples of prohibited (nonsensical) states that are invalid solutions to the packing problem. Since more than one rotamer is selected, these have no interpretation.

Designing heterochiral helical bundles with QPacker and the D-Wave 2000Q



With Michael Sawaya, Todd Yeates, Parmjit Arora, Haley Irene Merritt, and Hans Melo.

Self-assembling peptides designed with QPacker, with structures confirmed experimentally by x-ray crystallography

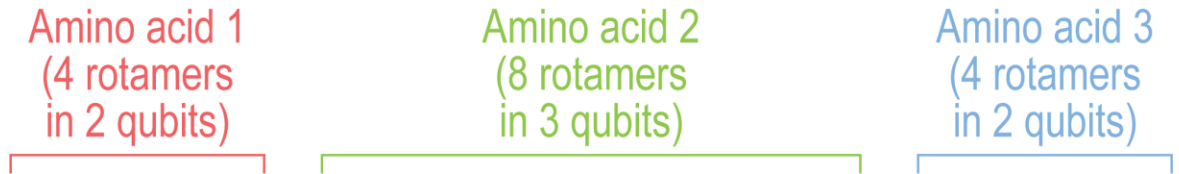


Can we scale this up for protein-sized heteropolymers?



With Michael Sawaya, Todd Yeates, Parmjit Arora, Haley Irene Merritt, and Hans Melo.

QPacker-B: Compressing problems to use $N \log_2 D$ qubits

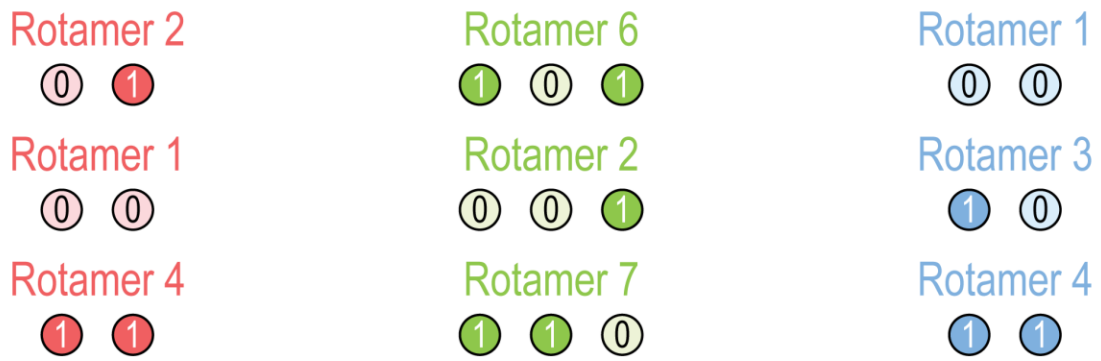


Single-qubit biases and two-qubit couplings for qubits for pairs of amino acids approximate all twobody rotamer pair energies.

Single-qubit biases and two-qubit couplings for a single amino acid's qubits approximate all onebody rotamer energies.



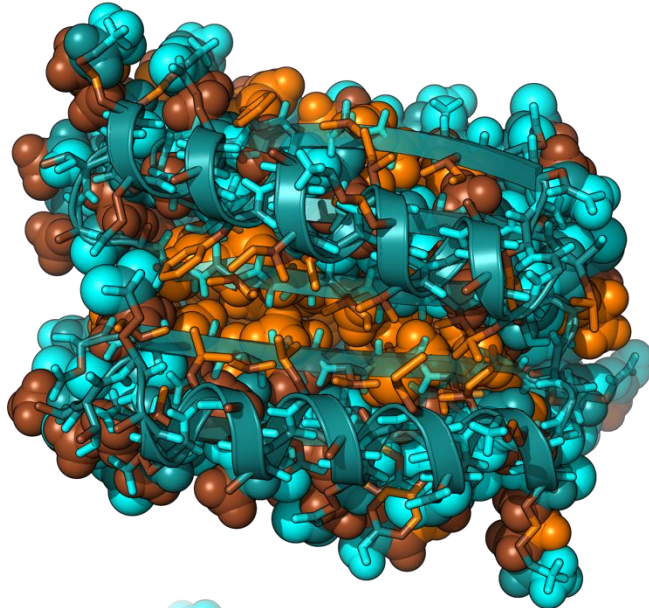
Tristan Zaborniak, U. Victoria



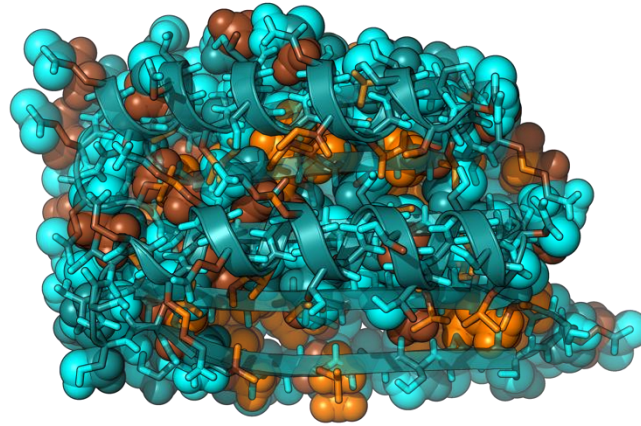
Example binary bitstrings representing valid possible solution states. Since every bitstring corresponds to a possible solution, there are no invalid states.

Full proteins designed on the D-Wave Advantage 6.4 Quantum Annealer

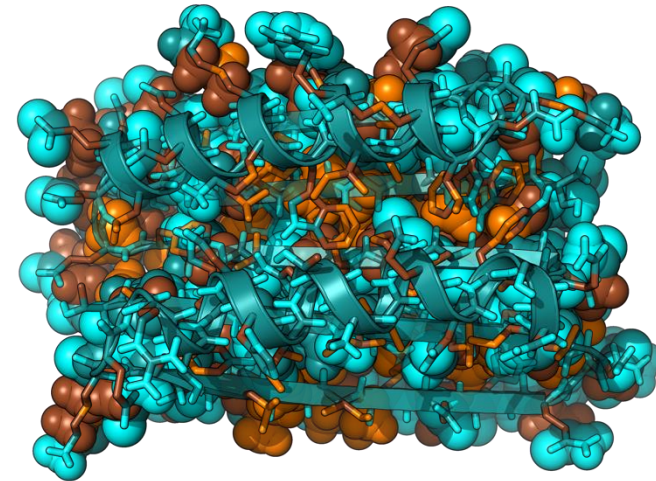
Top7 (classically designed in 2003)



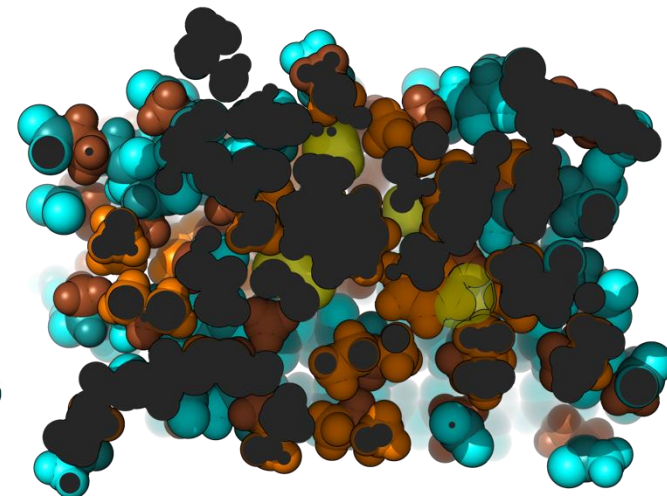
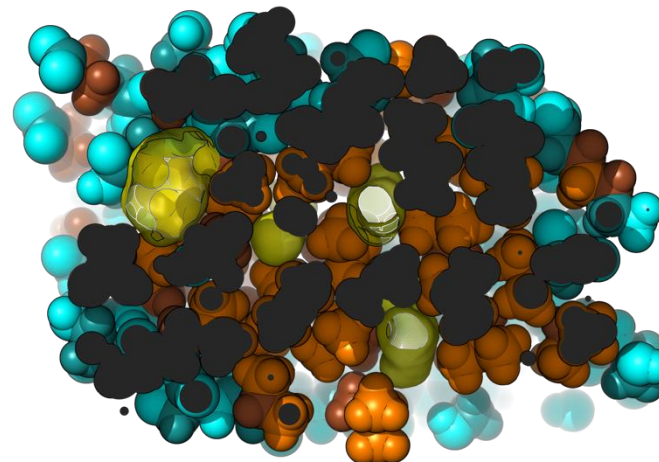
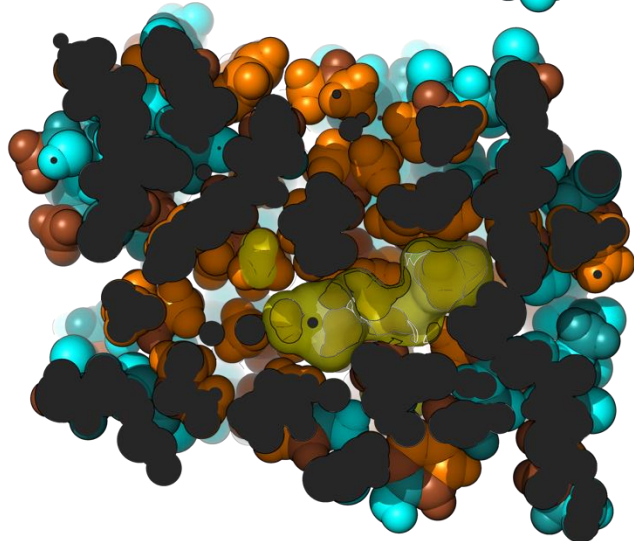
First quantum-designed protein
(15 QPU-seconds, Nov. 2024)



Second quantum-designed protein
(15 QPU-seconds, Nov. 2024)



Tristan
Zaborniak,
U. Victoria



Making enhanced classical and quantum peptide design available for everyone: The Masala software library



- Masala is a free and open-source successor to Rosetta under development at the Flatiron institute.
- It is structured to take full advantage of modern massively-parallel CPU and GPU hardware.
- It has a versatile plugin architecture permitting easy extensibility. Our QPU plugin permits design on quantum computers.
- It is intended to be used as standalone software *or* as a library in other projects. (Rosetta, for instance, can link Masala for high-efficiency design calculations.)
- *To be released shortly.*



P. Douglas
Renfrew, CCB



Noora Azadvari,
U. Oregon



Qiyao Zhu, CCB



S. M. Bargeen
Alam Turzo, CCB



Tristan Zaborniak,
U. Victoria

Acknowledgements

Rosetta senior developers:

- Andrew Leaver-Fay
- Sergey Lyskov
- Julia Koehler-Leman
- **P. Douglas Renfrew**
- Andrew M. Watkins
- Jason Labonte
- Jared Adolf-Bryfogle
- Steven M. Lewis
- Rocco Moretti

Biomolecular Design Group:

- **P. Douglas Renfrew**
- Bargeen Turzo
- Qiyao Zhu

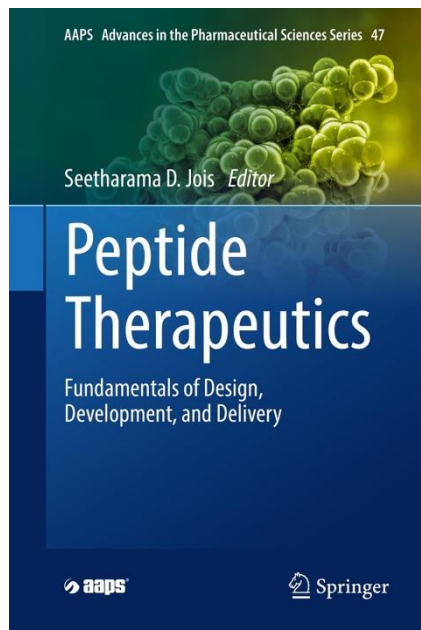
Interns and Visiting Scholars:

- Ekaterina Maximova
- **Tristan Zaborniak**
- Noora Azadvari
- Andrew Powers
- Allon Goldberg
- Rutika Patel

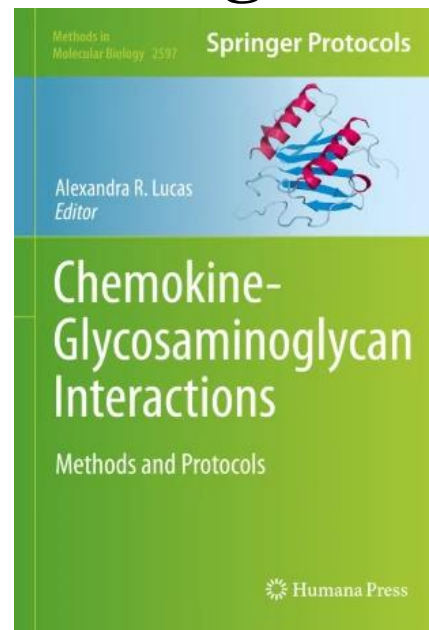
RosettaQM collaborators:

- **Benjamin Brown**
- Chris Jurich

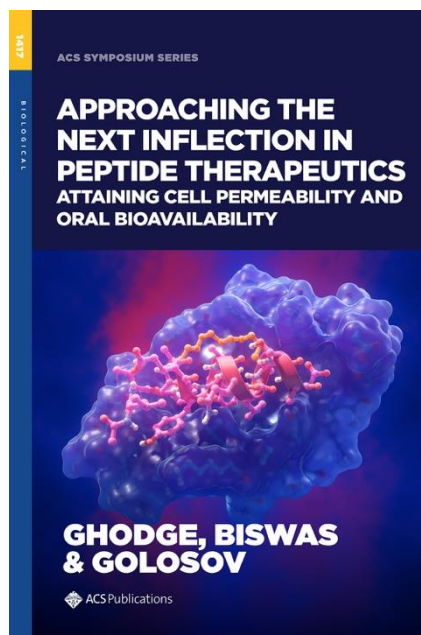
Further reading



V.K. Mulligan. Computational Methods for Peptide Macrocyclic Drug Design. Chapter in: S.D. Jois (Ed.), *Peptide Therapeutics: Fundamentals of Design, Development, and Delivery*, Springer International Publishing, New York, 2022: pp. 79–161.
https://doi.org/10.1007/978-3-031-04544-8_3.



J. Dodd-O, A.M. Acevedo-Jake, A.-R. Azizoglu, V.K. Mulligan, V.A. Kumar, *How to Design Peptides*, *Methods Mol Biol* 2597 (2023) 187–216.
https://doi.org/10.1007/978-1-0716-2835-5_15.

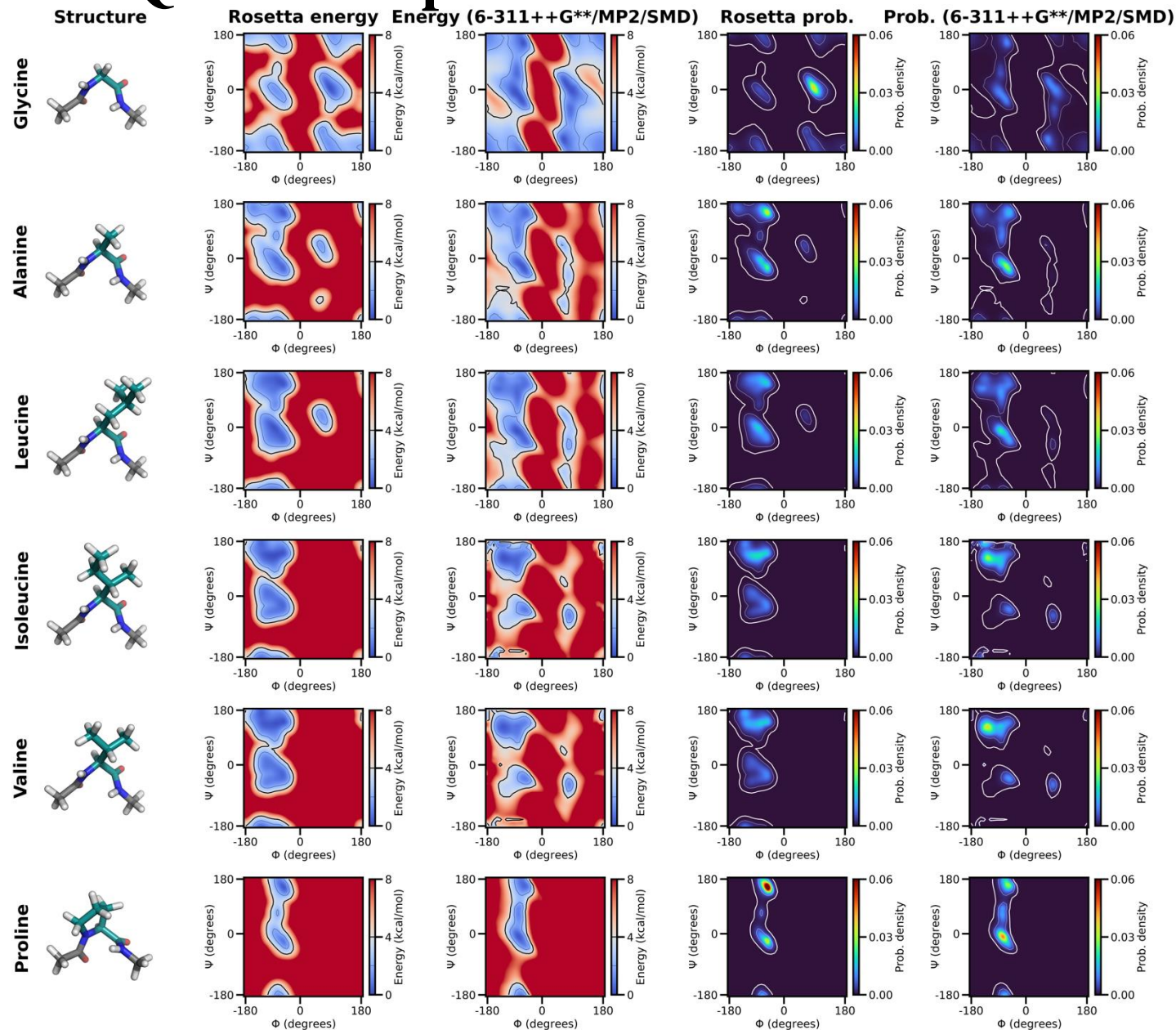


V.K. Mulligan, P. Hosseinzadeh. Computational Design of Peptide-Based Binders to Therapeutic Targets. Chapter in: S.V. Ghodse, K. Biswas, A.A. Golosov (Eds.), *Approaching the Next Inflection in Peptide Therapeutics: Attaining Cell Permeability and Oral Bioavailability*, American Chemical Society, Washington, DC, 2022: pp. 55–102.
<https://doi.org/10.1021/bk-2022-1417.ch003>.

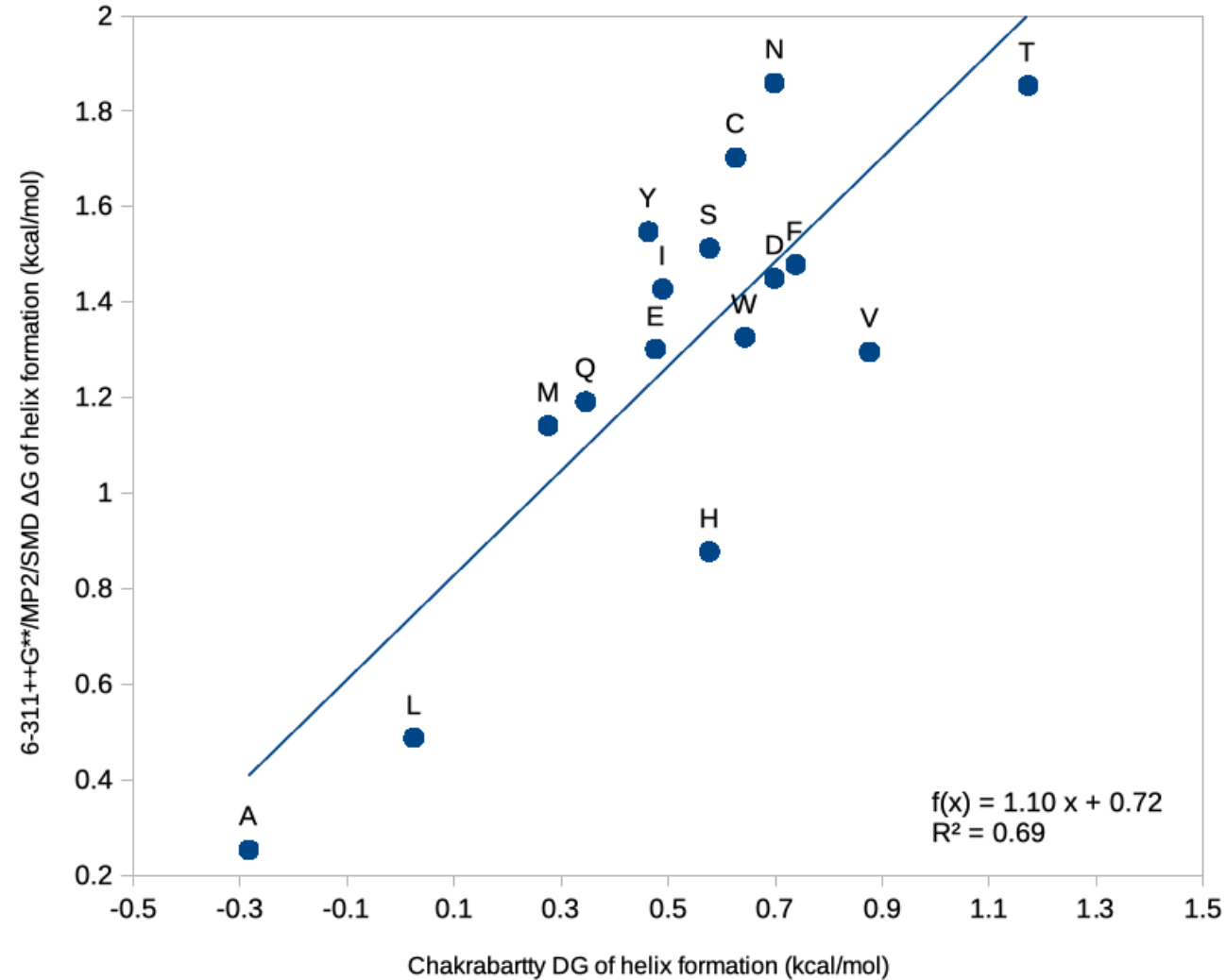
Or e-mail us: the Biomolecular Design Group, in the Flatiron Institute's Center for Computational Biology, is always looking for experimental collaborators.

vmulligan@flatironinstitute.org
pdrenfrew@flatironinstitute.org

Prediction of the Ramachandran map: RosettaQM-based production of better force fields



Prediction of the Ramachandran map: RosettaQM-based production of better force fields



Comparison to Chakrabarty, Kortemme, and Baldwin. (1994).
Protein Sci. 3(5)843-52. DOI: 10.1002/pro.5560030514.

How many qubits *should* the QPacker use?

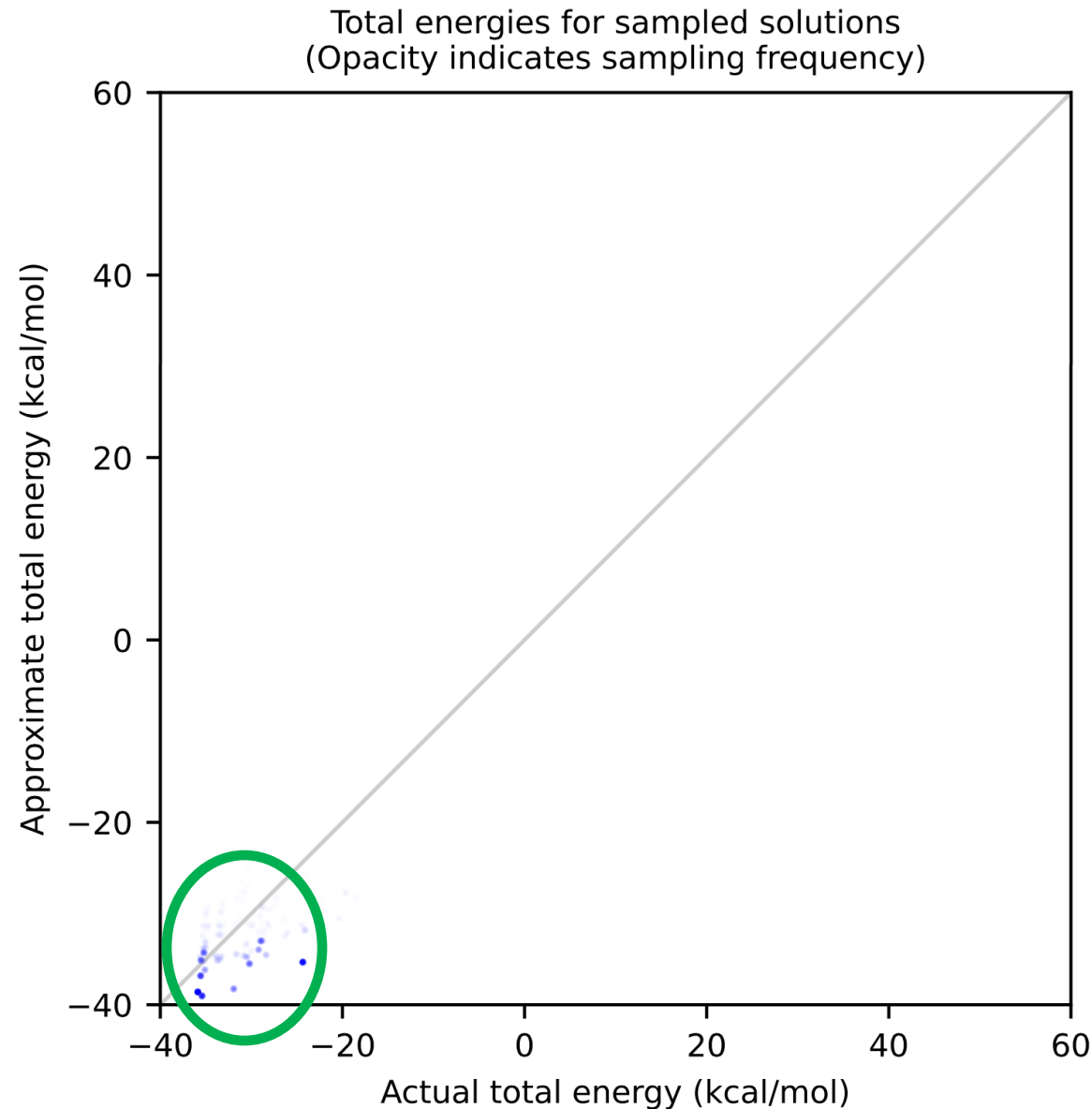
- A packing problem with N designable positions and D rotamers per position has D^N possible solutions.
- A register of Q qubits can exist in a superposition of 2^Q states (2^Q bitstrings). If we use qubits efficiently, then each bitstring will map to a unique solution.
- Let $2^Q = D^N$. Then $Q = N \log_2 D$.
- But we're using ND qubits, not $N \log_2 D$. One-hot encoding of rotamer selections is very inefficient. Can we do better?

QPacker-B: A classical approximation to compress one-body and two-body energies into fewer qubits (16-qubit example)

Actual samples on D-Wave (10,000 attempts)

Blue points: allowed solutions to design problem, each representing a rotamer selection

Red points: duplicates (since D is not an exact power of 2). We want to prohibit these.



Tristan
Zaborniak