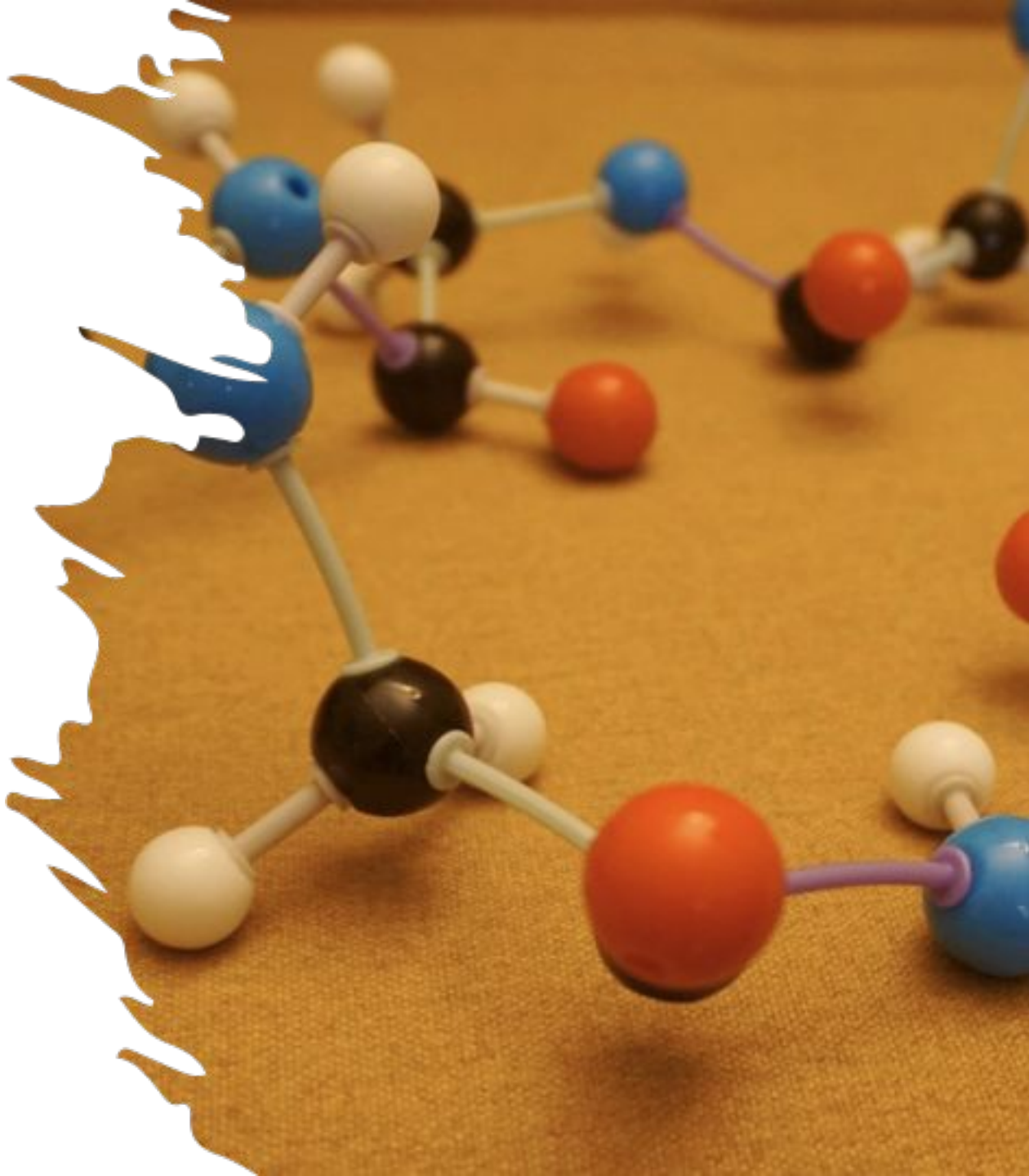
Heuristic energy-based cyclic peptide design

Qiyao Zhu (NYU, now Simons Foundation),

Vikram Mulligan (Simons Foundation), and

Dennis Shasha

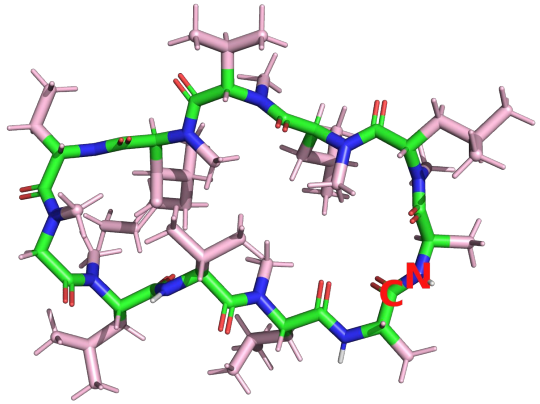
(New York University)



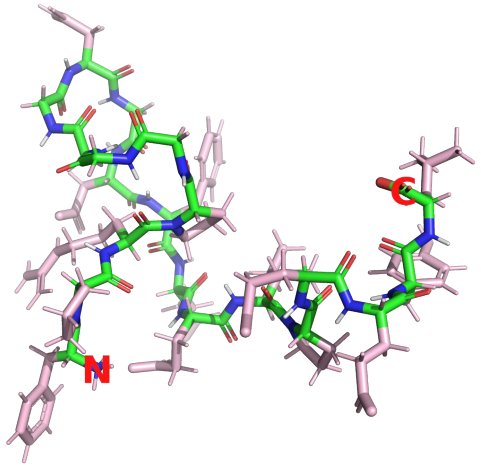
Why Cyclic Peptides?

- **Definition:** A chain of 7+ amino acids (residues), with the N- and C-termini connected to form a closed loop
- **Why good:**
  - More rigid conformations and resistance to degradation compared to linear peptides
  - Potentially superior binding affinity and selectivity compared to small molecules
  - Immune system (e.g. exopeptidase) doesn't recognize them as well as linear peptide (especially when mixing L- with D-).

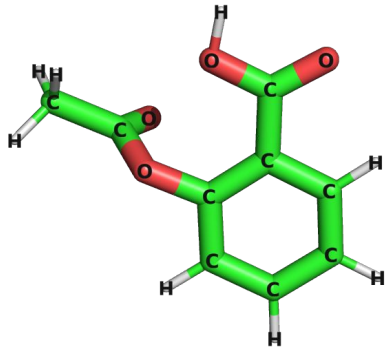
Cyclic peptide drug:
Cyclosporine (1.2 kDa)



Linear peptide drug:
Bivalirudin (2.18 kDa)

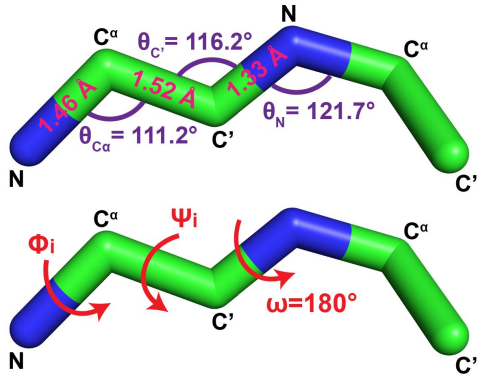


Small molecule drug:
Aspirin (0.18 kDa)

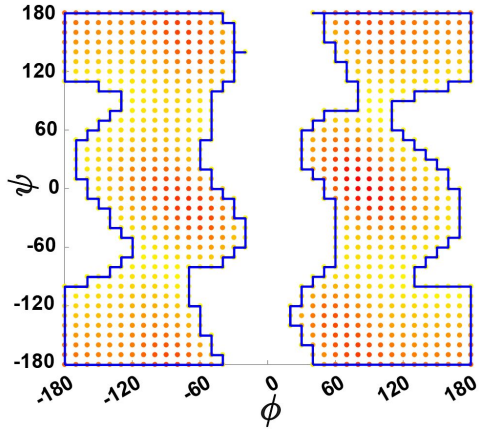


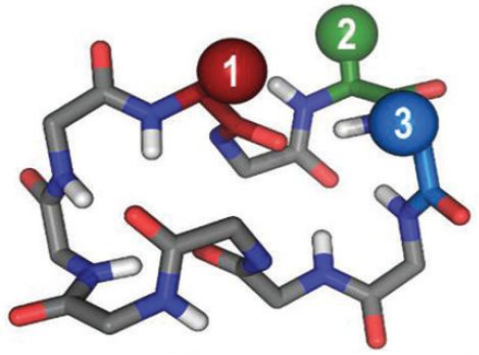
Cyclic Peptide Design Pipeline

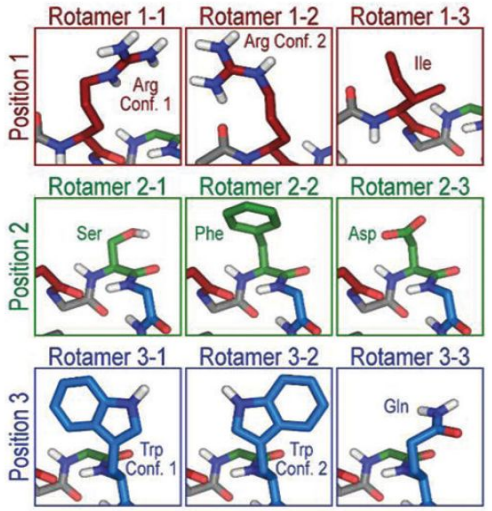
ϕ, ψ torsion angle variables



Ramachandran space







Backbone search: High dimensionality challenge

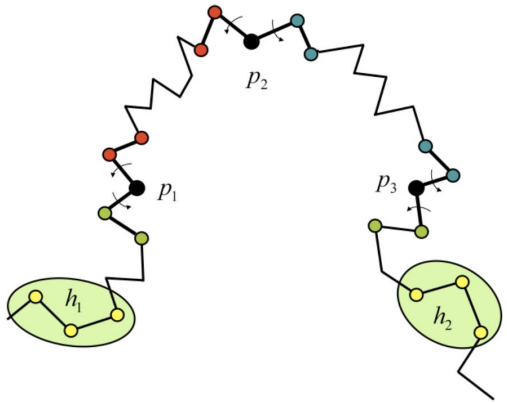
- For *n*-residue peptide, backbone has *n* pairs of **(ϕ, ψ)** torsion angle variables
- Each pair **(ϕ, ψ)** ∈ Ramachandran space
- The search space grows ***exponentially*** with peptide size

Side-chain optimization: exponential combinatorial problem

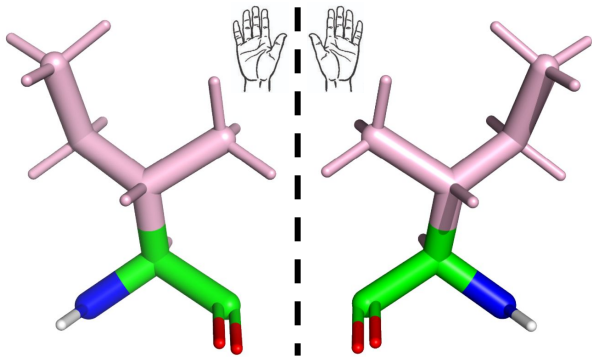
- There are 20 natural amino acids
- Sequence design has at least 20^n choices
- Many more if allow non-canonicals (e.g., D- and artificial amino acids)

Prior Work: Physics-based design in *Rosetta*

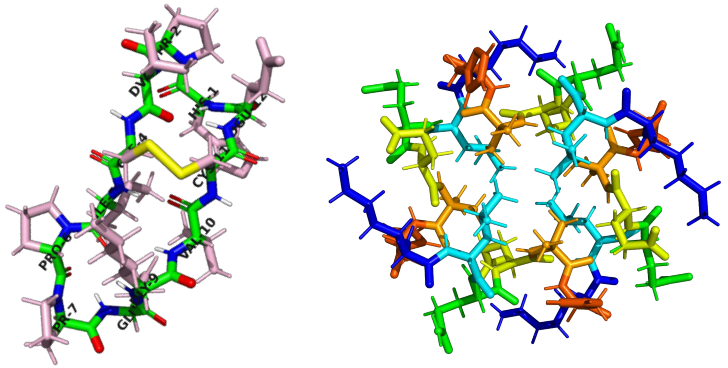
Kinematic loop closure



L-conformer **D-conformer**



Cross-link **S4 symmetry**



Algorithm

- Backbone: Use kinematic loop closure to algebraically solve the cyclic constraint [1]
- Side-chain: Use Monte Carlo simulated annealing to minimize the energy [2]

Pros

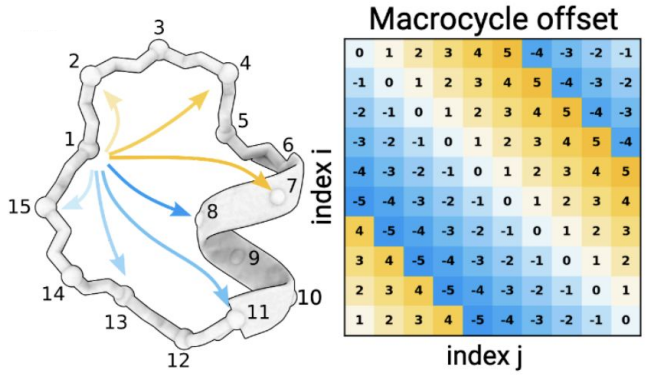
- Design can include D-amino acids, which are more resistant to peptidase degradation
- Can design non-canonicals, thus expanding chemical diversity even more

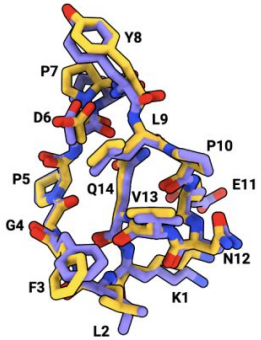
Cons

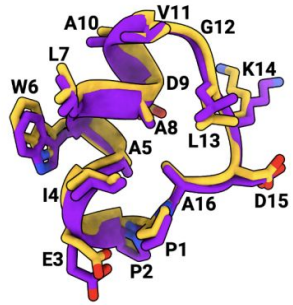
- Backbone sampling works efficiently only for 7-10 residues
- To exceed this size limit, researchers use disulfide cross-links (11-26 residues [3, 4]) or symmetry (15-24 residues [5])

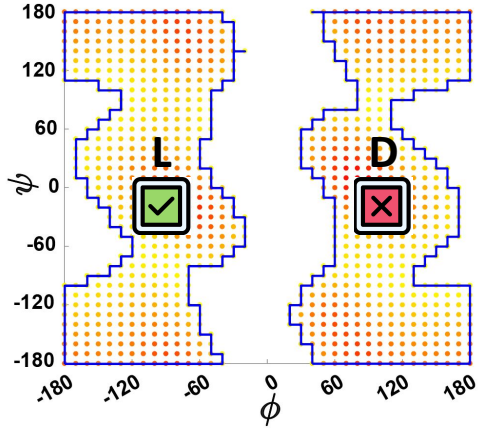
[1] D. Mandell, et al. *Nature Methods*, 6:551-552, 2009.
[2] A. Leaver-Fay, et al. *Methods Enzymol.*, 487:545-574, 2011.
[3] P. Hosseinzadeh, et al. *Science*, 358:1461–1466, 2017.
[4] G. Bhardwaj, et al. *Nature*,538:329–335, 2016.
[5] V. Mulligan, et al. *Protein Sci.*, 29:2433–2445, 2020.

Prior: Deep-learning design led by *AlphaFold*









Algorithm

- Base on *AlphaFold* or *RFdiffusion*
- Encode the cyclic backbone constraint into their amino acid relative position matrix

Pros

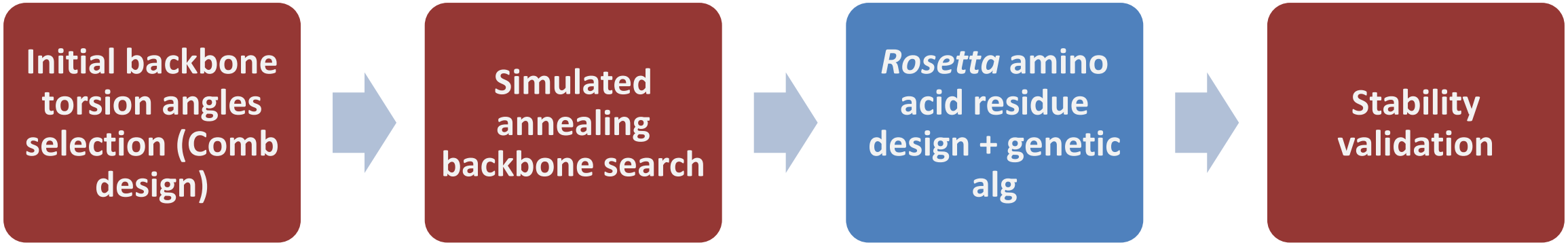
- Can design slightly larger sizes of 7-16 residues [1, 2] due to *AlphaFold*'s protein training data
- Can predict 12-39 residues [3]

Cons

- Trained on natural L-amino acids, so:
- Not good for designing non-canonicals or mixed chirality
- Therefore: restricted design search space

[1] S. Rettie, et al. *bioRxiv*, doi:10.1101/2023.02.25.529956, 2023.
[2] S. Rettie, et al. *bioRxiv*, doi:10.1101/2024.11.18.622547, 2024.
[3] C. Zhang, et al. *Brief Bioinform.*, 25:bbae215, 2024.

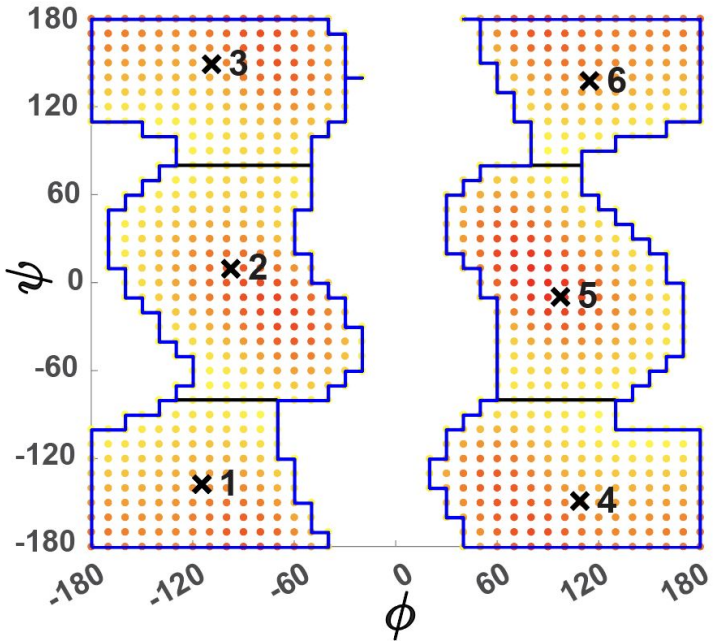
Our *CyclicChamp* Design Pipeline



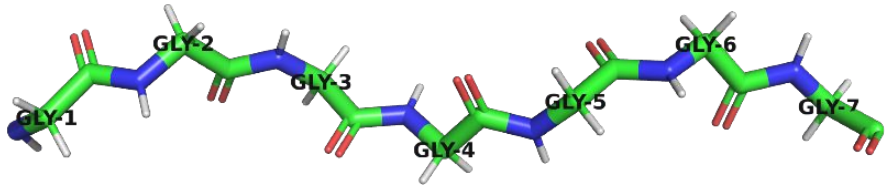
[1] Q. Zhu, et al. *PLoS Comput. Biol.* 2025.

Initial Backbone Torsion Angles Selection

- Initial backbones are chains of glycine residues, with (ϕ, ψ) chosen from the six torsion bin centers of the Rama space.
- For *n*-residue peptide, ~$6^n/n$ combinations subject to cyclic permutations.
- Use combinatorial design to obtain well-spaced random samples from all possible combinations [1].



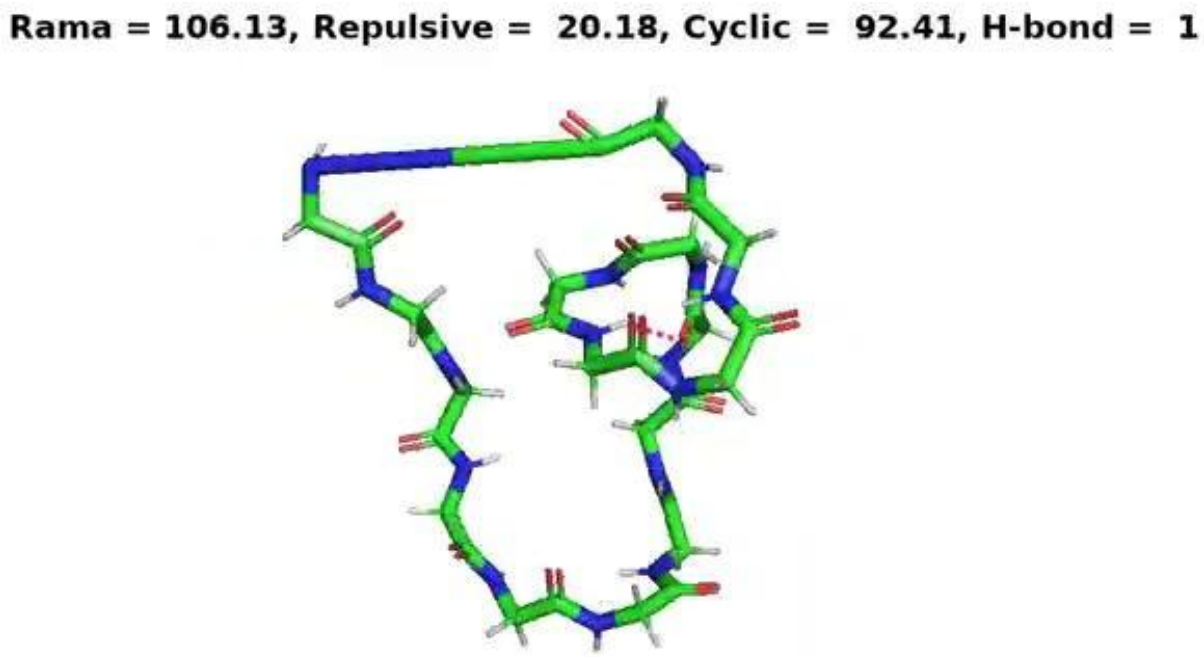
Example initial backbone with all torsion angles chosen as center 1



[1] C. Colbourn, et al. *J. Combin. Designs*, 14:124-138, 2006.

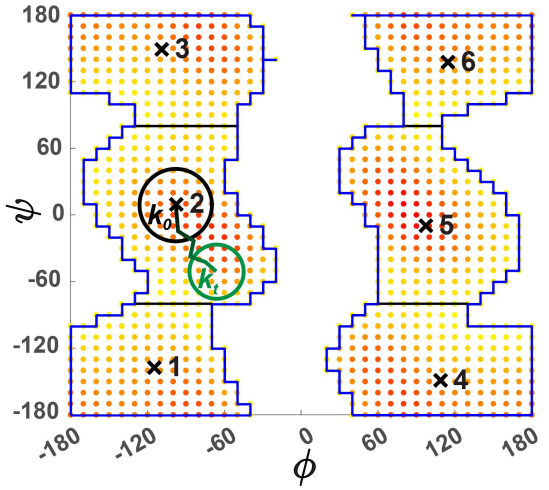
Active Search for Backbone: Simulated Annealing

- Start from “well-spaced” random initial configurations
- At each step, add perturbations and accept based on the Metropolis acceptance criterion
- Close the backbone into cycle, but also seek desirable features like hydrogen bonds and low steric clashes

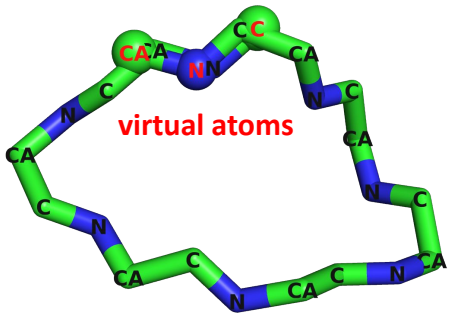


How *CyclicChamp* Simulated Annealing Works

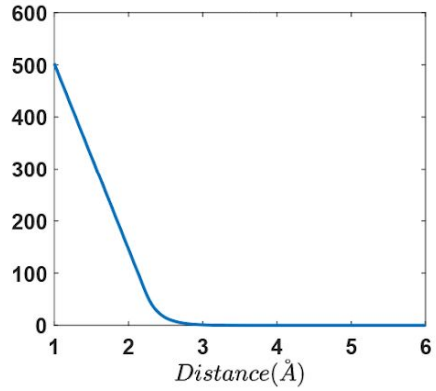
- Goal: find **cyclic** backbones with **low repulsive energy** and **sufficient hydrogen bonds**
- $E_{total} = w_{cyc} * E_{cyc} + w_{rep} * E_{rep} + w_{hbond} * E_{hbond}$
- At each step *t*, generate random moves in a shrinking disk of radius $k_t = \frac{k_0}{1 + b * t/M}$, M=10000 steps.
- Accept new configuration if (*Metropolis acceptance*)
  a) Its energy $E_{new} \leq E_{total}$
  b) Or $\text{rand}(0,1) \leq e^{(E_{total} - E_{new})/T_t}$, $T_t = \frac{T_0}{1 + c * t/M}$
- Parameters including *weights w*, *initial disk radius* k_0 and its *dropping rate b*, *initial temperature* T_0 and its *dropping rate c* vary using **combinatorial design** in test runs.



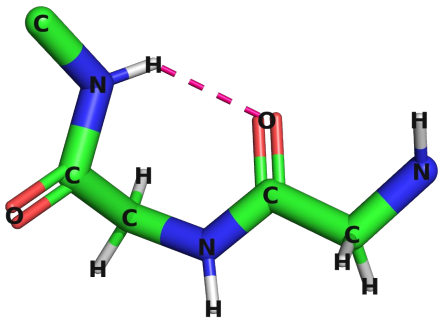
Cyclic error = 0.14



Repulsive energy [1]



Hydrogen bond [1]

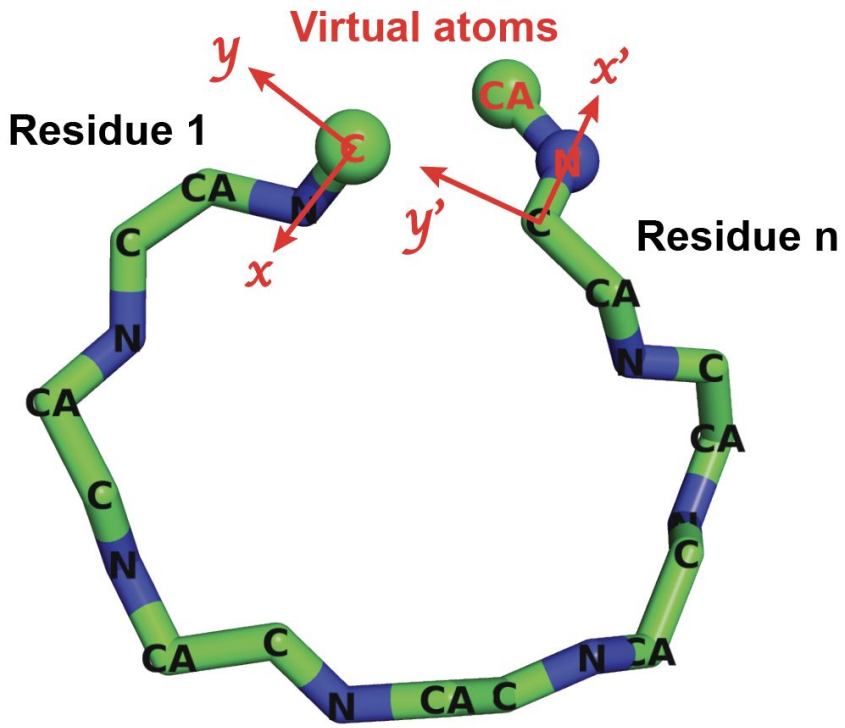


[1] R. Alford, et al. *J. Chem. Theory Comput.*, 13:3031–3048, 2017.

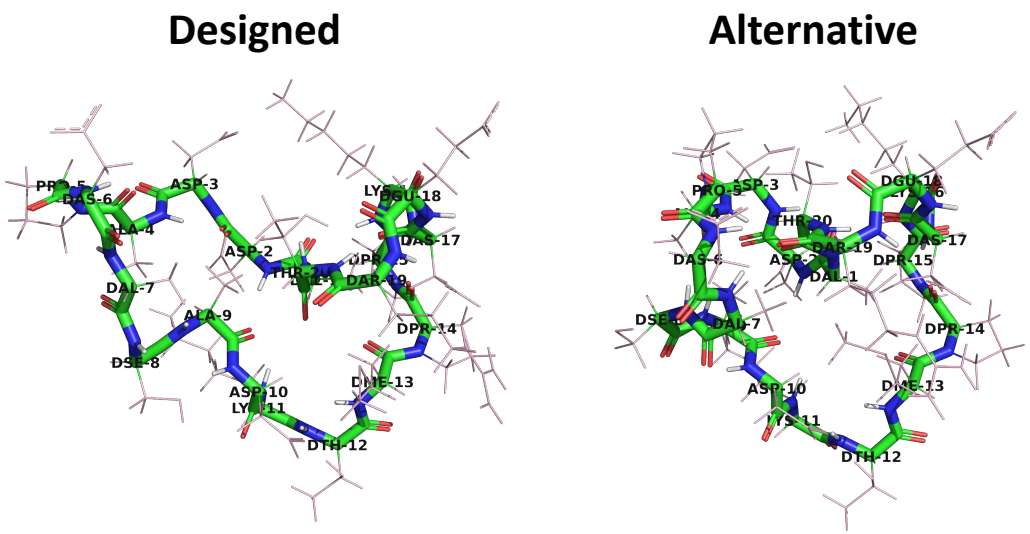
Cyclic Backbone Constraint Modeled as Cyclic Error

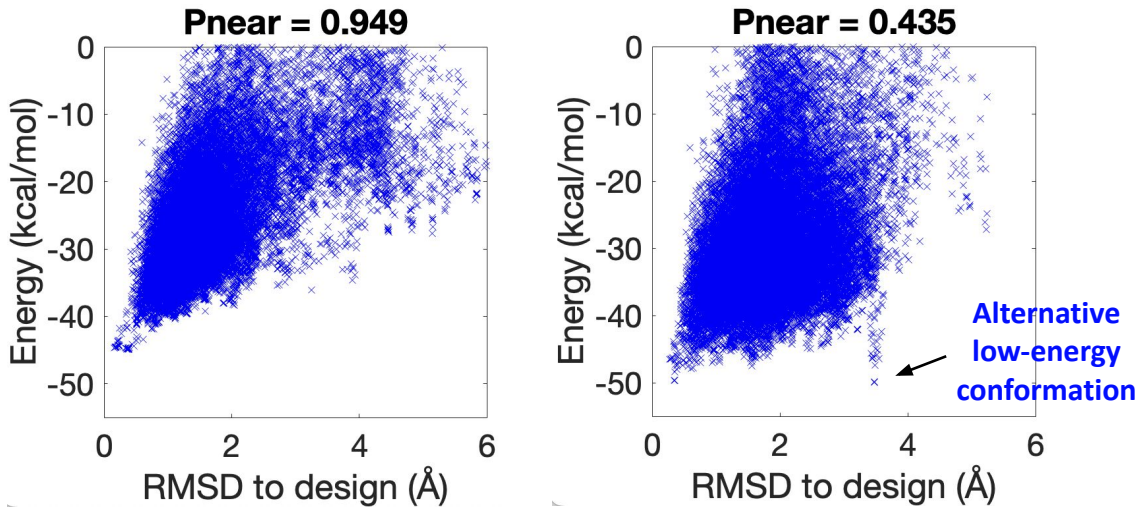
- Generate a virtual atom $C_{virtual}$ at the origin before the *N* terminus
- The first *N* atom lies on the standard *x*-axis, and the first *CA* atom lies in the standard *xy*-plane
- For a given set of ϕ, ψ torsion angle values, compute the coordinates of all backbone atoms using
  - Ideal bond angles and bond lengths
  - Ideal torsion ω = 180° at peptide bond (C-N)
- Two virtual atoms $N_{virtual}$, $CA_{virtual}$ after the *C* terminus also computed
- Construct unit vector *x'* pointing from C_n to $N_{virtual}$, and its perpendicular *y'* such that $CA_{virtual}$ lies in the *x'y'*-plane
- The cyclic error

$$E_{cyc} = \sqrt{\|C_n - C_{virtual}\|^2 + \|x - x'\|^2 + \|y - y'\|^2}$$



Stability Validation by Sampling Alternative Conformations





Sample energy landscapes

- For a designed amino acid sequence, the **full-energies** of alternative conformations, together with their backbone root-mean-square-deviations (**RMSDs**) from the designed structure, form the energy landscape
- **Stable** design if all low-energy conformations align closely to the backbone conformation

Stability = probability to fold into backbone

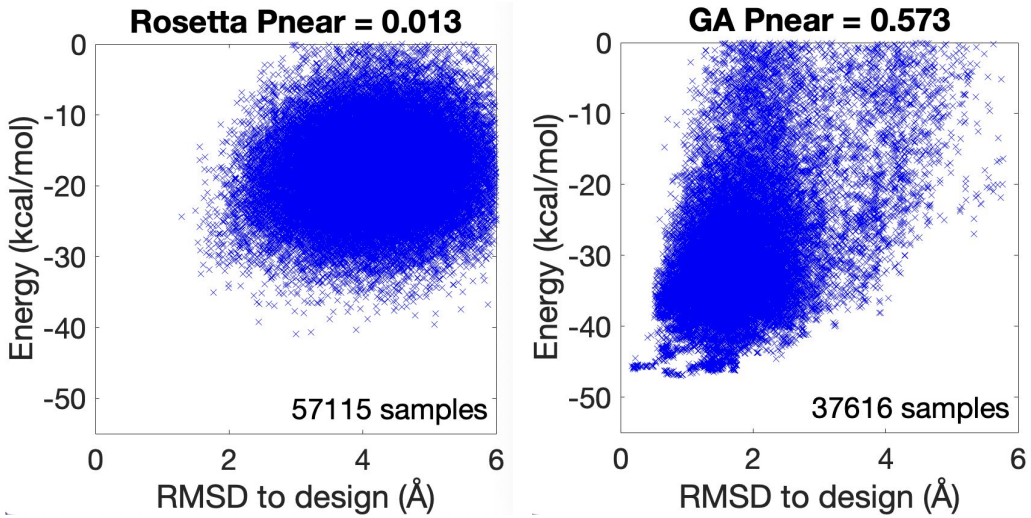
- To quantitatively measure stability, use

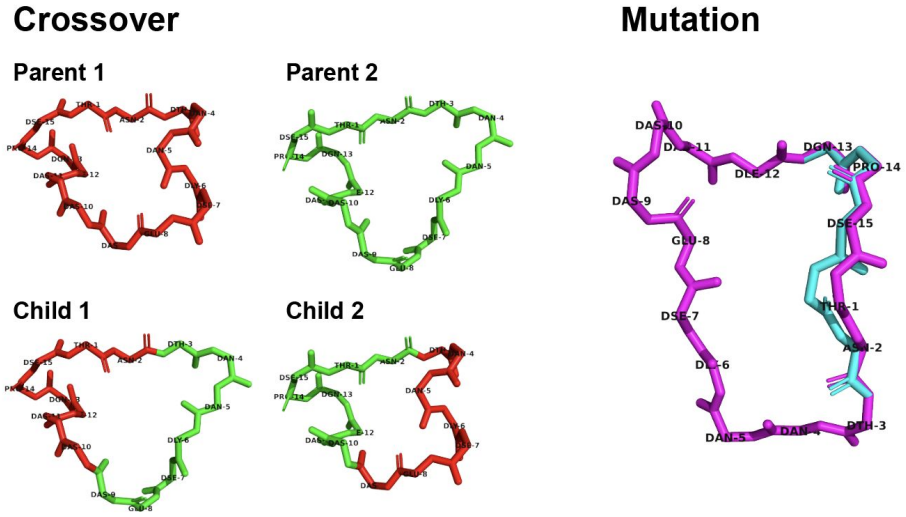
$$P_{Near} = \frac{\sum_{i=1}^{N} \exp\left(\frac{-RMSD_i^2}{\lambda^2}\right)\exp\left(-\frac{E_i}{k_B T}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{E_i}{k_B T}\right)},\ \lambda\sim 1,\ k_B T = 0.62$$

- P_{Near}**>0.9** experimentally shown to be indicative of stability [1]

[1] P. Hosseinzadeh, et al. *Science*, 358:1461–1466, 2017.

Genetic Algorithms to Avoid Local Minima





Rosetta's failure to sample thoroughly

- The conformational space grows **exponentially** with the peptide size.
- *Rosetta*'s **random sampling** approach [1] yields many false-negatives when validating 15 residue designs, and generates unusable energy landscapes for 20 and 24 residues.

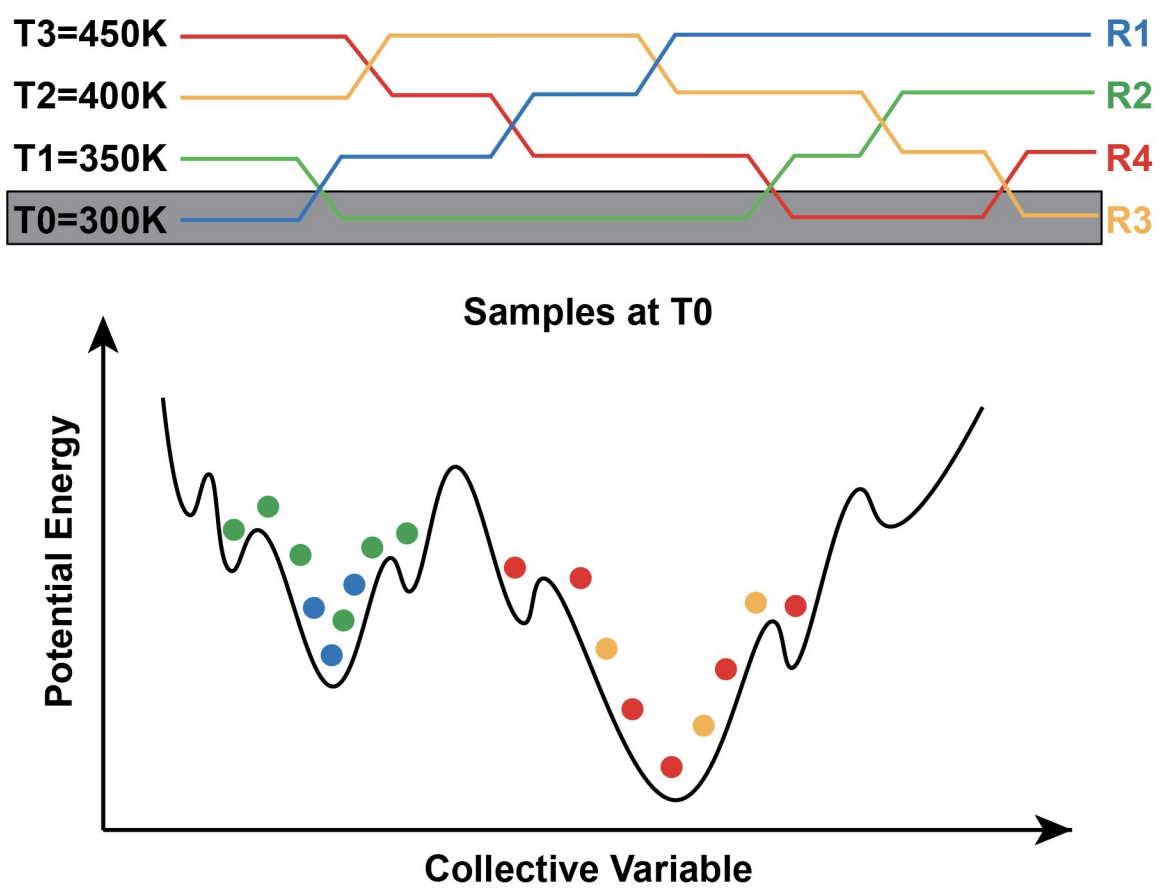
What genetic algorithms accomplish

- Using **simulated annealing** alone may find only shallow minima in a rugged energy landscape
- Explore potential energy minima via **genetic algorithms** that perform crossover and mutation on low-energy conformations

[1] P. Hosseinzadeh, et al. *Science*, 358:1461–1466, 2017.

Validation by Replica Exchange Molecular Dynamics

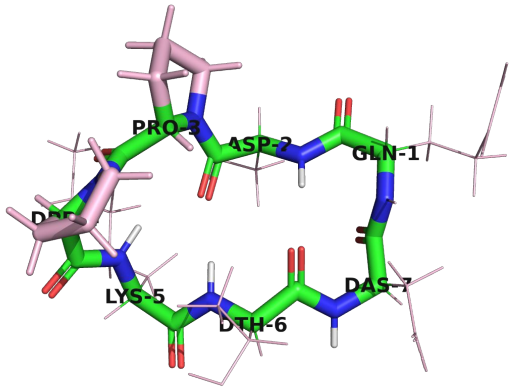
- For top designs with high P_{Near}, perform *replica exchange molecular dynamics (REMD)* simulations as a validation method
- Identical replicas of designed peptides in solvent (water) box copied for different temperatures, and undergo regular molecular dynamics simulations [1]
- Periodically ***swap*** temperatures for neighboring replicas, so that each replica has chance to explore behavior at high temperatures to overcome ***energy barriers***
- Collect and analyze conformations sampled at reference temperature *T0*



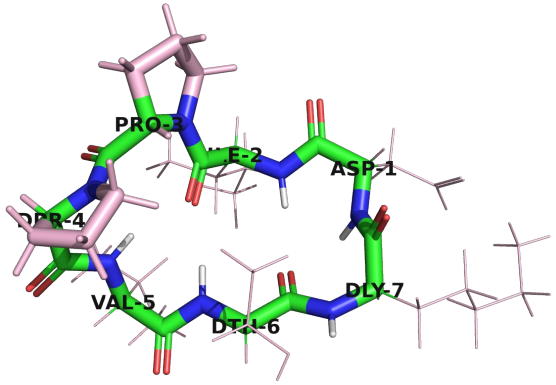
[1] Y. Sugita, et al. *Chem. Phys. Lett.*, 314:141–151, 1999.

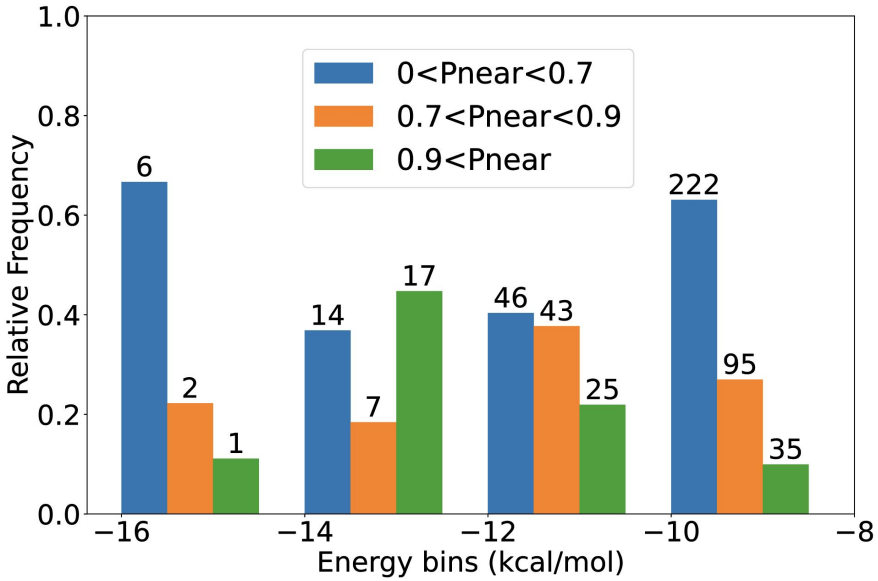
Faster 7-residue Macrocycle Design with P_{Near} Insights

PDB: 6BF3



CyclicChamp





Four times faster to design a stable peptide

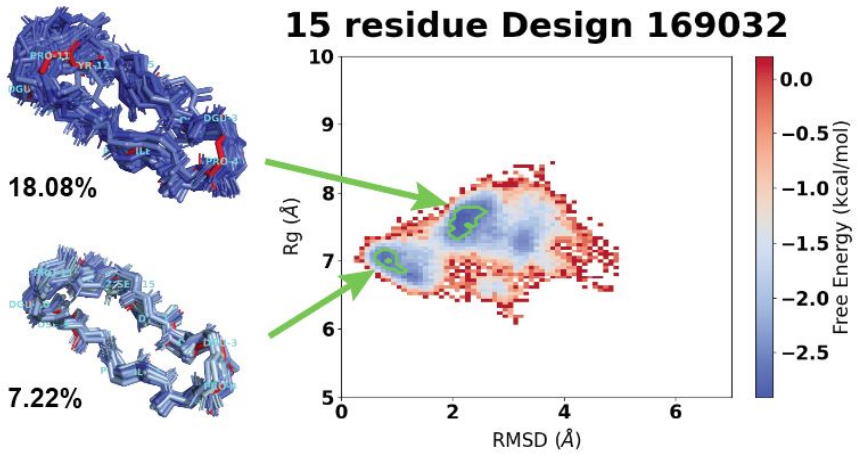
- Allocating equivalent computation time for backbone sampling, *CyclicChamp* found **4.5 times** as many stable designs as *Rosetta*.
- *CyclicChamp* also recovered the experimentally validated conformations designed by *Rosetta* [1].

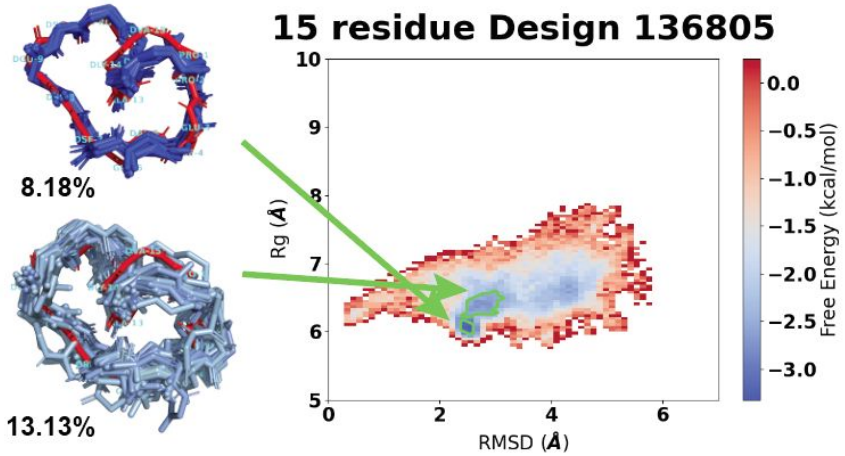
Low energy ≠ High P_{Near}

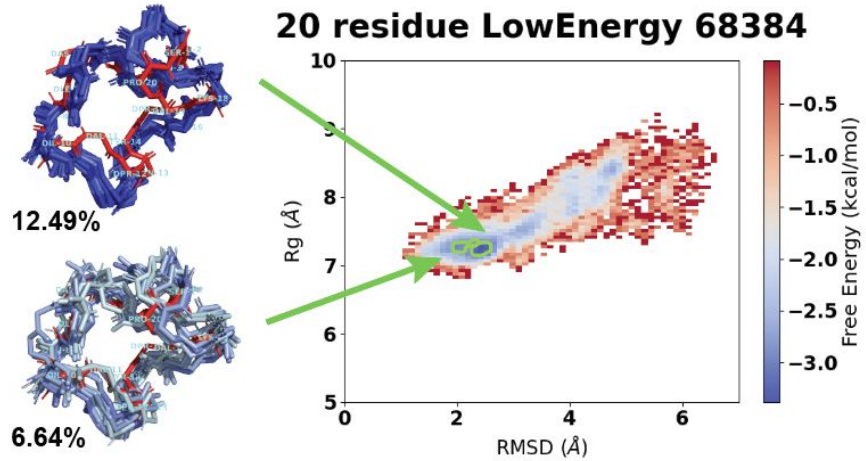
- The second-lowest energy bin had the largest probability of high P_{Near} values
- Low energy alone not sufficient for stability
- Other features like tight hydrogen bond networks often enhance stability

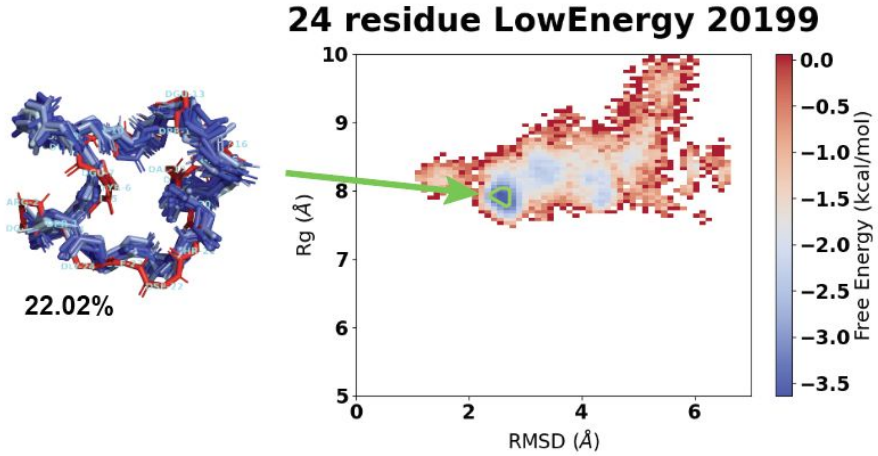
[1] P. Hosseinzadeh, et al. *Science*, 358:1461–1466, 2017.

REMD Validated 15-24 residue Large Macrocycle Designs

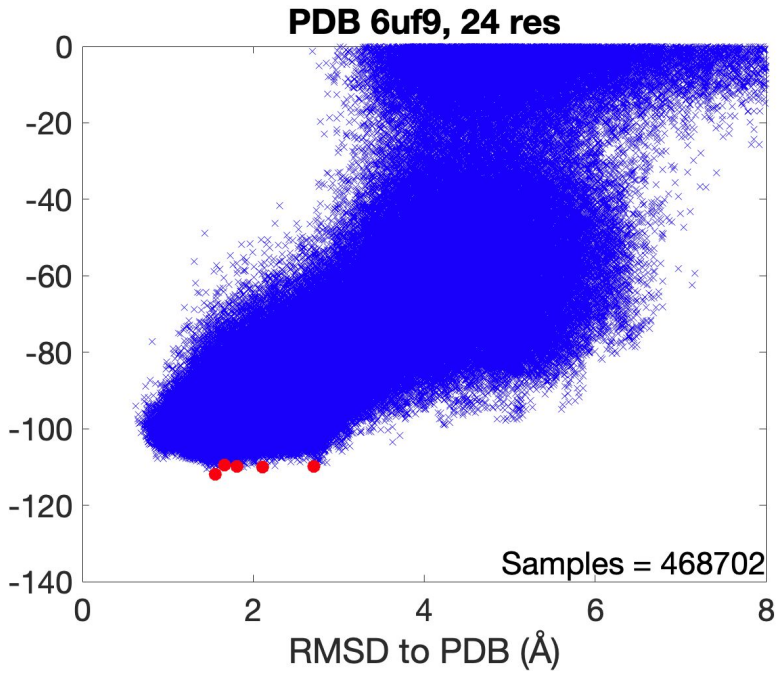


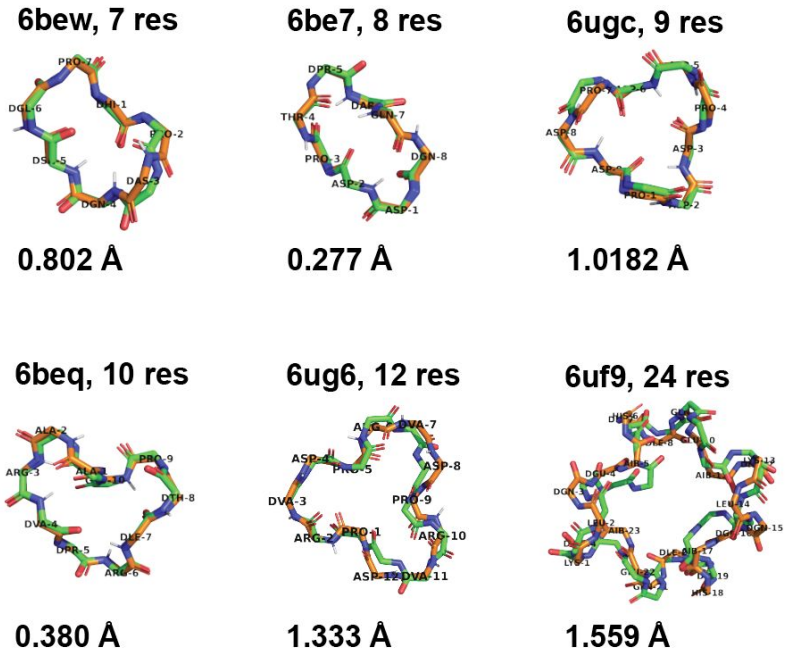






Structure Predictions for Previously Known Macrocycles





Only amino acid sequence known

- Generated energy landscapes using only the amino acid sequences as input
- Clustered the 50 lowest-energy conformations and selected the 5 lowest-energy cluster centers as predictions

Accurate predictions for 7-24 residues

- Predicted 17 existing macrocycles without cross-links, whose experimentally-solved structures deposited in the Protein Data Bank
- Predicted backbones have RMSDs of 0.277-1.836 Å from experimental structures

Conclusion

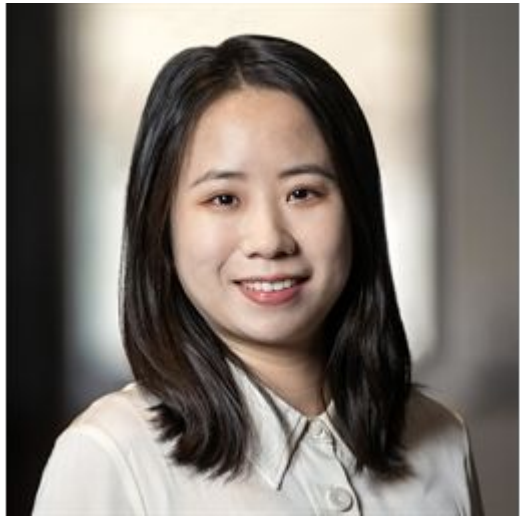
- We have addressed the high dimensionality challenge in backbone search by
  - Transforming the cyclic backbone constraint into an error function
  - Employing simulated annealing for backbone sampling
  - Employing genetic algorithms for thorough exploration of energy minima
  - Employing combinatorial design for initial configuration and parameter selection
- We have achieved
  - Faster speed for designing small (e.g. 7-residue) macrocycles
  - Unique ability to design 15-24 residue large macrocycles with mixed-chirality
  - Accurate predictions for existing macrocycles deposited in PDB
- Codes available at https://github.com/qiyaozhu/CyclicPeptide

Future Work

- Design cyclic peptide-binders to druggable protein sites as novel therapeutics
- Add option for sampling backbones with disulfide cross-links
- Incorporate into *Rosetta* for smooth transition between *CyclicChamp*'s backbone sampling and *Rosetta*'s sequence design, and user friendliness
- Add GPU support for further speed up

Acknowledgements

Vikram Mullligan
Research Scientist at Flatiron Institute



Qiyao Zhu
Research Fellow at Flatiron Institute

What is combinatorial design? Well-spaced sampling.

Suppose you are a thief…

Combinatorial Safe: 10 switches with 3 settings each. Over 59,000 (3^10) possible configurations. However there is a certain pair of switches (you don't know which pair) and a certain pair of values of those switches that will open the safe.

Illustration:

S1 S2 S3 S4 S5 S6 S7 S8 S9 S10

C A

Challenge:

Open the safe in as few switch configurations as possible. How many? How to do?

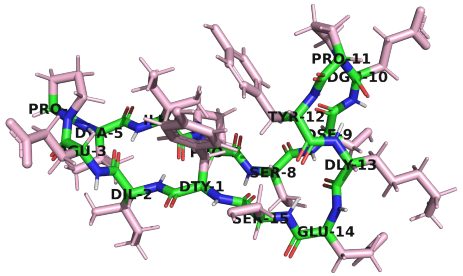
Safecracking Solution (X = Don’t care)

- **Number S1 S2 S3 S4 S5 S6 S7 S8 S9 S10**
- **1: A A A A A A A A A A**
- **2: A B B B B B B B B B**
- **3: A C C C C C C C C C**
- **4: B A B C A B C A B C**
- **5: B B C A B C A B C A**
- **6: B C A B C A B C A B**
- **7: C A C B A C B A C B**
- **8: C B A C B A C B A C**
- **9: C C B A C B A C B A**
- **10: X A A A B B B C C C**
- **11: X A A A C C C B B B**
- **12: X B B B A A A C C C**
- **13: X B B B C C C A A A**
- **14: X C C C A A A B B B**
- **15: X C C C B B B A A A**

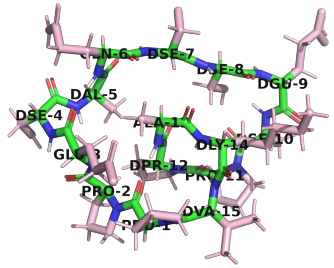
Faster design & Larger macrocycles

- For 7 residues, our *CyclicChamp* takes one-fourth the time of *Rosetta*'s to find a stable design. Advantage increases with size.
- *de novo* design of the first 15-24 residue mixed-chirality macrocycles, without relying on additional cross-links or symmetry
- Computationally validate using molecular dynamics

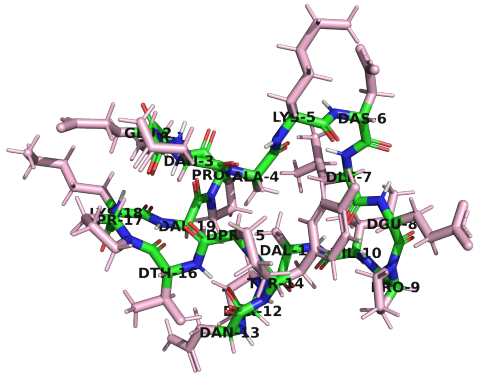
15 residue



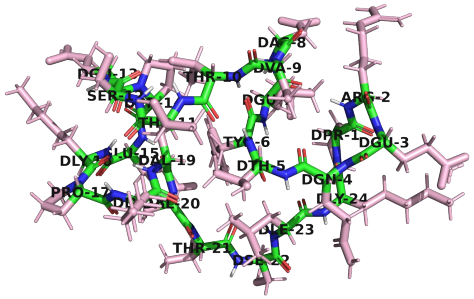
15 residue



20 residue

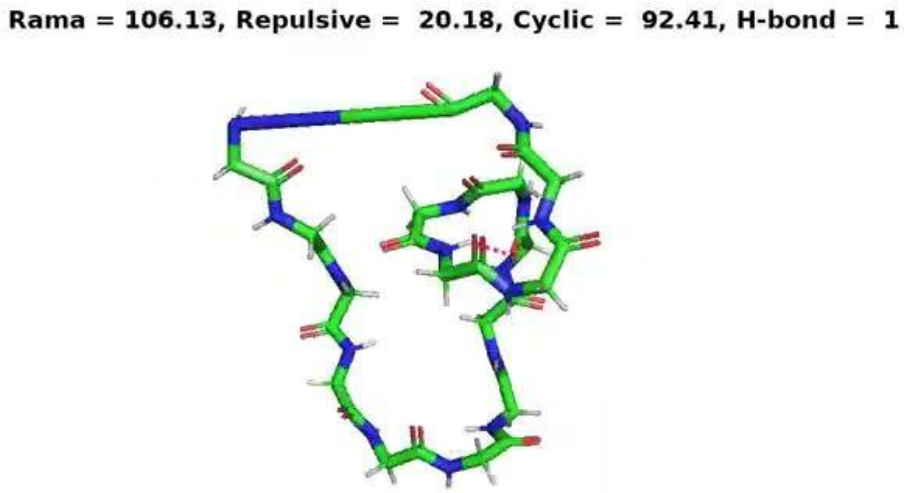


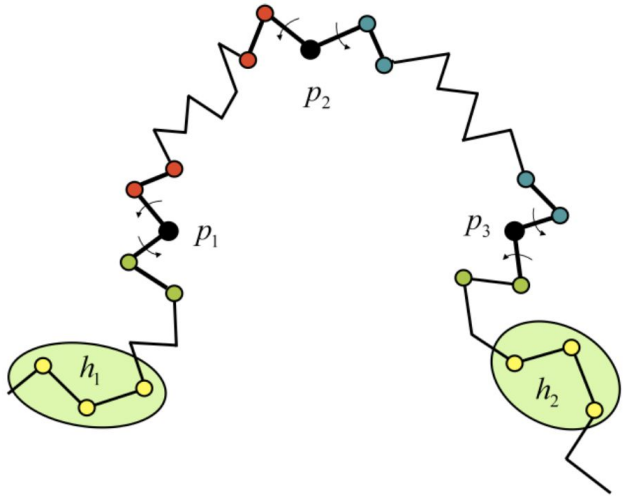
24 residue



[1] Q. Zhu, et al. *PLoS Comput. Biol.* 2025.

Our active search *vs. Rosetta*'s random sampling





Our backbone simulated annealing search [1]

- Start from "well-spaced" random initial configurations
- At each step, add perturbations and accept based on the Metropolis acceptance criterion
- Close the backbone into cycle, but also seek desirable features like hydrogen bonds

Rosetta's backbone kinematic closure [2]

- Besides the three pivot points *p1*, *p2*, *p3*, all other segments' torsion angles are randomly sampled
- Algebraically solve for the six torsions of *p1*, *p2*, *p3* to satisfy loop closure
- Fast solution, but hit desirable features by chance

[1] Q. Zhu, et al. *PLoS Comput. Biol.* 2025.
[2] D. Mandell, et al. *Nature Methods*, 6:551-552, 2009.