PREDICTING PROTEIN SIDECHAIN CONFORMATIONS – A NEW STRATEGY

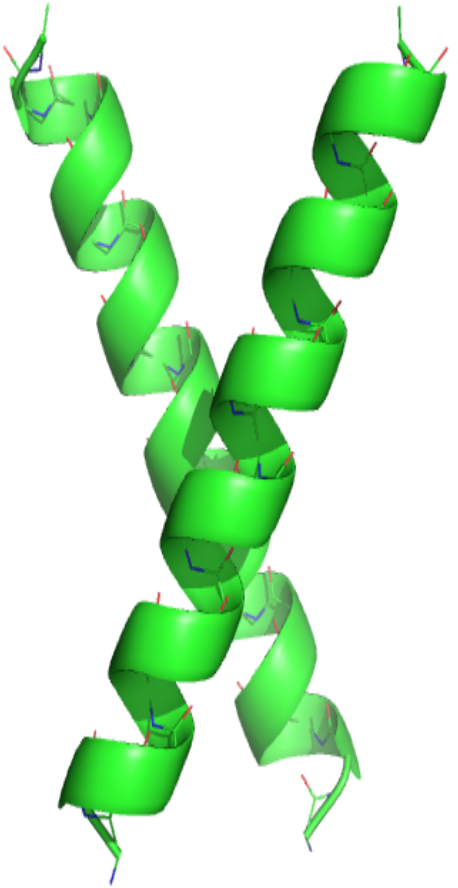
IPiB Seminar - 30th March 2012

Sabareesh Subramaniam
Senes Lab, UW Biochemistry

PREDICTING TRANSMEMBRANE HELICAL DIMERS

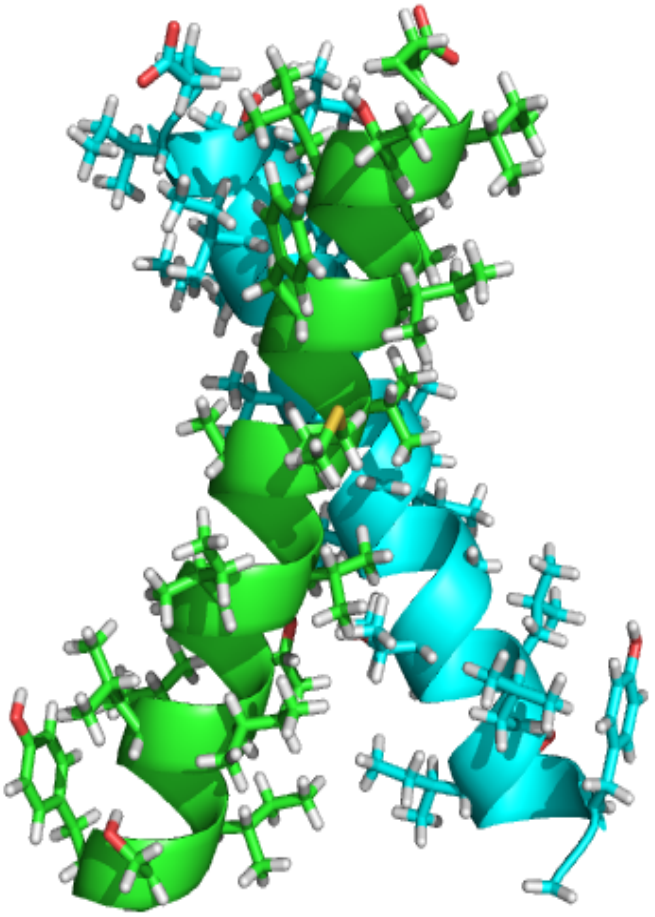
ITLIIFGVMAGVIGT

ITLIIFGVRAGRIGT



SEARCH DIMER SPACE

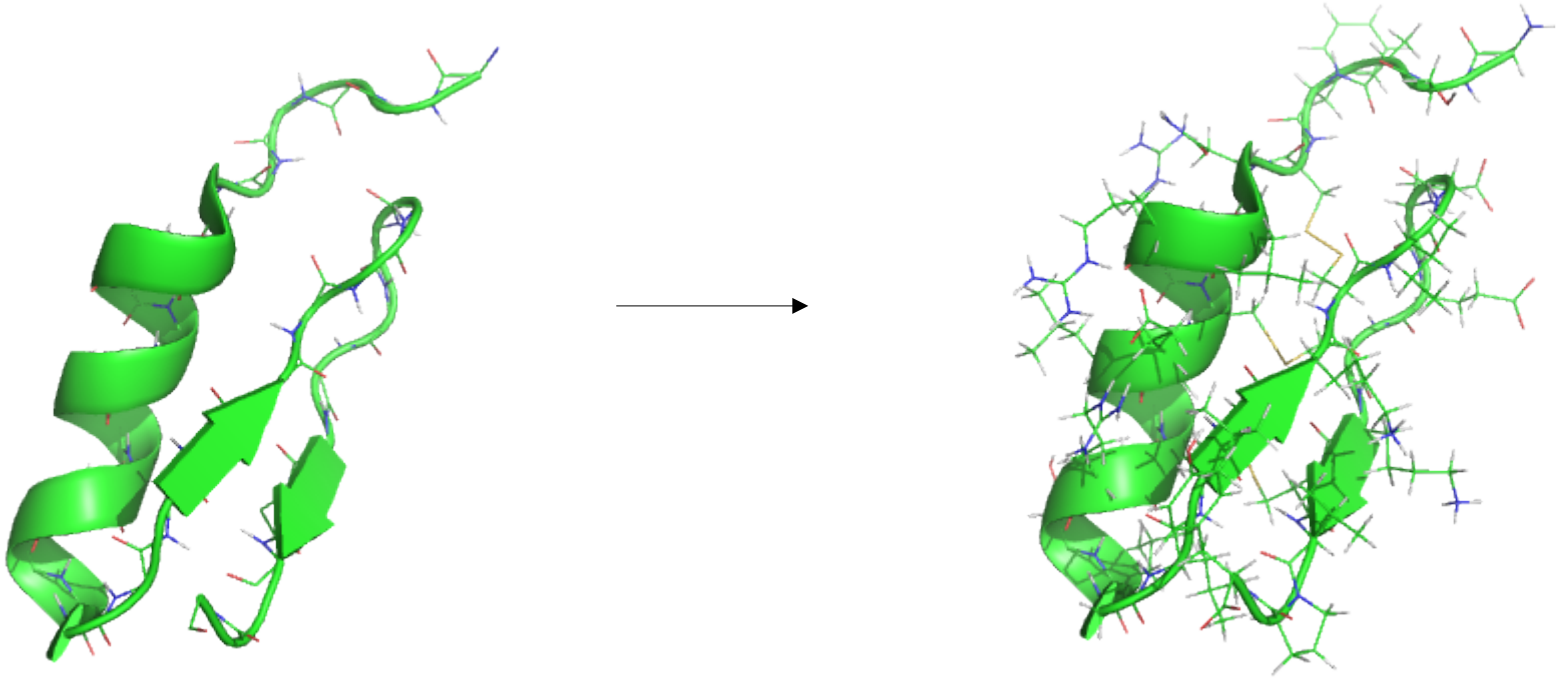
TO EVALUATE EACH BACKBONE ORIENTATION - MODEL SIDECHAINS



SIDECHAIN MODELING AS OPTIMIZATION

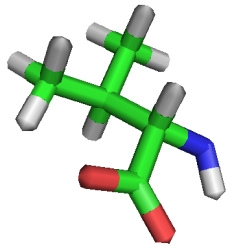
Backbone

Add sidechains to achieve minimum energy configuration

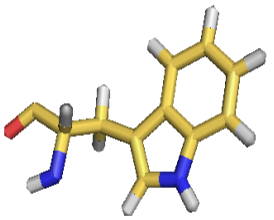


SIDECHAIN OPTIMIZATION

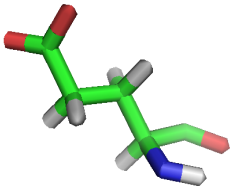
GLU

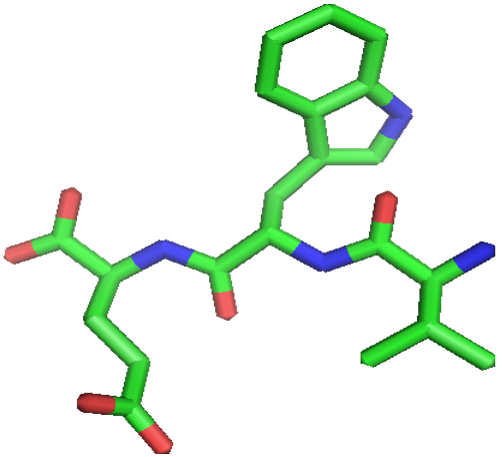


TRP

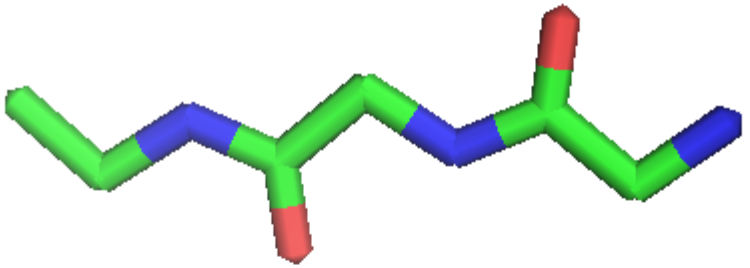


VAL

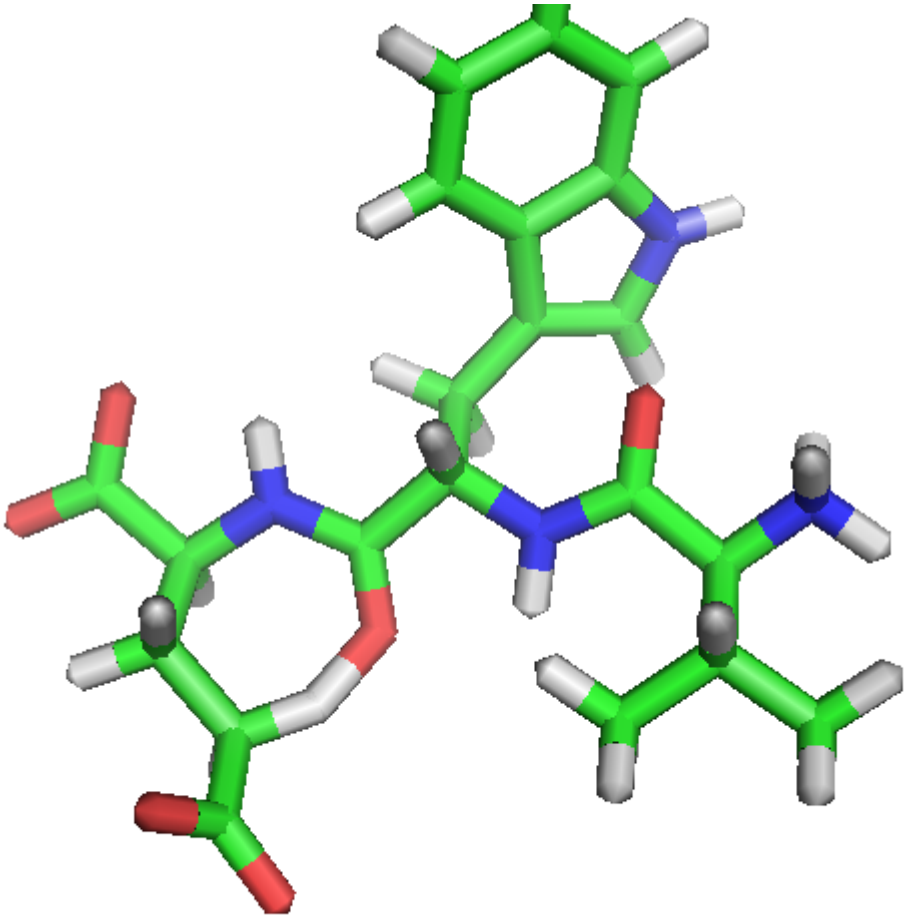




SIDECHAIN OPTIMIZATION

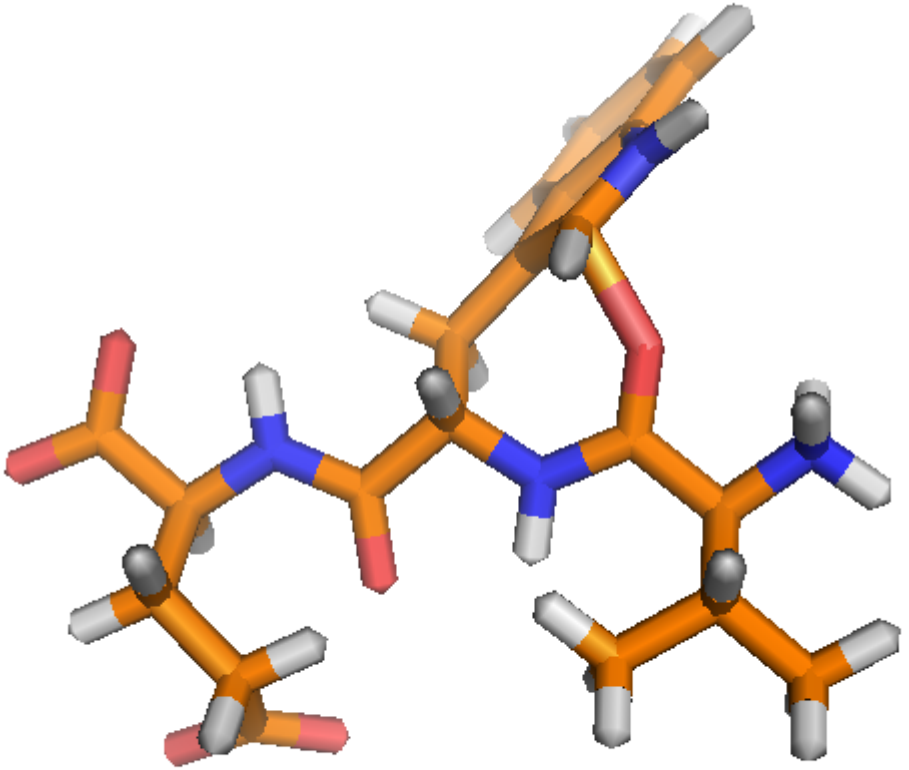


SIDECHAIN OPTIMIZATION



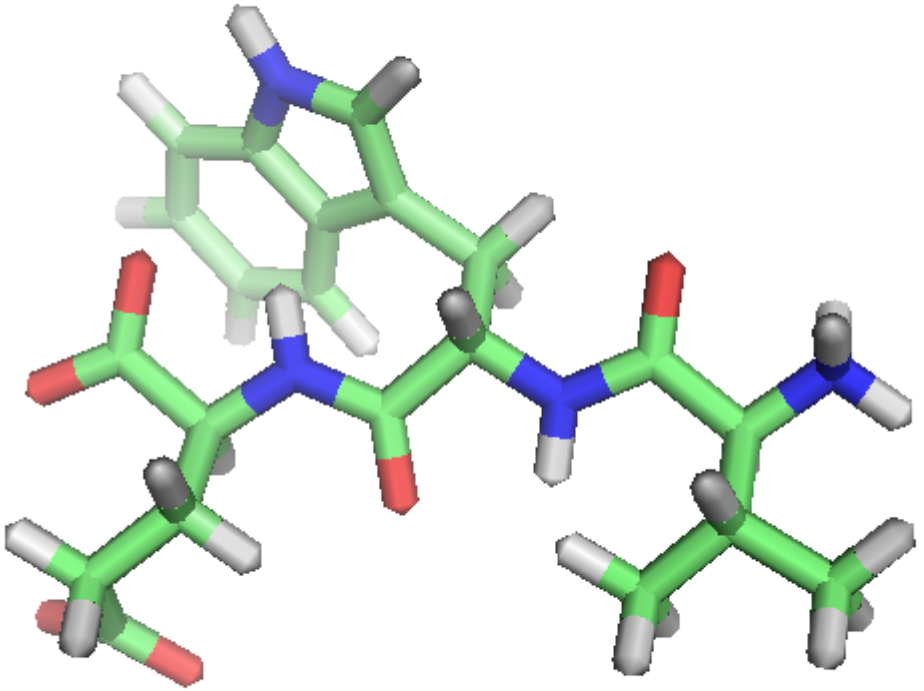
Energy = 40000 kcal / Mol

SIDECHAIN OPTIMIZATION



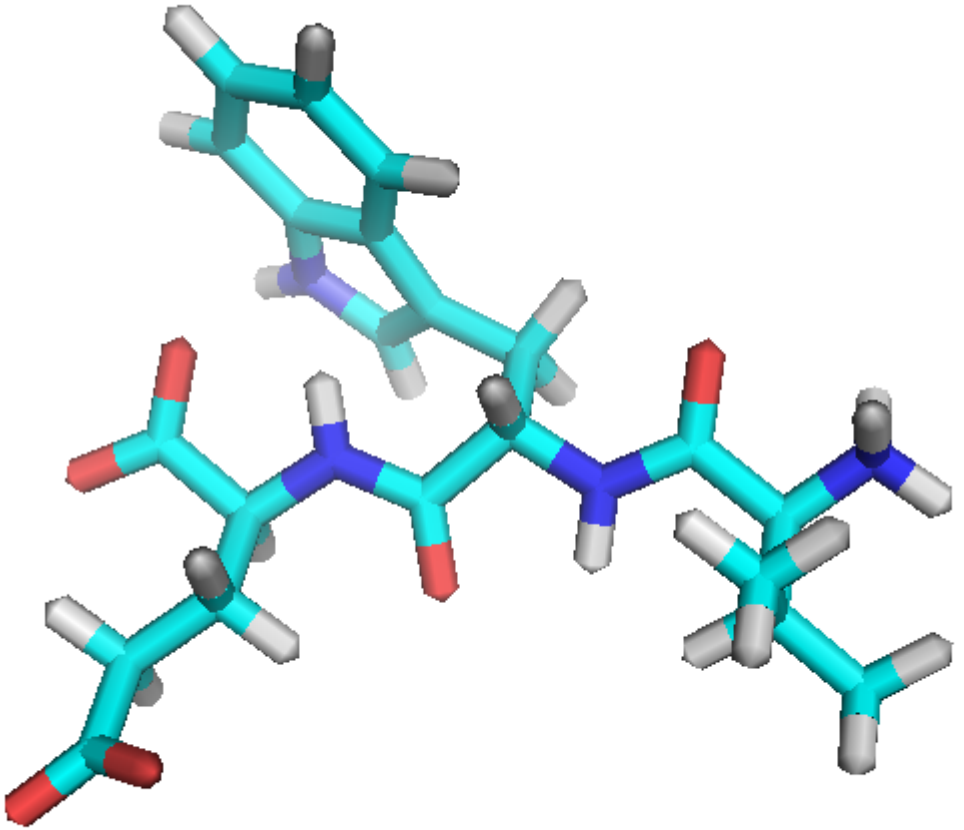
Energy = 30000 kcal / Mol

SIDECHAIN OPTIMIZATION



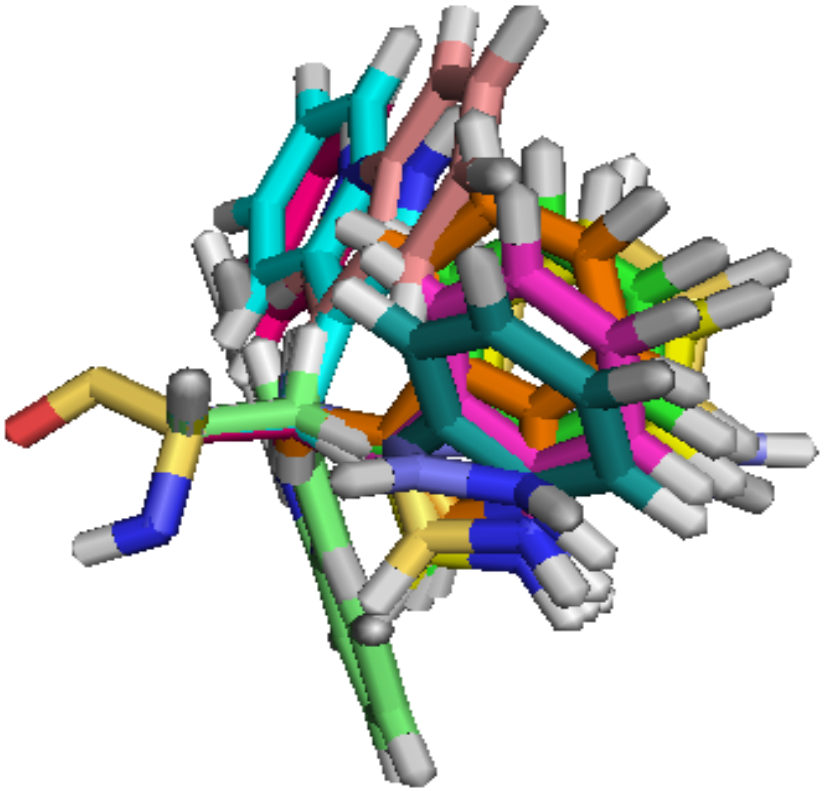
Energy = 70 kcal / Mol

SELECTED STRUCTURE – LOWEST ENERGY



Energy = 60 kcal / Mol

DISCRETIZED CONFORMATION LIBRARIES

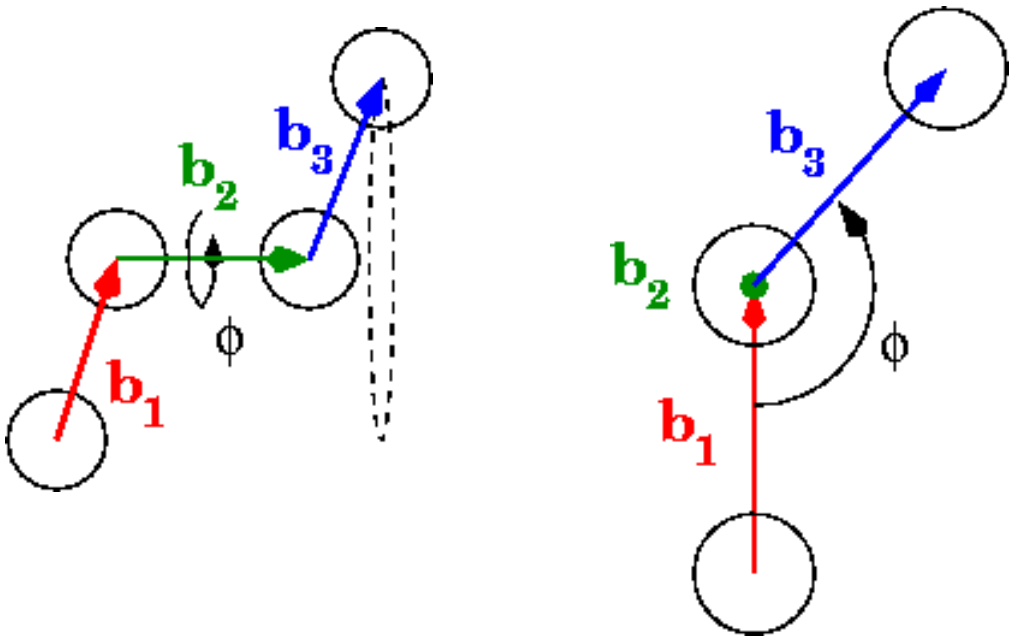


DEGREES OF FREEDOM

Bond distances

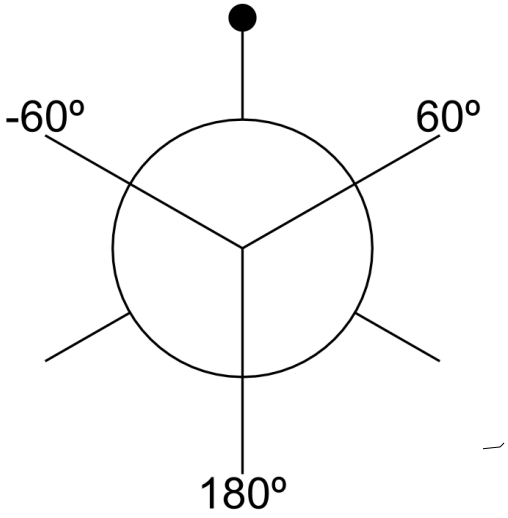
Bond Angles

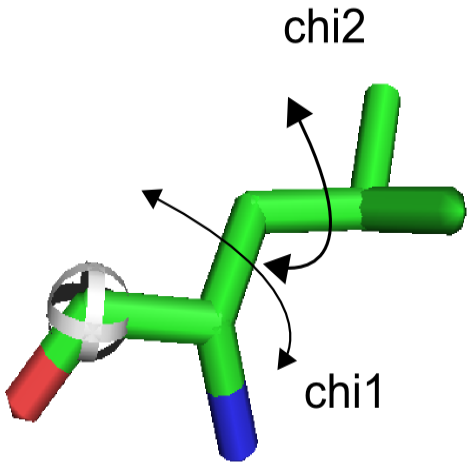
Dihedral or torsional angles

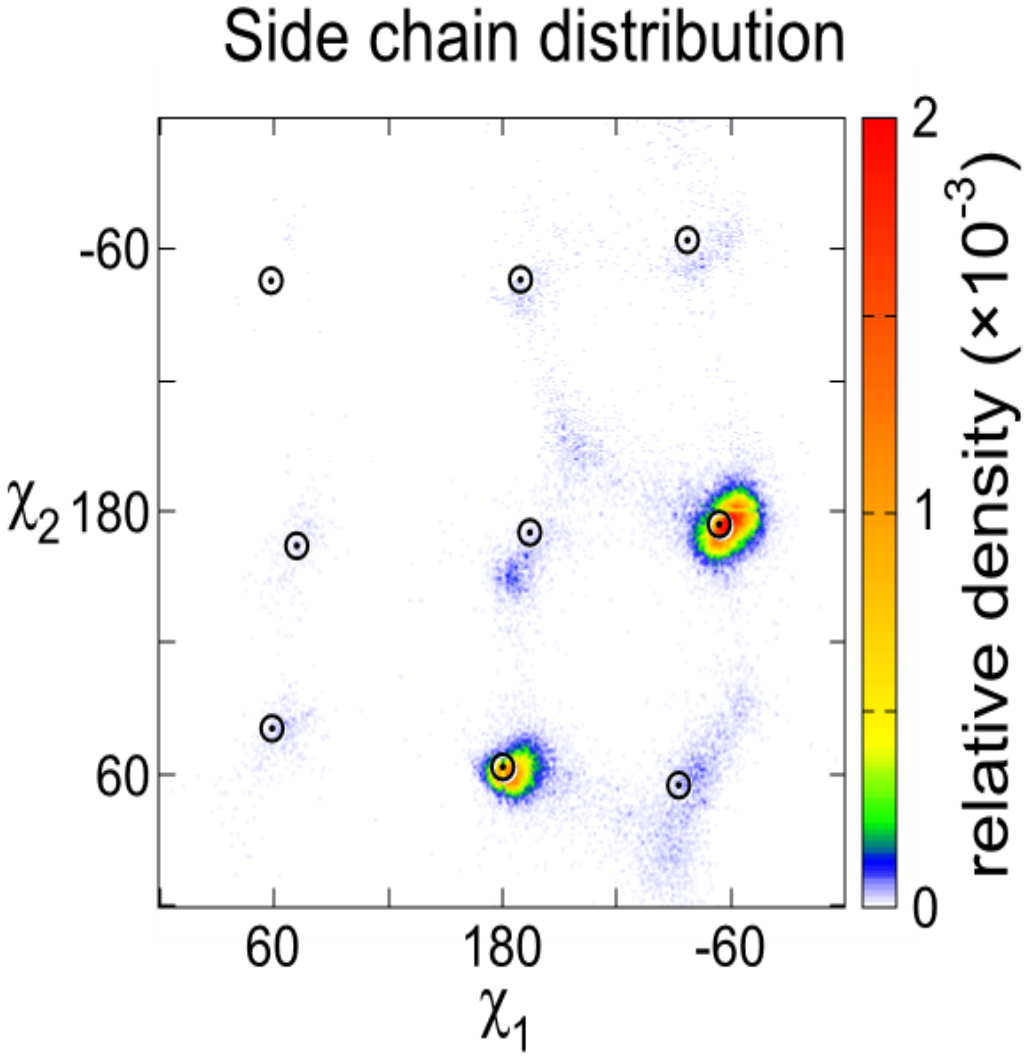


*Figures from Wikipedia

STATISTICS OF DIHEDRAL ANGLES





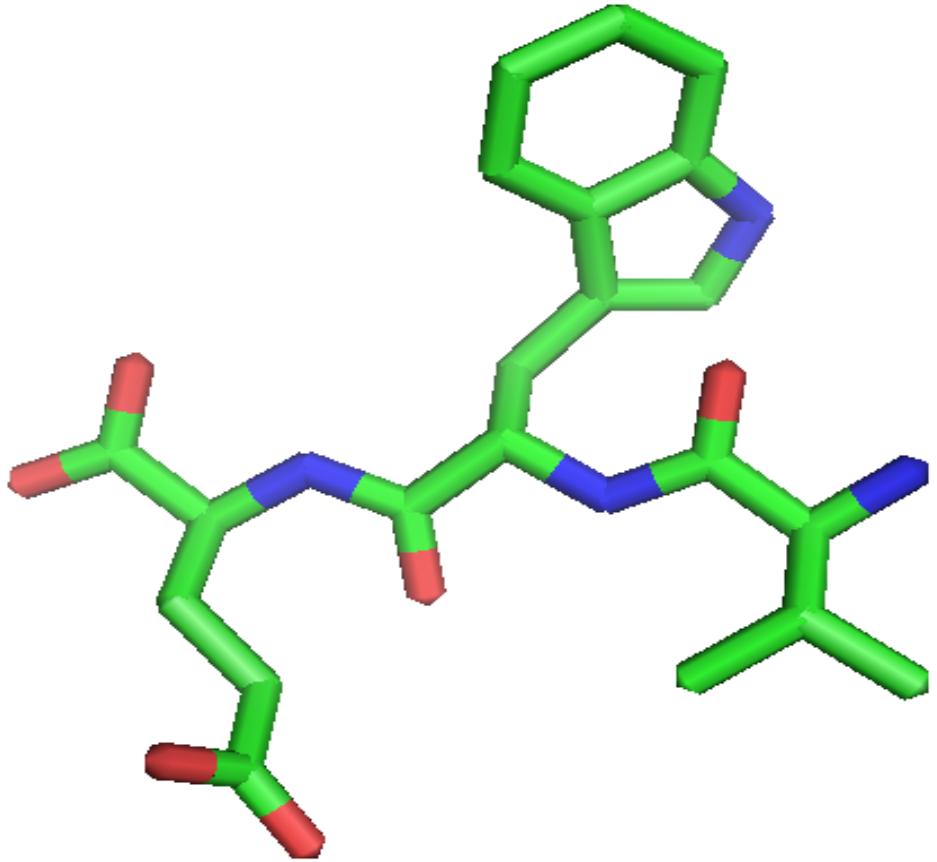


SIDECHAIN OPTIMIZATION

GLU

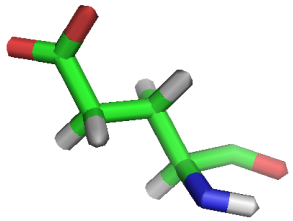
TRP

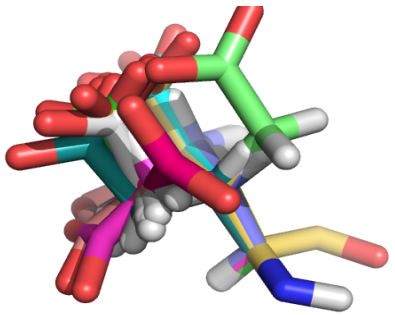
VAL



COMBINATORIAL SEARCH SPACE (3-D JIGSAW PUZZLE)

GLU



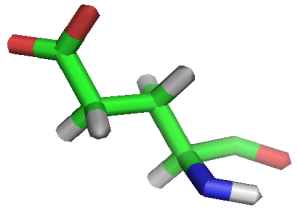


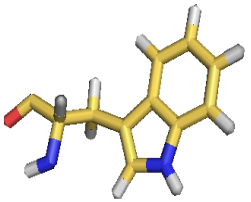
No of conformations to search 36

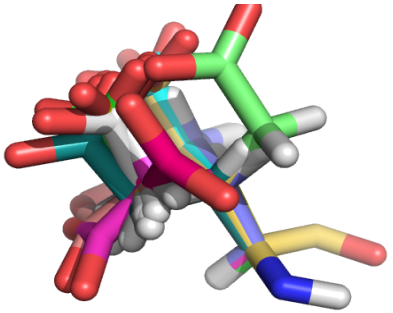
COMBINATORIAL SEARCH SPACE (3-D JIGSAW PUZZLE)

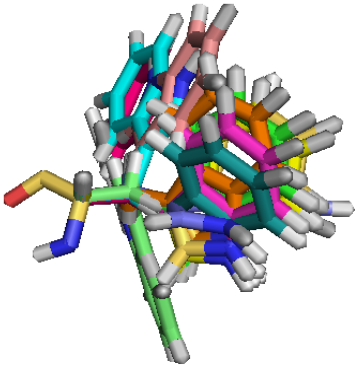
GLU

TRP





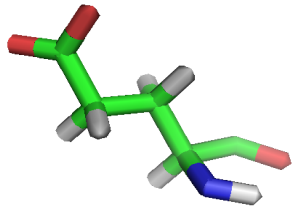


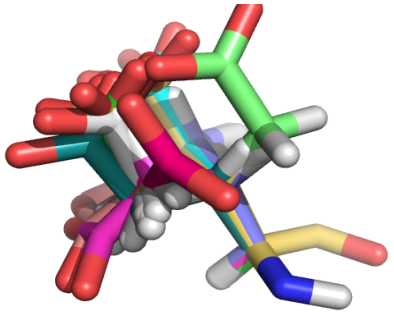


No of conformations to search 36* 54

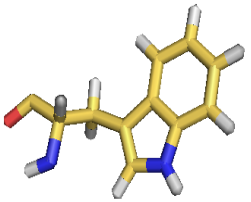
COMBINATORIAL SEARCH SPACE (3-D JIGSAW PUZZLE)

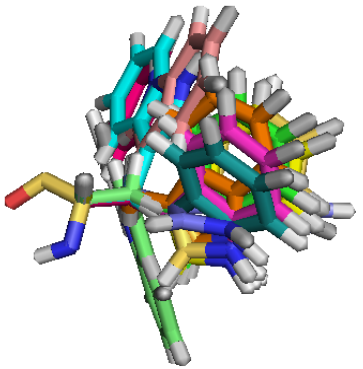
GLU



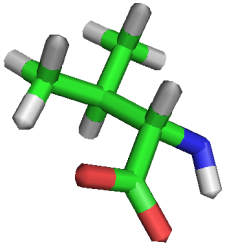


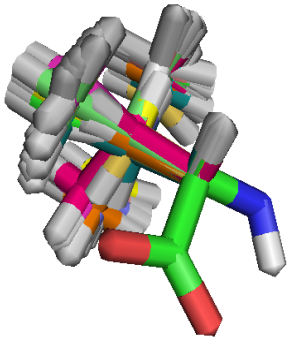
TRP





VAL



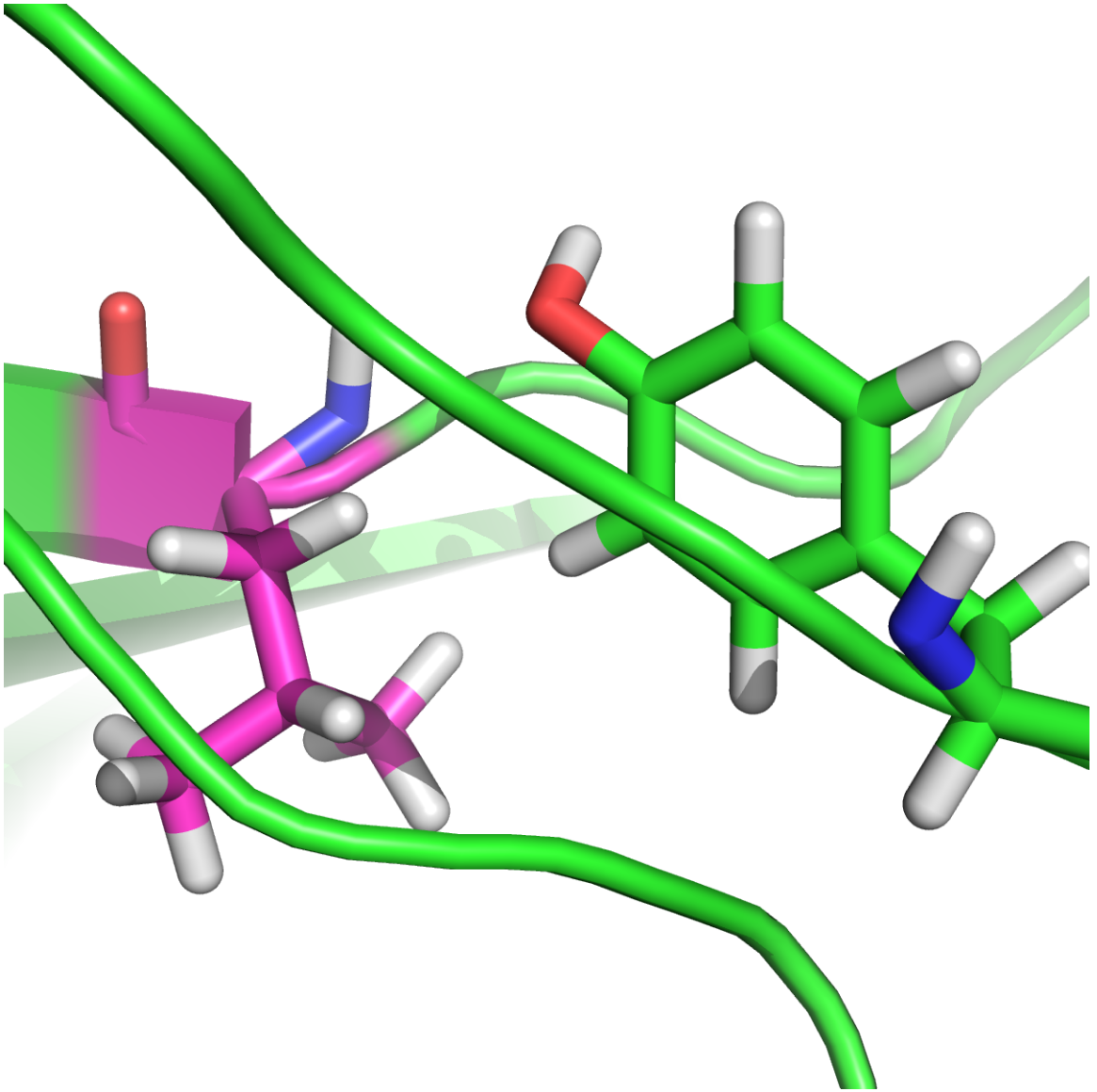


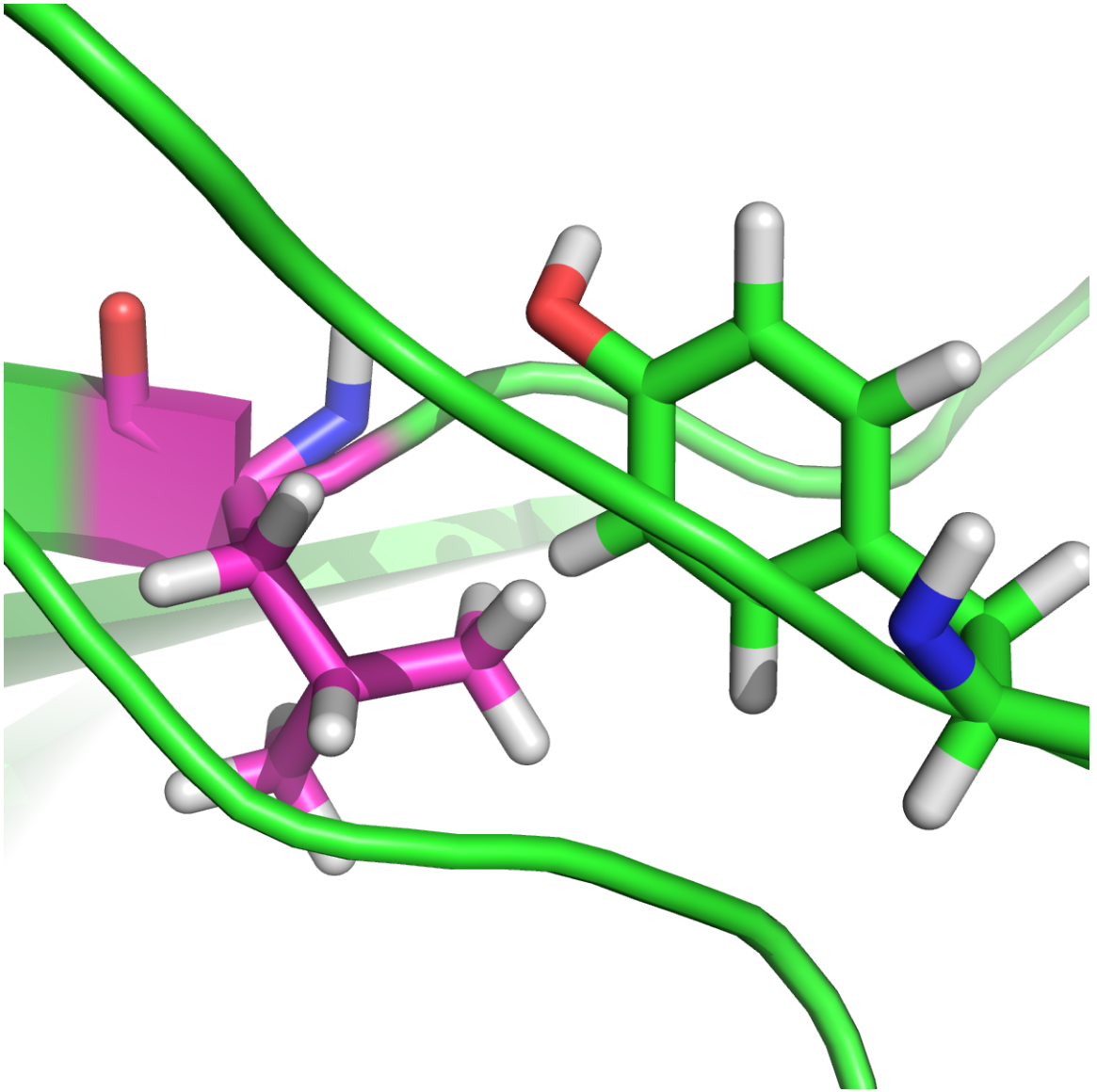
No of conformations to search 36* 54 * 3 = 5832
Typically > 10 ^ 60

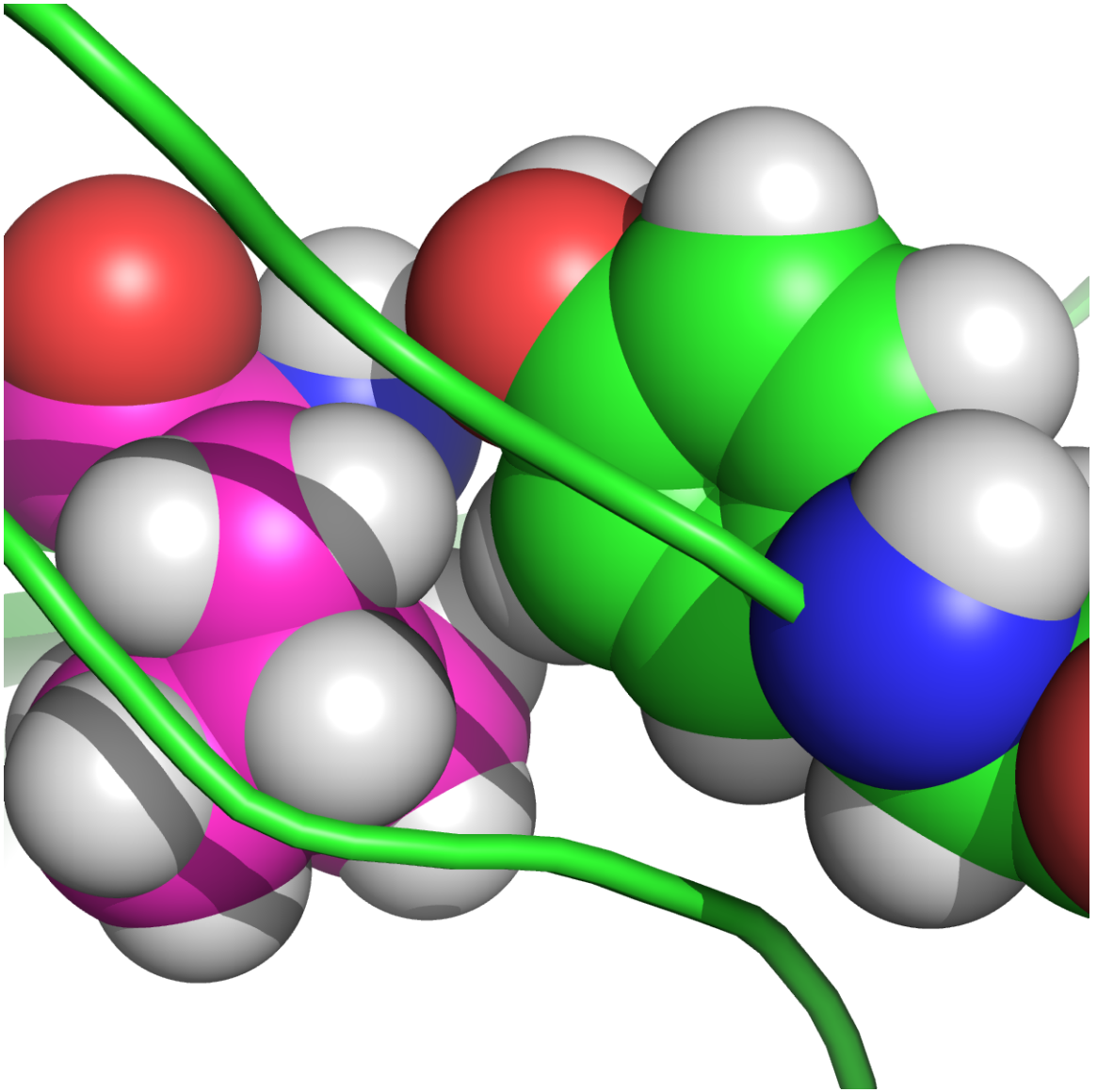
ALGORITHMS TO SEARCH ROTAMER SPACE

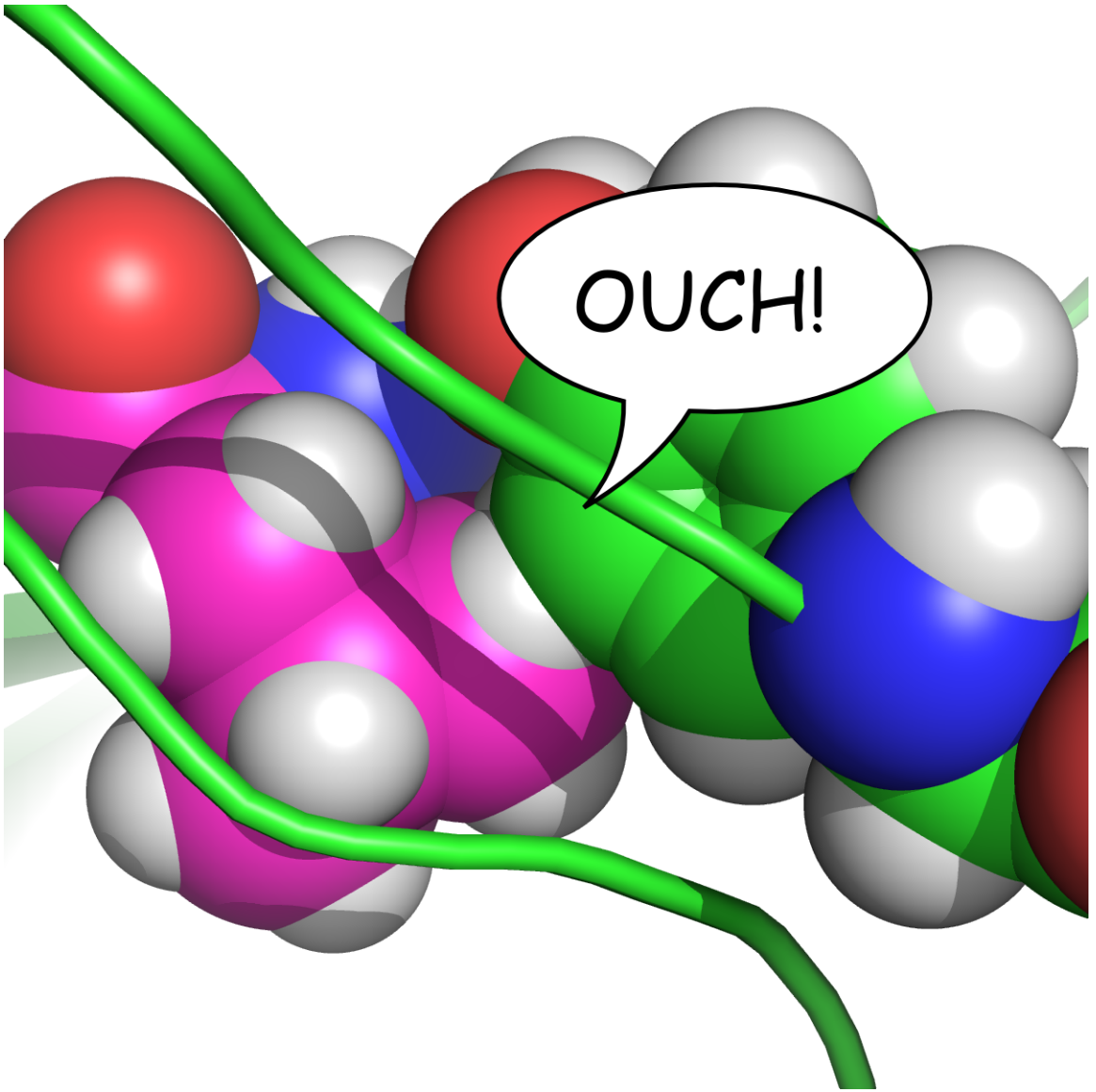
- Dead End Elimination
- Self Consistent Mean Field
- A* search
- Monte Carlo Simulated Annealing
- Graph decomposition

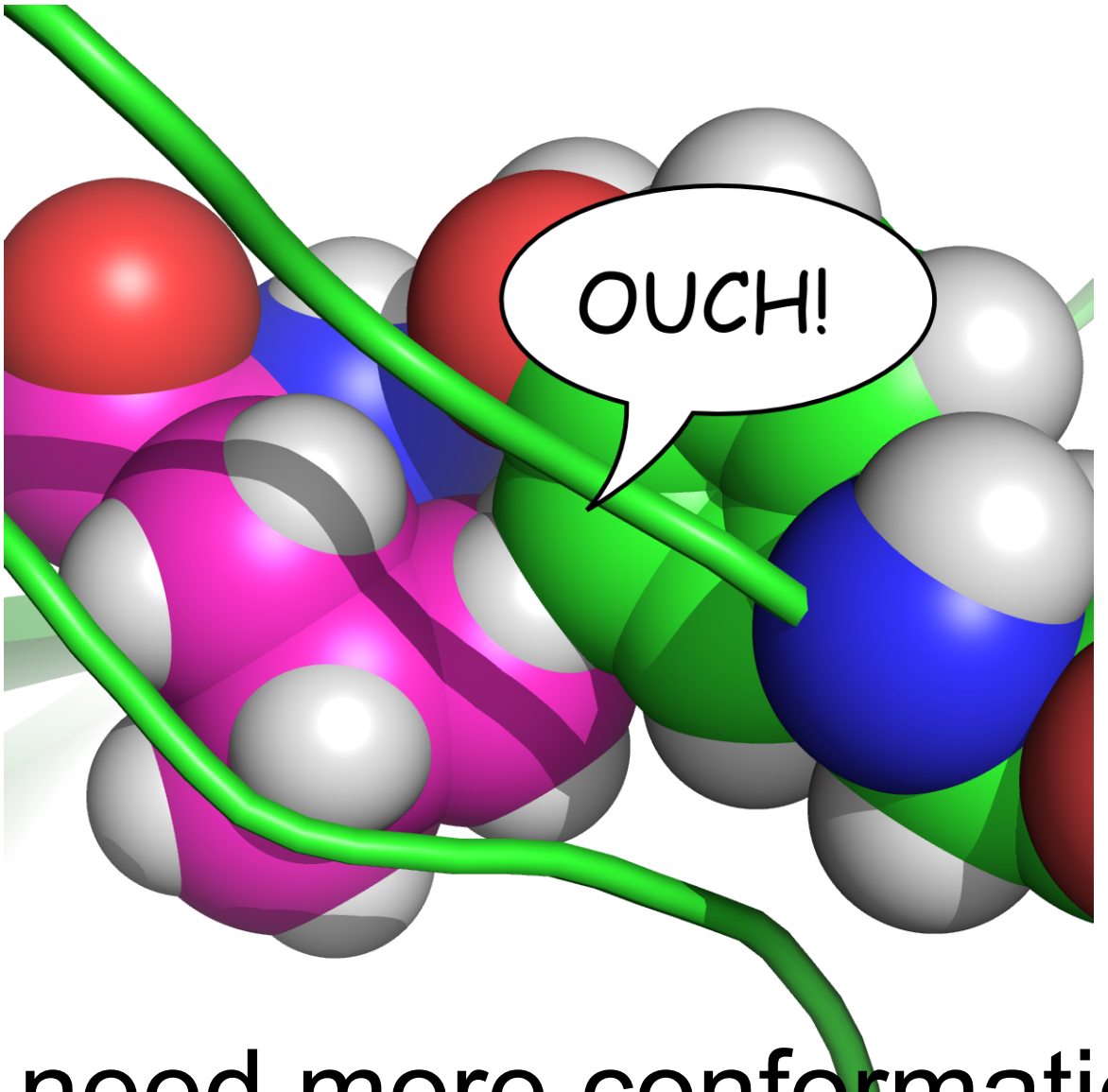
SO WHAT IS THE PROBLEM?





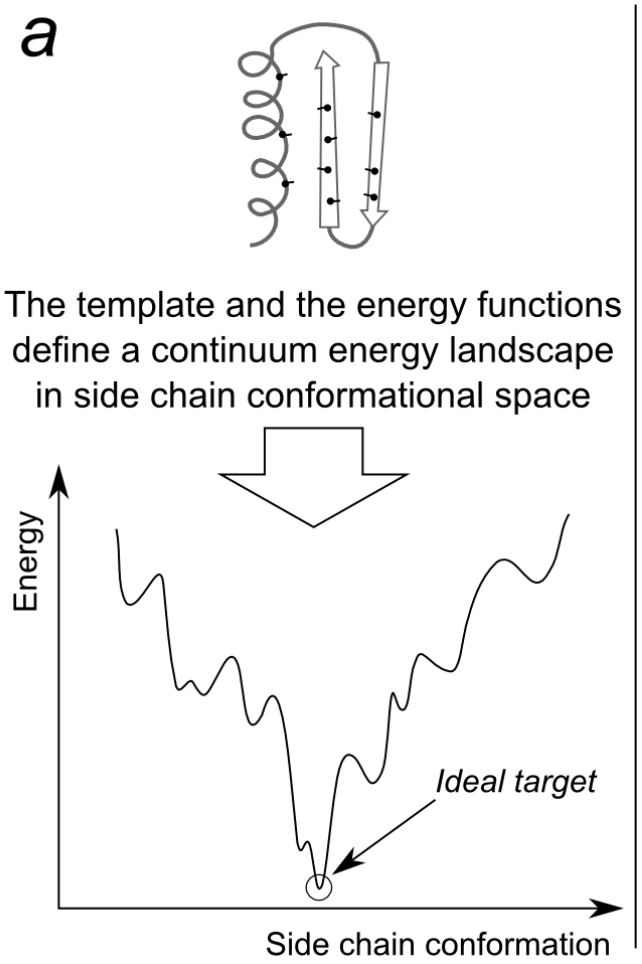



OUCH!

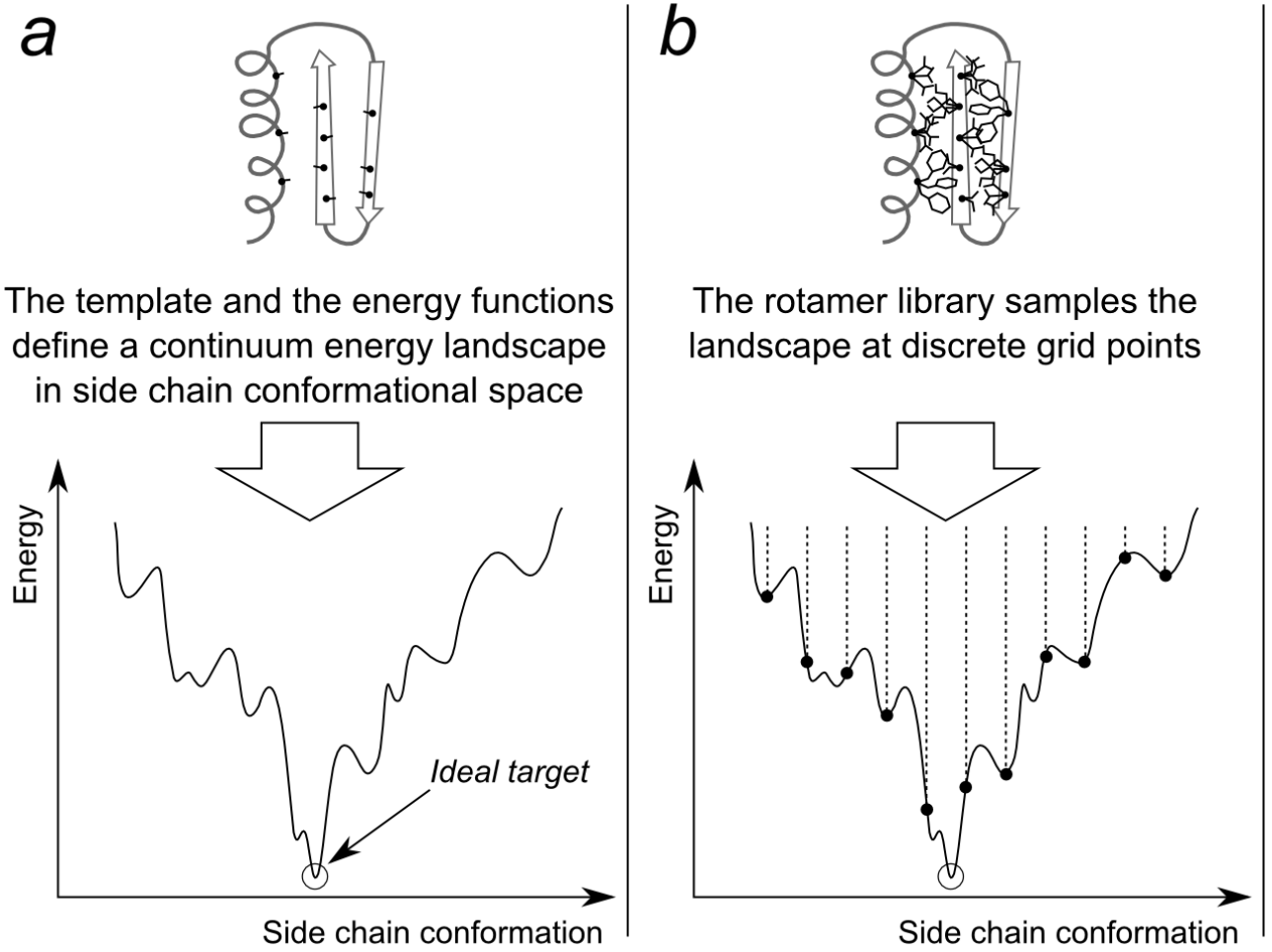


We need more conformations (sampling)

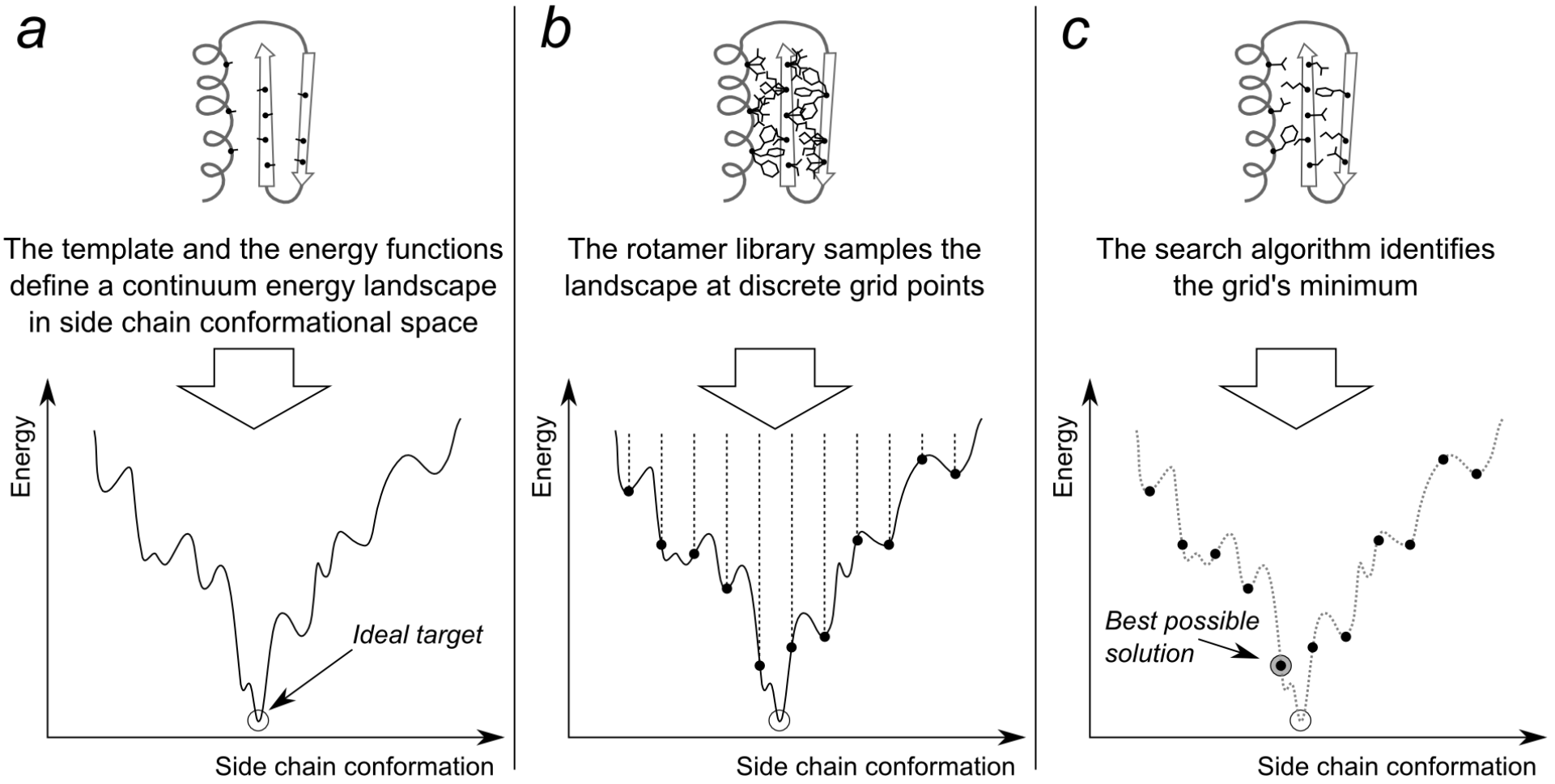
ROTAMER LIBRARY DETERMINES QUALITY OF SOLUTION

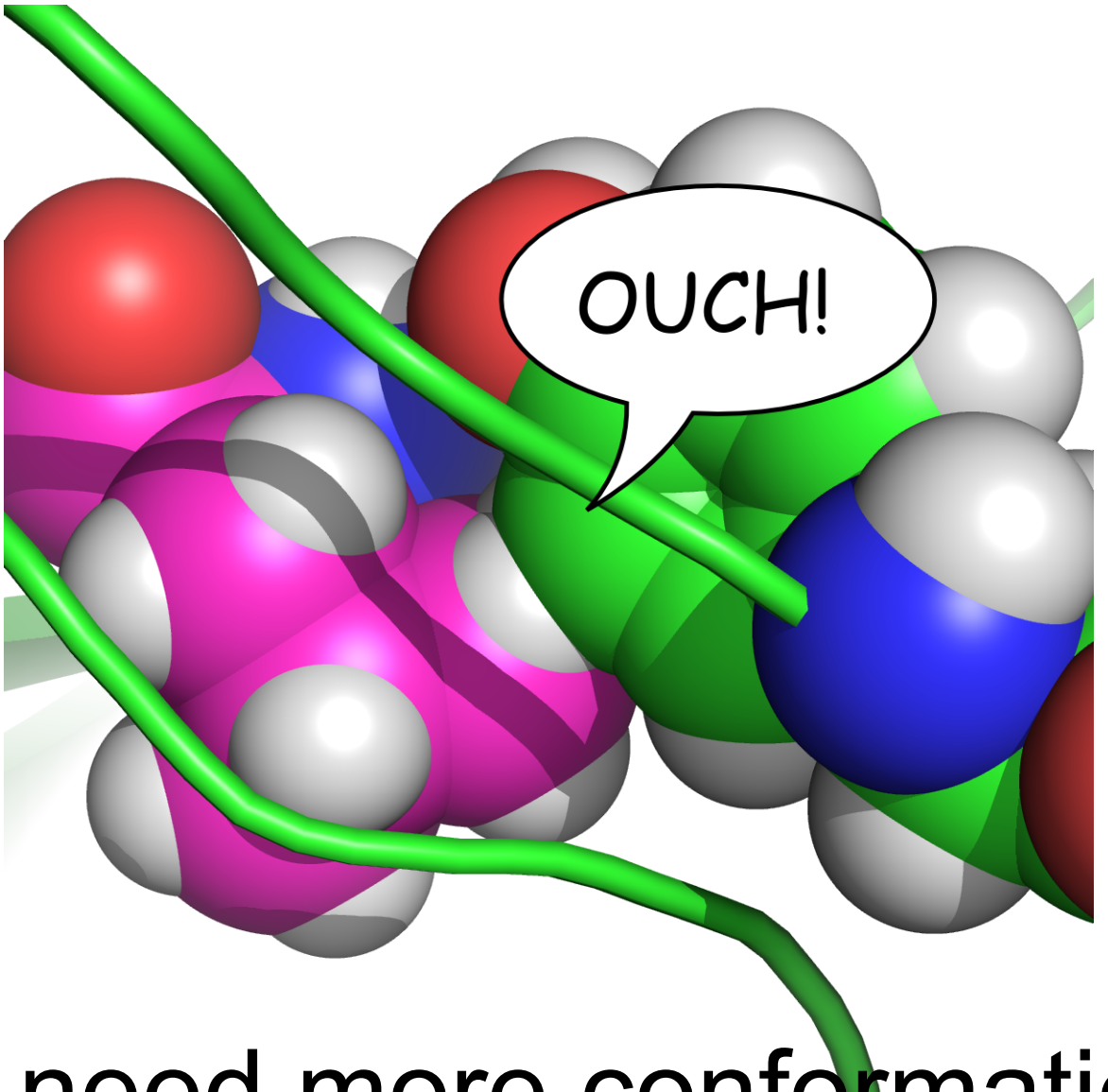


ROTAMER LIBRARY DETERMINES QUALITY OF SOLUTION



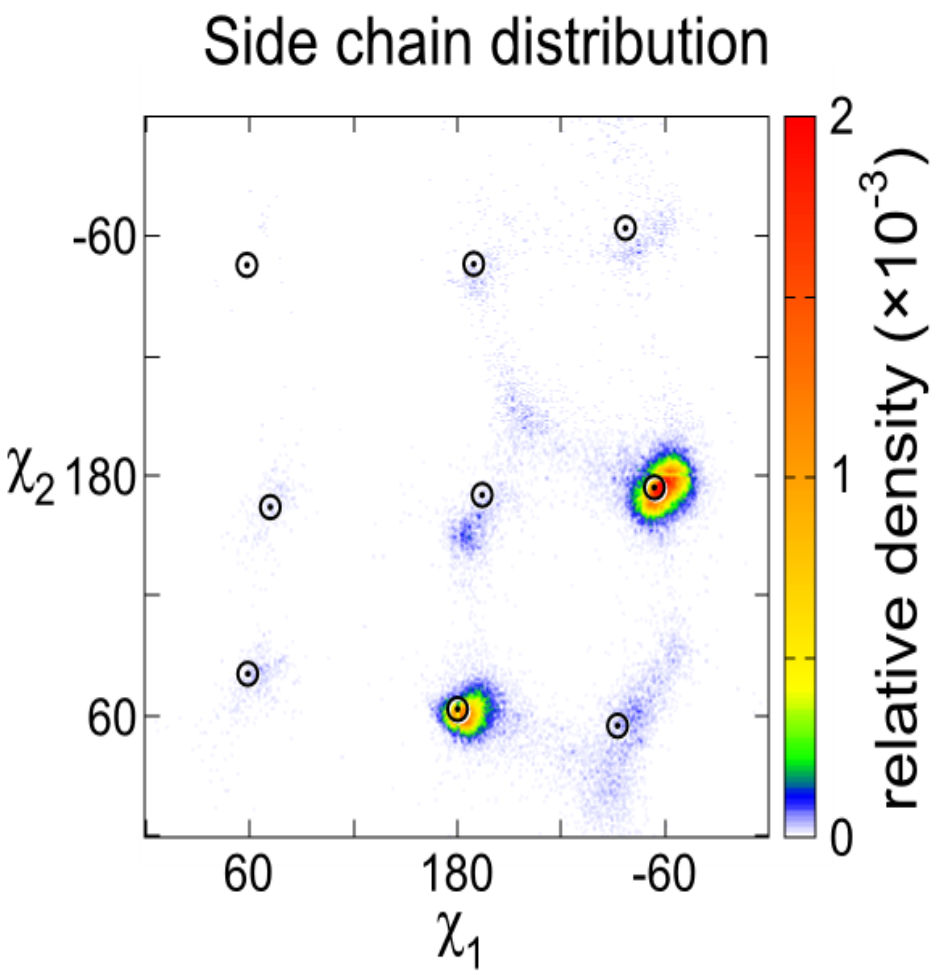
ROTAMER LIBRARY DETERMINES QUALITY OF SOLUTION



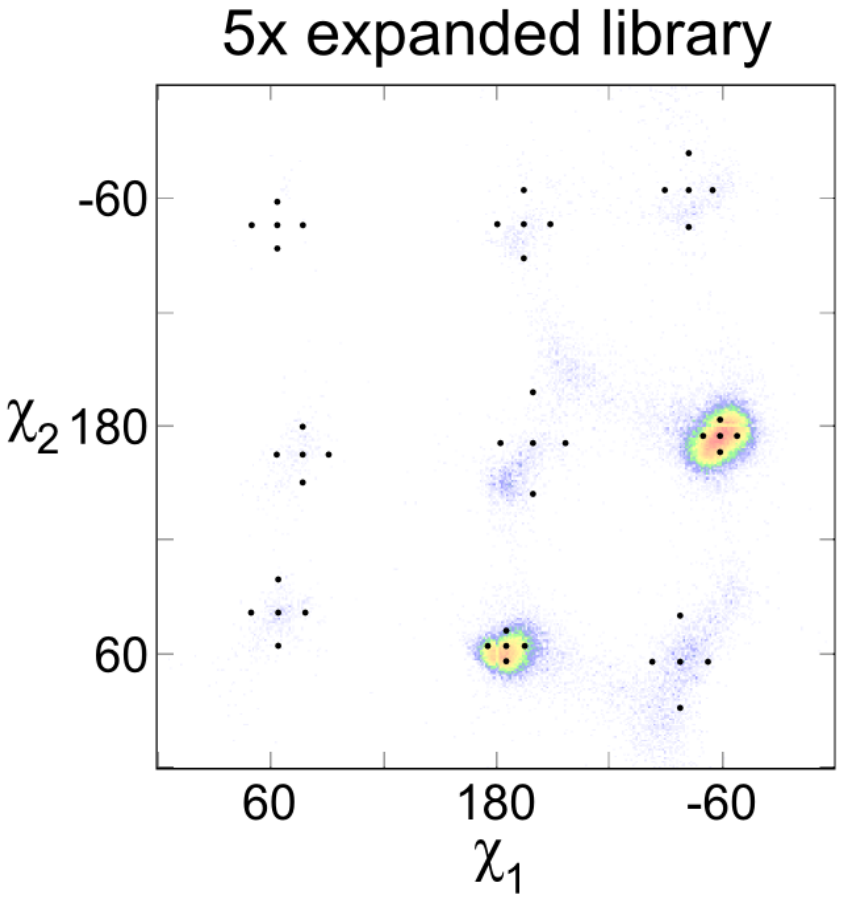


We need more conformations (sampling)

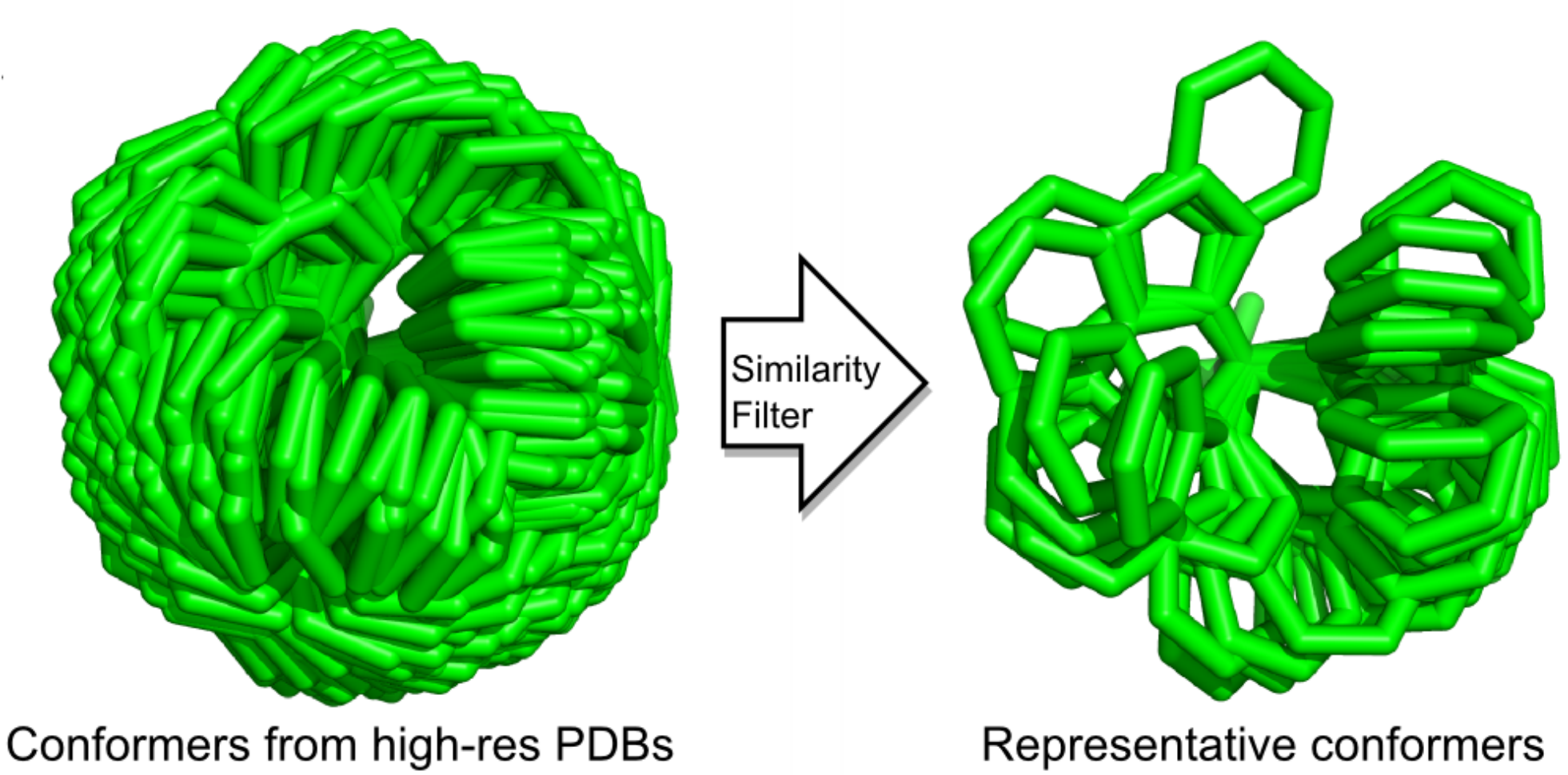
MORE SAMPLING



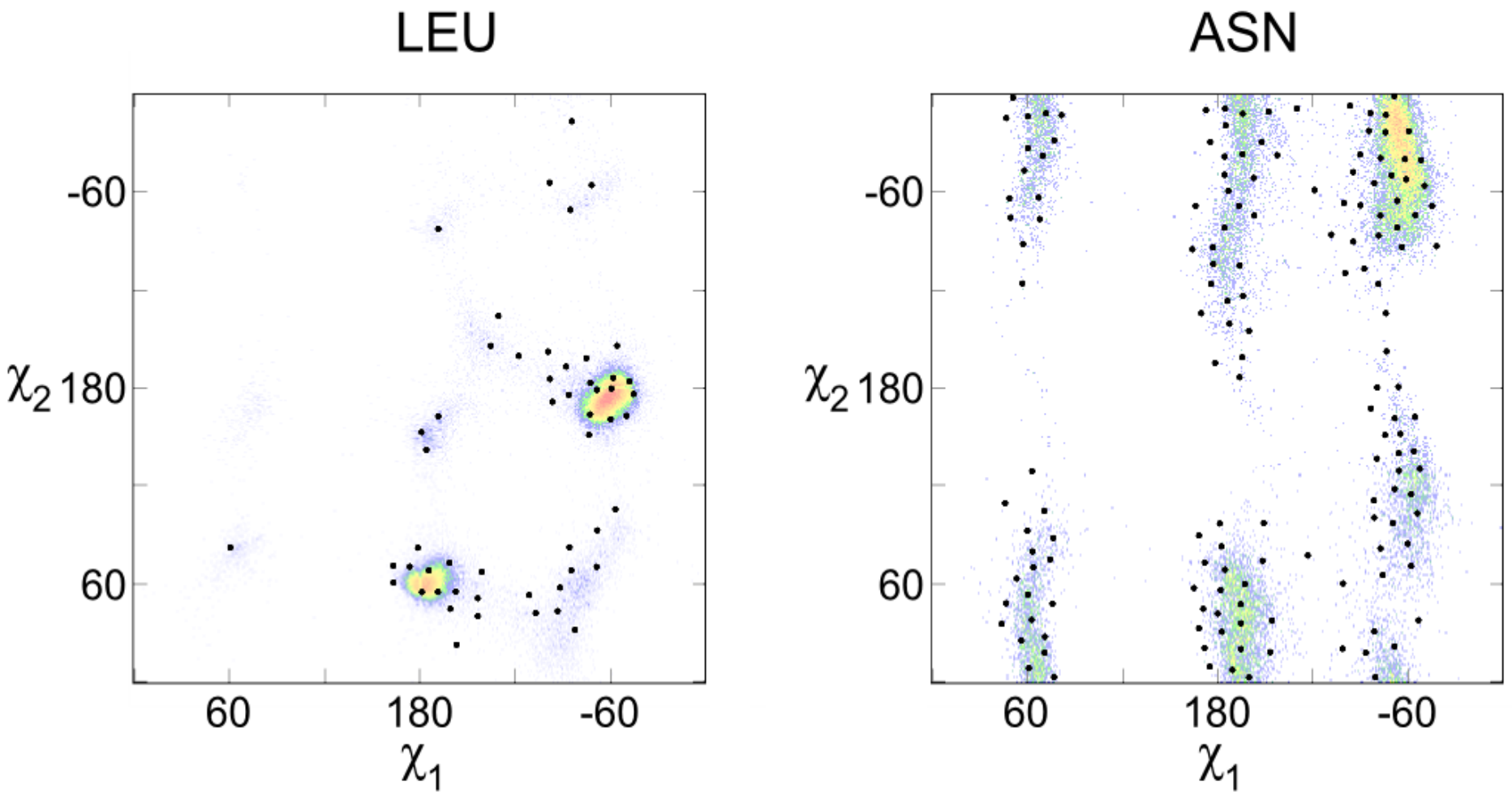
MORE SAMPLING



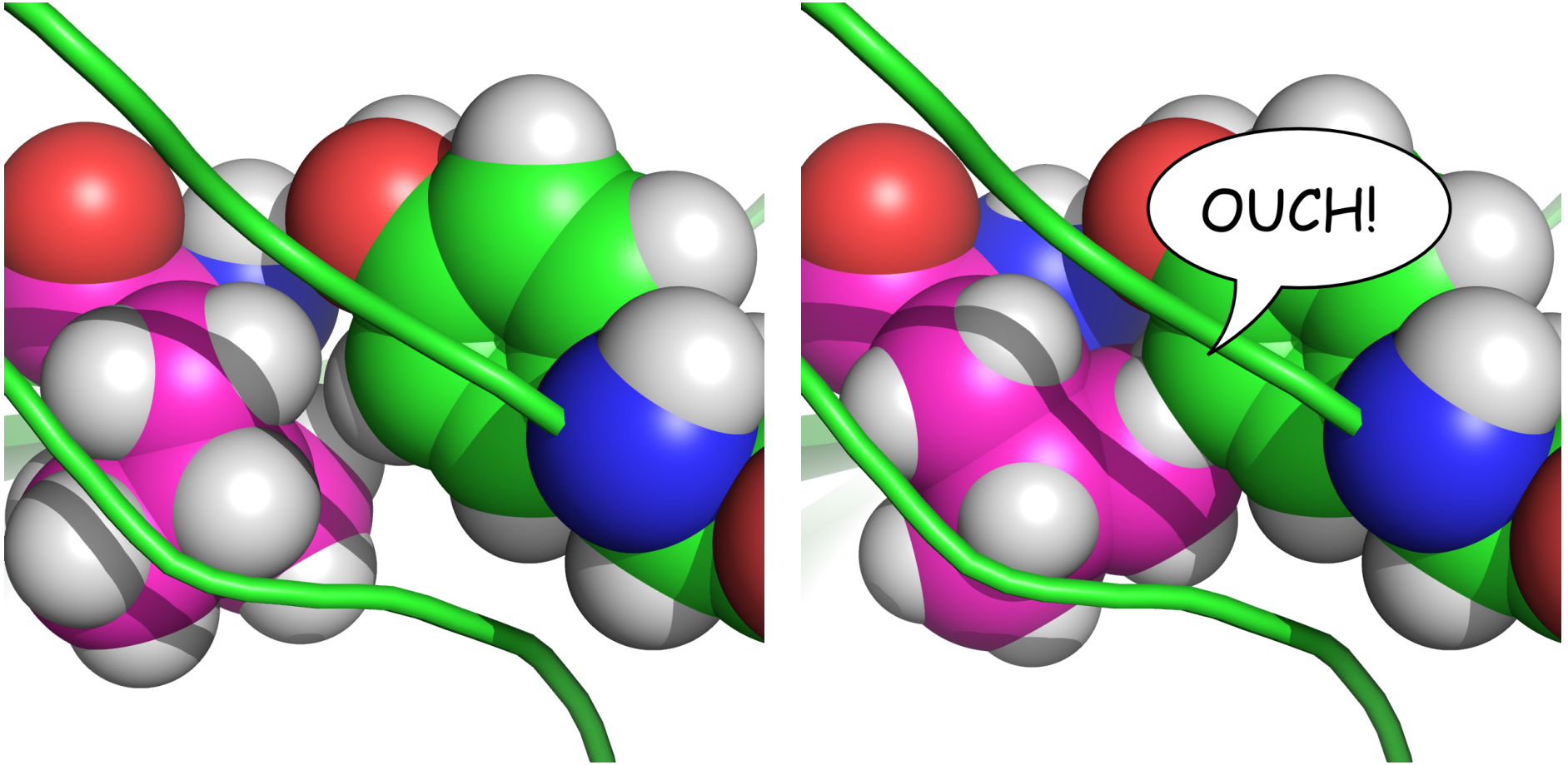
GEOMETRIC FILTERS LEAD TO CONFORMER LIBRARIES



IGNORES THE NATURAL DISTRIBUTION



- People have been looking for solutions using the statistical distribution in structures
- However, the problem is that sampling is related to the energetics in a way that is difficult to predict



- People have been looking for solutions using the statistical distribution in structures
- However, the problem is that sampling is related to the energies in a way that is difficult to predict
- Solution: use energetics to identify the best sampling strategy for side chain optimization

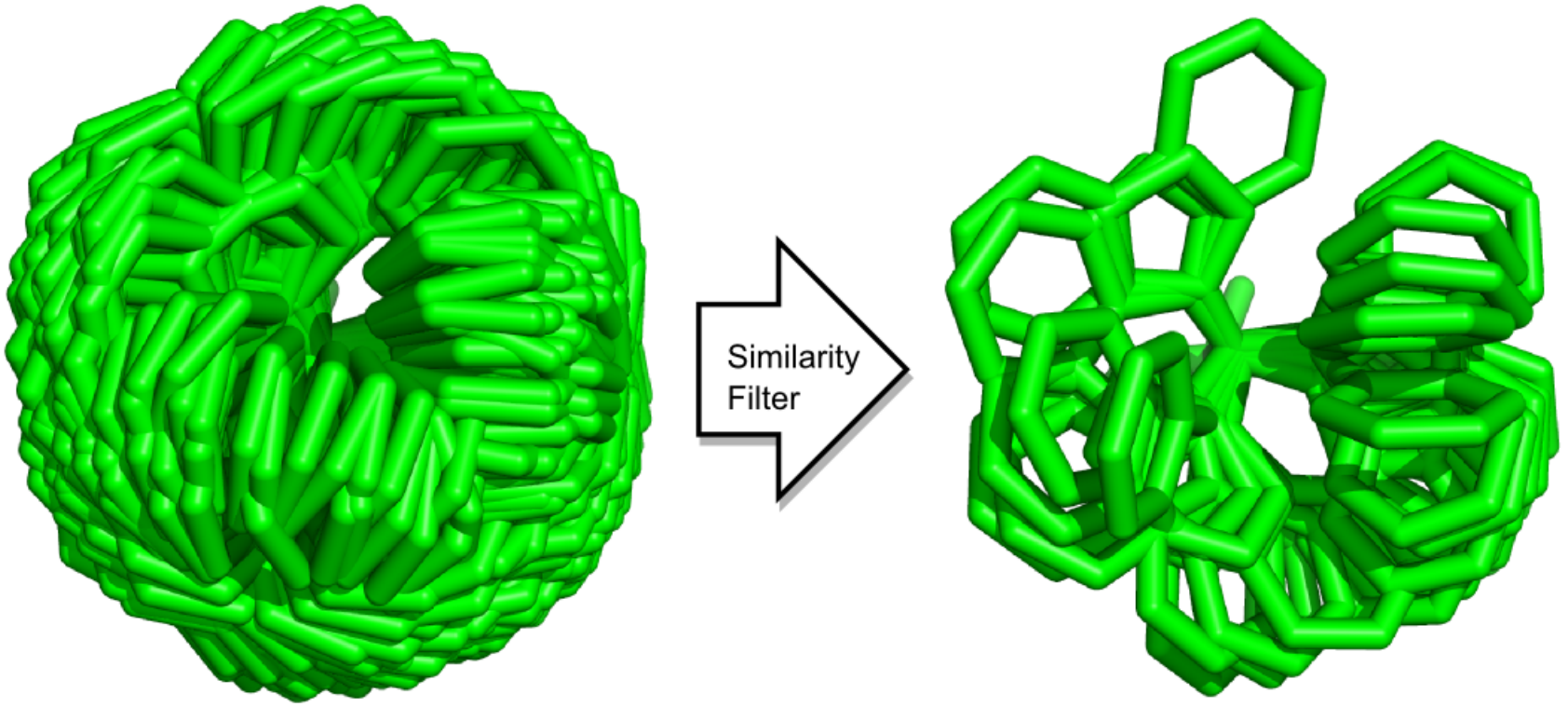
BUILDING AN ENERGY BASED CONFORMER LIBRARY

GOALS

Can we create a library that can outperform existing libraries in terms of speed and/or Energies ?

Can we create a flexible library where the conformers are in some useful order?

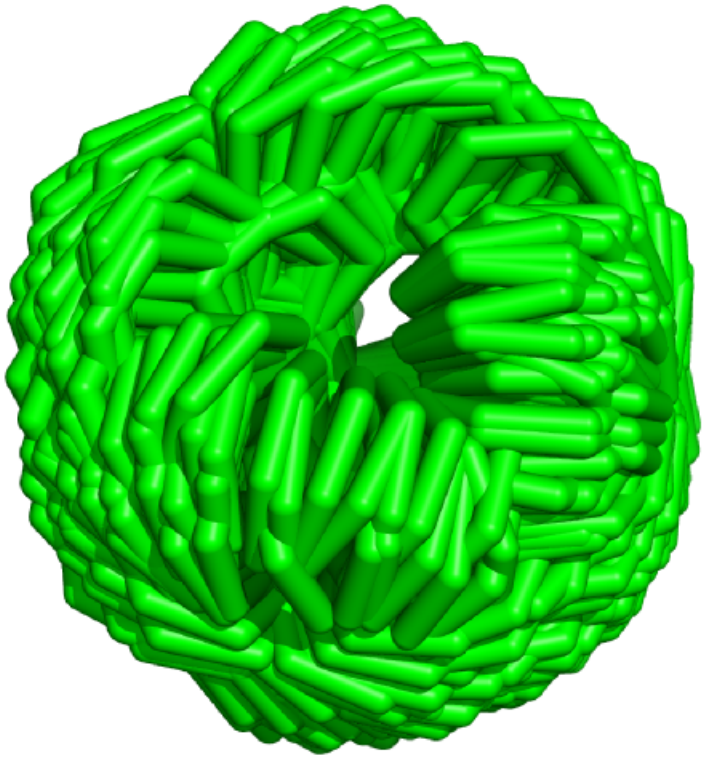
Can we sort the conformers instead of extracting a fixed-size subset?



5,000 conformers from high-res PDBs

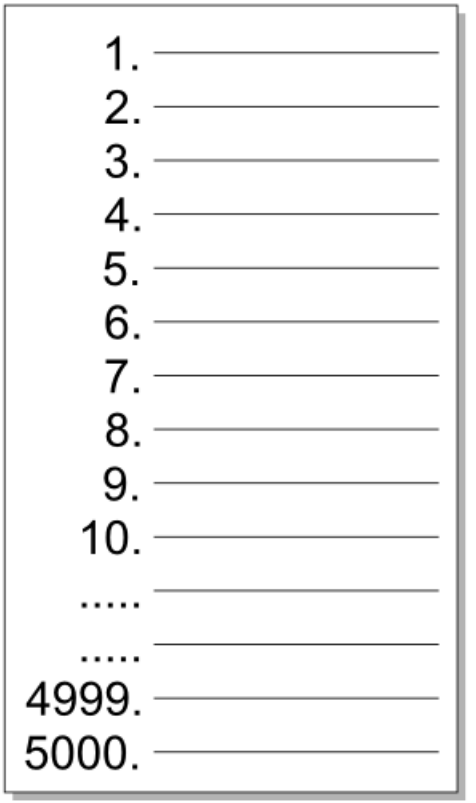
50 representative conformers

Can we sort the conformers instead of extracting a fixed-size subset?



5,000 conformers from high-res PDBs

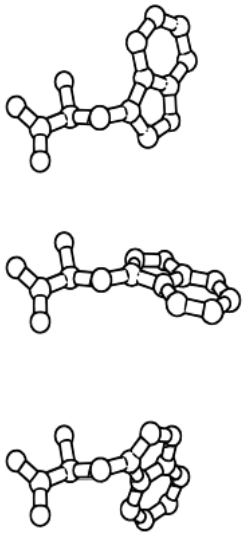
Energy based sorting



Sorted list of 5,000 conformers

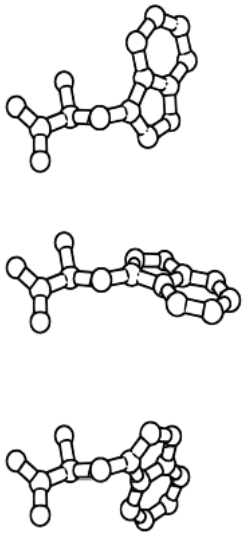
- Use energetics to identify the best sampling strategy for side chain optimization

Conformers

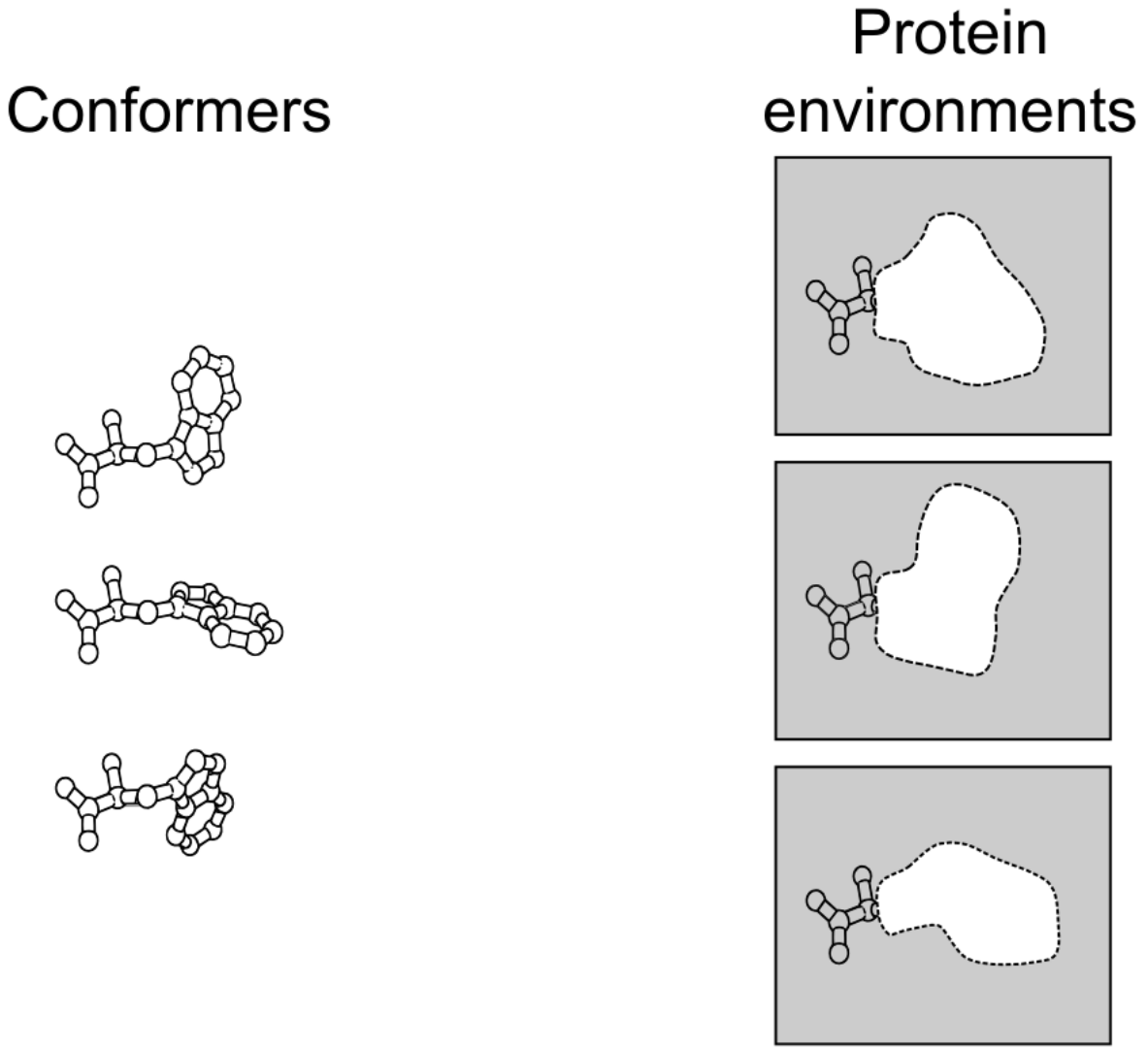


- Use energetics to identify the best sampling strategy for side chain optimization

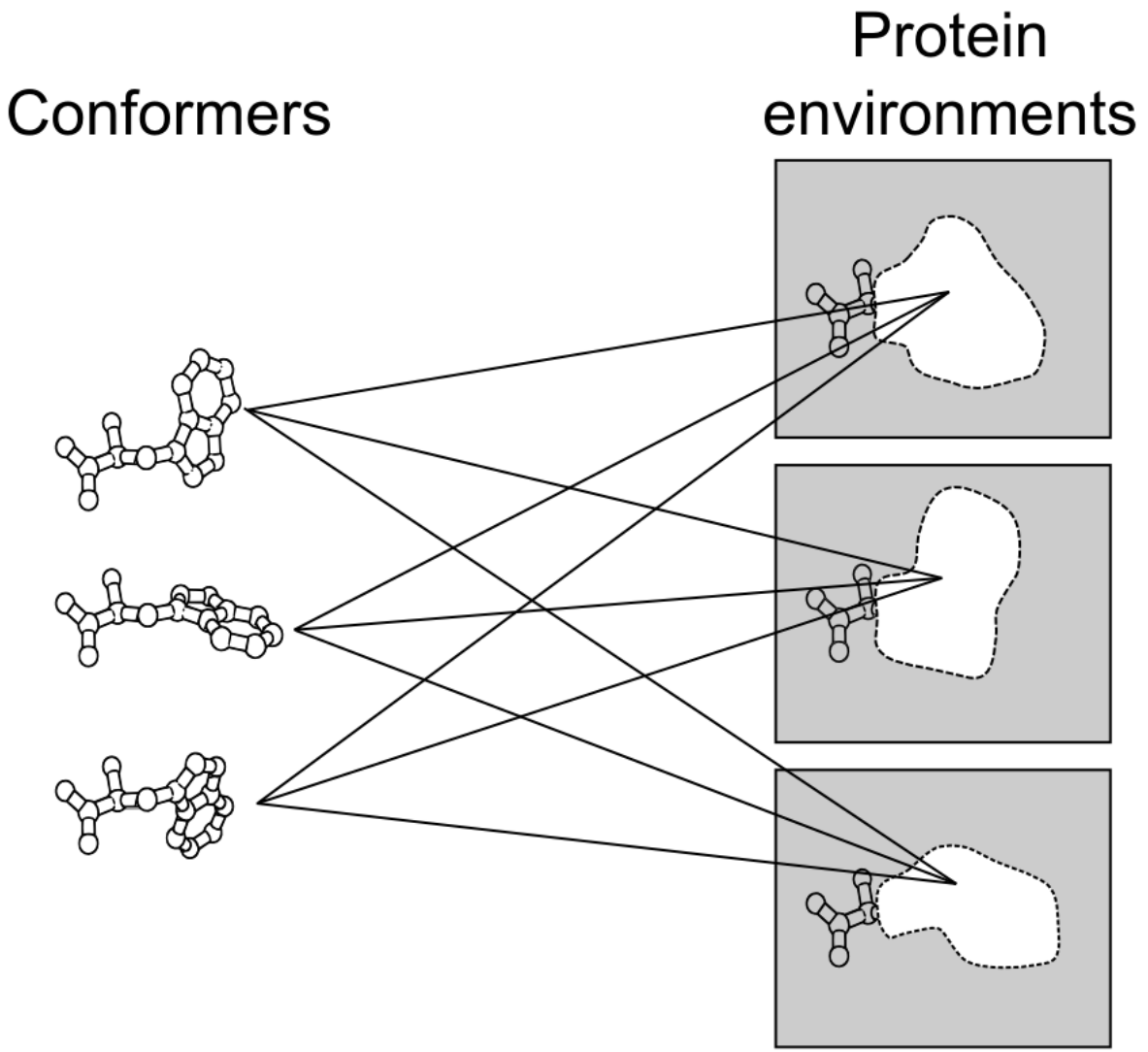
Conformers



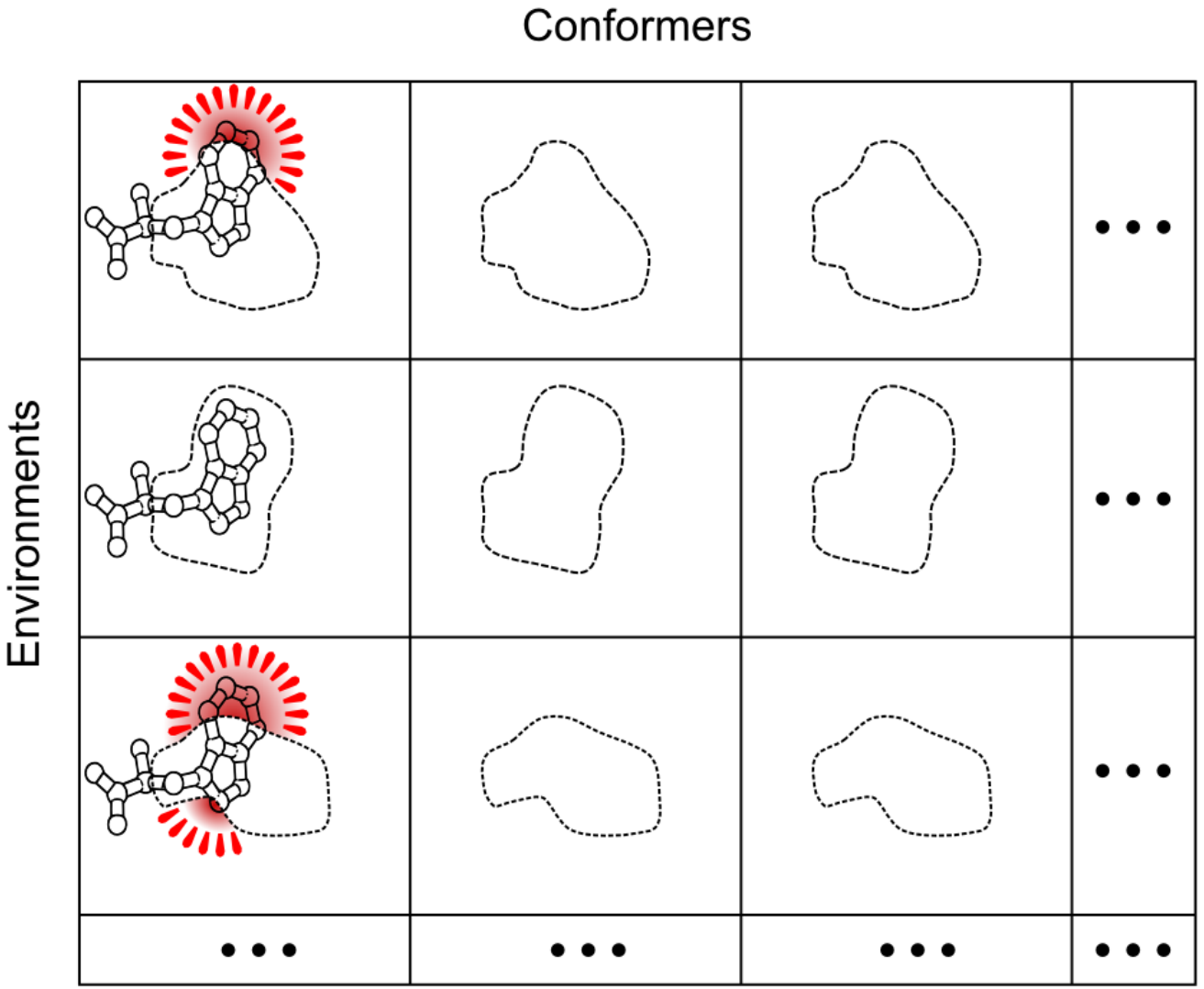
- Use energetics to identify the best sampling strategy for side chain optimization



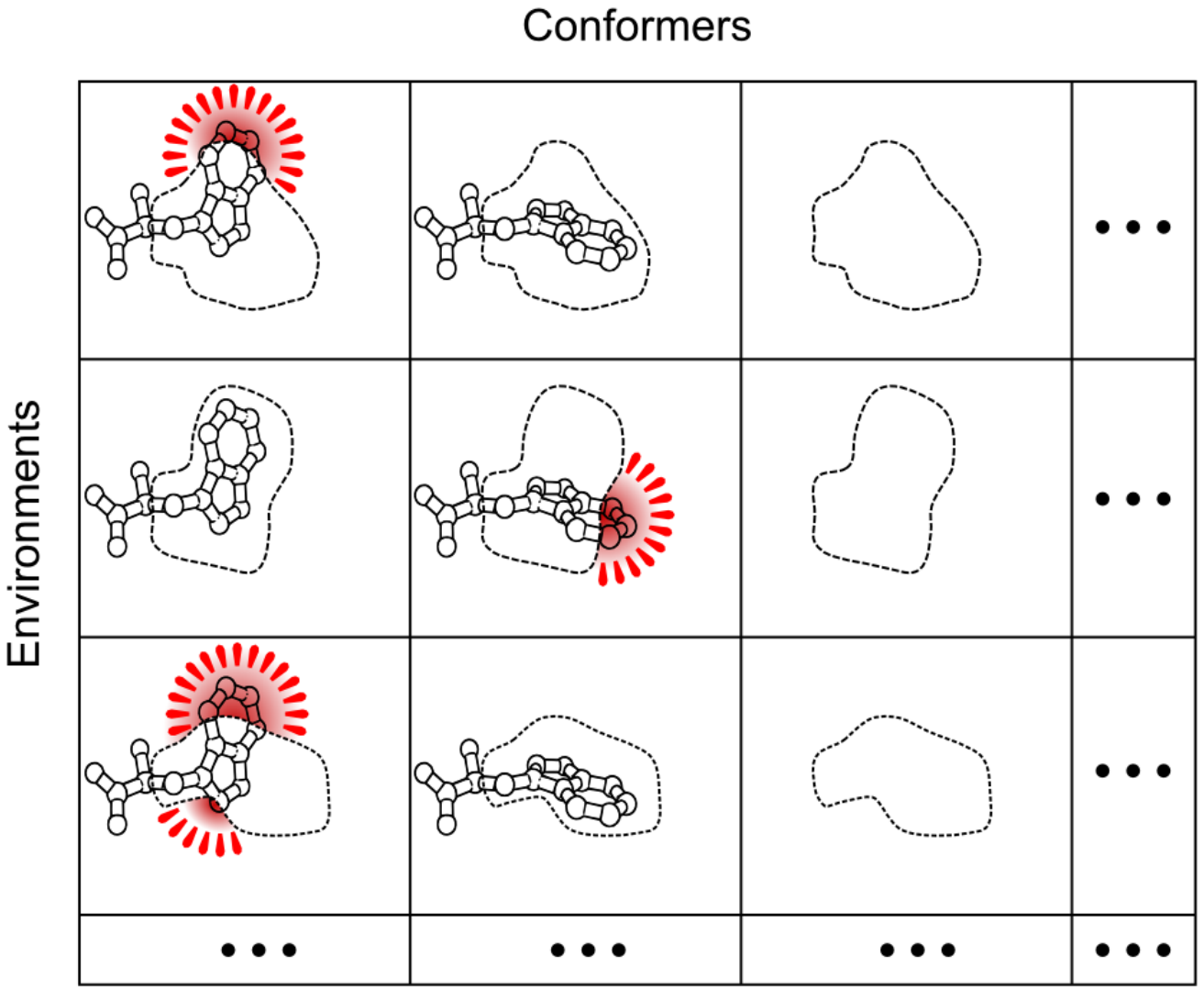
- Use energetics to identify the best sampling strategy for side chain optimization



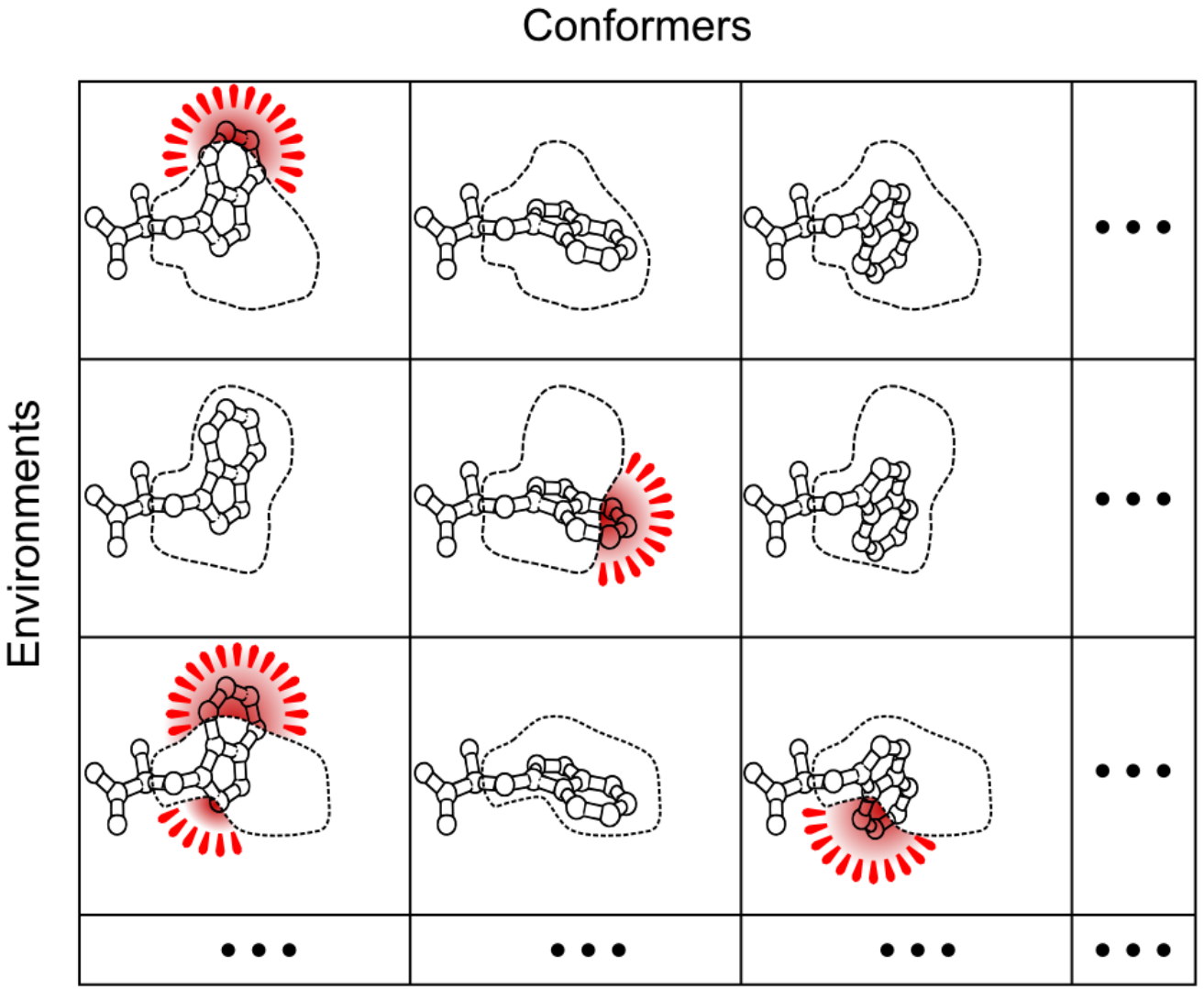
- Use energetics to identify the best sampling strategy for side chain optimization



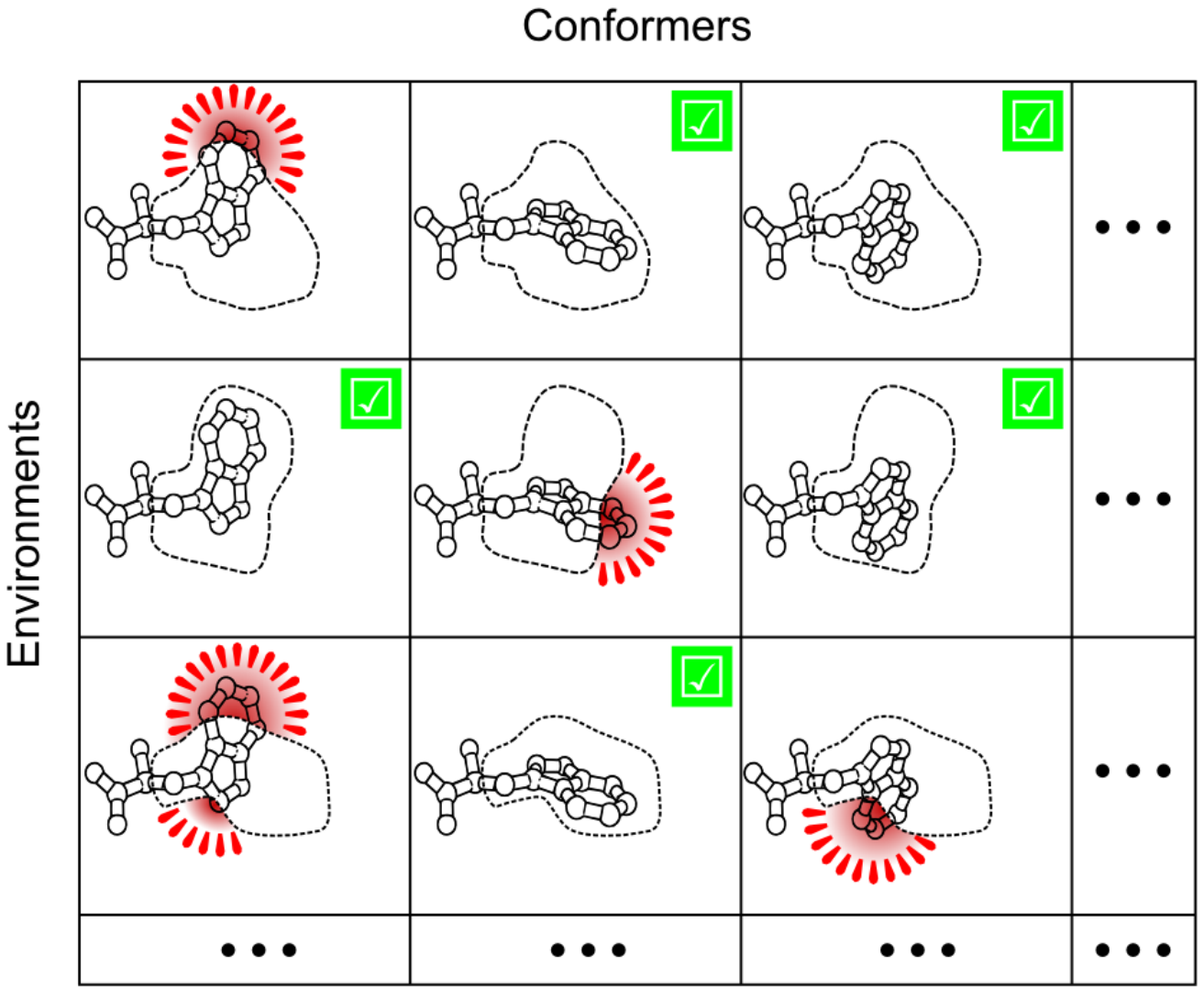
- Use energetics to identify the best sampling strategy for side chain optimization

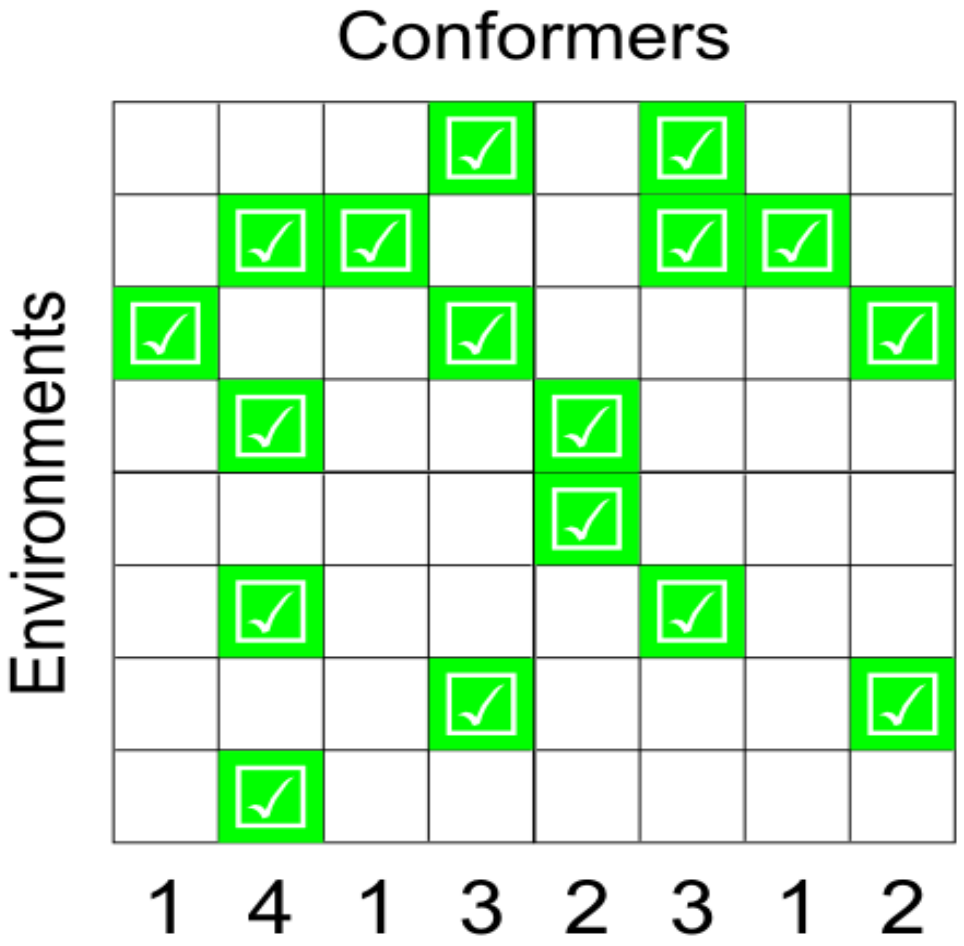


- Use energetics to identify the best sampling strategy for side chain optimization

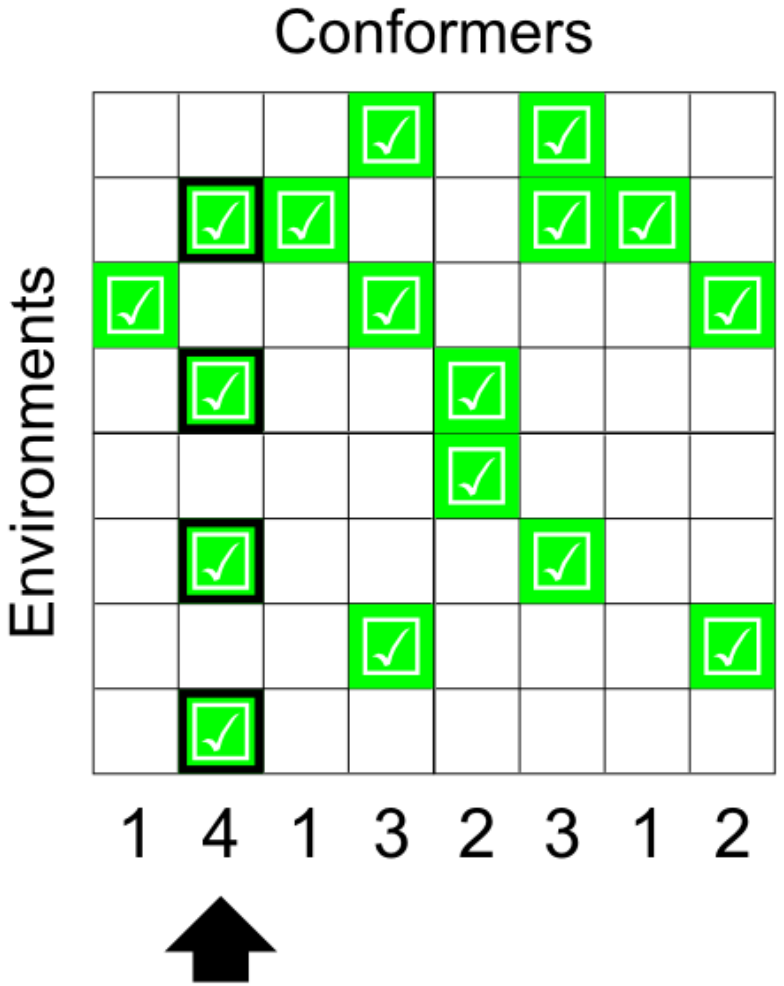


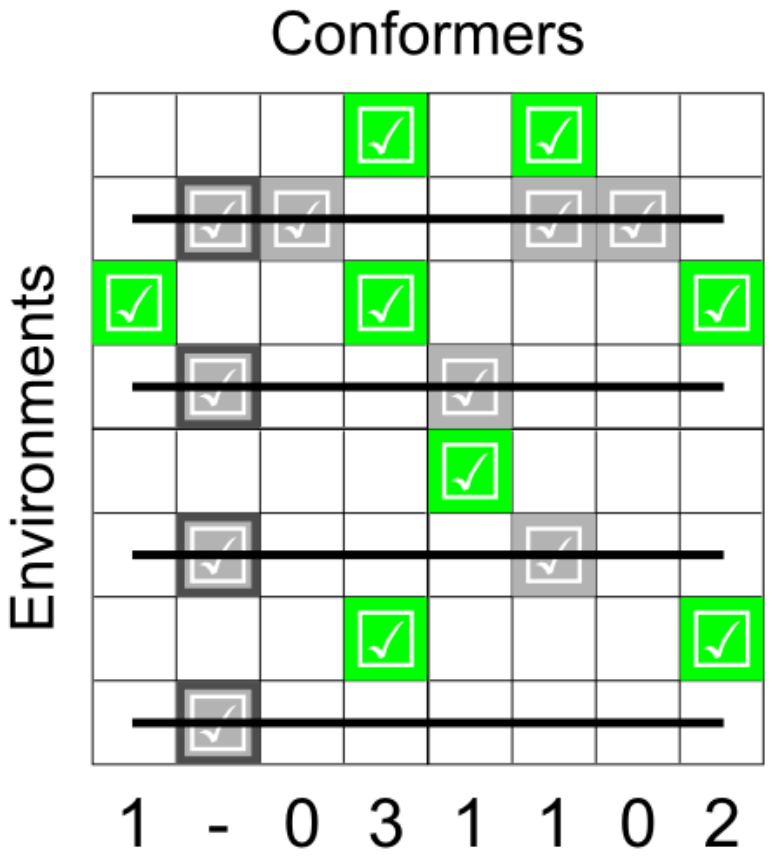
- Use energetics to identify the best sampling strategy for side chain optimization




Conformers
Environments
1 4 1 3 2 3 1 2

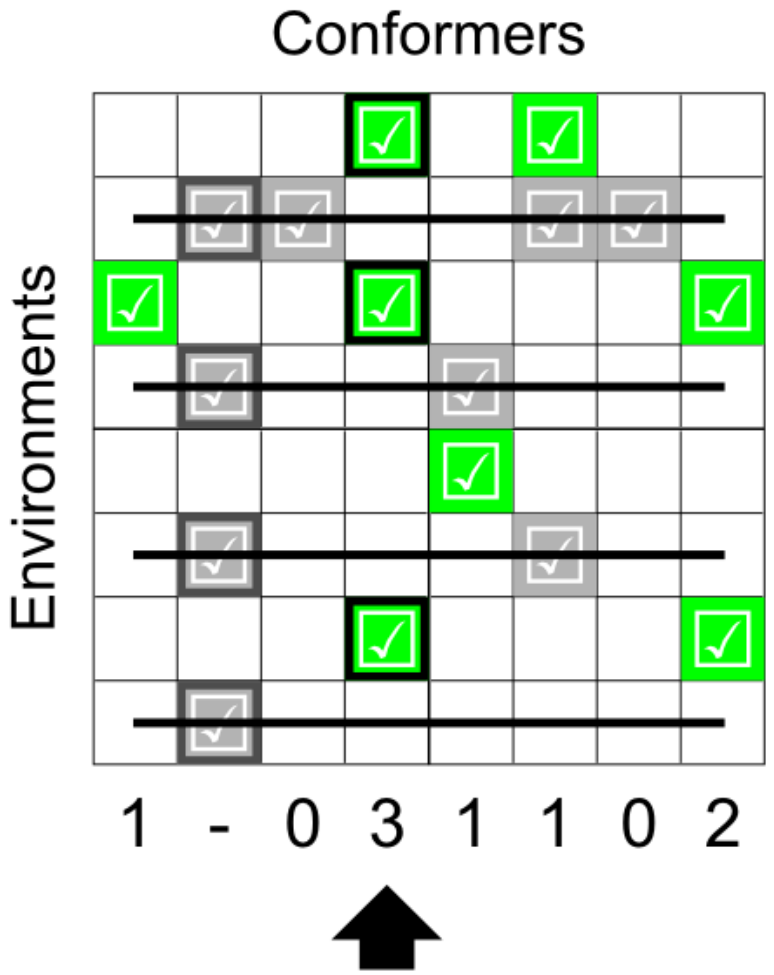
THE first conformer

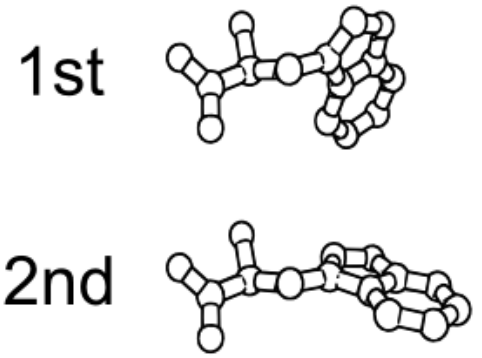




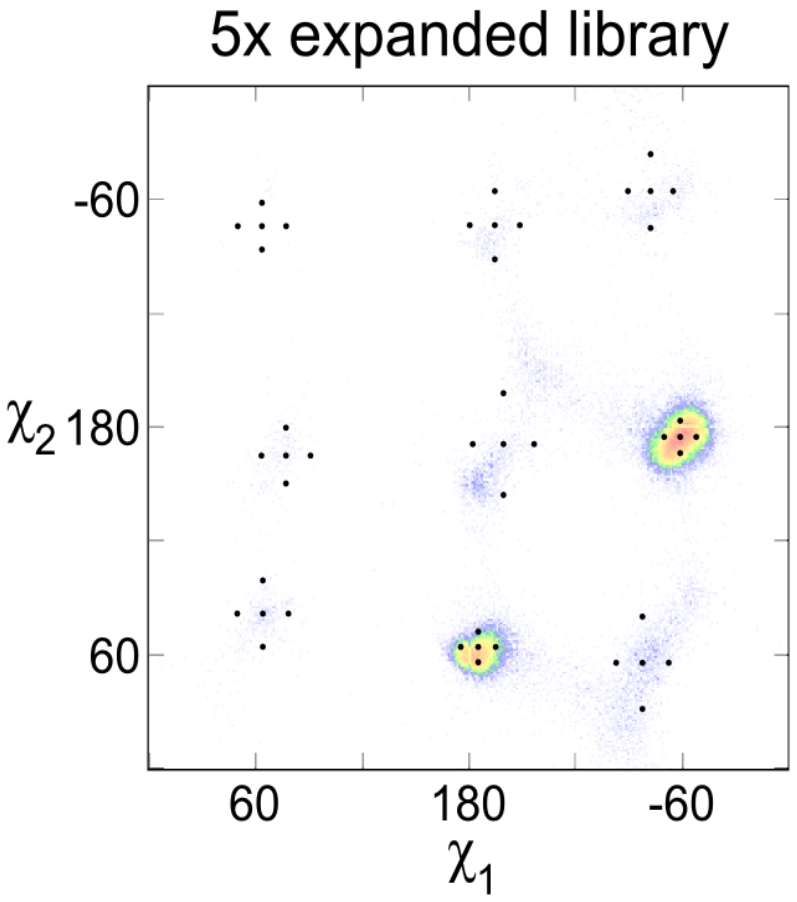
1st

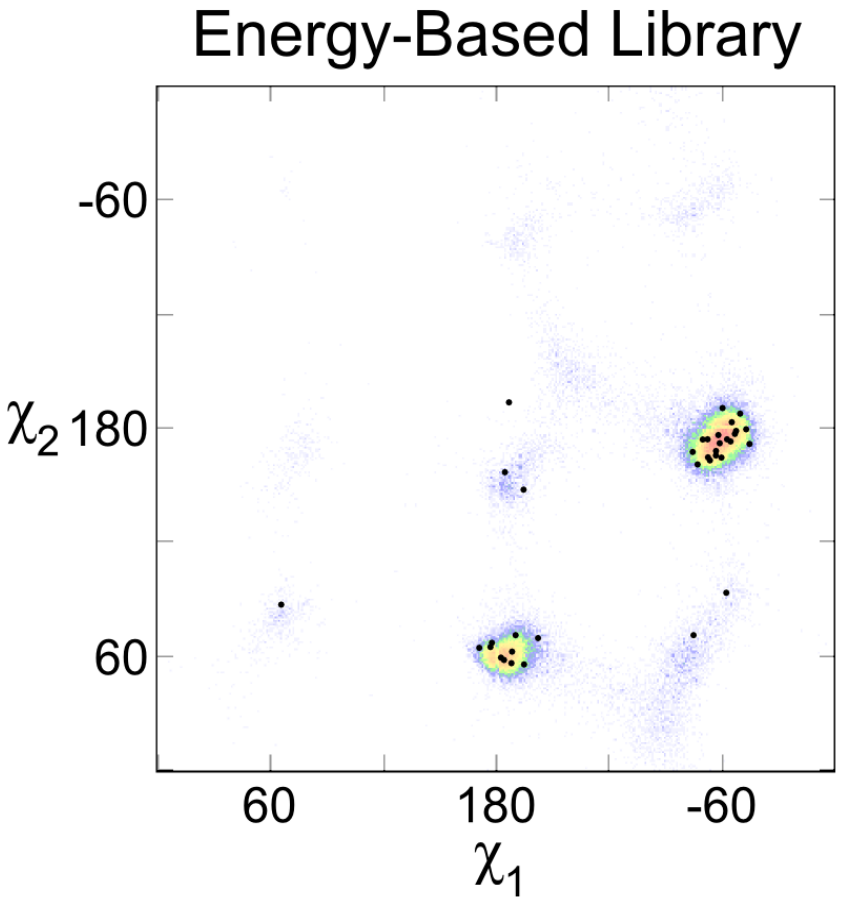
THE second conformer

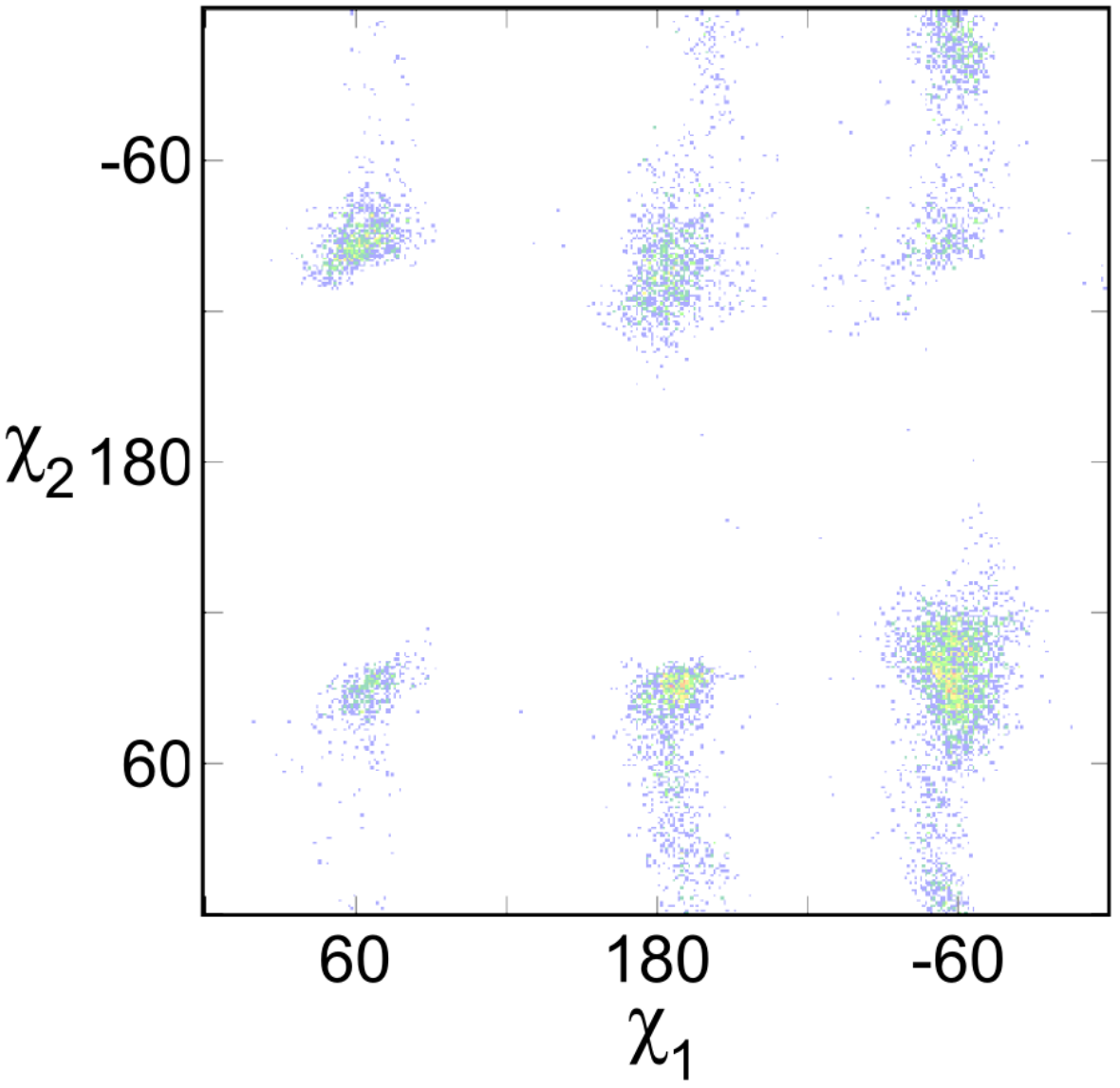




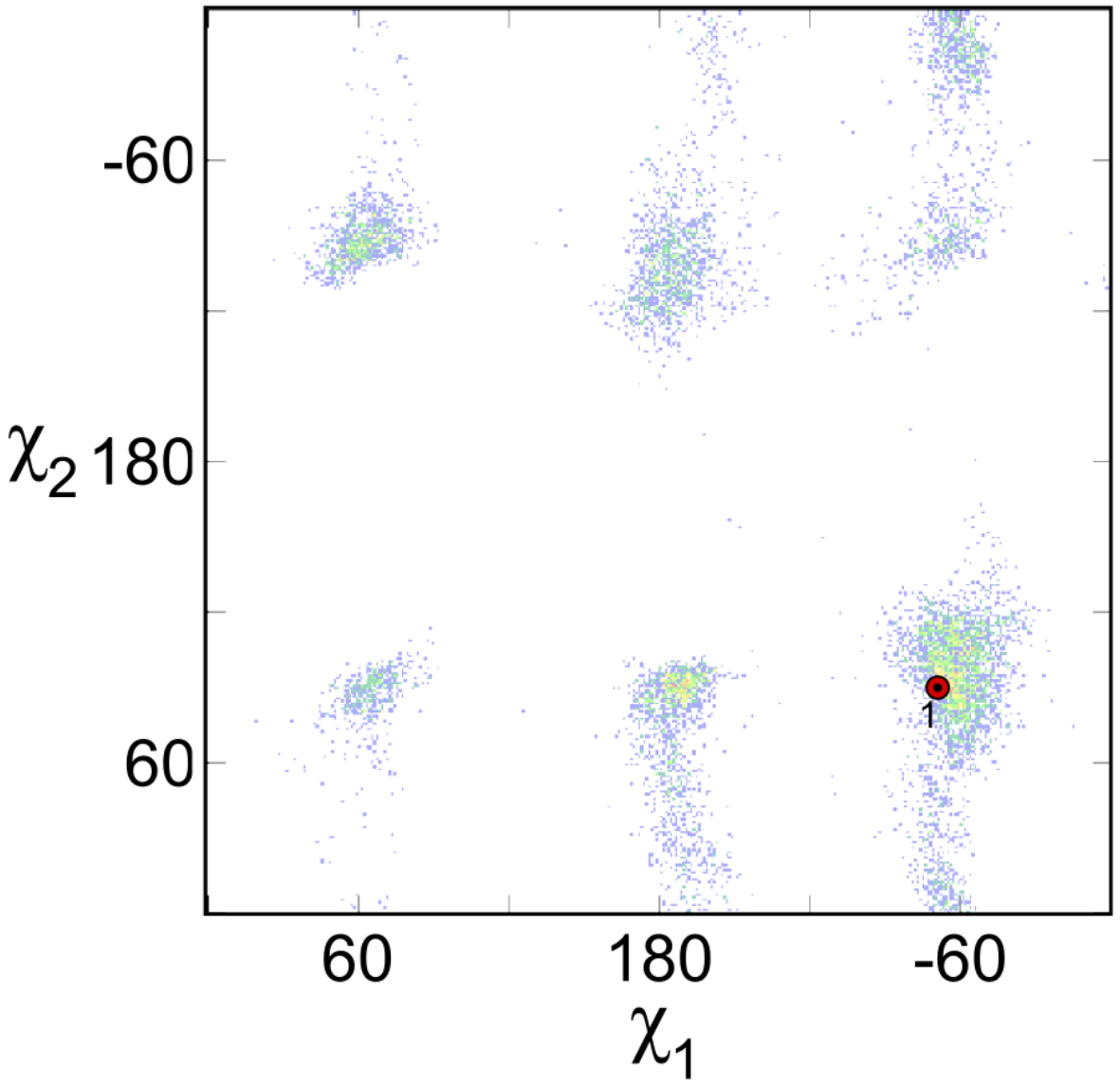
CONFORMERS IN PROPORTION TO DISTRIBUTION

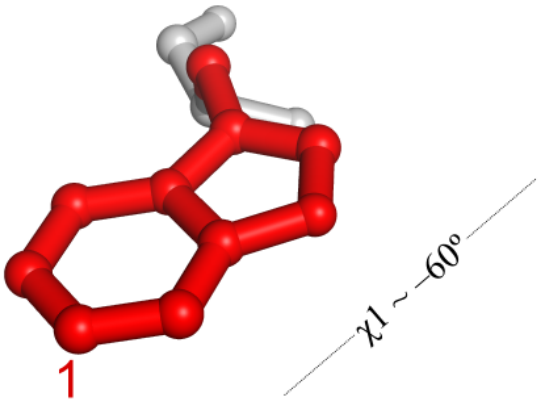


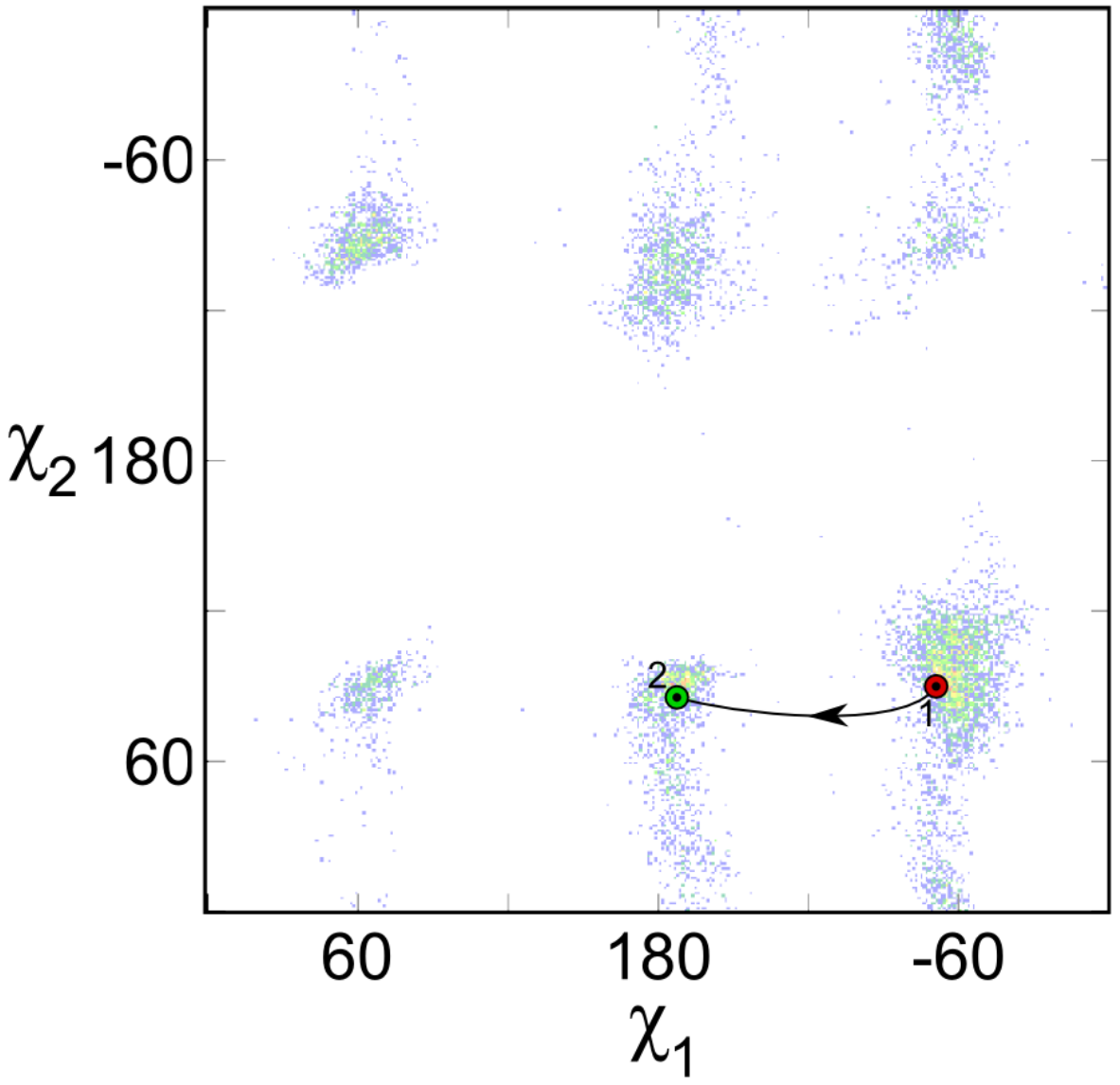


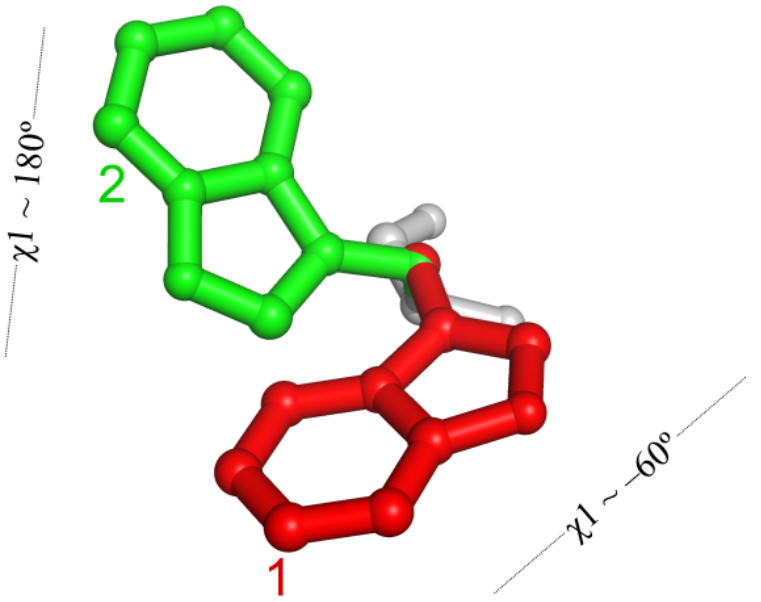


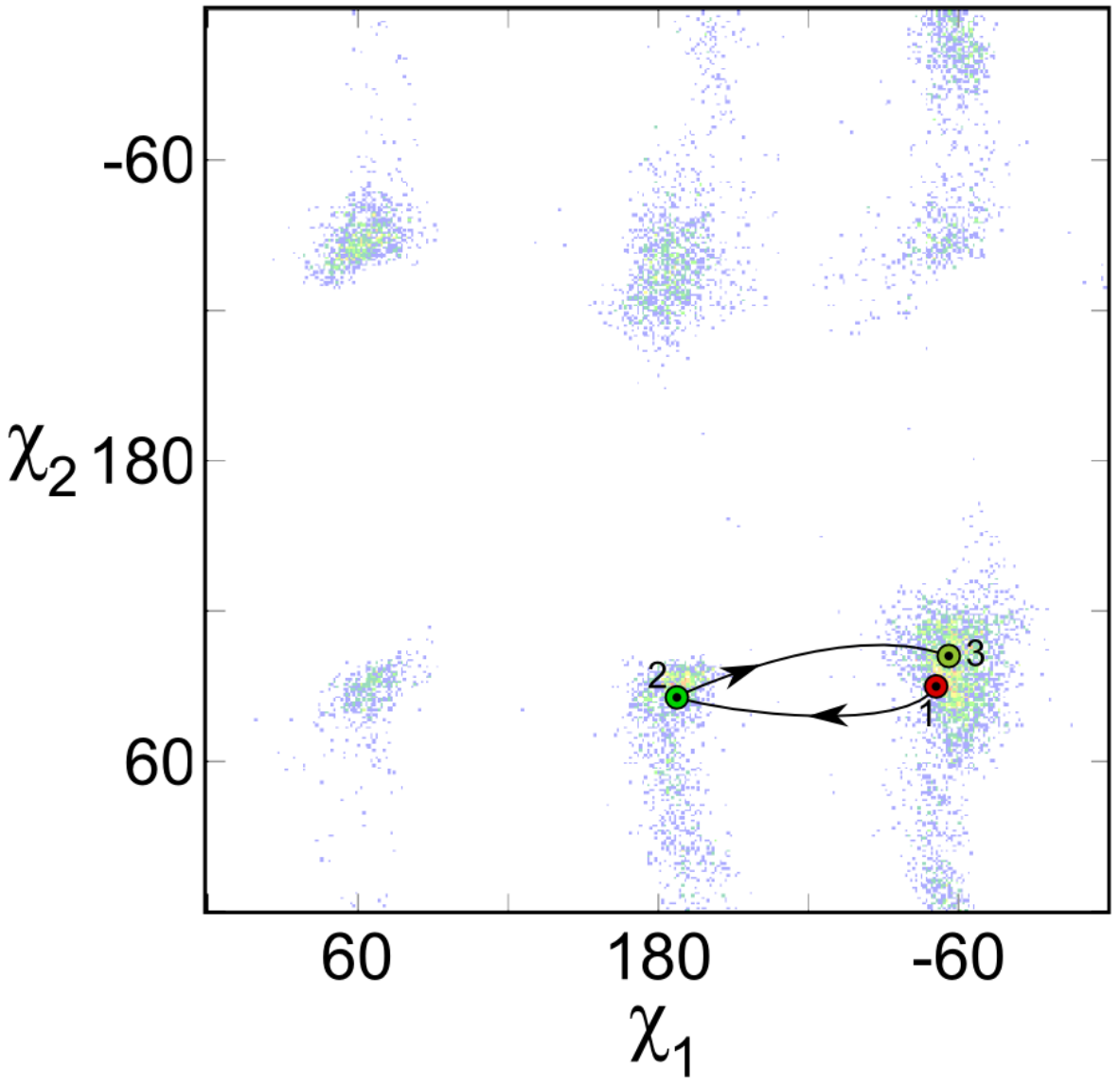
A walk in Trp space

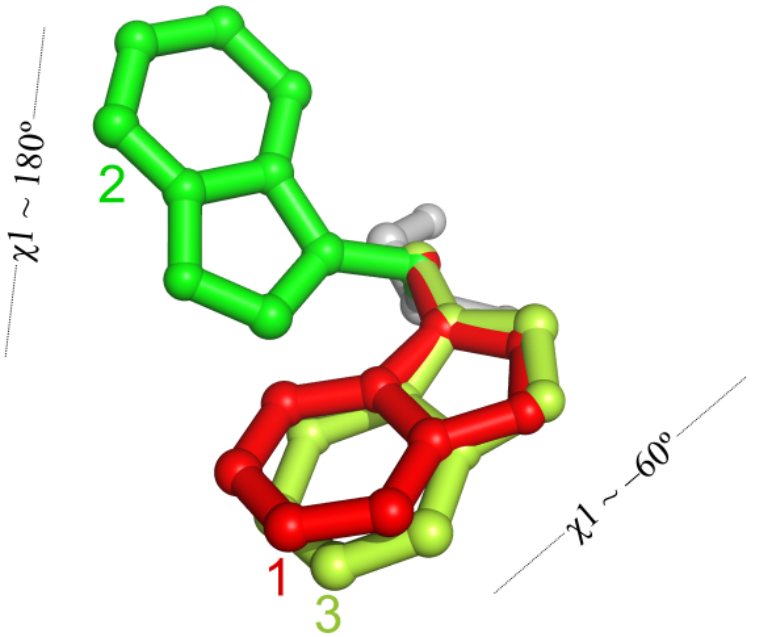

-60
χ_2 180
60
60
180
-60
χ_1
1

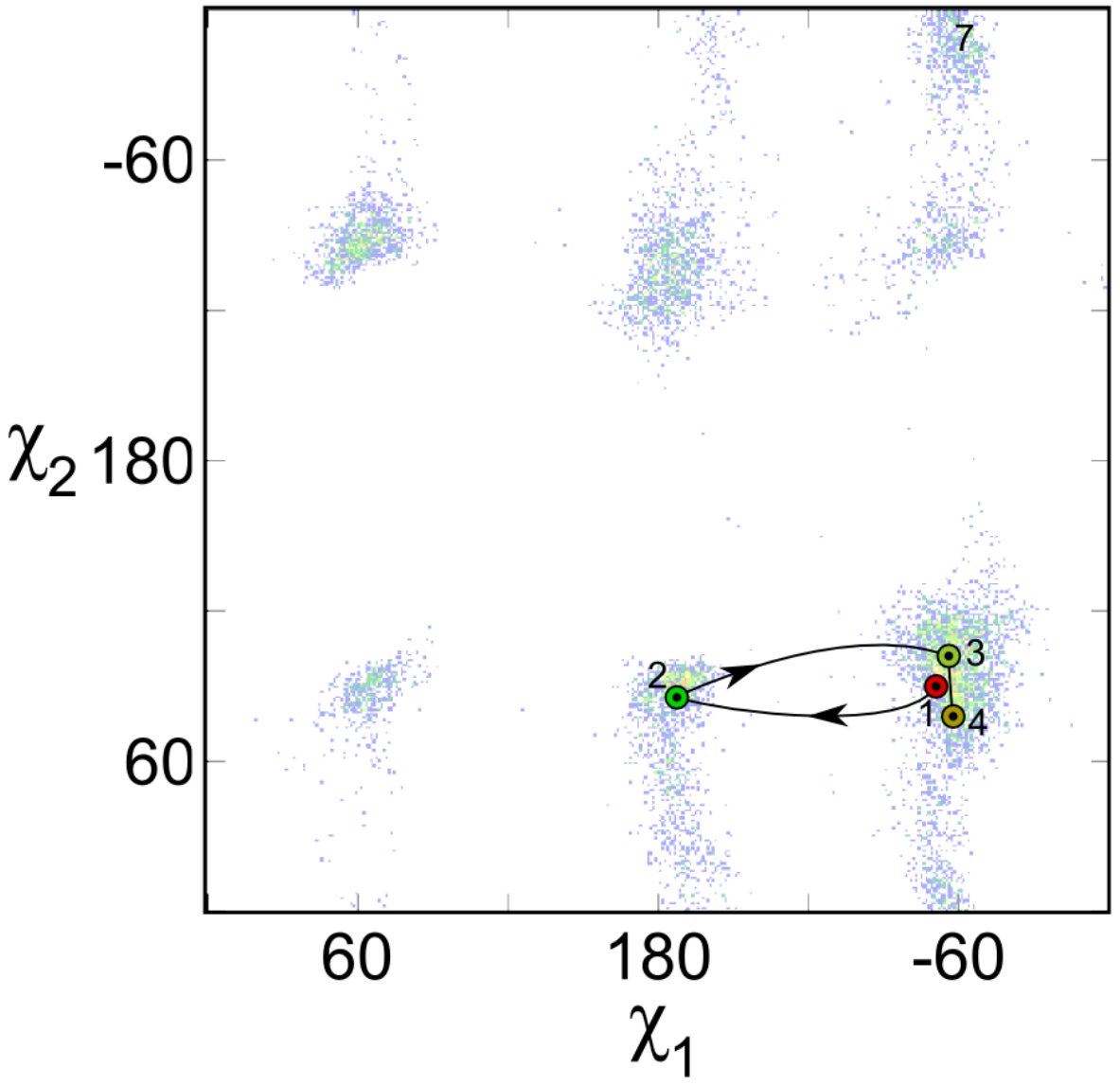

1
χ1 ~ -60°

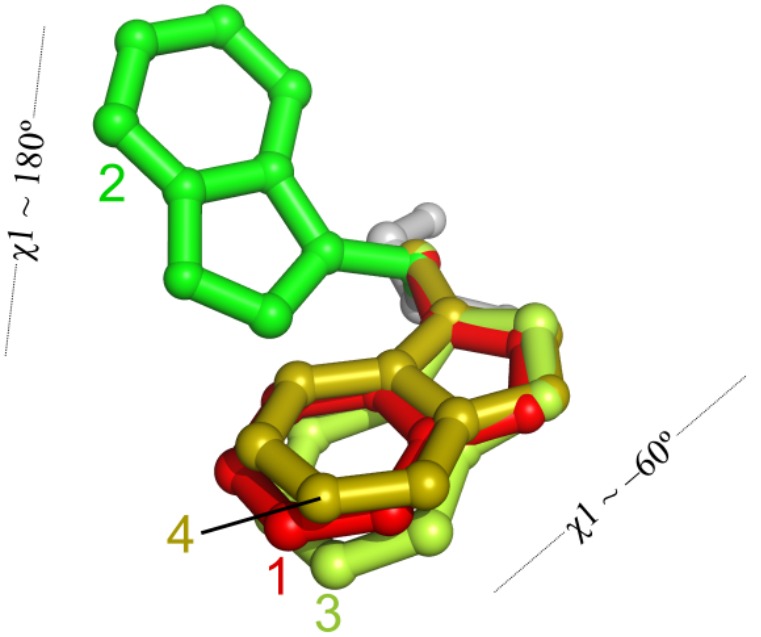

-60
χ_2 180
60
60
180
-60
χ_1
2
1

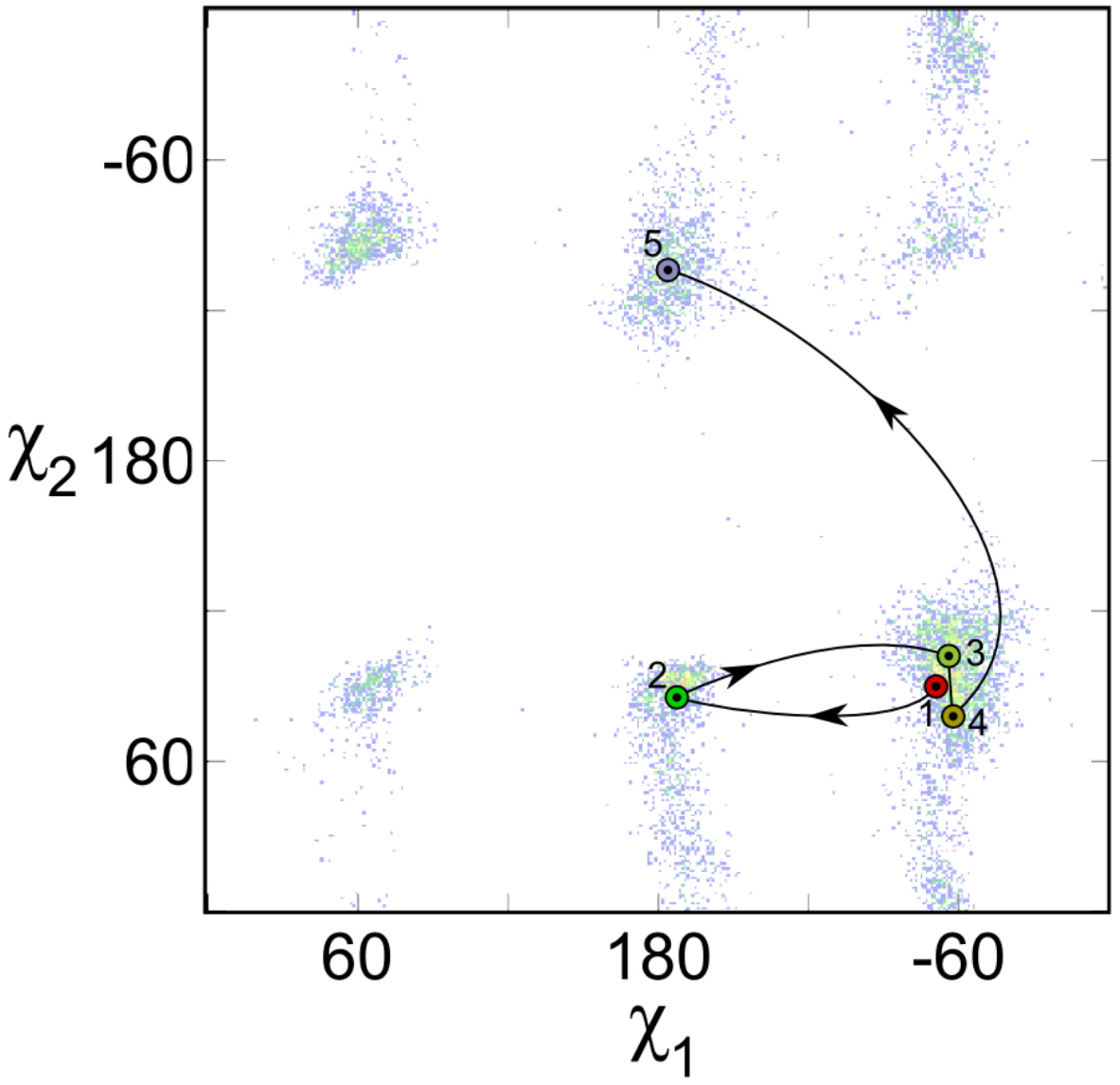

χ1 ~ 180°
2
1
χ1 ~ −60°

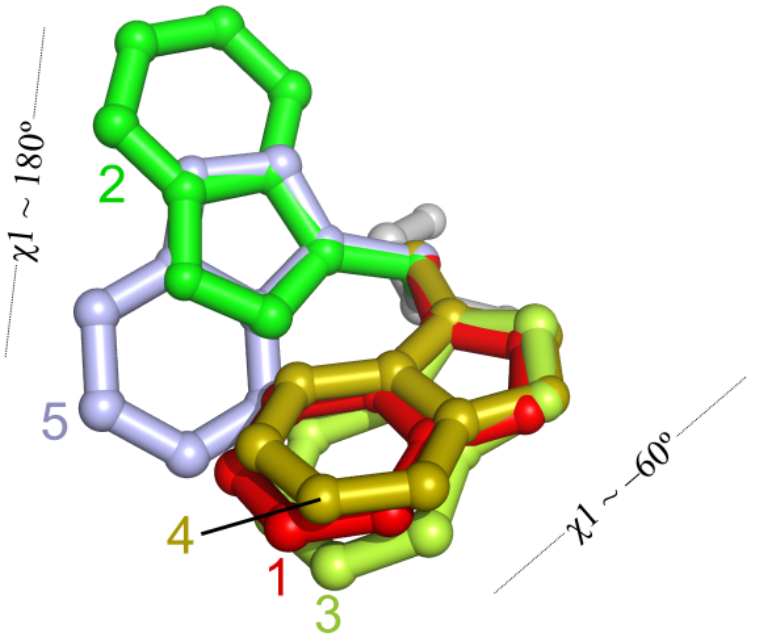

-60
χ_2 180
60
60
180
-60
χ_1
2
3
1

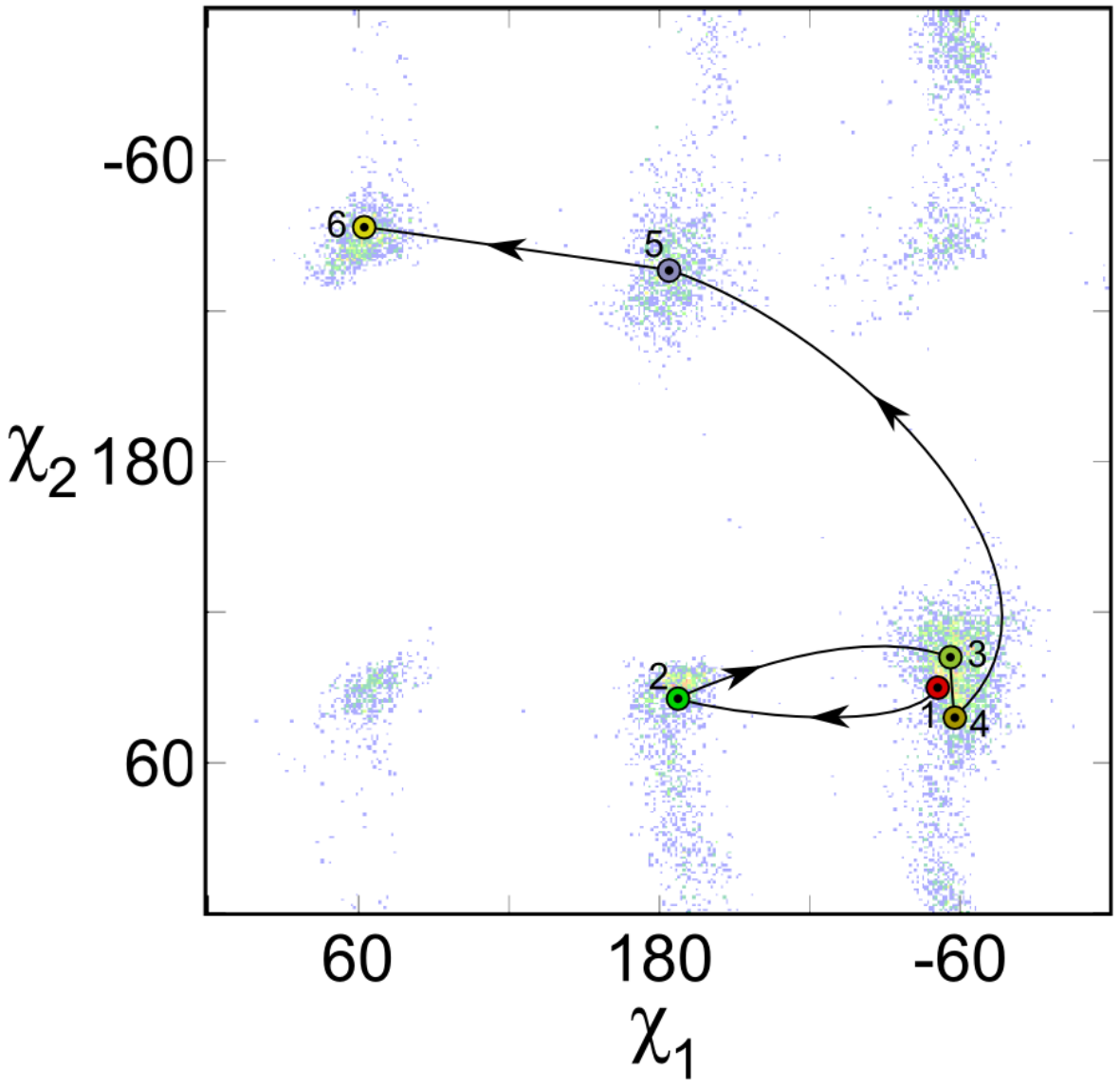

χ1 ~ 180°
2
1
3
χ1 ~ −60°

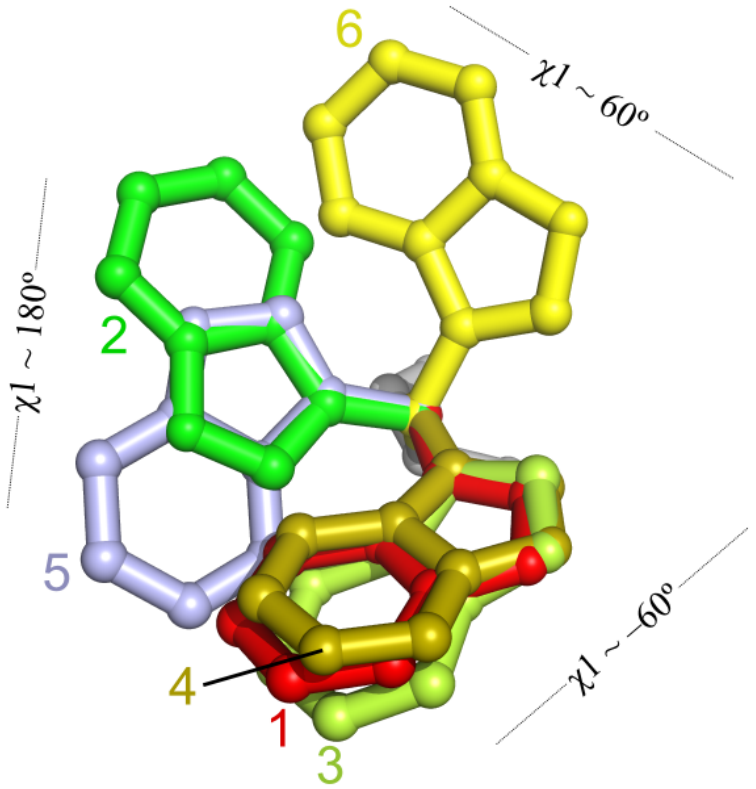

-60
χ_2 180
60
60
180
-60
χ_1
7
2
3
1
4

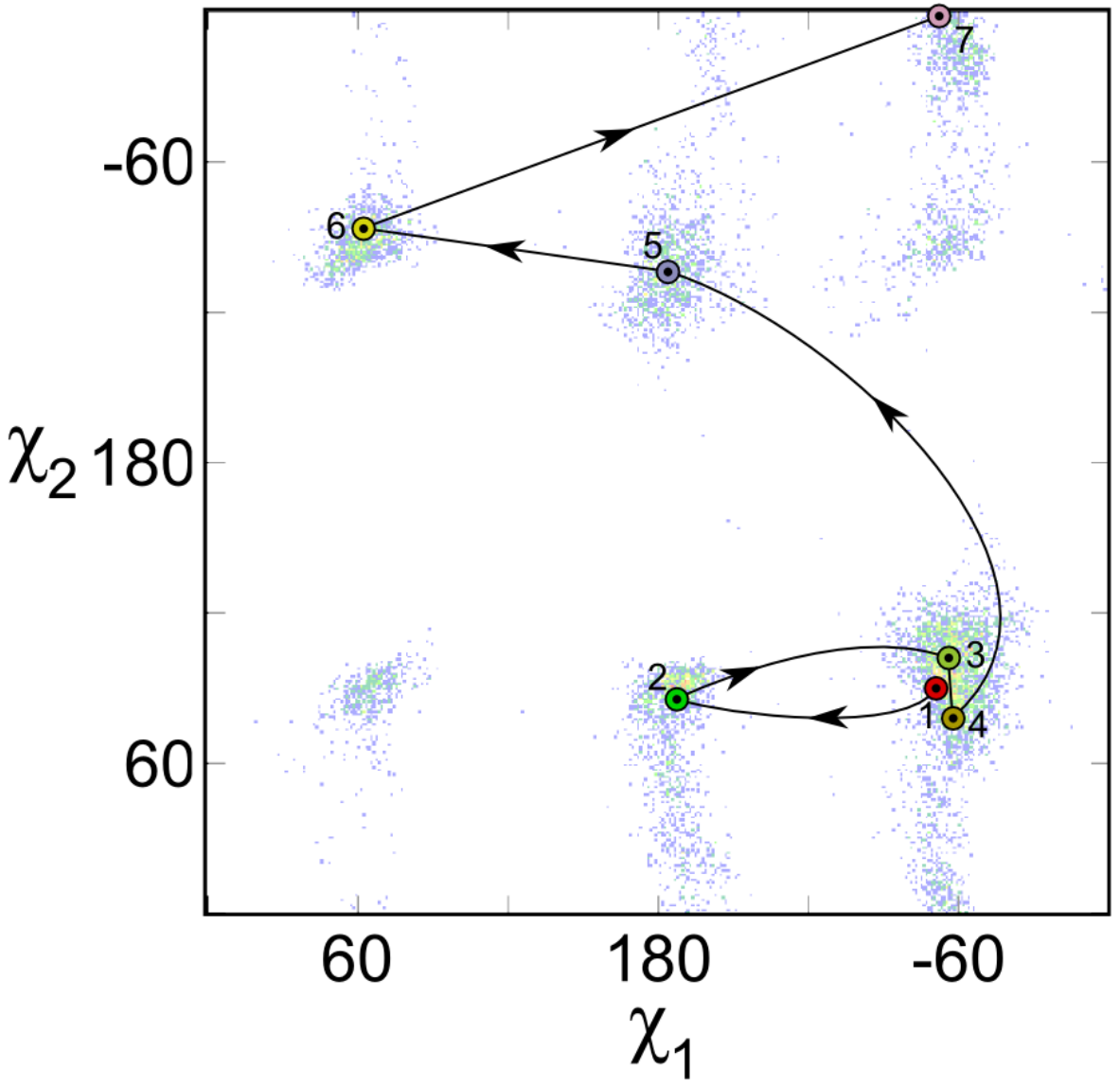

χ1 ~ 180°
2
4
1
3
χ1 ~ −60°

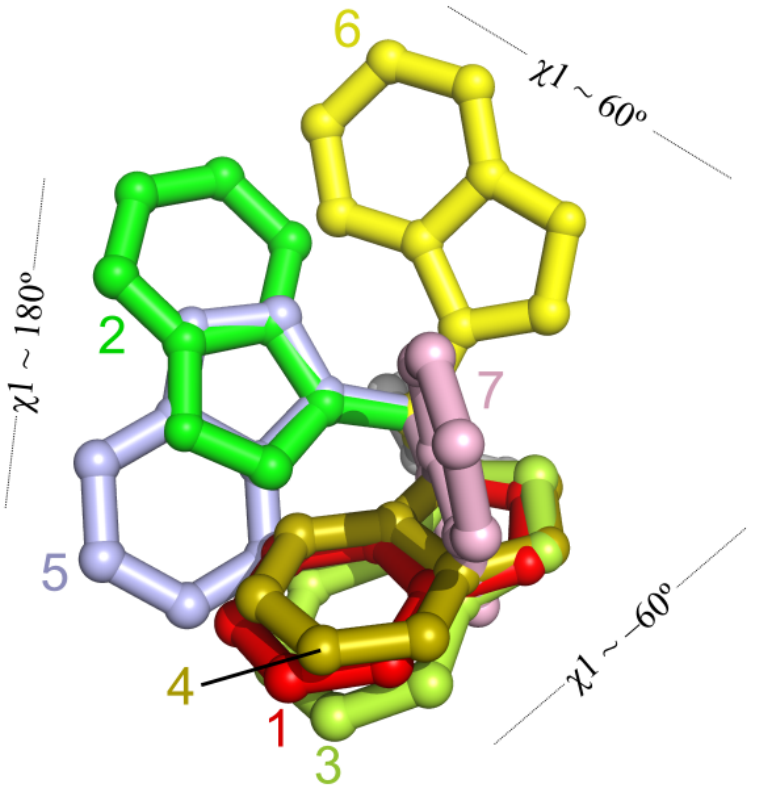

-60
χ_2 180
60
60
180
-60
χ_1
5
2
3
1
4

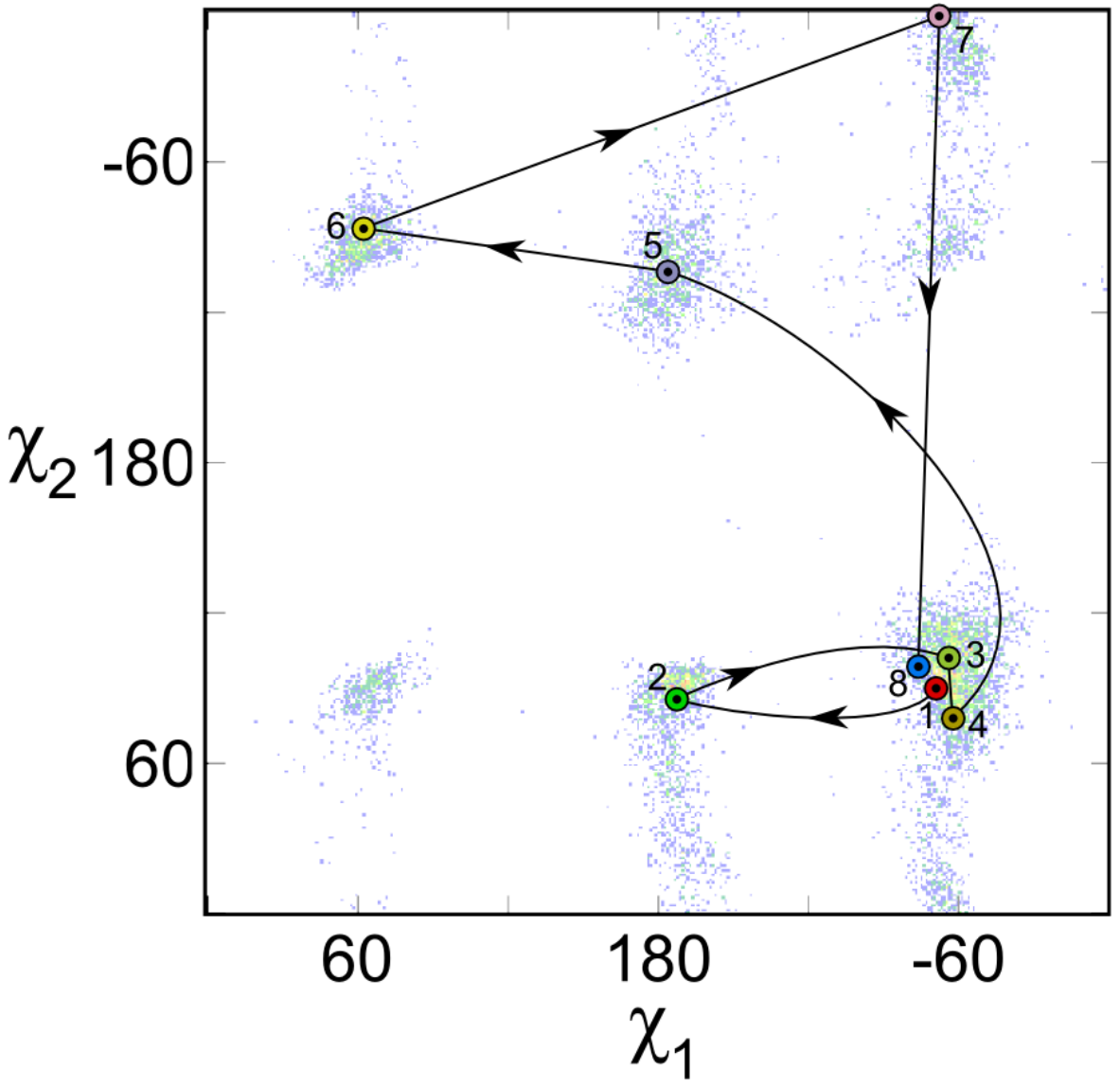

χ1 ~ 180°
2
5
4
1
3
χ1 ~ −60°

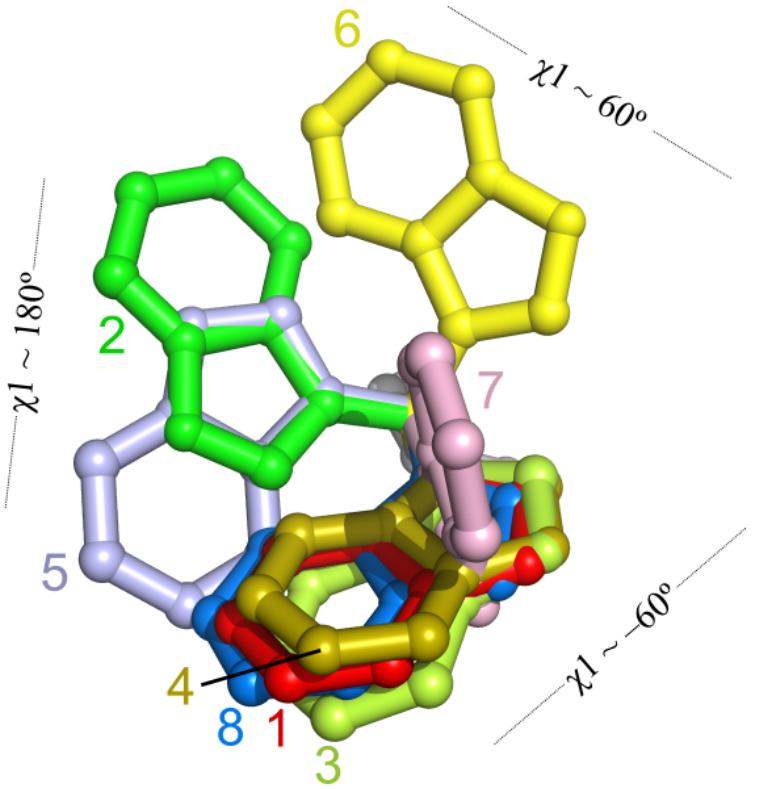

-60
χ_2 180
60
60
180
-60
χ_1
6
5
2
3
1
4

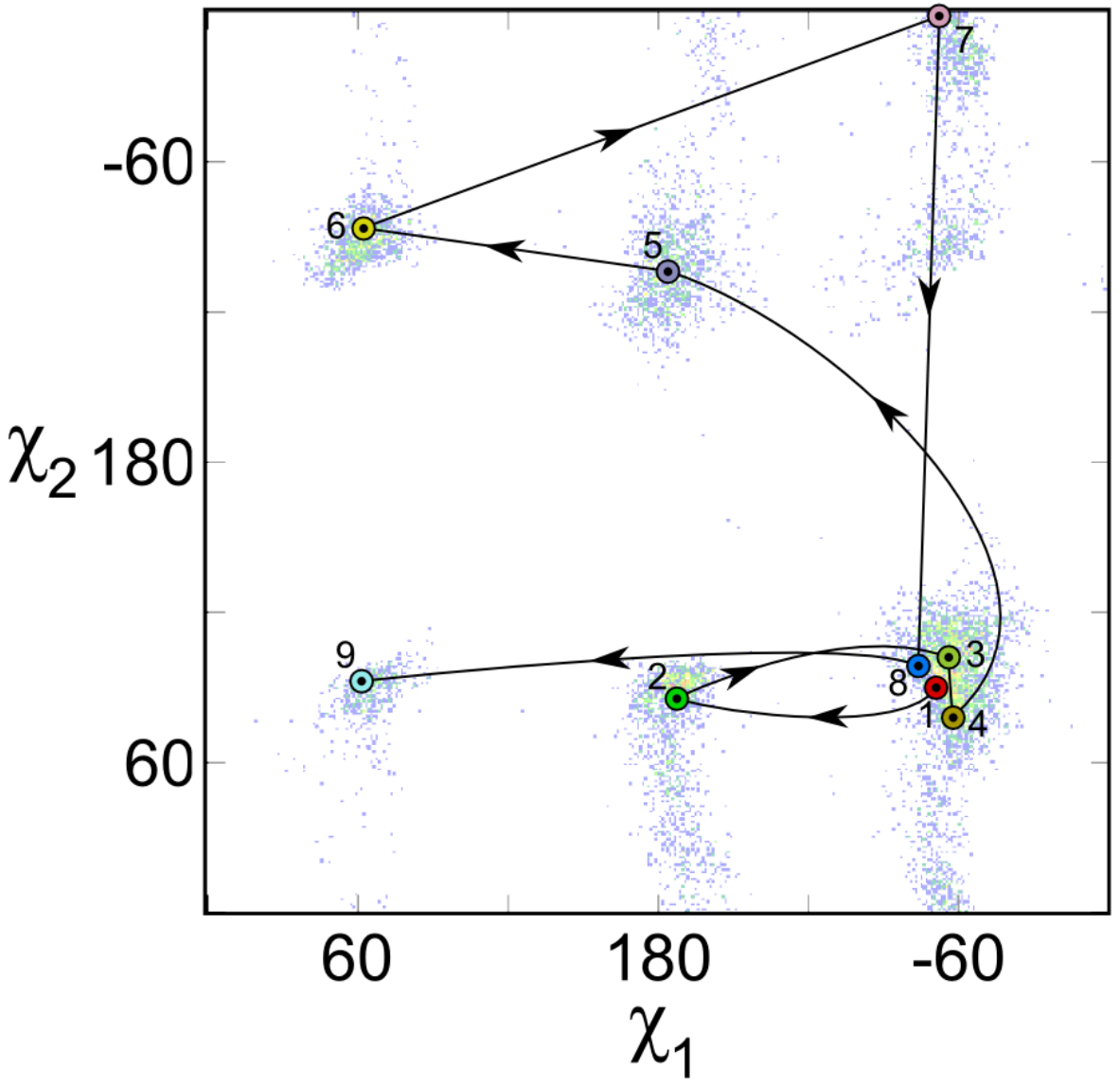

6
χ1 ~ 60°
χ1 ~ 180°
2
5
4
1
3
χ1 ~ -60°

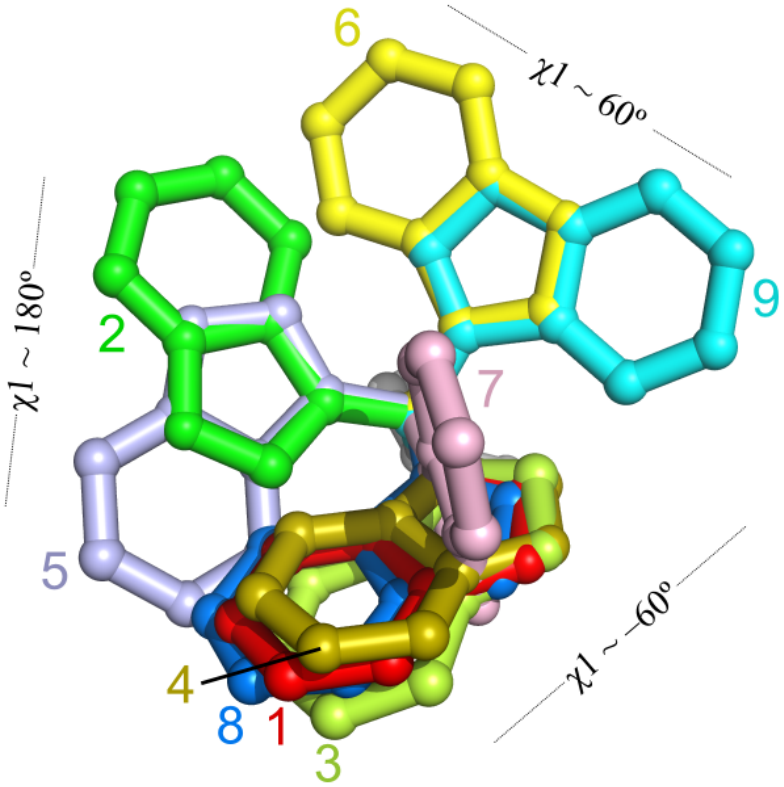

-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
2
3
1
4

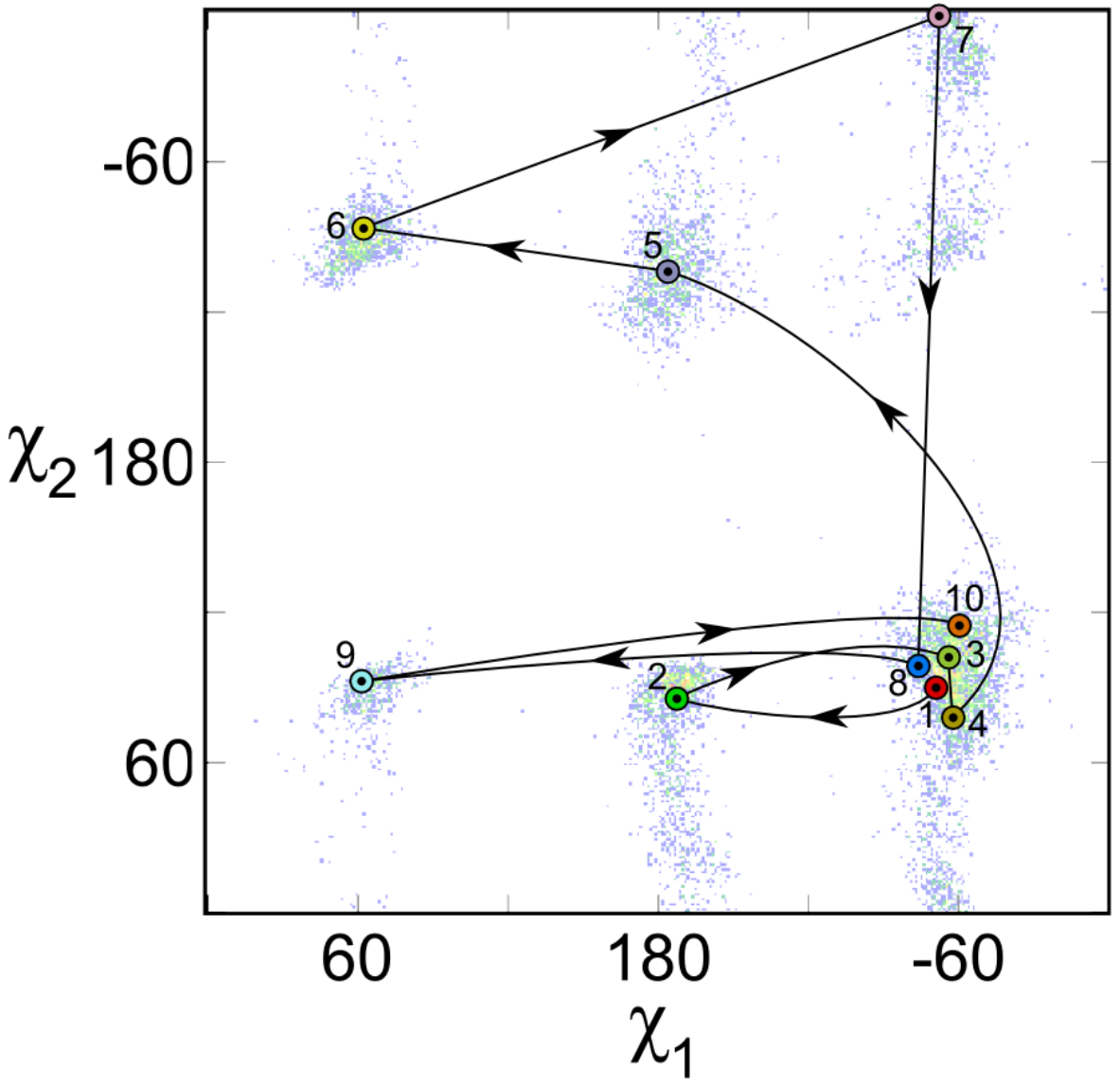

6
χ1 ~ 60°
χ1 ~ 180°
2
7
5
4
1
3
χ1 ~ -60°

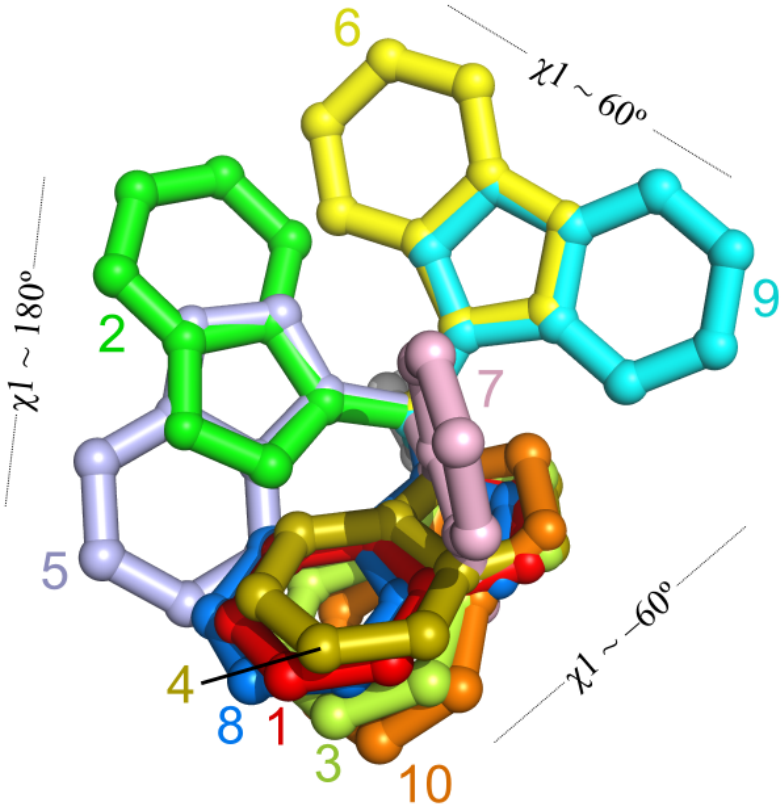

-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
2
8
1
3
4

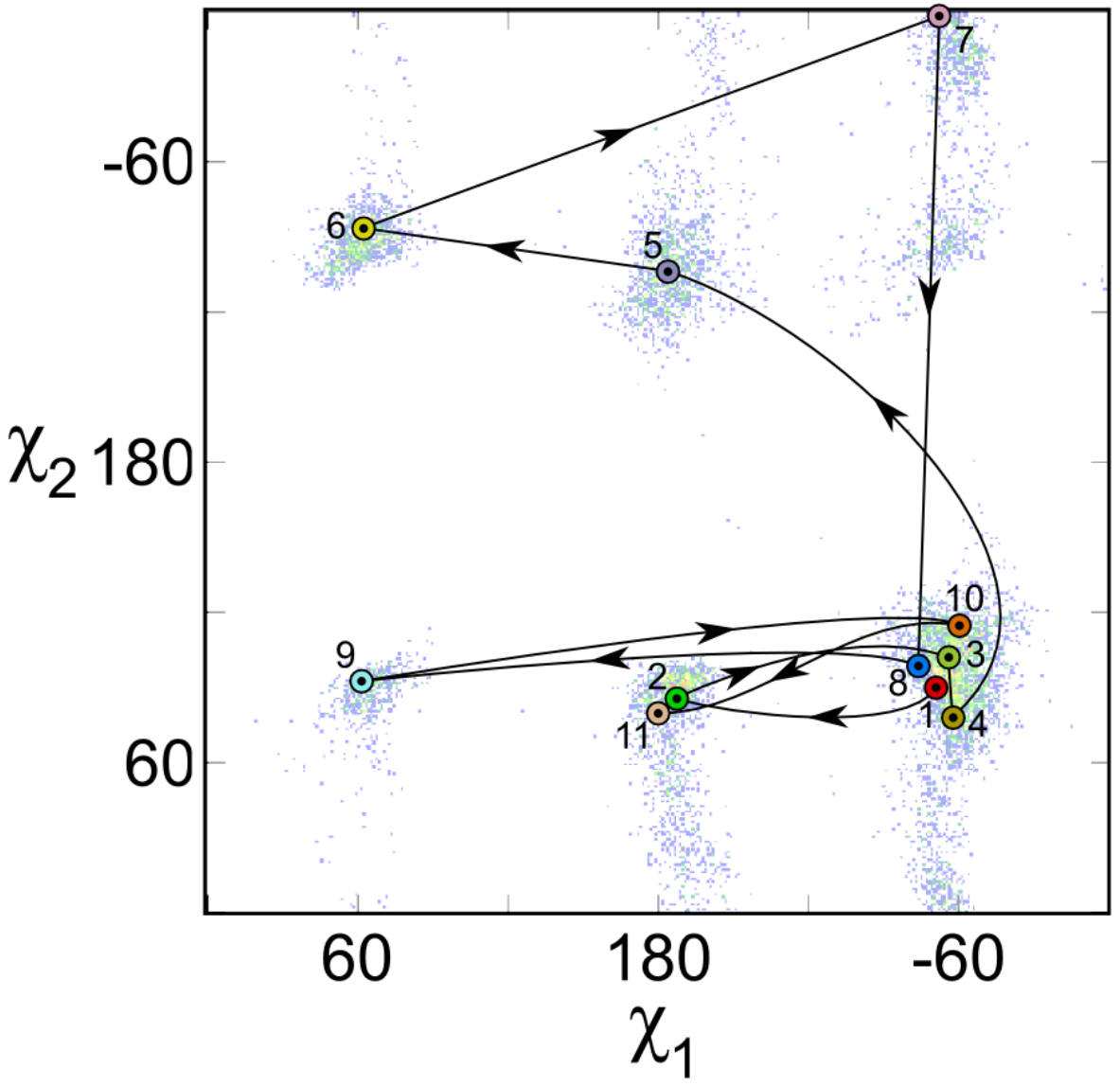

6
χ1 ~ 60°
χ1 ~ 180°
2
7
5
4
8
1
3
χ1 ~ -60°

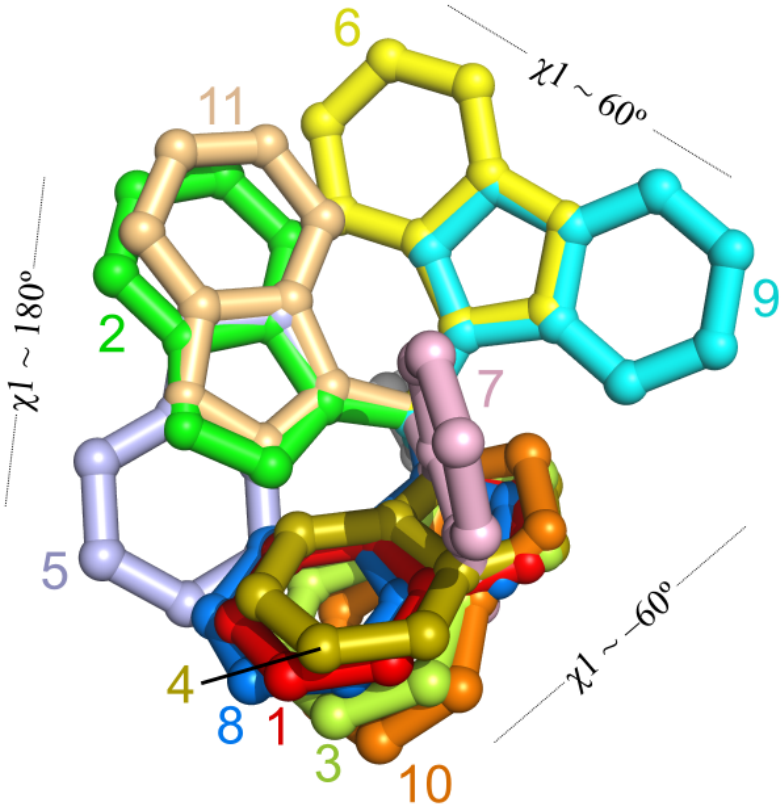

-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
9
2
8
1
3
4

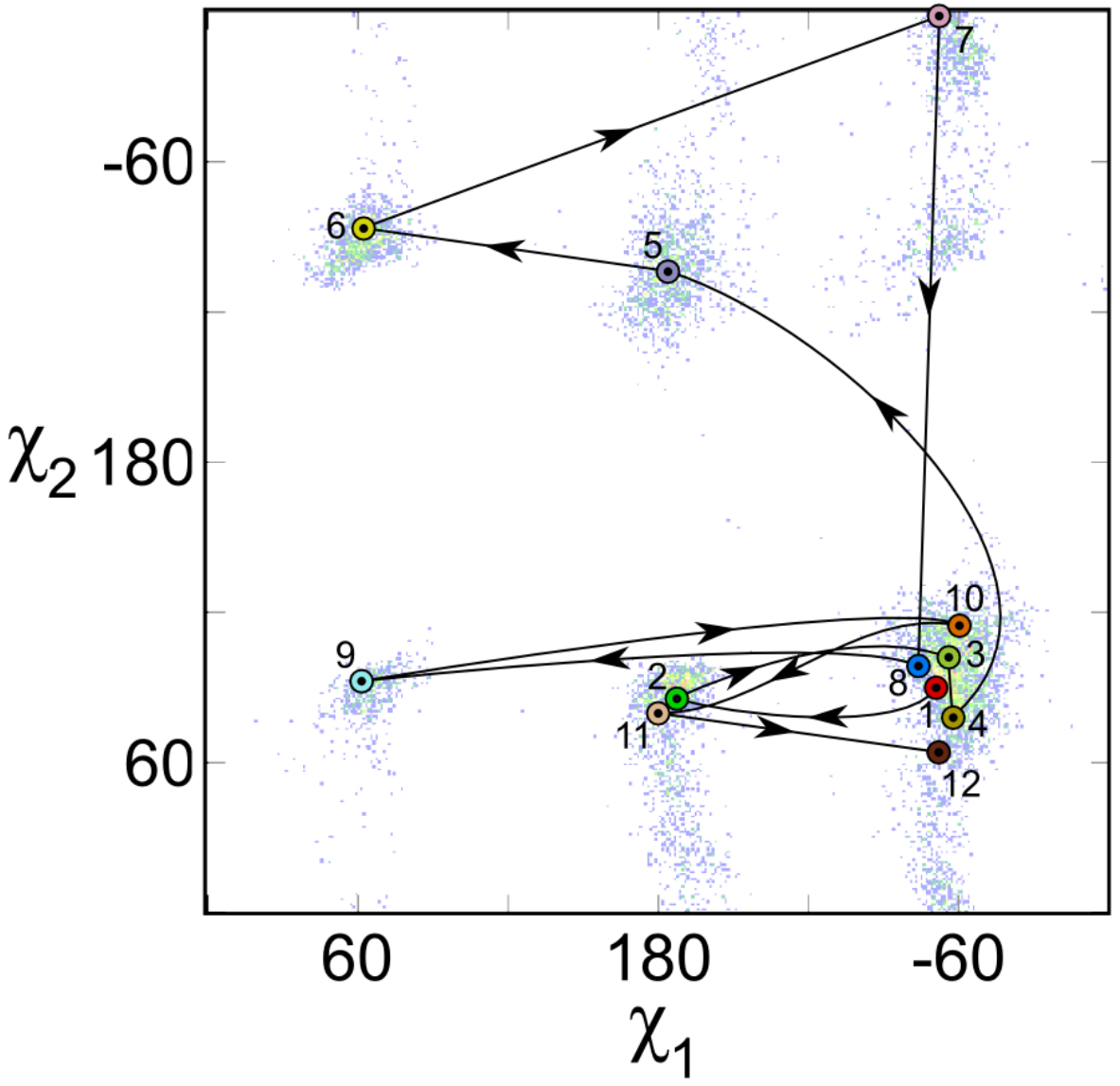

6
χ1 ~ 60°
9
χ1 ~ 180°
2
7
5
χ1 ~ -60°
4
8
1
3

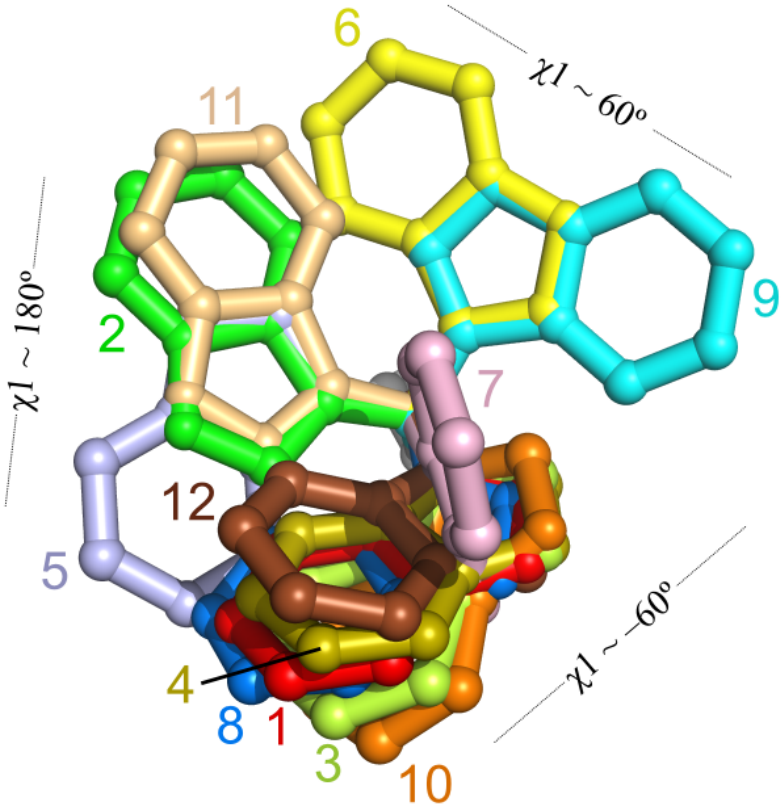

-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
9
2
10
3
8
1
4

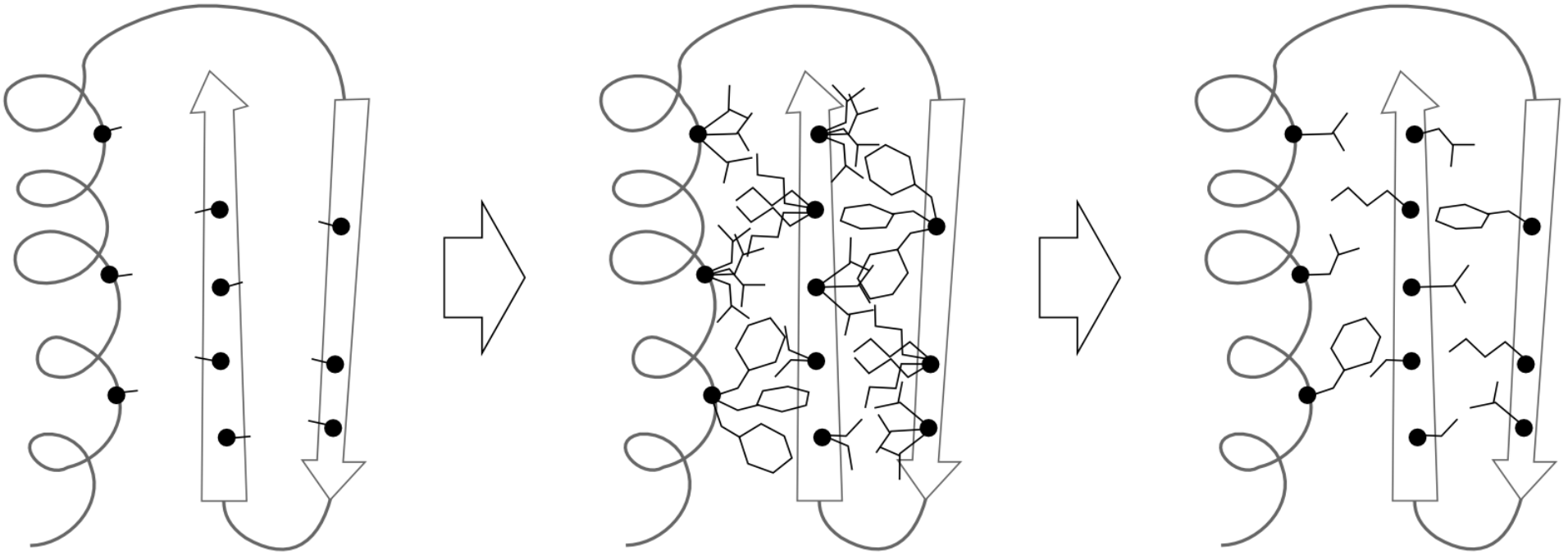

6
χ1 ~ 60°
9
χ1 ~ 180°
2
7
5
4
8
1
3
10
χ1 ~ -60°


-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
9
2
11
10
3
8
1
4

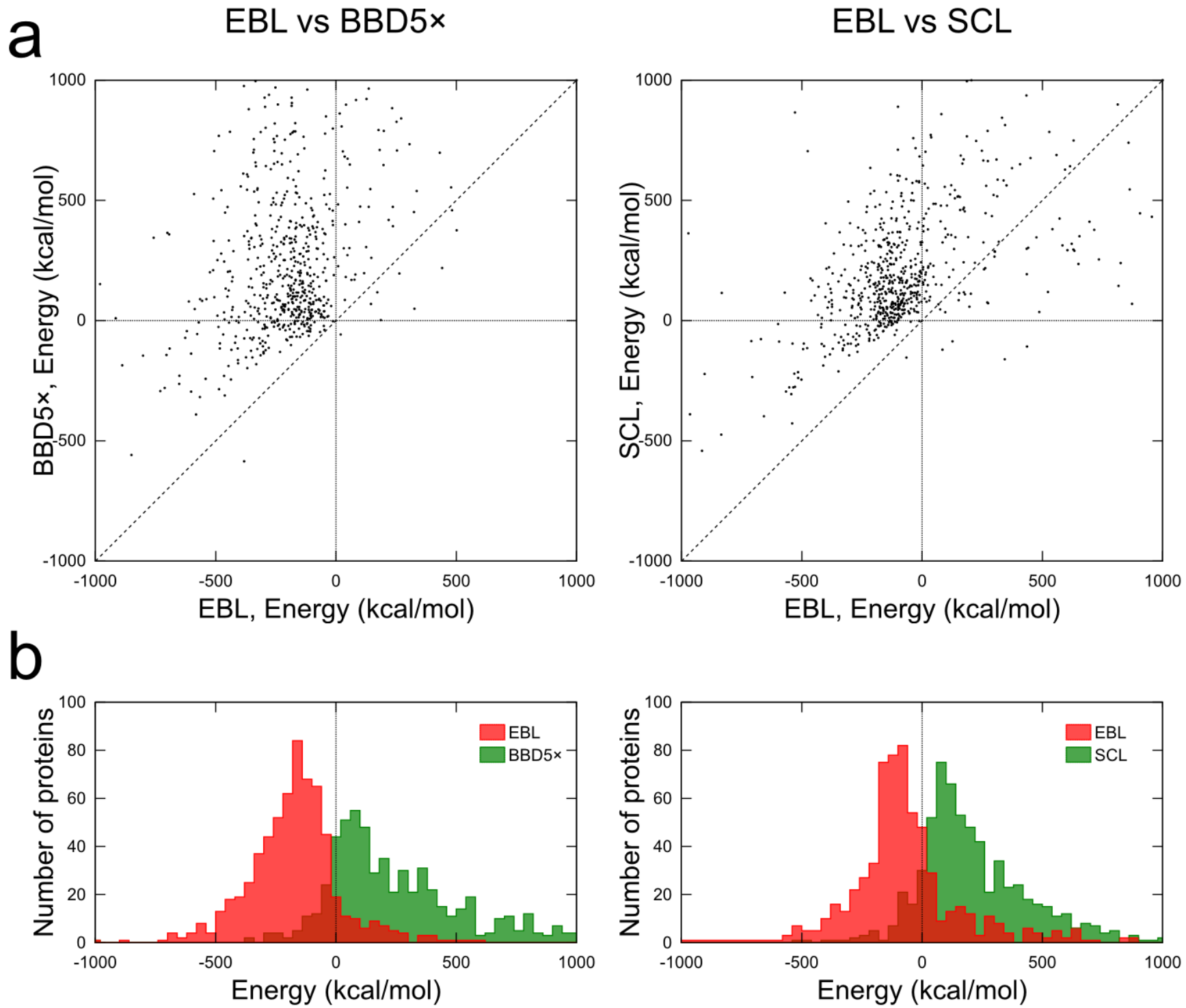

6
11
χ1 ~ 60°
χ1 ~ 180°
2
9
7
5
4
8
1
3
10
χ1 ~ -60°


-60
χ_2 180
60
60
180
-60
χ_1
6
5
7
9
2
11
10
3
8
1
4
12

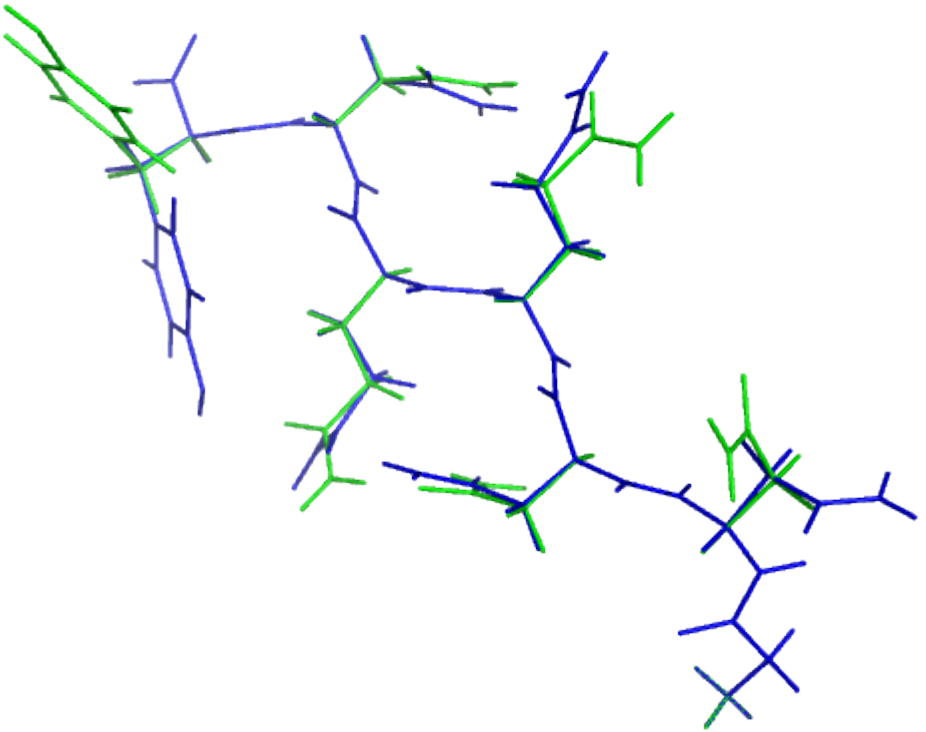

6
11
χ1 ~ 60°
χ1 ~ 180°
9
2
7
12
5
4
8
1
3
10
χ1 ~ -60°



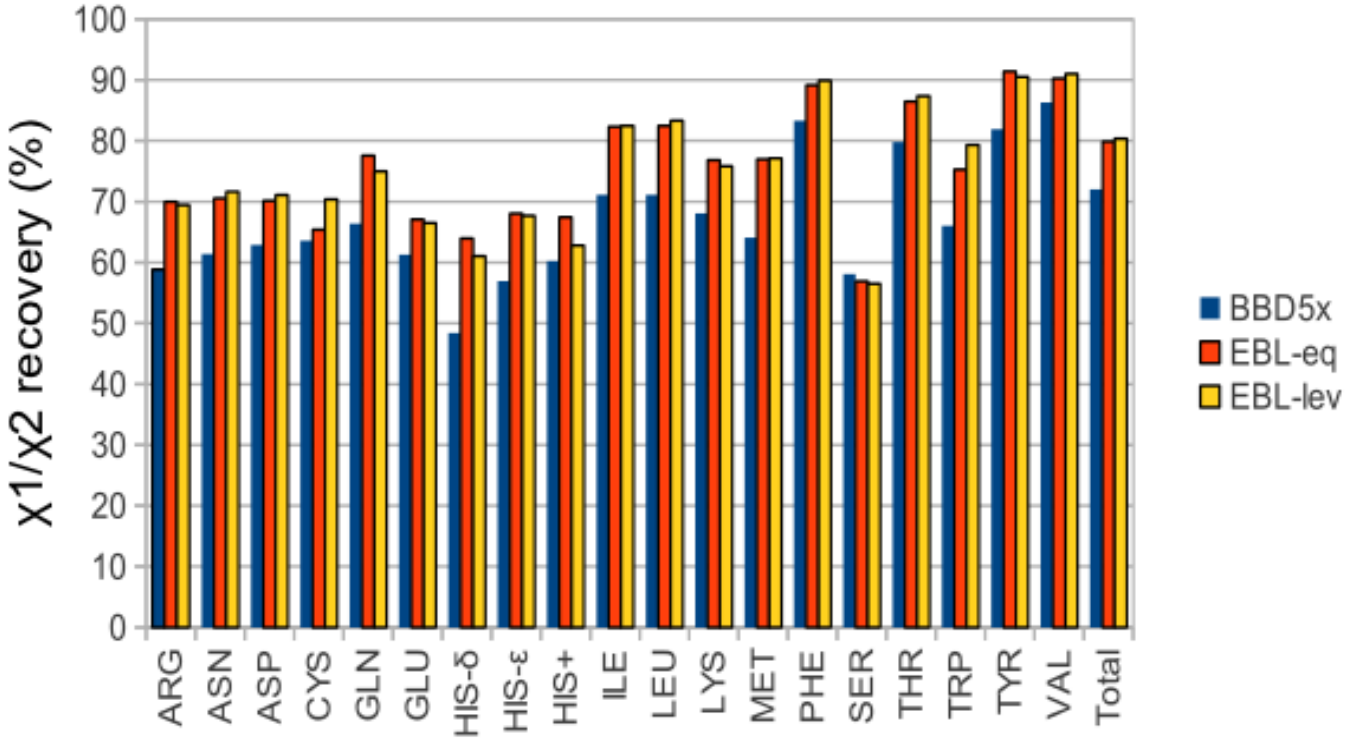
560 complete protein repacks.
The lower the energy, the better.


a
EBL vs BBD5×
EBL vs SCL
BBD5×, Energy (kcal/mol)
SCL, Energy (kcal/mol)
EBL, Energy (kcal/mol)
EBL, Energy (kcal/mol)
1000
500
0
-500
-1000
b
Number of proteins
Energy (kcal/mol)
EBL
BBD5×
SCL
100
80
60
40
20
0

DIHEDRAL RECOVERY



DIHEDRAL RECOVERY RESULTS

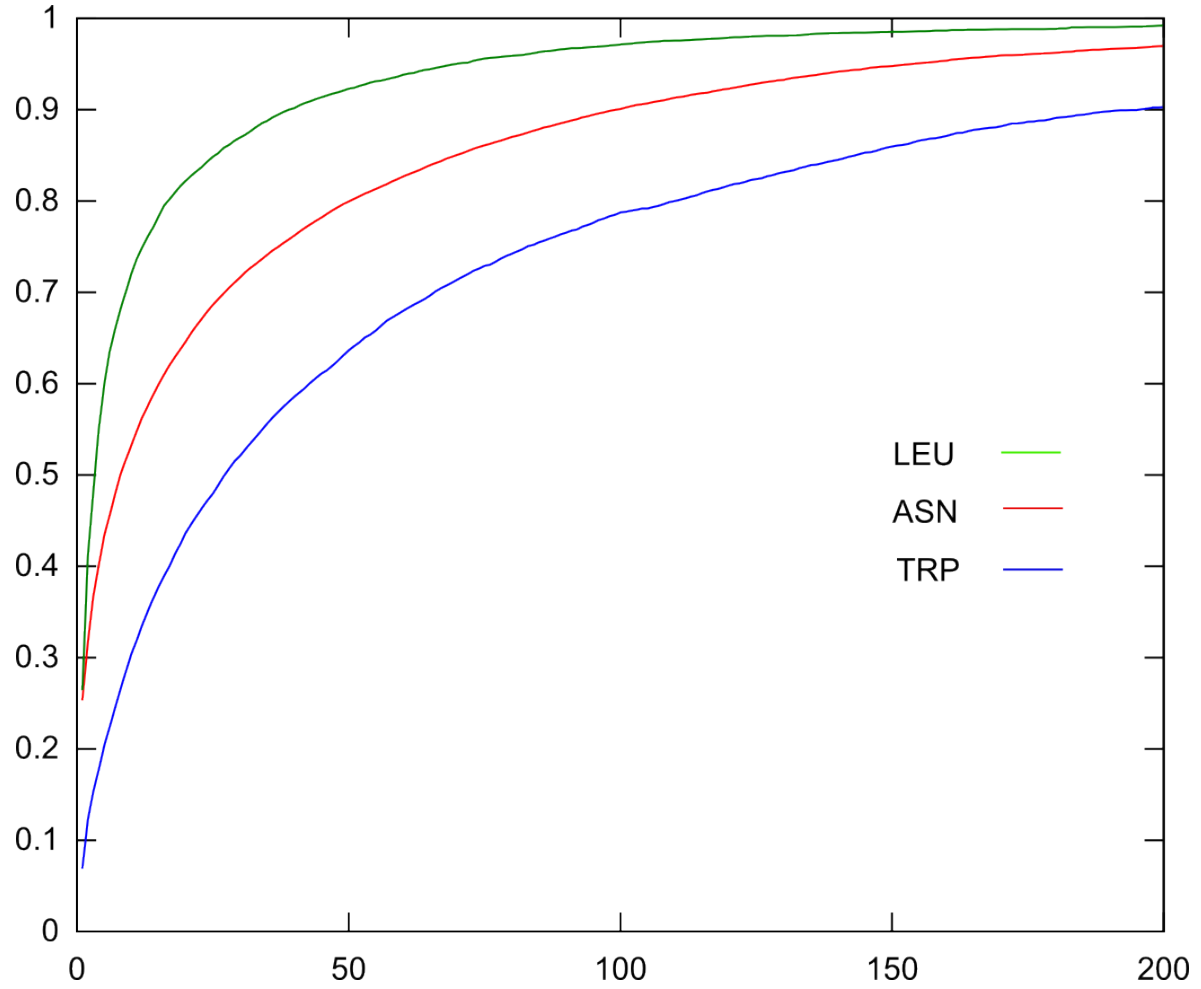


ADVANTAGES OF ENERGY BASED LIBRARY

- Helps achieve lower energy structures
- Better dihedral recovery
- Unprecedented flexibility – the first 'n' conformers is probably the best set of 'n' conformers

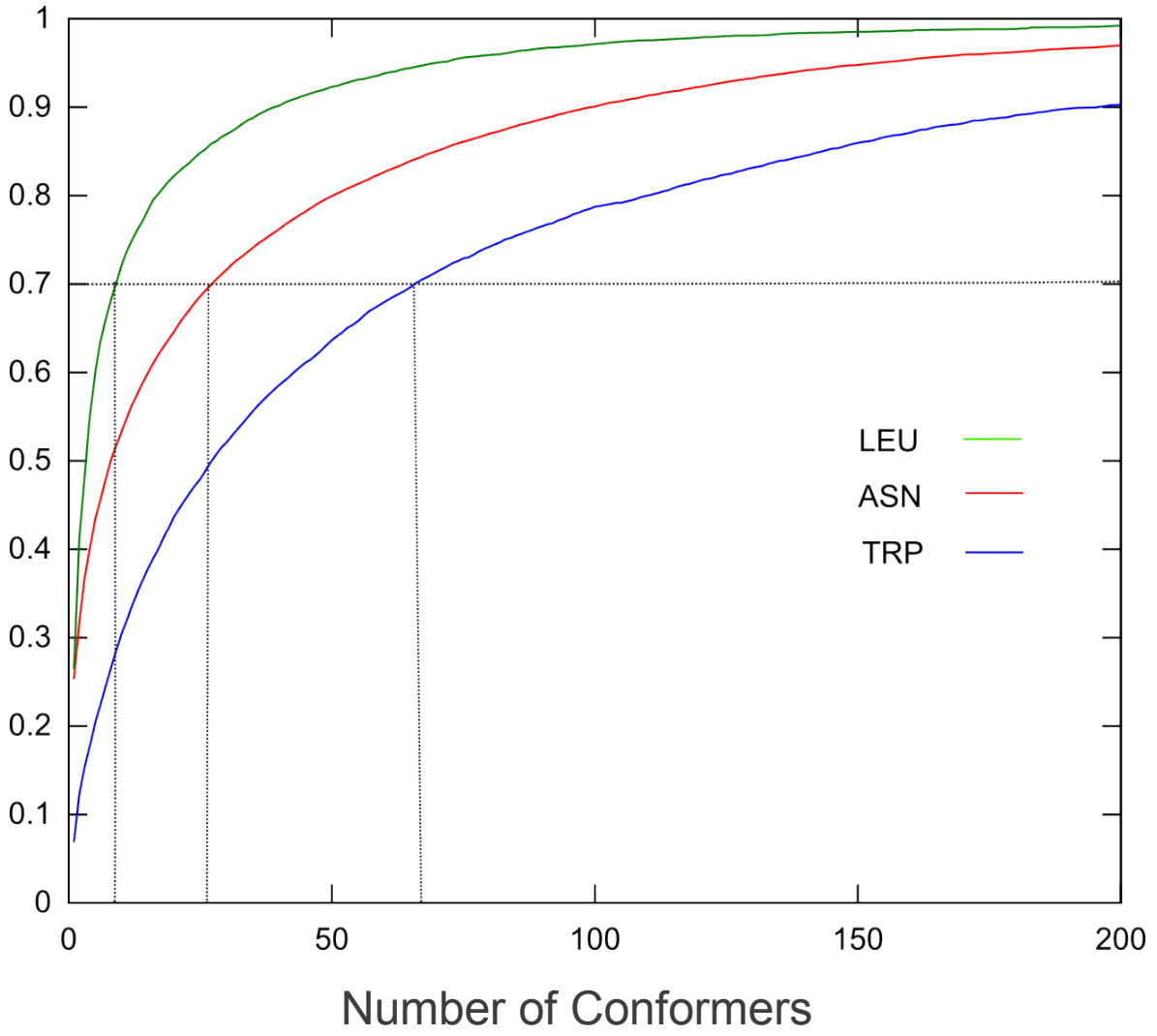
How to balance sampling across residue types?

NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE



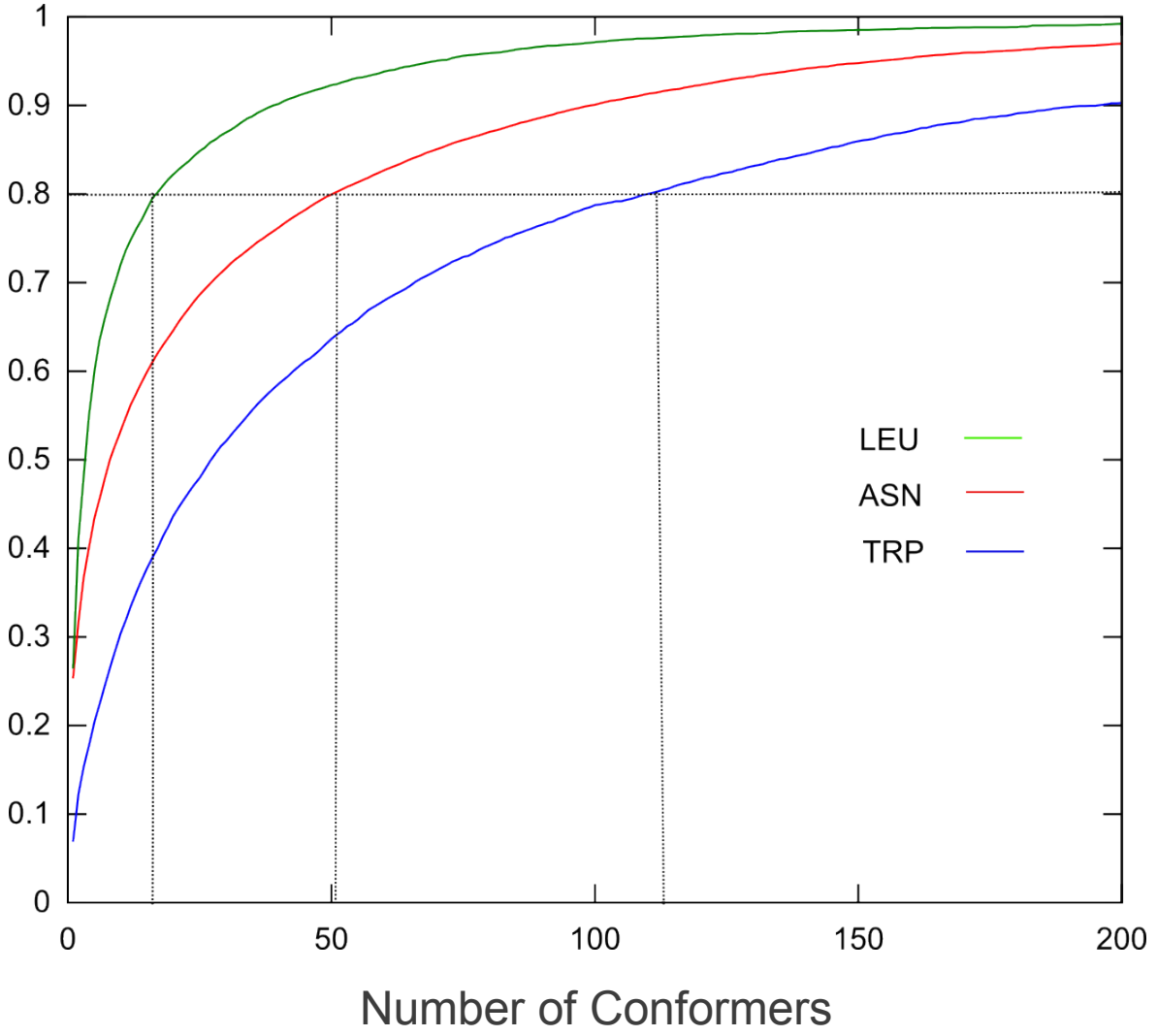
NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

Single Repack Coverage



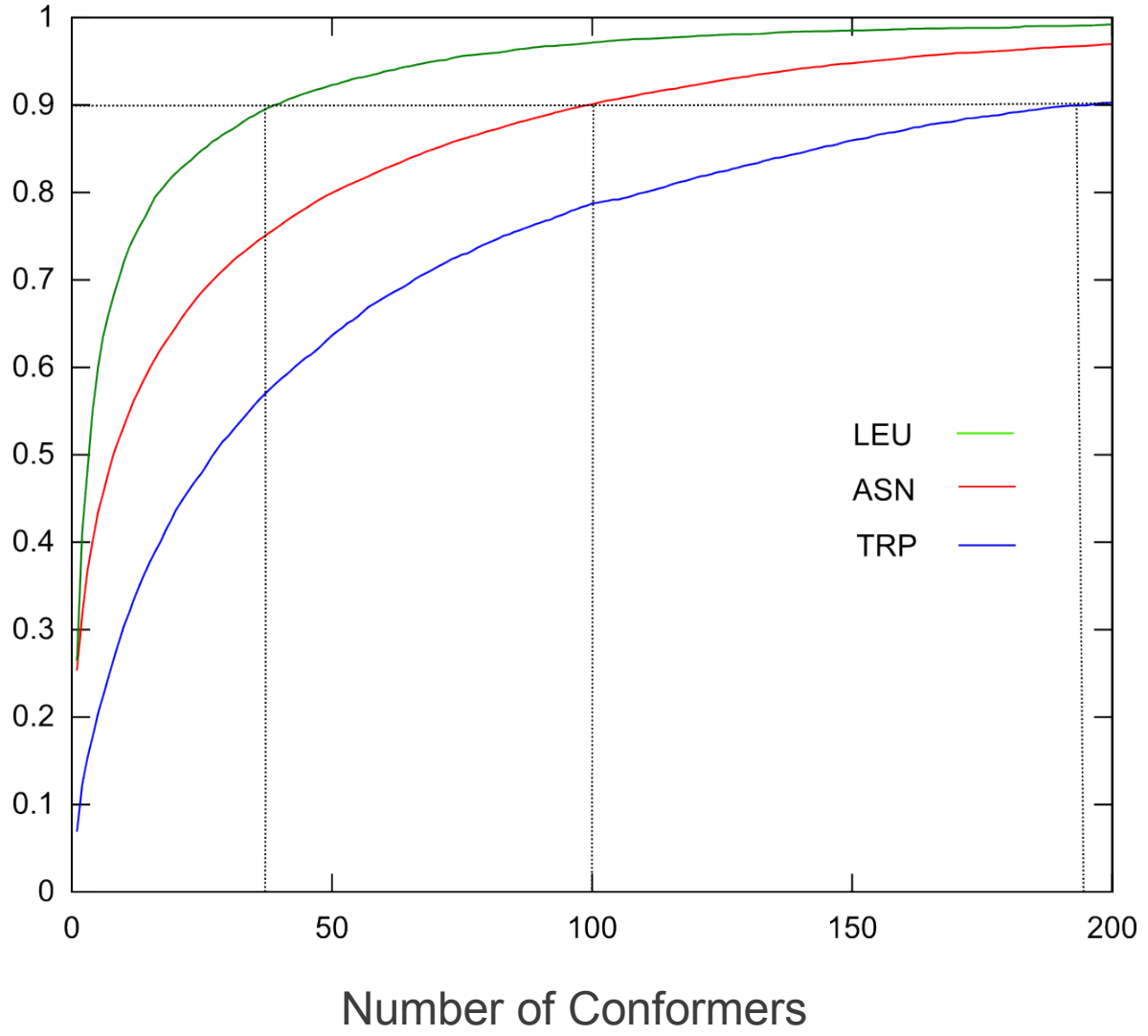
NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

Single Repack Coverage



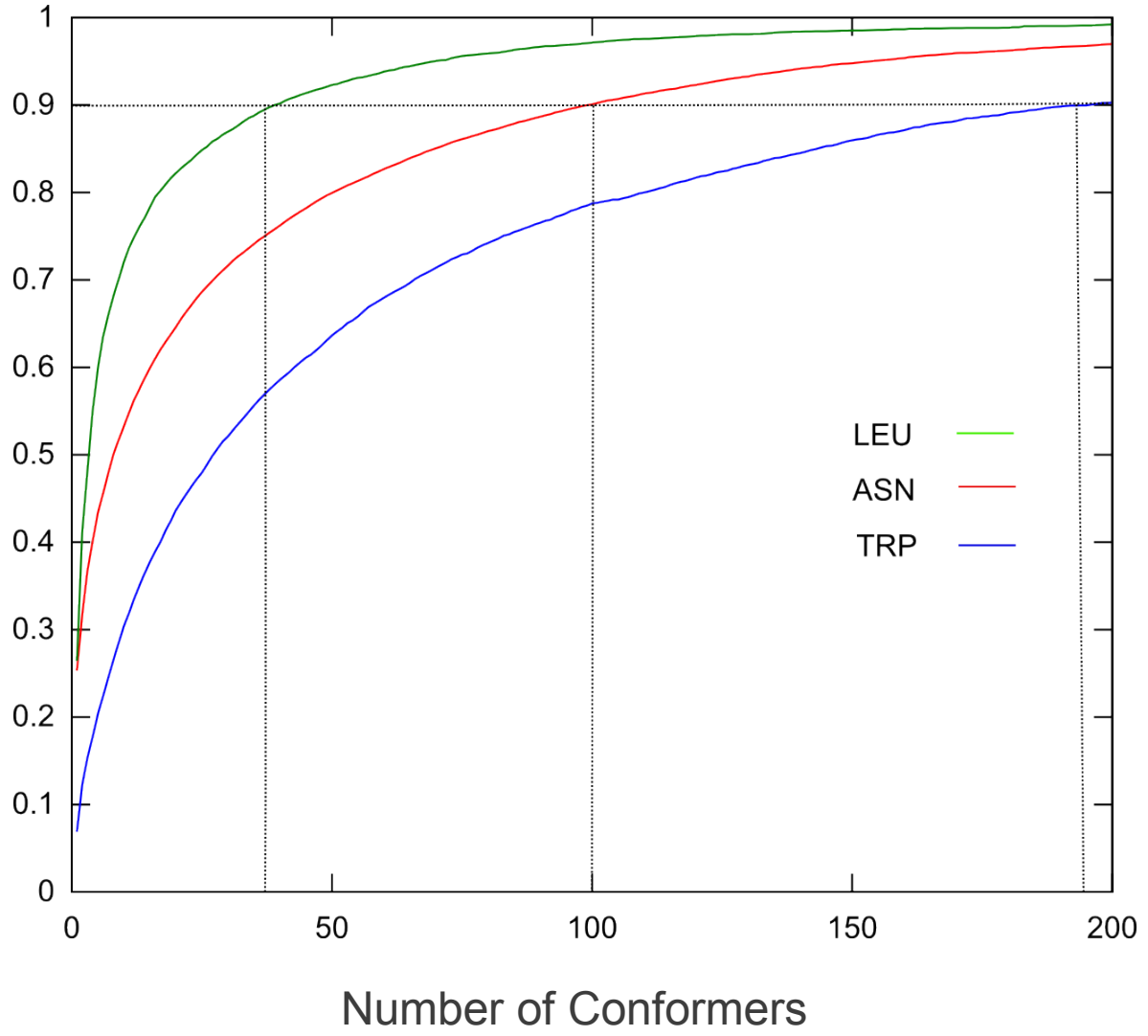
NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

Single Repack Coverage



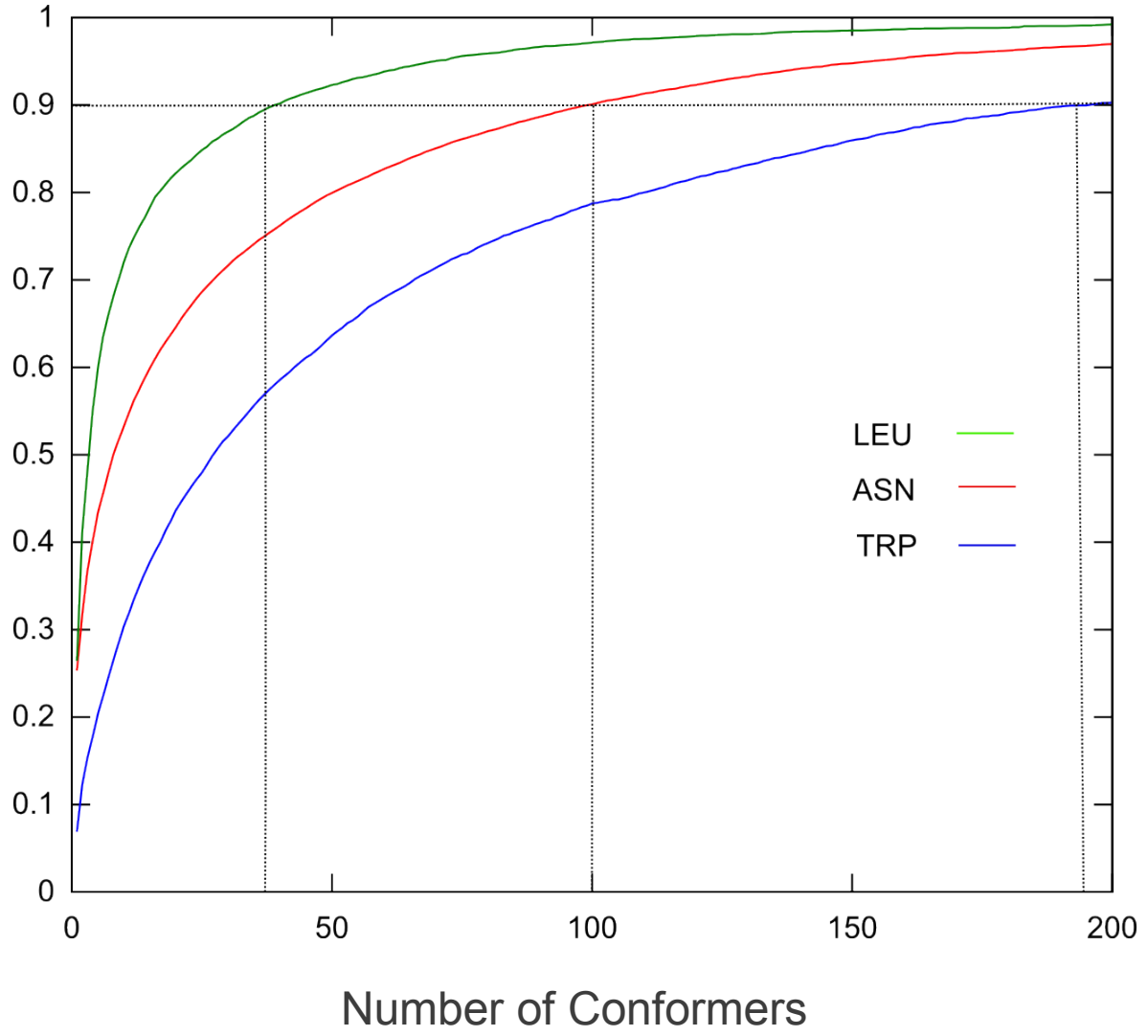
NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

Single Repack Coverage

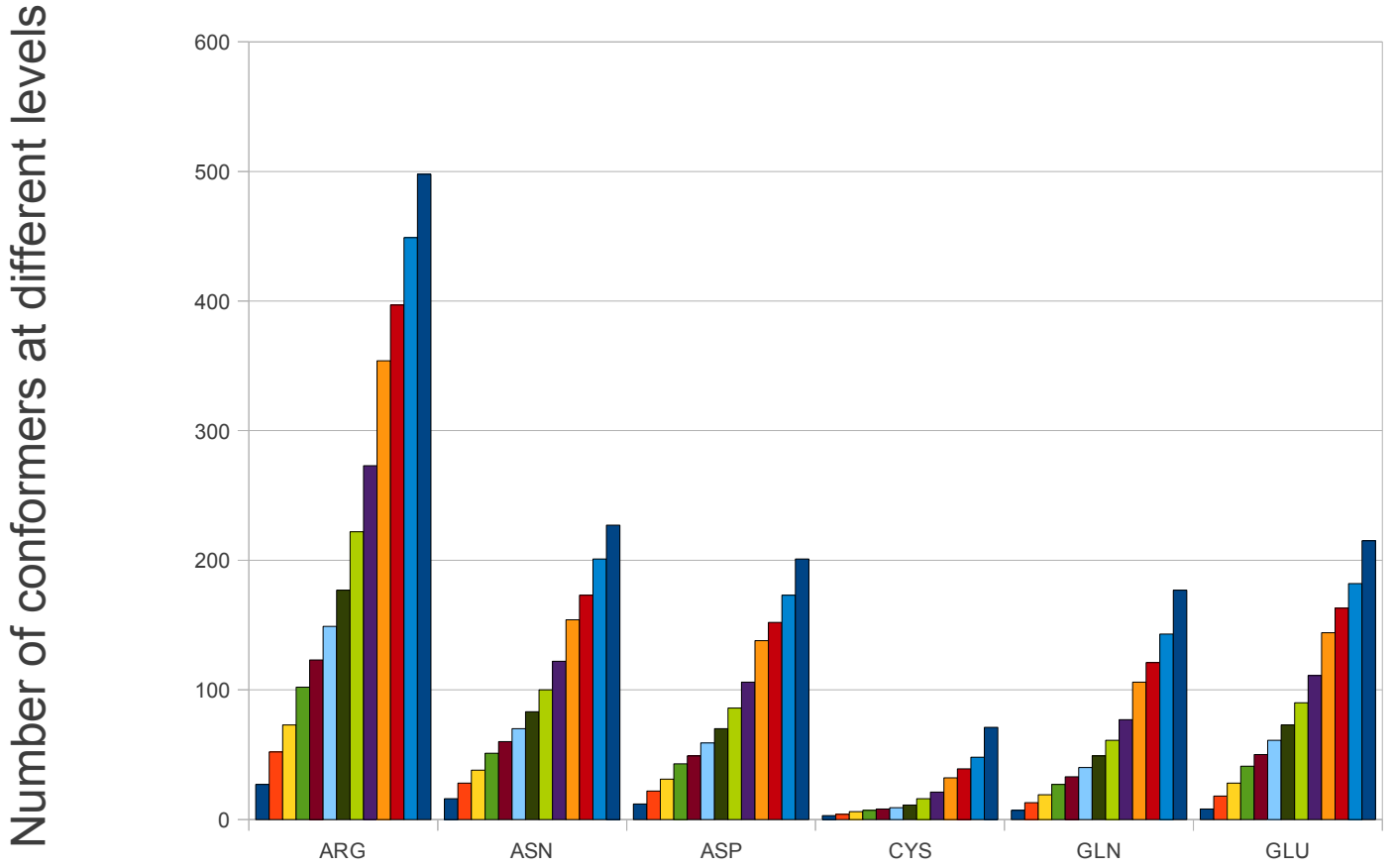


NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

Single Repack Coverage

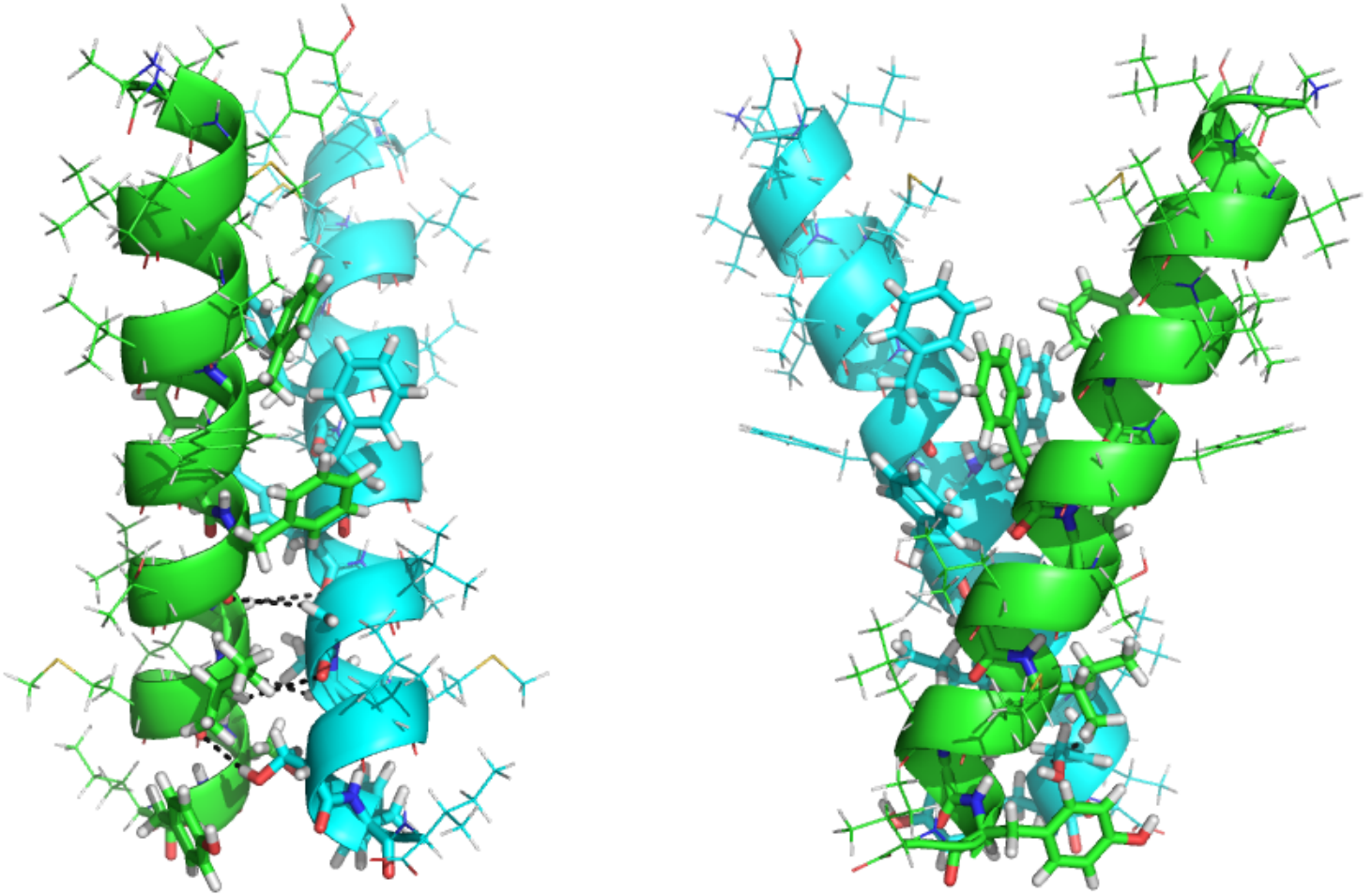


NUMBER OF CONFORMERS FOR EACH AMINO ACID TYPE

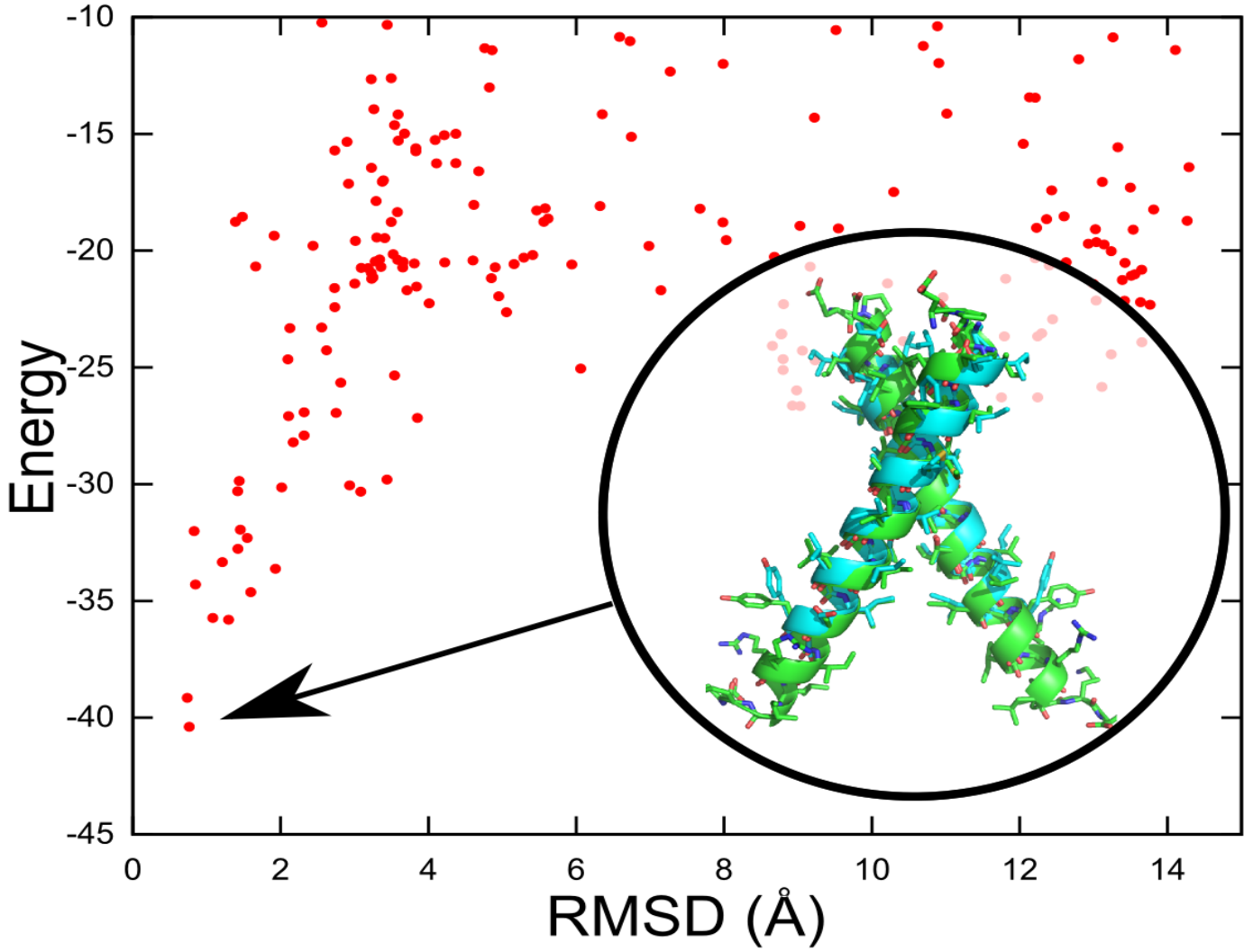


SEARCH DIMER SPACE

USE DIFFERENT LEVELS TO IMPROVE SPEED AND ACCURACY



RESULTS OF GLYCOPHORIN A



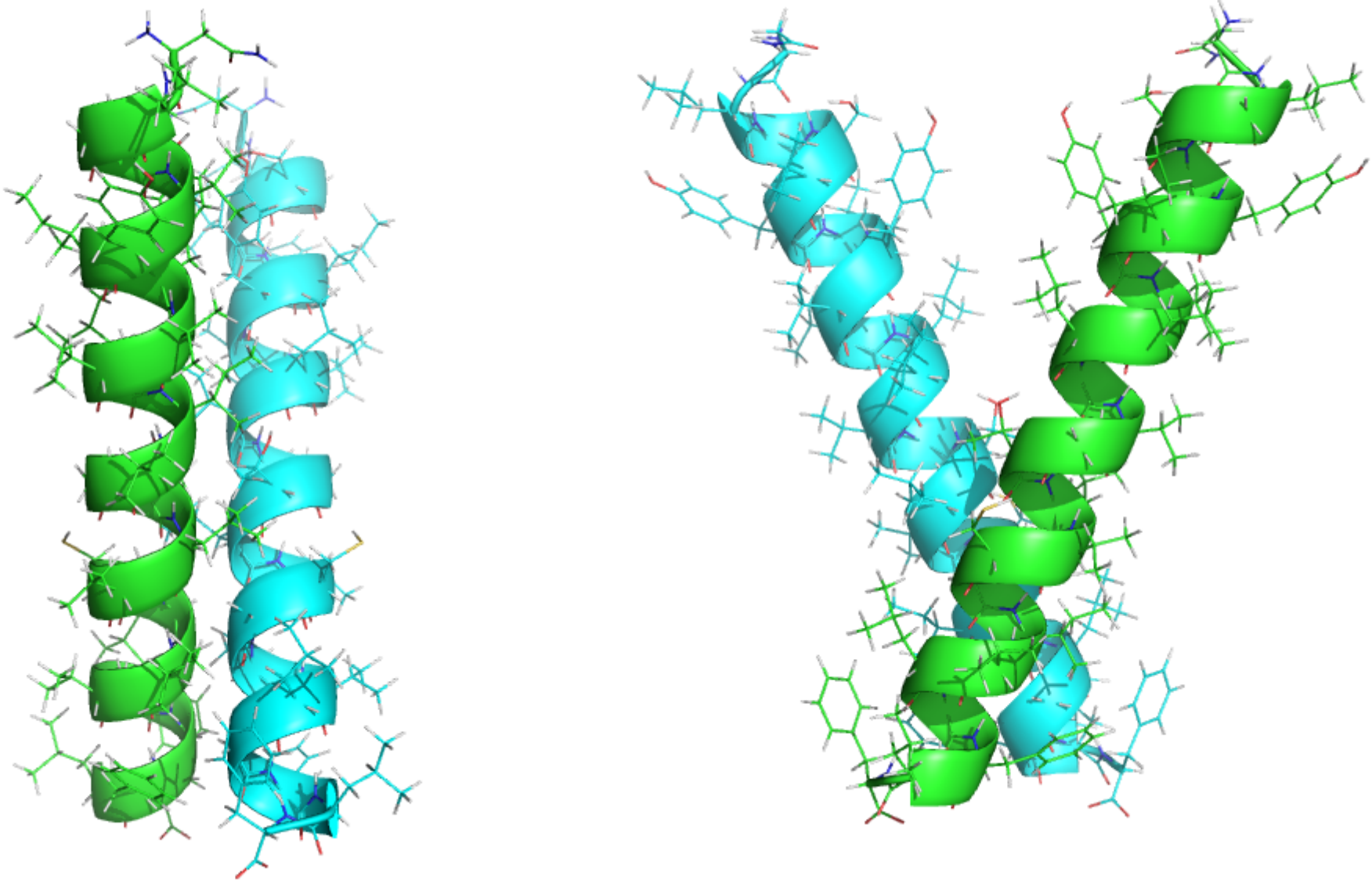
PERFORMANCE

- Predict a structure in 10 mins
- On a cluster with 128 cores
- Sidechain optimization ~ 3-8 seconds
- Glycophorin A
  - < 0.7 A RMSD from NMR structure

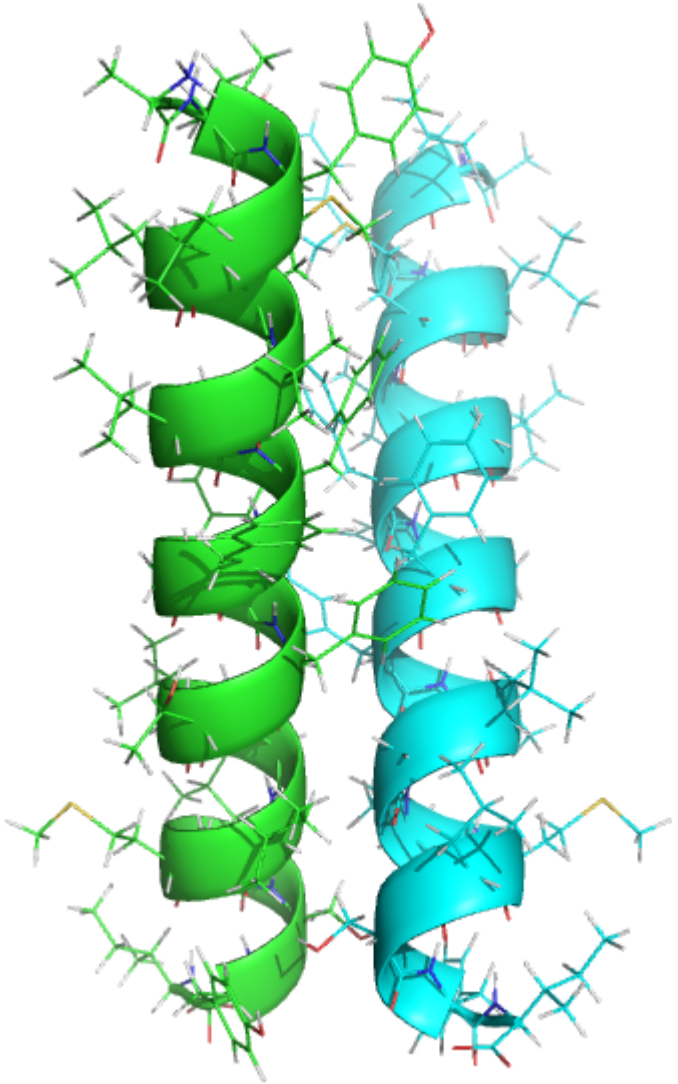
RESULTS ON THE HUMAN GENOME

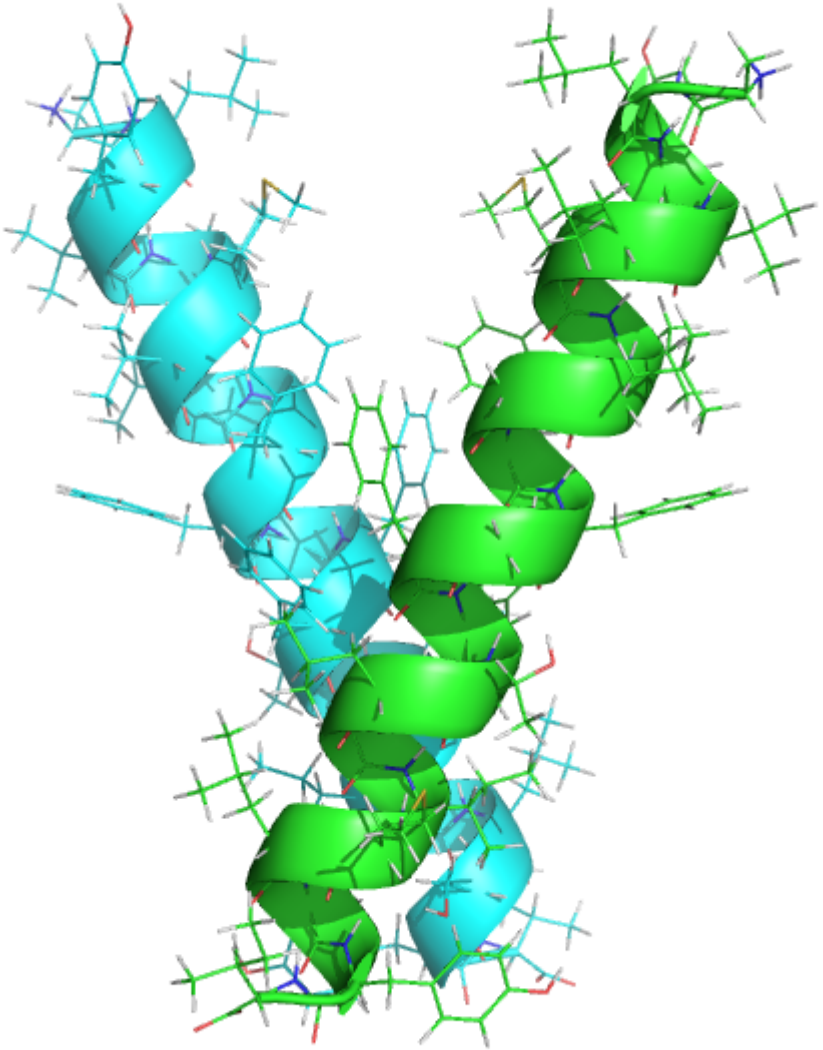
- High throughput screening of the human genome
  - CSF2R - NLGSVYIYVLLIVGTLVCGIVLGFLF
  - KCNE1 - ALYVLMVLGFFGFFTLGIMLSYI
  - LECT1 - VVLISGAVLLLFGAIGAFYFW
  - MEP1A - QVHGSVLGMVIGGTAGVIFLTFSIIAIL
  - NRP1 - ILITIIAMSALGVLLGAVCGVVL

CSF2R - NLGSVYIYVLLIVGTLVCGIVLGFLF

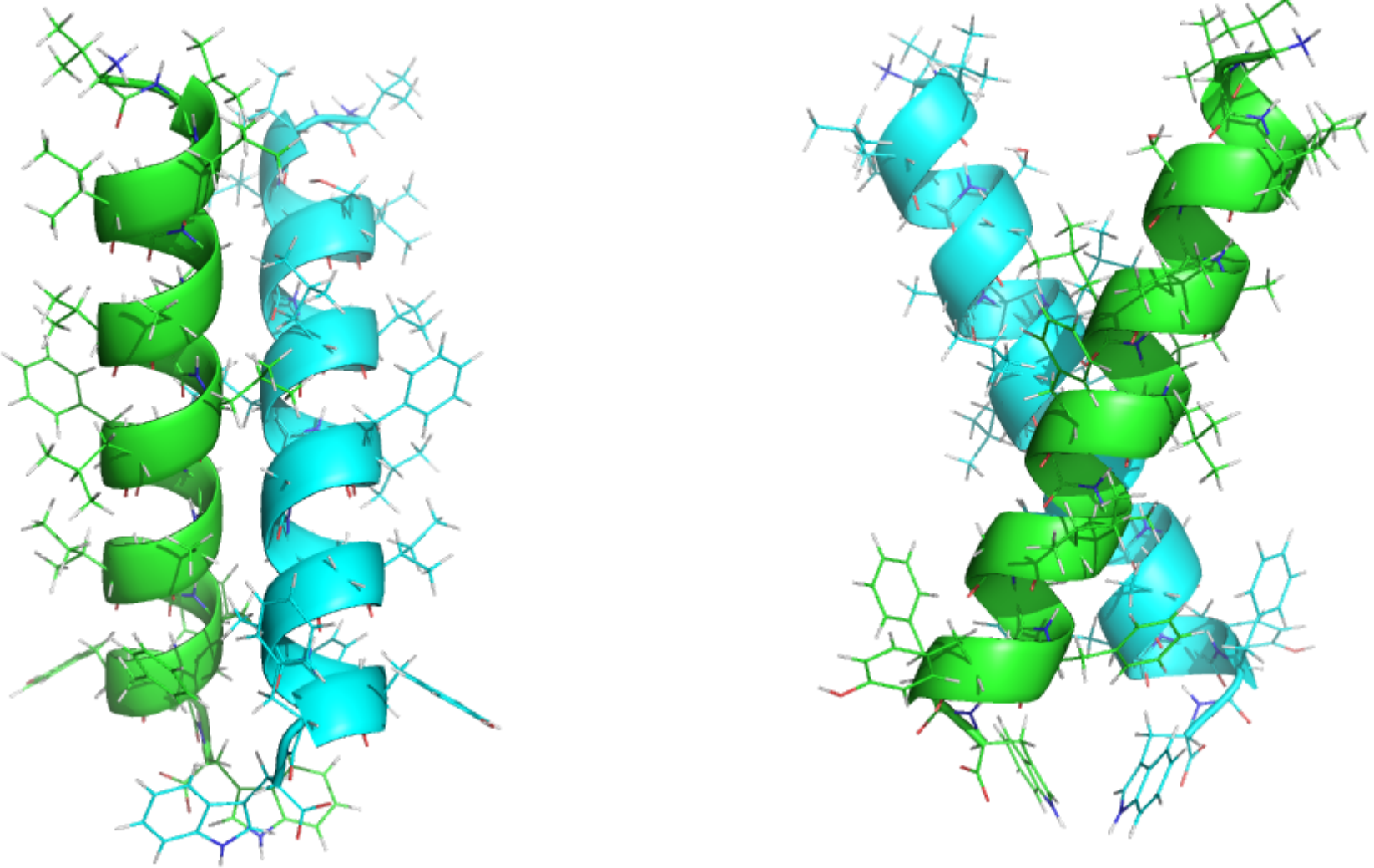


KCNE1 - ALYVLMVLGFFGFFTLGIMLSYI

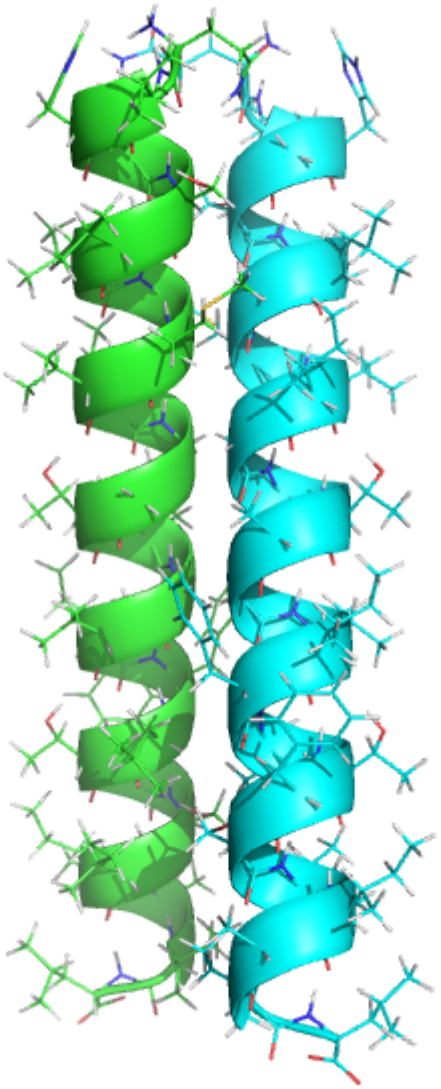


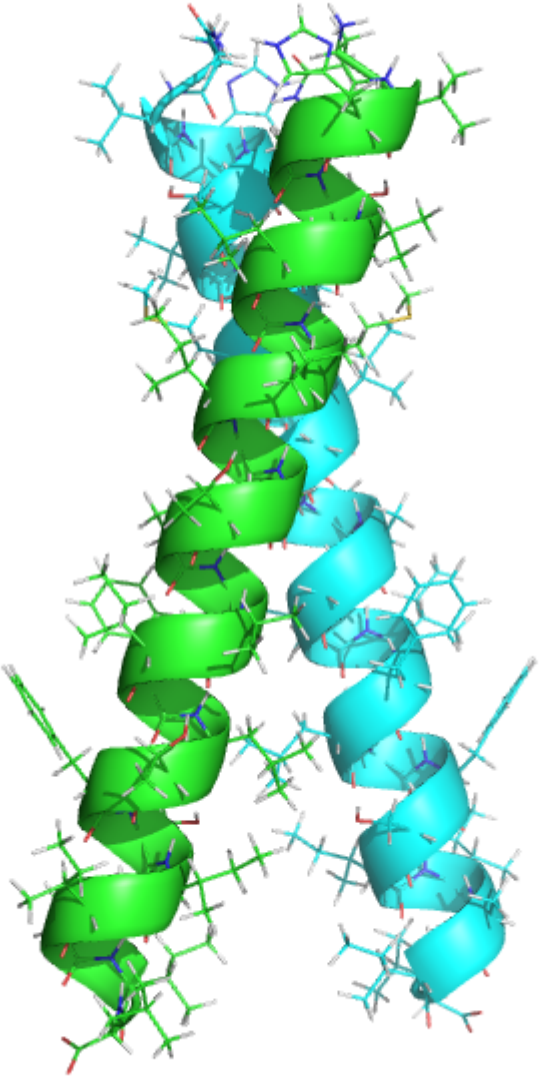


LECT1 - VVLISGAVLLLFGAIGAGAFYFW

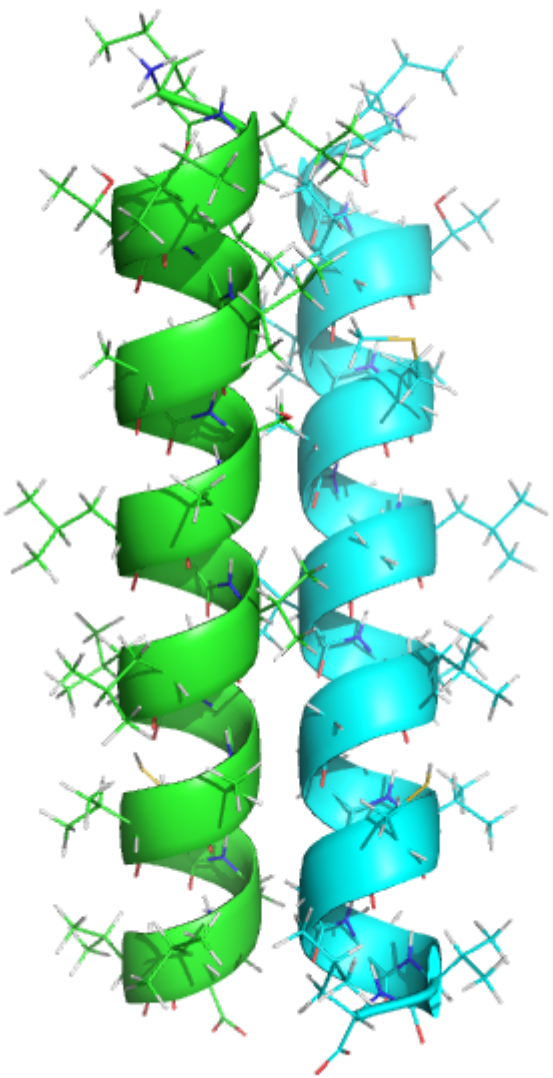


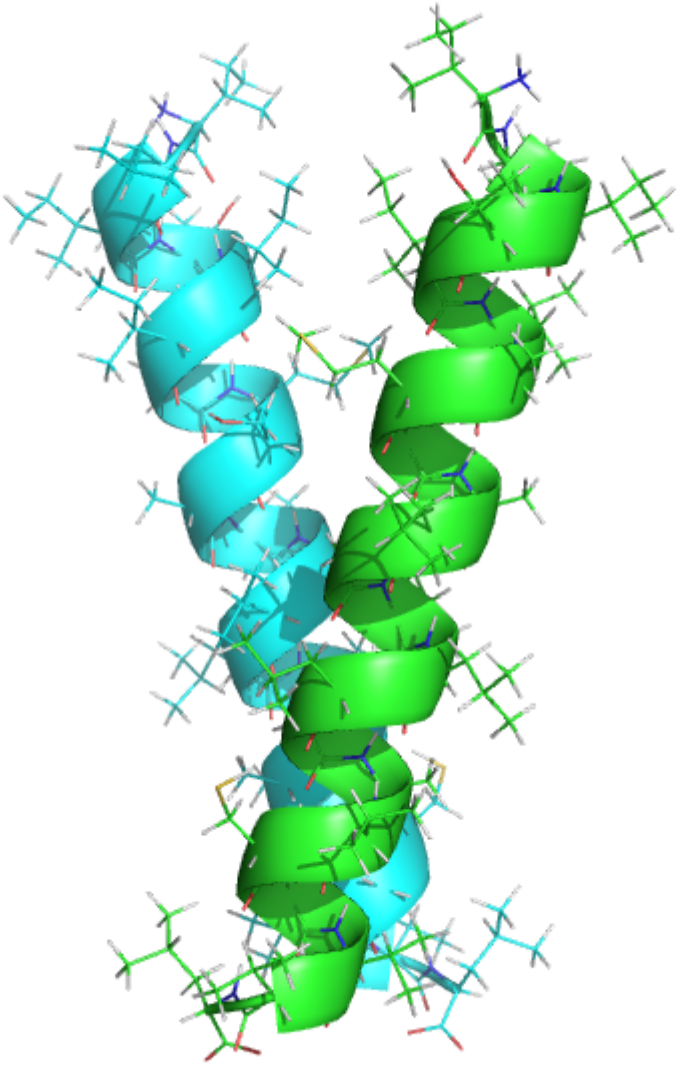
MEP1A - QVHGSVLGMVIGGTAGVIFLTFSIIAIL





NRP1 - ILITIIAMSALGVLLGAVCGVVL





Thank You

Alessandro Senes

Ambalika Khadria
Loren LaPointe
Ben Mueller

Center for High Throughput Computing
University of Wisconsin-Madison