

Evaluating molecular dynamics force fields using computed NMR chemical shifts

David Ryan Koes
John K. Vries

Computational and Systems Biology
University of Pittsburgh

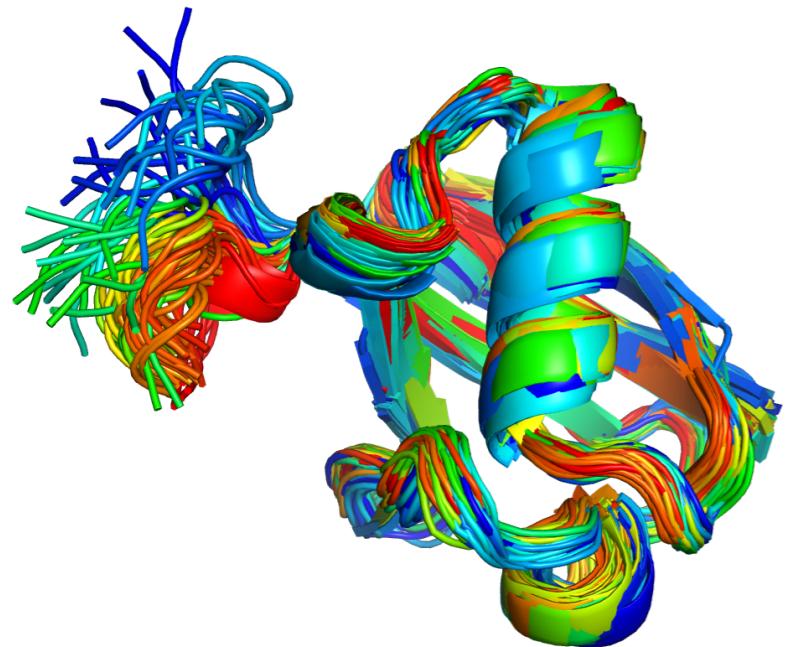


Evaluating molecular dynamics force fields using computed NMR chemical shifts

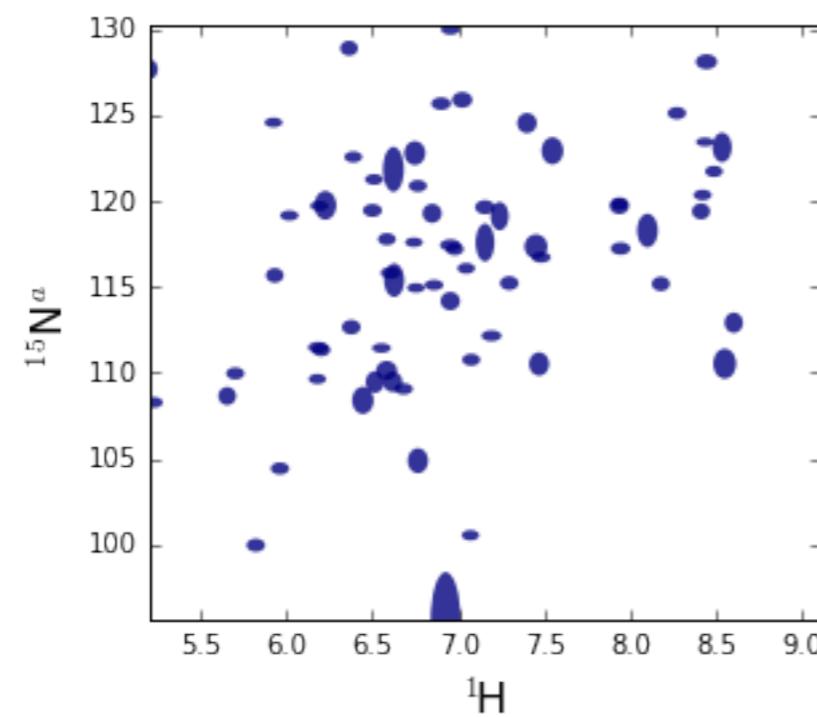
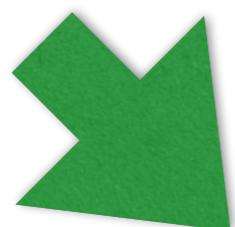
David Ryan Koes
John K. Vries

Computational and Systems Biology
University of Pittsburgh

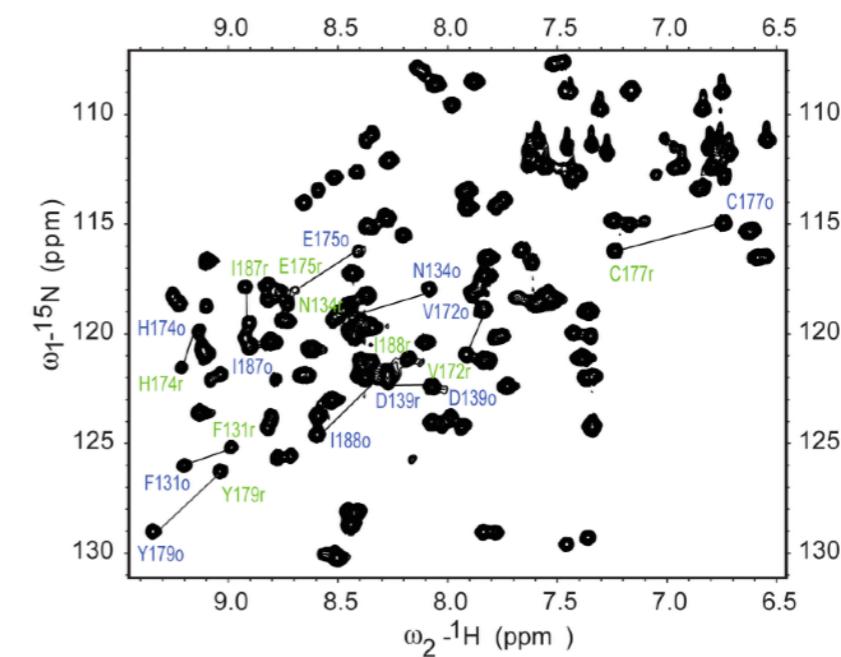




Molecular
Dynamics



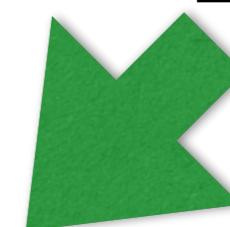
?
==

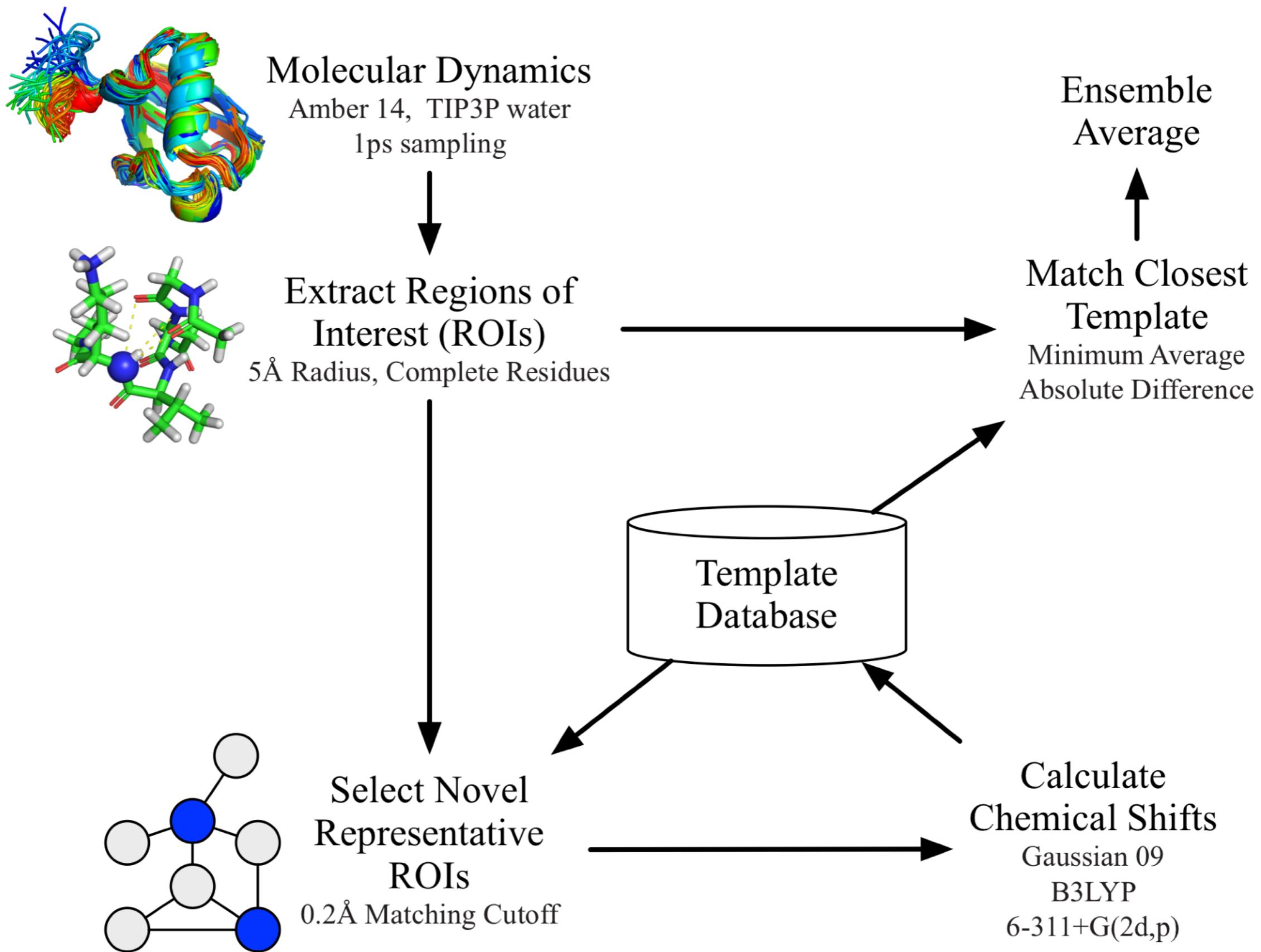


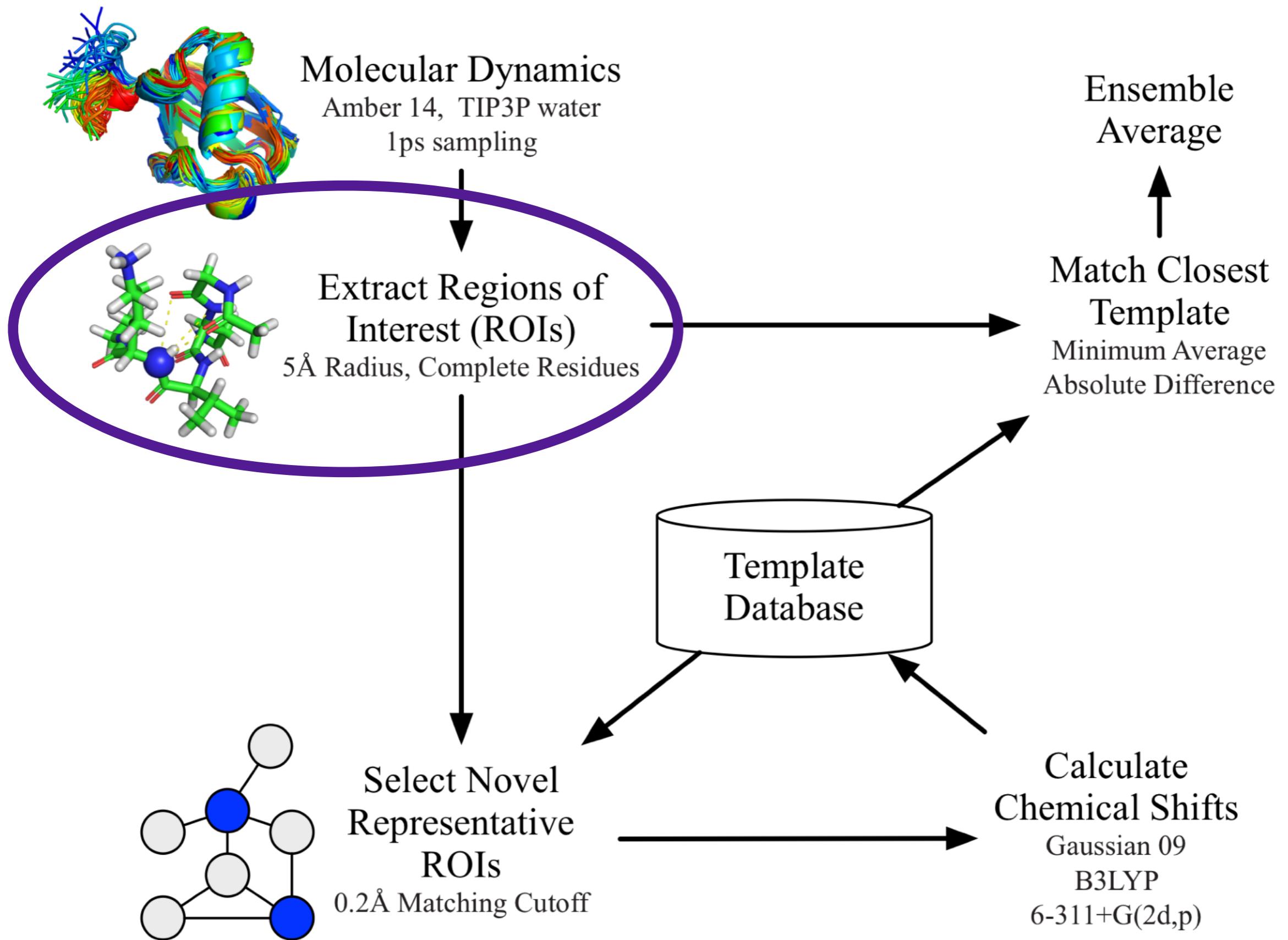
. DOI:10.1371/journal.ppat.1000960



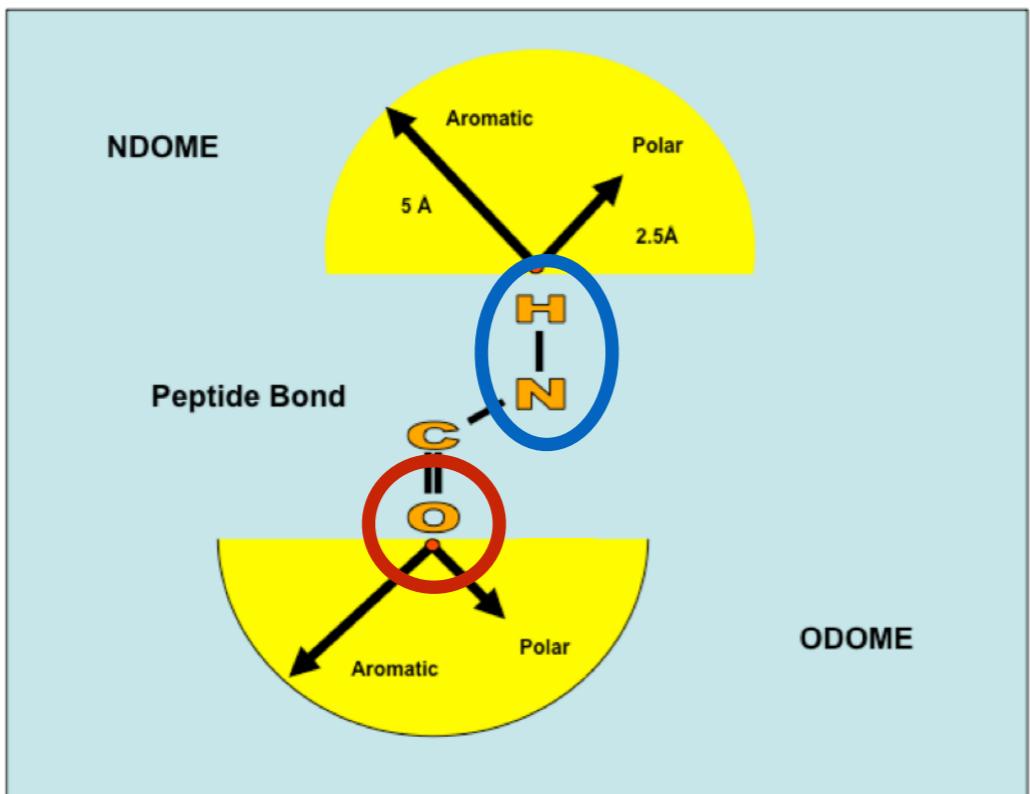
Experiment



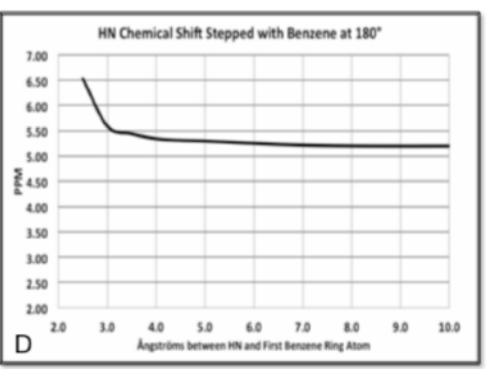
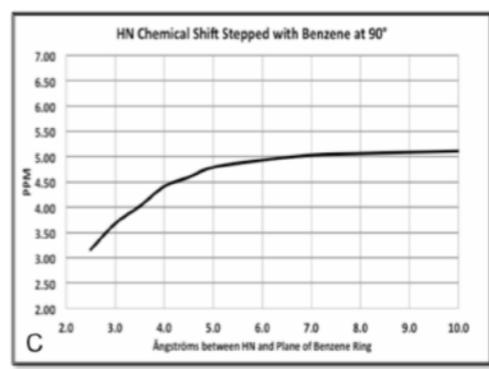
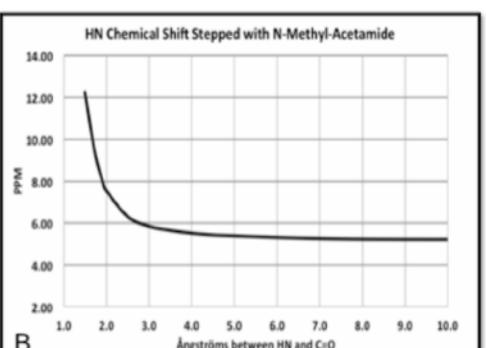
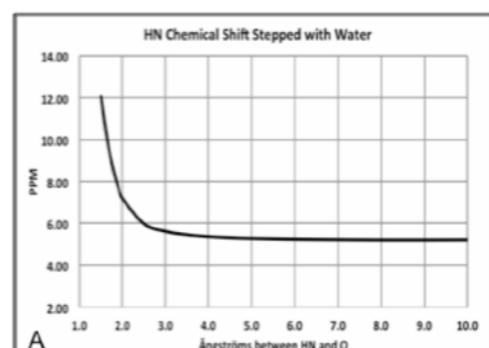
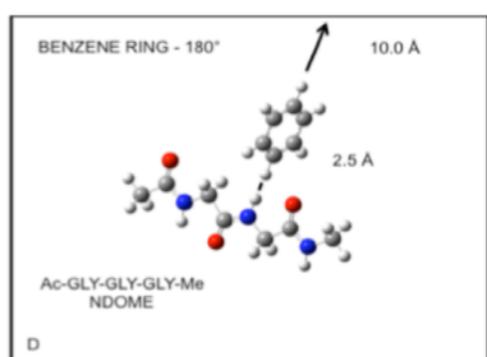
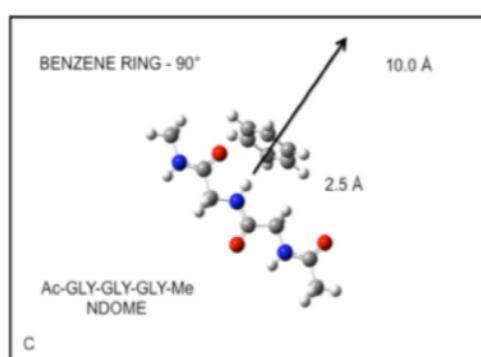
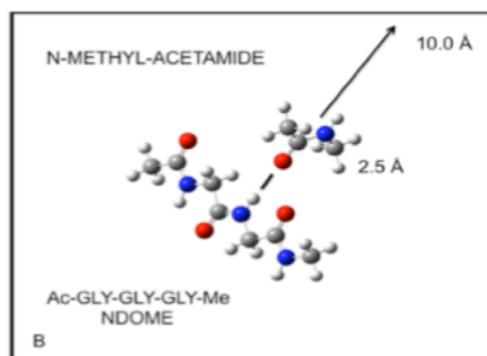
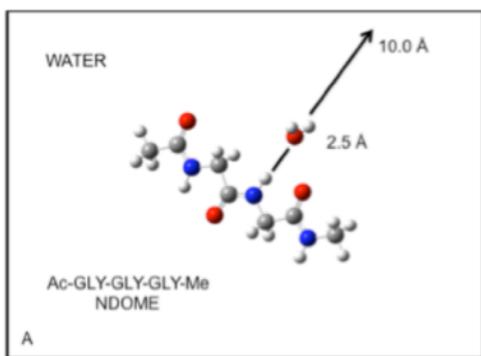




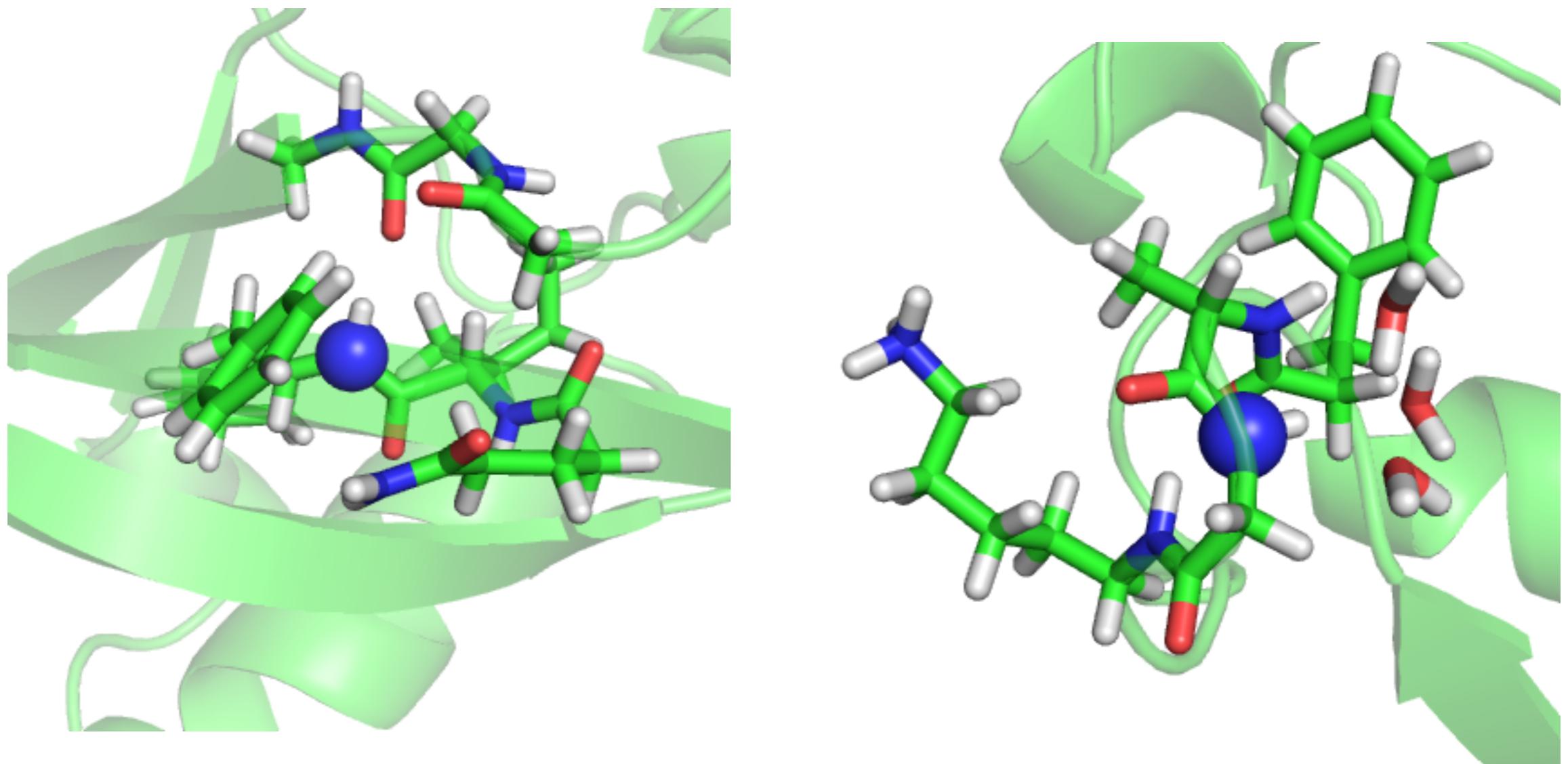
Regions of Interest (ROI)



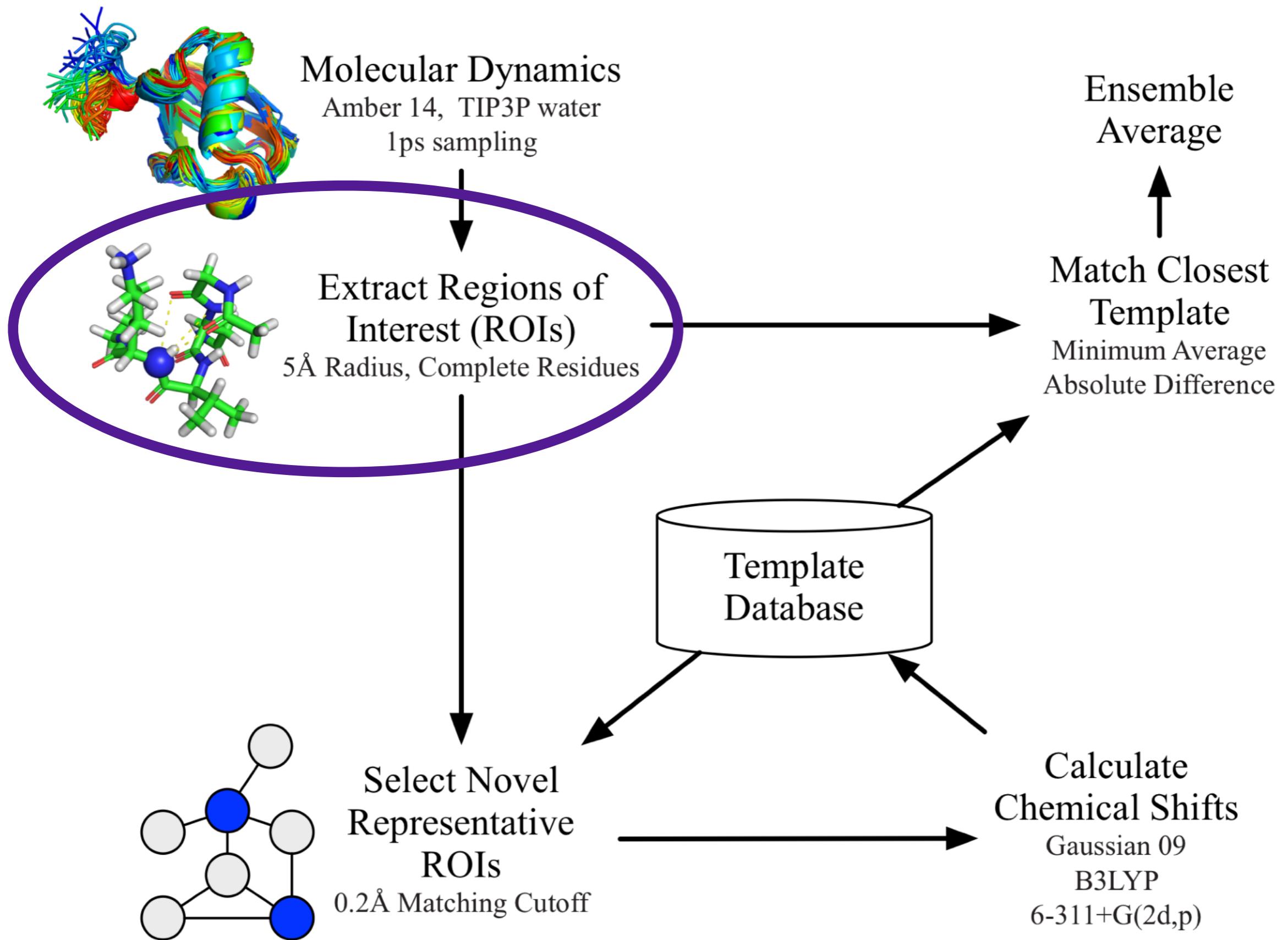
- Identify all nearby polar and aromatic atoms
- Separate analysis for O and NH

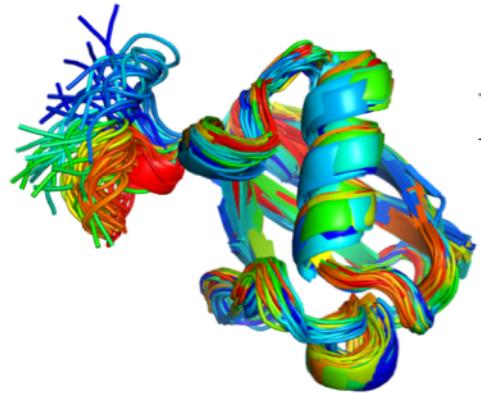


ROI Conformer

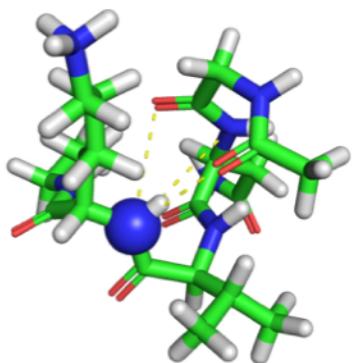


- Extend target residue to capped 4-residue peptide
- Extend nearby polar/aromatic atoms to 3-residue peptide
- Include nearby waters

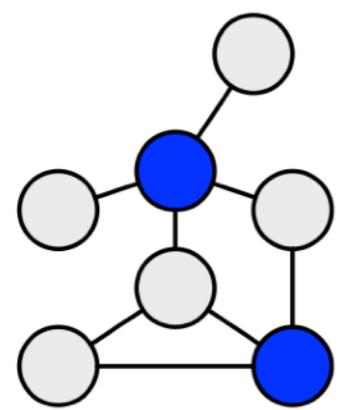




Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



Extract Regions of
Interest (ROIs)
5Å Radius, Complete Residues



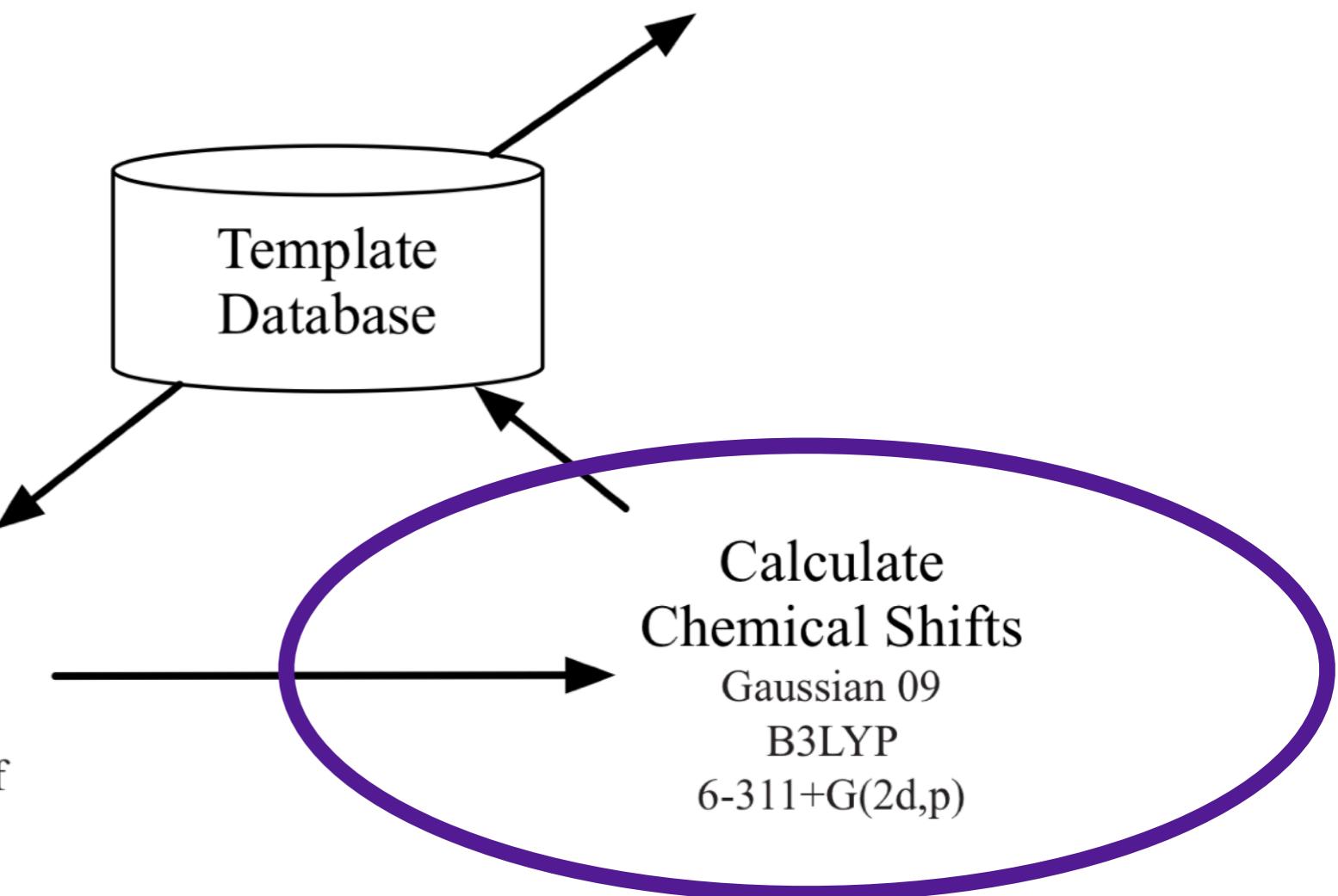
Select Novel
Representative
ROIs
0.2Å Matching Cutoff

Template
Database

Ensemble
Average

Match Closest
Template
Minimum Average
Absolute Difference

Calculate
Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)





HartreeFocker @edsherer · 4h

#compchem when you tell someone you're using B3LYP #acsdc



4

9





HartreeFocker @edsherer · 4h

#compchem when you tell someone you're using B3LYP #acsdc

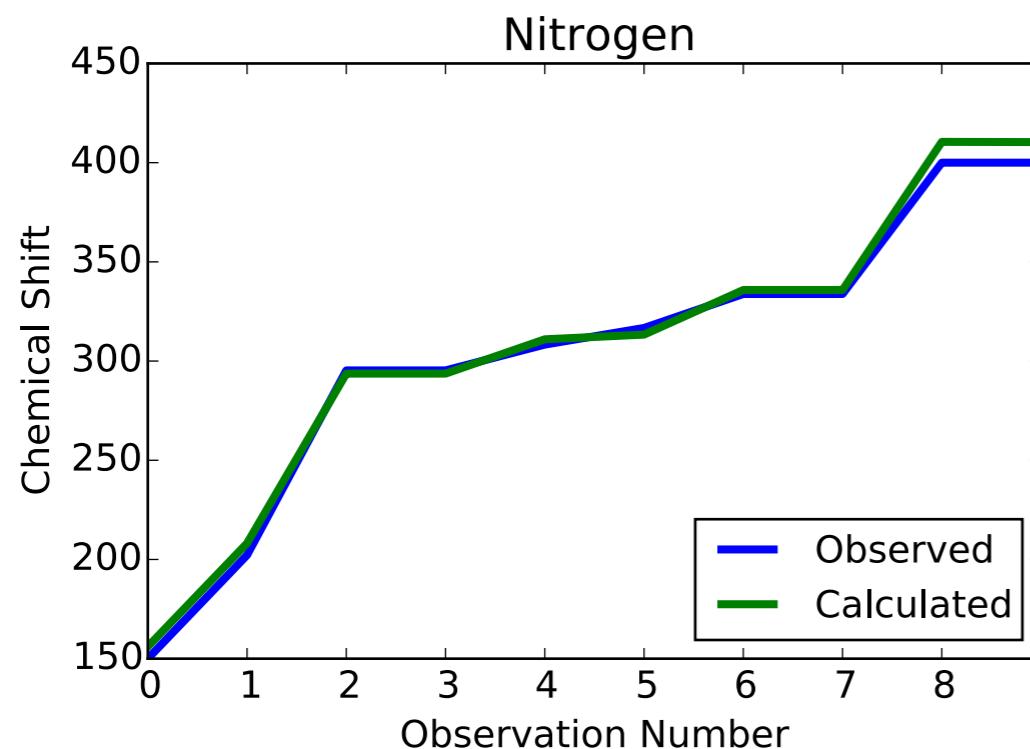
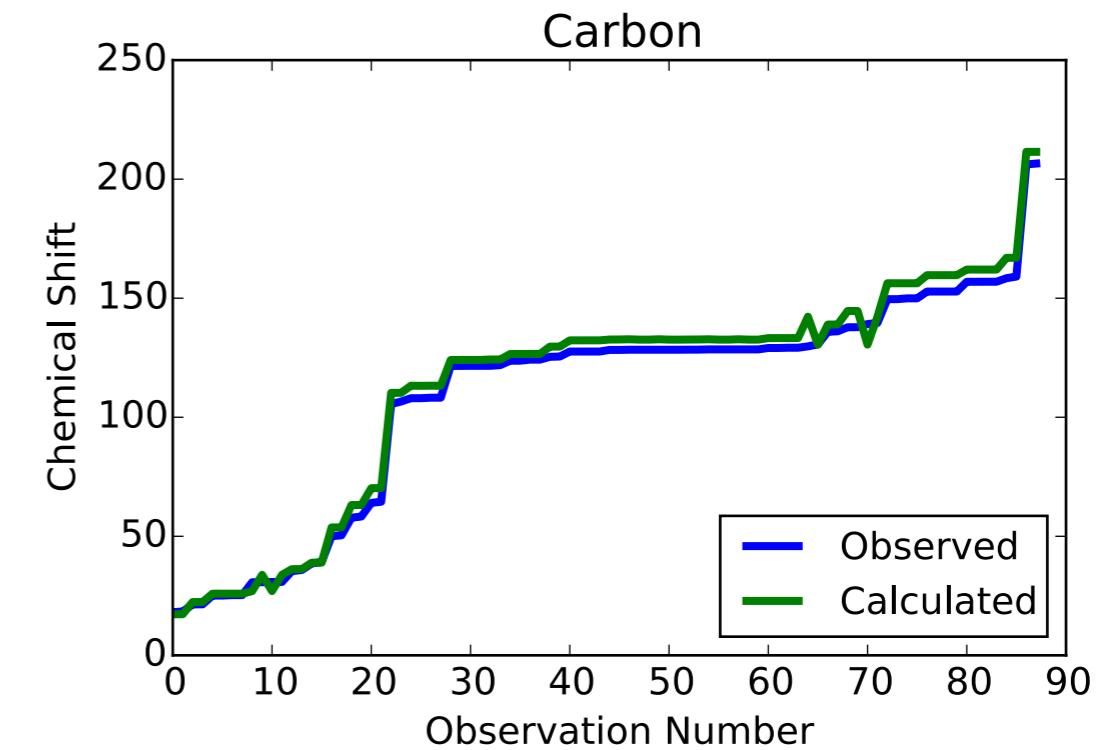
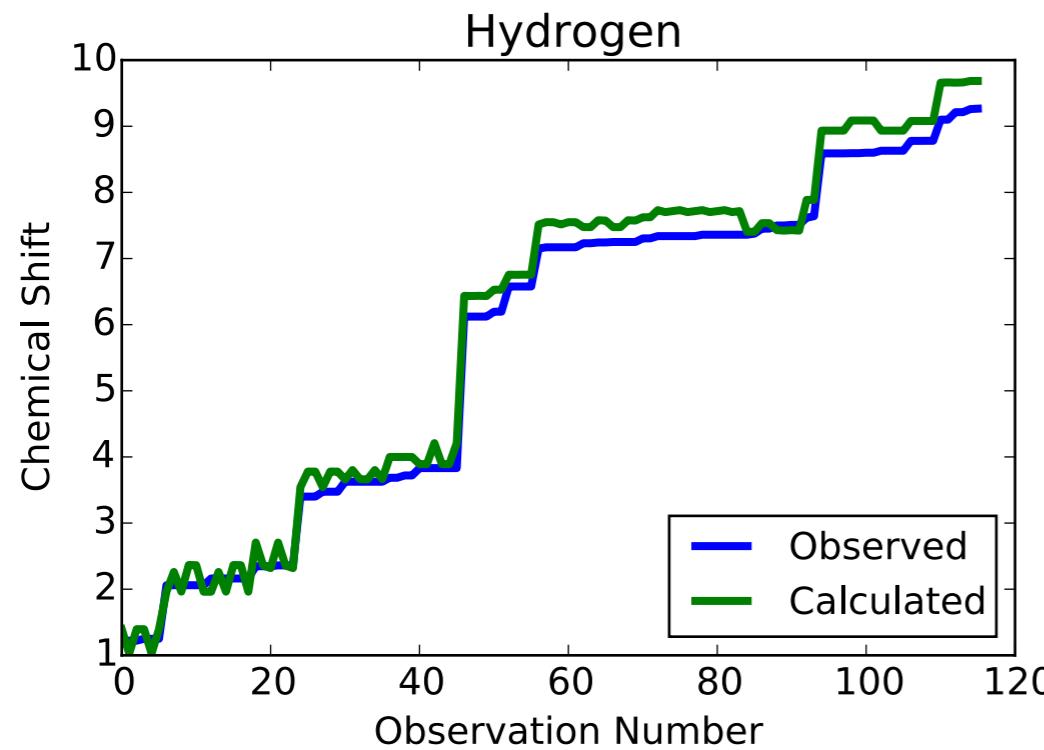


4

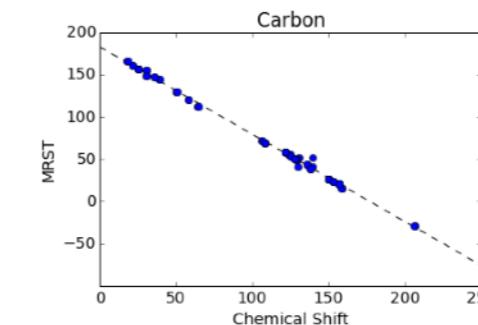
9



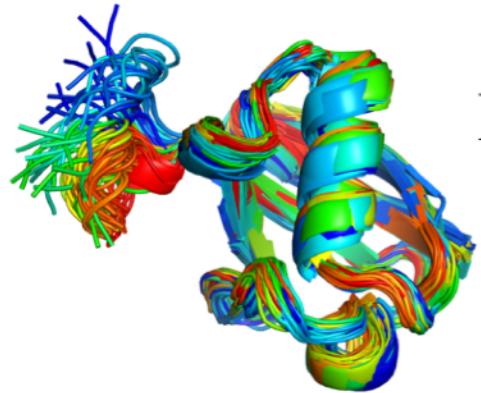
Calculate Chemical Shifts



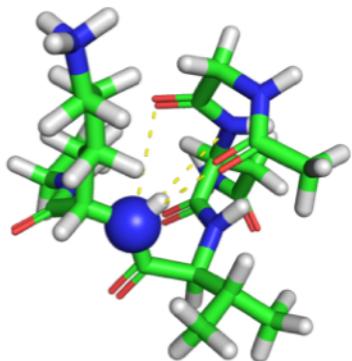
Gaussian G09
6-311+G(2d,p)
Reference Values:
Organic Compounds + PCM



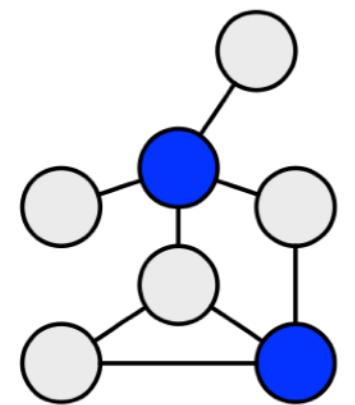
$$\begin{aligned} {}^1\text{H} &= 32.023 \\ {}^{13}\text{C} &= 182.629 \\ {}^{15}\text{N} &= 233.153 \end{aligned}$$



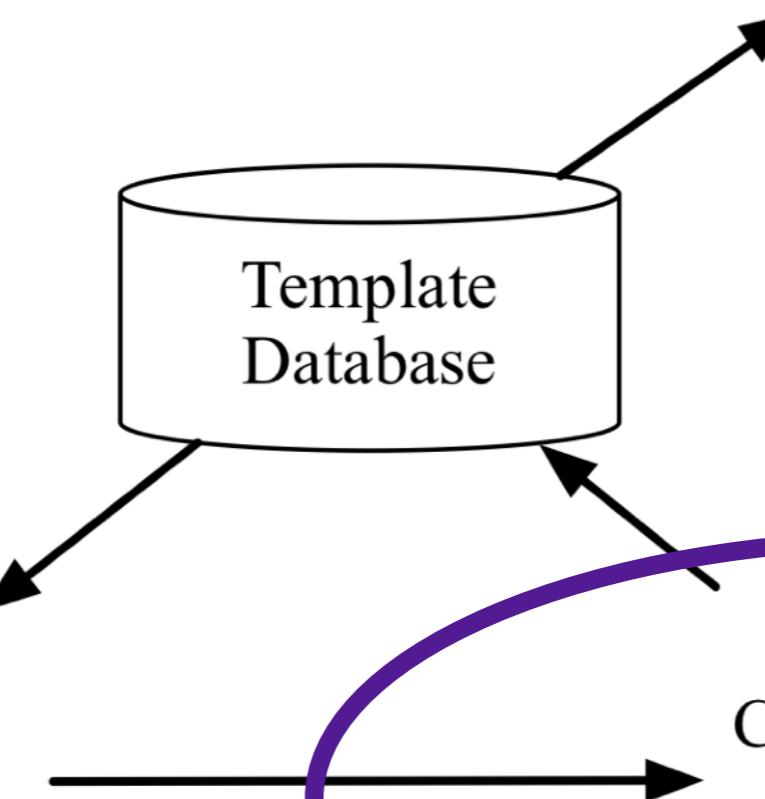
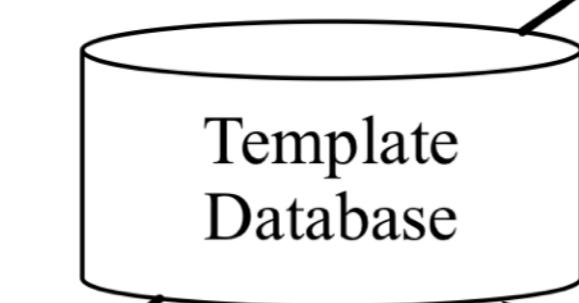
Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



Extract Regions of
Interest (ROIs)
5Å Radius, Complete Residues

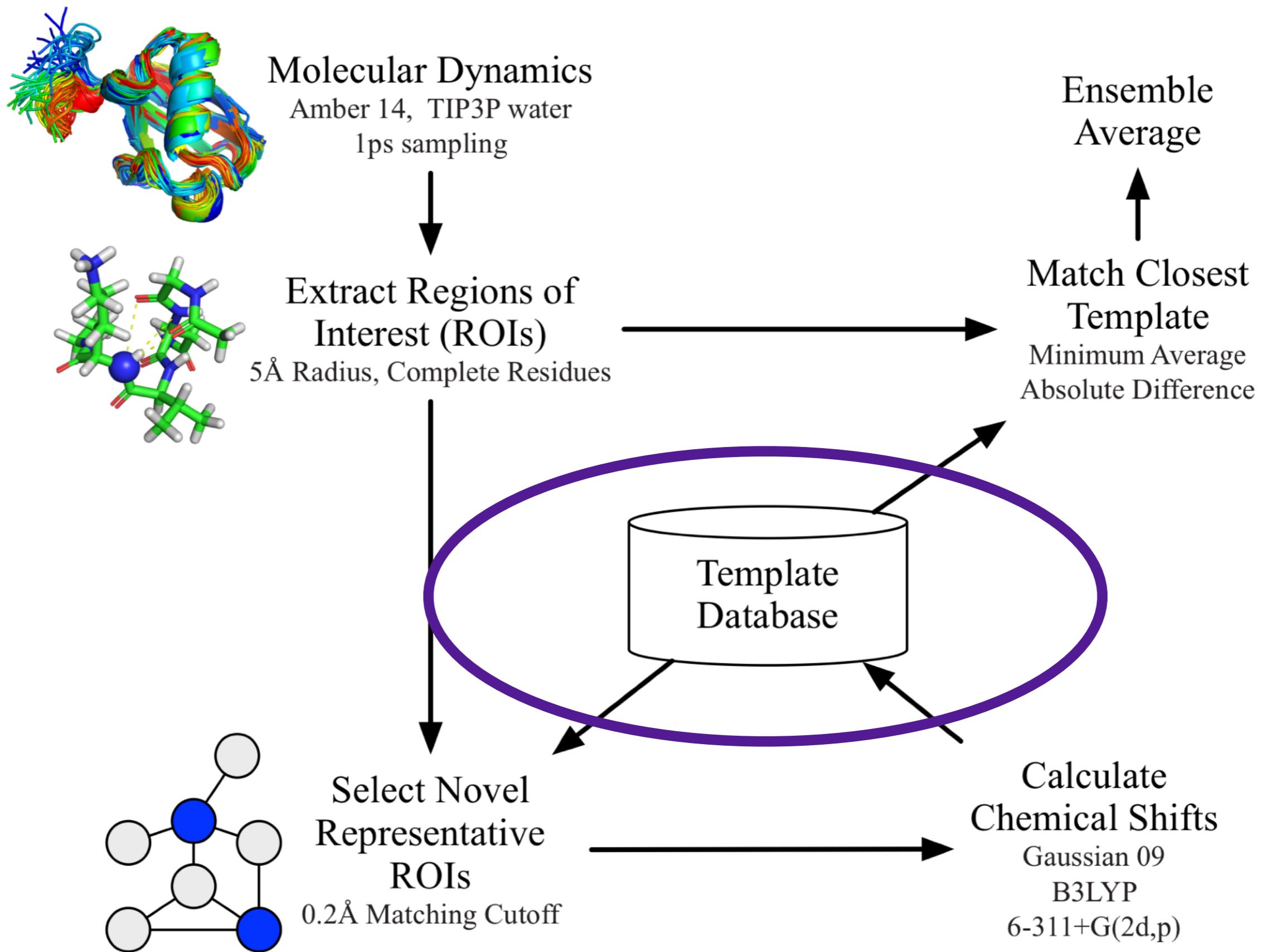


Select Novel
Representative
ROIs
0.2Å Matching Cutoff

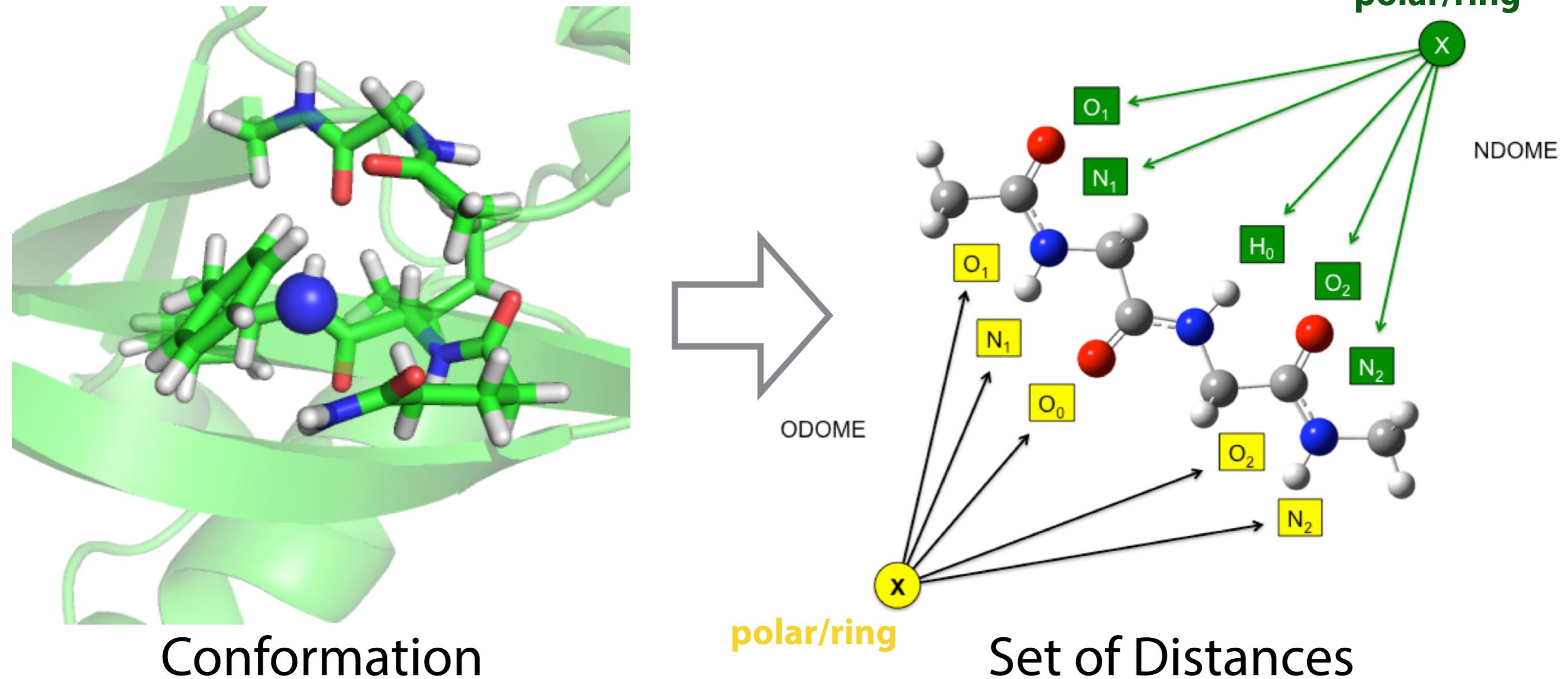


Ensemble
Average
↑
Match Closest
Template
Minimum Average
Absolute Difference

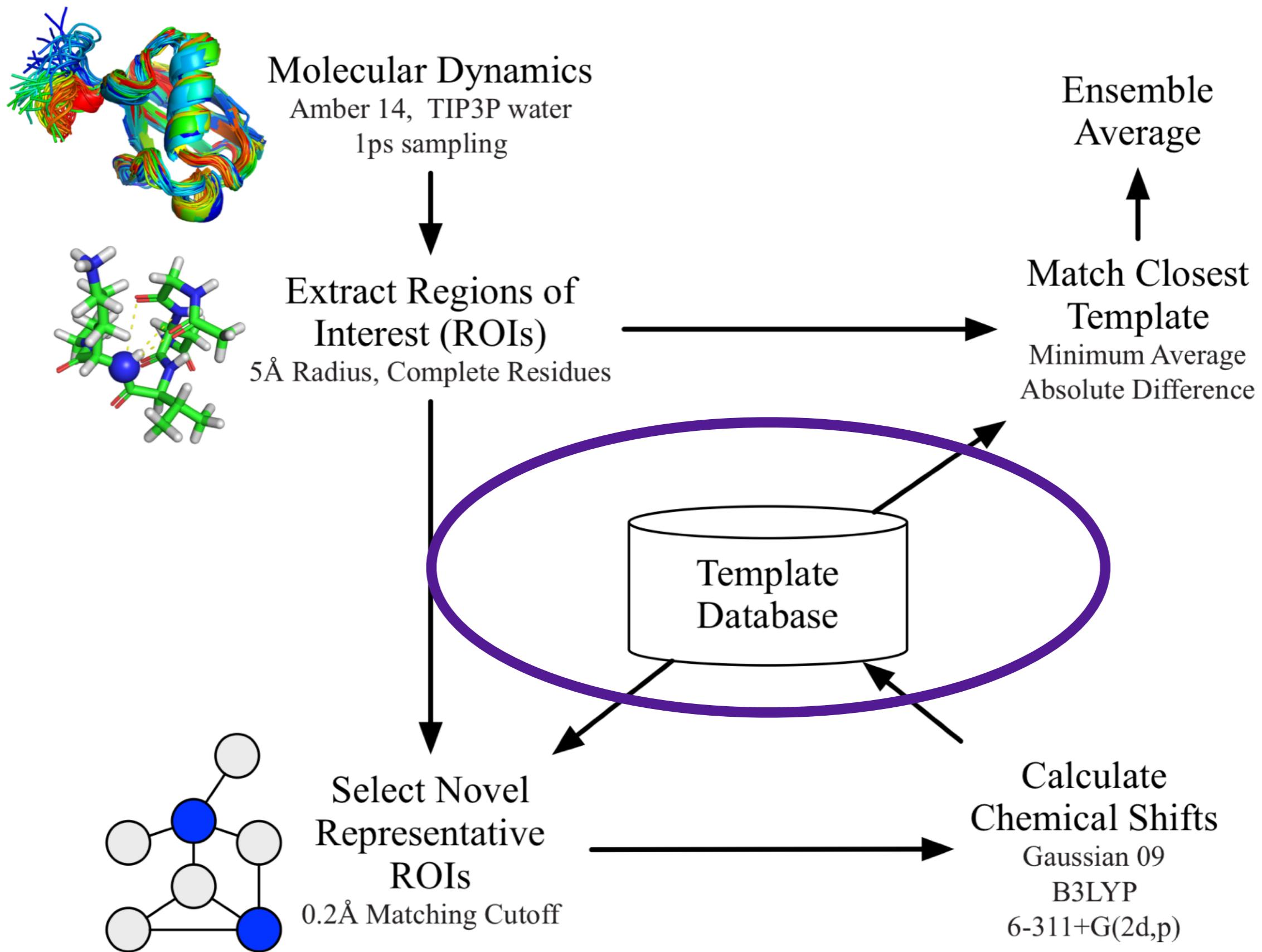
Calculate
Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

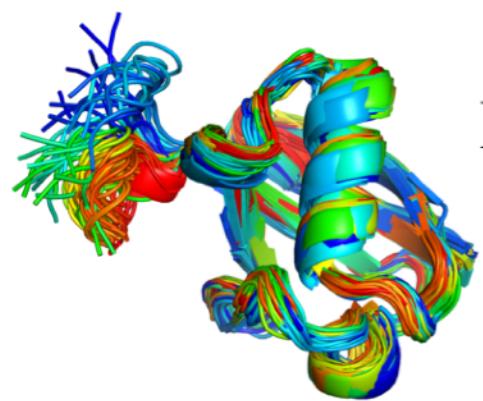


Pattern Templates

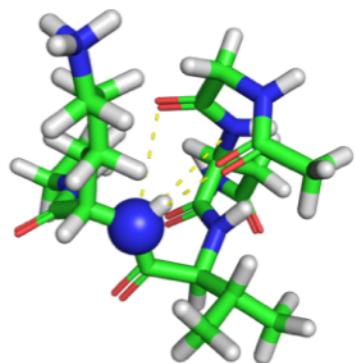


O:R:R|3.512|3.408|3.131|3.295|4.399|2.129|4.802|3.906|4.912|5.907|4.884|5.752|5.908|6.788|8.
O|3.361|2.866|3.072|3.337|4.310|1.945|4.619|3.614|4.312|5.528
Z|3.652|3.518|3.071|3.151|4.390

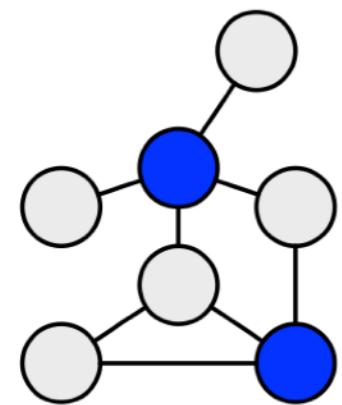




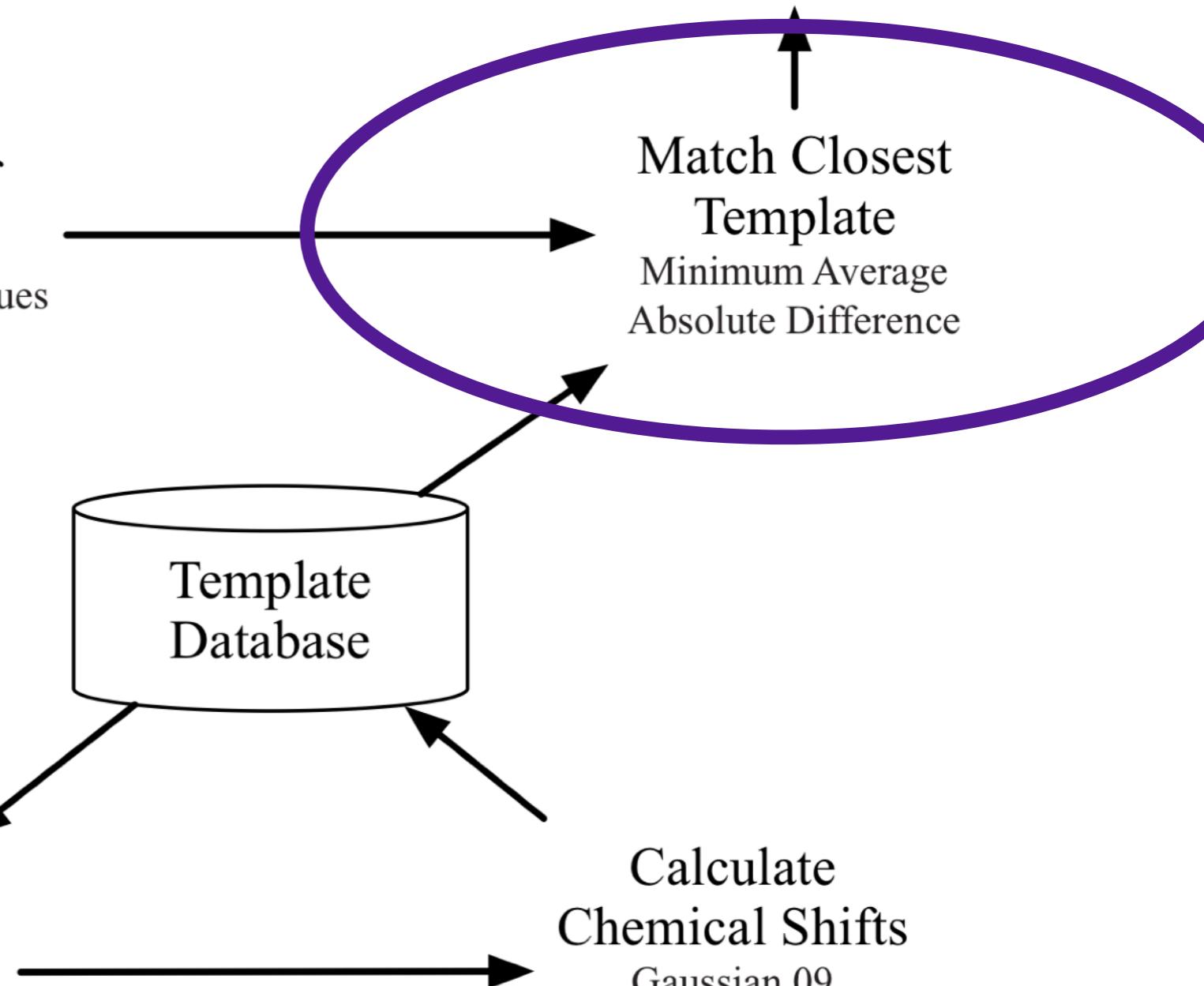
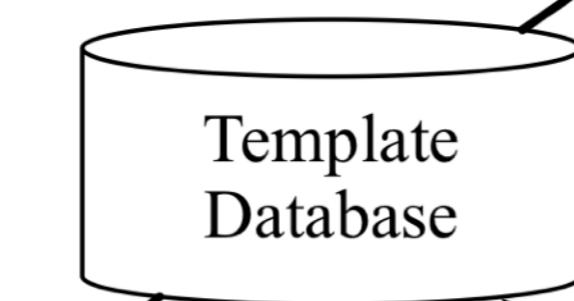
Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



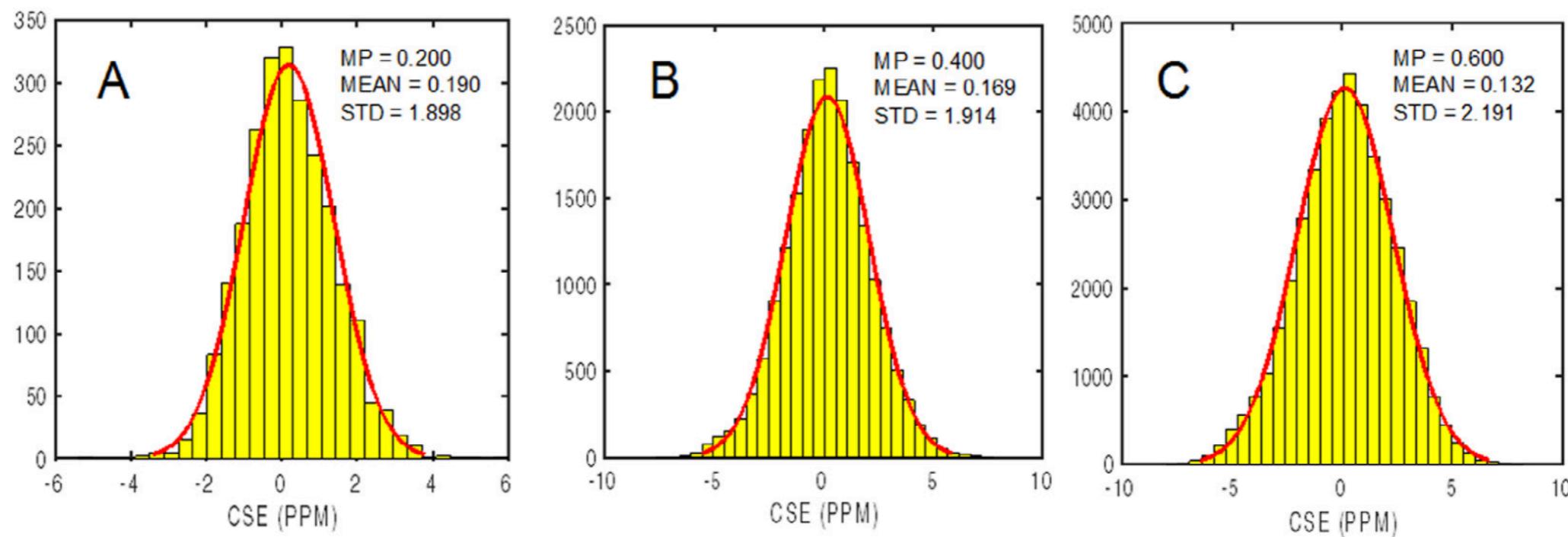
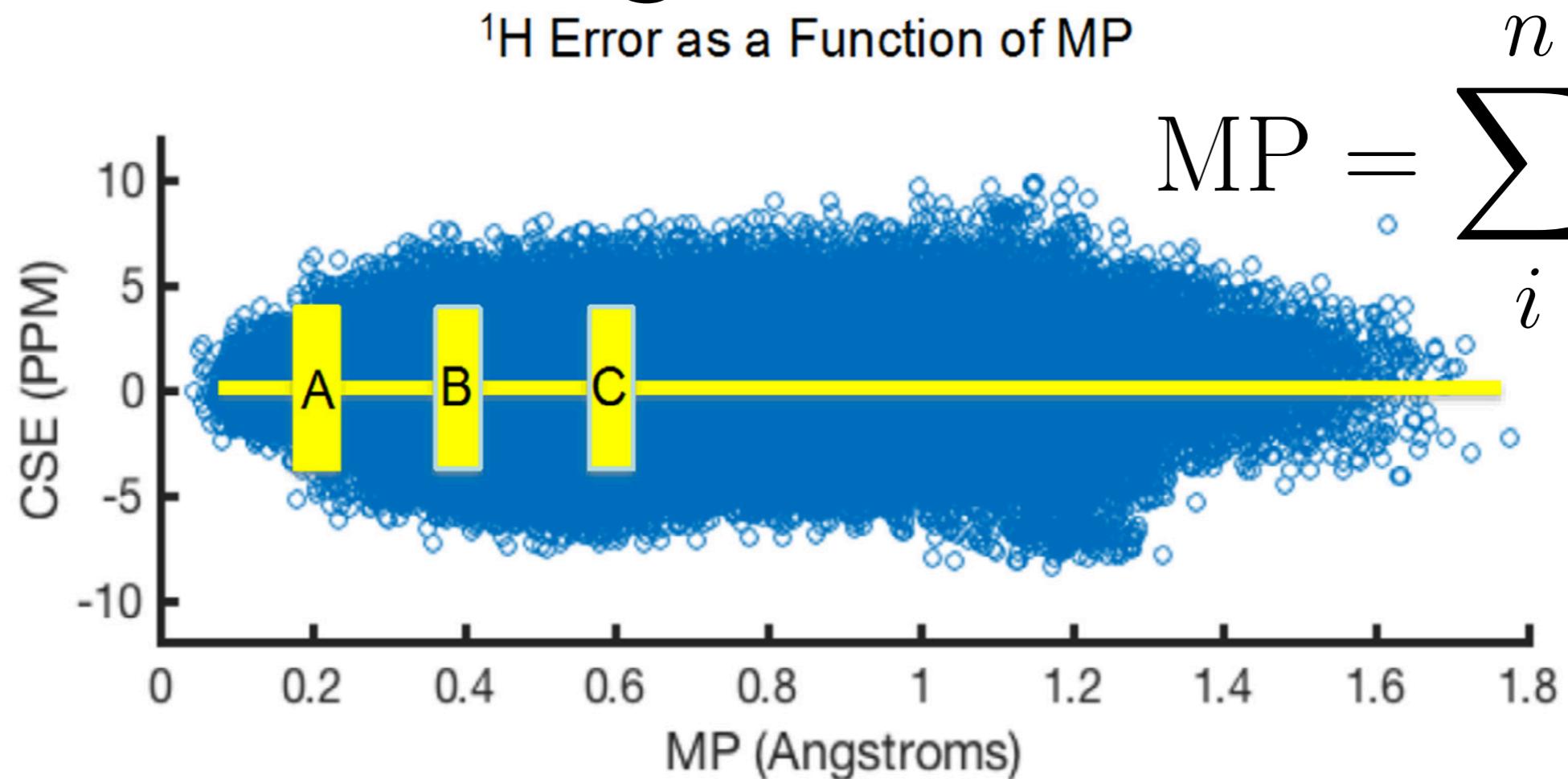
Extract Regions of
Interest (ROIs)
5Å Radius, Complete Residues

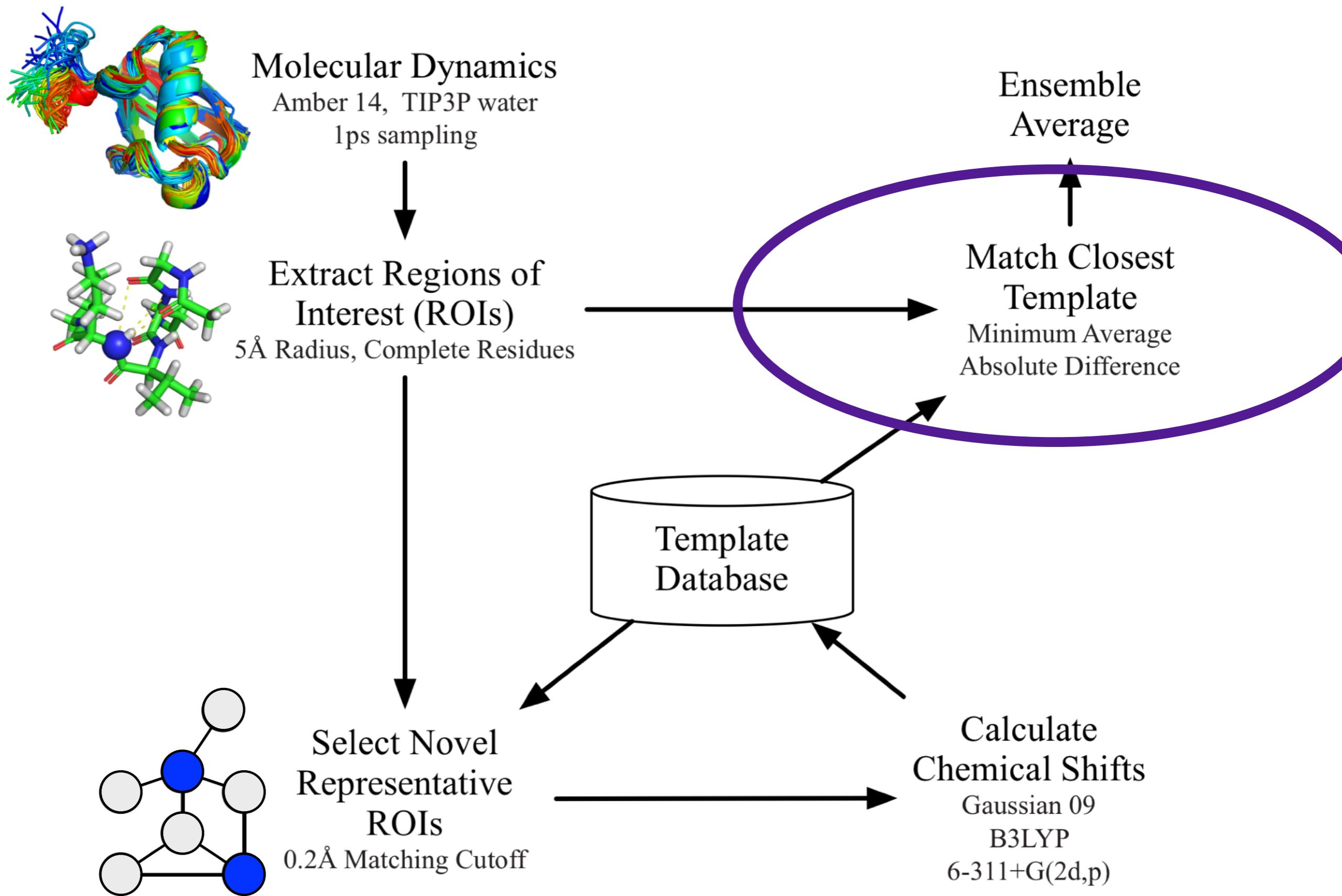


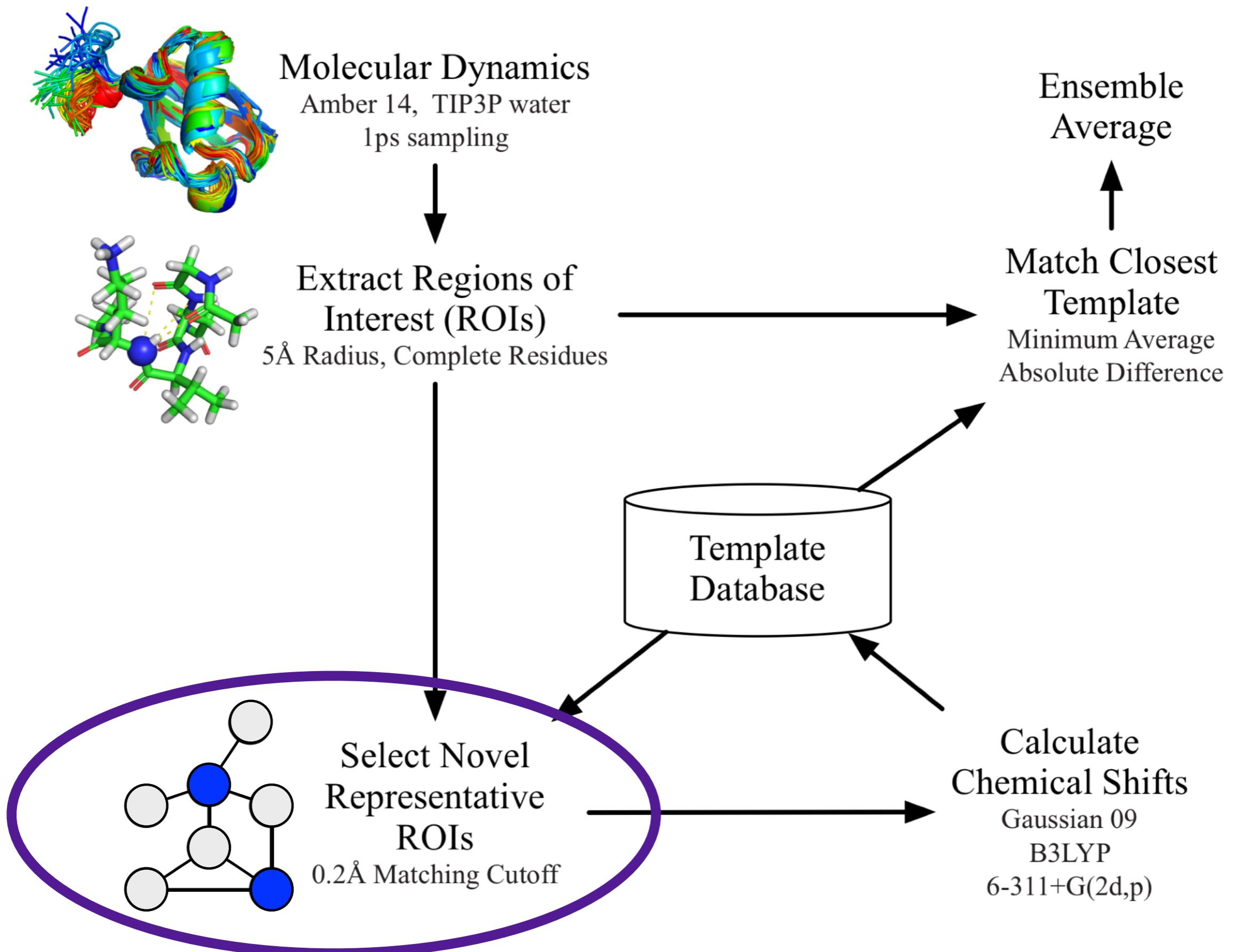
Select Novel
Representative
ROIs
0.2Å Matching Cutoff



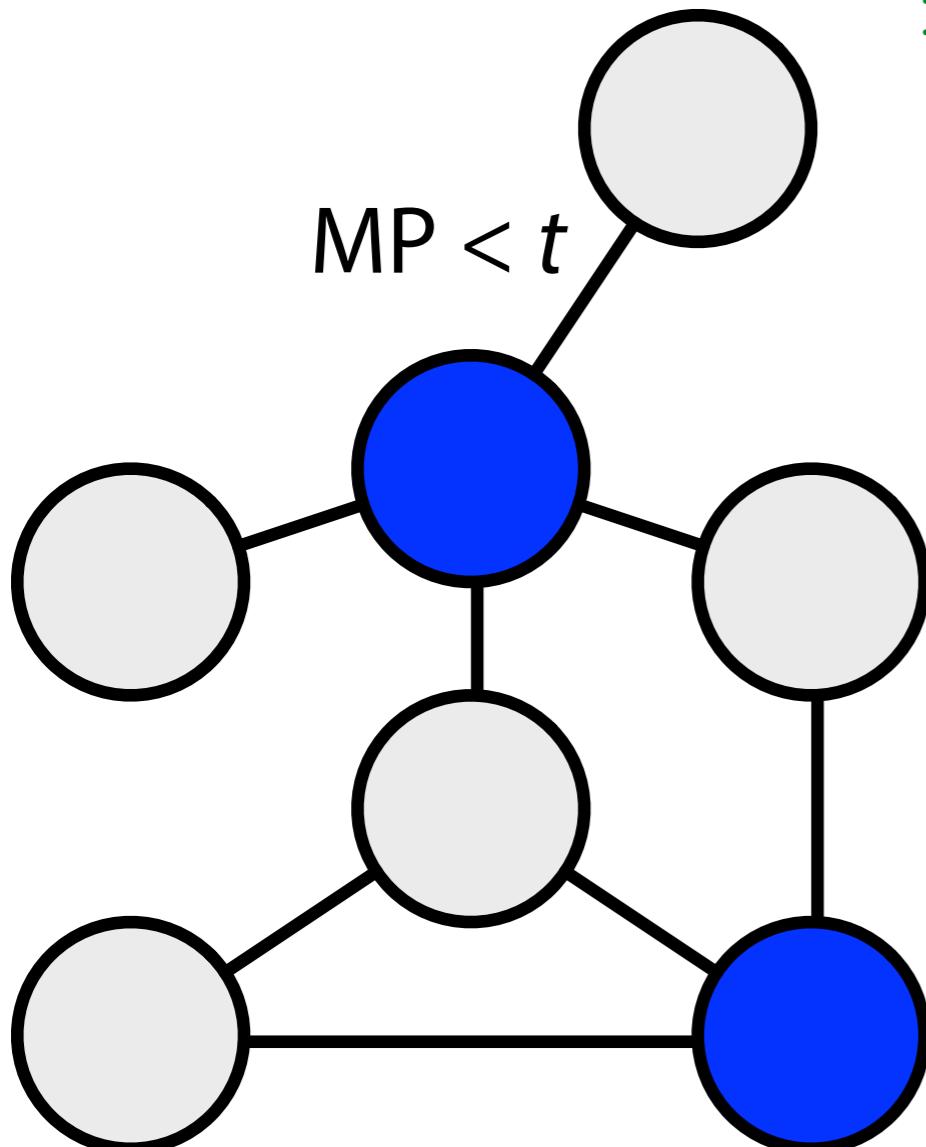
Matching Parameter







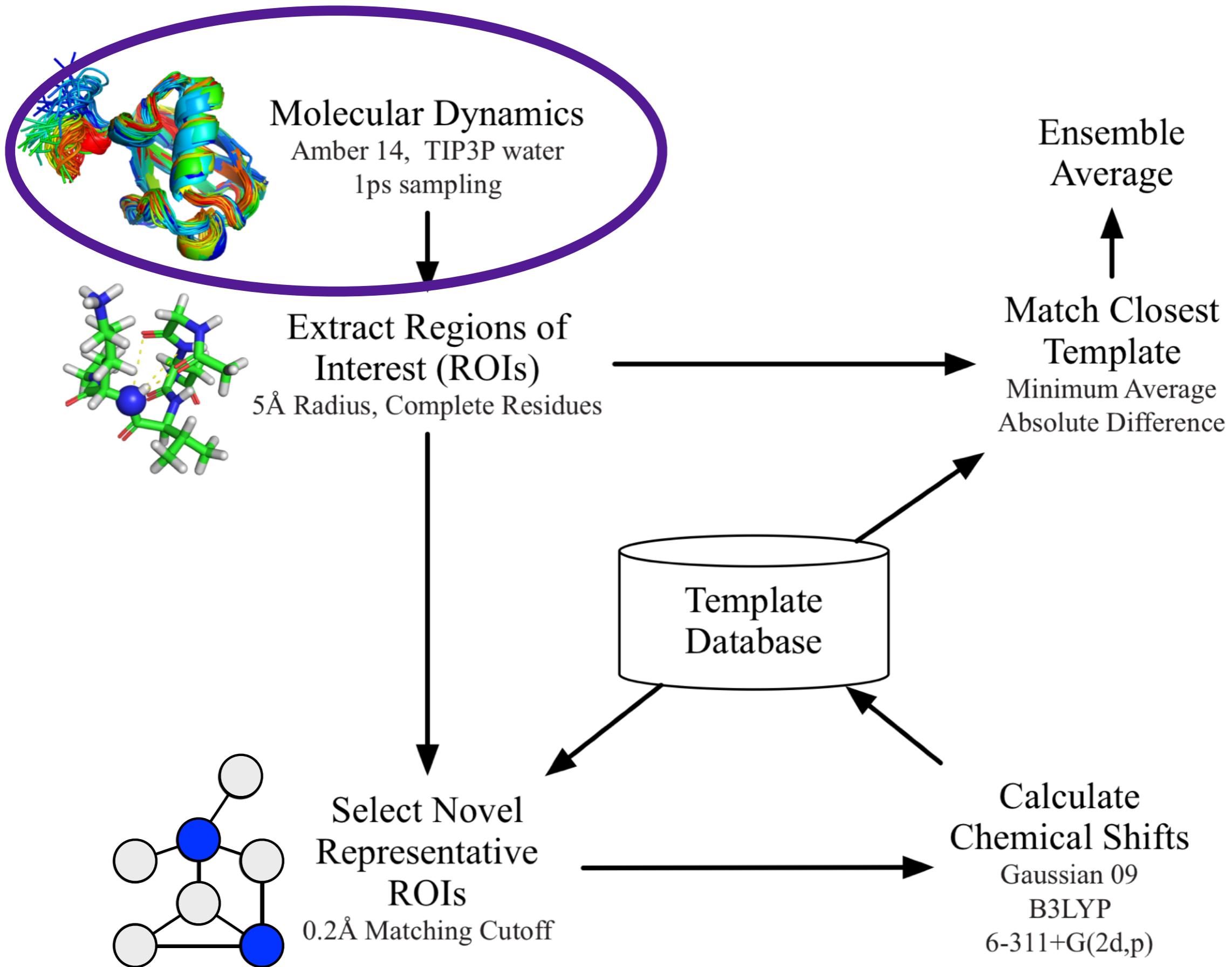
ROI Selection

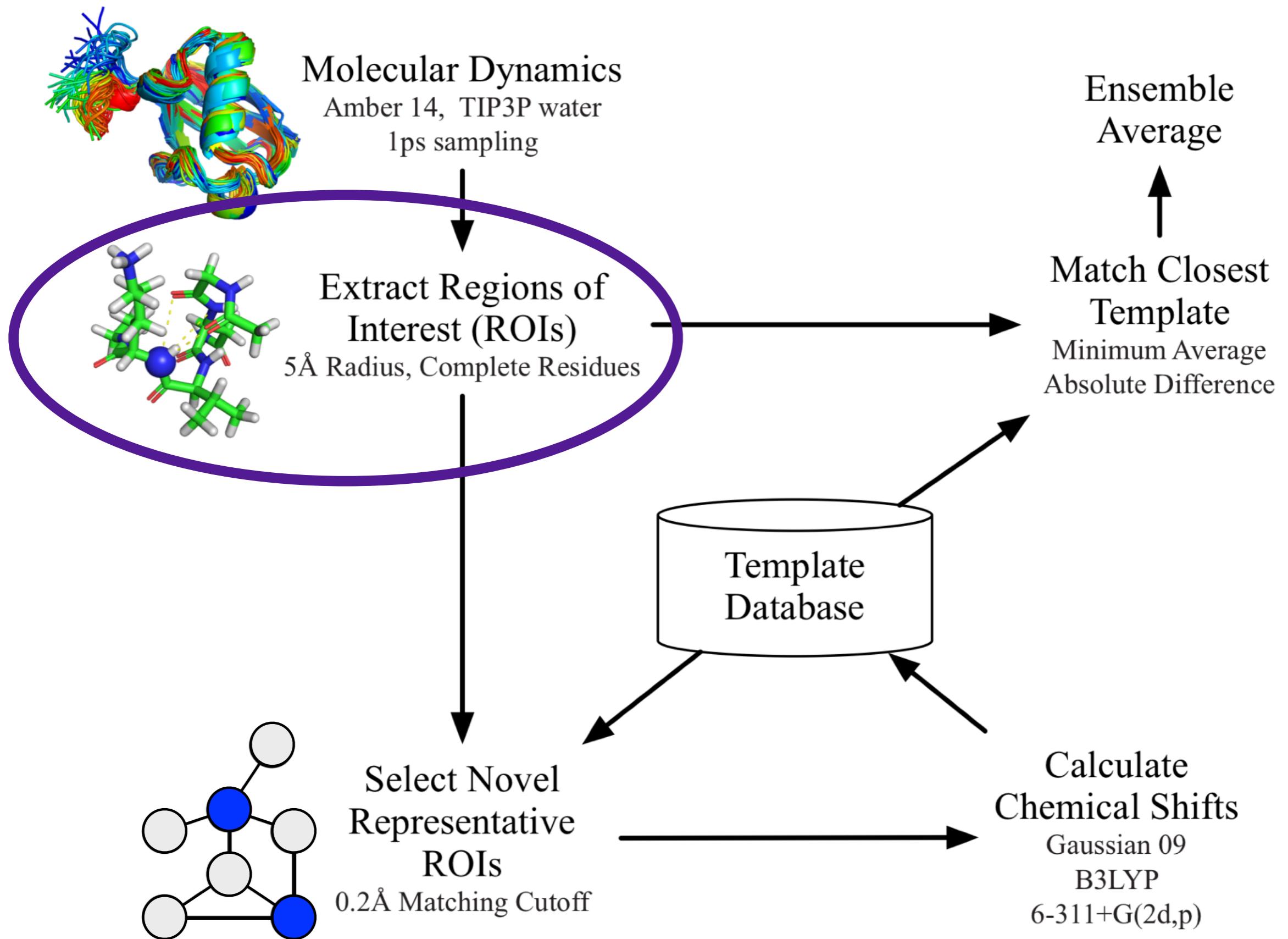


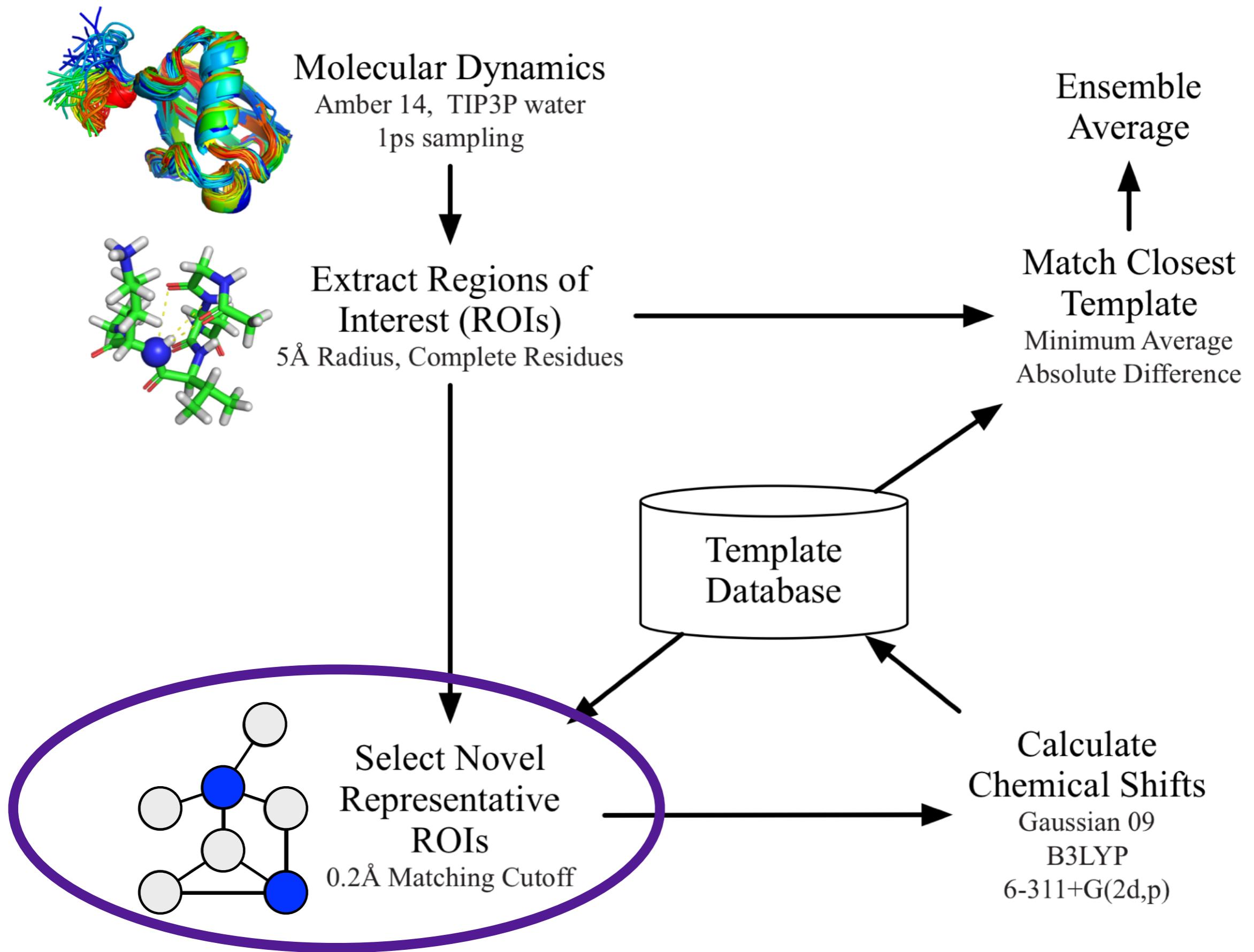
```
foreach pattern
  create_graph  $G_{pattern}$ 
  foreach frame in simulation
    foreach ROI with pattern
      create_node  $n$  in  $G_{pattern}$ 
      foreach  $m$  in  $G_{pattern}$ 
        if  $MP(n, m) < threshold$ 
          add_edge  $(n, m)$  to  $G_{pattern}$ 
```

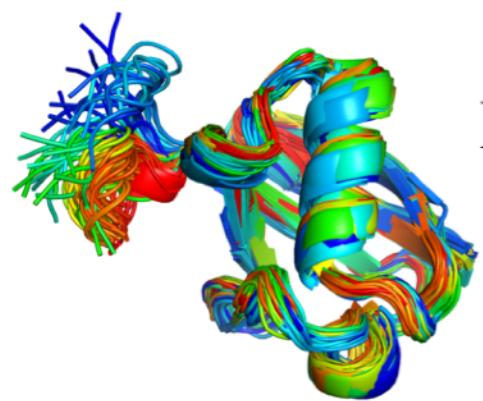
Solve *dominating set problem*
Smallest subset of nodes D such
that every node of G is either in D
or is adjacent to a node in D

NP-hard, but good approximation algorithm

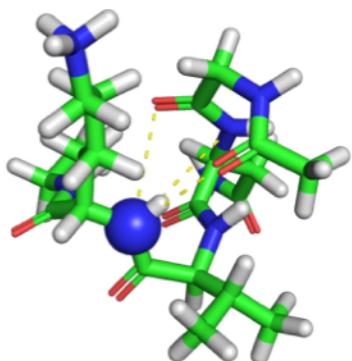




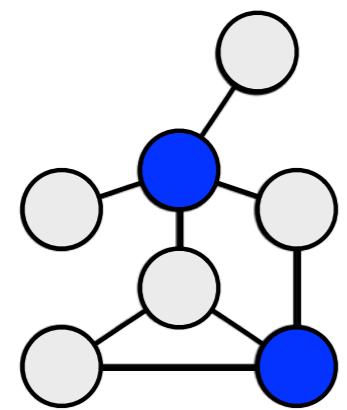




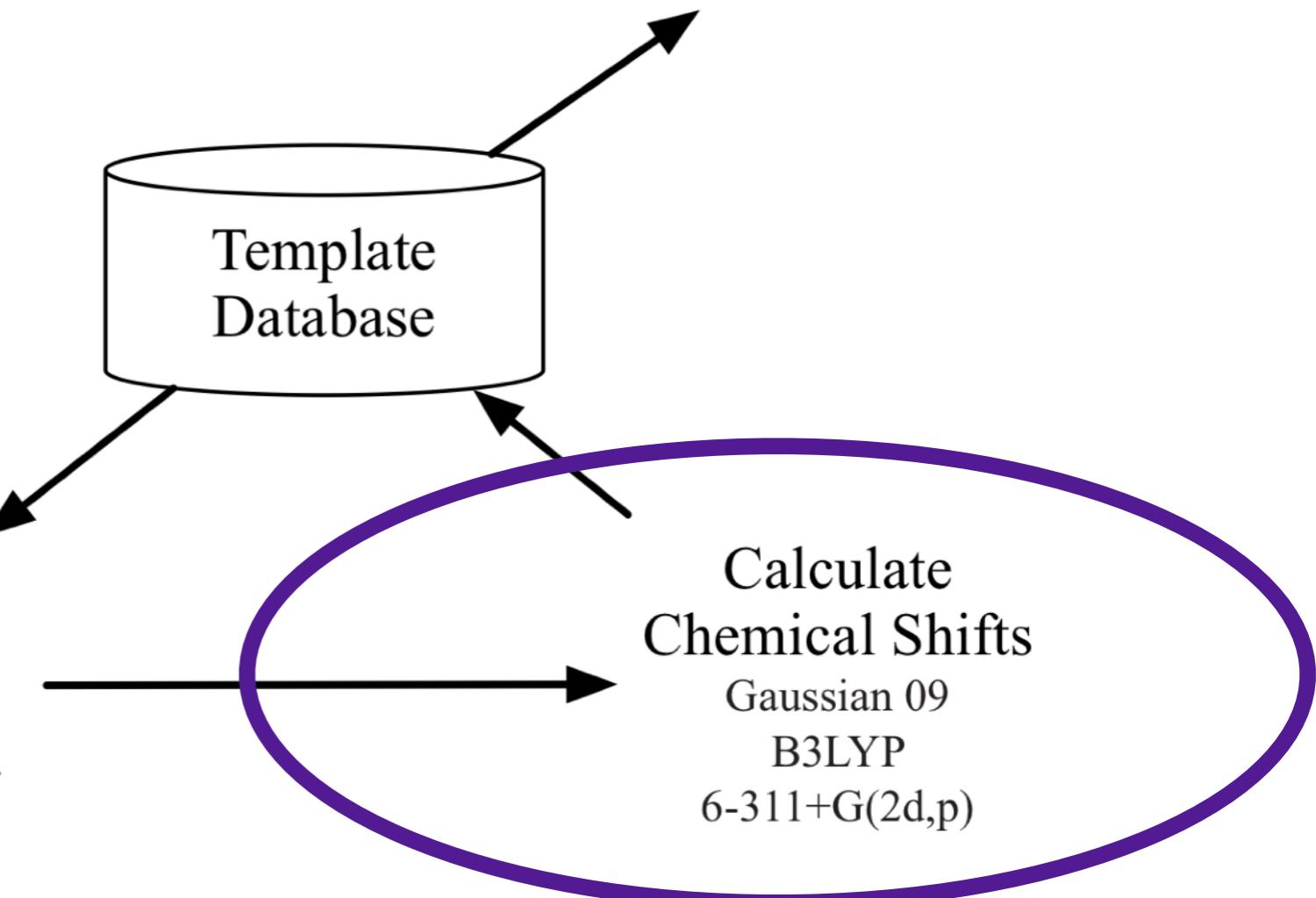
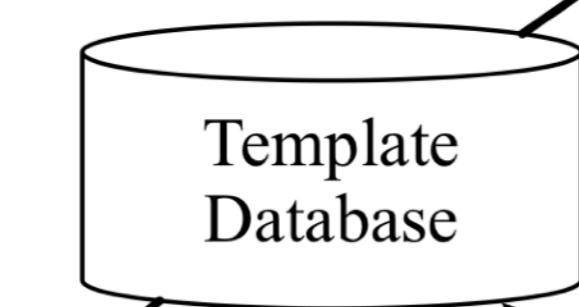
Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



Extract Regions of
Interest (ROIs)
5Å Radius, Complete Residues

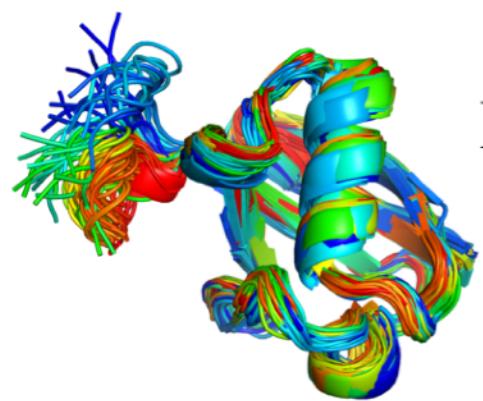


Select Novel
Representative
ROIs
0.2Å Matching Cutoff

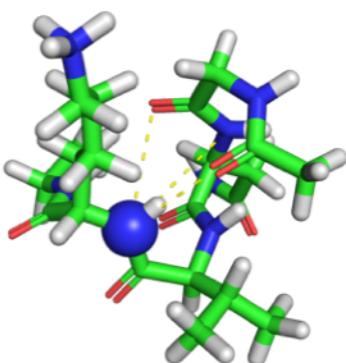


Ensemble
Average

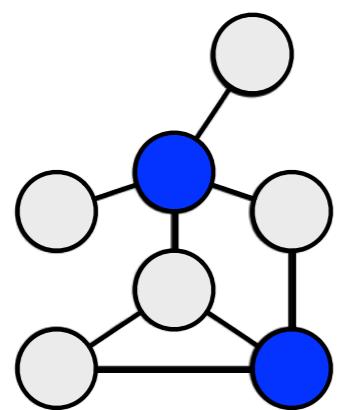
Match Closest
Template
Minimum Average
Absolute Difference



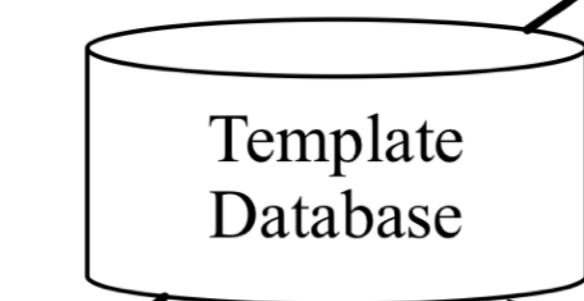
Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



Extract Regions of
Interest (ROIs)
5Å Radius, Complete Residues

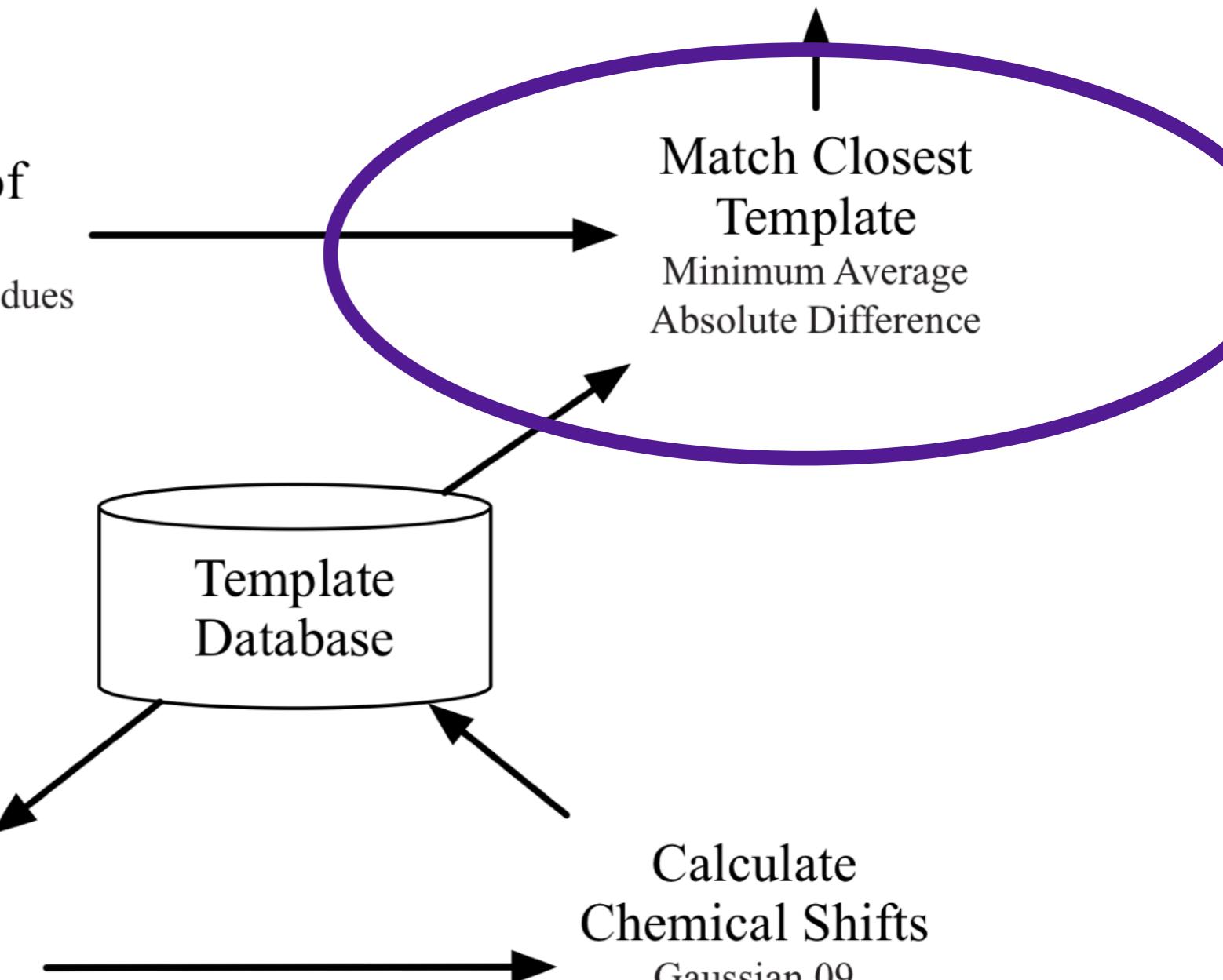


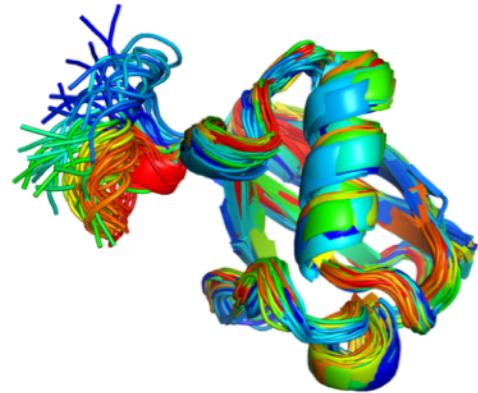
Select Novel
Representative
ROIs
0.2Å Matching Cutoff



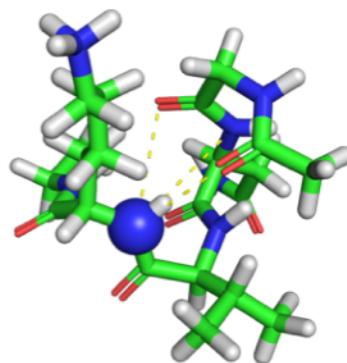
Ensemble
Average
Match Closest
Template
Minimum Average
Absolute Difference

Calculate
Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

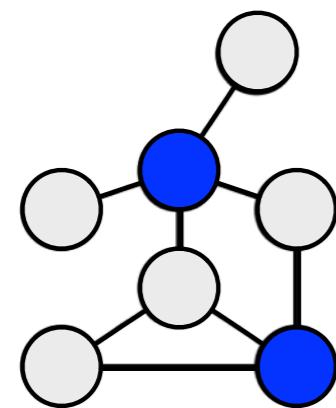




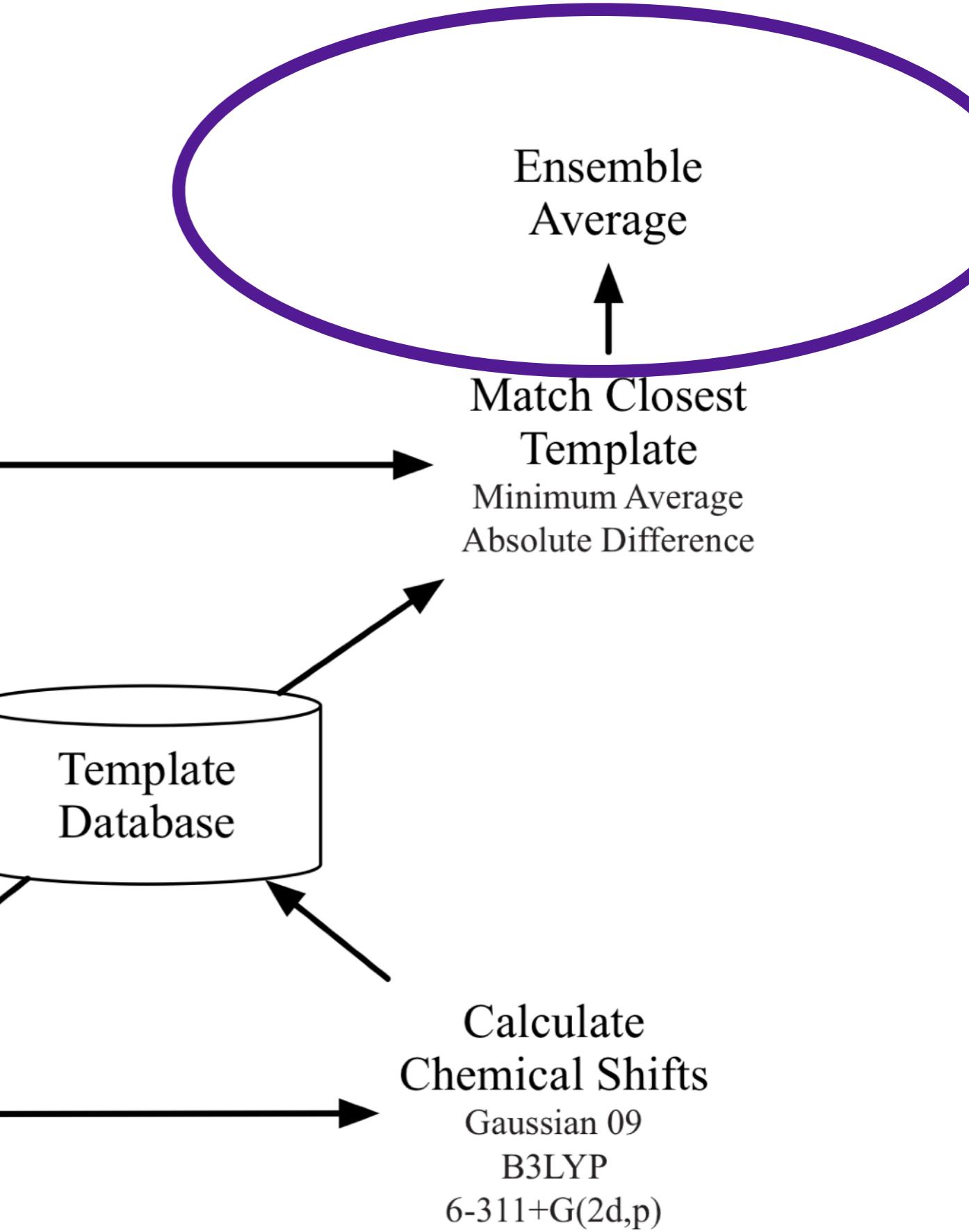
Molecular Dynamics
Amber 14, TIP3P water
1ps sampling



Extract Regions of
Interest (ROIs)
5 Å Radius, Complete Residues



Select Novel
Representative
ROIs
0.2 Å Matching Cutoff



Sources of Error

Level of Quantum Theory
Size of ROI

Reduction to Pattern
Closeness of Match

Sampling of MD
Accuracy of MD

Molecular Dynamics Evaluation

Sampling

- 30 x 10ns simulations
- independent equilibration, random velocities

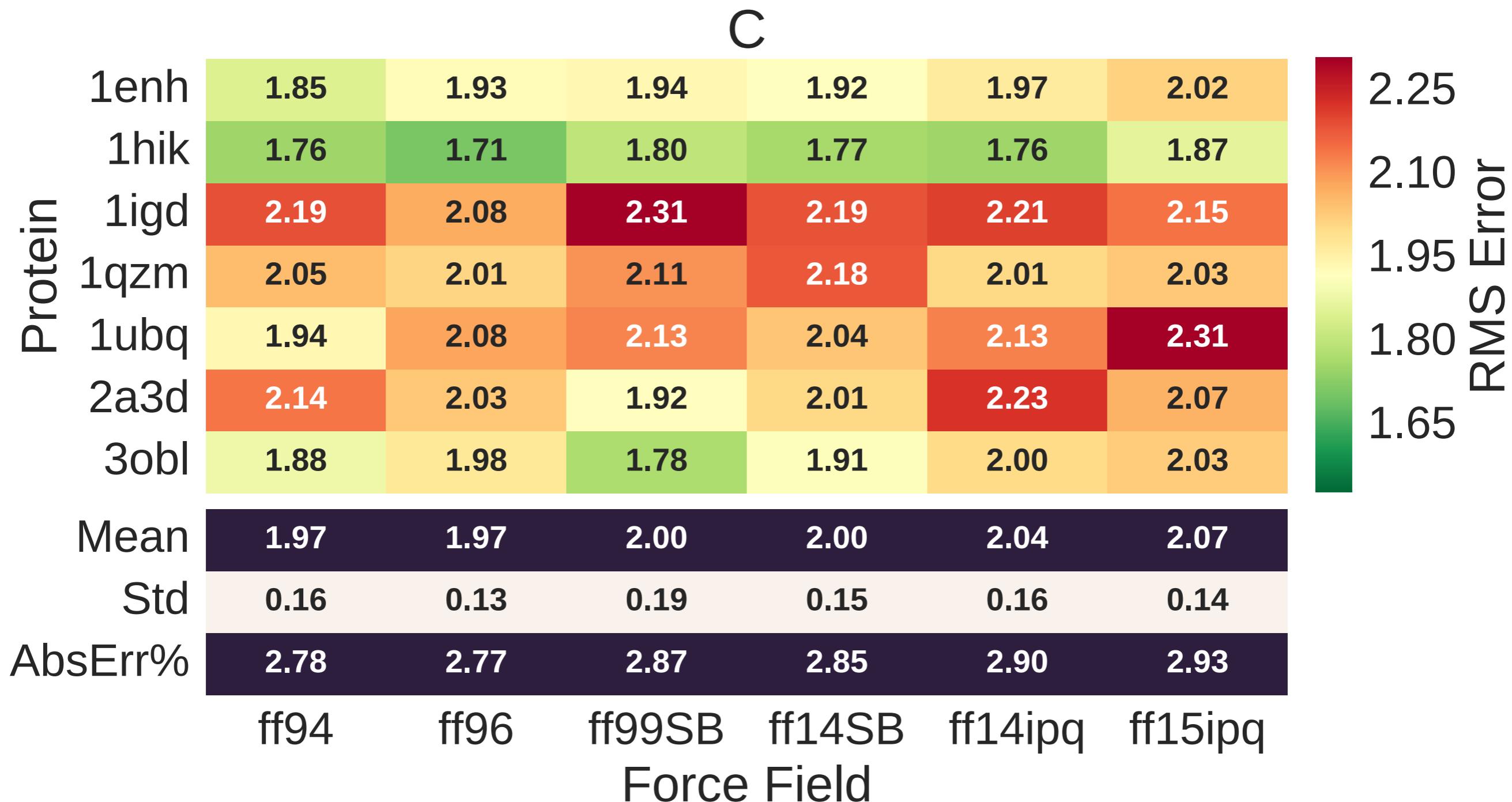
Amber Force Fields

- ff94
- ff96
- ff99SB
- ff14SB
- ff14ipq
- ff15ipq

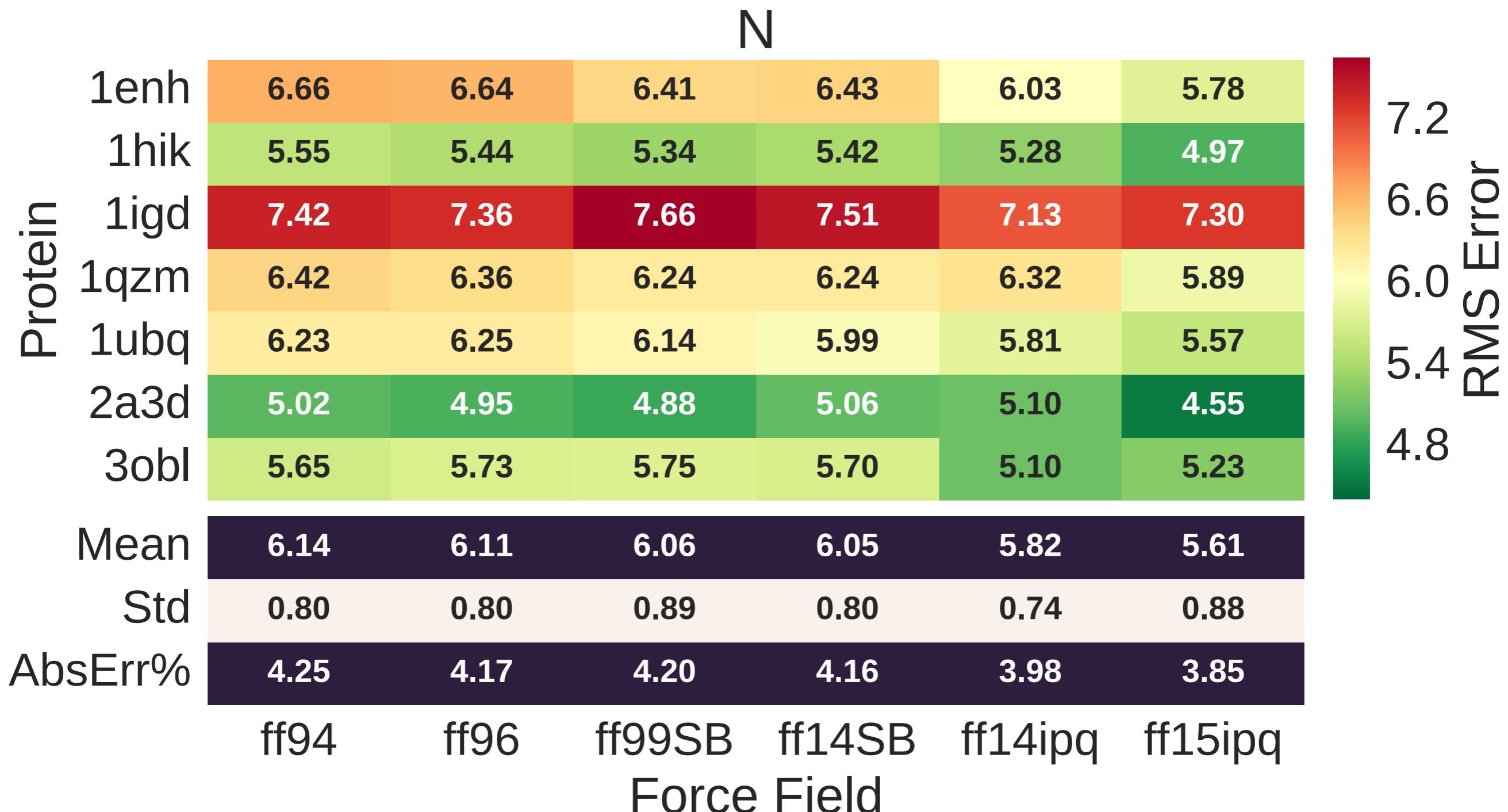
Proteins (BMRB)

1enh	2a3d
1ubq	3obl
1qzm	1hik
1igd	1l2y

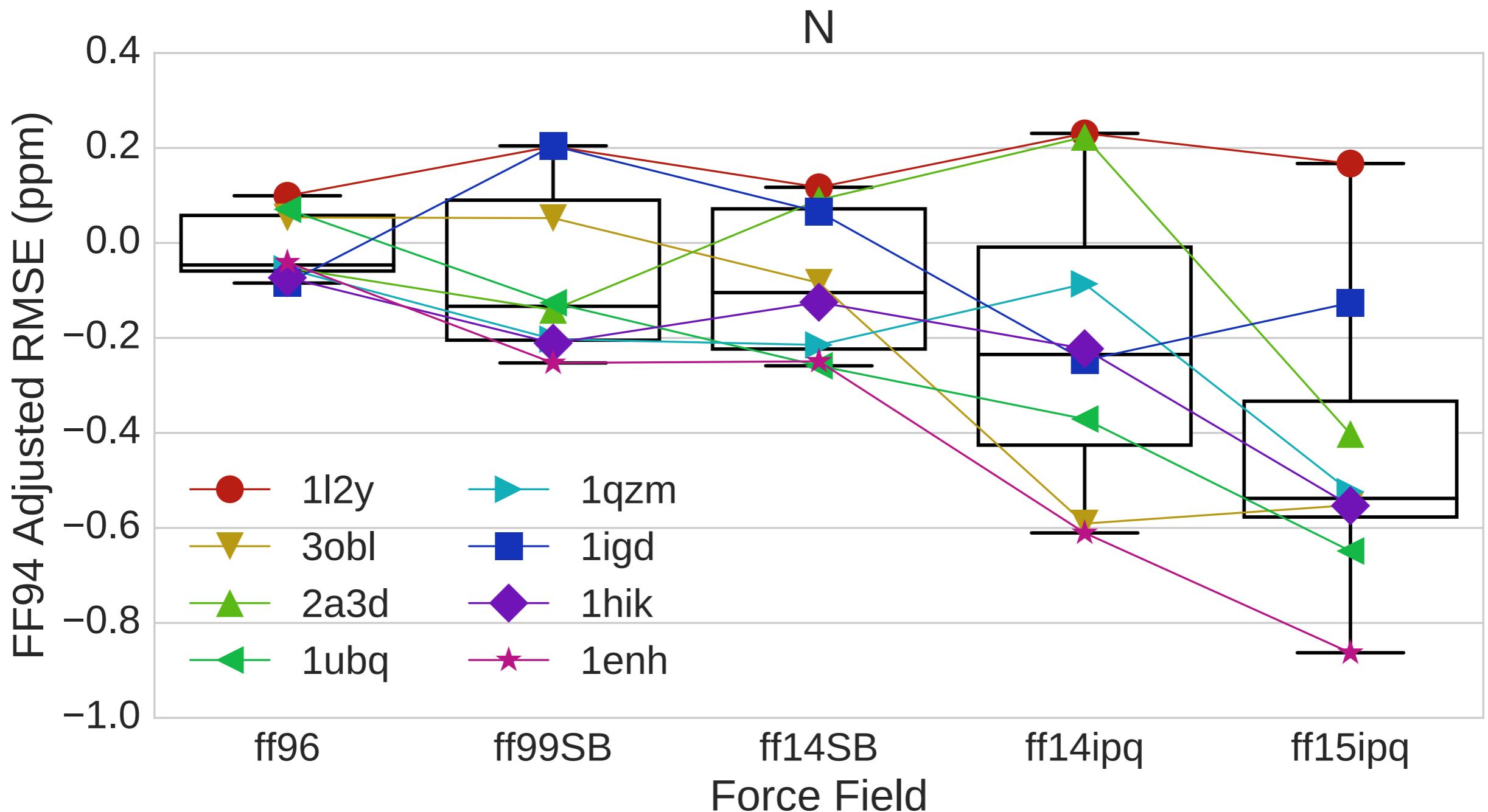
$^{13}\text{C}^\alpha$ RMS Error



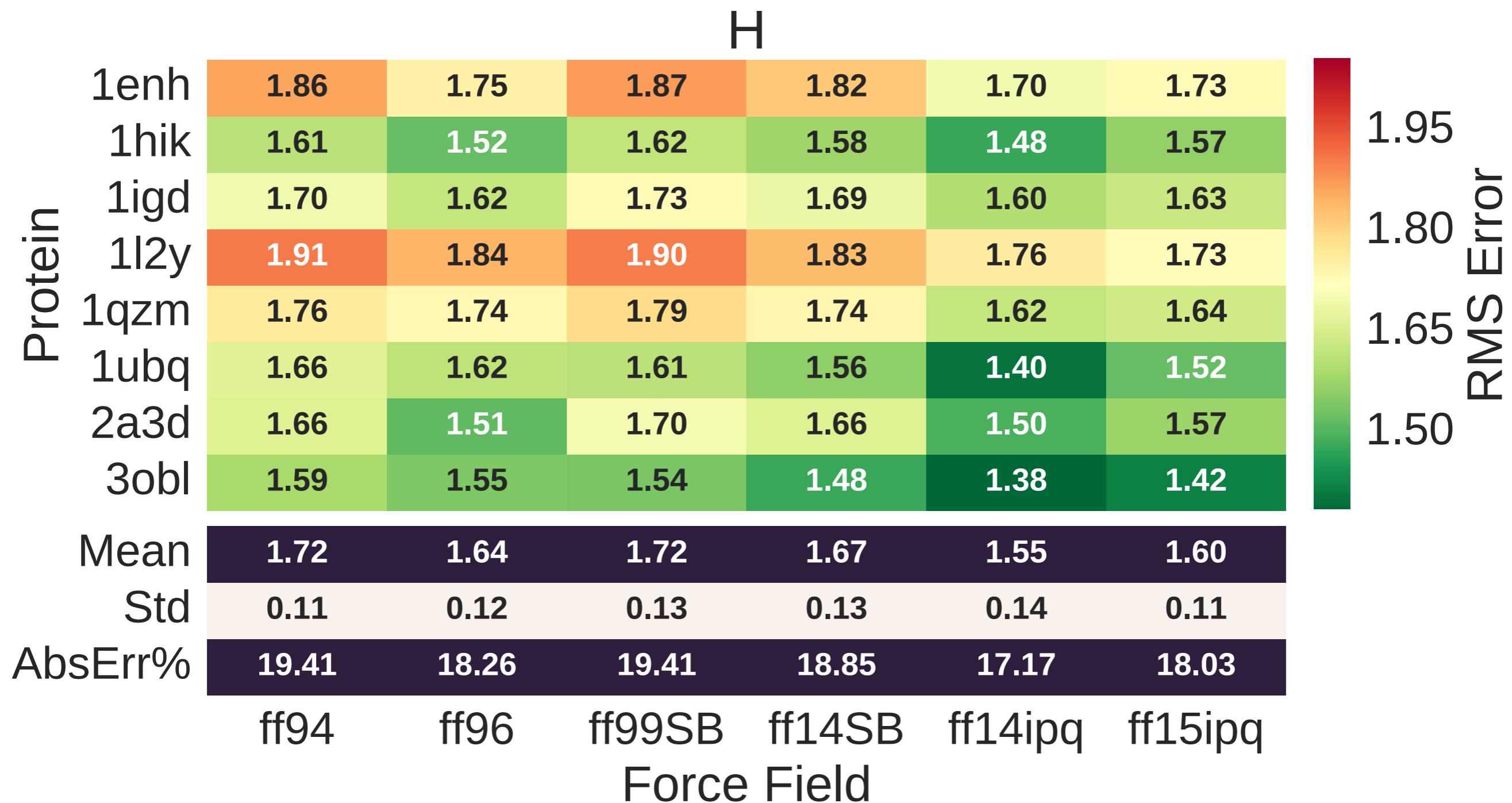
¹⁵N RMS Error



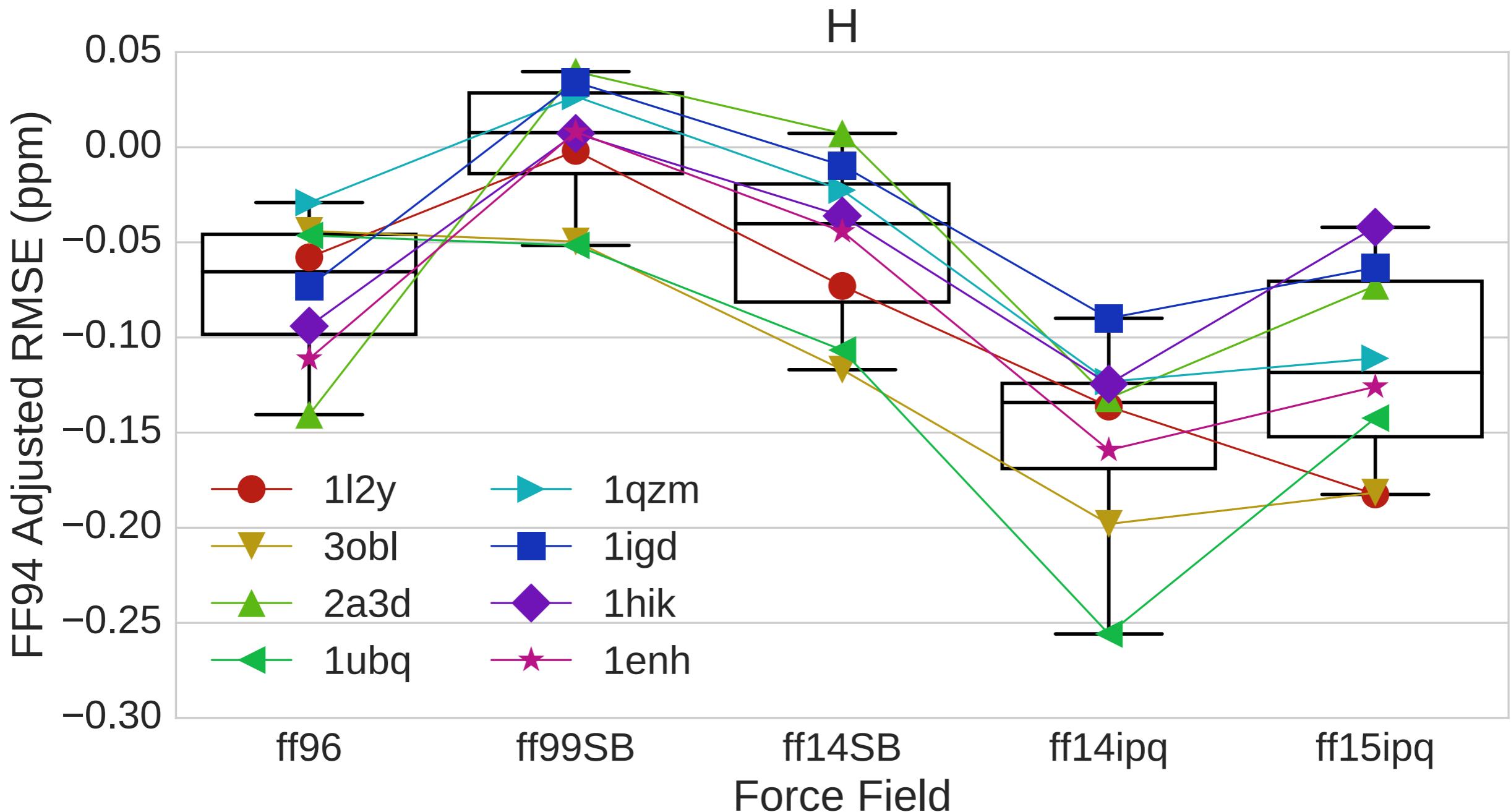
^{15}N RMS Error



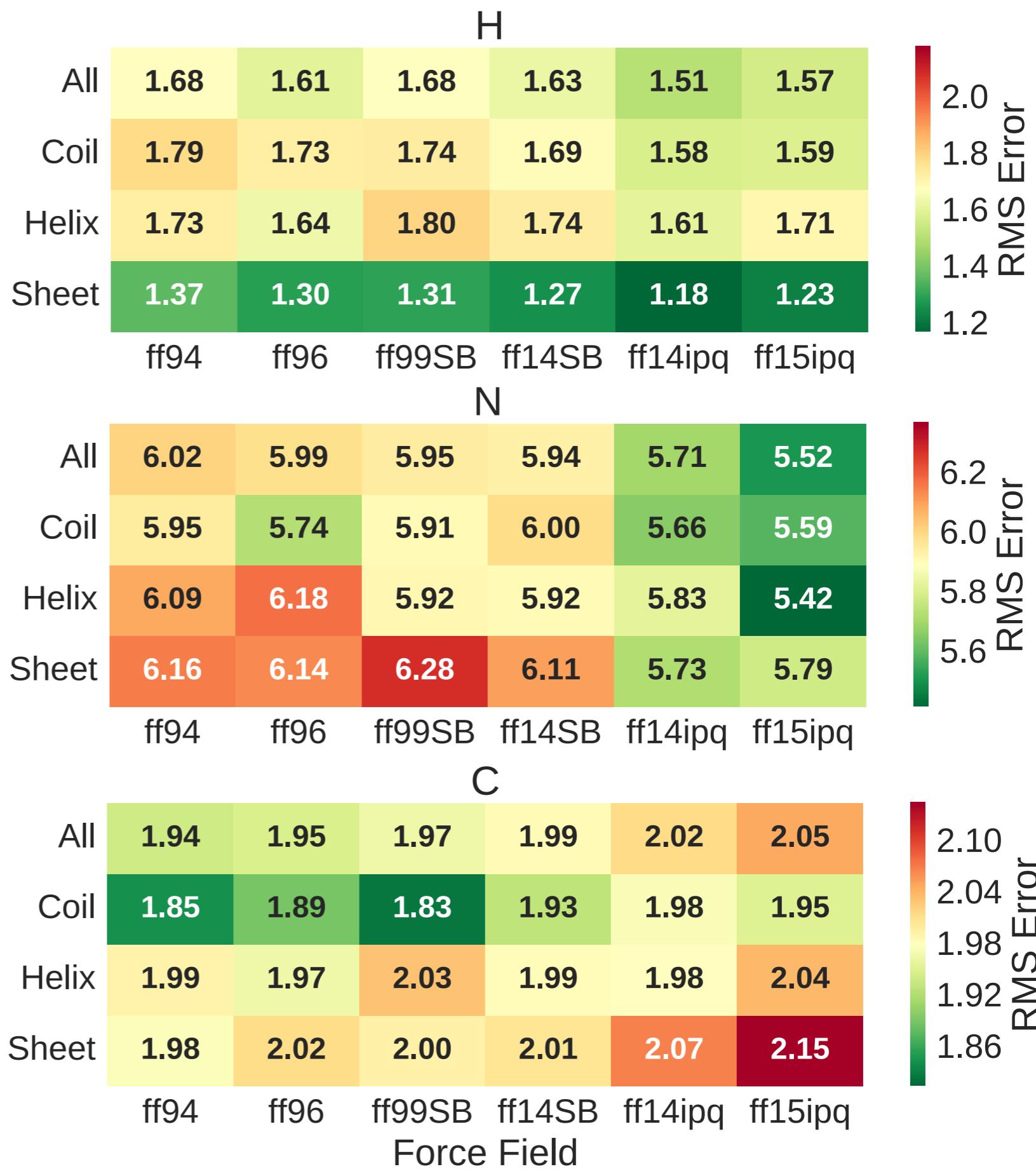
^1H RMS Error



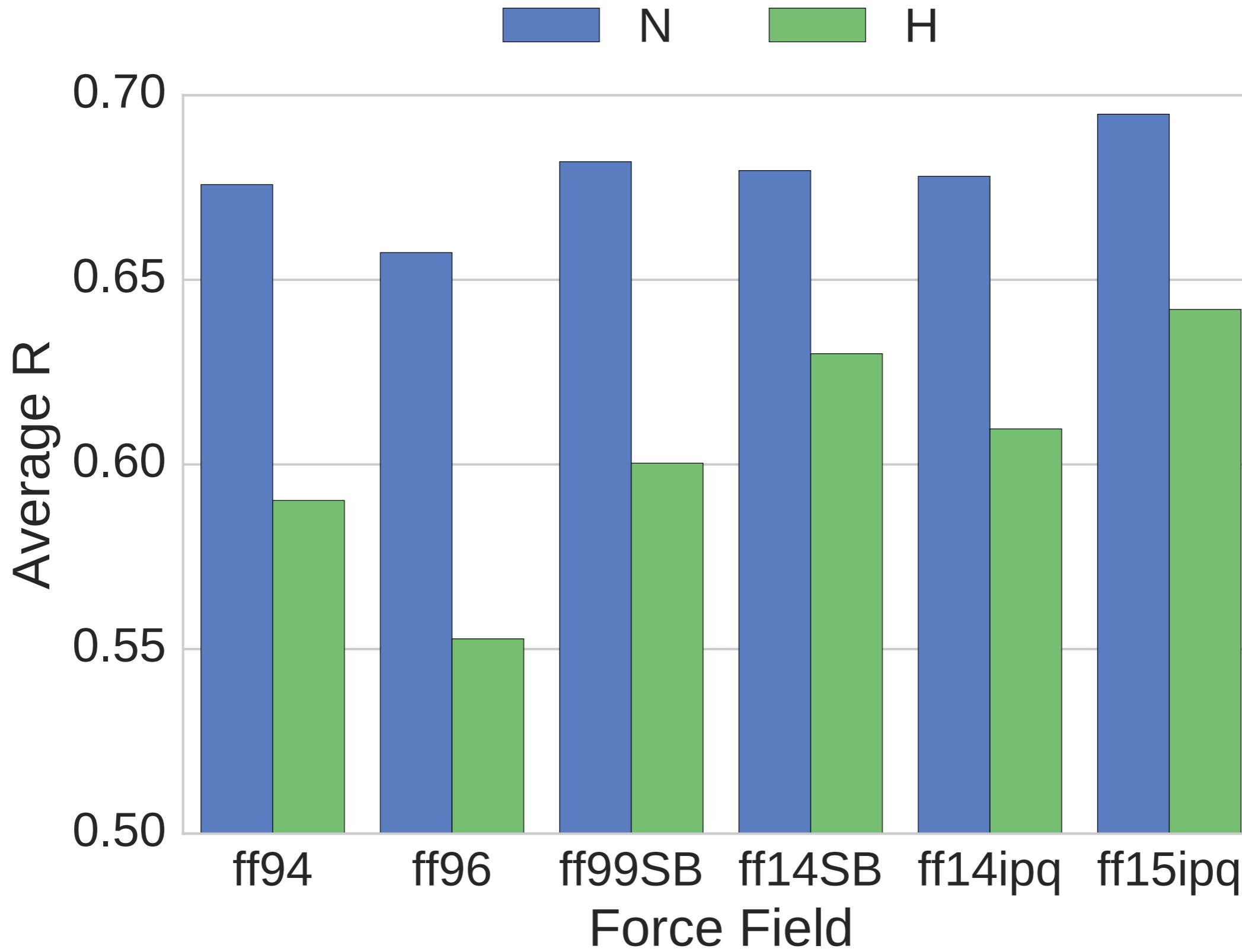
^1H RMS Error



Secondary Structure

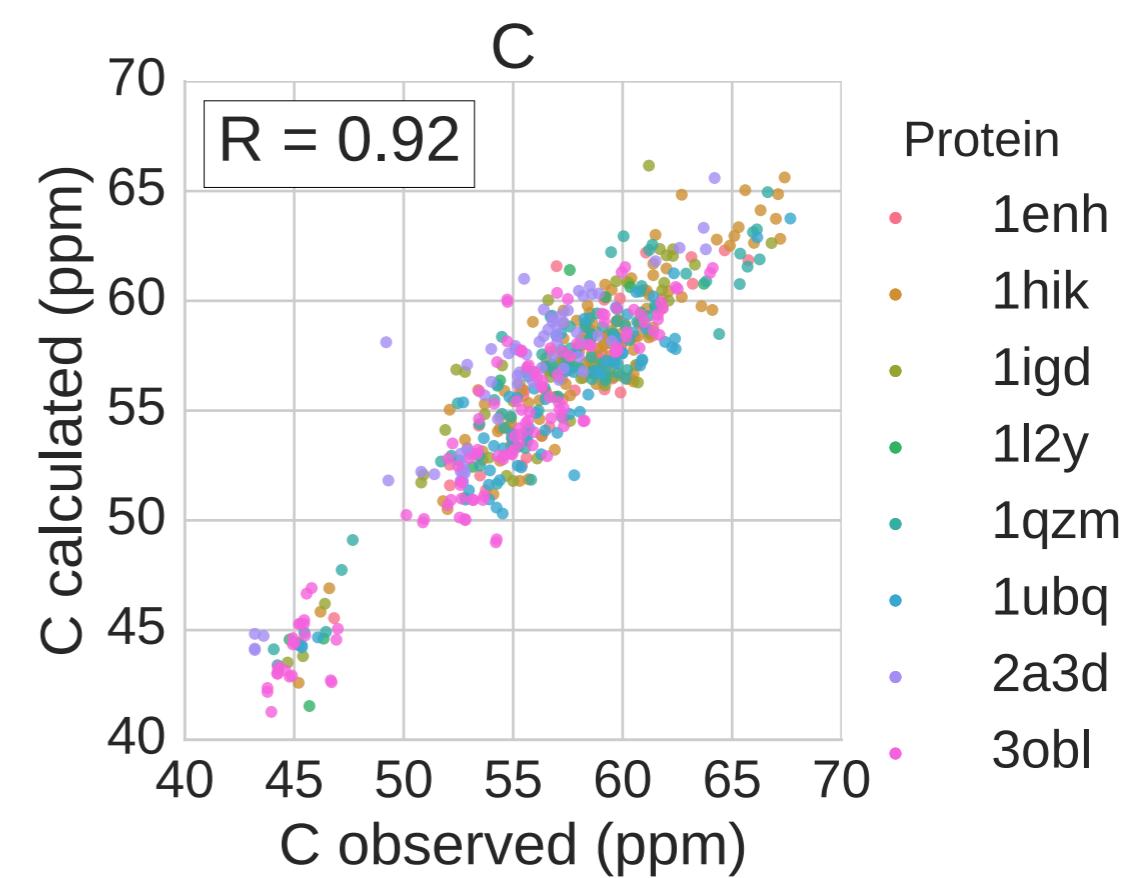
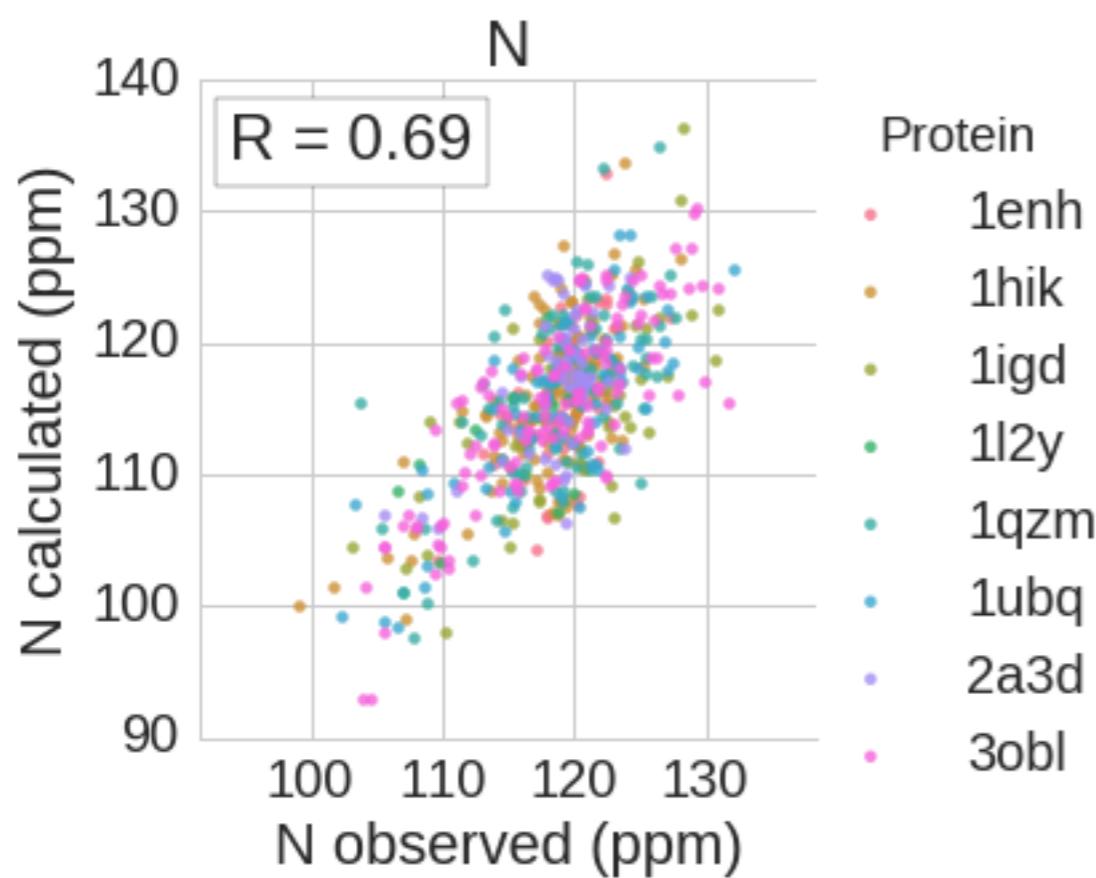
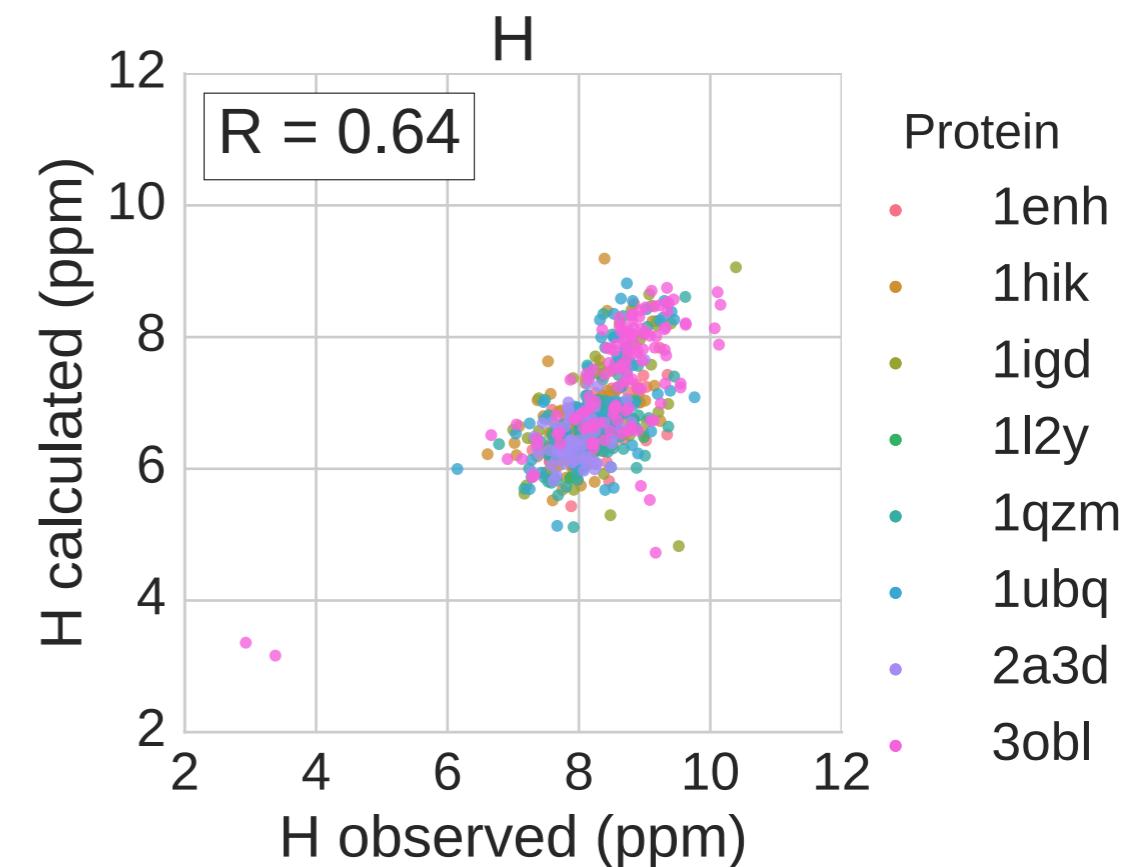


Correlation

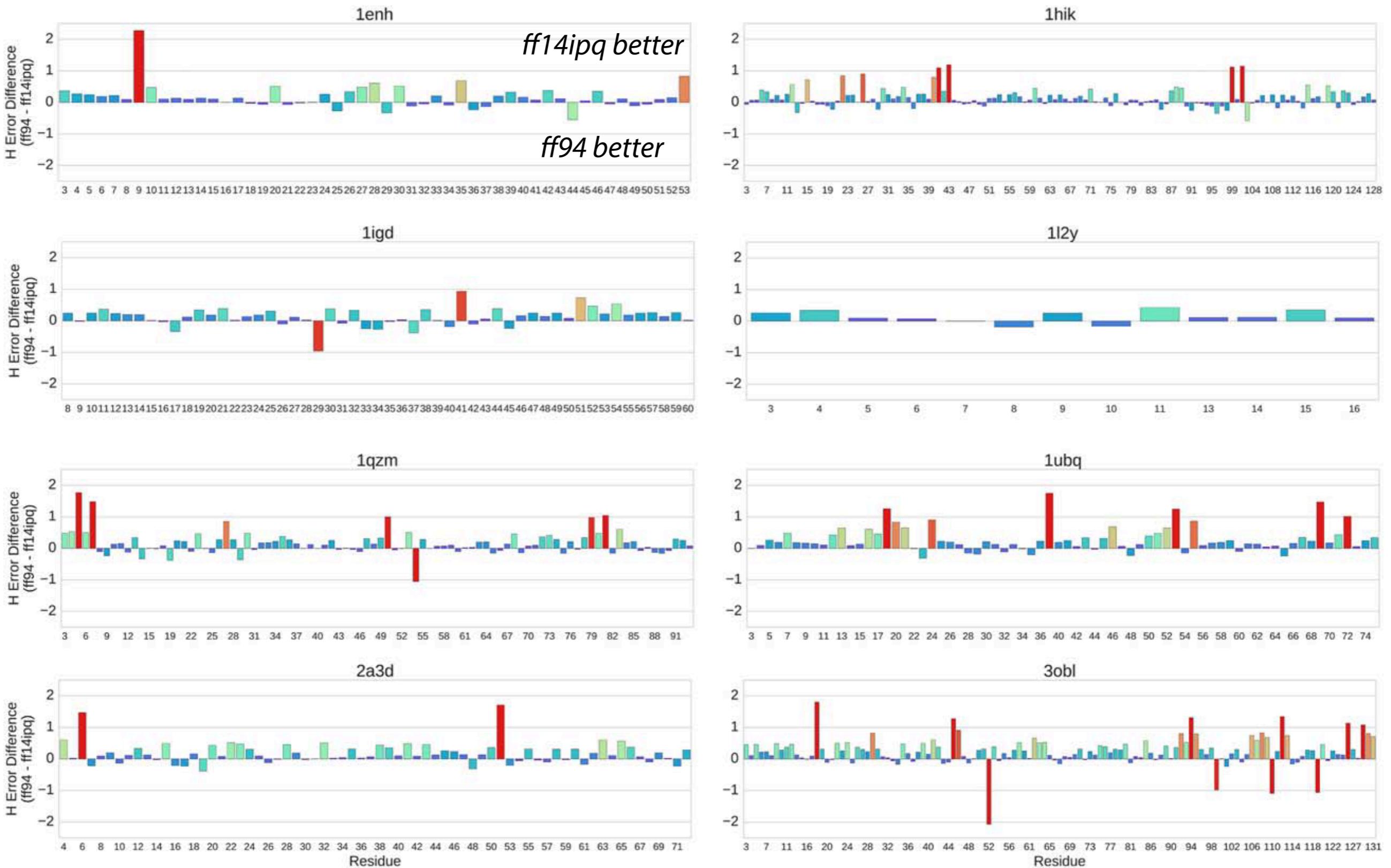


ff15ipq

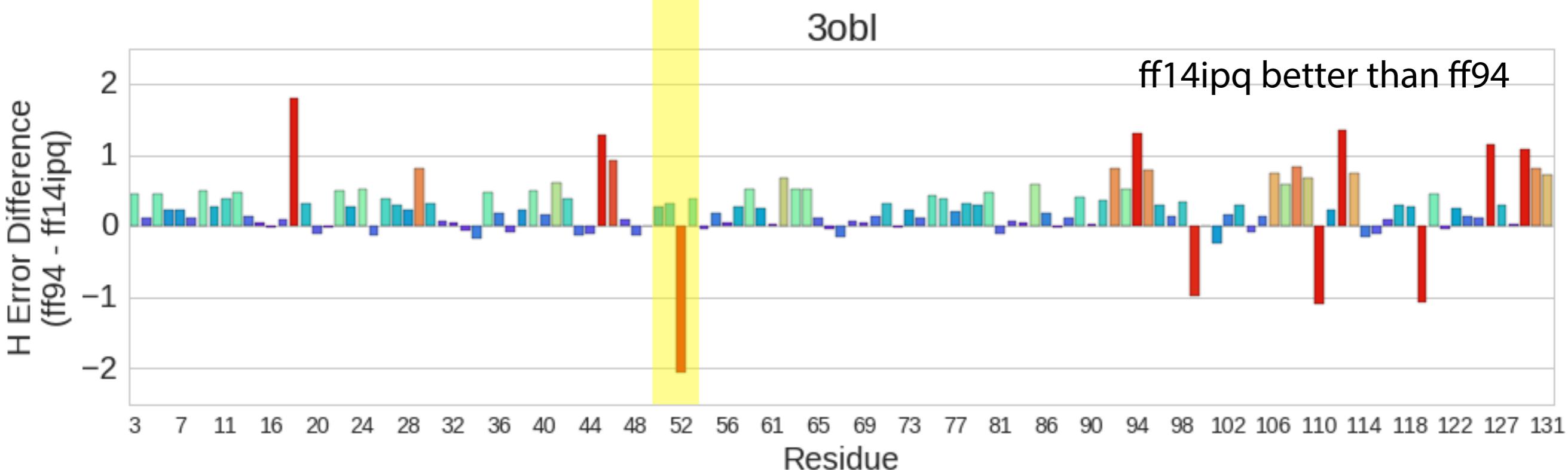
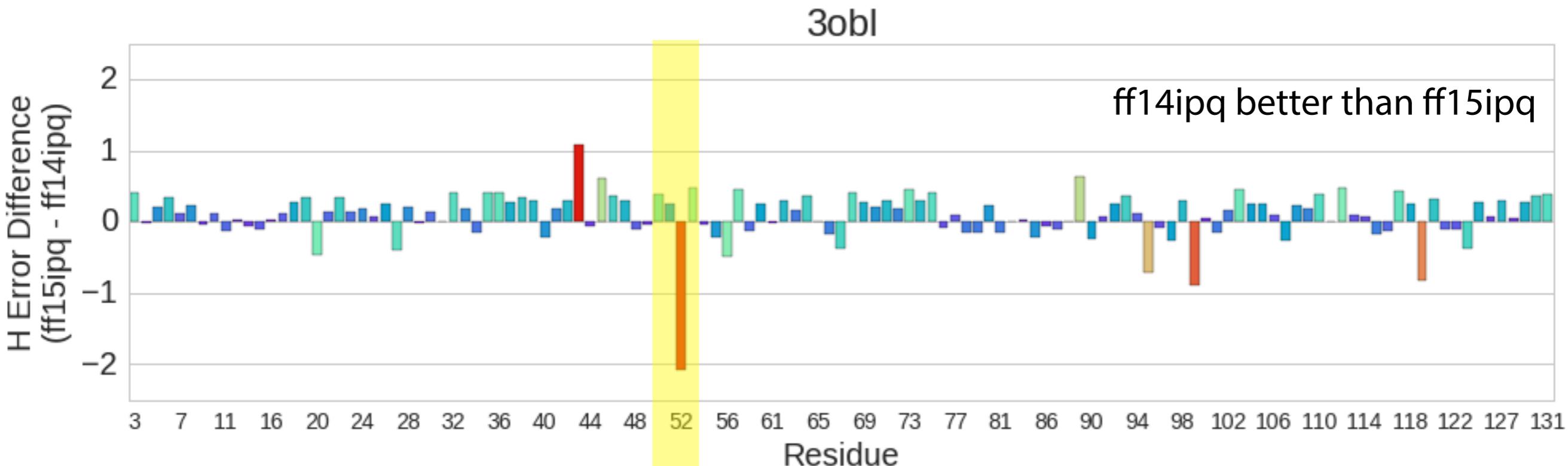
Correlation



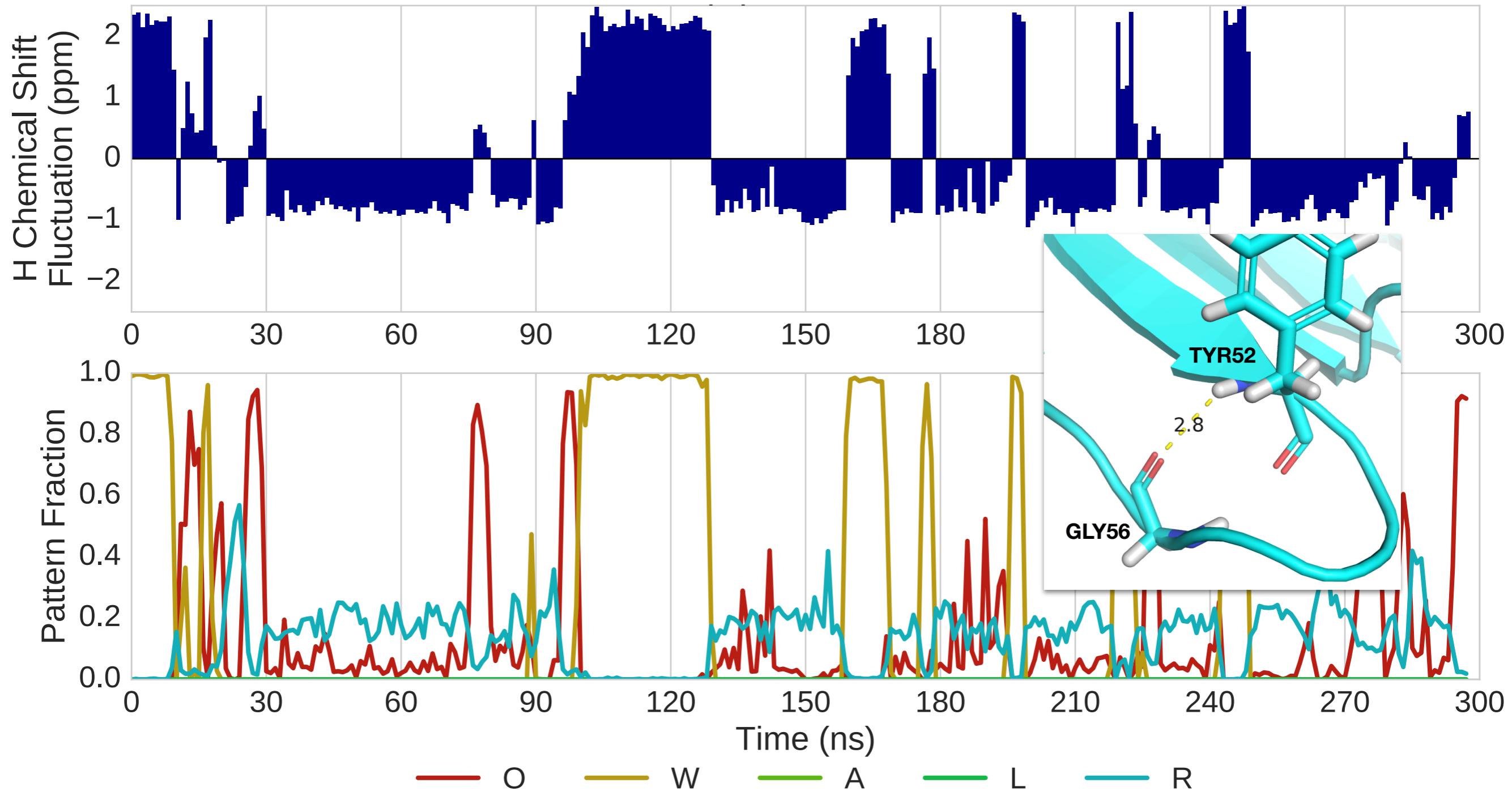
ff94 vs ff14ipq



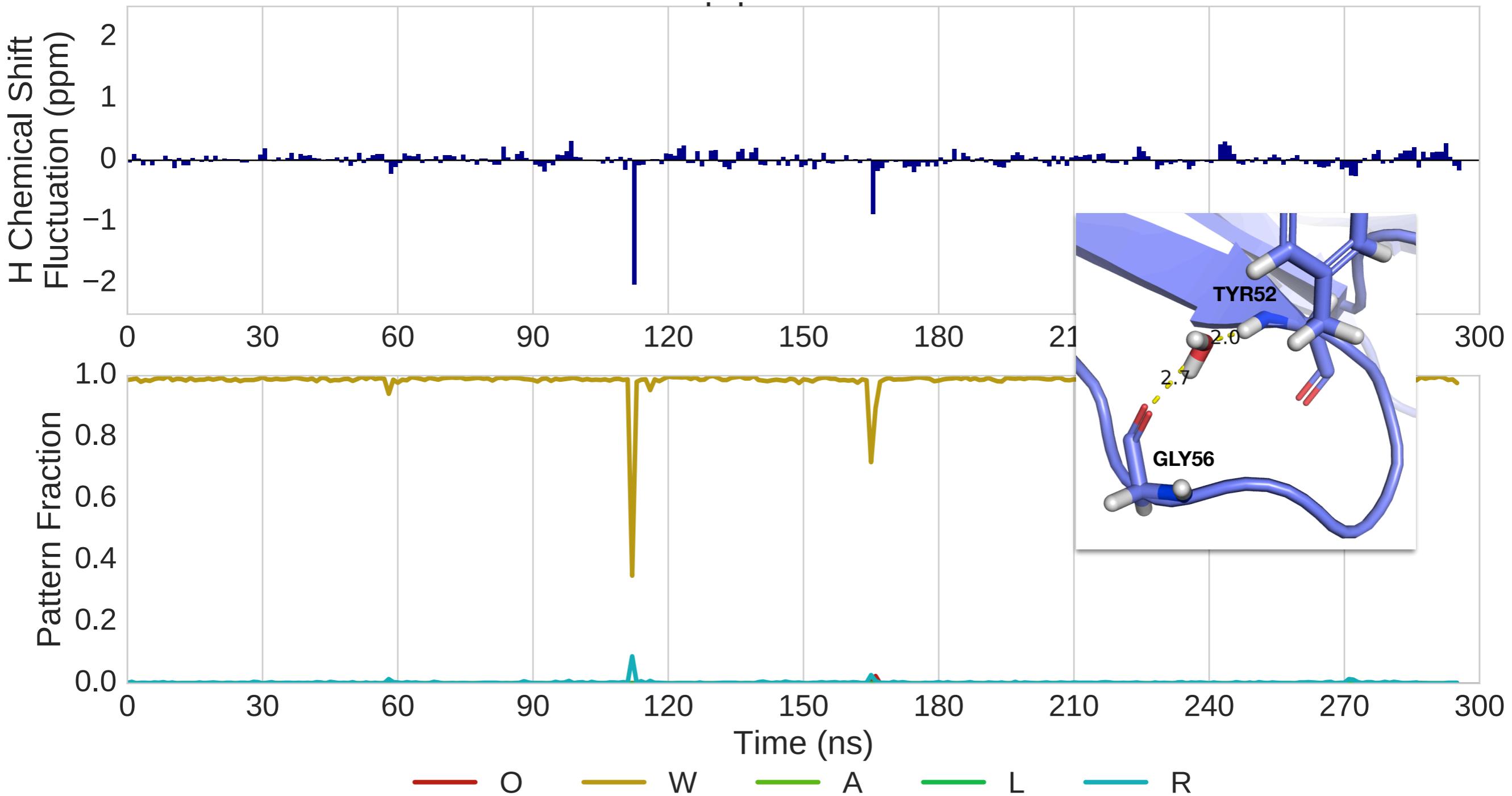
Drilling Down: 3OBL TYR52



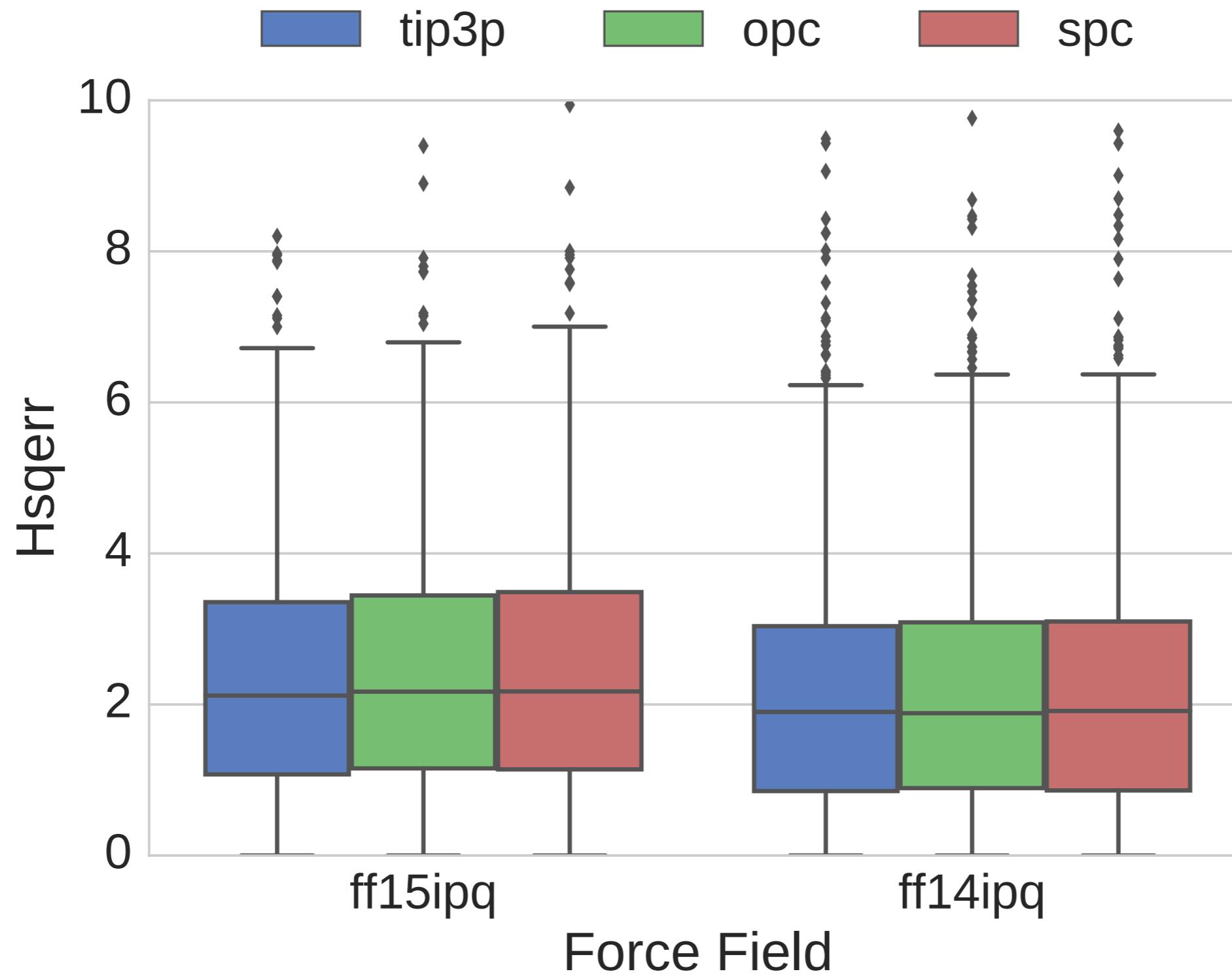
ff14ipq 3OBL:TYR52



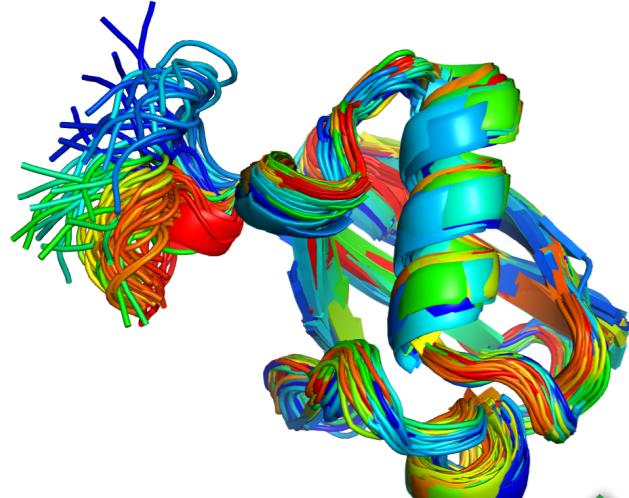
ff15ipq 3OBL:TYR52



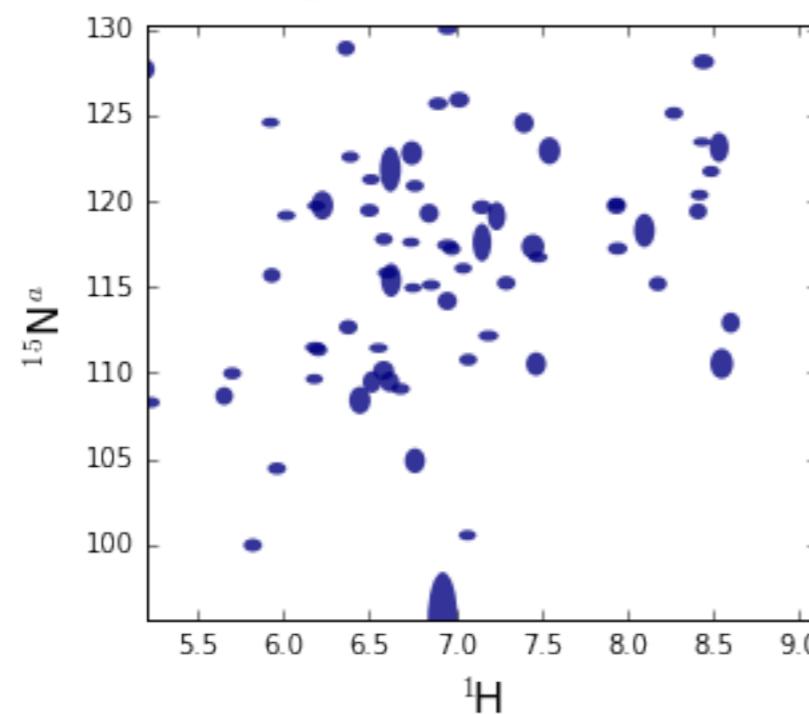
Water Models



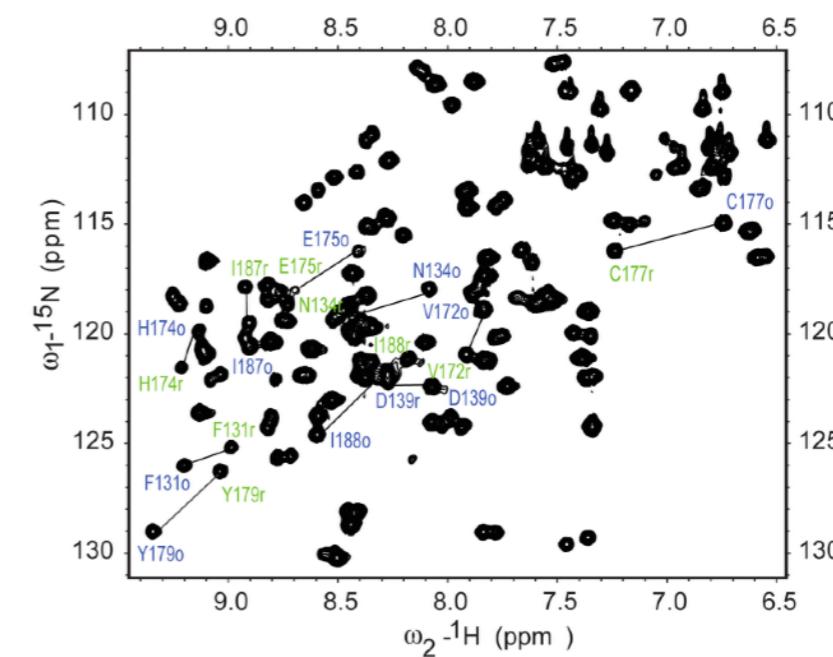
Conclusions



*According to our method
Force fields are improving
Use **ff15ipq**
TIP3P is okay*



?
==



DOI:10.1371/journal.ppat.1000960

Acknowledgements

John Vries



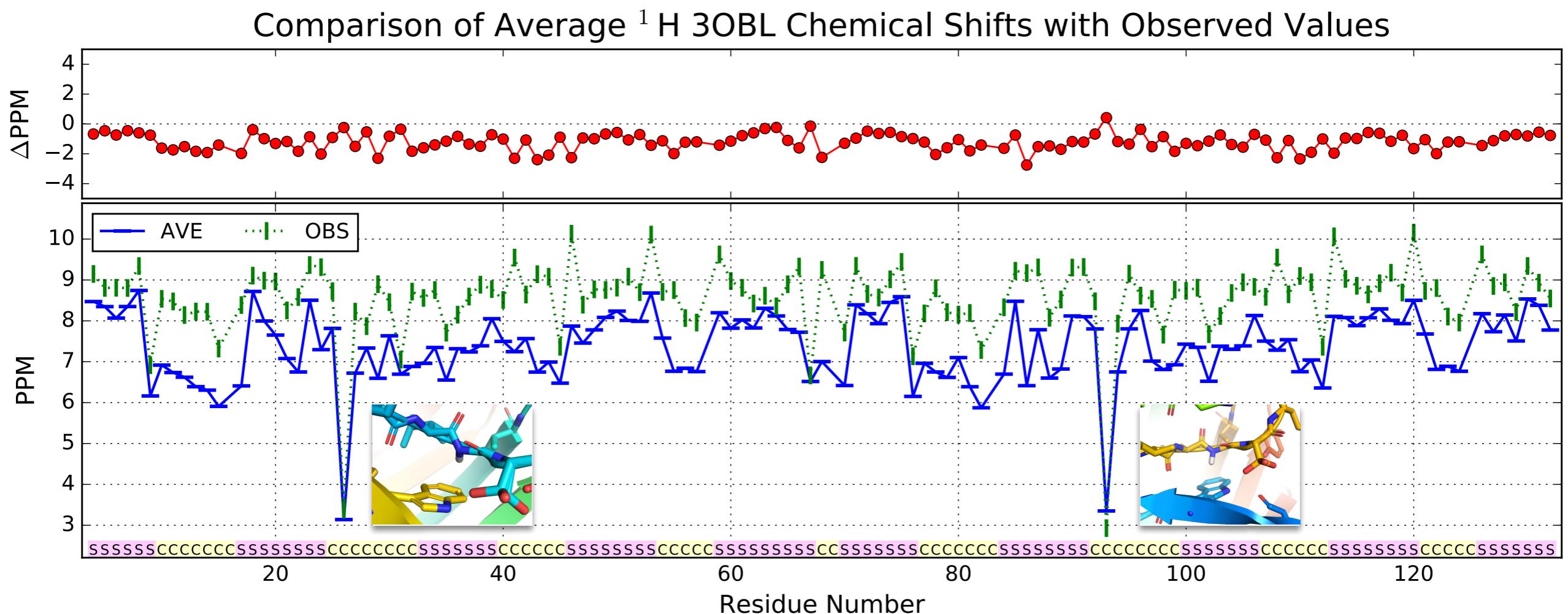
Department of
Computational and
Systems Biology
Center for Simulation and
Modeling

Angela Gronenborn
Leonard Koharudin
Antonio Ferreira
Karl Debiec
Krishna Mallupeddi
Steven DeMarco
Tejas Badgujar

Funding
R01GM108340

Questions?

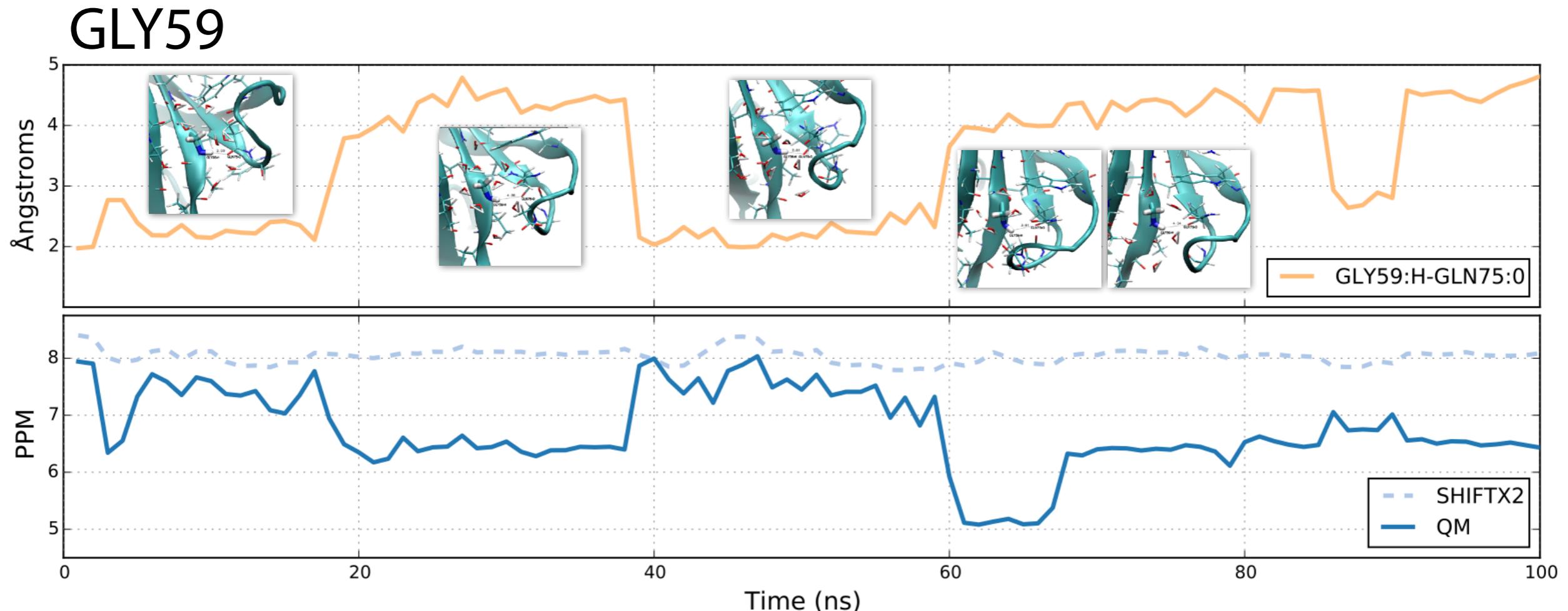
<https://github.com/dkoes/MD2NMR>



SHIFT X2 Comparison

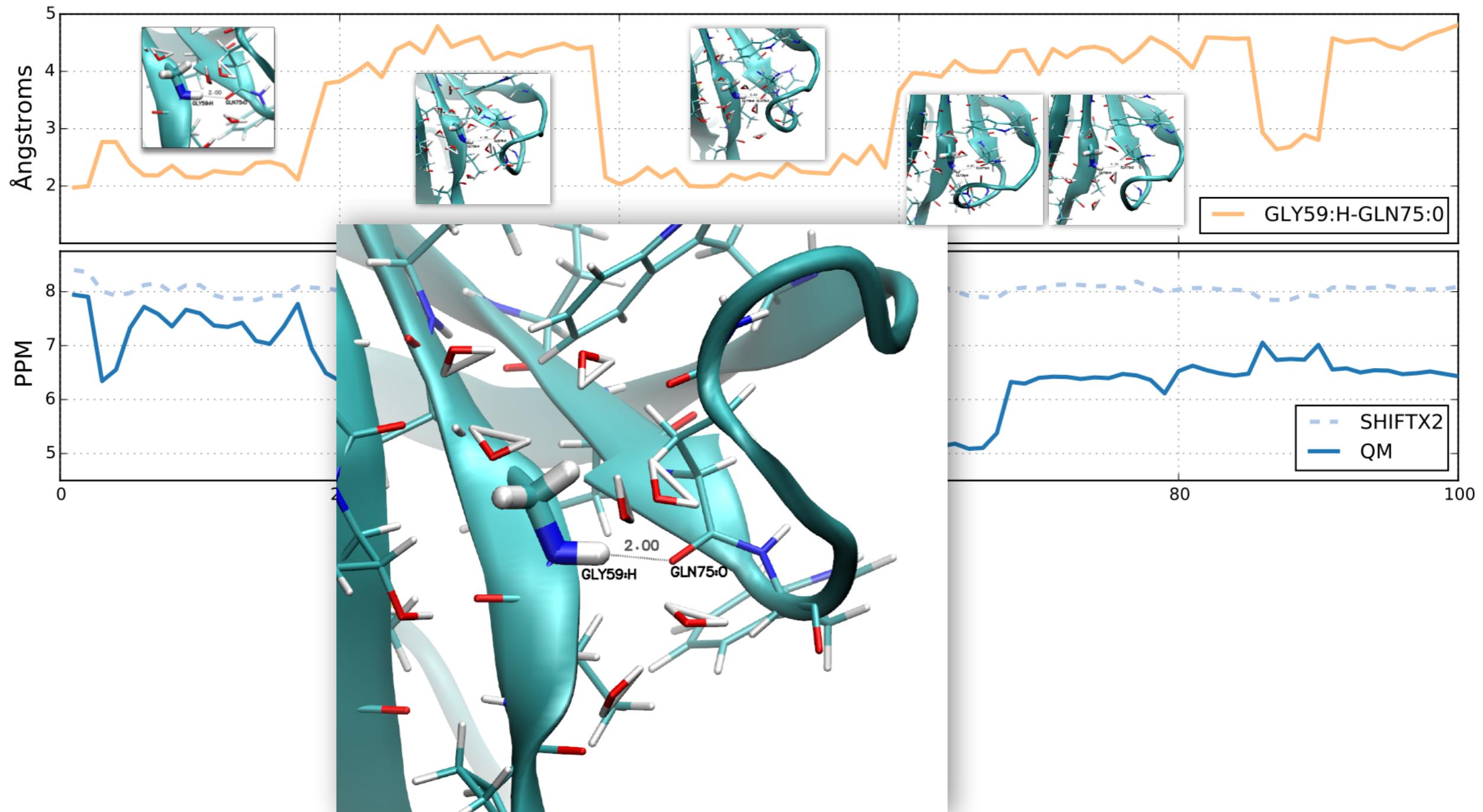
	ab initio						H	SHIFTX2						
	ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq		ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq	
ff94	1.000	0.949	0.965	0.951	0.875	0.878		1.000	0.962	0.973	0.970	0.951	0.949	ff94
ff96	0.949	1.000	0.943	0.927	0.851	0.848		0.962	1.000	0.969	0.967	0.944	0.943	ff96
ff99SB	0.965	0.943	1.000	0.956	0.888	0.902		0.973	0.969	1.000	0.980	0.955	0.964	ff99SB
ff14SB	0.951	0.927	0.956	1.000	0.916	0.932		0.970	0.967	0.980	1.000	0.966	0.969	ff14SB
ff14ipq	0.875	0.851	0.888	0.916	1.000	0.936		0.951	0.944	0.955	0.966	1.000	0.972	ff14ipq
ff15ipq	0.878	0.848	0.902	0.932	0.936	1.000		0.949	0.943	0.964	0.969	0.972	1.000	ff15ipq
	ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq		ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq	
	ab initio						Z	SHIFTX2						
ff94	1.000	0.964	0.977	0.968	0.949	0.957		1.000	0.982	0.992	0.986	0.975	0.980	ff94
ff96	0.964	1.000	0.961	0.950	0.938	0.938		0.982	1.000	0.982	0.975	0.972	0.971	ff96
ff99SB	0.977	0.961	1.000	0.980	0.954	0.967		0.992	0.982	1.000	0.988	0.975	0.983	ff99SB
ff14SB	0.968	0.950	0.980	1.000	0.963	0.975		0.986	0.975	0.988	1.000	0.981	0.988	ff14SB
ff14ipq	0.949	0.938	0.954	0.963	1.000	0.974		0.975	0.972	0.975	0.981	1.000	0.986	ff14ipq
ff15ipq	0.957	0.938	0.967	0.975	0.974	1.000		0.980	0.971	0.983	0.988	0.986	1.000	ff15ipq
	ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq		ff94	ff96	ff99SB	ff14SB	ff14ipq	ff15ipq	

3OBL (cyanobacterial lectin)

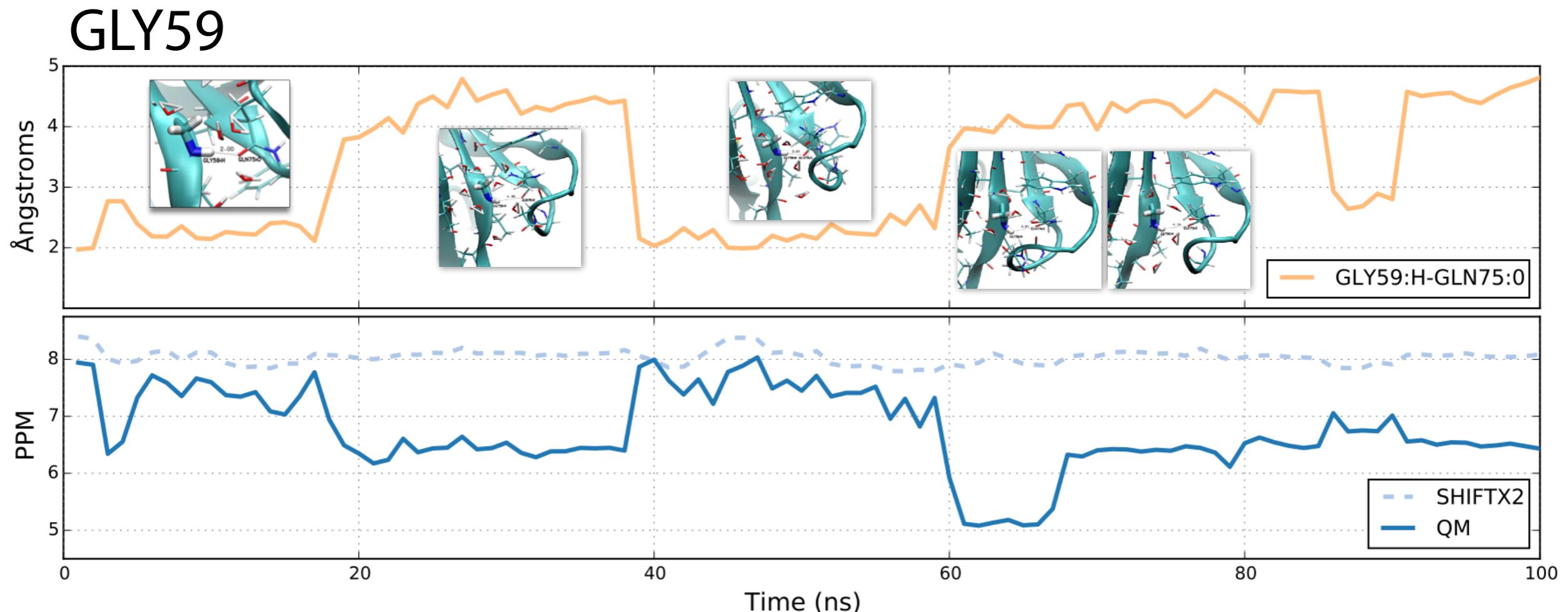


3OBL (cyanobacterial lectin)

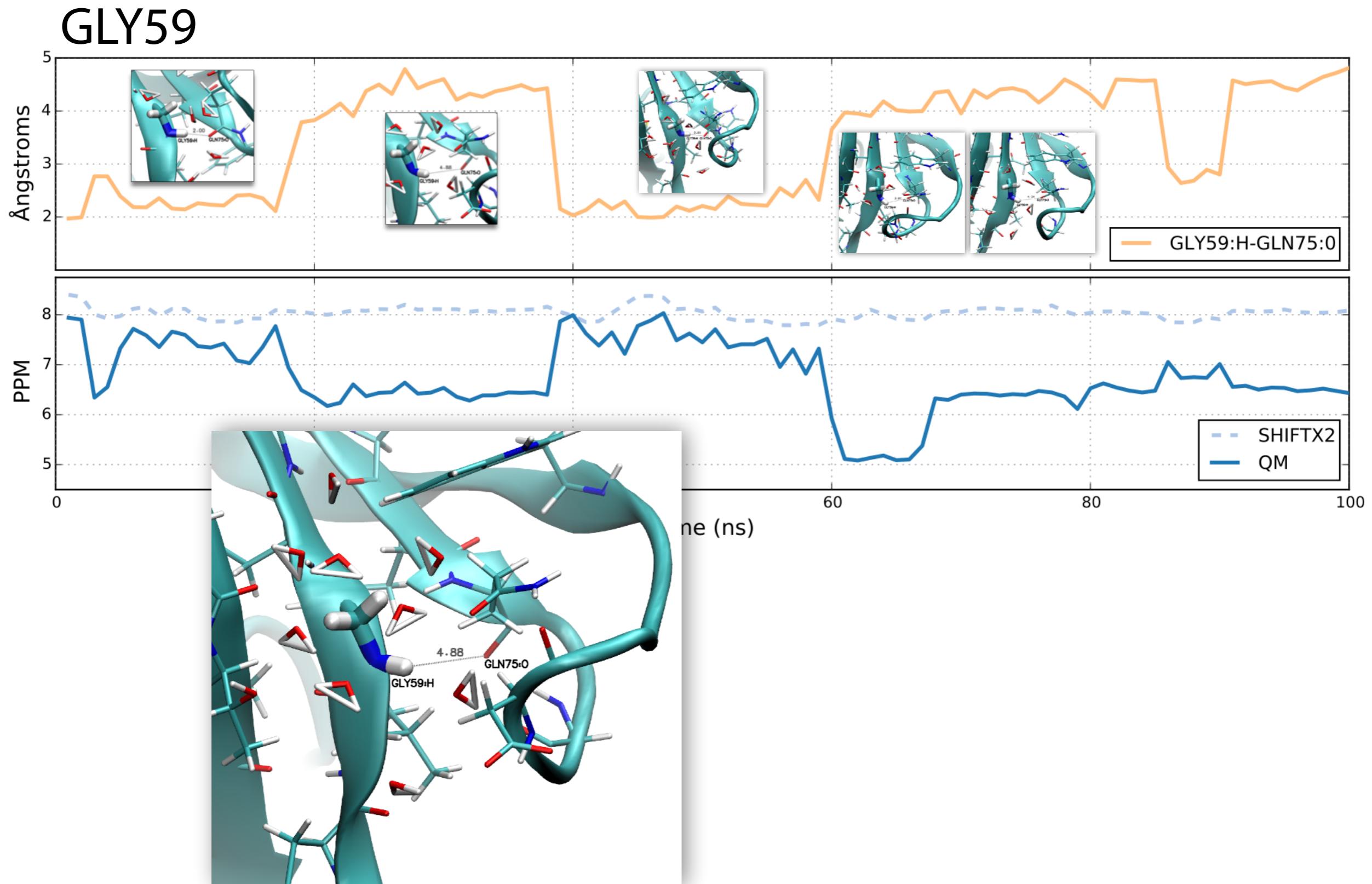
GLY59



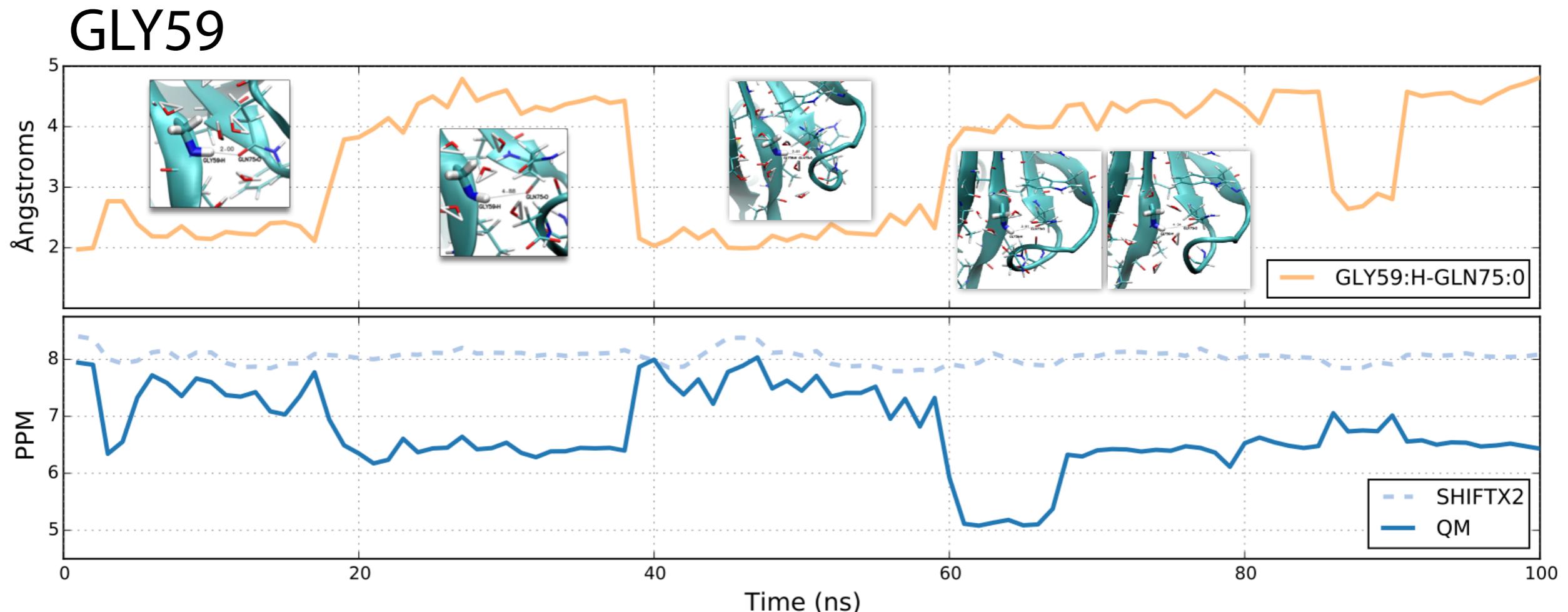
3OBL (cyanobacterial lectin)



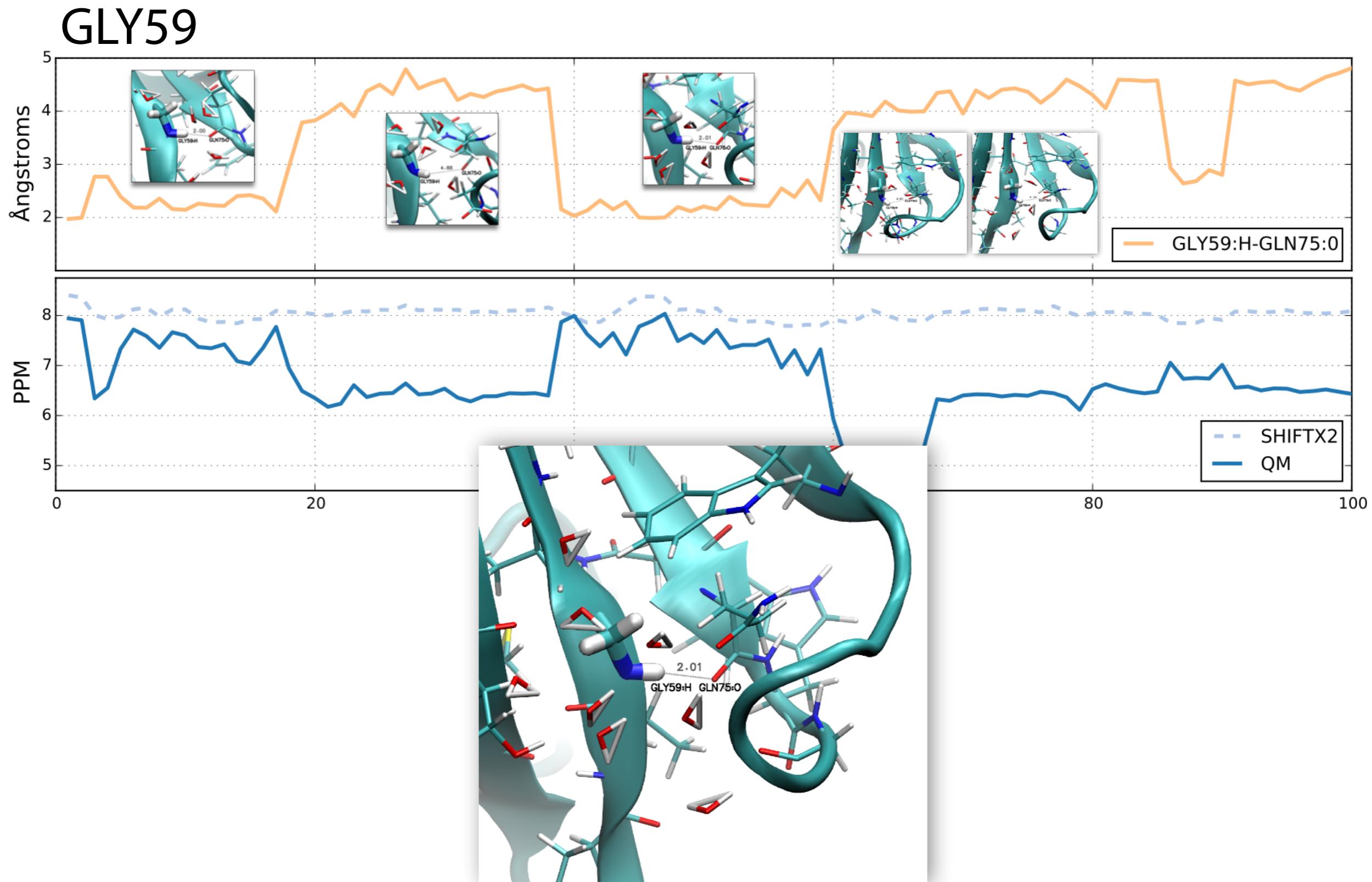
3OBL (cyanobacterial lectin)



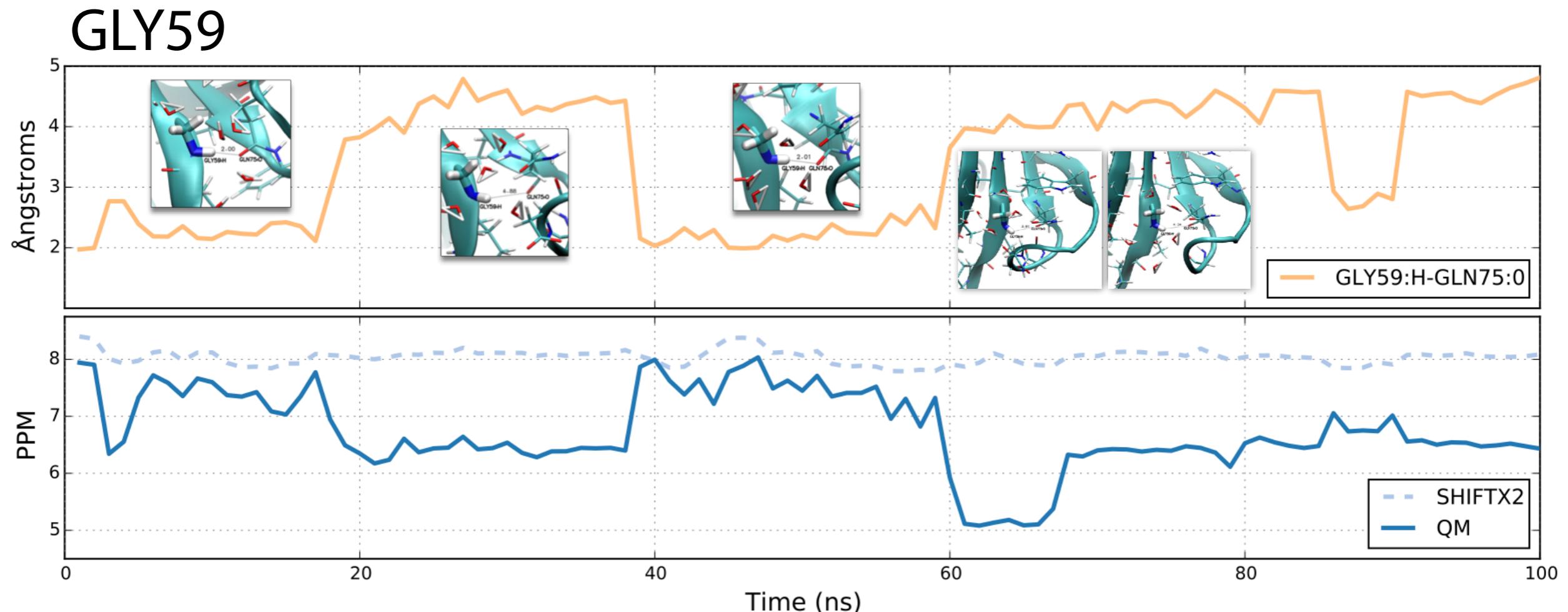
3OBL (cyanobacterial lectin)



3OBL (cyanobacterial lectin)

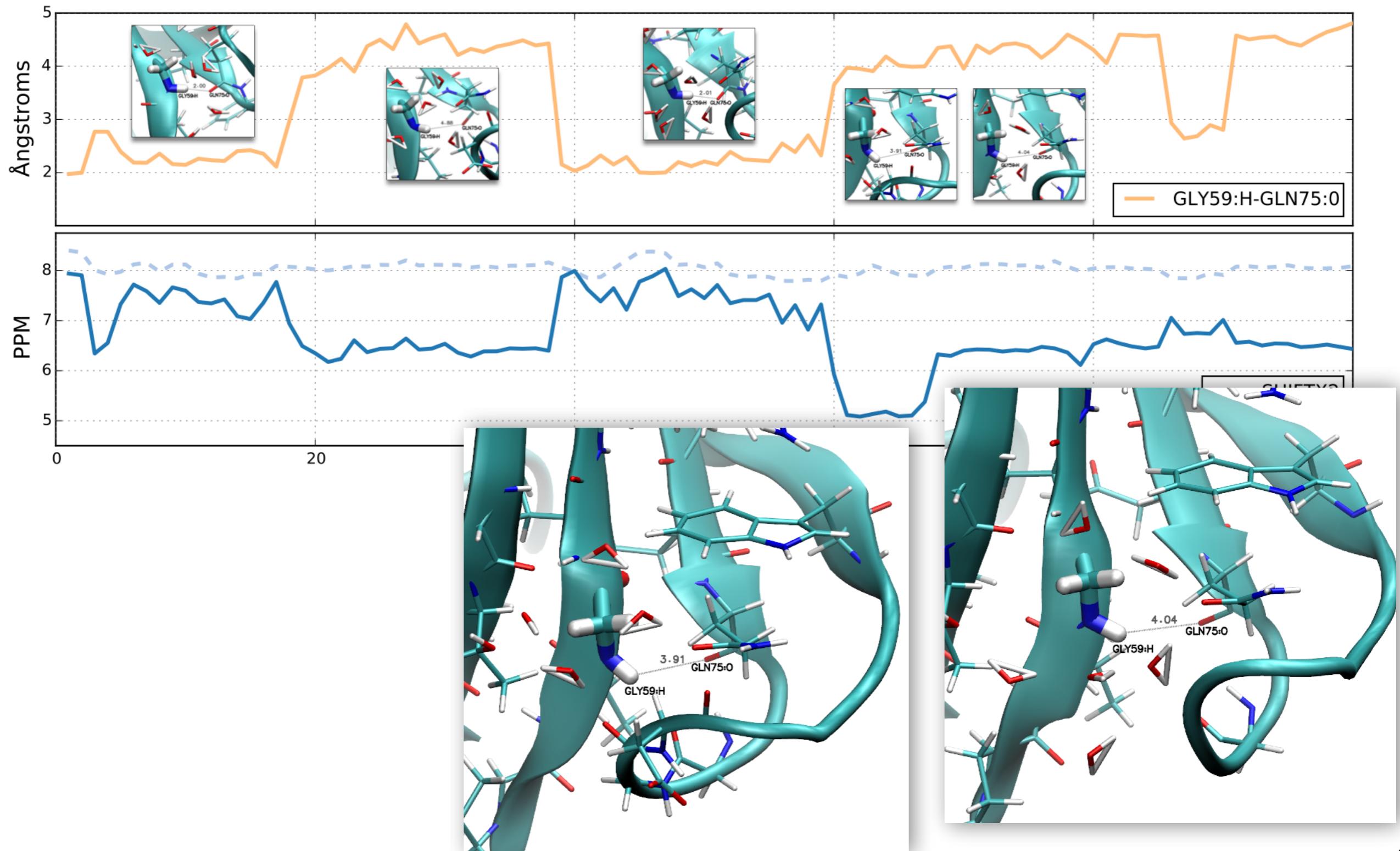


3OBL (cyanobacterial lectin)



3OBL (cyanobacterial lectin)

GLY59



3OBL (cyanobacterial lectin)

