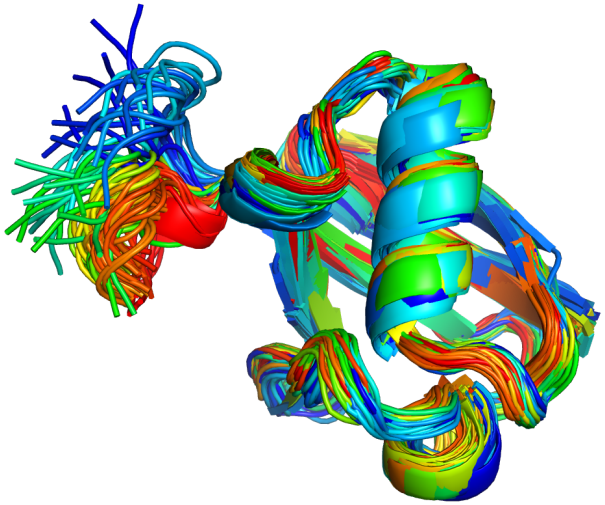
Quantum chemical approach for evaluating molecular mechanics force fields based on comparison of computed and observed NMR chemical shifts

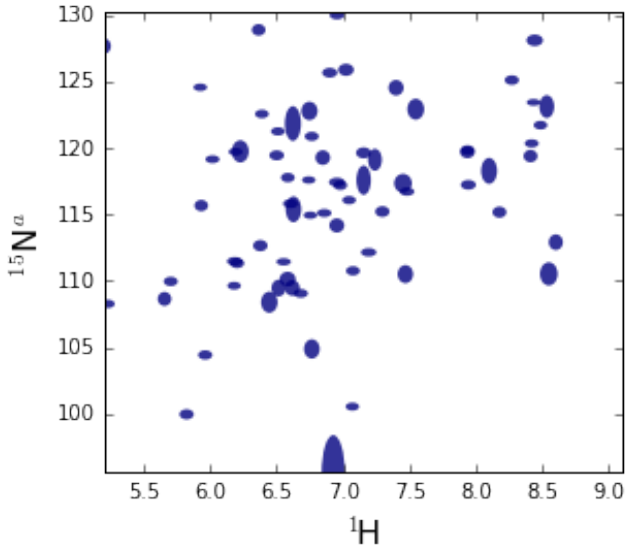
David Ryan Koes
John K. Vries

Computational and Systems Biology
University of Pittsburgh



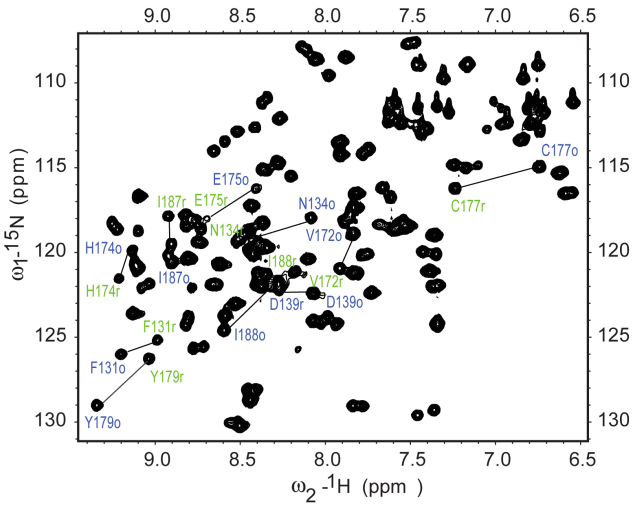
Molecular Dynamics

Experiment

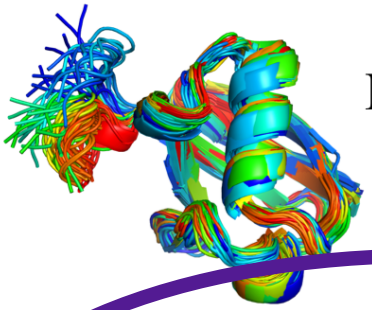


?

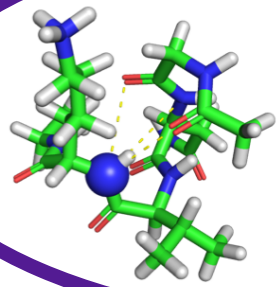
=



. DOI:10.1371/journal.ppat.1000960



Molecular Dynamics
Amber 14, TIP3P water
1ps sampling

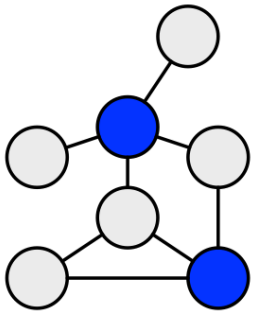


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Ensemble Average

Match Closest Template
Minimum Average Absolute Difference

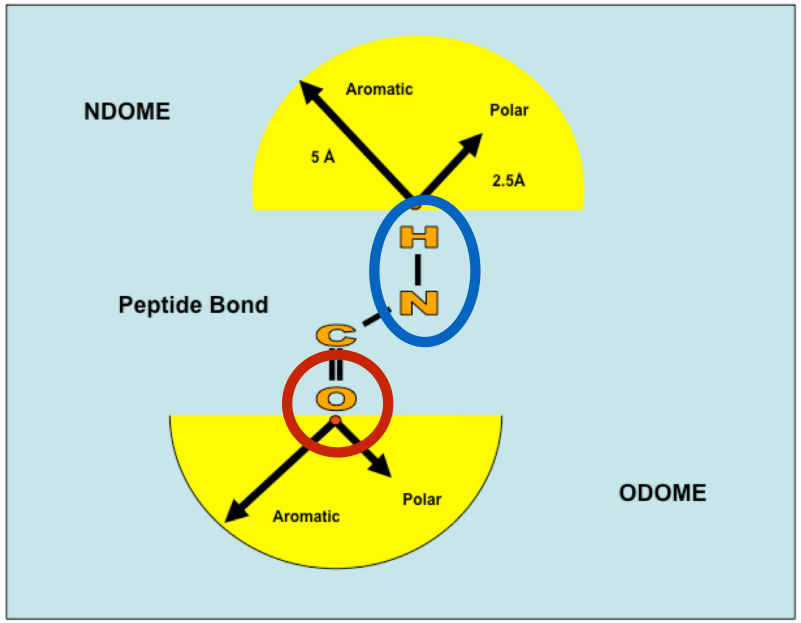
Template Database



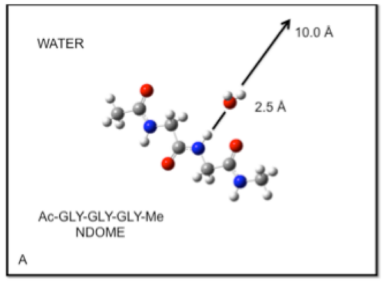
Select Novel Representative ROIs
0.2Å Matching Cutoff

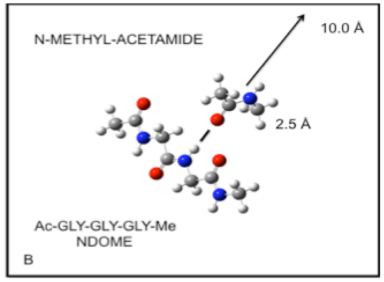
Calculate Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

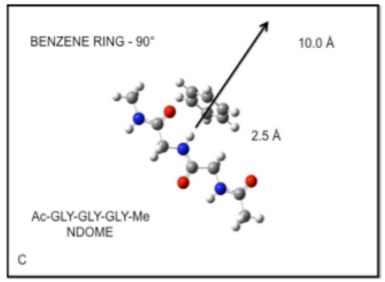
Regions of Interest (ROI)

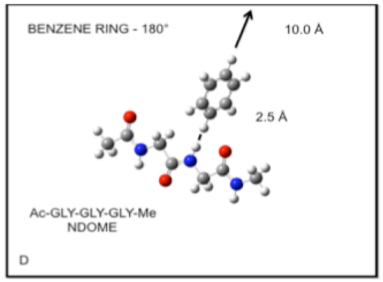


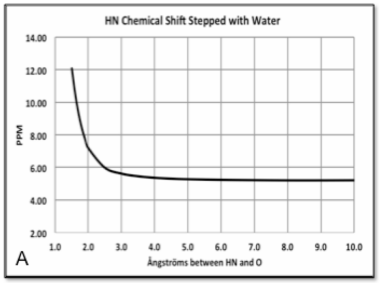
- Identify all nearby polar and aromatic atoms
- Separate analysis for O and NH

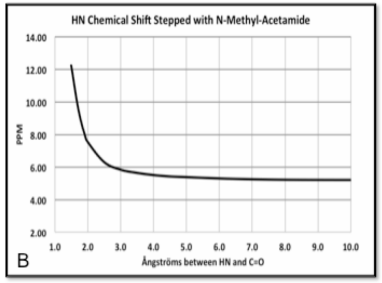


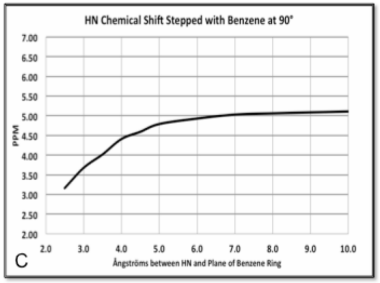


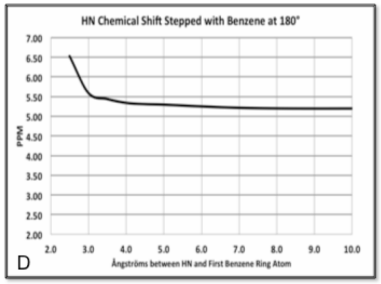




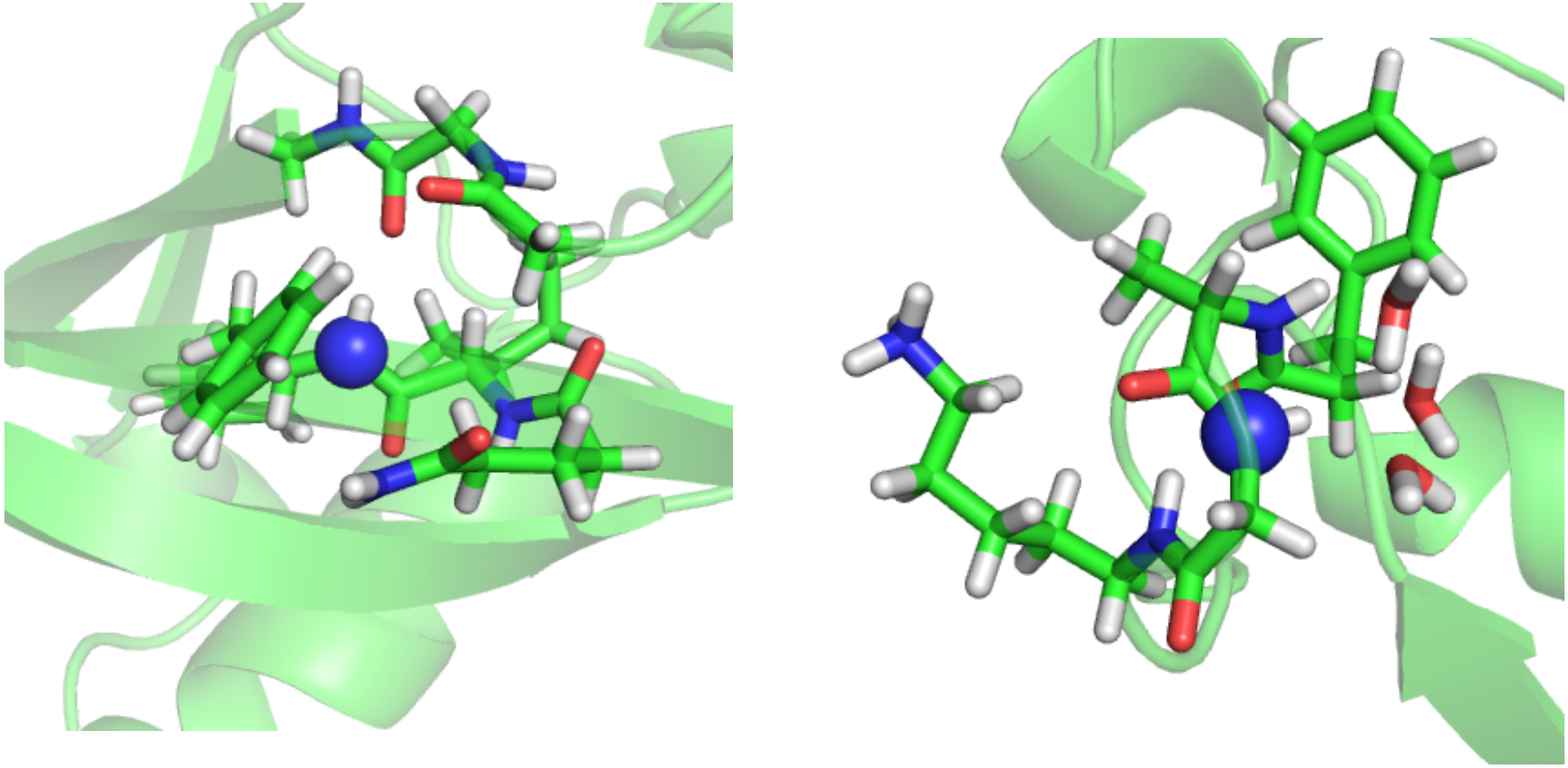




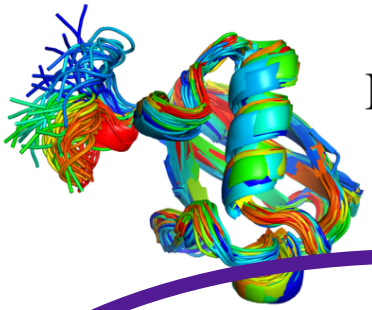




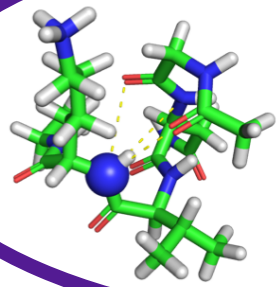
ROI Conformer



- 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

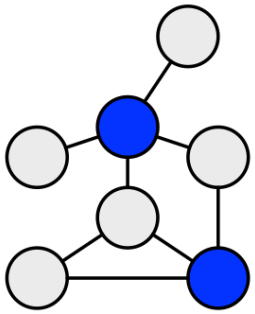


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Ensemble Average

Match Closest Template
Minimum Average Absolute Difference

Template Database



Select Novel Representative ROIs
0.2Å Matching Cutoff

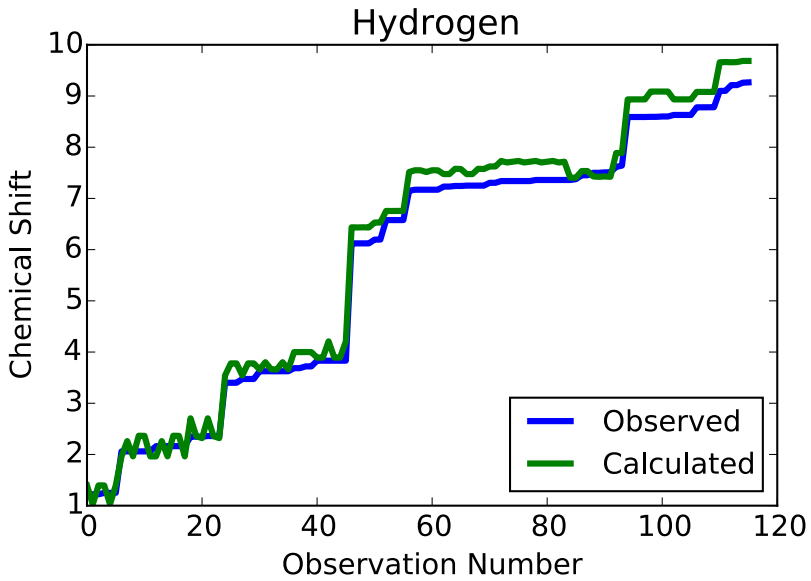
Calculate Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

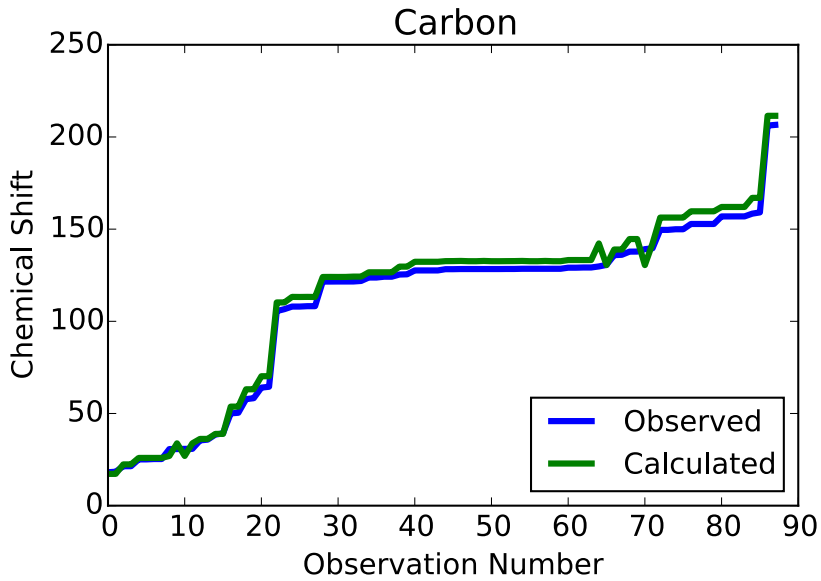
Calculate Chemical Shifts

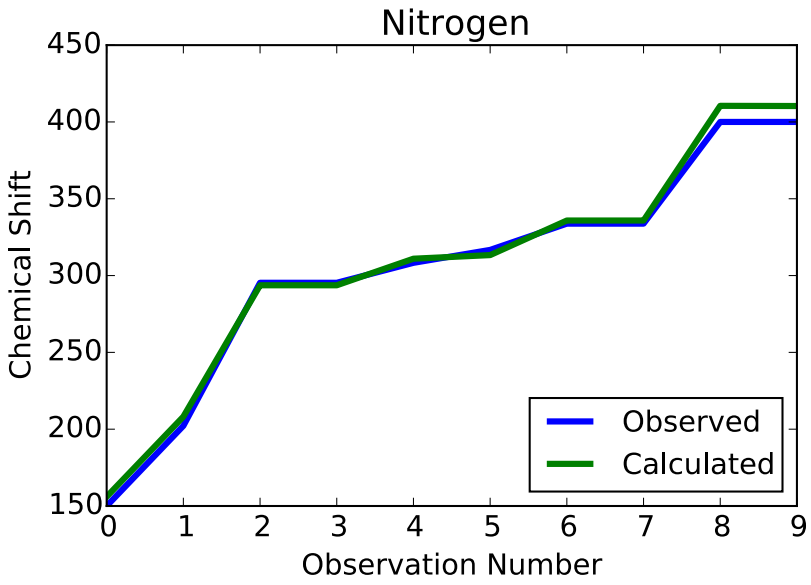
Magnetic resonance shielding tensors

QM Models	*6-311++ G(2d,p)/ 3-21G*	*6-311+ G(2d,p)/ 3-21G*	*6-311++ G(2d,p)*	*6-311+ G(2d,p)*
TMS ^{13}C	183.980	184.009	183.329	183.361
TSP ^{13}C	184.815	184.836	184.084	184.249
DSS ^{13}C	185.015	184.982	184.312	184.472
TMS ^{1}H	31.950	31.950	31.942	31.944
TSP ^{1}H	32.136	32.134	32.131	32.135
DSS ^{1}H	32.050	32.048	32.053	32.056

Calculate Chemical Shifts



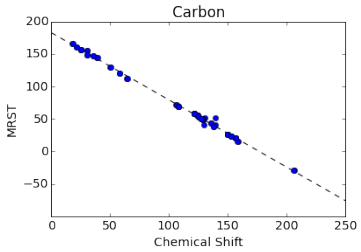


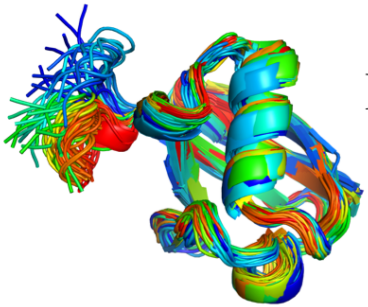


Gaussian G09

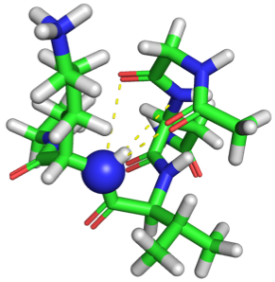
6-311+G(2d,p)

Reference Values:
Organic Compounds + PCM





Molecular Dynamics
Amber 14, TIP3P water
1ps sampling

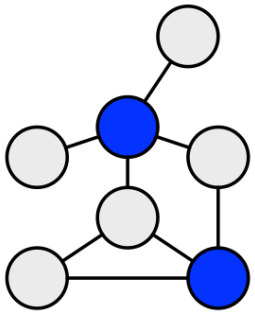


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Match Closest Template
Minimum Average Absolute Difference

Ensemble Average

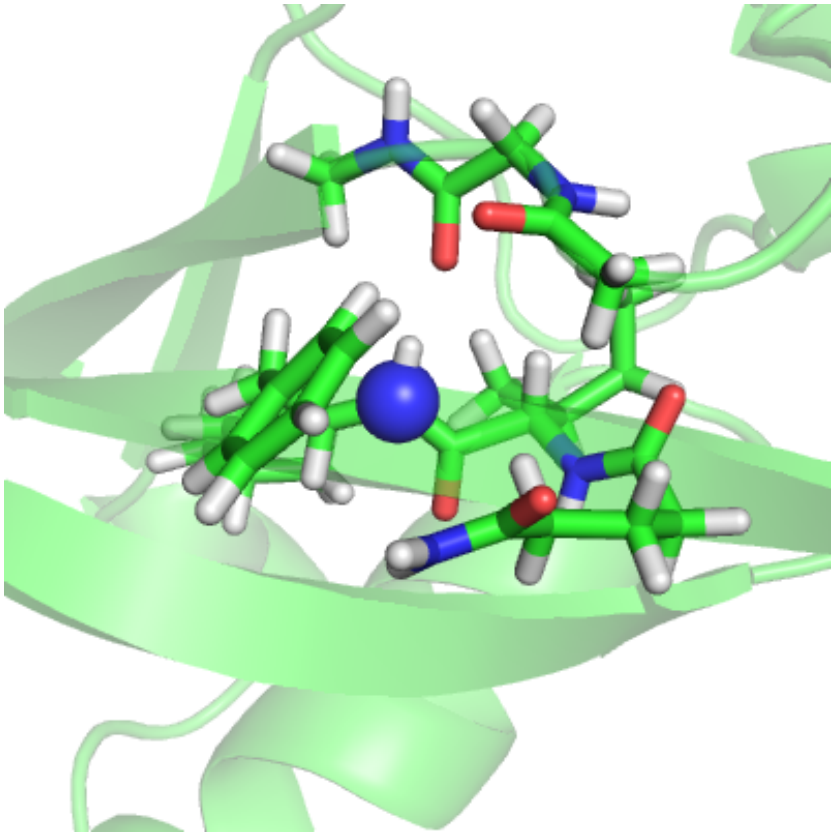
Template Database



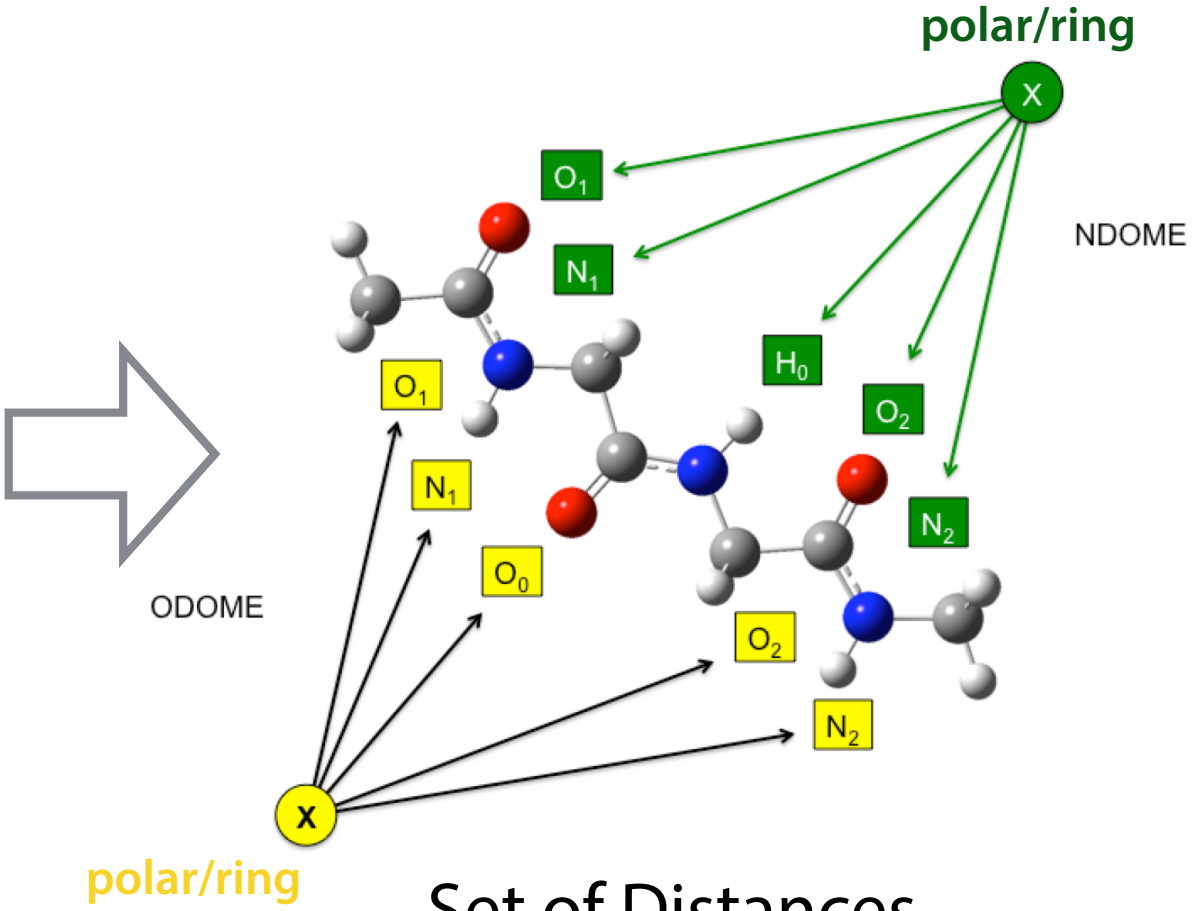
Select Novel Representative ROIs
0.2Å Matching Cutoff

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

Pattern Templates

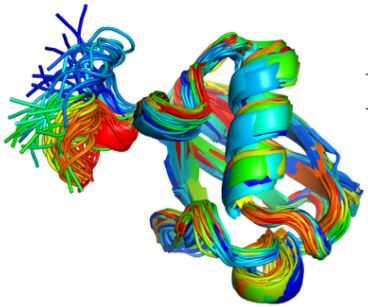


Conformation

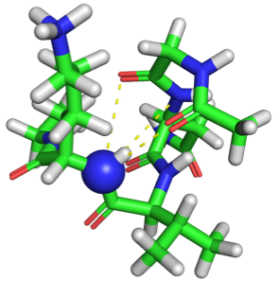


Set of Distances

```
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

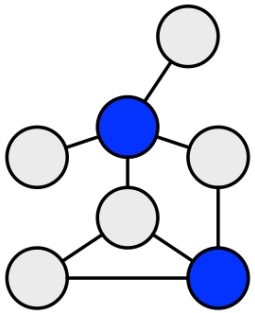


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Match Closest Template
Minimum Average Absolute Difference

Ensemble Average

Template Database

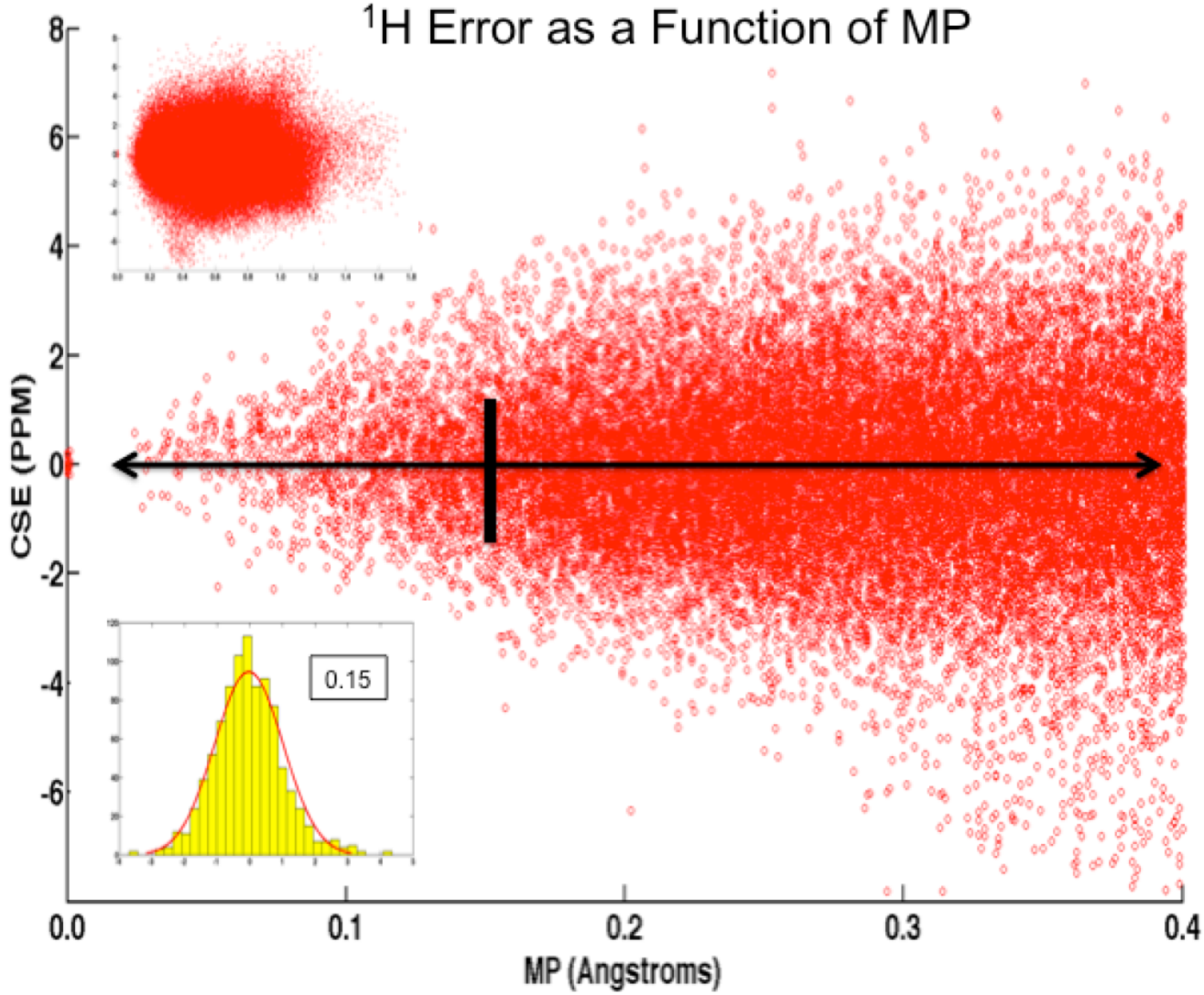


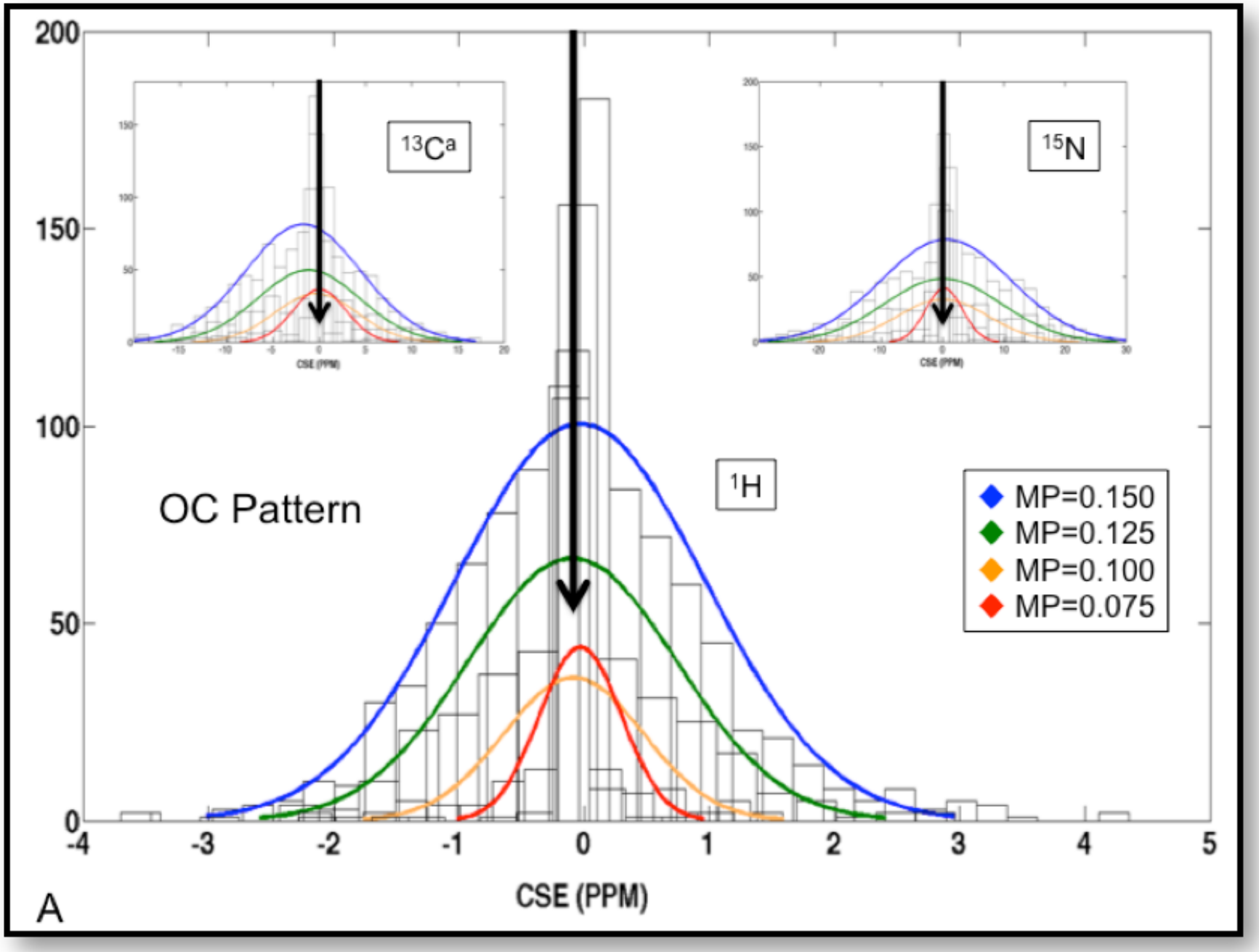
Select Novel Representative ROIs
0.2Å Matching Cutoff

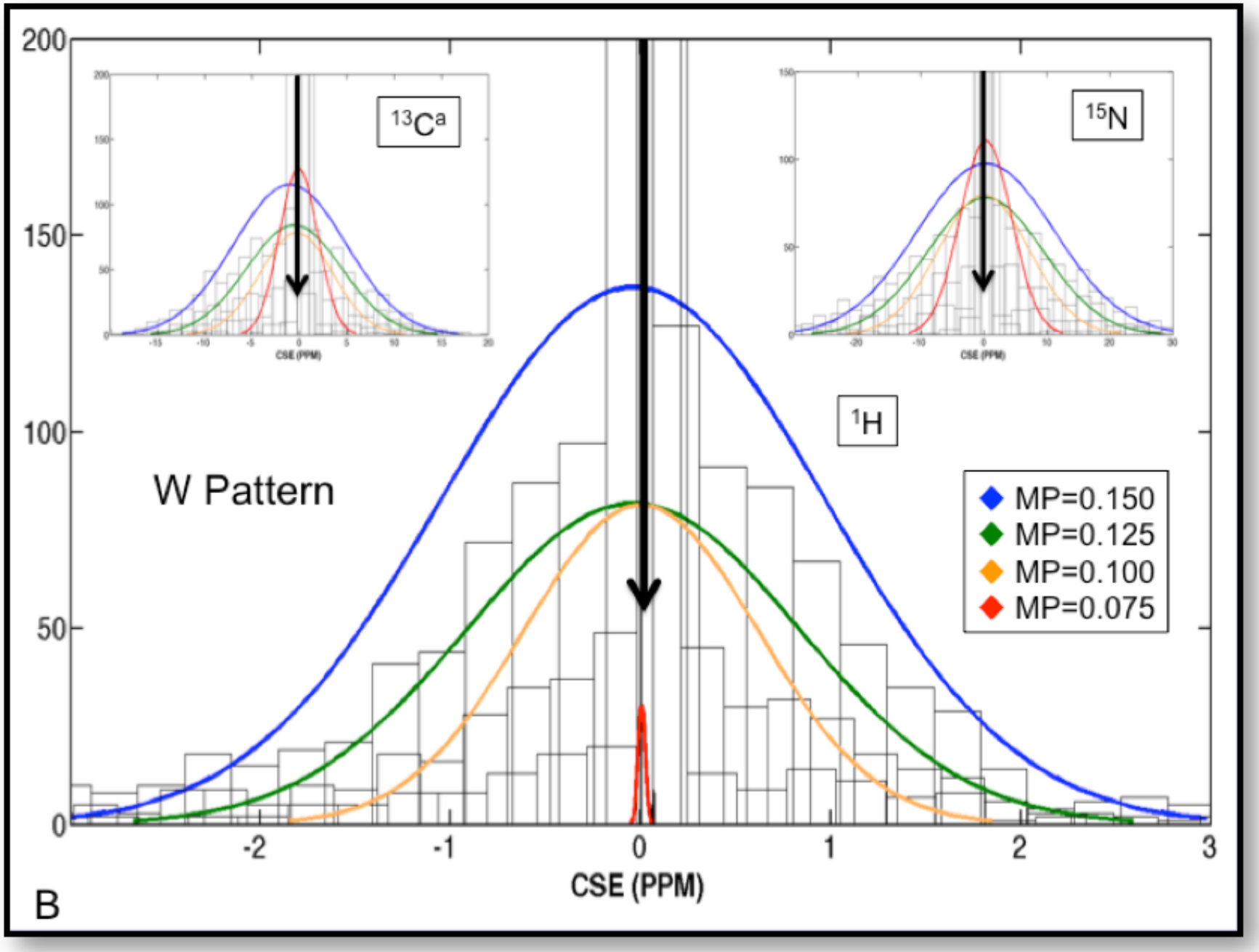
Calculate Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

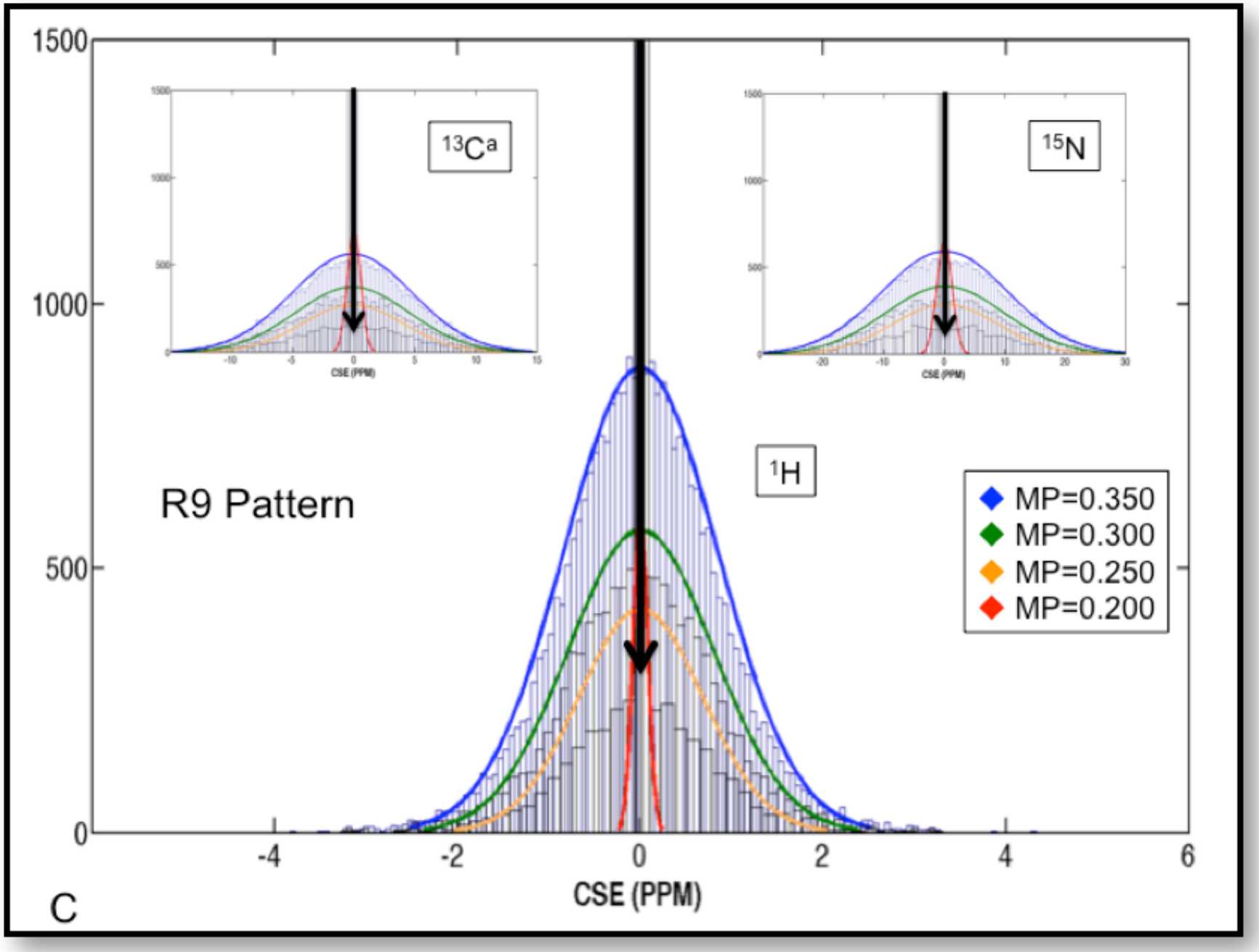
Matching Parameter

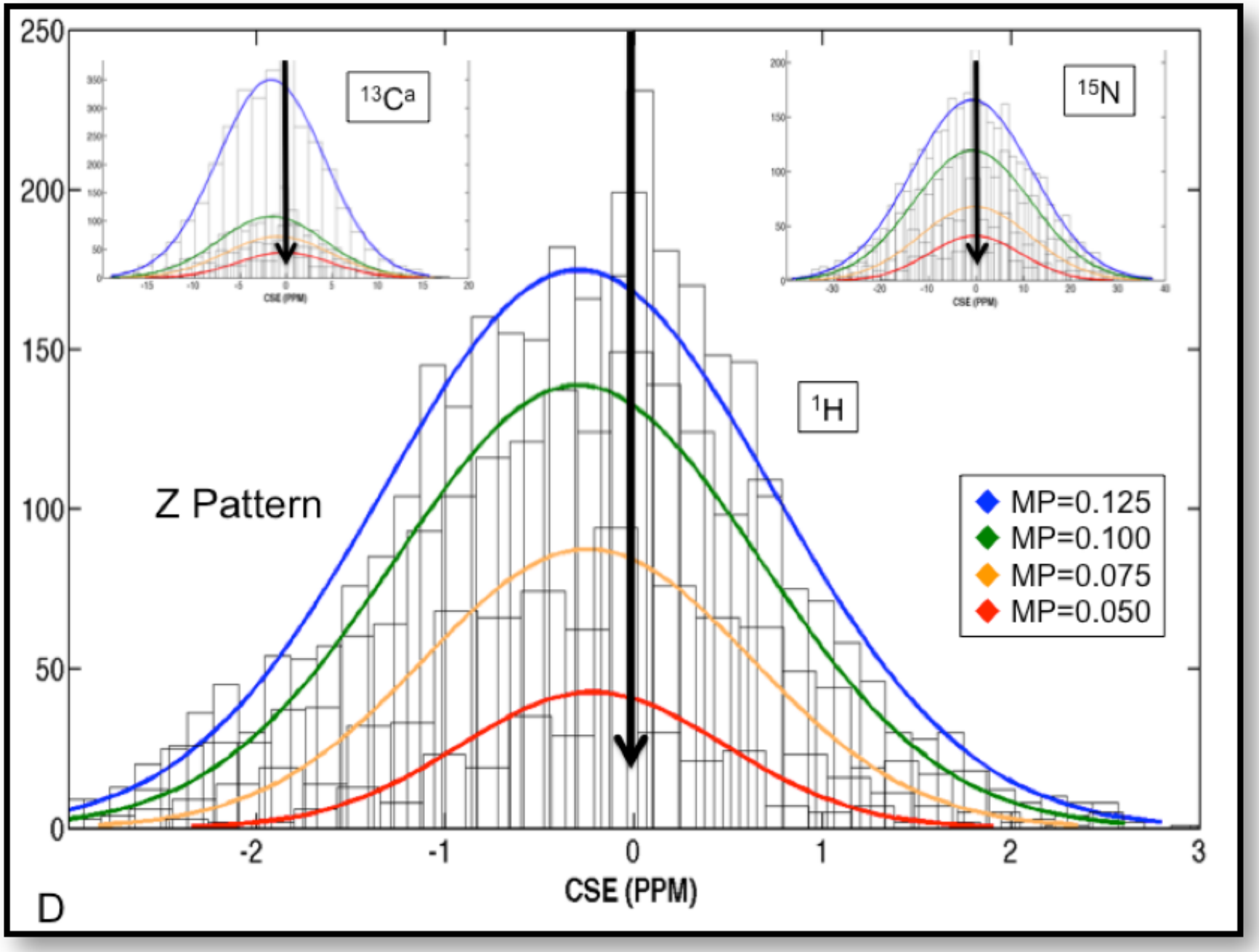
$$\mathrm{MP} = \sum_{i}^{n} \frac{|d_i|}{n}$$

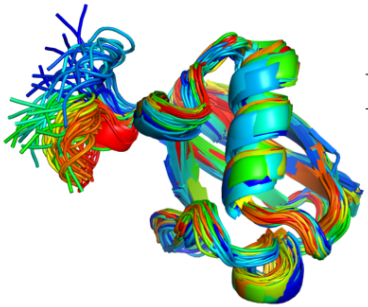



200
150
100
50
0
-4
-3
-2
-1
0
1
2
3
4
5
CSE (PPM)
A
OC Pattern
^{1}H
^{13}Ca
^{15}N
MP=0.150
MP=0.125
MP=0.100
MP=0.075

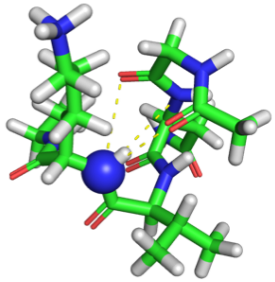

200
150
100
50
0
-2
-1
0
1
2
3
CSE (PPM)
B
W Pattern
$^{13}C^{a}$
^{15}N
^{1}H
MP=0.150
MP=0.125
MP=0.100
MP=0.075


1500
1000
500
0
-4
-2
0
2
4
6
CSE (PPM)
C
R9 Pattern
$^{13}C^{a}$
^{15}N
^{1}H
MP=0.350
MP=0.300
MP=0.250
MP=0.200


250
200
150
100
50
0
-2
-1
0
1
2
3
CSE (PPM)
D
Z Pattern
^{1}H
^{13}Cα
^{15}N
MP=0.125
MP=0.100
MP=0.075
MP=0.050



Molecular Dynamics
Amber 14, TIP3P water
1ps sampling

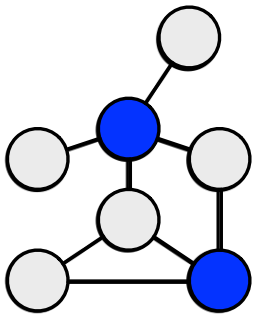


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Ensemble Average

Match Closest Template
Minimum Average Absolute Difference

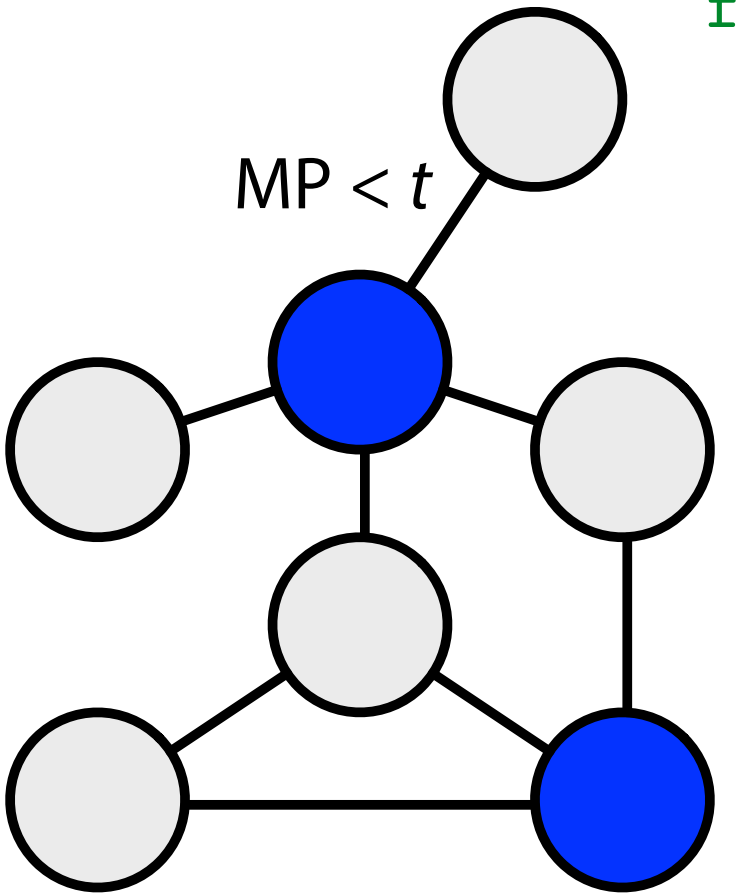
Template Database



Select Novel Representative ROIs
0.2Å Matching Cutoff

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

ROI Selection

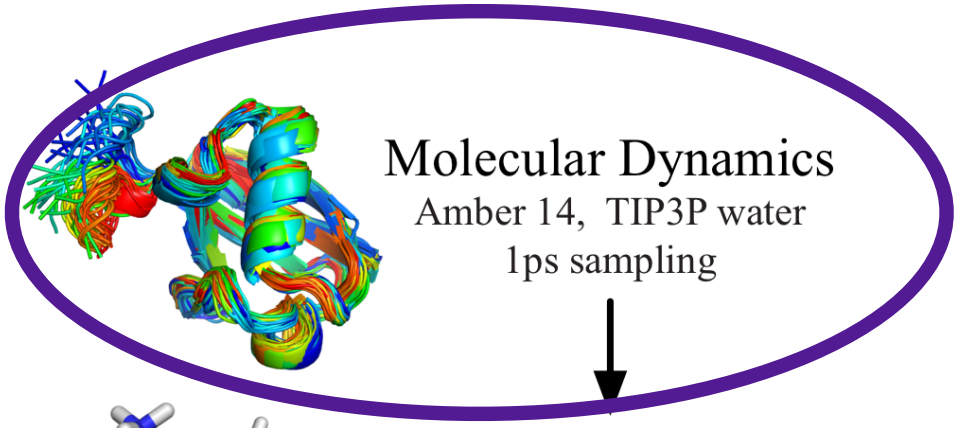



```
foreach pattern
 create_graph G_pattern
 foreach frame in simulation
  foreach ROI with pattern
   create_node n in G_pattern
   foreach m in G_pattern
    if MP(n,m) < threshold
     add_edge (n,m) to G_pattern
```

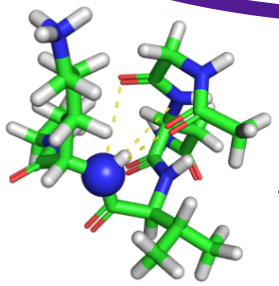
Solve *dominating set problem*

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

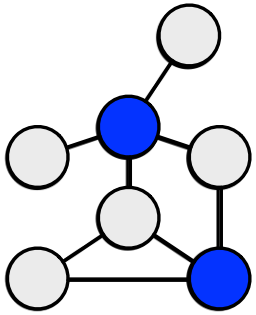


Extract Regions of Interest (ROIs)
5Å Radius, Complete Residues

Match Closest Template
Minimum Average Absolute Difference

Ensemble Average

Template Database



Select Novel Representative ROIs
0.2Å Matching Cutoff

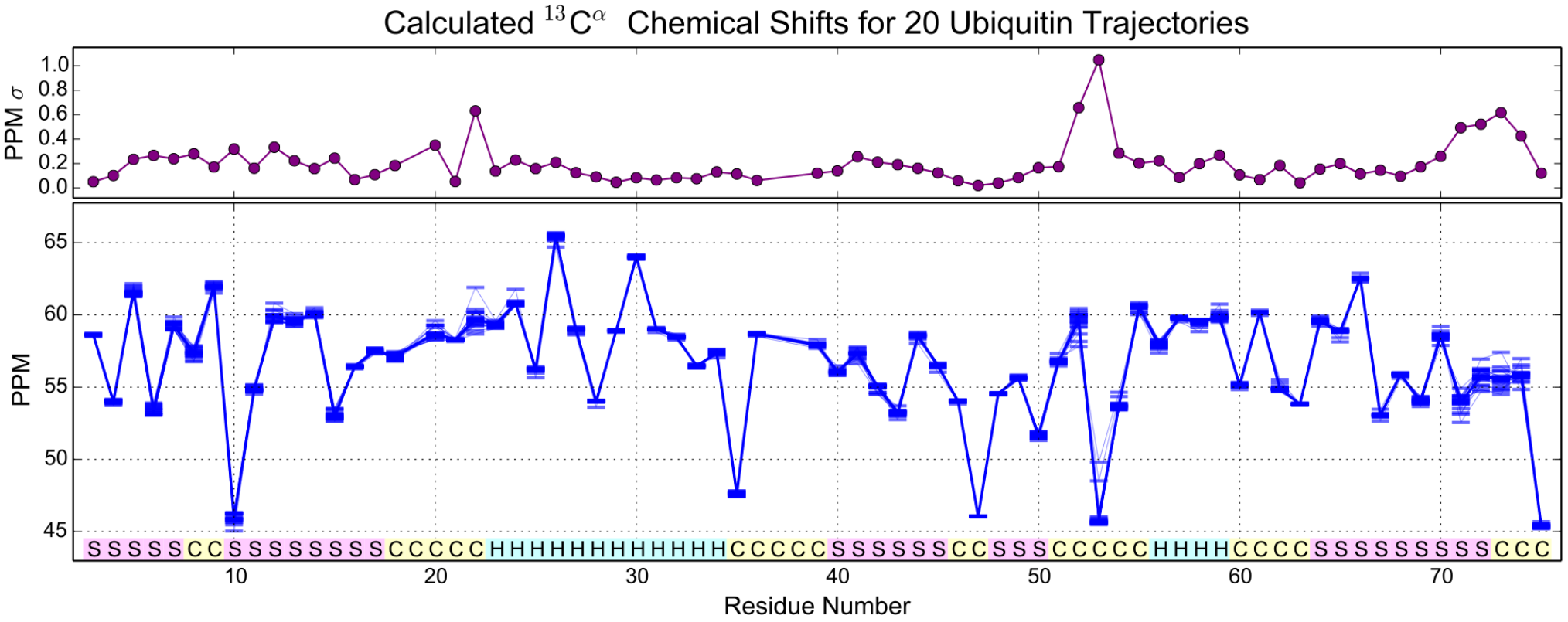
Calculate Chemical Shifts
Gaussian 09
B3LYP
6-311+G(2d,p)

Sources of Error

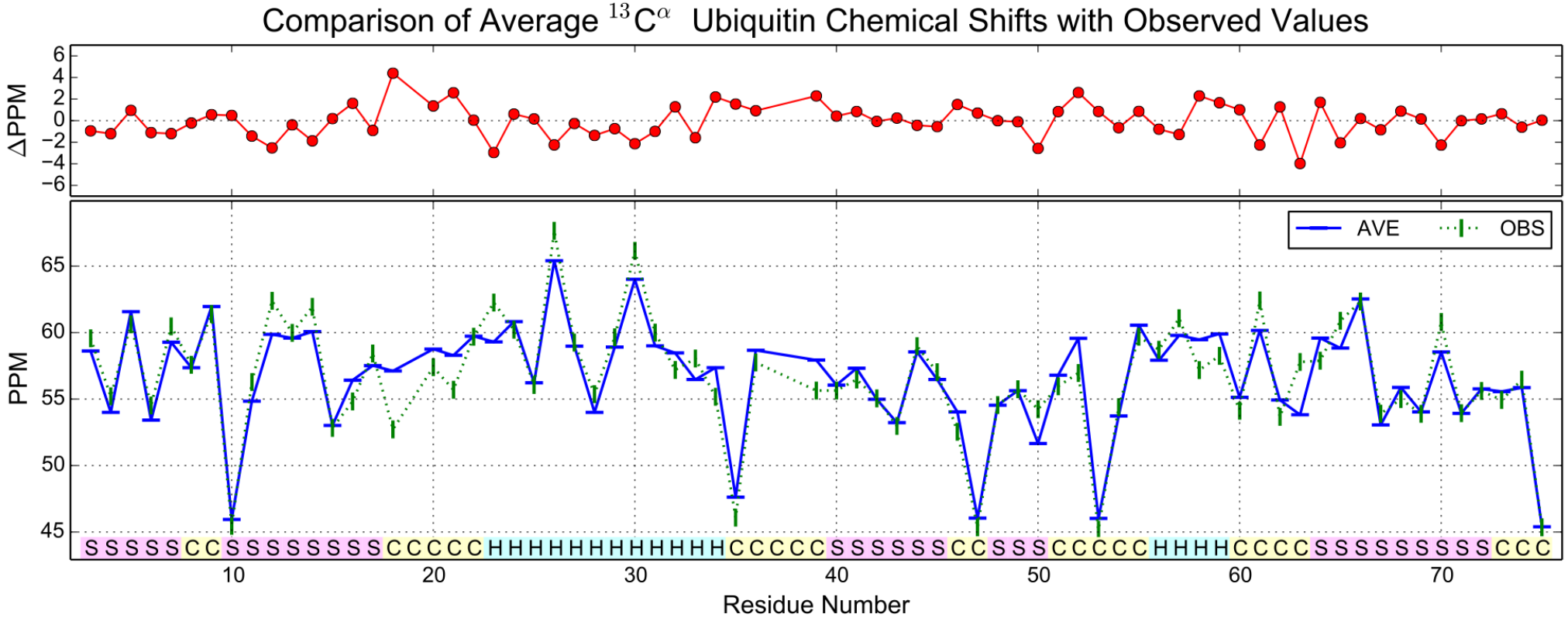
Level of Quantum Theory
Size of ROI
Reduction to Pattern
Closeness of Match

Sampling of MD
Accuracy of MD

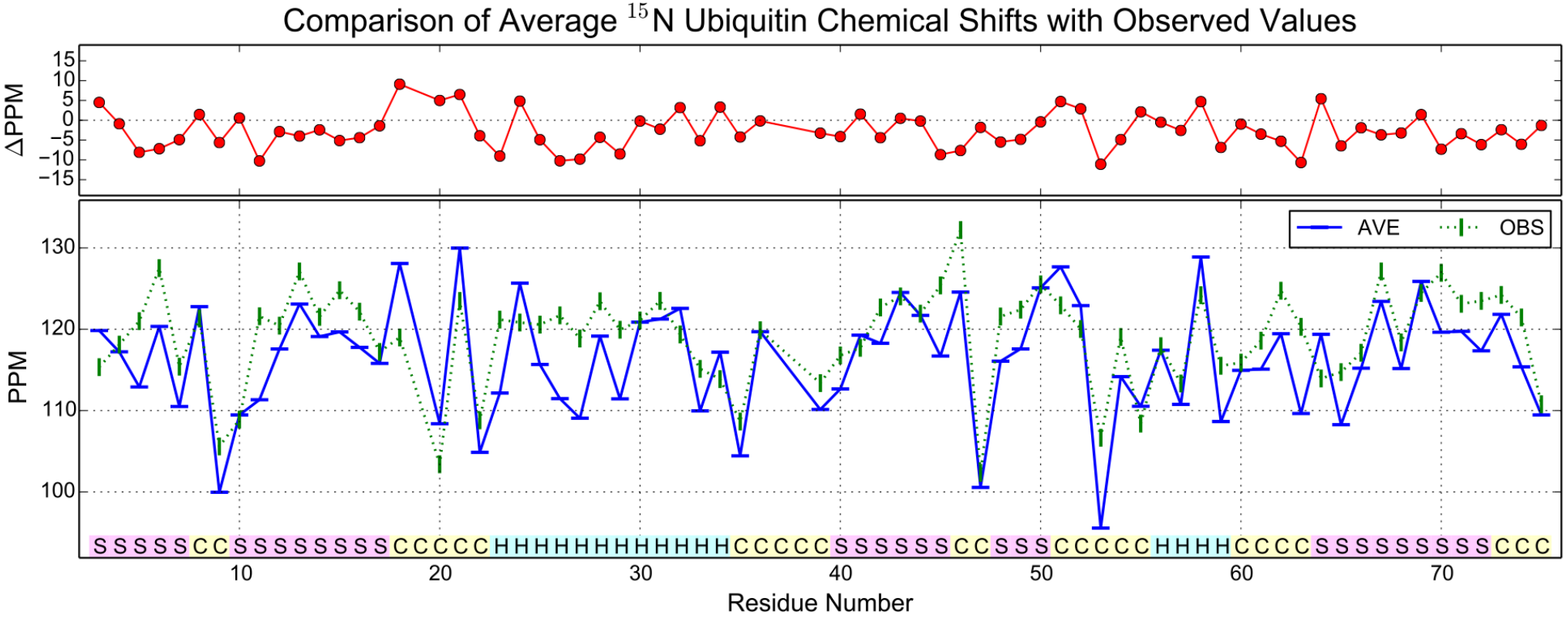
Ubiquitin


Calculated $^{13}C^{\alpha}$ Chemical Shifts for 20 Ubiquitin Trajectories
PPM σ
1.0
0.8
0.6
0.4
0.2
0.0
PPM
65
60
55
50
45
S S S S S C C S S S S S S S S C C C C C H H H H H H H H H H H C C C C C S S S S S S C C S S S C C C C C H H H H C C C C S S S S S S S S S C C C
10
20
30
40
50
60
70
Residue Number

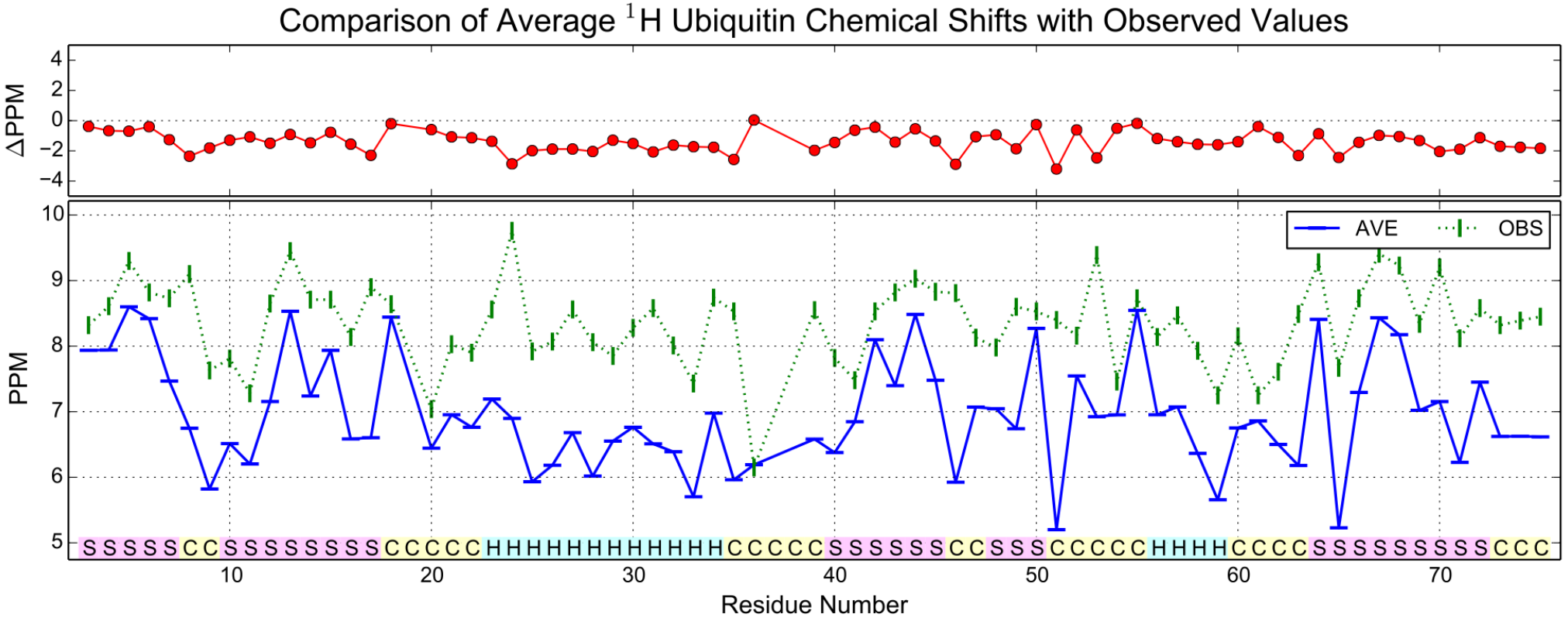
Ubiquitin


Comparison of Average $^{13}C^{\alpha}$ Ubiquitin Chemical Shifts with Observed Values
ΔPPM
6
4
2
0
−2
−4
−6
PPM
AVE
OBS
65
60
55
50
45
S S S S S C C S S S S S S S S C C C C C H H H H H H H H H H H C C C C C S S S S S C C S S S C C C C C H H H H C C C C S S S S S S S S S C C C
10
20
30
40
50
60
70
Residue Number

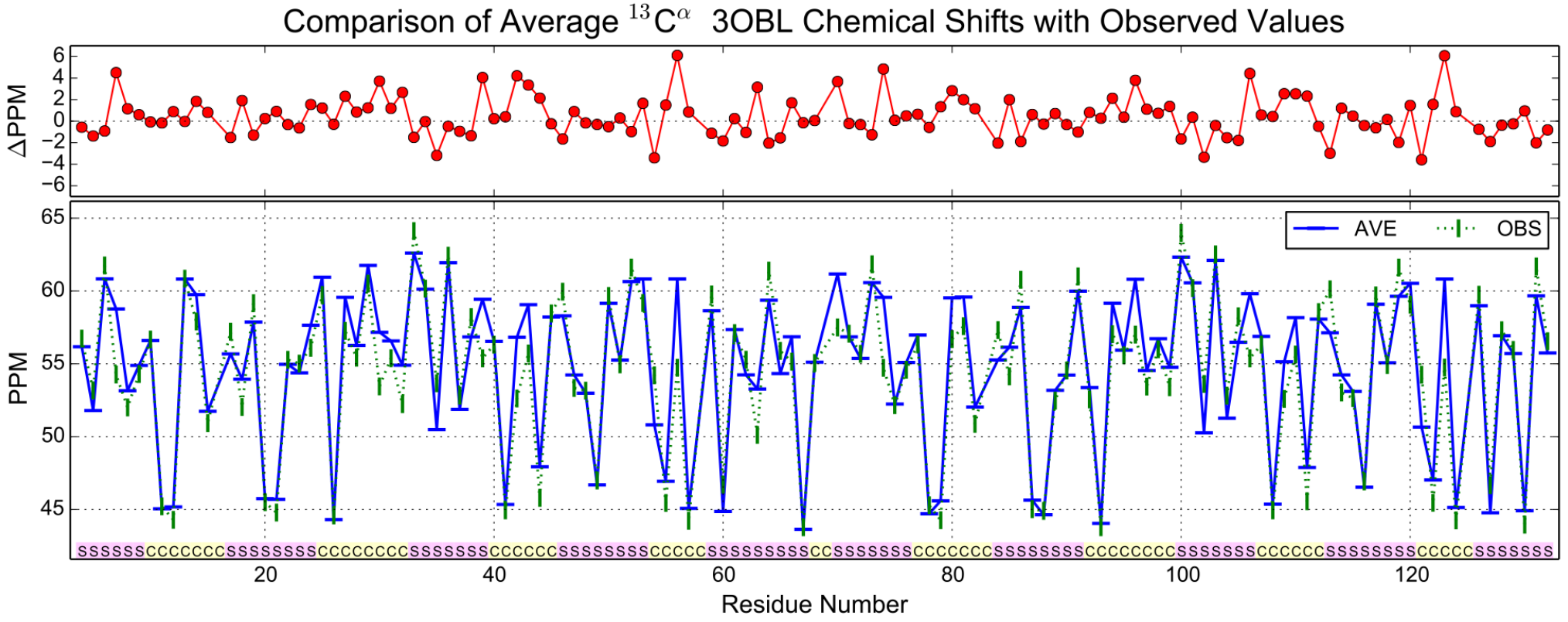
Ubiquitin


Comparison of Average ^{15}N Ubiquitin Chemical Shifts with Observed Values
ΔPPM
15
10
5
0
−5
−10
−15
PPM
130
120
110
100
AVE
OBS
S S S S S C C S S S S S S S S C C C C C H H H H H H H H H H H C C C C C S S S S S S S C C S S S C C C C C H H H H C C C C S S S S S S S S S C C C
10
20
30
40
50
60
70
Residue Number

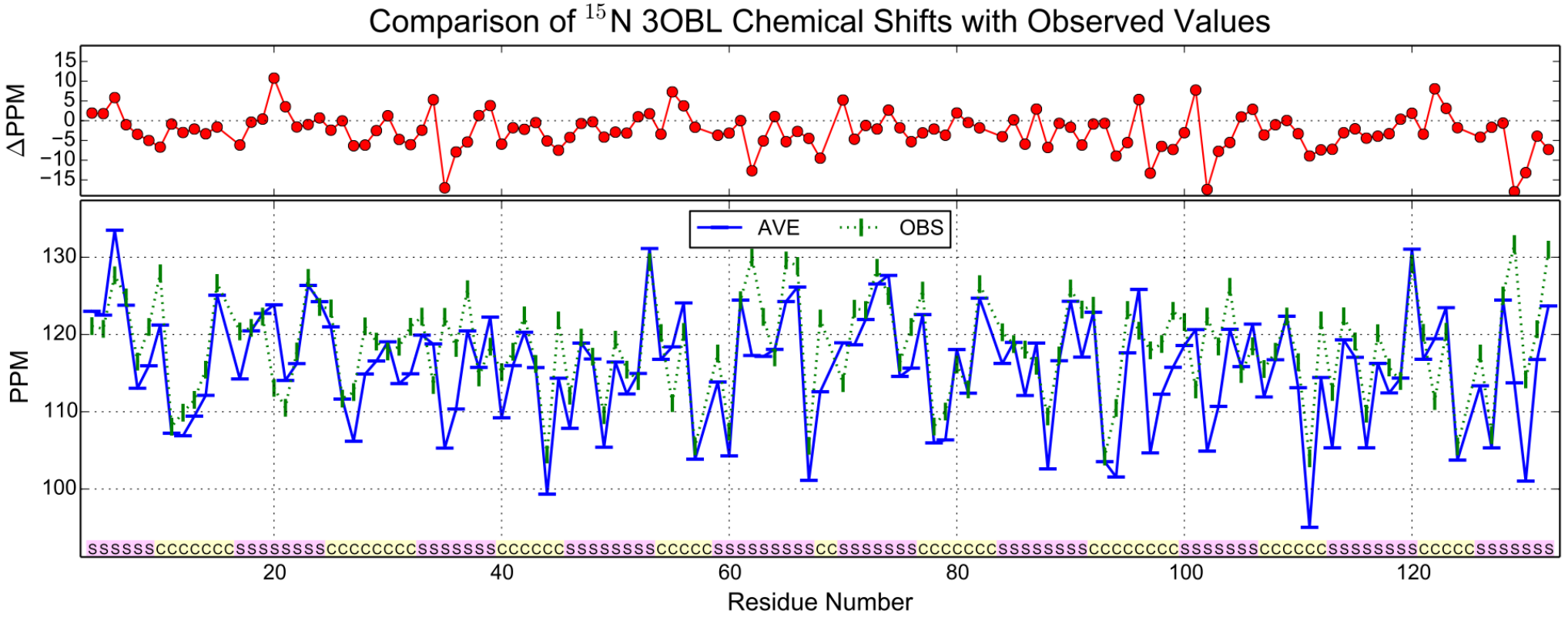
Ubiquitin


Comparison of Average ^{1}H Ubiquitin Chemical Shifts with Observed Values
ΔPPM
PPM
AVE
OBS
Residue Number
SSSSSCCSSSSSSSSCCCCCCHHHHHHHHHHHHHCCCCCSSSSSSCCSSSCCCCCCHHHHCCCCSSSSSSSSSCCC

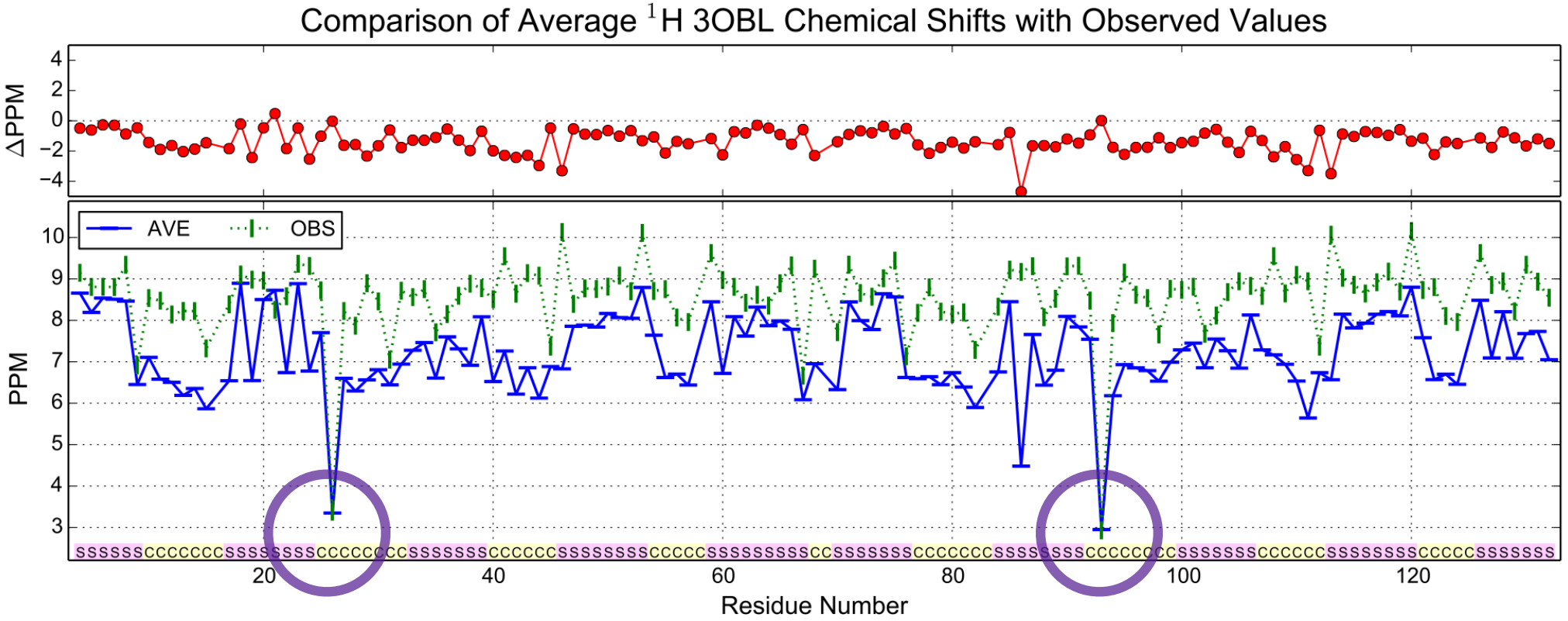
3OBL (cyanobacterial lectin)



3OBL (cyanobacterial lectin)

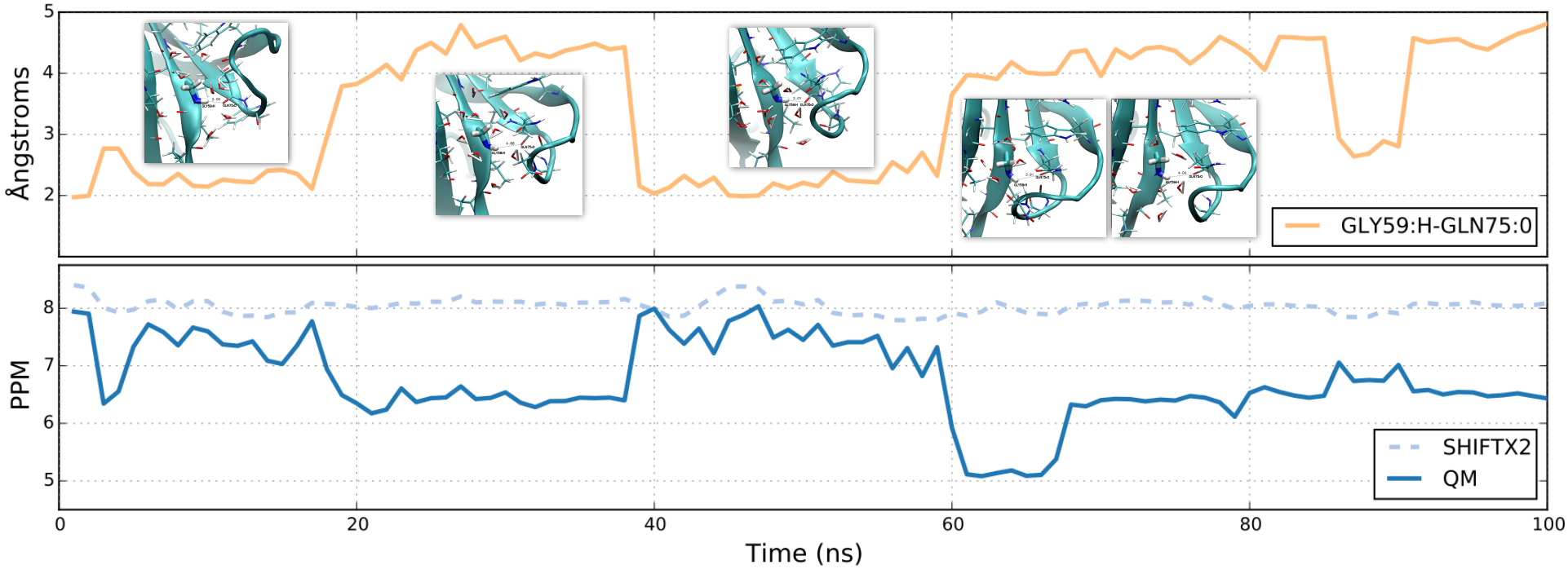


3OBL (cyanobacterial lectin)



3OBL (cyanobacterial lectin)

GLY59



Overall Results

PROTEIN	N-ERR%	N-RMSD	H-ERR%	H-RMSD	C-ERR%	C-RMSD
1ENH	3.9%	5.795	21.3%	1.848	2.1%	1.568
1HIK	3.1%	4.660	18.4%	1.578	2.4%	1.721
1IGD	4.6%	6.484	17.5%	1.651	2.9%	2.240
3OBL	3.5%	5.418	15.6%	1.558	2.6%	1.914
1QZM	3.8%	5.644	19.0%	1.678	3.1%	2.268
1UBQ	3.7%	5.197	16.6%	1.559	2.1%	1.526
MEAN	**3.8%**	**5.533**	**18.2%**	**1.645**	**2.5%**	**1.873**
STD	0.5%	0.612	2.2%	0.111	0.4%	0.325
1UBQ20	3.7%	5.261	16.8%	1.555	2.1%	1.503

Conclusions

- Template-based approach is feasible and conformation sensitive
- Large systematic error with H shifts
  - only in H-bond context
  - reducing H-bond length reduces error

Happy to share data - >170,000 Gaussian ROI calculations

Acknowledgements

John Vries

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

Department of Computational and Systems Biology

Center for Simulation and Modeling

Funding
R01GM108340

Questions?

Presentation CINF: Division	CINF 37: 3Dmol.js: Chemical structure visualization for the modern web 6:30pm-8:30pm Sun, Mar 13	Jasmine Collins
Presentation COMP: Divisio..	COMP 91: Quantum chemical approach for evaluating molecular mechanics force fields based on comparison… 10:30am-10:55am Mon, Mar 14	David Koes
Presentation COMP: Divisio..	COMP 232: GPU implementation of energy minimization for virtual screening 8:00pm-10:00pm Mon, Mar 14	Jocelyn Sunseri
Presentation COMP: Divisio..	COMP 165: Pharmit: Interactive exploration of chemical space 11:25am-11:45am Tue, Mar 15	David Koes
Presentation COMP: Divisio..	COMP 271: Convolutional neural networks for protein-ligand scoring 6:00pm-8:00pm Tue, Mar 15	Matthew Ragoza
Presentation COMP: Divisio..	COMP 374: Benchmarking computational methods for binding free-energy estimation 6:00pm-8:00pm Tue, Mar 15	Jocelyn Sunseri
Presentation COMP: Divisio..	COMP 377: Fragment oriented molecular shape (FOMS) search: A novel shape-based virtual screening method 6:00pm-8:00pm Tue, Mar 15	Ethan Hain
Presentation COMP: Divisio..	COMP 232: GPU implementation of energy minimization for virtual screening 6:00pm-8:00pm Tue, Mar 15	Jocelyn Sunseri