Supporting Information for

Forcefield_PTM: *Ab Initio* Charge and AMBER Forcefield Parameters for Frequently Occurring Post-Translational Modifications

George A. Khoury, Jeff P. Thompson, James Smadbeck,

Chris A. Kieslich, and Christodoulos A. Floudas*

Department of Chemical and Biological Engineering, Princeton, NJ, USA

E-mail: floudas@titan.princeton.edu

*To whom correspondence should be addressed

1 Supplementary Instructions and Parameters

1.1 Instructions for importing new parameters into AMBER

AMBER must be installed and the environment variable $AMBERHOME must be set to the directory in which AMBER is installed locally. Please see the AMBER manual regarding how to do this. The archive **ffptm.zip** contains all the files related to the forcefield (ffptm.in and all .frcmod files as described below).

Before running any simulations, a user must prepare their protein structure for simulation in AMBER. This is typically done in tleap. The following is the contents of the file **tleap-min.in** which contains the syntax needed to import the parameters to perform calculations using FFPTM's parameterization of Phosphoserine (SEP) on a protein structure containing one or more SEP residues. /DirectoryContainingFFPTM/ refers to the path to the directory containing FFPTM (where **ffptm.zip** is unzipped). In this particular example it contains ffptm.in and SEP.frcmod.

File: tleapmin.in

```
source leaprc.ff03.r1   # imports ff03 parameters


# loads FFPTM
loadAmberPrep /DirectoryContainingFFPTM/ffptm.in
# loads any residue-specific parameters needed for SEP
frcmod = loadAmberParams /DirectoryContainingFFPTM/SEP.frcmod
protein = loadpdb test.pdb
check protein
saveamberparm protein test.prmtop test.inpcrd
quit
```

This file produces test.prmtop (a topology/parameter file) and test.inpcrd (an input coordinate file) which are needed by sander or pmemd to perform AMBER simulations.

To call **tleapmin.in** you must use the command:

```
COMMAND: Serial call to tleap

_____________________________________________________

$AMBERHOME/bin/tleap -f tleapmin.in
```

Note, if your structure has multiple post-translational modifications, you may add more lines to tleapmin.in as per the following command example.

```
COMMAND: Example of how to load in multiple frcmod files to tleapmin.in

_____________________________________________________

$AMBERHOME/bin/tleap -f tleapmin.in

frcmoda = loadAmberParams /DirectoryContainingFFPTM/SEP.frcmod

frcmodb = loadAmberParams /DirectoryContainingFFPTM/M3L.frcmod
```

Using the newly created prmtop and inpcrd files generated using **tleapmin.in**, you can now perform a local minimization of your protein containing one or more SEP residues using the following file min1.in.

```
File: min1.in

_____________________________________________________

Stage 1 - minimization of structure

 &cntrl

 imin = 1, maxcyc = 10000, ncyc = 6000,

 cut = 16., rgbmax = 16., igb=5, ntb = 0,

 ntpr = 100,

/
```

To perform the minimization, you must call sander or pmemd. The following calls sander to minimize the structure prepared above.

```
COMMAND: Serial call to sander

_____________________________________________________


$AMBERHOME/exe/sander -O -i min1.in -o min1.out -p test.prmtop -c \

test.inpcrd -r min1.rst -ref test.inpcrd

ambpdb -p test.prmtop <min1.rst > test_min.pdb


COMMAND: Parallel call to sander.MPI

_____________________________________________________

mpirun -np 4 $AMBERHOME/exe/sander.MPI -O -i min1.in -o min1.out -p test.prmtop \

-c test.inpcrd -r min1.rst -ref test.inpcrd

ambpdb -p test.prmtop <min1.rst > test_min.pdb
```

The above commands will produce for you a locally minimized protein structure containing the SEP post-translational modification. A note of caution: please make sure the atoms in the modified residue you are simulating have names matching those atom names in the forcefield parameterization. If the atom names do not match, change the naming in the PDB file to match that in the forcefield. For PTMs contained within the PDB that were parameterized, efforts were undertaken to make sure the atom namings matched. In some cases though, PDB file atom namings were not identical even for the same modification.

One can perform unrestrained molecular dynamics in implicit solvent as was done in this paper by using the following input files to sander.

```
File: heat1.in

_____________________________________________________

Stage 1 heating of protein 0 to 50K

 &cntrl

  imin=0, irest=0, ntx=1,

  nstlim=100000, dt=0.0005,
```

```
 ntc=2, ntf=2,
 ntt=3, gamma_ln = 5,
 tempi=0.0, temp0=50.0,
 ntpr=50, ntwx=500,
 ntb=0, igb=5,ig=-1,
 cut=16.,rgbmax=16.
/
```

The stepsize can be increased from dt=0.0005 to 0.002 if the user wishes. The protein can be subsequently heated by adapting the file above and creating the temperature intervals 50-100, 100-150, 150-200, 200-250, 250-300. Then, the protein can be submitted for production using the input file below.

File: prod.in

```
Production run implicit solvent 0-5ns
 &cntrl
  imin=0, irest=1, ntx=5,
  nstlim=2500000, dt=0.002,
  ntc=2, ntf=2,
  ntt=3, gamma_ln = 5,
  tempi=300.0, temp0=300.0,
  ntpr=500, ntwx=500,
  ntb=0, igb=5,ig=-1,
  cut=16.,rgbmax=16.
 /
```

1.2 Explanation of contents of AMBER prep file

There are a number of parameters contained in the AMBER prep file which contains the charges calculated in this work. Using the instructions provided, one can easily import the new parameters into AMBER. We present as an example the explanation of the parameters for Anti-symmetric dimethylarginine which is contained in the forcefield.

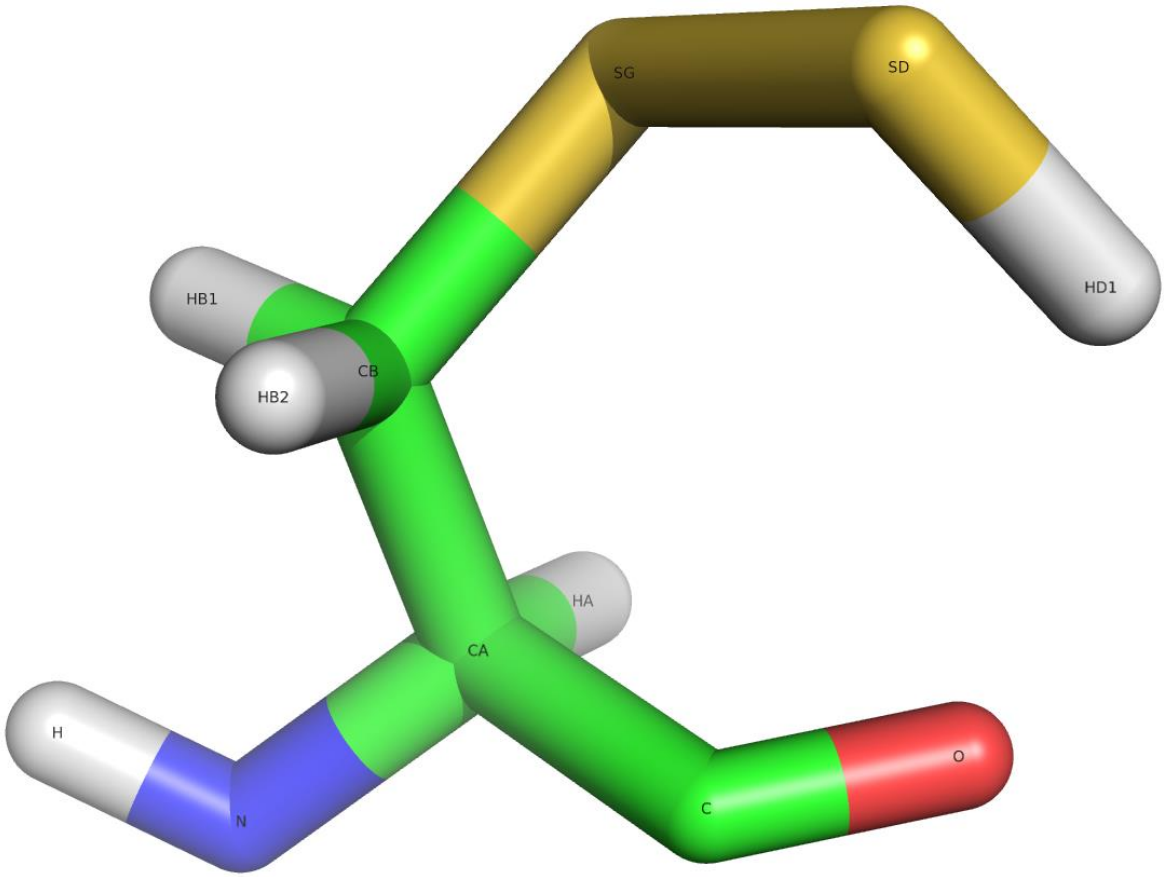
```
Dimethylarginine (Anti-symmetric) (name of the residue)
dimethylarg.res  (Name of the output file generated in from Antechamber)
DA2   INT  0     (Unique 3-letter code naming the amino acid)
CORRECT     OMIT DU   BEG
  0.0000
(1) # of atoms in tree
(2) Unique atom name for atom I
(3) Symbol for atom I which defines its forcefield atom type
and is used in the module PARM for assigning the forcefield parameters
(4) The topological type (tree symbol) for atom I
(5) The atom number to which atom I is connected
(6) The atom number to which atom I makes an angle along with (5)
(7) The atom number which atom I makes a dihedral along with (5) and (6)
(8) The bond length between atoms I and (5)
(9) The bond angle between the bond angle between (6),(5) and (1)
(10) The dihedral angle between (7), (6), (5), and (1)
(11) The partial atomic charge on atom I generated from the RESP-fitting procedure.

 (1)   (2)   (3)    (4)   (5)  (6)  (7)      (8)        (9)        (10)       (11)
   1   DUMM  DU     M     0   -1   -2      0.000       .0         .0       .00000
   2   DUMM  DU     M     1    0   -1      1.449       .0         .0       .00000
   3   DUMM  DU     M     2    1    0      1.522    111.1         .0       .00000
   4   N     N      M     3    2    1      1.540    111.208    180.000 -0.376741
```

5	H	H	E	4	3	2	0.994	107.922	-18.167	0.270465
6	CA	CT	M	4	3	2	1.448	10.429	150.011	-0.119858
7	CB	CT	3	6	4	3	1.539	110.455	4.603	-0.129377
8	CG	CT	3	7	6	4	1.540	113.981	-168.737	-0.016346
9	CD	CT	3	8	7	6	1.534	114.694	-91.332	-0.092919
10	NE	N2	B	9	8	7	1.465	112.221	171.176	-0.288262

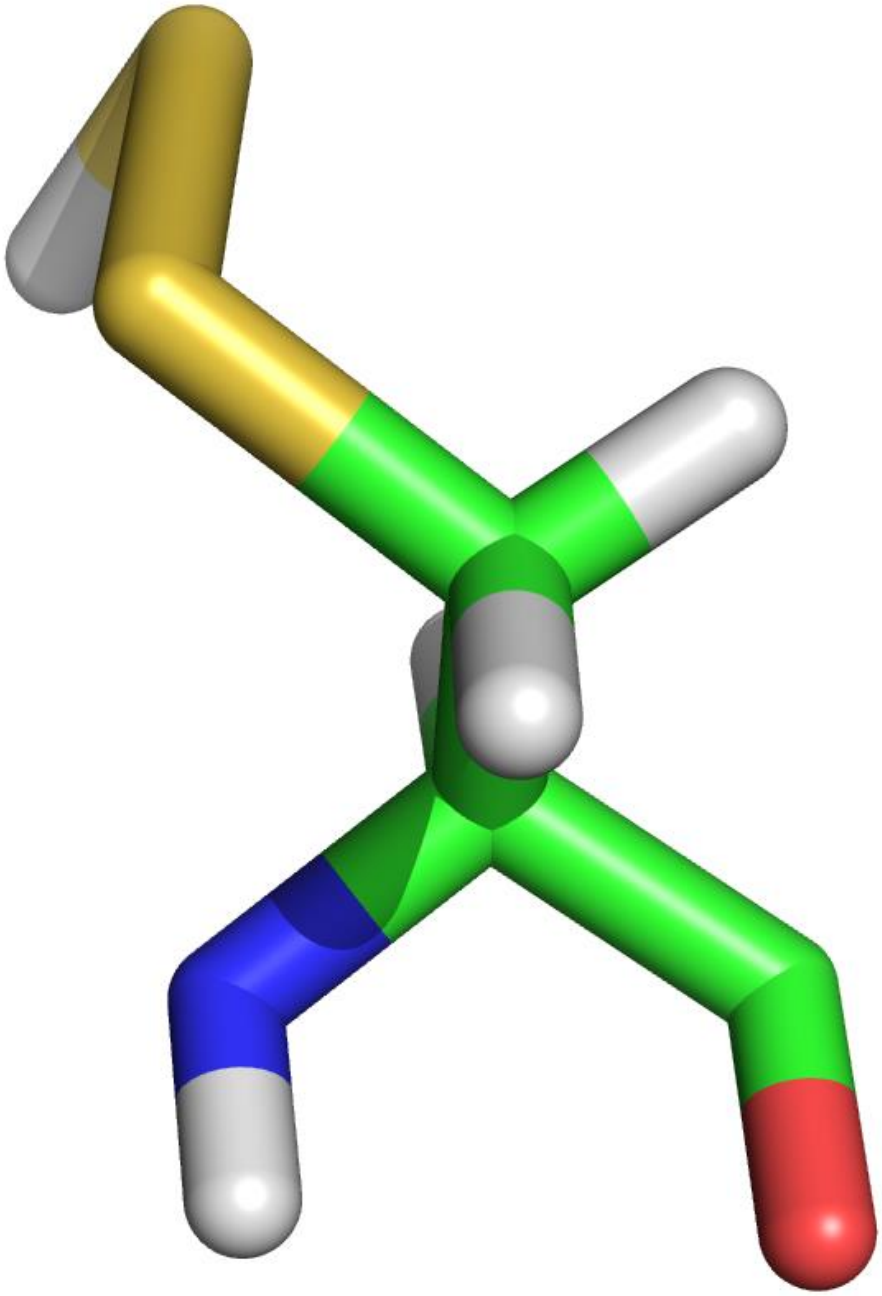
1.3 Images of Each Parameterized Post-Translational Modification Grouped by Scaffold Residue

Images of the HF/6-31G** optimized α-helical and β-strand conformer of each post-translational modification are presented, grouped by scaffold residue to serve as reference for the presented parameters. The atom names following PDB conventions are labeled for each PTM as well in the α-helical conformation.

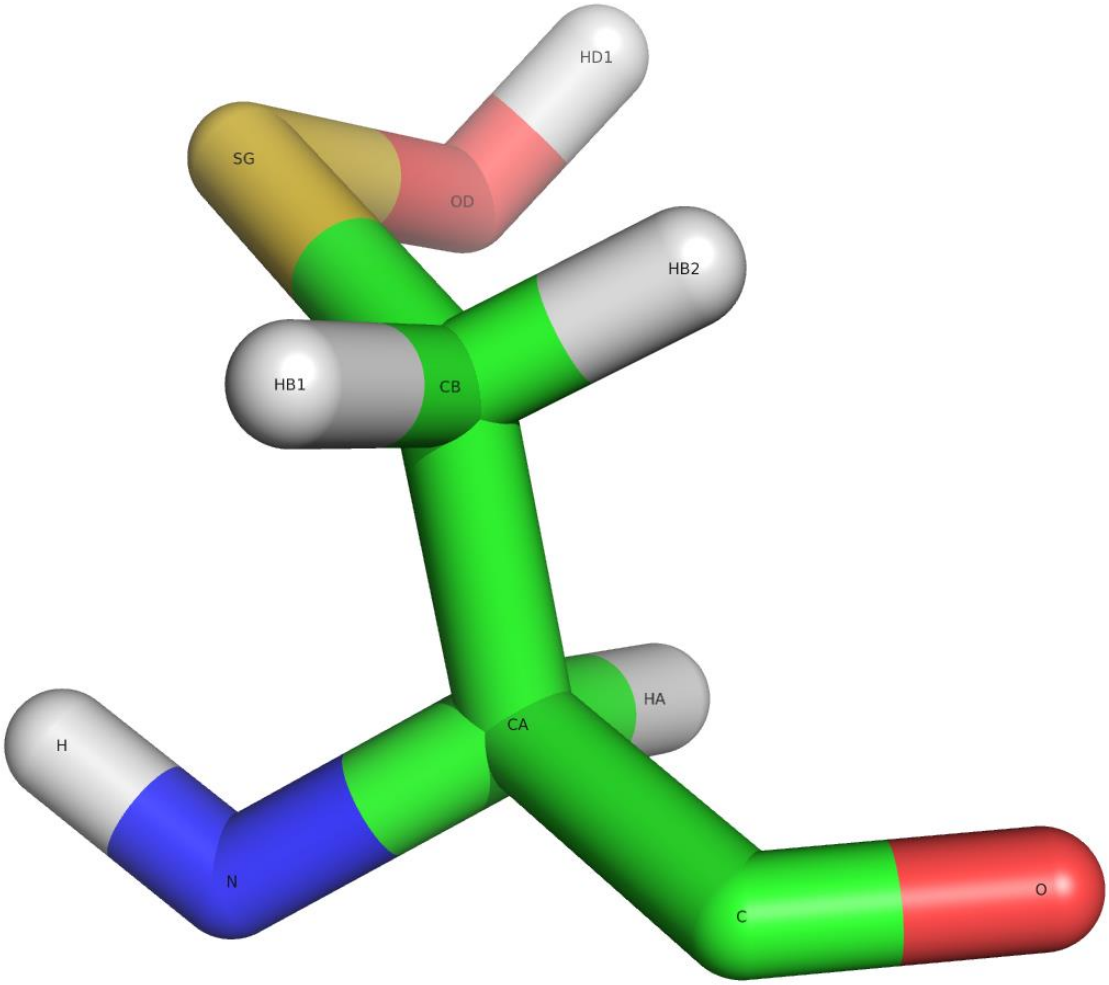
Cysteine Modifications
Cysteinepersulfide (CSS) α-conformer



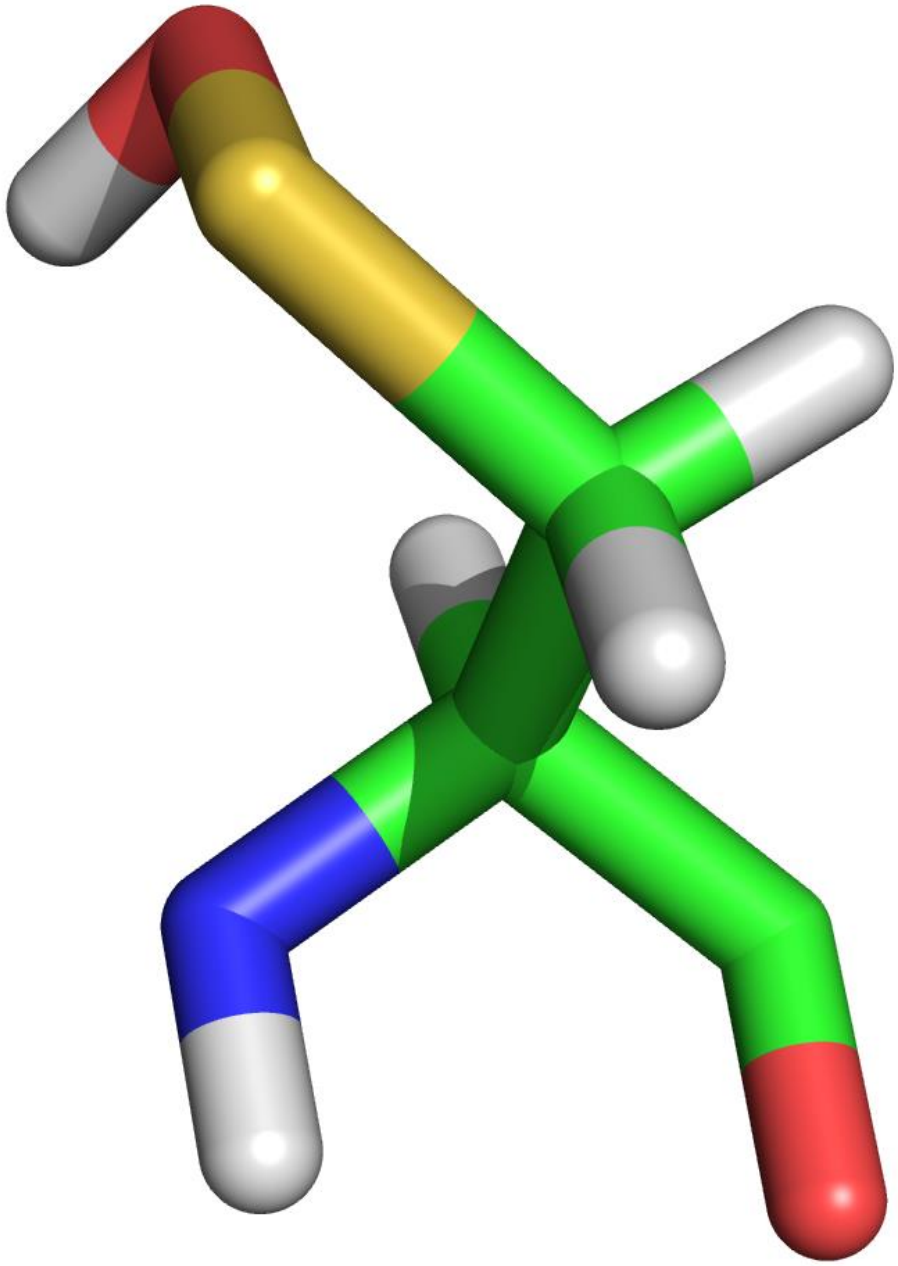
Cysteinepersulfide (CSS) β-conformer



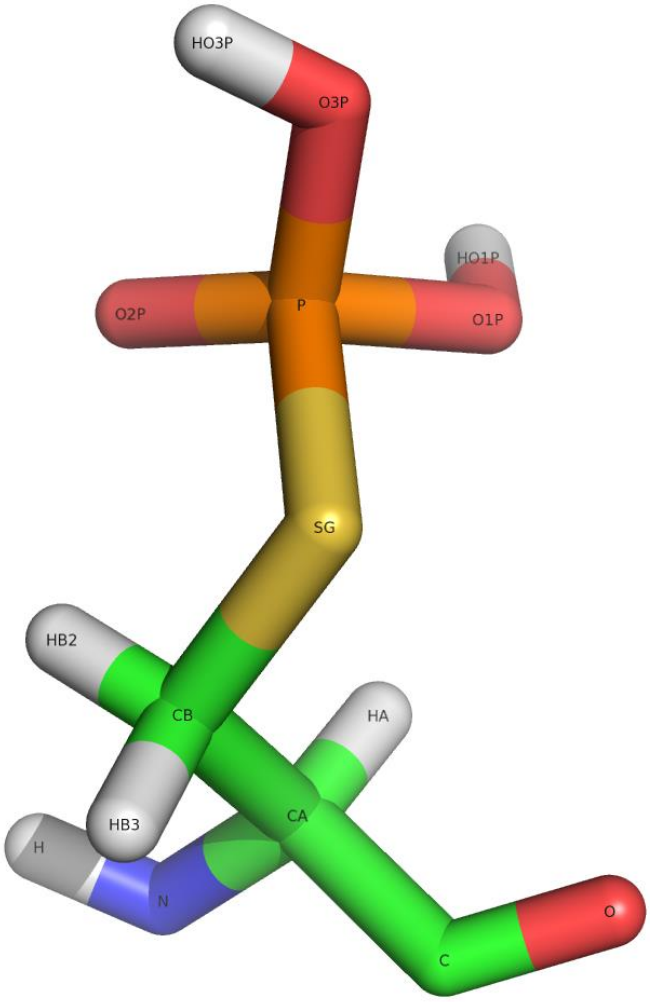
Cysteine sulfenic acid (CSO) α-conformer



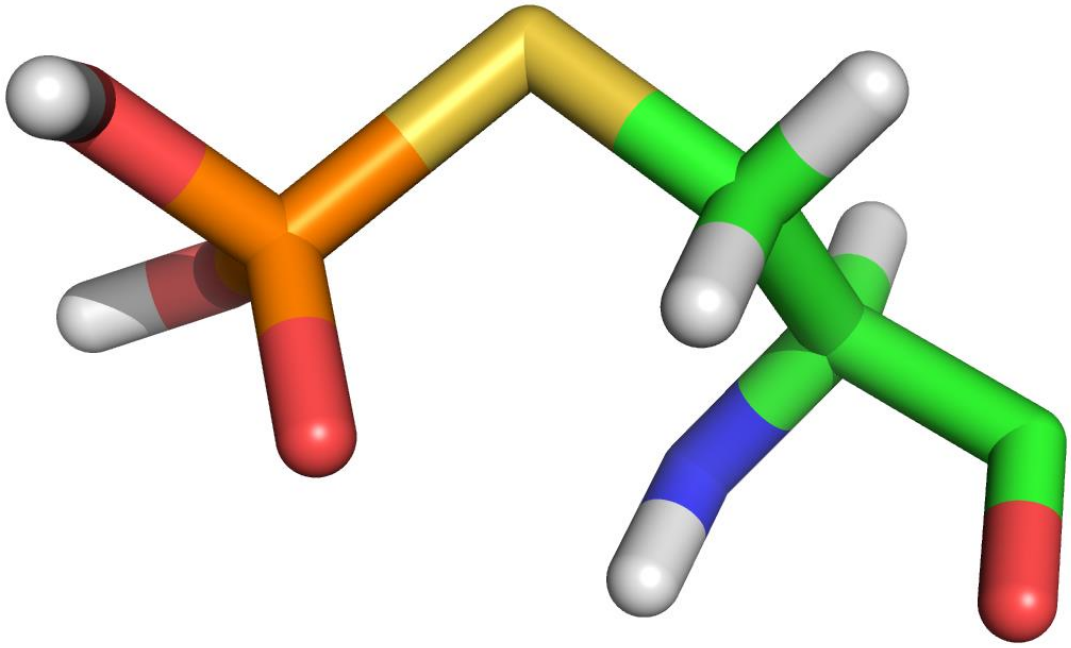
Cysteine sulfenic acid (CSO) β-conformer



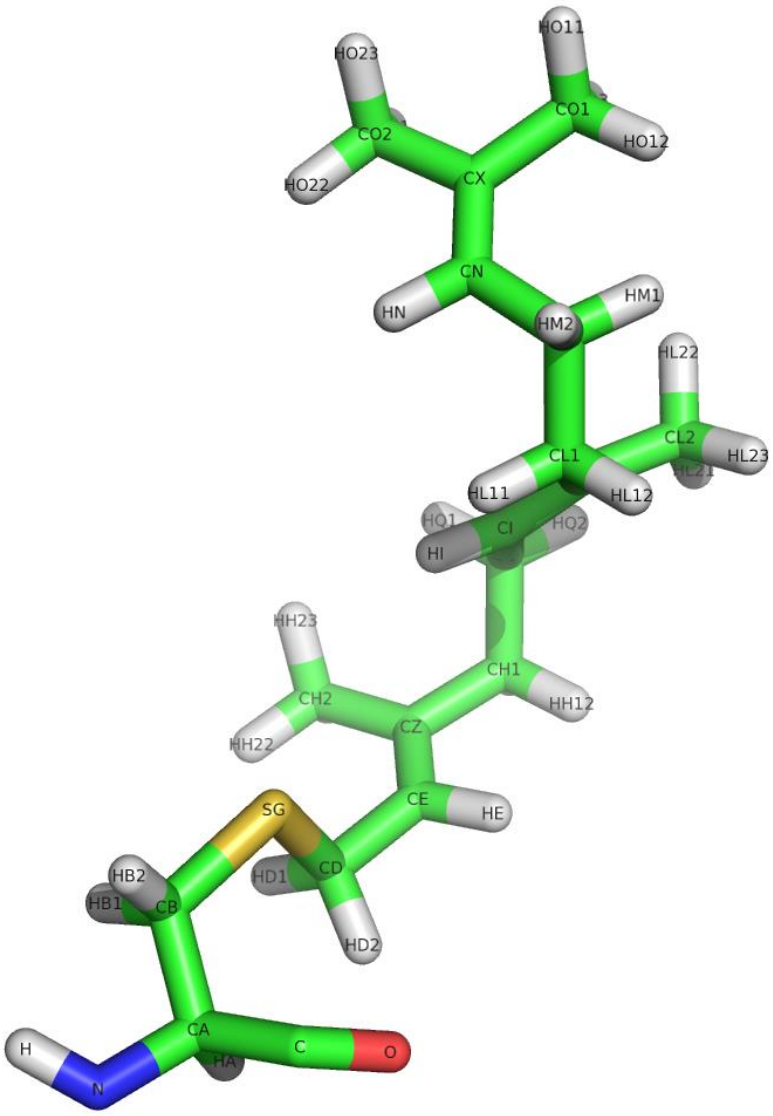
Phosphocysteine (neutral) (CSP) α-conformer



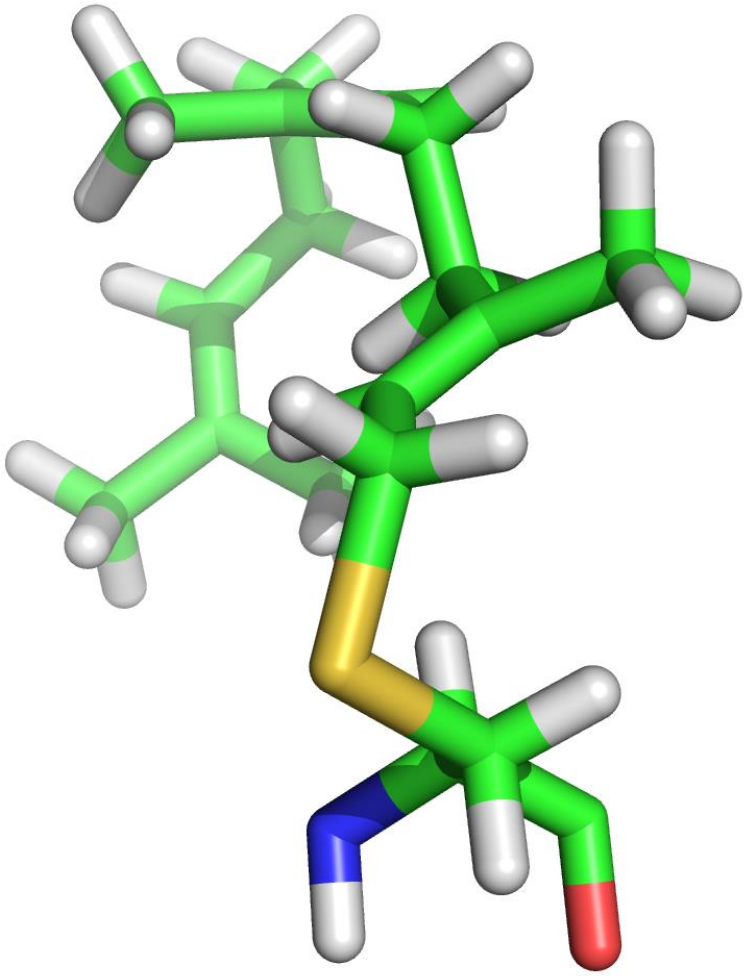
Phosphocysteine (neutral) (CSP) β-conformer



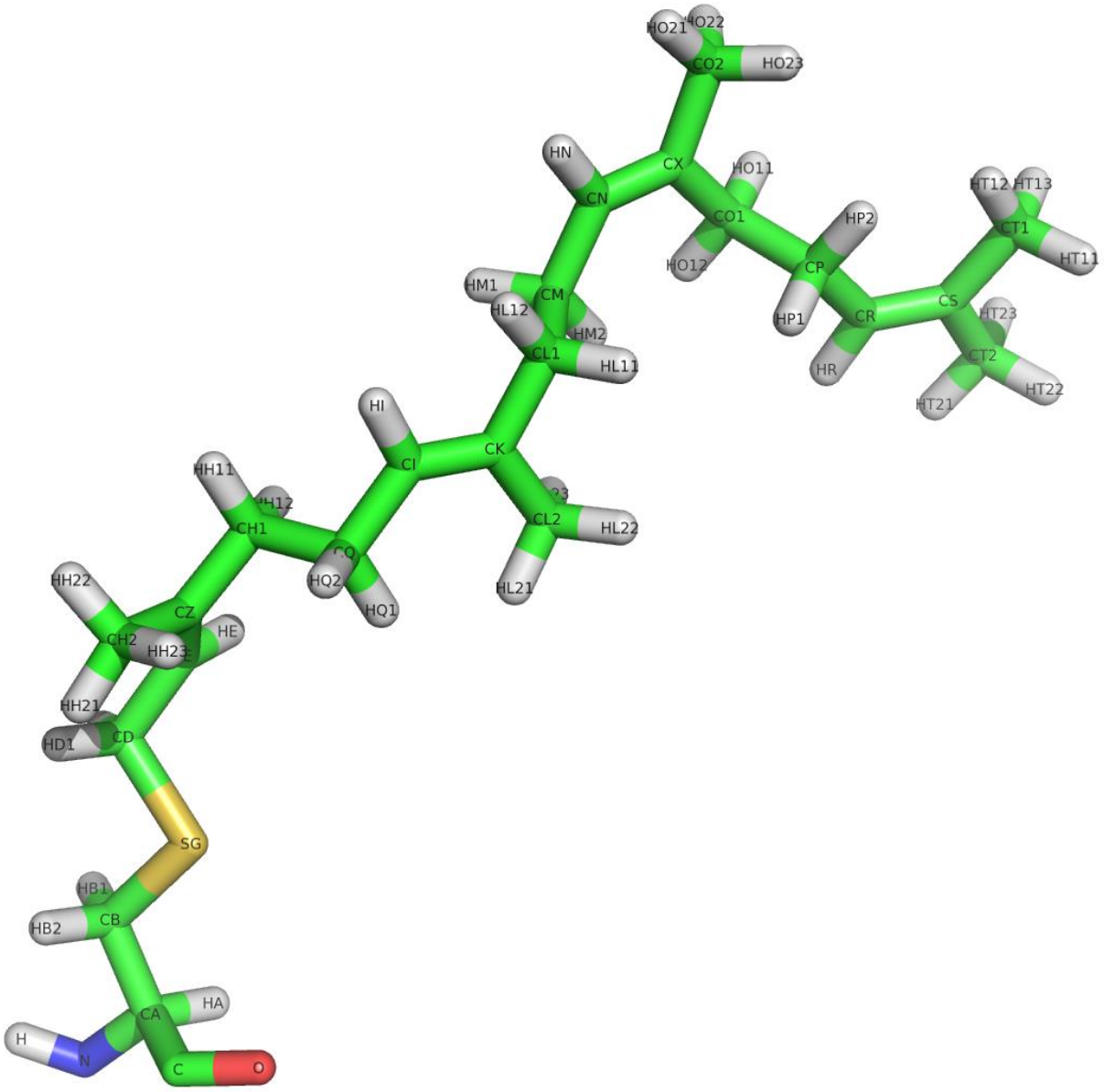
S-farnesylcysteine (FCY) α-conformer



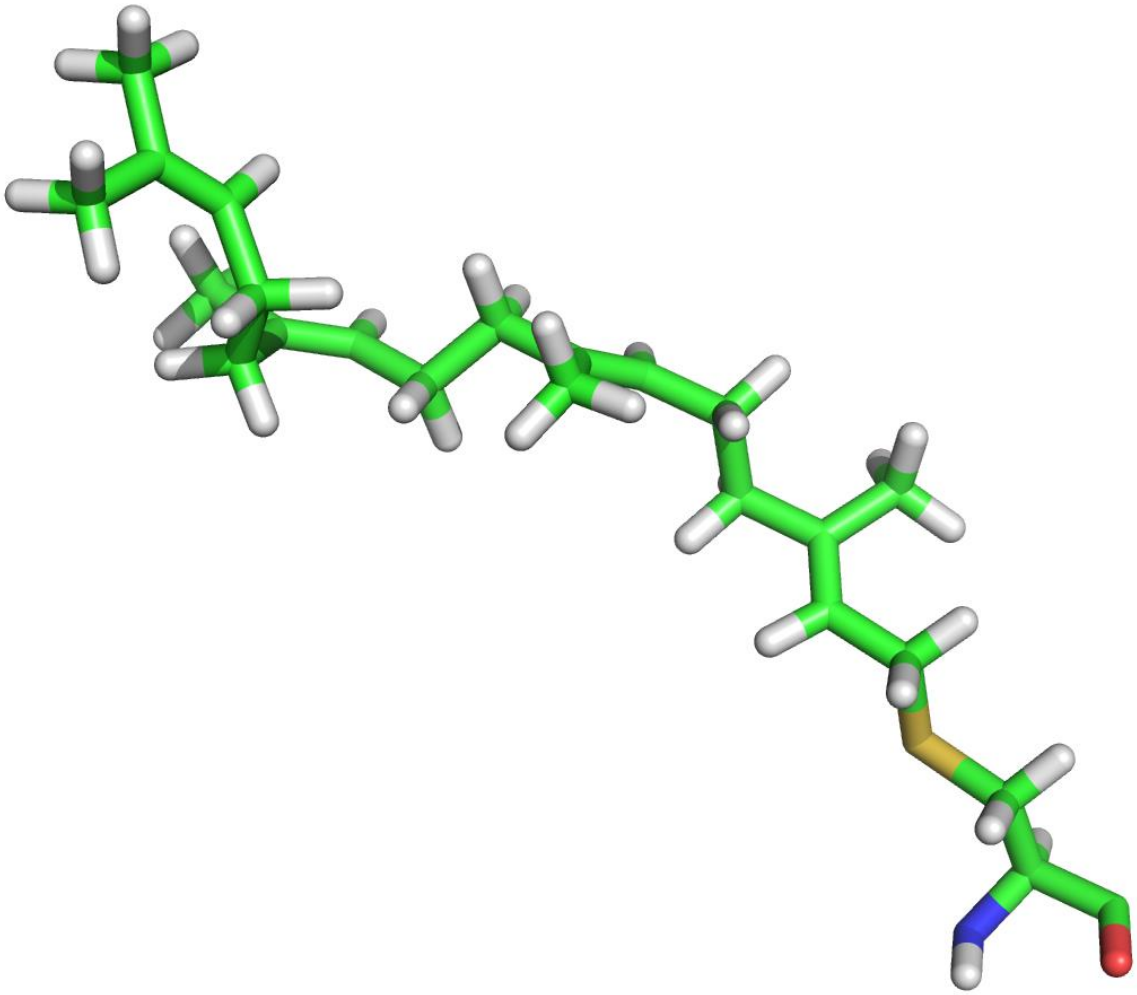
S-farnesylcysteine (FCY) β-conformer



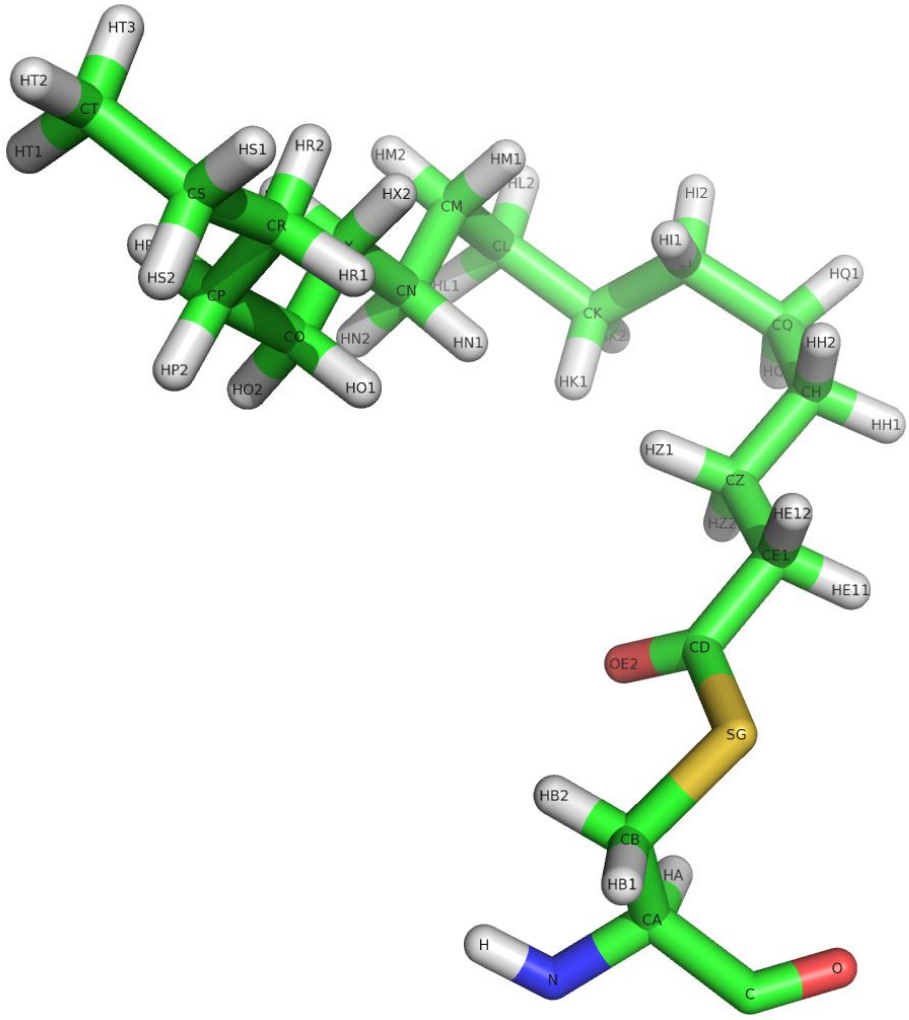
S-geranylgeranylcysteine (GCY) α-conformer



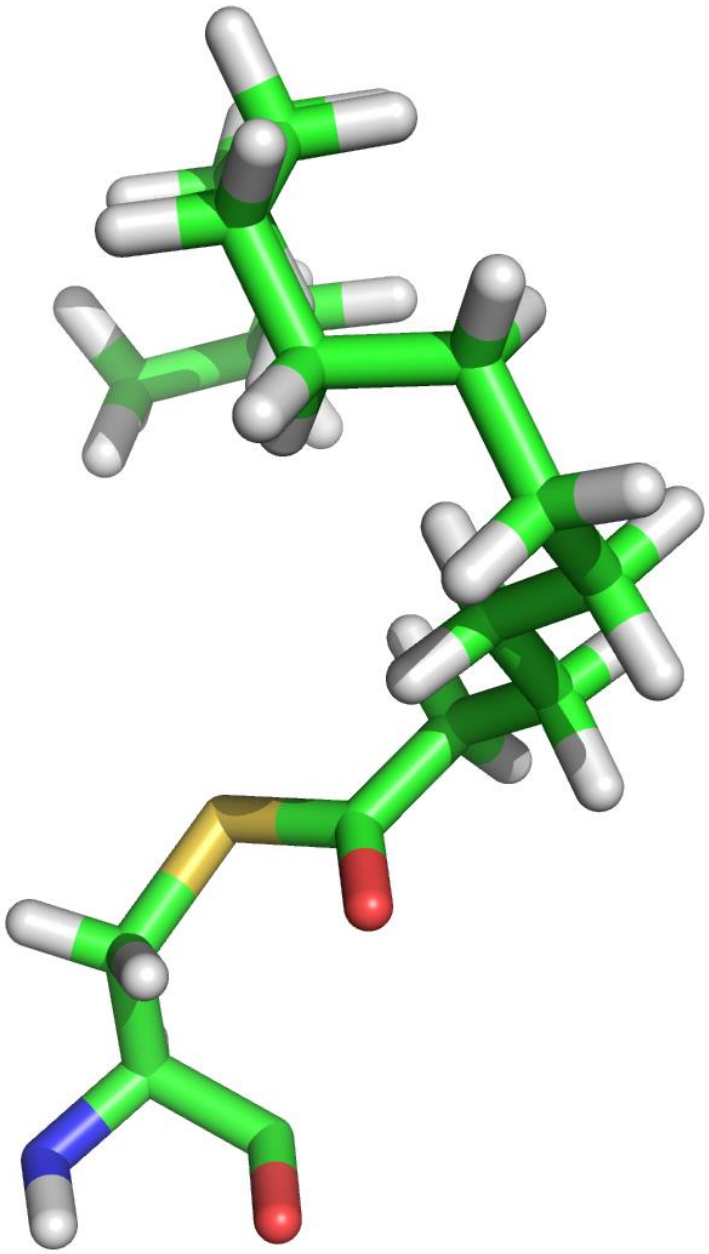
S-geranylgeranylcysteine (GCY) β-conformer



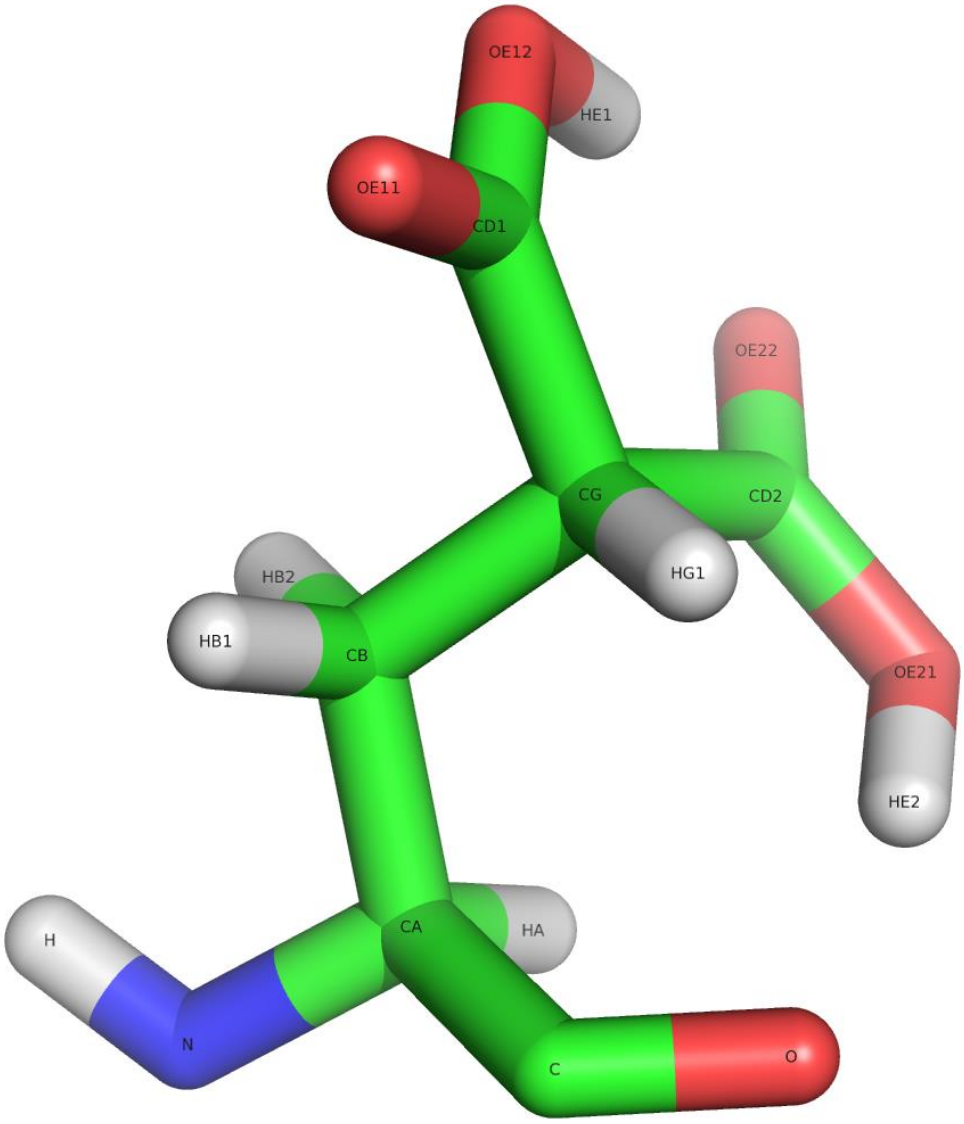
S-palmitoylcysteine (SPC) α-conformer



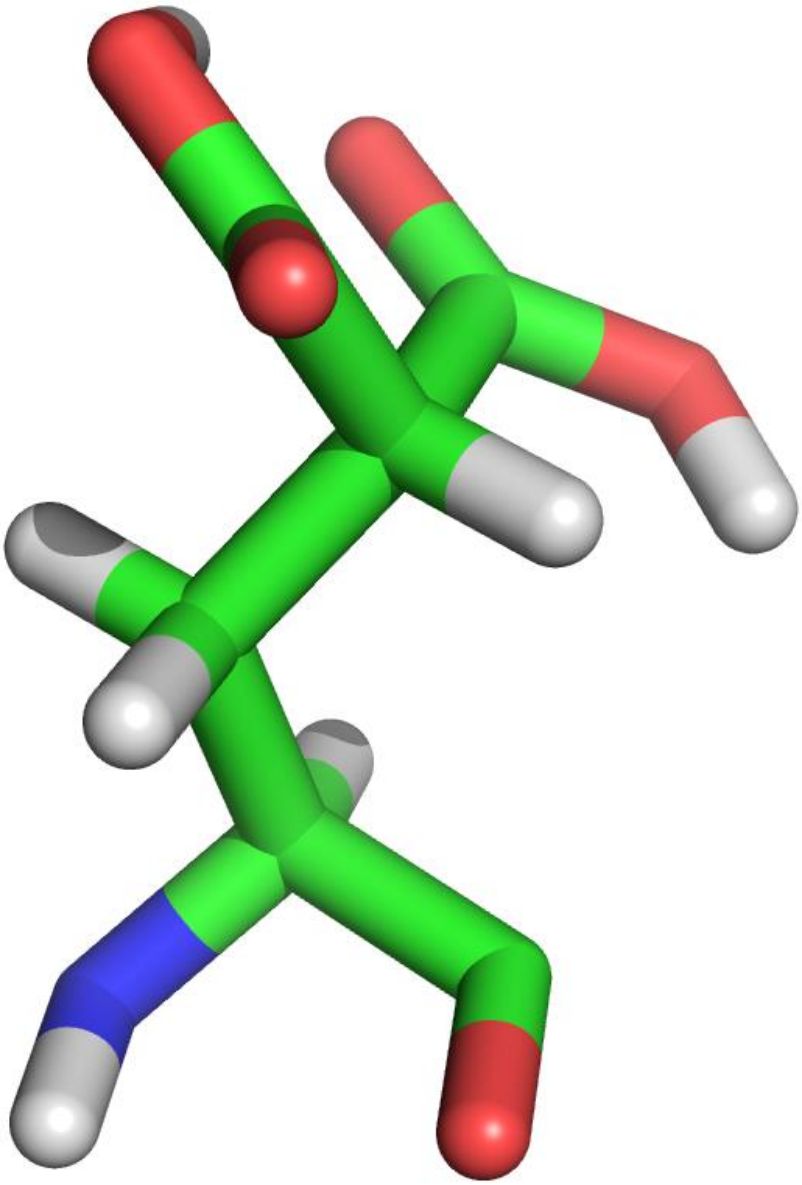
S-palmitoylcysteine (SPC) β-conformer



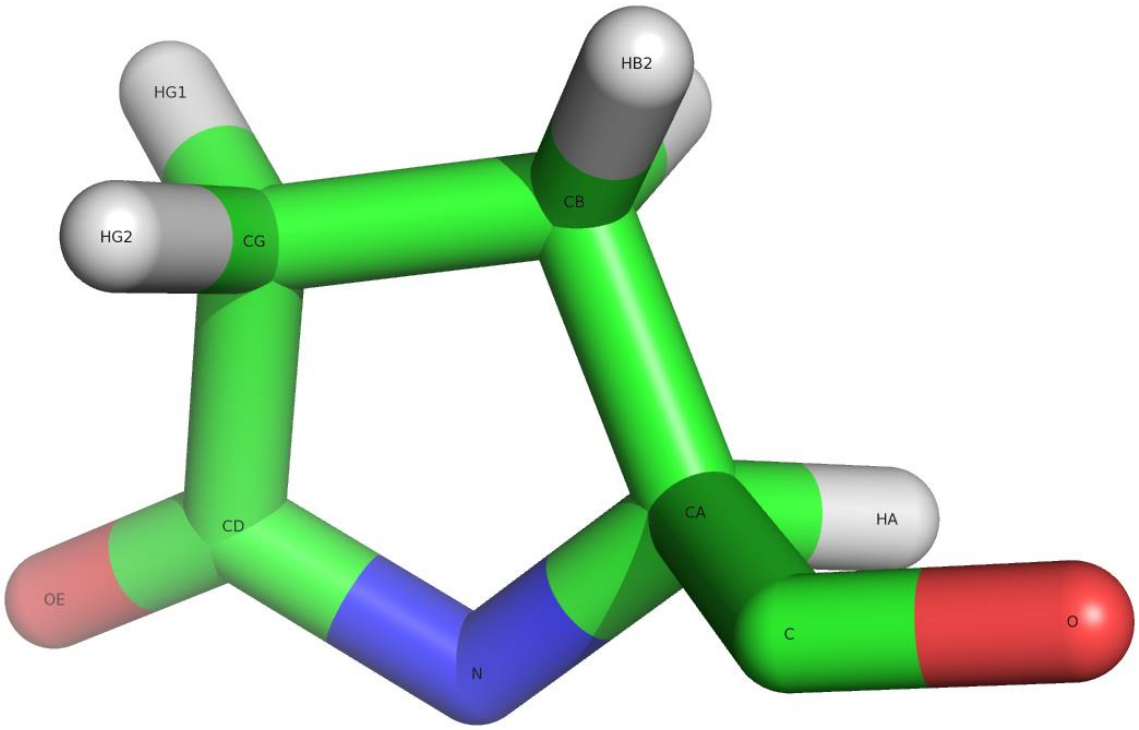
Glutamic Acid Modifications
1-carboxyglutamic acid (neutral) (CGU) α-conformer



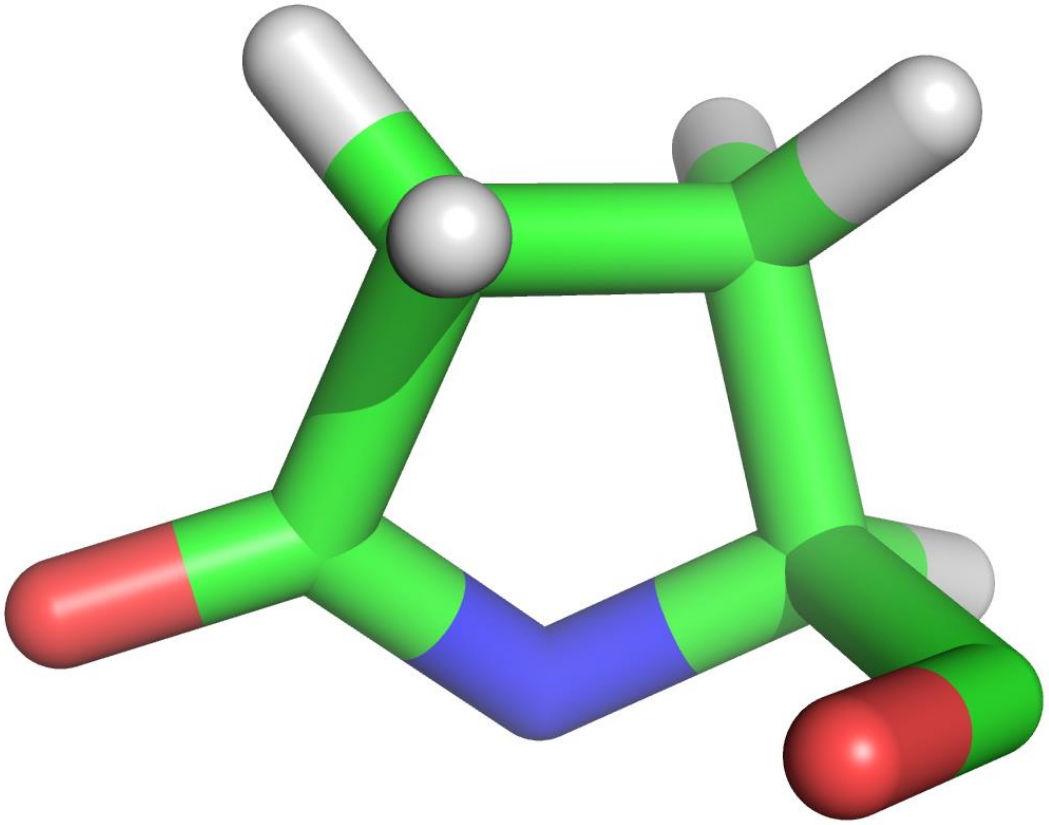
1-carboxyglutamic acid (neutral) (CGU) β-conformer



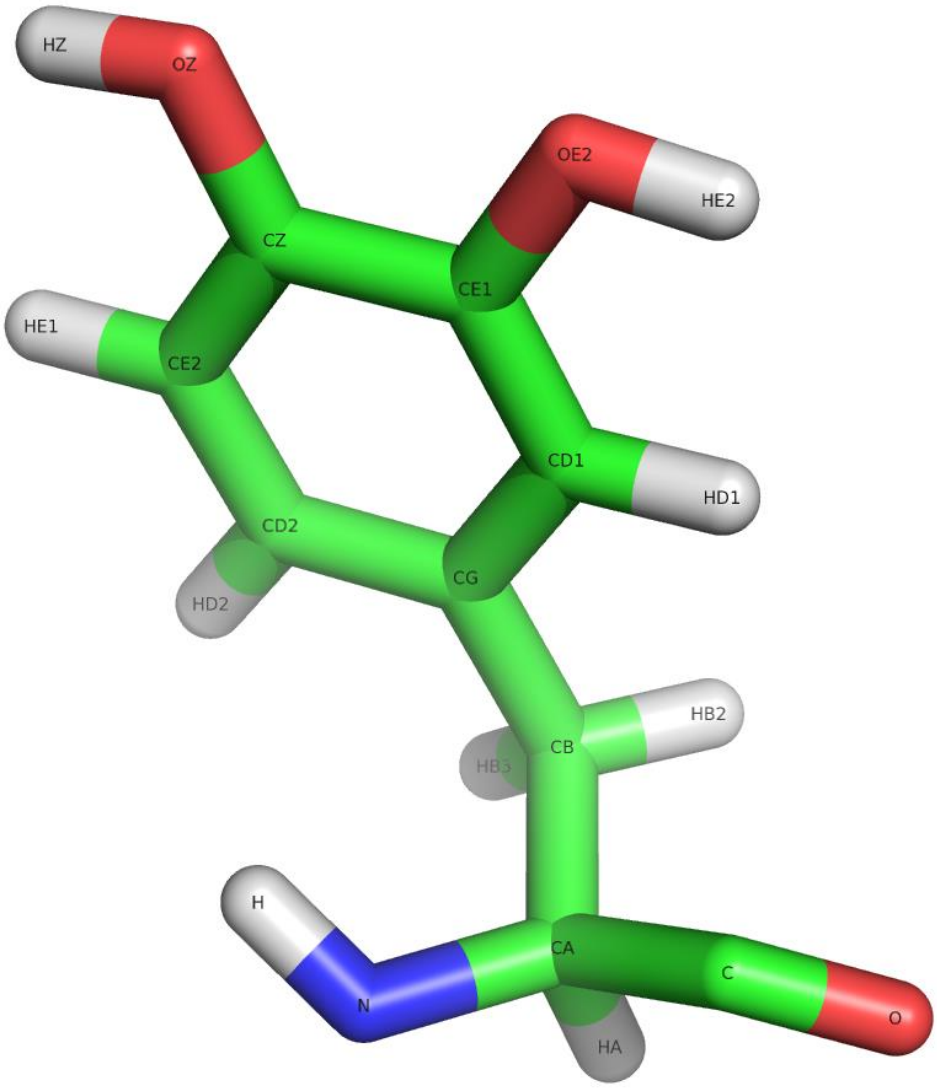
Pyroglutamic Acid (PCA) α-conformer



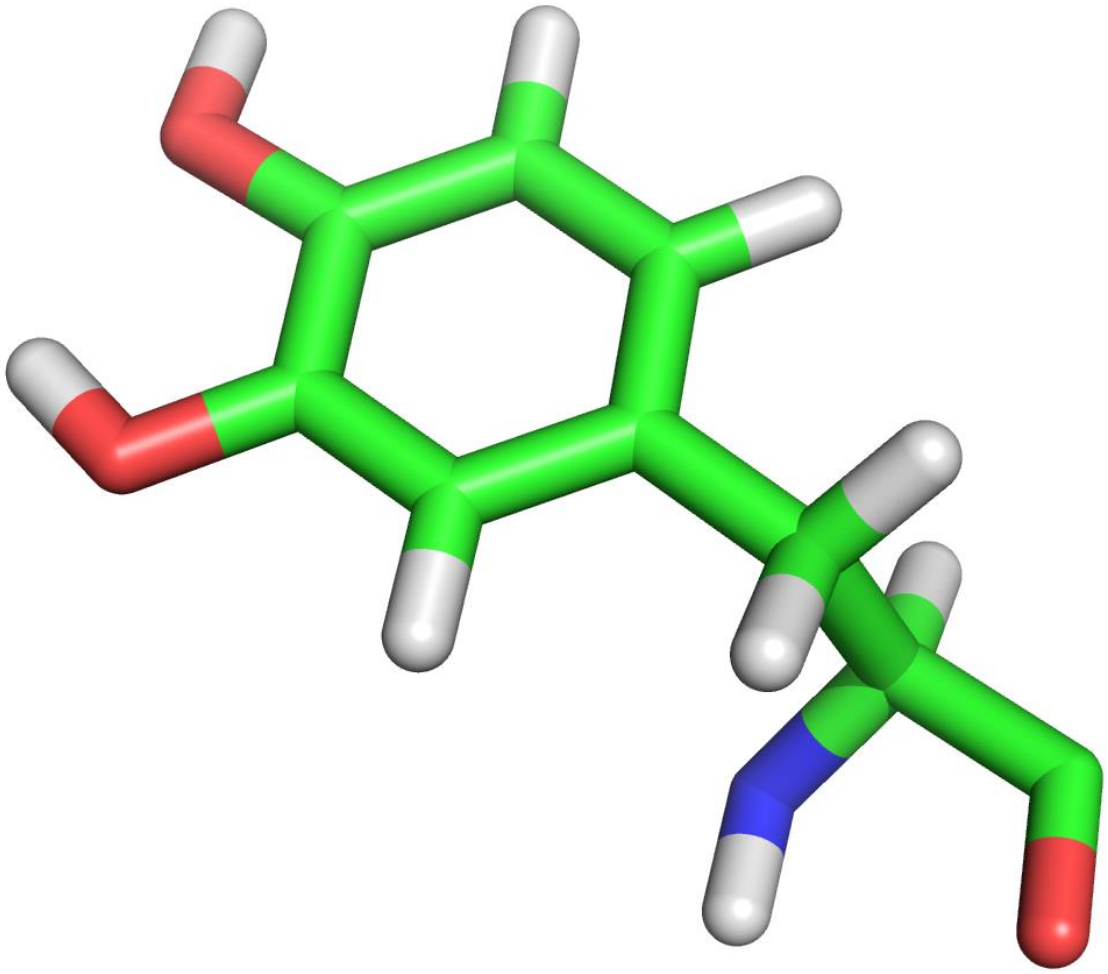
Pyroglutamic Acid (PCA) β-conformer



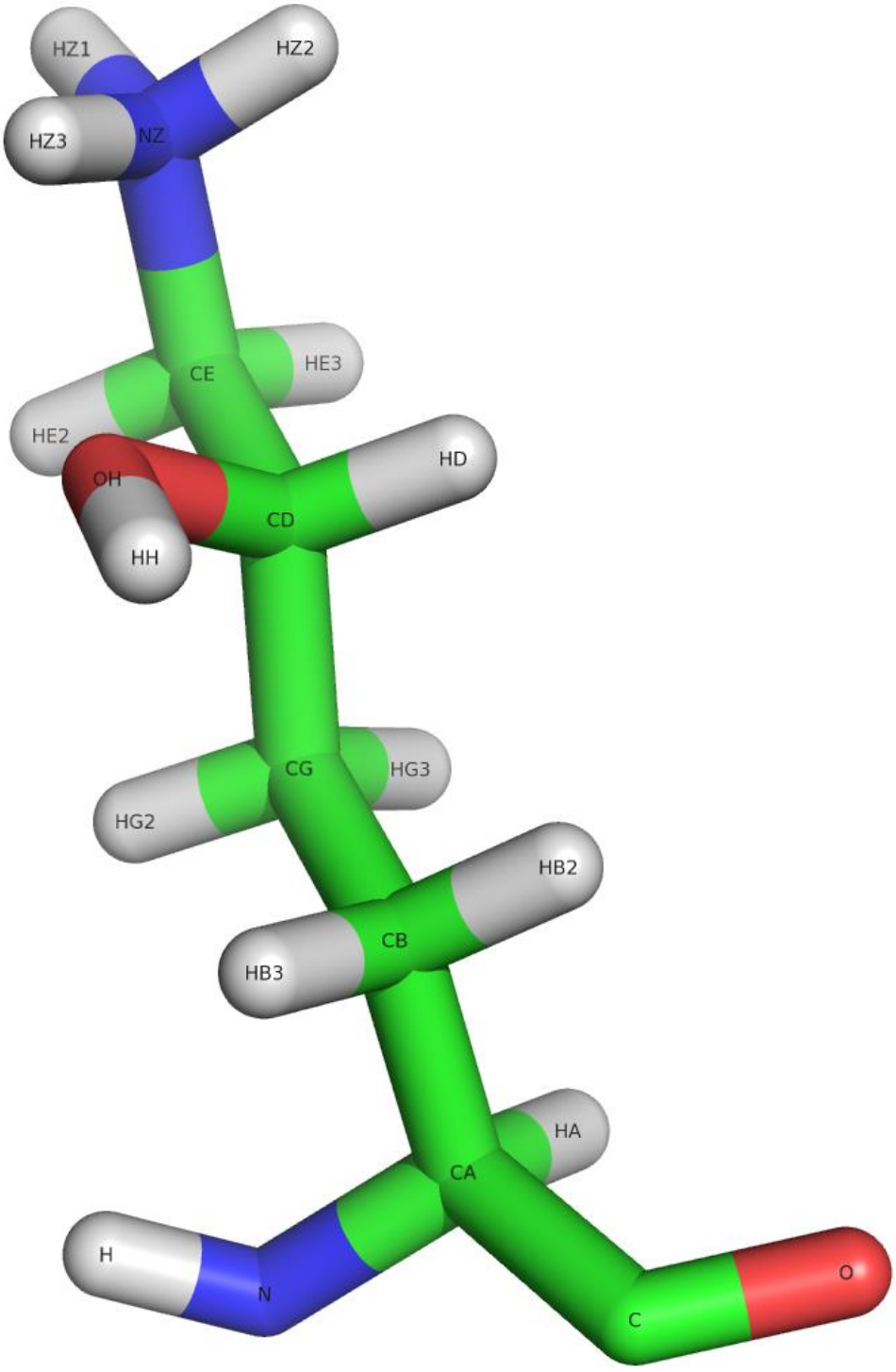
Phenylalanine Modifications
Dihydroxyphenylalanine (DAH) α-conformer



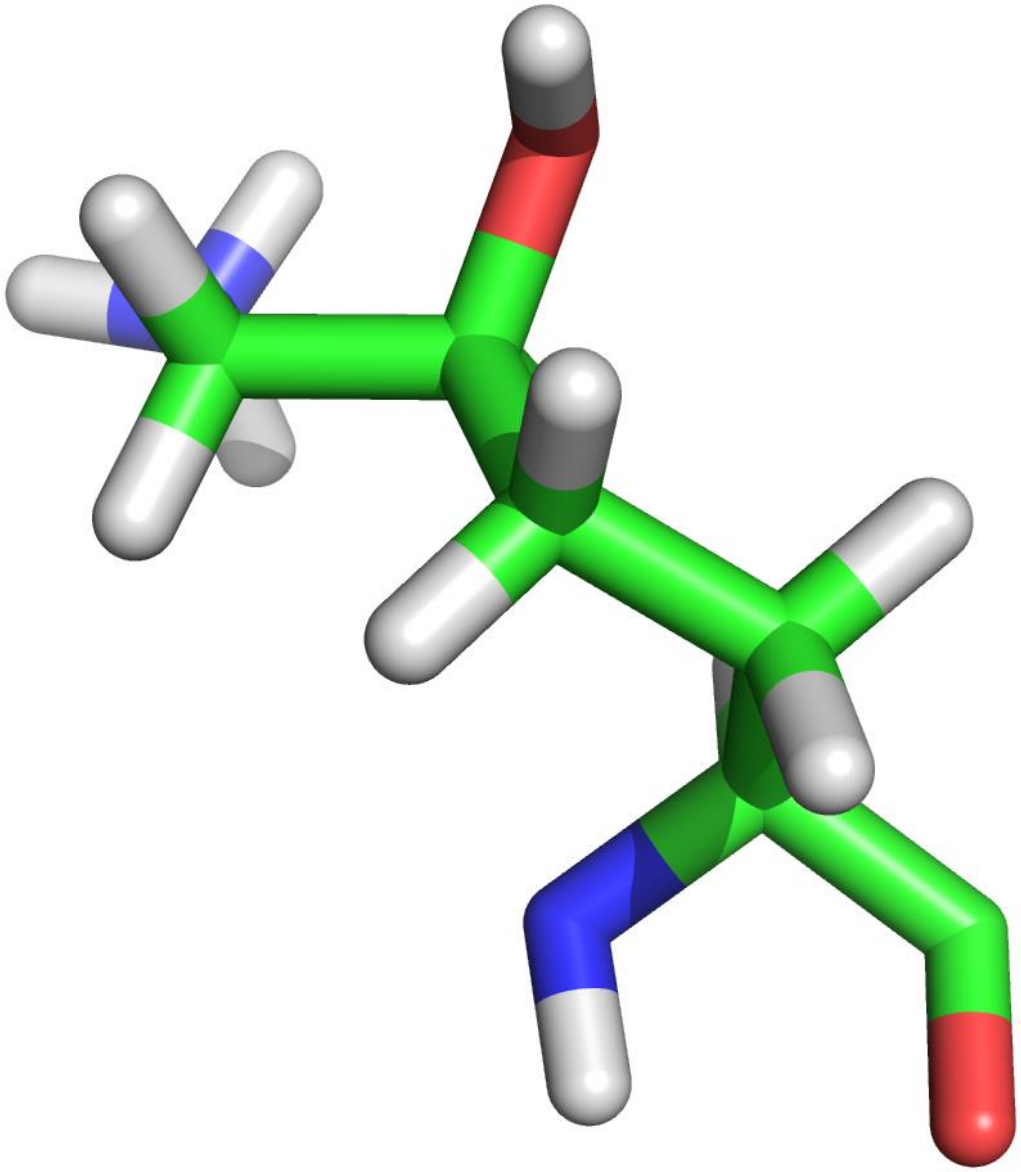
Dihydroxyphenylalanine (DAH) β-conformer



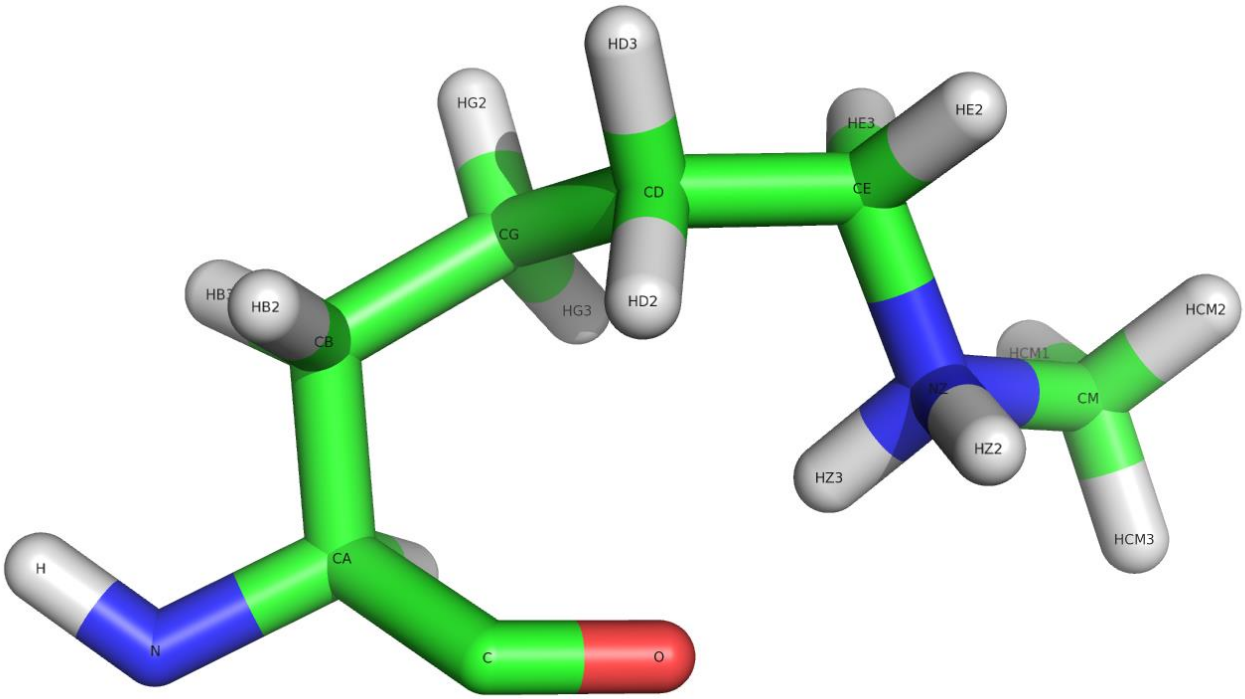
Lysine Modifications
5-hydroxylysine (LYZ) α-conformer



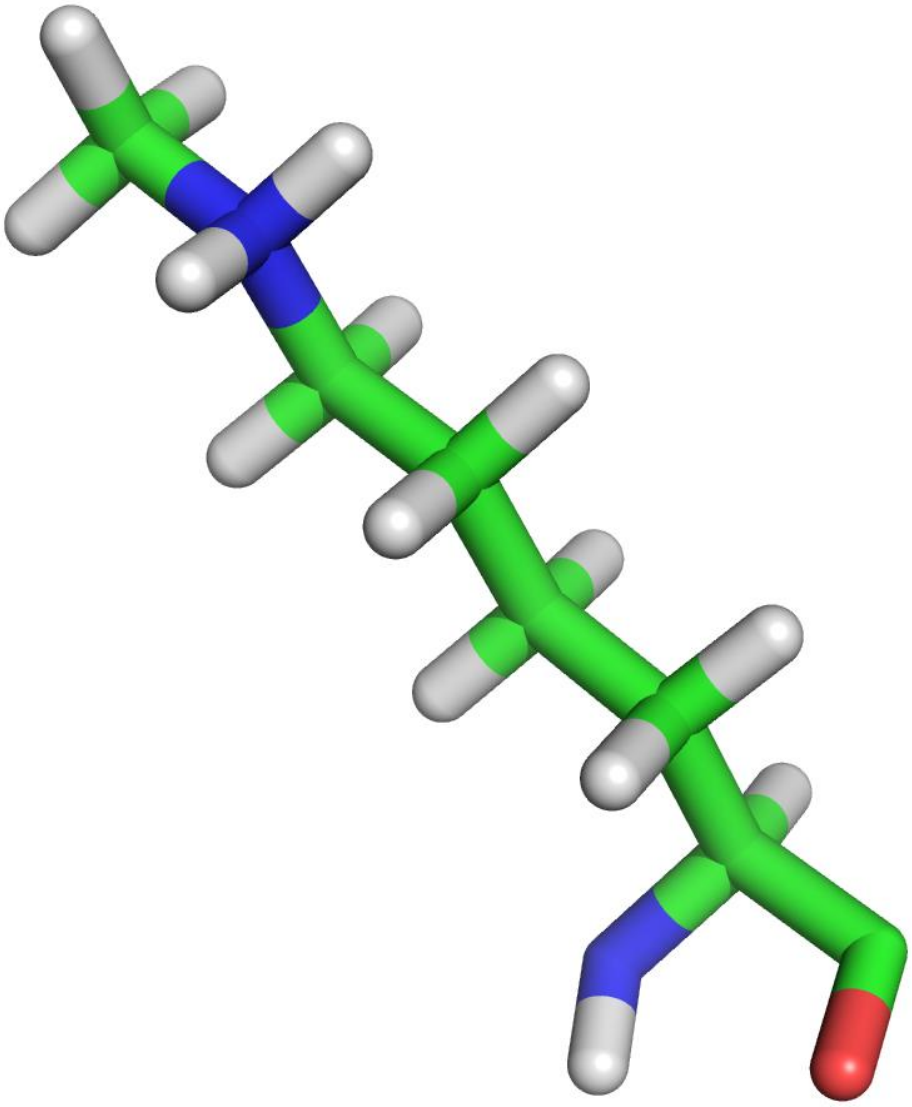
5-hydroxylysine (LYZ) β-conformer



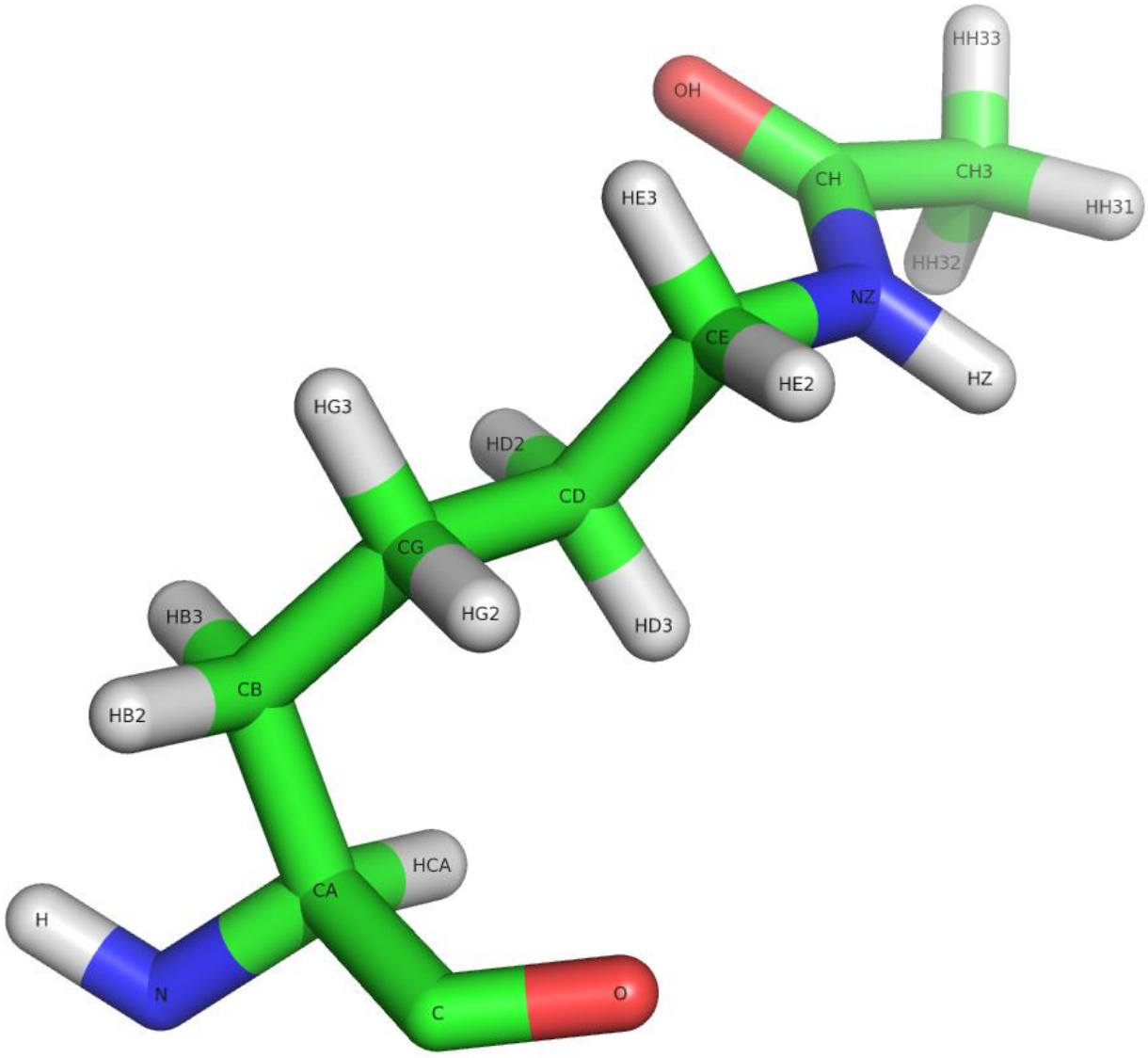
ε-N-methyllysine (MLZ) α-conformer



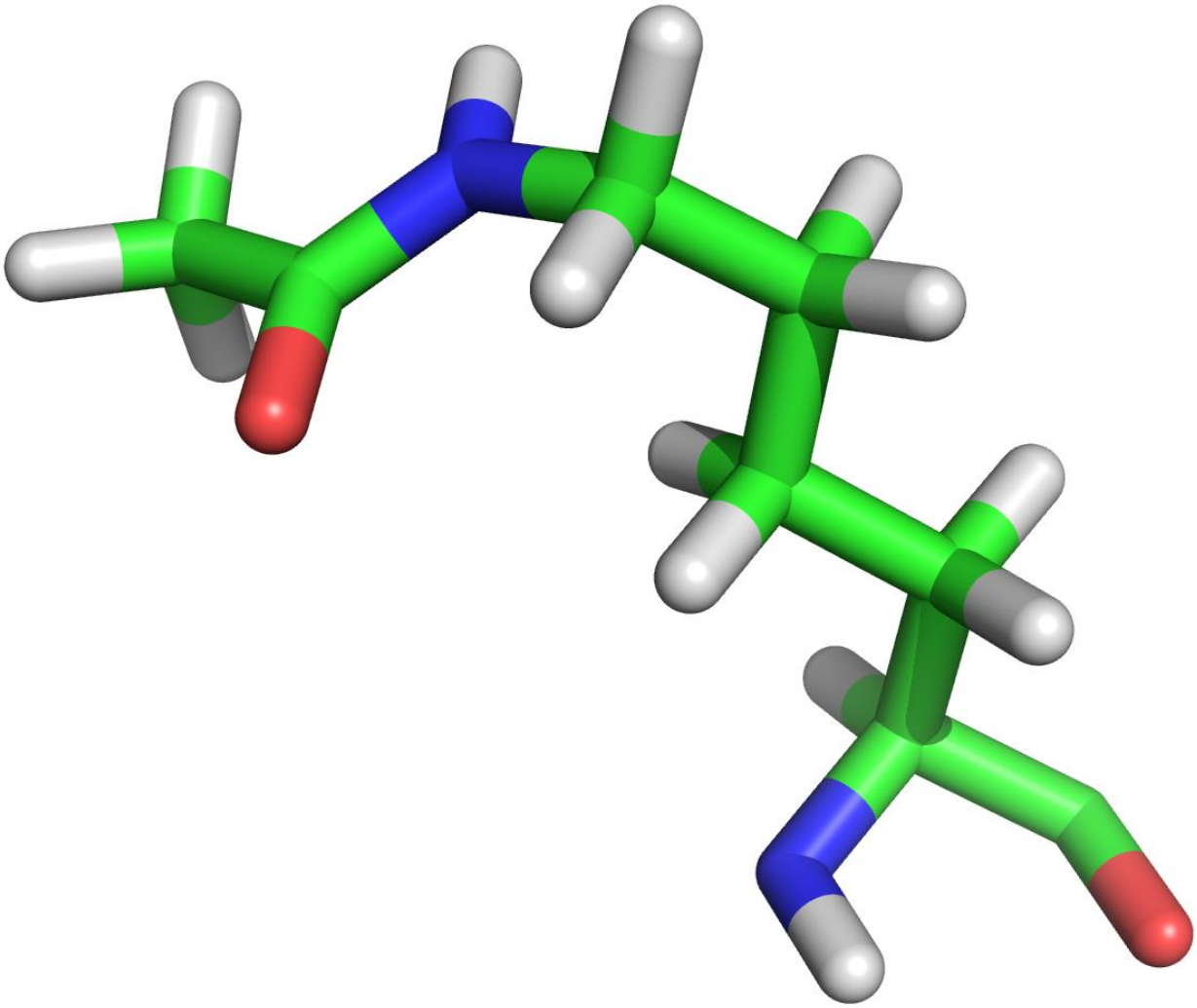
ε-N-methyllysine (MLZ) β-conformer



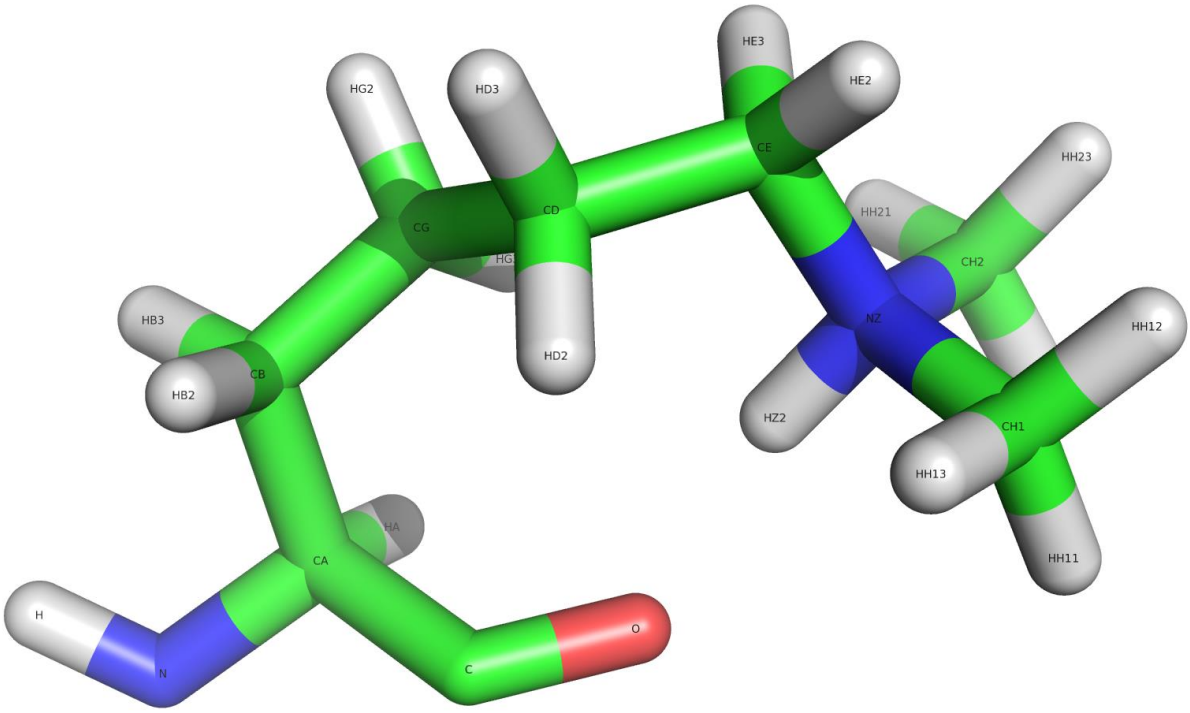
N-ε-acetyllysine (ALY) α-conformer



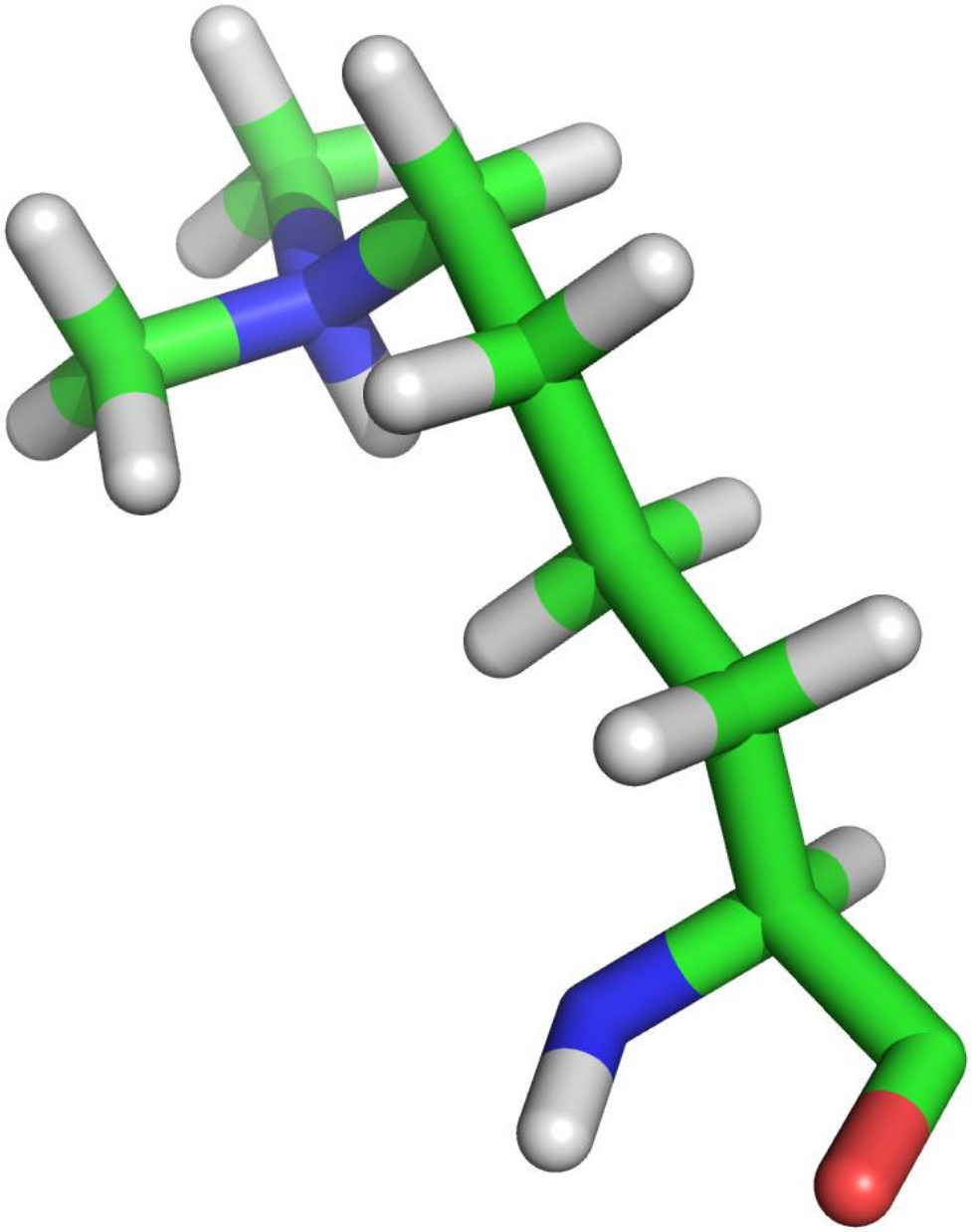
N-ε-acetyllysine (ALY) β-conformer



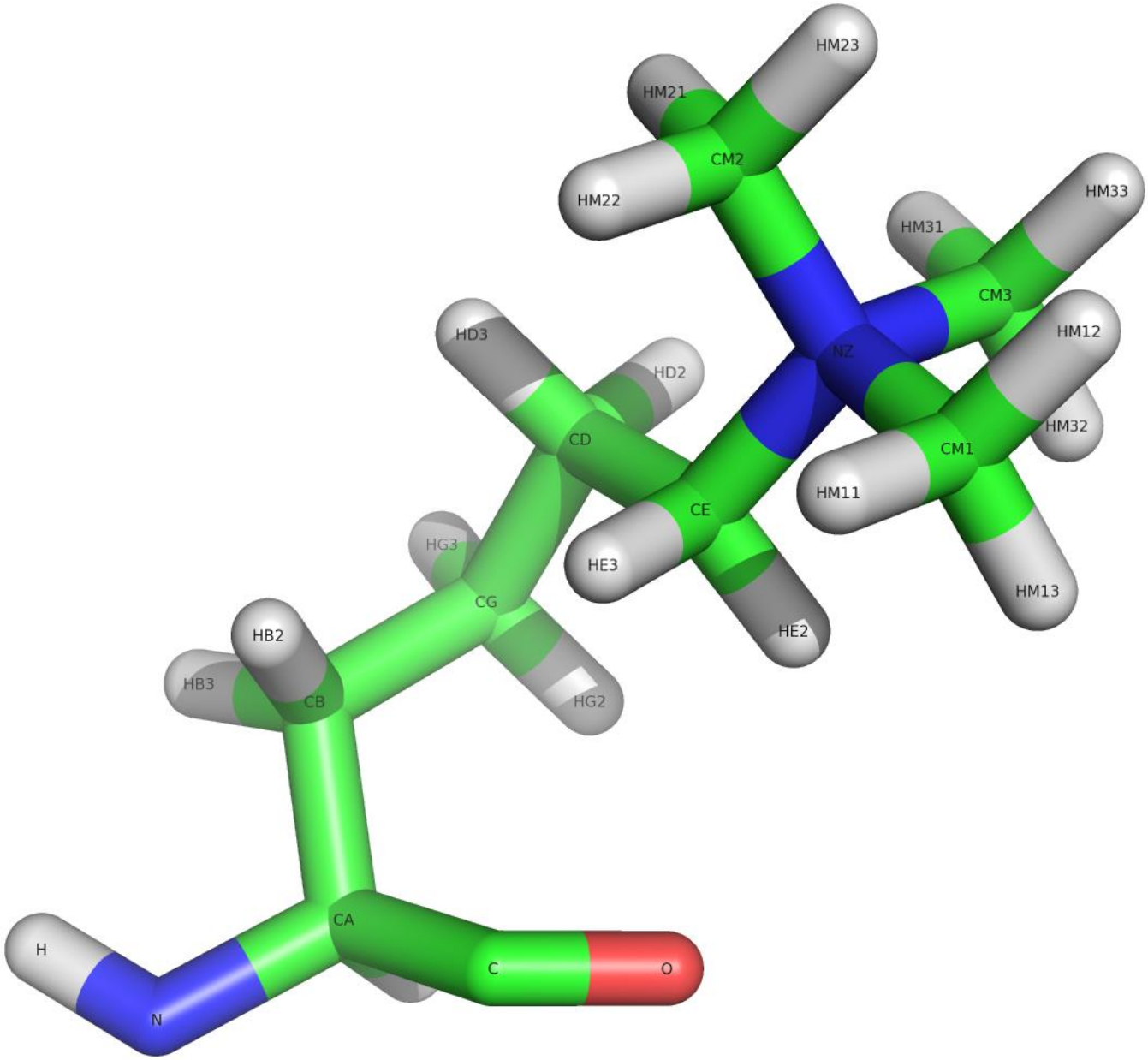
N6,N6-dimethyllysine (MLY) α-conformer



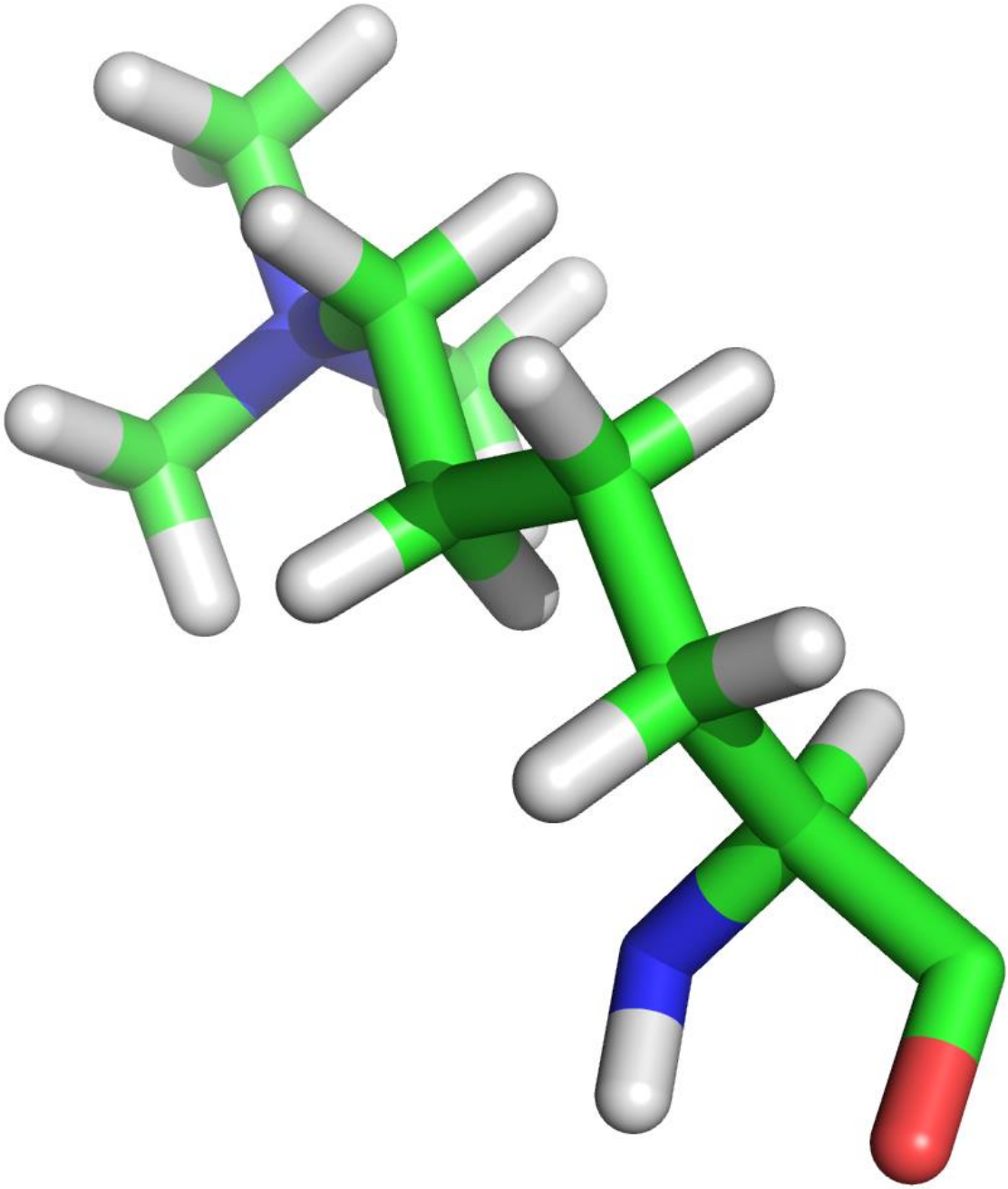
N6,N6-dimethyllysine (MLY) β-conformer



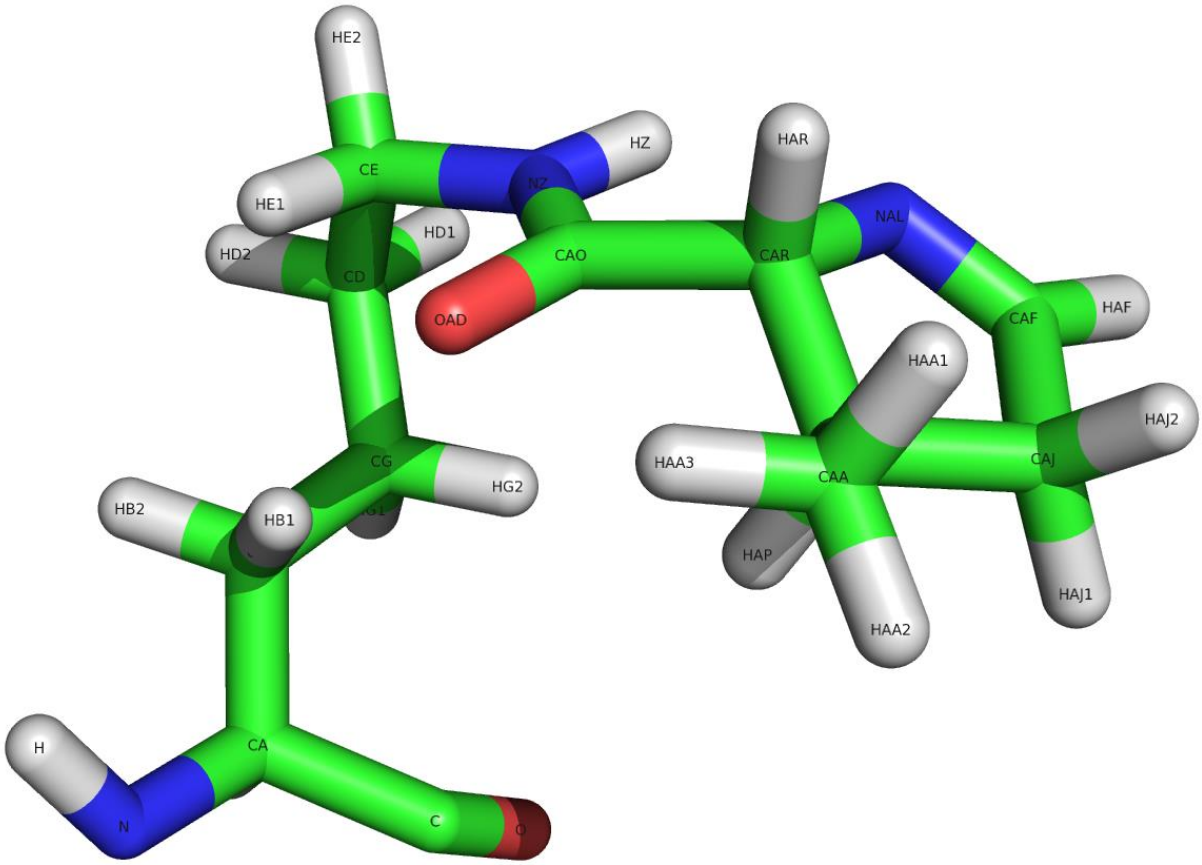
N6,N6,N6-trimethyllysine (M3L) α-conformer



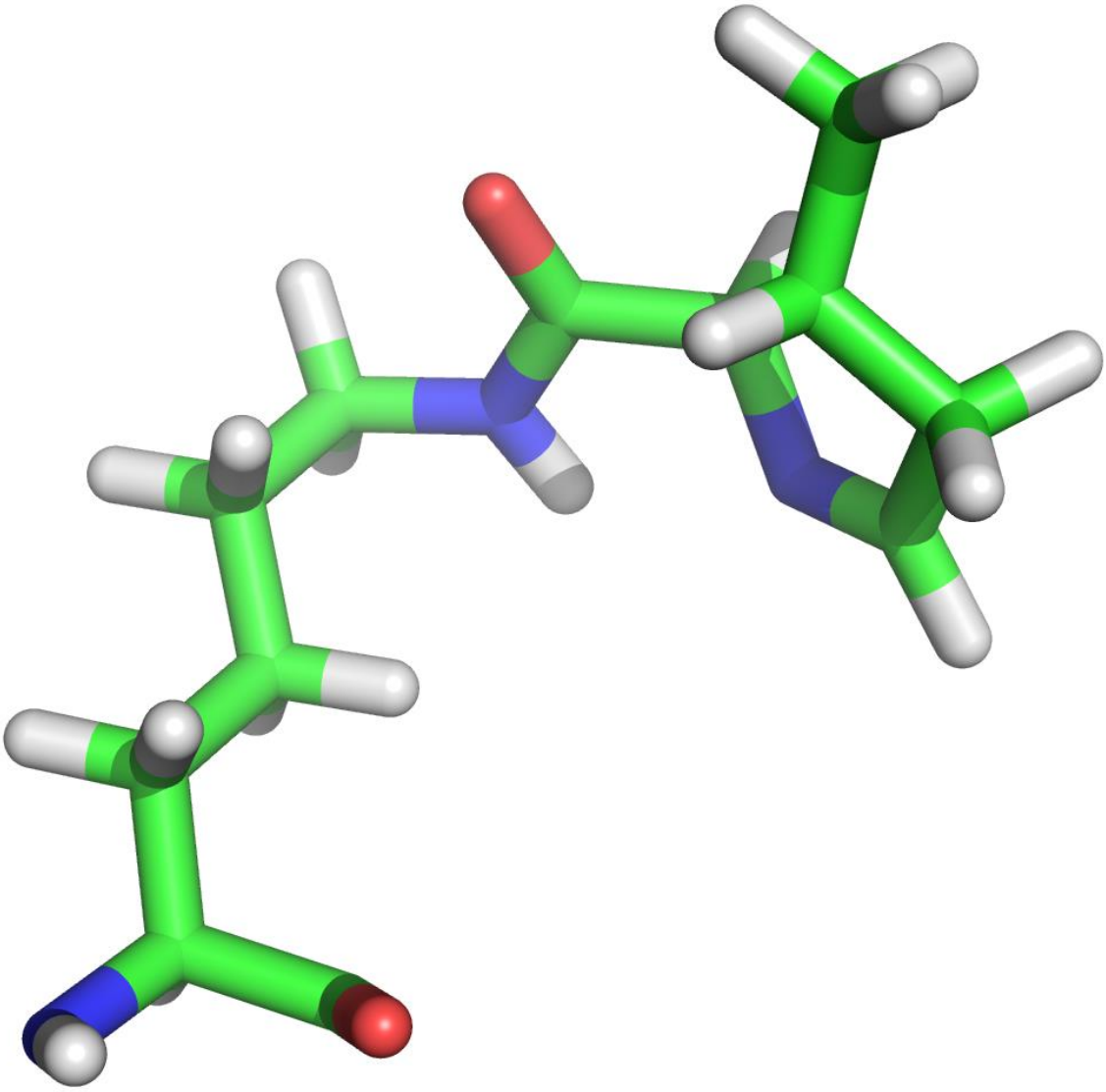
N6,N6,N6-trimethyllysine (M3L) β-conformer



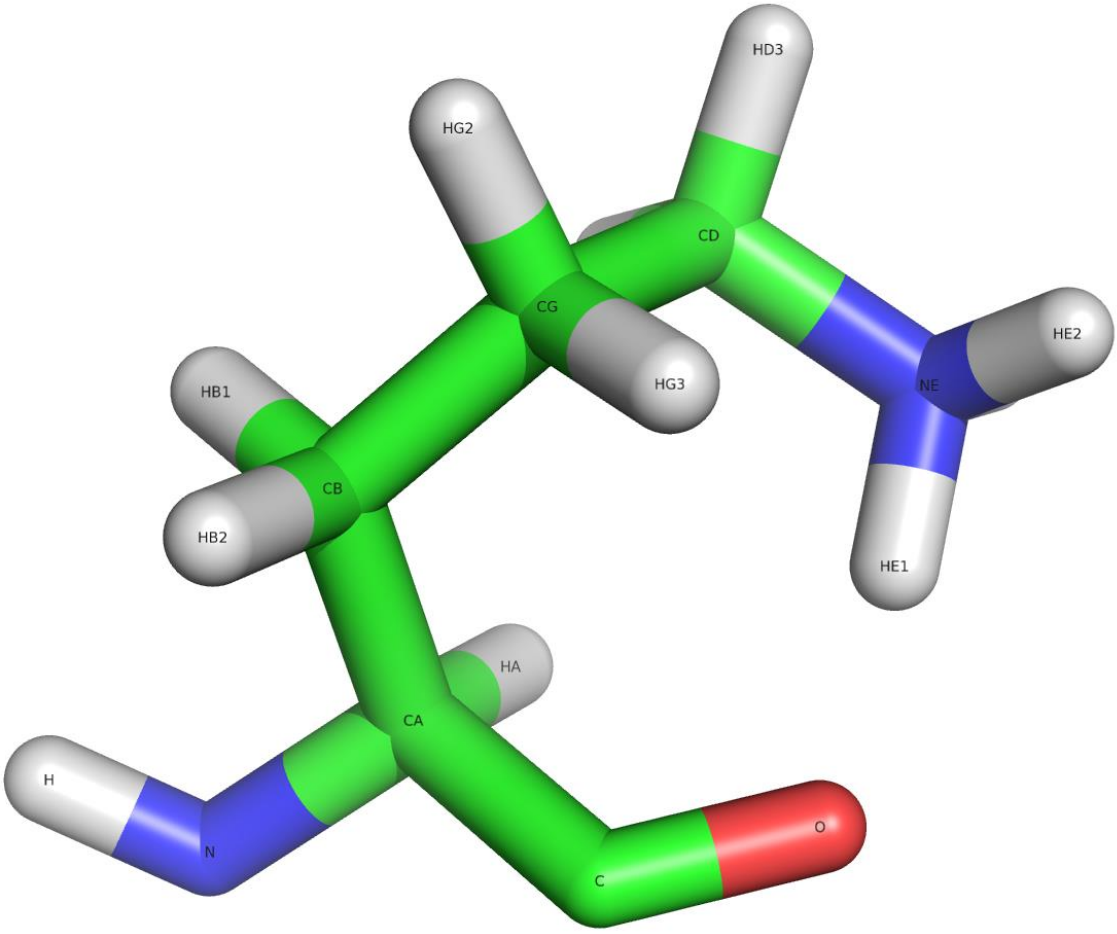
Pyrrolysine (PYH) α-conformer



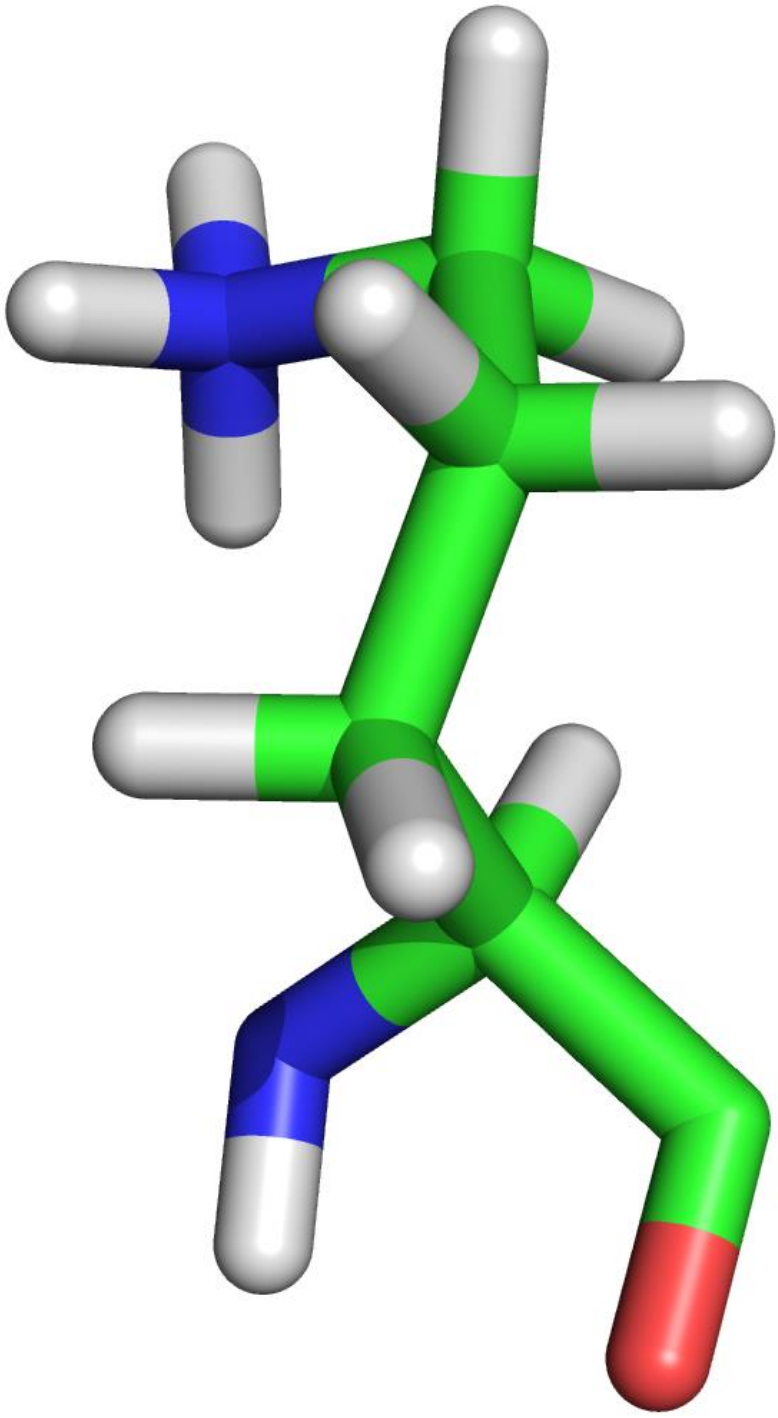
Pyrrolysine (PYH) β-conformer



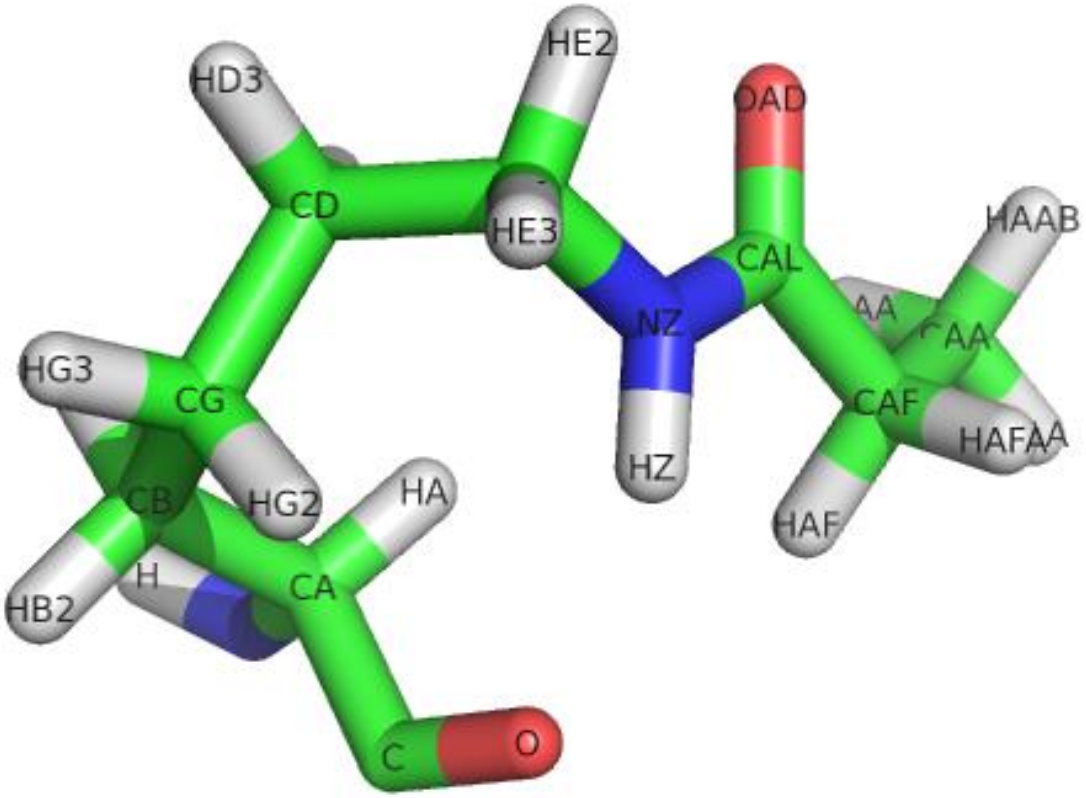
Ornithine (ORN) α-conformer



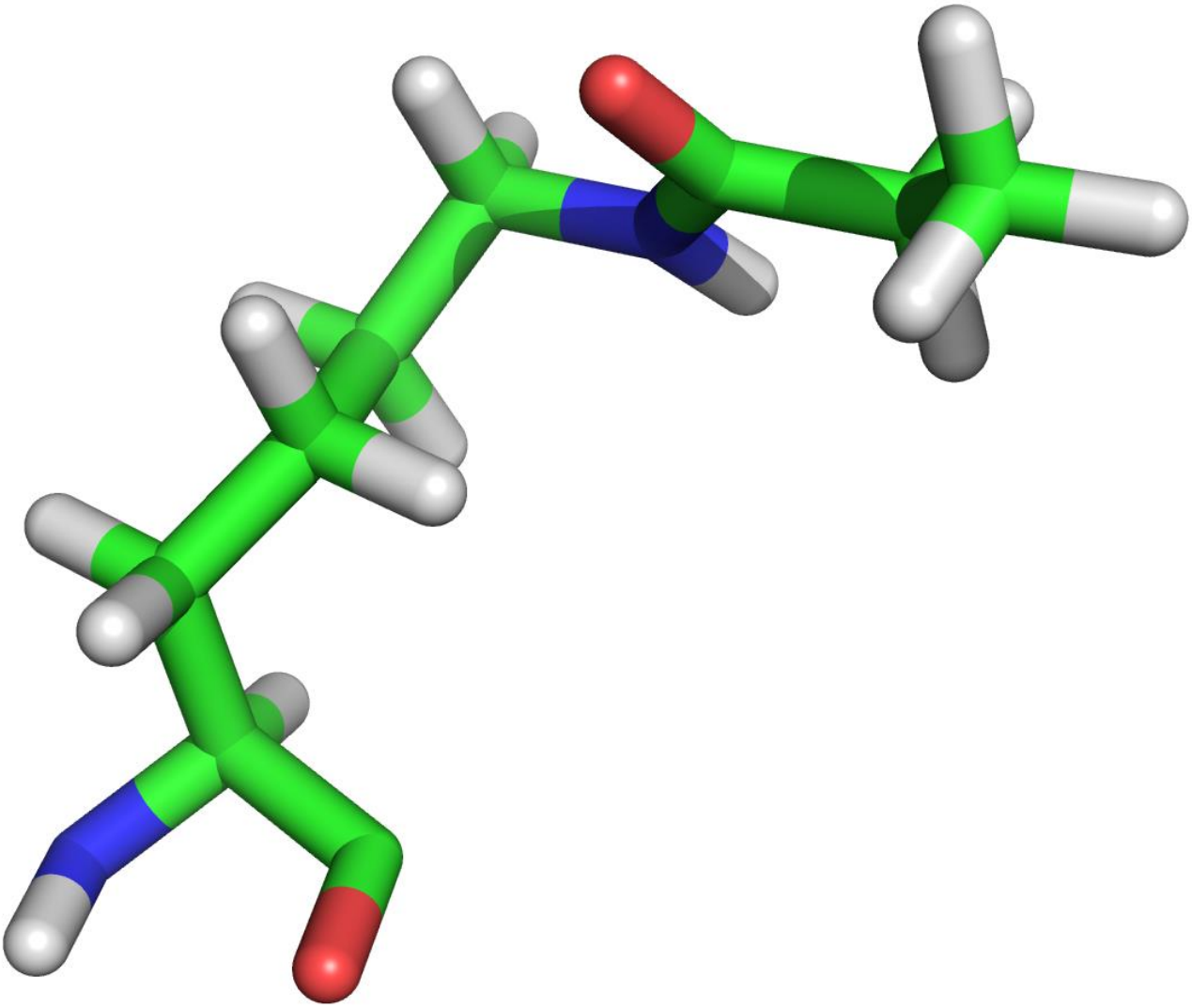
Ornithine (ORN) β-conformer



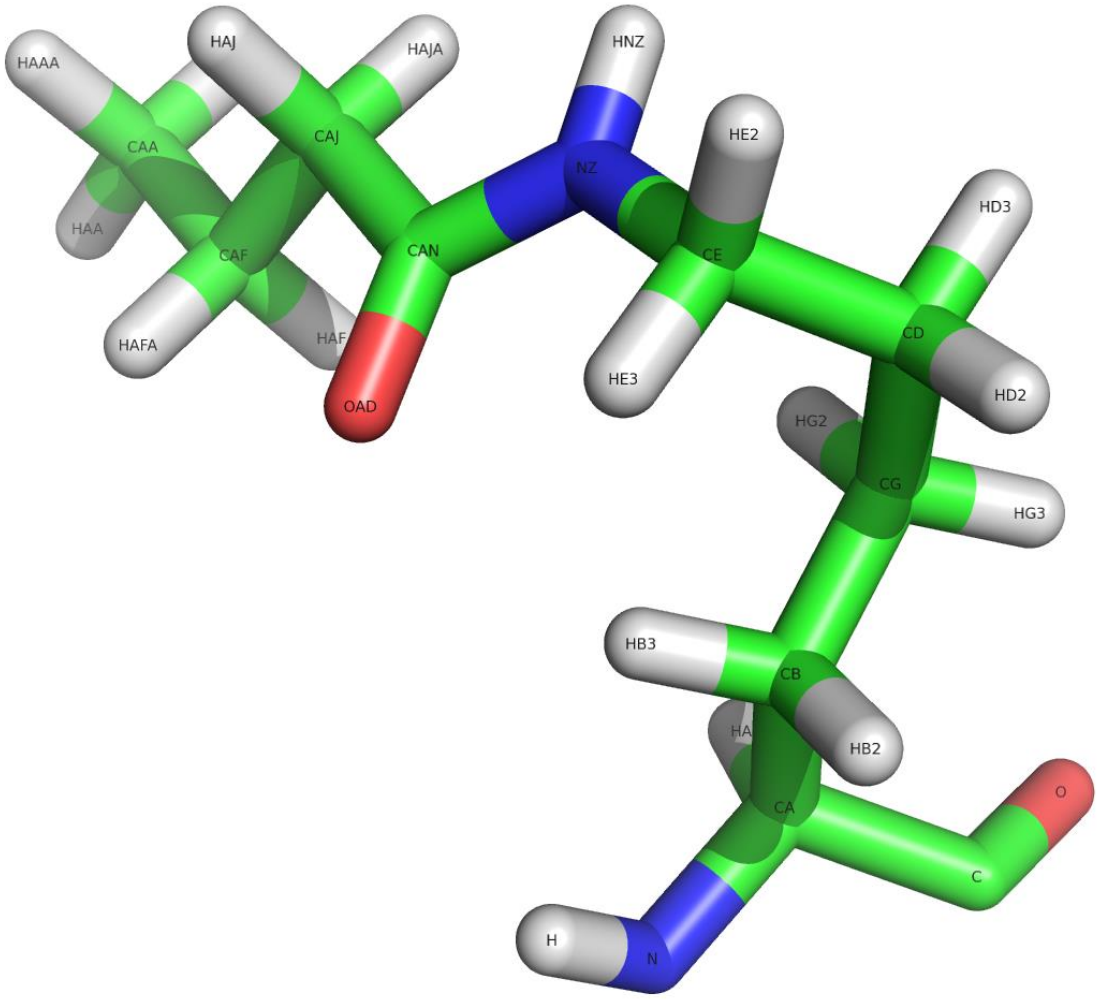
N6-propanoyllysine (PRK) α-conformer



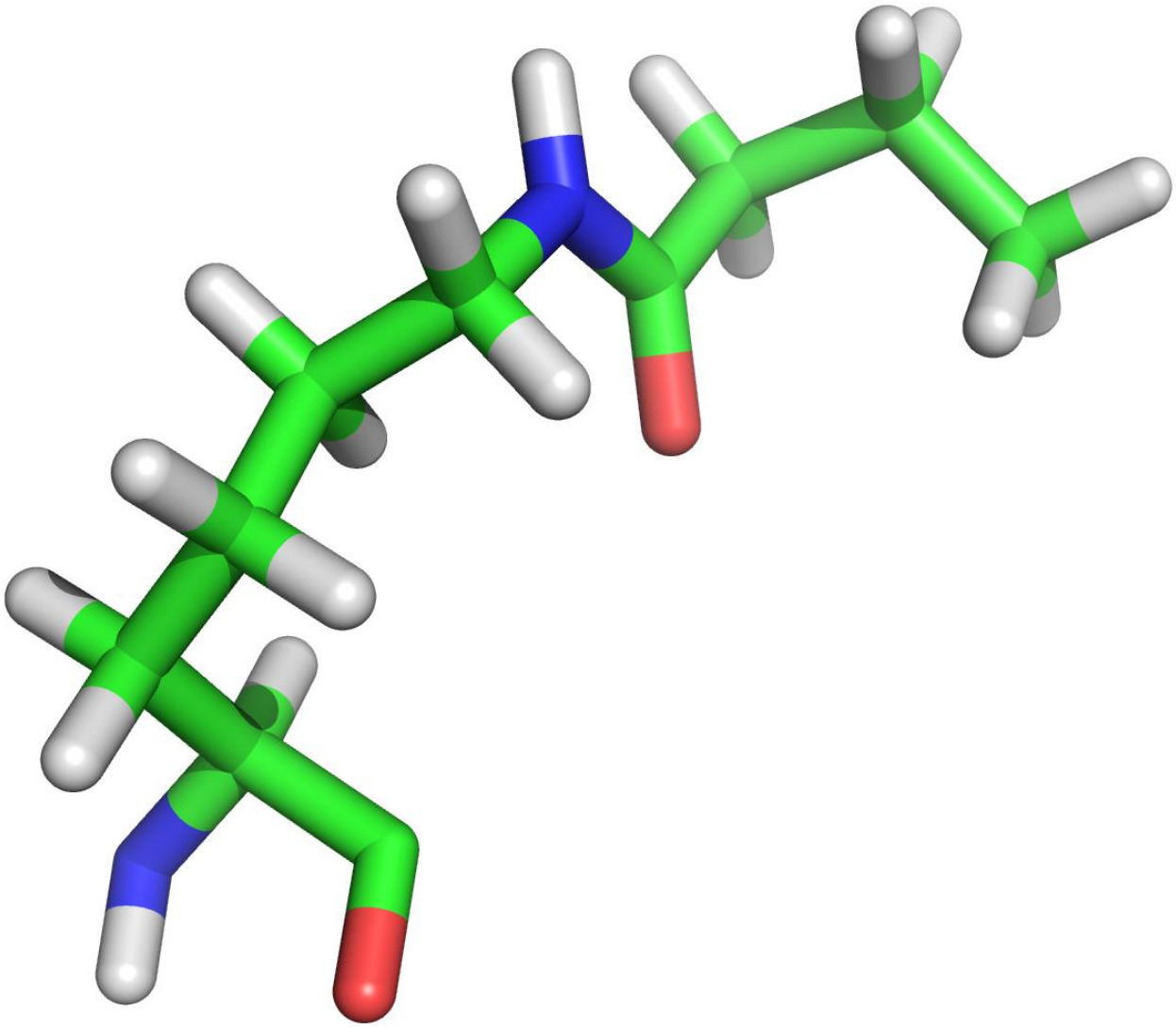
N6-propanoyllysine (PRK) β-conformer



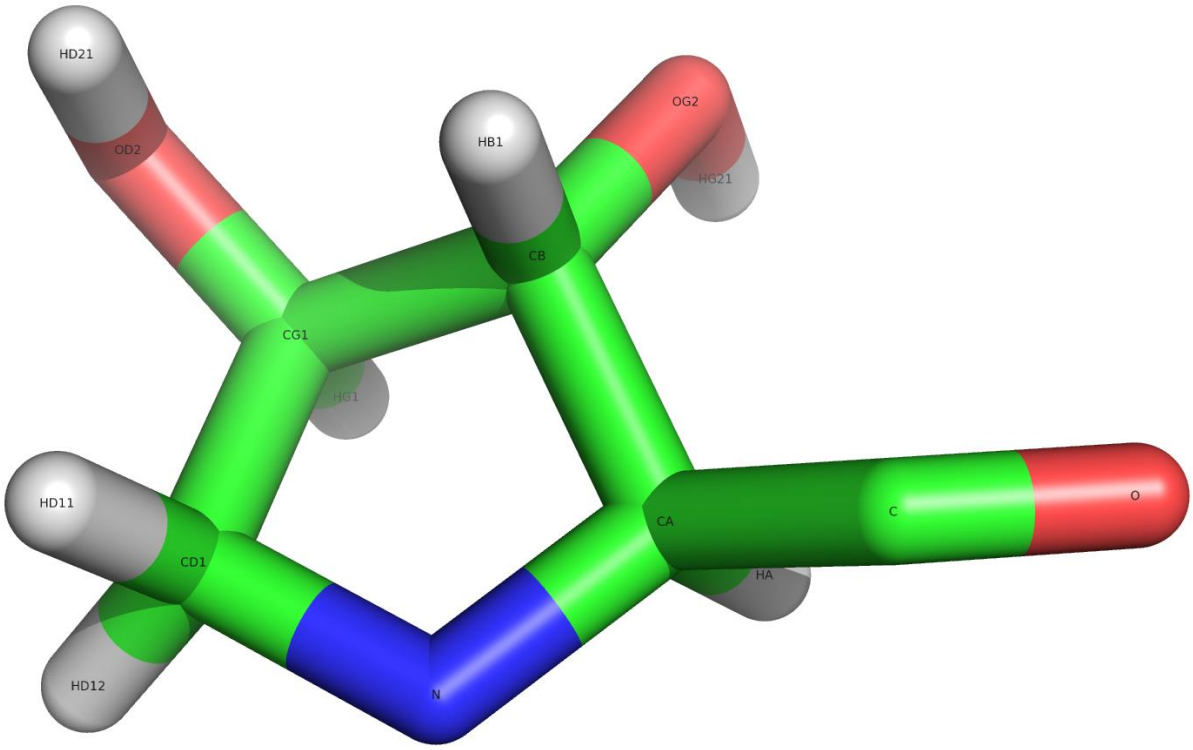
N6-butanoyllysine (BTK) α-conformer



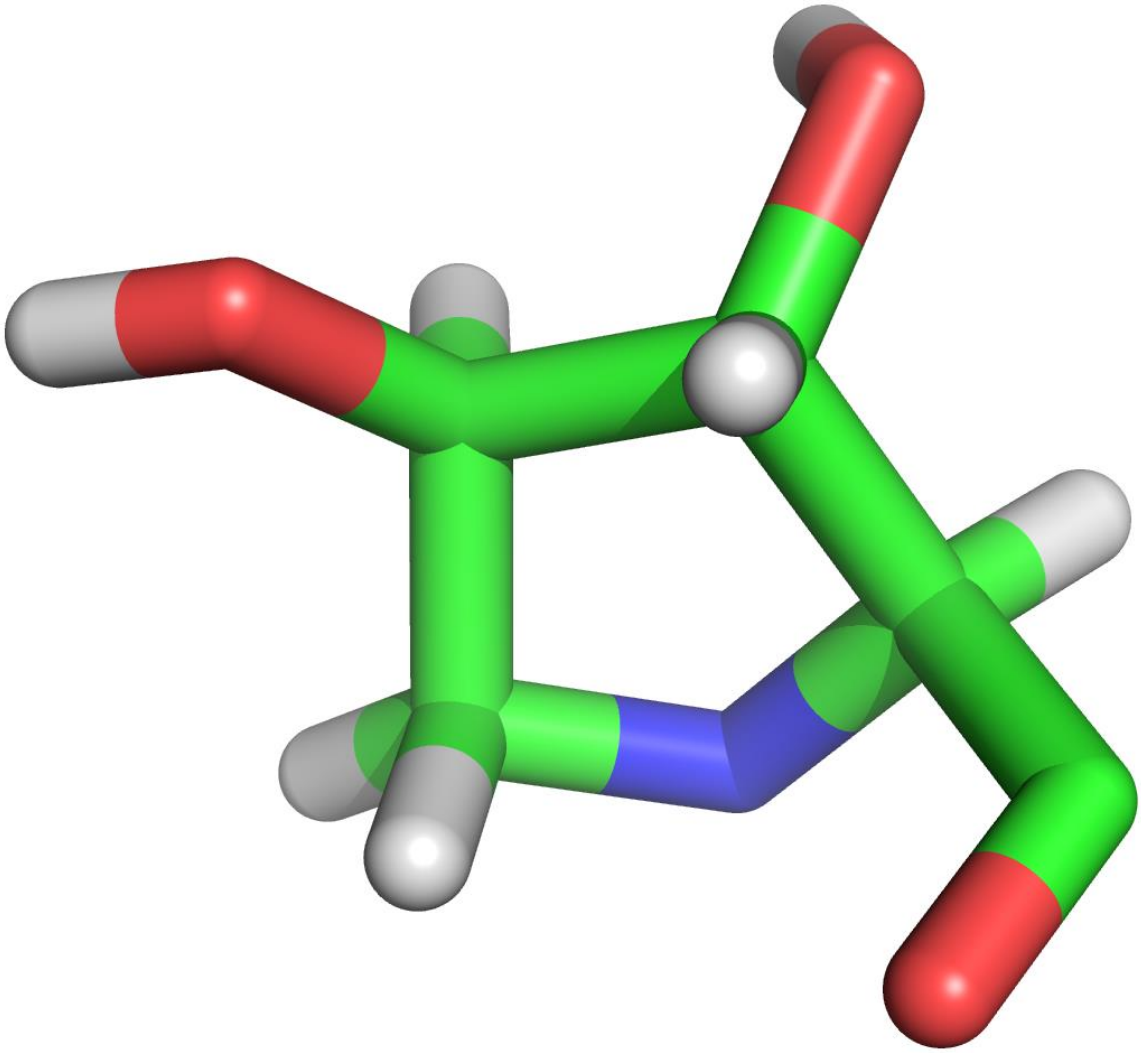
N6-butanoyllysine (BTK) β-conformer



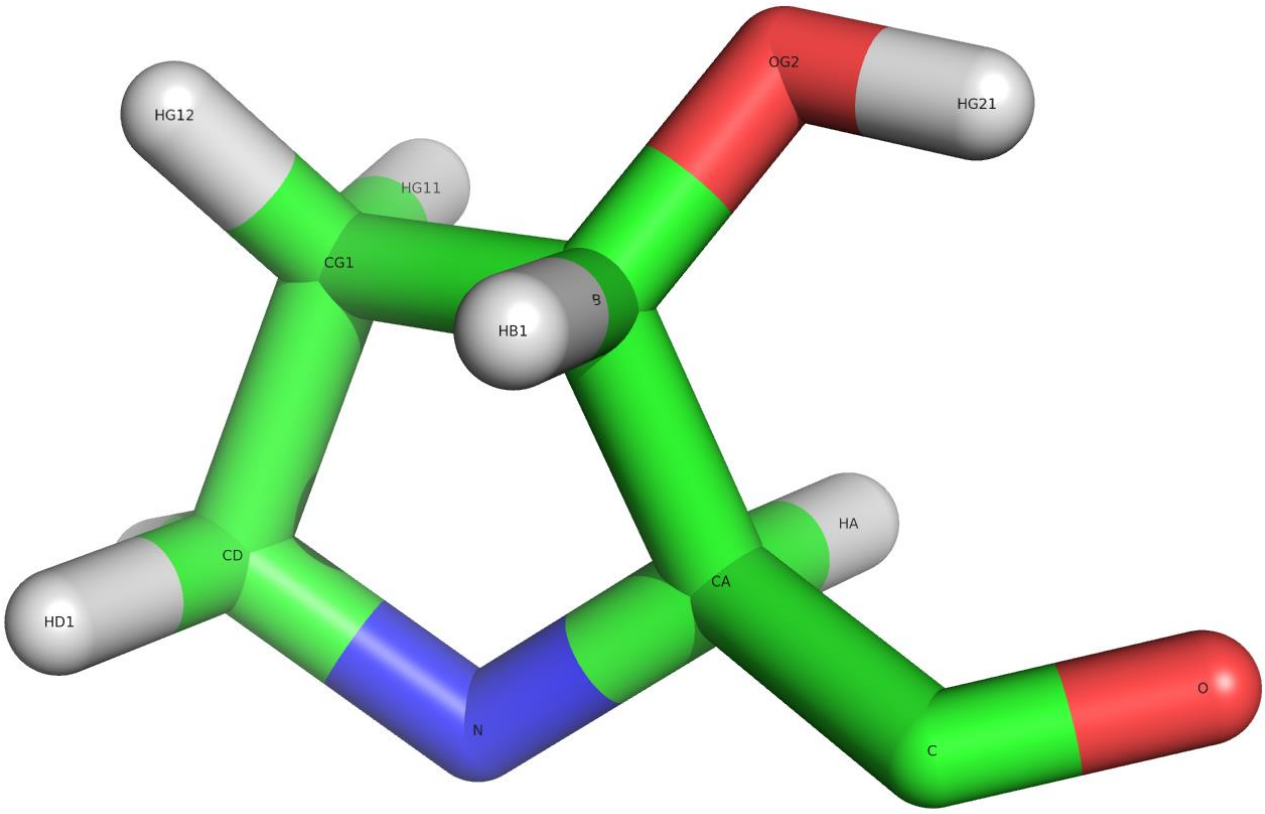
Proline Modifications
3,4-dihydroxyproline (DHP) α-conformer



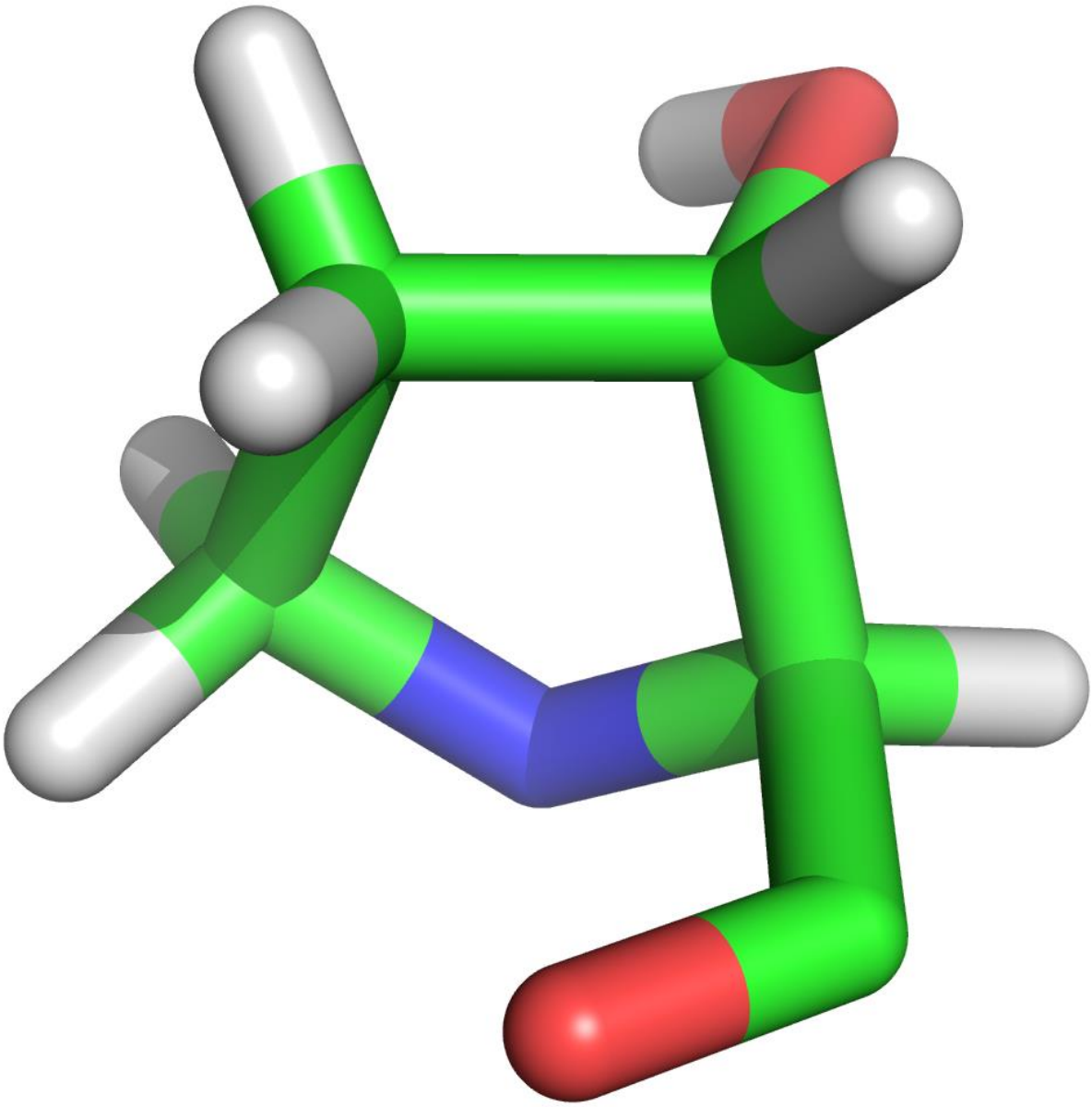
3,4-dihydroxyproline (DHP) β-conformer



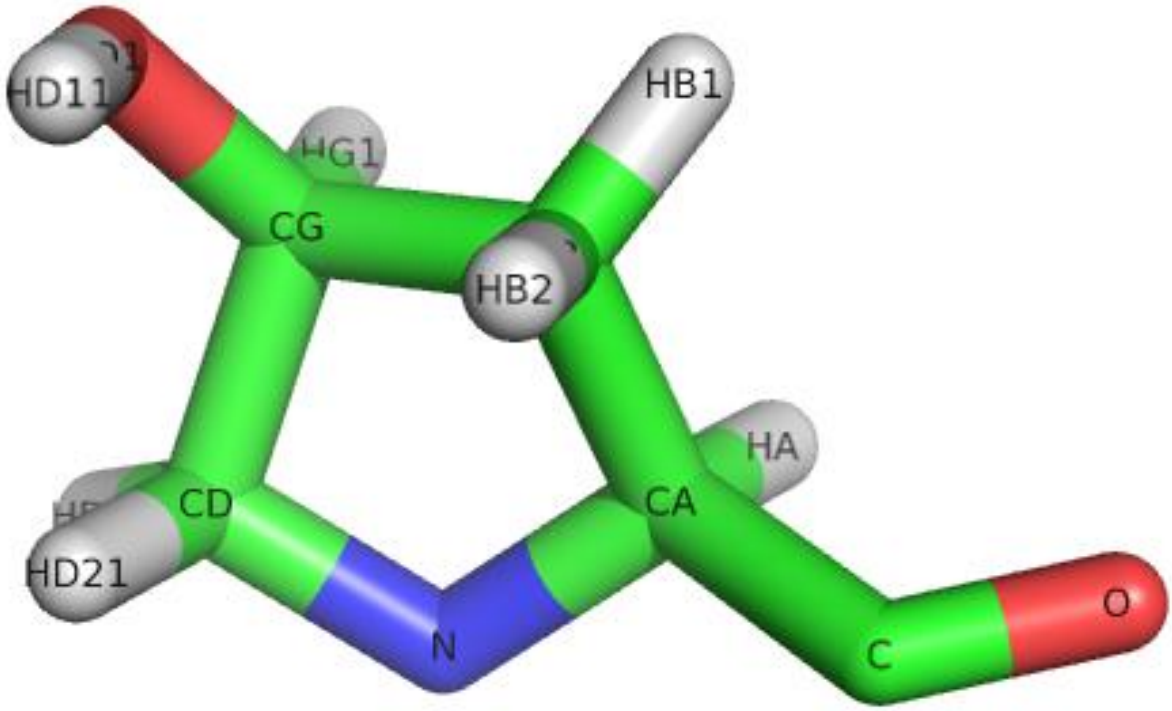
3-hydroxyproline (HY3) α-conformer



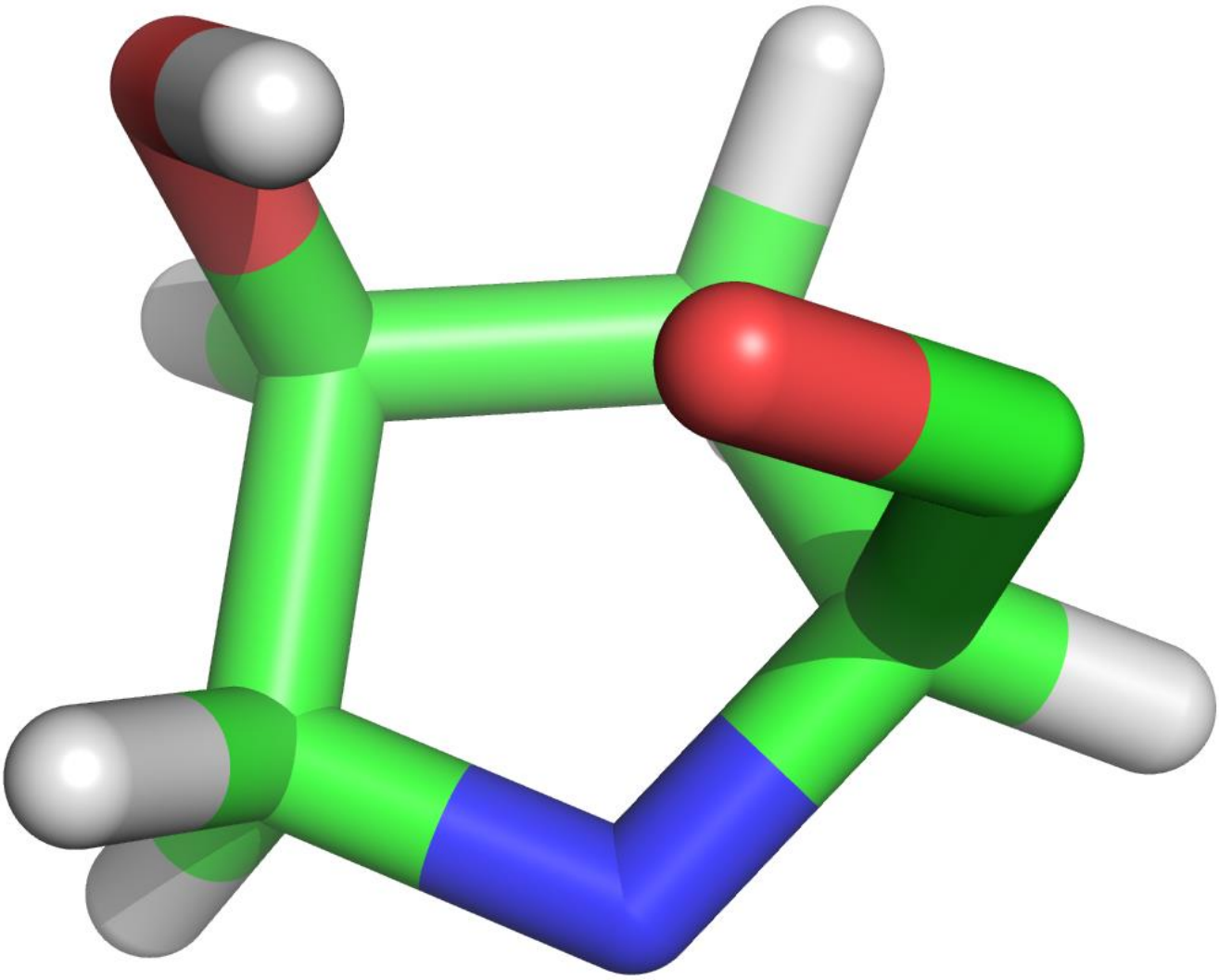
3-hydroxyproline (HY3) β-conformer



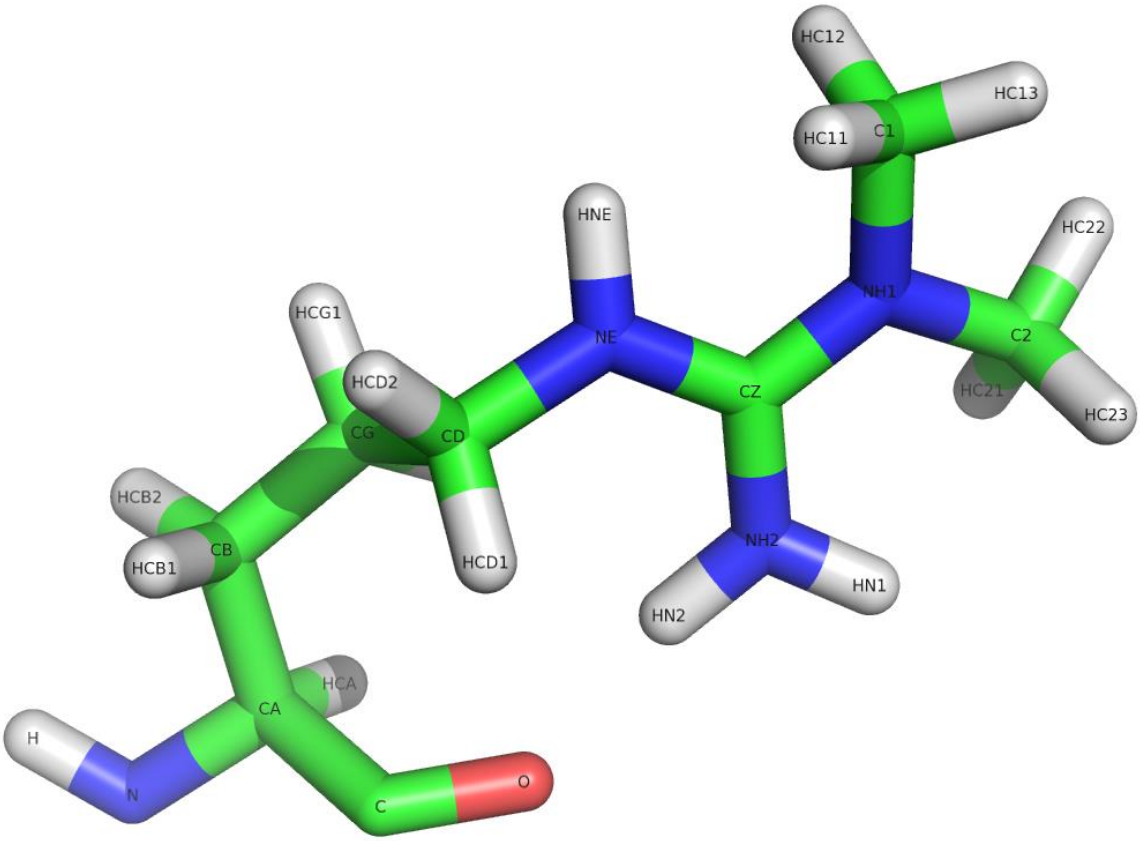
4-hydroxyproline (HYP) α-conformer



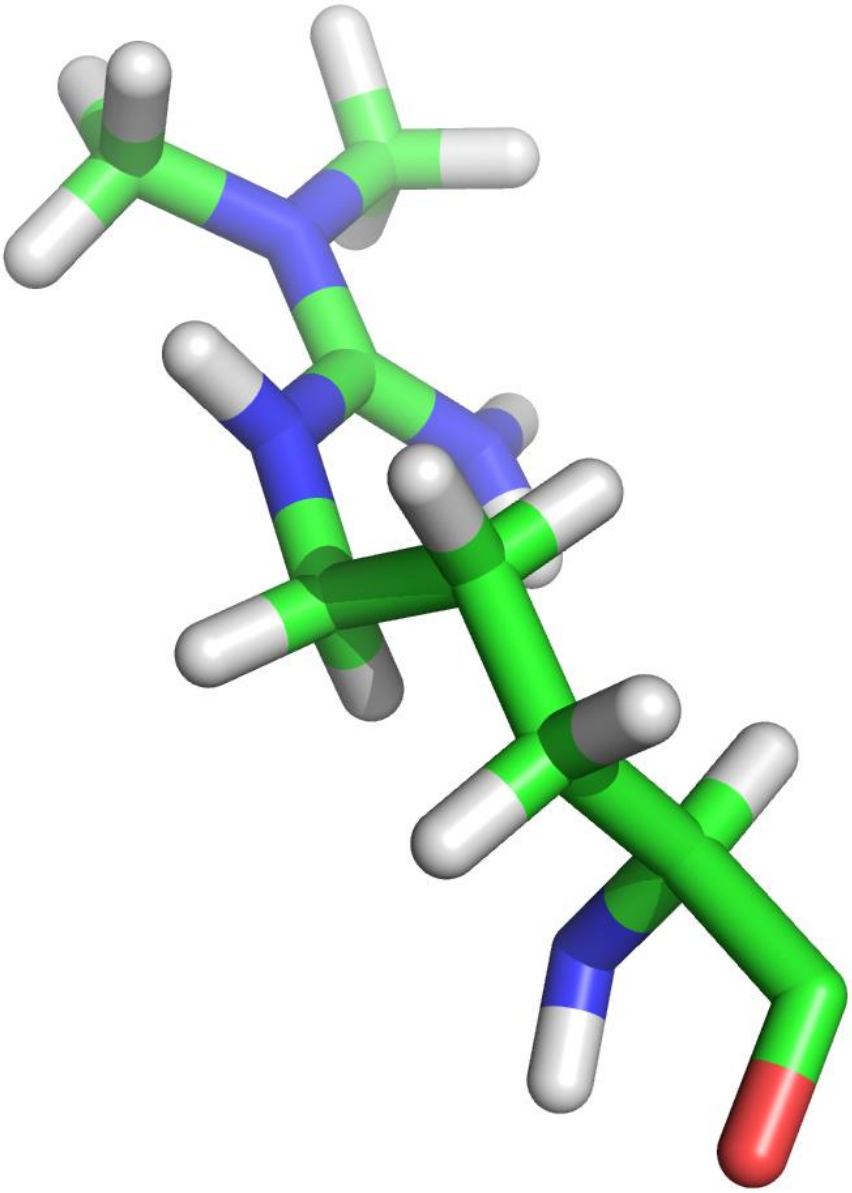
4-hydroxyproline (HYP) β-conformer



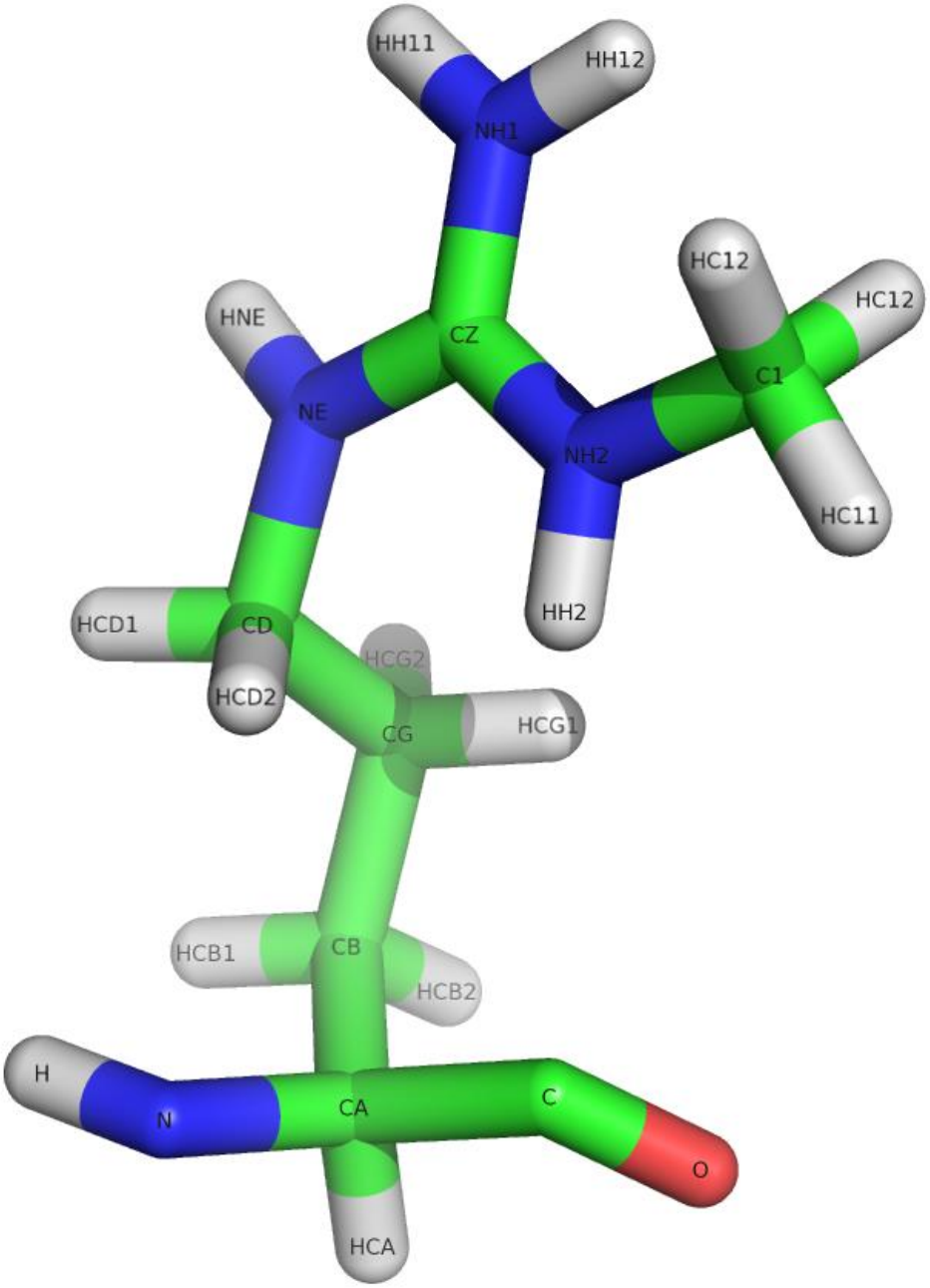
Arginine Modifications
Dimethylarginine (Anti-symmetric) (DA2) α-conformer



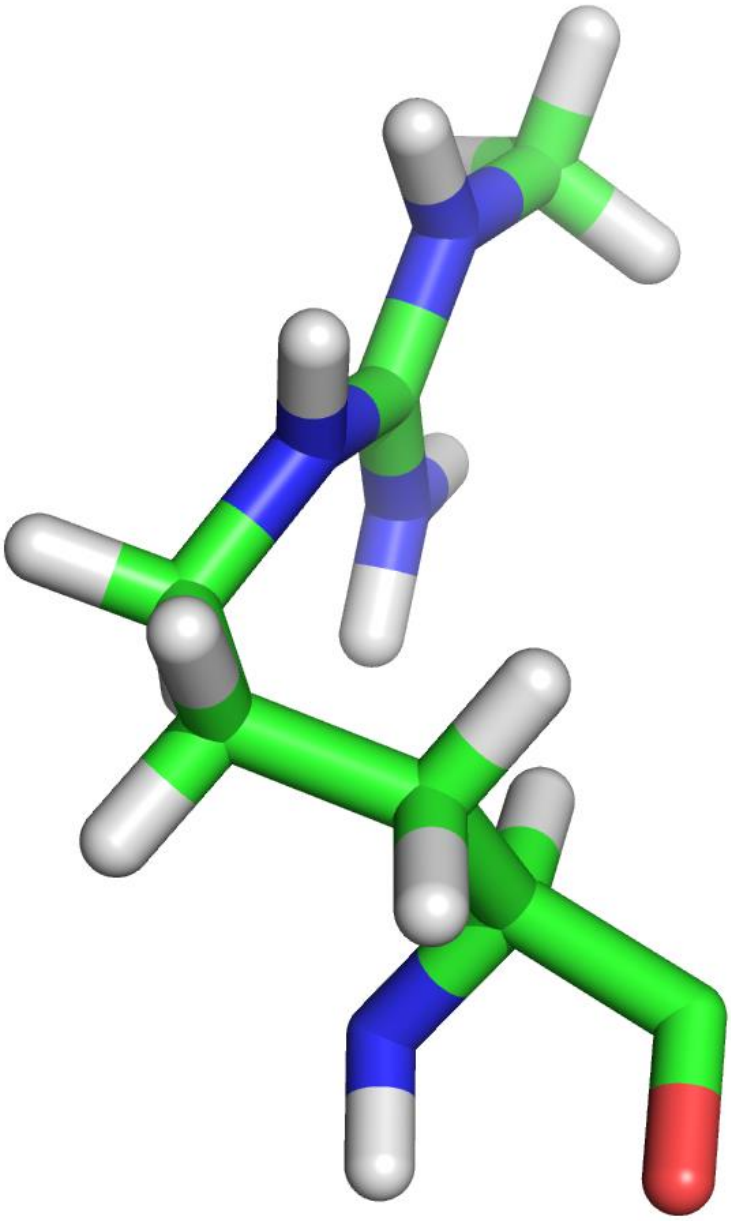
Dimethylarginine (Anti-symmetric) (DA2) β-conformer



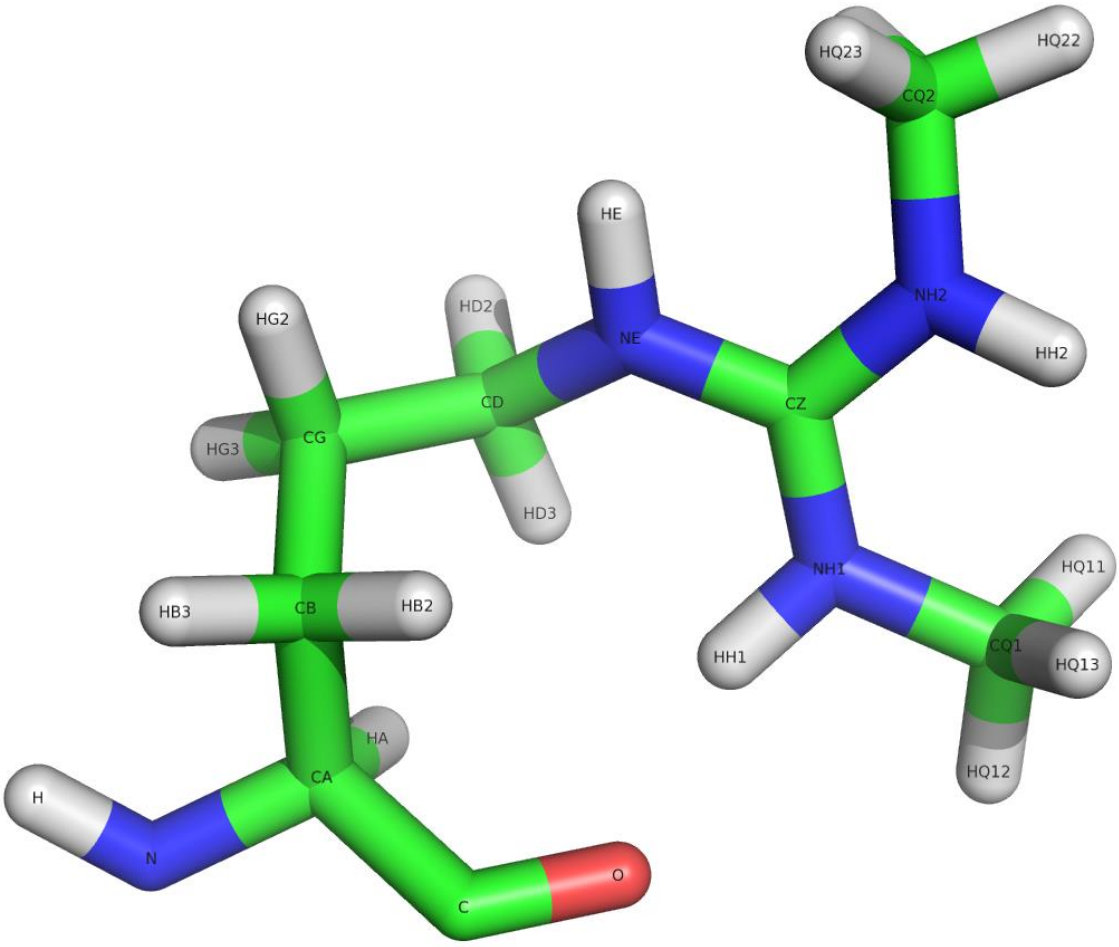
ω-methylarginine (DA1) α-conformer



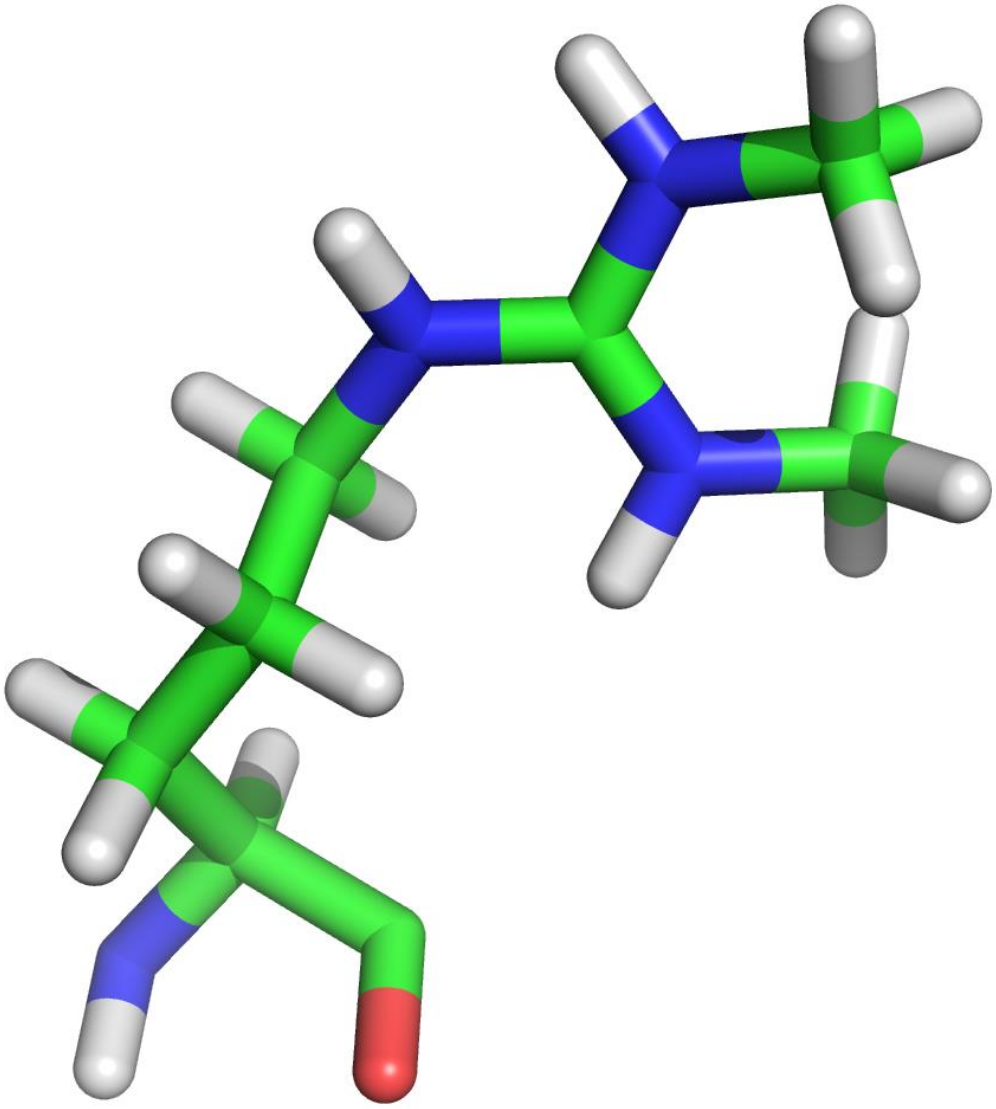
ω-methylarginine (DA1) β-conformer



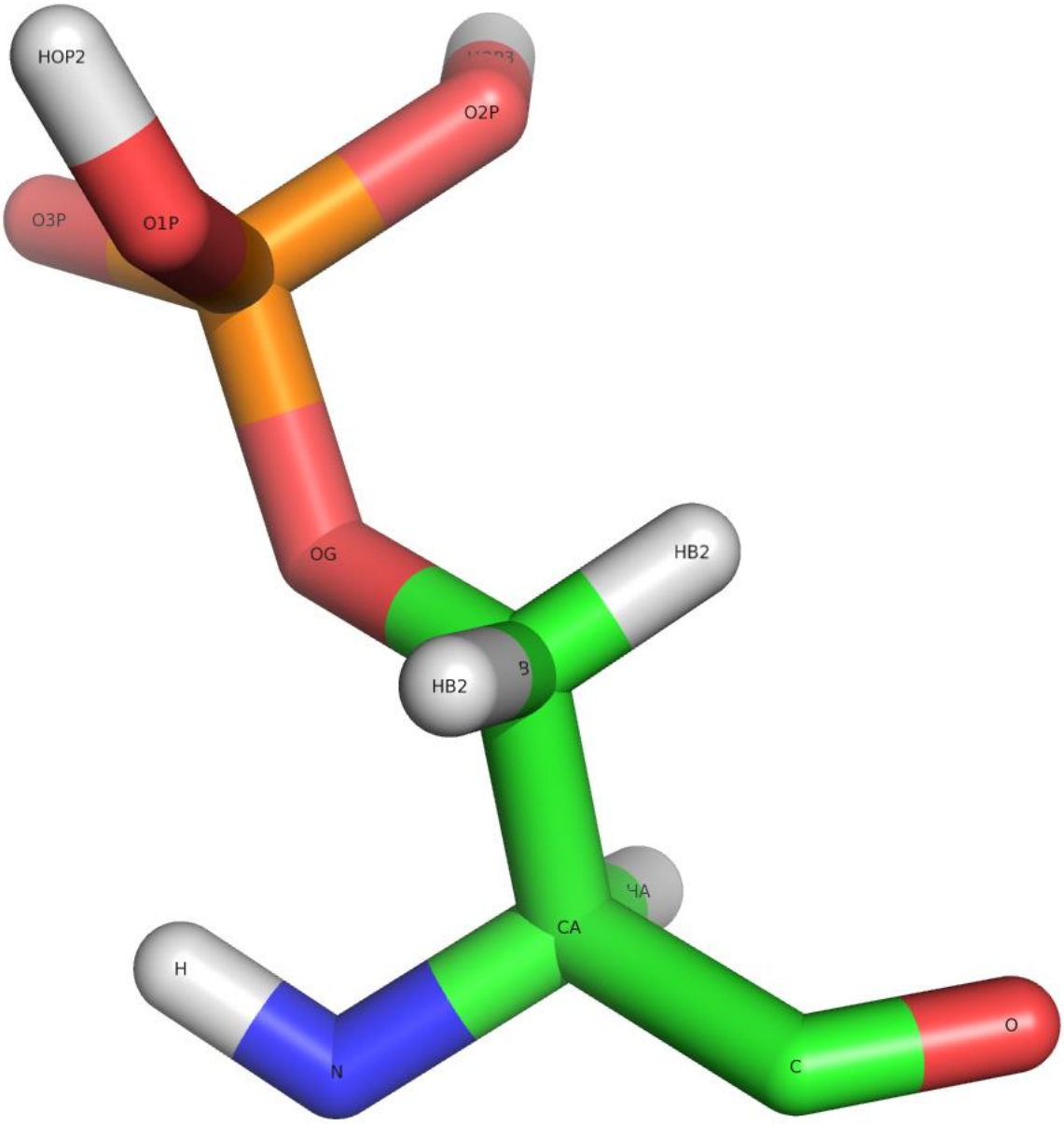
Dimethylarginine (Symmetric) (2MR) α-conformer



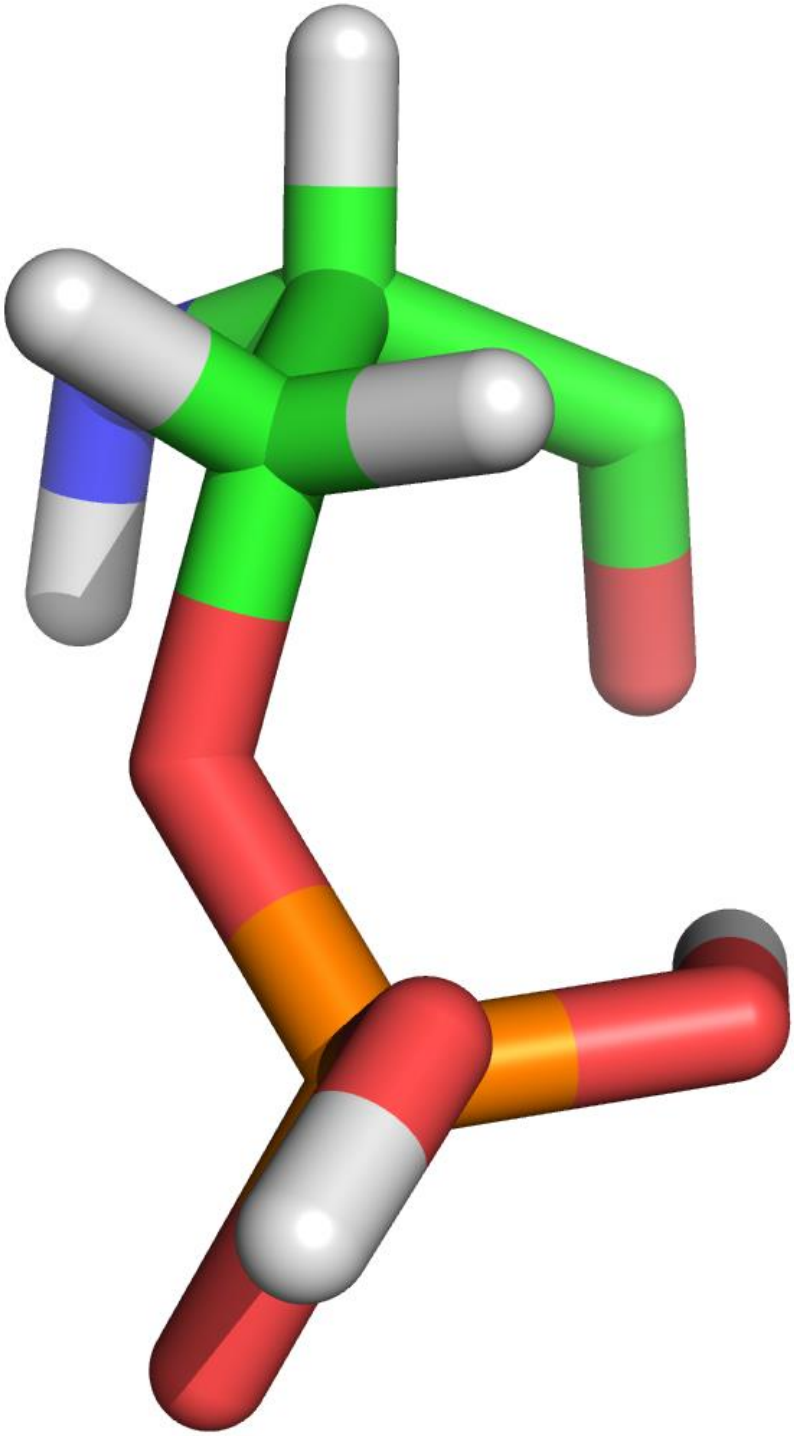
Dimethylarginine (Symmetric) (2MR) β-conformer



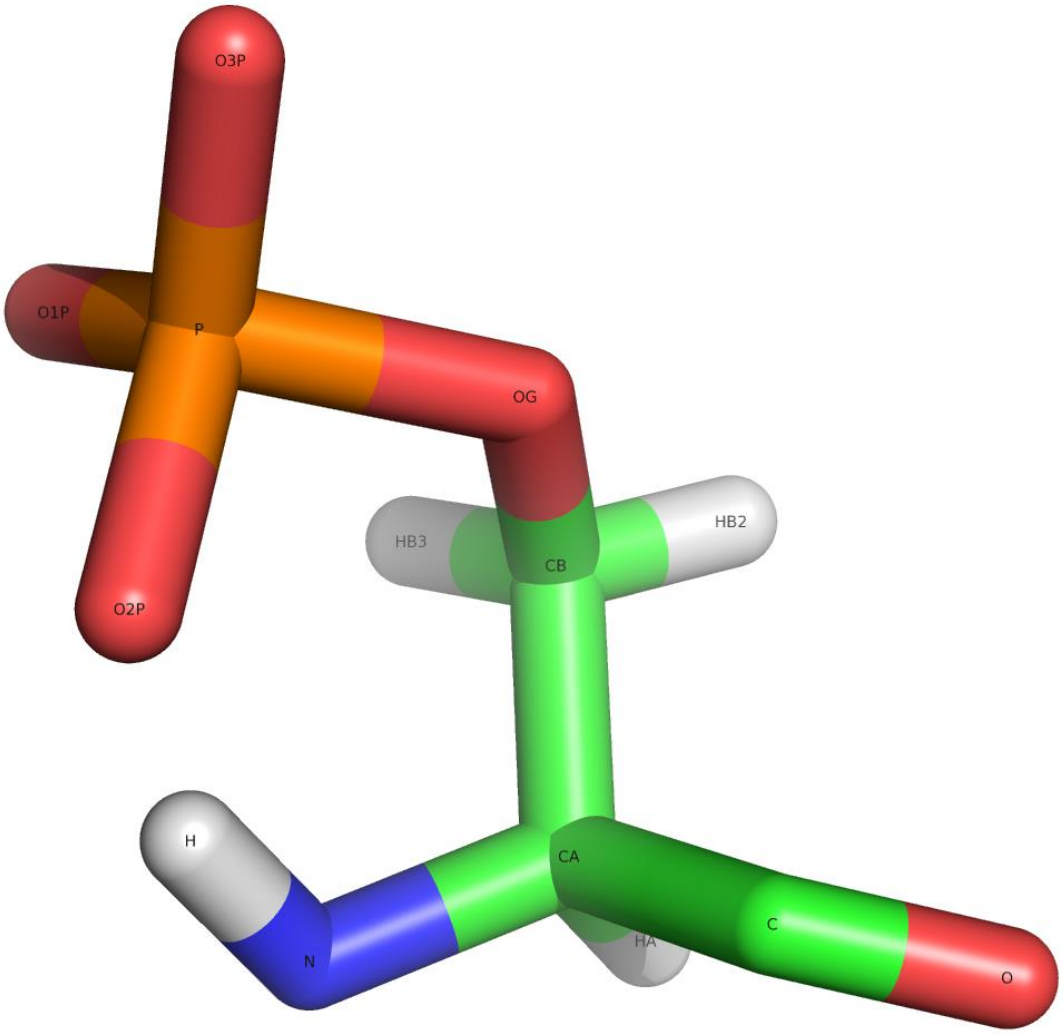
Serine Modifications
Phosphoserine (neutral) (SEN) α-conformer



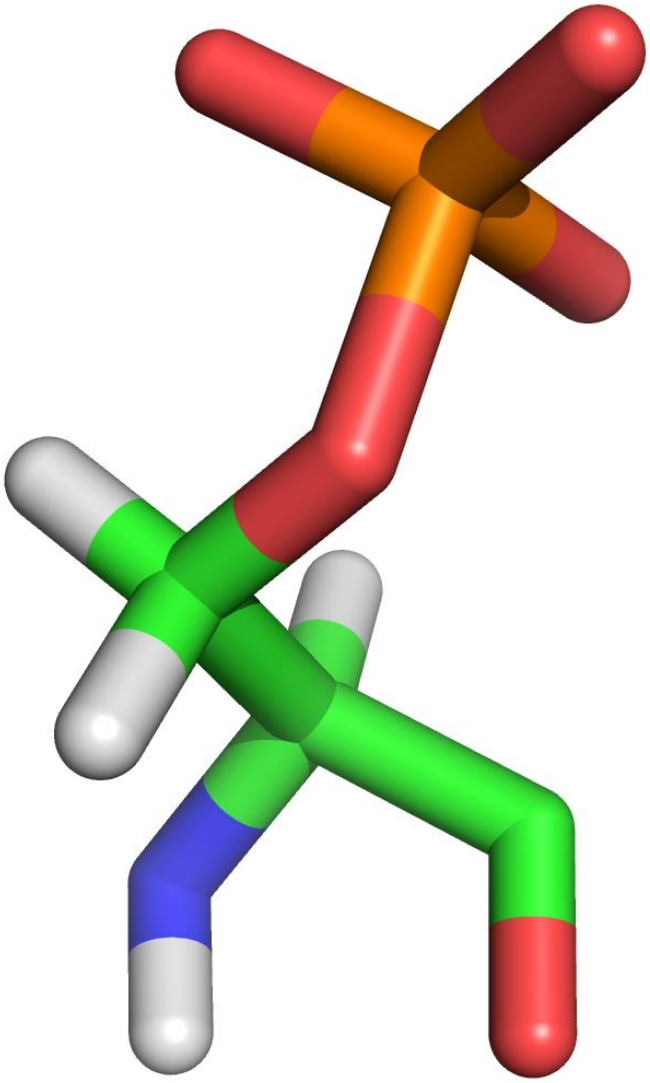
Phosphoserine (neutral) (SEN) β-conformer



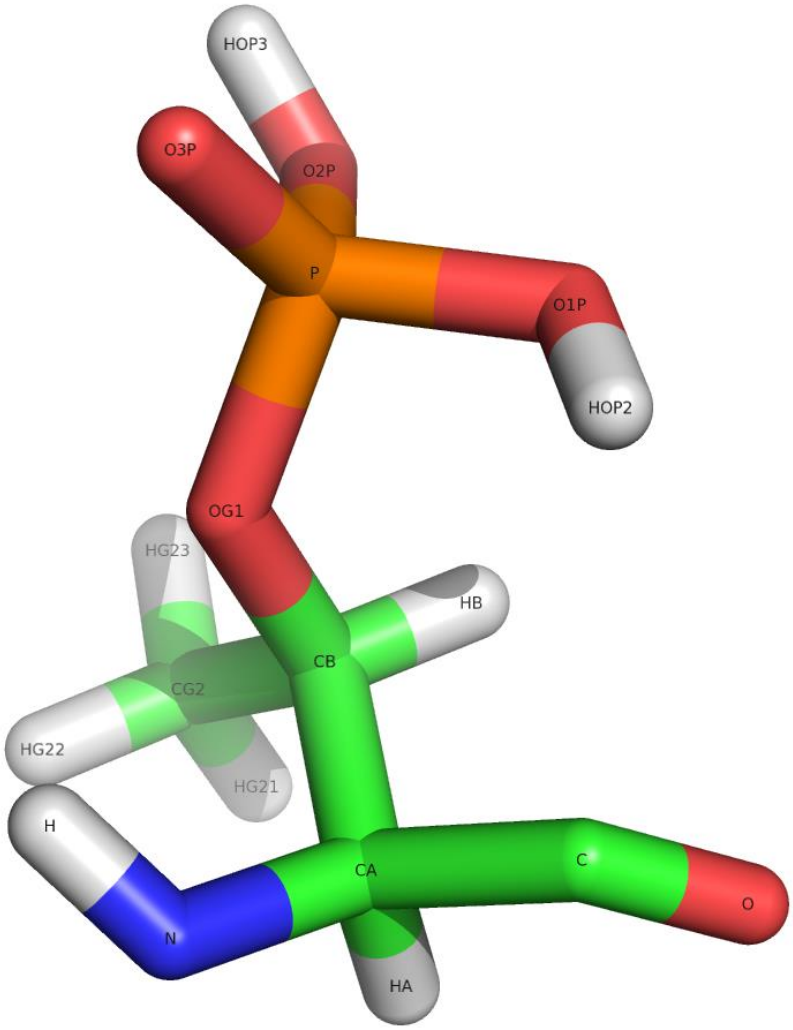
Phosphoserine (-2 charge) (SEP) α-conformer



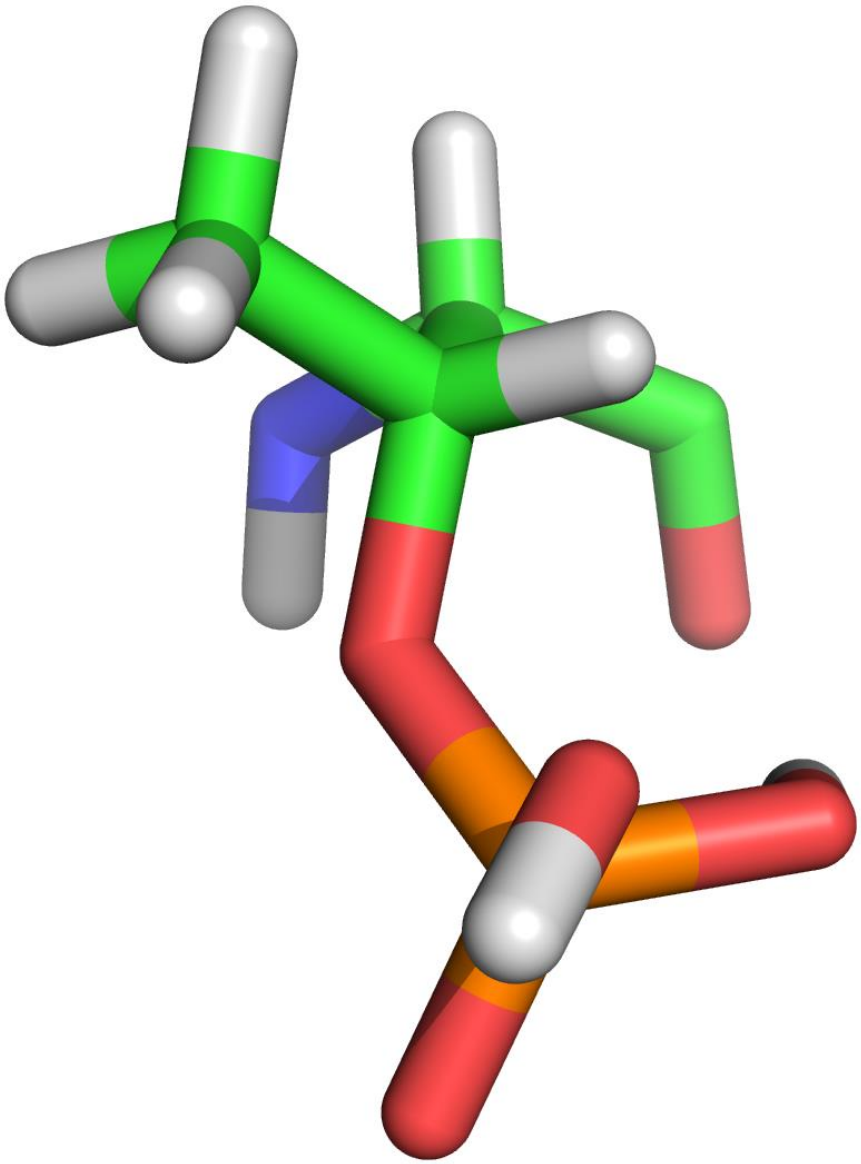
Phosphoserine (-2 charge) (SEP) β-conformer



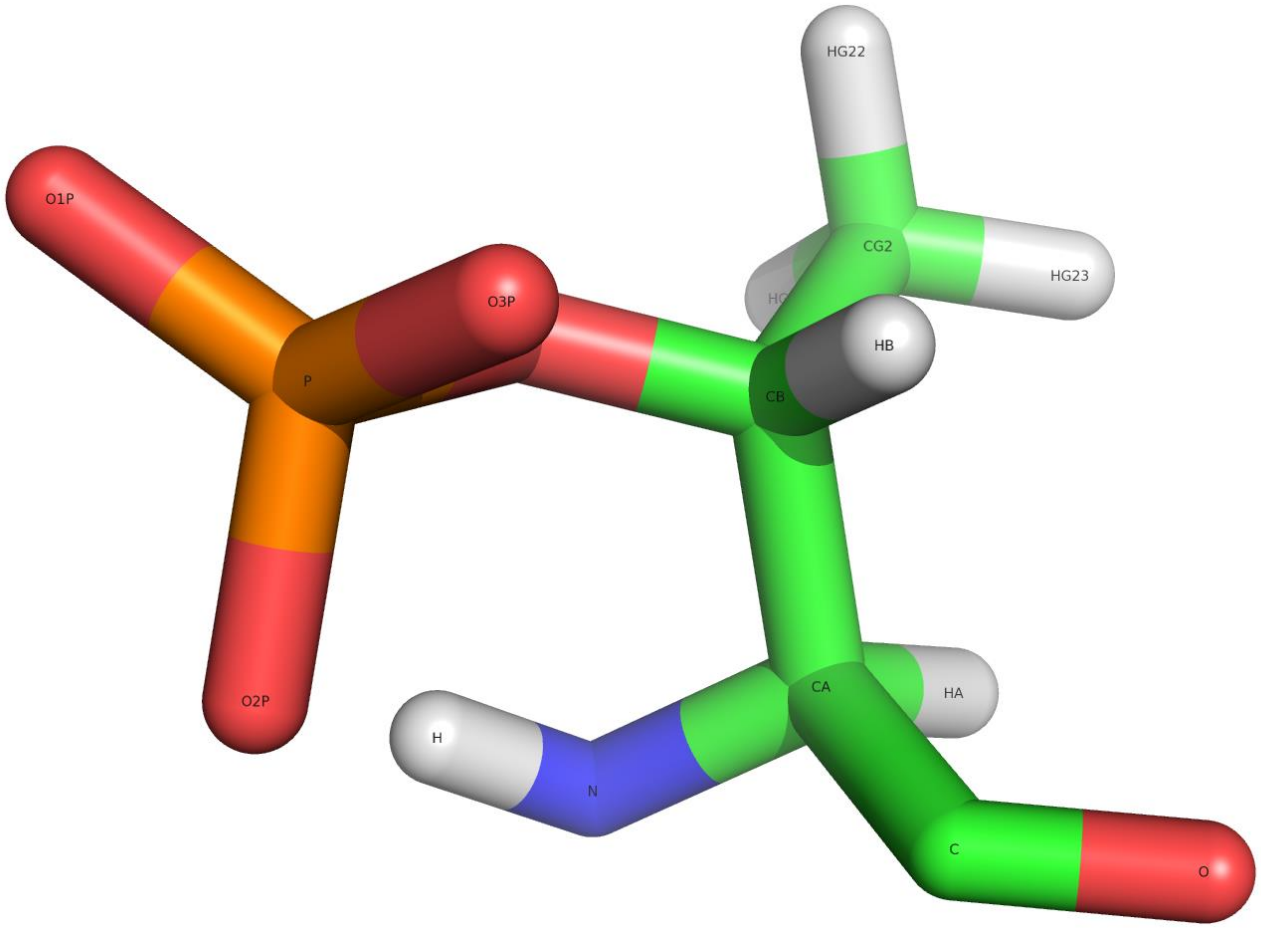
Threonine Modifications
Phosphothreonine (neutral) (TON) α-conformer



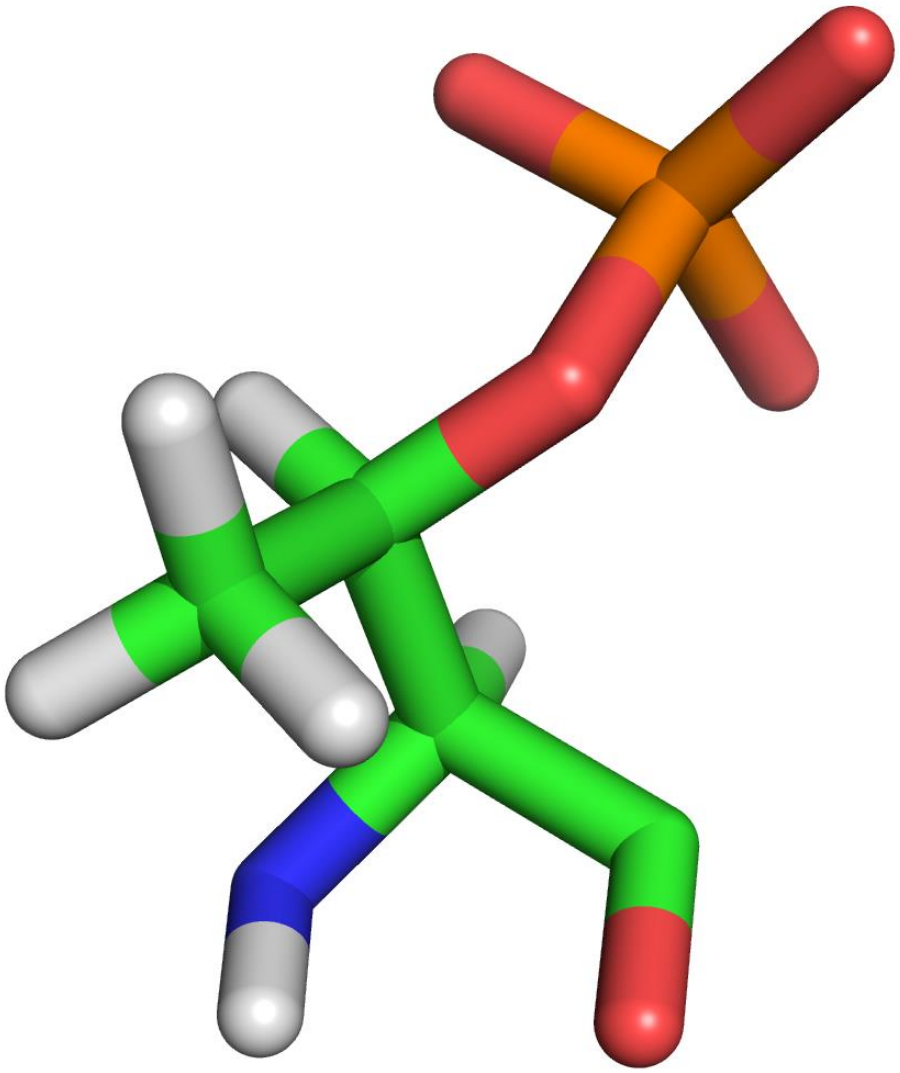
Phosphothreonine (neutral) (TON) β-conformer



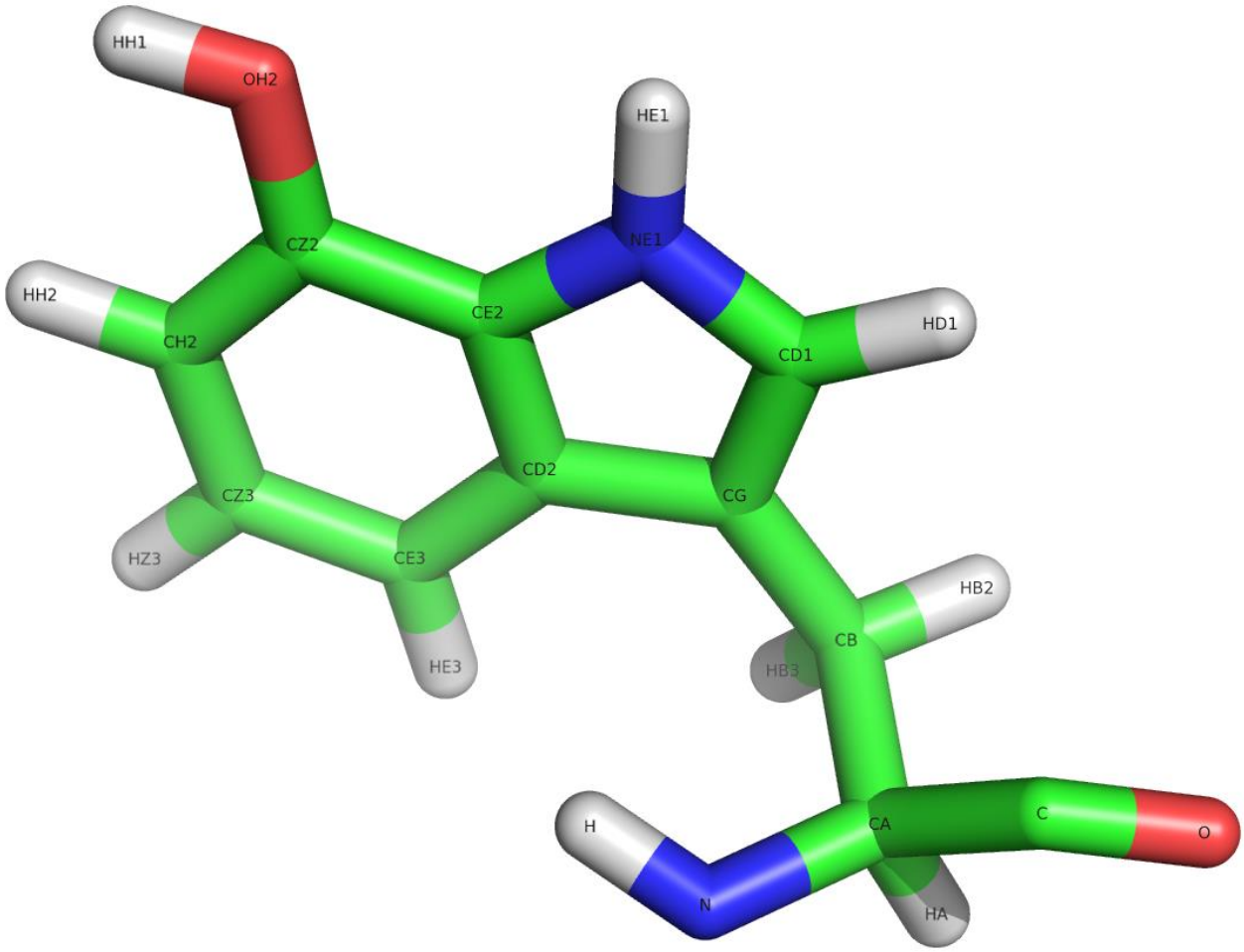
Phosphothreonine (-2 charge) (TPO) α-conformer



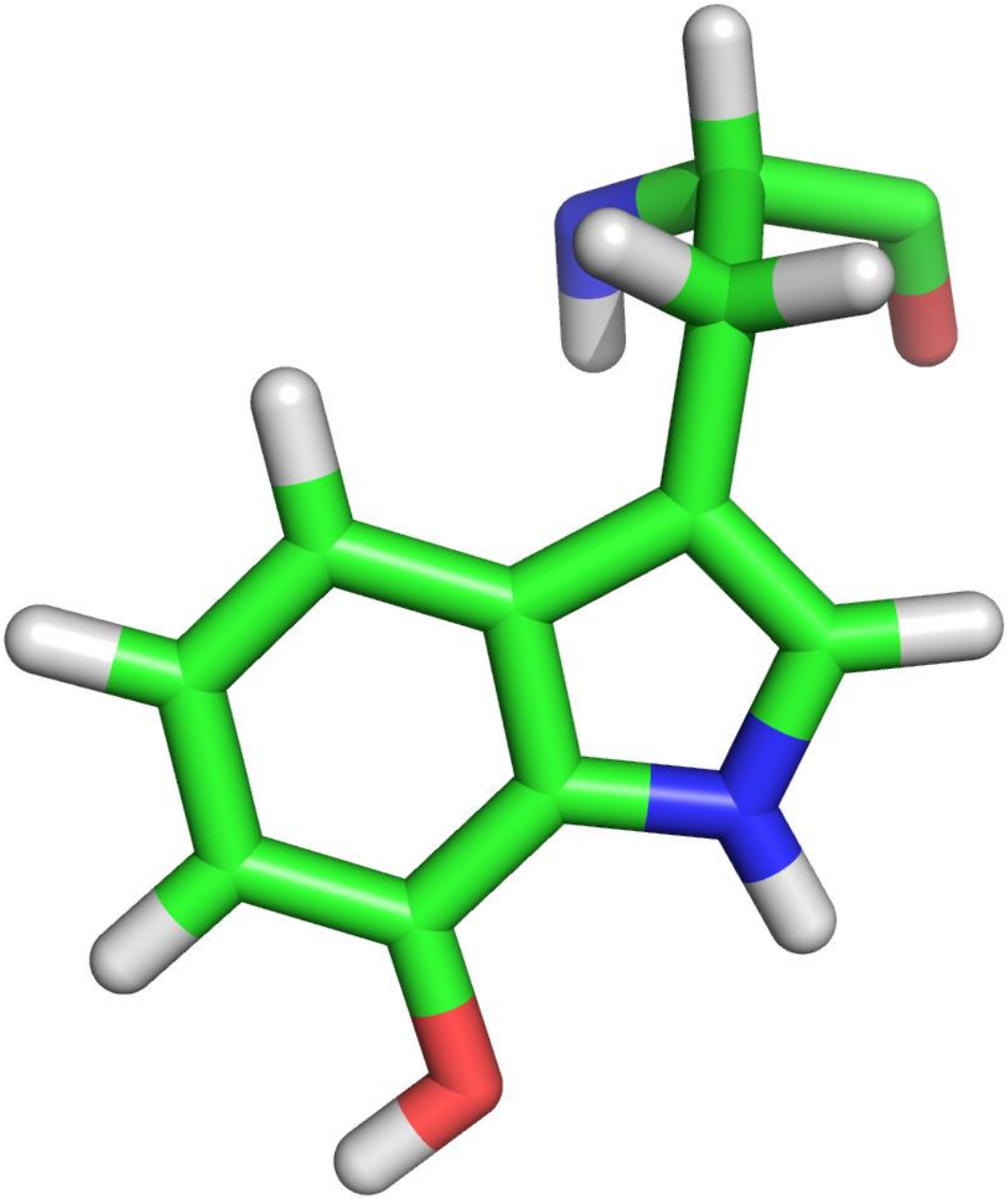
Phosphothreonine (-2 charge) (TPO) β-conformer



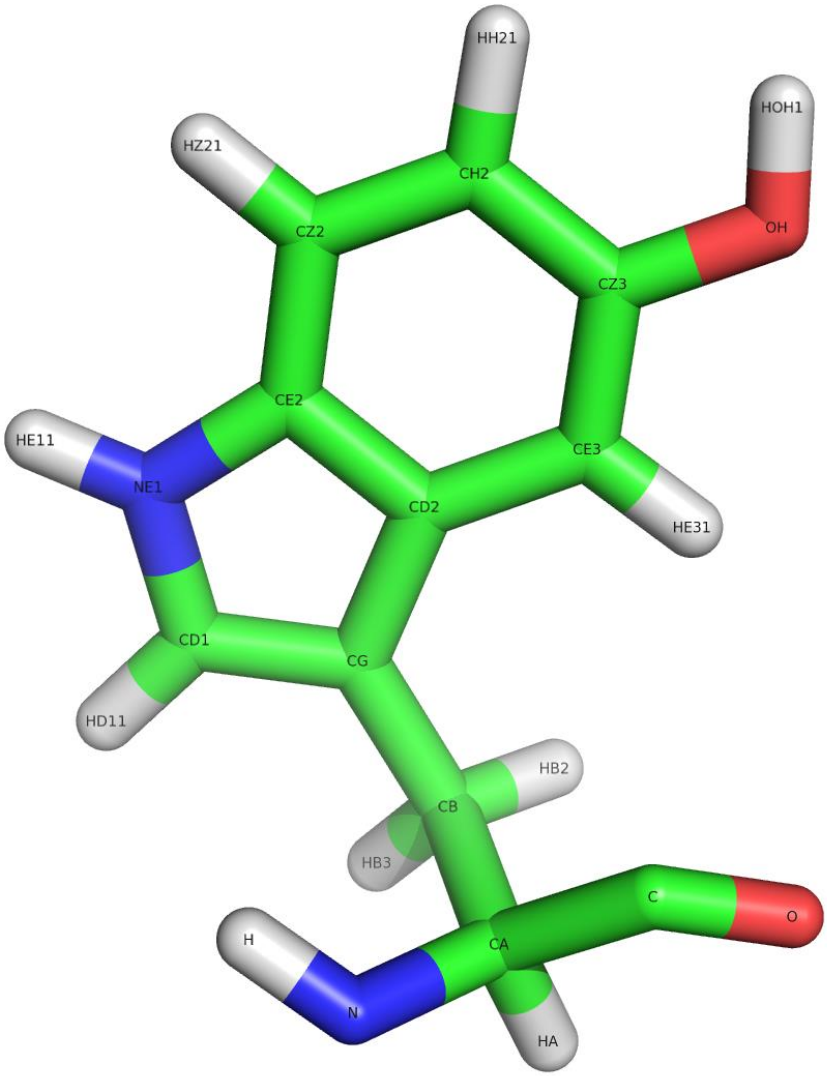
Tryptophan Modifications
7-hydroxytryptophan (0AF) α-conformer



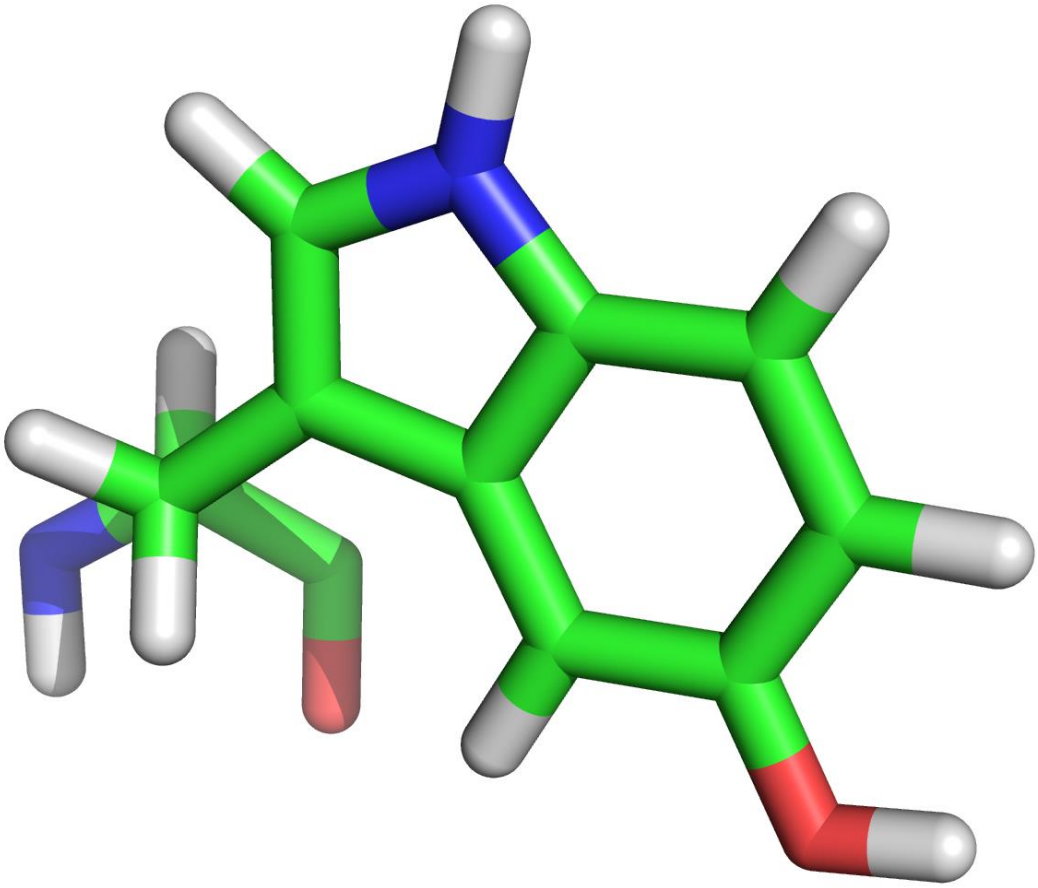
7-hydroxytryptophan (0AF) β-conformer



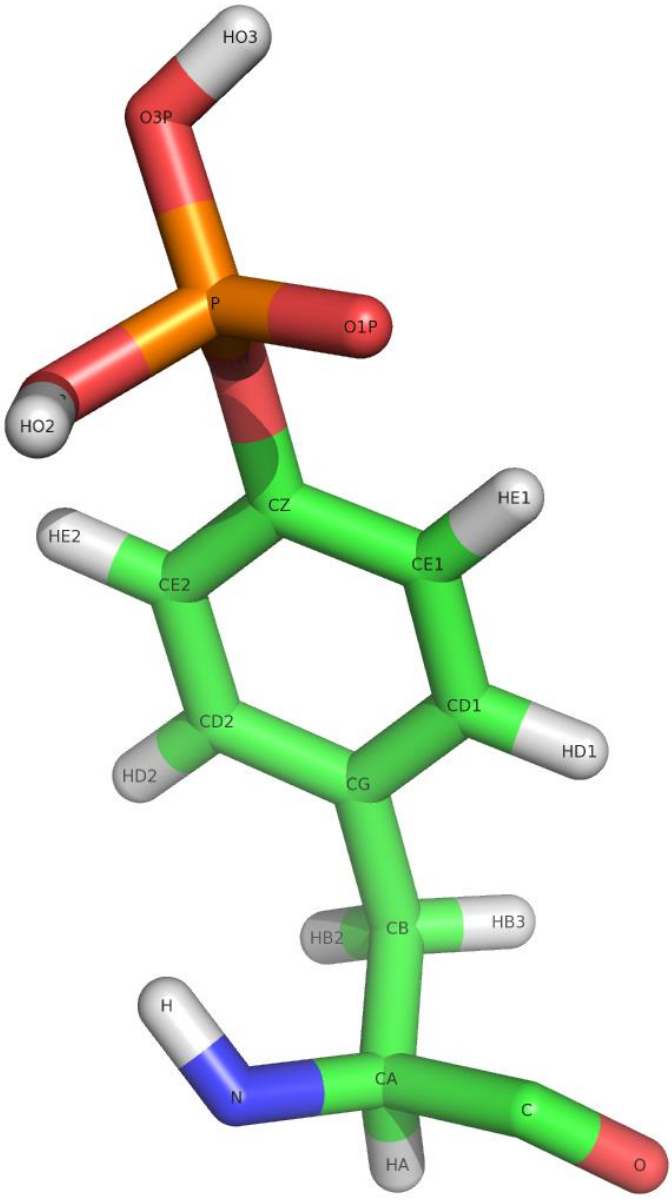
5-hydroxytryptophan (HTR) α-conformer



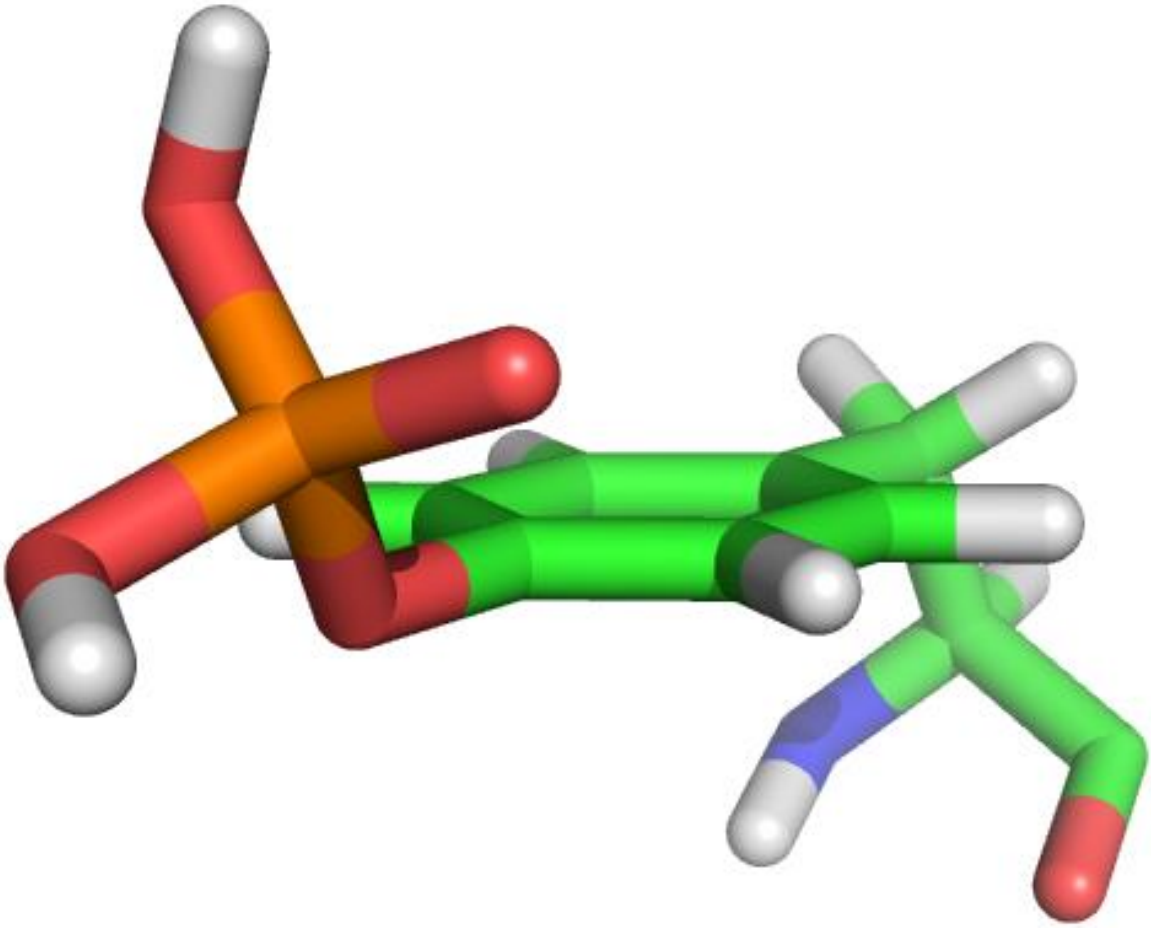
5-hydroxytryptophan (HTR) β-conformer



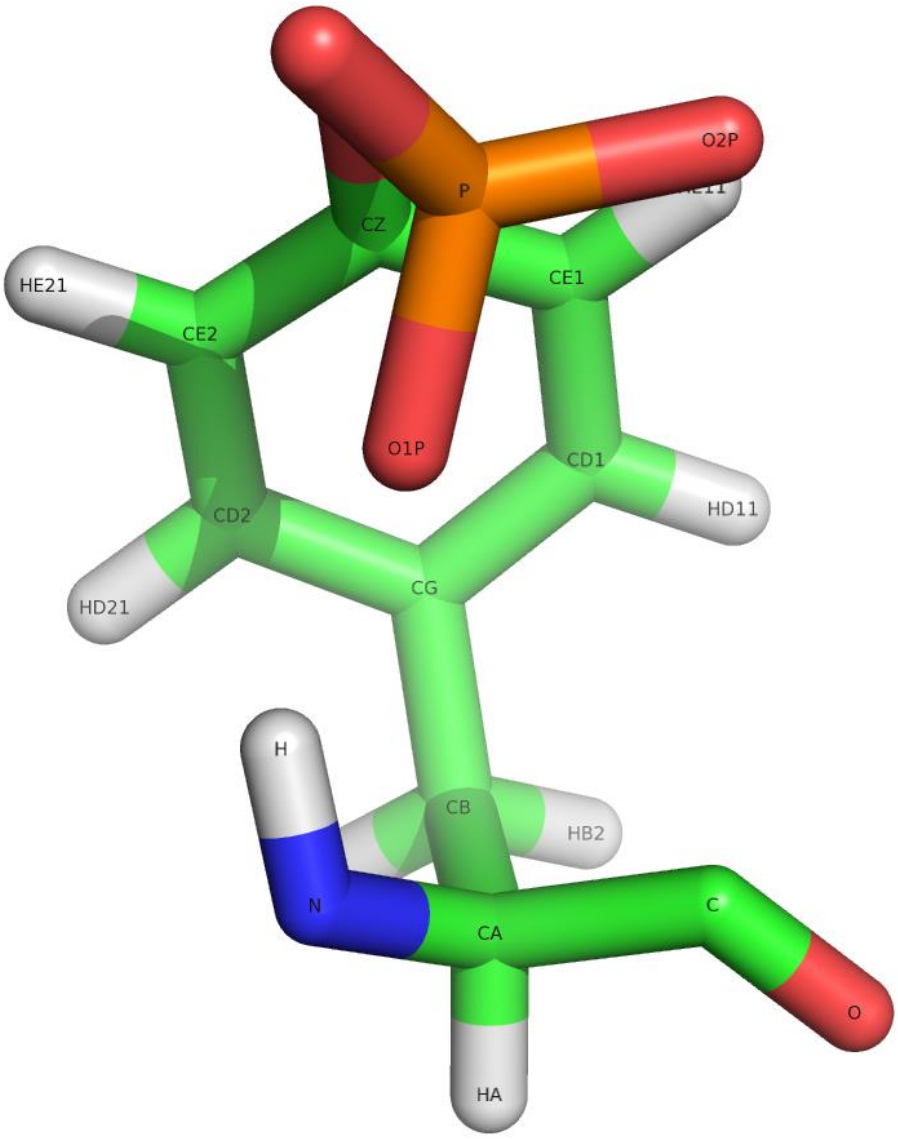
Tyrosine Modifications
Phosphotyrosine (neutral) (PTN) α-conformer



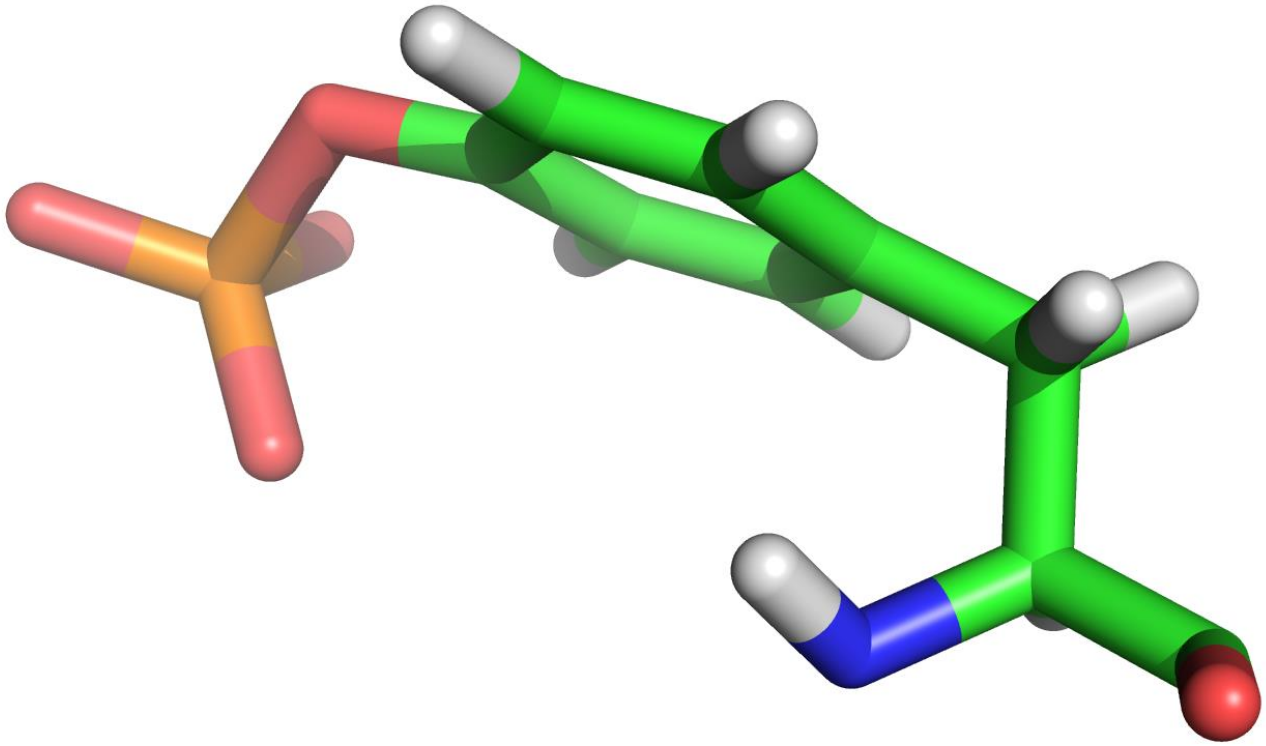
Phosphotyrosine (neutral) (PTN) β-conformer



Phosphotyrosine (-2 charge) (PTR) α-conformer



Phosphotyrosine (-2 charge) (PTR) β-conformer



1.4 Forcefield Parameters for Each Post-Translational Modification Grouped by Scaffold Residue

1.4.1 Cysteine Modifications

Table 1: Partial charges assigned to Cysteinepersulfide (CSS)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.31632
H	H	0.246008
CA	CT	-0.11831
CB	CT	-0.03683
SG	S	-0.16663
SD	SH	-0.16936
HD1	HS	0.187015
HB1	H1	0.099243
HB2	H1	0.099243
HA	H1	0.121987
C	C	0.631964
O	O	-0.57801

Table 2: New parameters assigned to Cysteinepersulfide (CSS)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
S -SH	155.8	2.067	same as sh-ss	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CT-S -SH	47.8	101.93	same as c3-ss-sh	
S -SH-HS	33.73	99.17	same as hs-sh-ss	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-S-SH-HS	1.623	140.097	-1	fitted,GK13
CT-S-SH-HS	3.652	-6.4855	2	fitted,GK13
CT-CT-S-SH	7.362	170.865	-1	fitted,GK13
CT-CT-S-SH	5.361	139.062	-2	fitted,GK13
CT-CT-S-SH	4.914	118.59	-3	fitted,GK13
CT-CT-S-SH	4.319	78.606	-4	fitted,GK13
CT-CT-S-SH	3.673	43.02	-5	fitted,GK13
CT-CT-S-SH	2.847	10.87	-6	fitted,GK13
CT-CT-S-SH	2.0516	-18.906	-7	fitted,GK13
CT-CT-S-SH	1.365	-48.47	-8	fitted,GK13
CT-CT-S-SH	0.8598	-76.77	-9	fitted,GK13
CT-CT-S-SH	0.5347	-104.95	10	fitted,GK13

Table 3: Partial charges assigned to Cysteine sulfenic acid (CSO)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.60655
H	H	0.336844
CA	CT	0.092053
CB	CT	-0.31688
SG	S	-0.04901
OD	OH	-0.43568
HD1	HO	0.411219
HB1	H1	0.164503
HB2	H1	0.164503
HA	H1	0.093011
C	C	0.759105
O	O	-0.61311

Table 4: New parameters assigned to Cysteine sulfenic acid (CSO)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
S -OH	265.6	1.682	same as oh-ss	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CT-S -OH	40.85	98.28	same as c3-ss-oh	
S -OH-HO	54.54	107.06	same as ho-oh-ss	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-S -OH-HO	2.4	0	2	same as X -oh-ss-X
CT-CT-S-OH	2.255	172.8	-1	fitted,GK13
CT-CT-S-OH	2.093	148.25	-2	fitted,GK13
CT-CT-S-OH	1.403	-95.47	-3	fitted,GK13
CT-CT-S-OH	1.178	-81.75	4	fitted,GK13

Table 5: Partial charges assigned to Phosphocysteine (CSP)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.59379
H	H	0.33177
CA	CT	0.032976
CB	CT	-0.17276
SG	S	-0.1766
P	P	0.787836
O1P	OH	-0.50119
HO1P	HO	0.428307
O3P	OH	-0.50119
HO3P	HO	0.428307
O2P	O2	-0.59723
HB2	H1	0.109868
HB3	H1	0.109868
HA	H1	0.07981
C	C	0.917801
O	O	-0.68378

Table 6: New parameters assigned to Phosphocysteine (CSP)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
S -P	167	2.104	same as p4-ss	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
S -P -OH	80.64	100.06	Calculated empirically	
OH-P -OH	72.99	95.71	same as oh-p4-oh	
CT-S -P	38.03	98.16	same as c3-ss-p4	
S -P -O2	75.8	116.14	same as o -p4-ss	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-S-P-O2	0.42884	-6.1956	-1	fitted,GK13
CT-S-P-O2	1.099	-144.466	-2	fitted,GK13
CT-S-P-O2	0.894	-0.54663	-3	fitted,GK13
CT-S-P-O2	0.574	145.56	4	fitted,GK13
CT-S-P-OH	0.6	180	2	same as X -p4-ss-X

Table 7: Partial charges assigned to S-farnesylcysteine (FCY)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.34734
H	H	0.264484
CA	CT	0.031692
CB	CT	-0.26402
SG	S	-0.20184
CD	CT	-0.10409
CE	CM	-0.21892
CZ	CM	0.07085
CH1	CT	-0.04273
CQ	CT	0.027635
CI	CM	-0.35699
CK	CM	0.090323
CL1	CT	-0.04274
CM	CT	0.103421
CN	CM	-0.40464
CX	CM	0.29378
CO1	CT	-0.44051
HO11	HC	0.121343
HO12	HC	0.121343
HO13	HC	0.121343
CO2	CT	-0.44051
HO21	HC	0.121343
HO22	HC	0.121343
HO23	HC	0.121343
HN	HA	0.141831
HM1	HC	0.015946
HM2	HC	0.015946
HL12	HC	0.019779
HL11	HC	0.019779
CL2	CT	-0.2463
HL21	HC	0.082942
HL22	HC	0.082942
HL23	HC	0.082942
HI	HA	0.141983
HQ1	HC	0.042875
HQ2	HC	0.042875
HH11	HC	0.036698
HH12	HC	0.036698
CH2	CT	-0.37643
HH21	HC	0.118037
HH22	HC	0.118037
HH23	HC	0.118037
HE	HA	0.148475
HD1	H1	0.115815
HD2	H1	0.115815
HB1	H1	0.137424
HB2	H1	0.137424
HA	H1	0.082107
C	C	0.59478
O	O	-0.57239

Table 8: New parameters assigned to S-farnesylcysteine (FCY)

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
CT-CM-CT	62.7	116.52	same as c3-c2-c3
S -CT-CM	80.52	104.97	same as c2-c3-ss

Table 9: Partial charges assigned to S-geranylgeranylcysteine (GCY)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.32209
H	H	0.25507
CA	CT	-0.03453
CB	CT	-0.2502
SG	S	-0.20495
CD	CT	-0.03557
CE	CM	-0.365
CZ	CM	0.09559
CH1	CT	-0.05223
CQ	CT	0.104001
CI	CM	-0.459
CK	CM	0.129409
CL1	CT	-0.10997
CM	CT	0.013532
CN	CM	-0.42434
CX	CM	0.137989
CO1	CT	-0.08978
CP	CT	-0.05759
CR	CM	-0.44096
CS	CM	0.356221
CT1	CT	-0.44037
HT11	HC	0.117773
HT12	HC	0.117773
HT13	HC	0.117773
CT2	CT	-0.44037
HT21	HC	0.117773
HT22	HC	0.117773
HT23	HC	0.117773
HR	HA	0.164711
HP1	HC	0.077291
HP2	HC	0.077291
HO11	HC	0.06016
HO12	HC	0.06016
CO2	CT	-0.29253
HO21	HC	0.086703
HO22	HC	0.086703
HO23	HC	0.086703
HN	HA	0.170225
HM1	HC	0.055852
HM2	HC	0.055852
HL11	HC	0.060298
HL12	HC	0.060298
CL2	CT	-0.27798
HL21	HC	0.087864
HL22	HC	0.087864
HL23	HC	0.087864
HI	HA	0.16437
HQ1	HC	0.041753
HQ2	HC	0.041753
HH11	HC	0.04612
HH12	HC	0.04612
CH2	CT	-0.26946
HH21	HC	0.087994
HH22	HC	0.087994
HH23	HC	0.087994
HE	HA	0.174645
HD1	H1	0.1087
HD2	H1	0.1087
HB1	H1	0.13367
HB2	H1	0.13367
HA	H1	0.090017
C	C	0.609172
O	O	-0.56005

Table 10: New parameters assigned to S-geranylgeranylcysteine (GCY)

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
S -CT-CM	80.52	104.97	same as c2-c3-ss
CT-CM-CT	62.7	116.52	same as c3-c2-c3

Table 11: Partial charges assigned to S-palmitoylcysteine (SPC)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.5412
H	H	0.287316
CA	CT	0.015923
CB	CT	0.033953
SG	S	-0.25961
CD	C	0.338563
OE2	O	-0.39585
CE1	CT	0.003578
CZ	CT	0.03369
CH	CT	0.030442
CQ	CT	-0.01959
CI	CT	-0.01288
CK	CT	0.006472
CL	CT	-0.01233
CM	CT	-0.02952
CN	CT	0.011246
CX	CT	0.067084
CO	CT	-0.02967
CP	CT	-0.01553
CR	CT	-0.01001
CS	CT	0.011184
CT	CT	-0.10496
HT1	HC	0.025133
HT2	HC	0.025133
HT3	HC	0.025133
HS1	HC	0.004001
HS2	HC	0.004001
HR1	HC	0.005534
HR2	HC	0.005534
HP1	HC	0.01281
HP2	HC	0.01281
HO1	HC	0.000367
HO2	HC	0.000367
HX1	HC	-0.01186
HX2	HC	-0.01186
HN1	HC	-0.00526
HN2	HC	-0.00526
HM1	HC	0.003749
HM2	HC	0.003749
HL1	HC	0.004846
HL2	HC	0.004846
HK1	HC	0.001841
HK2	HC	0.001841
HI1	HC	0.008669
HI2	HC	0.008669
HQ1	HC	0.002787
HQ2	HC	0.002787
HH1	HC	-0.00472
HH2	HC	-0.00472
HZ2	HC	0.000833
HZ1	HC	0.000833
HE12	HC	0.041258
HE11	HC	0.041258
HB1	H1	0.076471
HB2	H1	0.076471
HA	H1	0.084341
C	C	0.757334
O	O	-0.60806

Table 12: New parameters assigned to S-palmitoylcysteine (SPC)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
S -C	261.9	1.762	same as c -ss	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
S -C -O	81.78	122.29	same as o -c -ss	
S -C -CT	78.99	114.32	same as c3-c -ss	
CT-S -C	38.47	100.29	same as c -ss-c3	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-S -C -O	3.1	180	2	same as X -c -ss-X
CT-S -C -CT	3.1	180	2	same as X -c -ss-X

1.4.2 Glutamic Acid Modifications

Table 13: Partial charges assigned to 1-Carboxyglutamic Acid (CGU)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.36665
H	H	0.258072
CA	CT	0.01712
CB	CT	-0.04268
CG	CT	-0.13309
CD1	C	0.626223
OE12	OH	-0.53738
HE1	HO	0.403134
OE11	O	-0.51498
CD2	C	0.626223
OE21	OH	-0.53738
HE2	HO	0.403134
OE22	O	-0.51498
HG1	HC	0.048251
HB1	HC	0.052153
HB2	HC	0.052153
HA	H1	0.06349
C	C	0.6582
O	O	-0.56102

Table 14: New parameters assigned to 1-Carboxyglutamic Acid (CGU)

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
C -CT-C	64.06	111.61	same as c -c3-c

Table 15: Partial charges assigned to Pyroglutamic Acid (PCA)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.07406
CD	C	0.477825
OE	O	-0.51306
CG	CT	-0.07989
HG1	HC	0.06335
HG2	HC	0.06335
CB	CT	-0.12097
HB1	HC	0.084138
HB2	HC	0.084138
CA	CT	-0.21908
HA	H1	0.096398
C	C	0.720032
O	O	-0.58216

Table 16: New parameters assigned to Pyroglutamic Acid (PCA)

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
C -N -C	65.33	127.14	same as c -n -c

1.4.3 Phenylalanine Modifications

Table 17: Partial charges assigned to Dihydroxyphenylalanine (DAH)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.25072
H	H	0.231704
CA	CT	-0.12992
HA	H1	0.117003
CB	CT	-0.15007
HB2	HC	0.100437
HB3	HC	0.100437
CG	CA	-0.03932
CD1	CA	-0.25106
HD1	HA	0.167655
CE1	CA	0.193233
OE2	OH	-0.47707
HE2	HO	0.384123
CZ	CA	0.304404
OZ	OH	-0.55575
HZ	HO	0.43265
CE2	CA	-0.3885
HE1	HA	0.188964
CD2	CA	-0.1197
HD2	HA	0.142592
C	C	0.572011
O	O	-0.57312

1.4.4 Lysine Modifications

Table 18: Partial charges assigned to 5-hydroxylysine (LYZ)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.56082
H	H	0.353086
CA	CT	-0.21727
HA	H1	0.174249
CB	CT	-0.11071
HB2	HC	0.074178
HB3	HC	0.074178
CG	CT	-0.04314
HG2	HC	0.077764
HG3	HC	0.077764
CD	CT	0.012735
HD	H1	0.121196
OH	OH	-0.54028
HH	HO	0.387547
CE	CT	-0.03624
HE2	HP	0.118663
HE3	HP	0.118663
NZ	N3	-0.22955
HZ1	H	0.306198
HZ2	H	0.306198
HZ3	H	0.306198
C	C	0.853628
O	O	-0.62425

Table 19: Partial charges assigned to ε-N-methyllysine (MLZ)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.29213
H	H	0.250079
CA	CT	-0.13624
HA	H1	0.103386
CB	CT	-0.16062
HB2	HC	0.090555
HB3	HC	0.090555
CG	CT	0.001635
HG2	HC	0.039204
HG3	HC	0.039204
CD	CT	-0.05403
HD2	HC	0.041825
HD3	HC	0.041825
CE	CT	-0.02444
HE2	HP	0.094314
HE3	HP	0.094314
NZ	N3	-0.09565
CM	CT	-0.24004
HCM1	HP	0.146379
HCM2	HP	0.146379
HCM3	HP	0.146379
HZ2	H	0.303569
HZ3	H	0.303569
C	C	0.722968
O	O	-0.65299

Table 20: Partial charges assigned to N-ε-acetyllysine (ALY)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.38381
H	H	0.271318
CA	CT	-0.06271
HCA	H1	0.098964
CB	CT	-0.08326
HB2	HC	0.028979
HB3	HC	0.028979
CG	CT	0.067202
HG2	HC	-0.00372
HG3	HC	-0.00372
CD	CT	0.153608
HD2	HC	-0.00958
HD3	HC	-0.00958
CE	CT	-0.21575
HE2	H1	0.118563
HE3	H1	0.118563
NZ	N	-0.47859
HZ	H	0.328702
CH	C	0.633076
CH3	CT	-0.3464
HH31	HC	0.097261
HH32	HC	0.097261
HH33	HC	0.097261
OH	O	-0.57744
C	C	0.603212
O	O	-0.56837

Table 21: Partial charges assigned to N6,N6-dimethyllysine (MLY)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.28707
H	H	0.254897
CA	CT	-0.29887
HA	H1	0.140152
CB	CT	-0.04291
HB2	HC	0.070701
HB3	HC	0.070701
CG	CT	-0.02107
HG2	HC	0.038868
HG3	HC	0.038868
CD	CT	-0.1335
HD2	HC	0.074847
HD3	HC	0.074847
CE	CT	-0.02158
HE2	HP	0.102404
HE3	HP	0.102404
NZ	N3	0.024782
CH1	CT	-0.25932
HH11	HP	0.149515
HH12	HP	0.149515
HH13	HP	0.149515
CH2	CT	-0.25932
HH21	HP	0.149515
HH22	HP	0.149515
HH23	HP	0.149515
HZ2	H	0.285085
C	C	0.775625
O	O	-0.62765

Table 22: Partial charges assigned to N6,N6,N6-trimethyllysine (M3L)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.519066
H	H	0.321251
CA	CT	-0.073404
HA	H1	0.117605
CB	CT	-0.1614
HB2	HC	0.073951
HB3	HC	0.073951
CG	CT	0.006026
HG2	HC	0.044747
HG3	HC	0.044747
CD	CT	-0.029166
HD2	HC	0.052321
HD3	HC	0.052321
CE	CT	-0.144224
HE2	HP	0.111726
HE3	HP	0.111726
NZ	N3	0.188994
CM1	CT	-0.276341
HM11	HP	0.153734
HM12	HP	0.153734
HM13	HP	0.153734
CM2	CT	-0.276341
HM21	HP	0.153734
HM22	HP	0.153734
HM23	HP	0.153734
CM3	CT	-0.276341
HM31	HP	0.153734
HM32	HP	0.153734
HM33	HP	0.153734
C	C	0.801193
O	O	-0.627883

Table 23: Partial charges assigned to Pyrrolysine (PYH)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.47961
H	H	0.292739
CA	CT	0.030245
CB	CT	-0.02873
CG	CT	0.087505
CD	CT	-0.0075
CE	CT	0.012157
NZ	N	-0.21197
CAO	C	0.258518
OAD	O	-0.51949
CAR	CT	0.191249
NAL	N*	-0.6286
CAF	CM	0.306859
CAJ	CT	-0.20624
CAP	CT	0.283346
CAA	CT	-0.43443
HAA1	HC	0.097475
HAA2	HC	0.097475
HAA3	HC	0.097475
HAP	HC	0.033557
HAJ1	HC	0.061823
HAJ2	HC	0.061823
HAF	HA	0.059111
HAR	H1	0.089927
HZ	H	0.216413
HE1	H1	0.047707
HE2	H1	0.047707
HD1	HC	0.018973
HD2	HC	0.018973
HG1	HC	-0.01941
HG2	HC	-0.01941
HB1	HC	0.016614
HB2	HC	0.016614
HA	H1	0.073177
C	C	0.631538
O	O	-0.59361

Table 24: New parameters assigned to Pyrrolysine (PYH)

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
N*-CM-CT	64.95	122.54	same as c3-c2-na
N*-CM-HA	49.73	119.1	same as hc-c2-na
C -CT-N*	66.81	111.37	same as c -c3-na

Table 25: Partial charges assigned to Ornithine (ORN)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.34585
H	H	0.264421
CA	CT	-0.0387
CB	CT	-0.08467
CG	CT	0.01586
CD	CT	0.036406
NE	N3	-0.31589
HE1	H	0.304808
HE2	H	0.304808
HE3	H	0.304808
HD2	HP	0.083019
HD3	HP	0.083019
HG2	HC	0.036471
HG3	HC	0.036471
HB1	HC	0.05758
HB2	HC	0.05758
HA	H1	0.052081
C	C	0.765002
O	O	-0.61723

Table 26: Partial charges assigned to N6-propanoyllysine (PRK)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.53557
H	H	0.284431
CA	CT	0.173956
CB	CT	-0.04782
CG	CT	0.03087
CD	CT	0.057953
CE	CT	0.10608
NZ	N	-0.56466
CAL	C	0.586851
CAF	CT	0.108525
CAA	CT	-0.12553
HAA	HC	0.027248
HAAA	HC	0.027248
HAAB	HC	0.027248
HAF	HC	-0.01501
HAFA	HC	-0.01501
OAD	O	-0.58184
HZ	H	0.304349
HE2	H1	0.031469
HE3	H1	0.031469
HD2	HC	-0.00579
HD3	HC	-0.00579
HG2	HC	0.000862
HG3	HC	0.000862
HB2	HC	0.019933
HB3	HC	0.019933
HA	H1	-0.00996
C	C	0.668112
O	O	-0.60042

Table 27: Partial charges assigned to N6-butanoyllysine (BTK)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.29559
H	H	0.260432
CA	CT	-0.65983
CB	CT	-0.02307
CG	CT	-0.02617
CD	CT	-0.06617
CE	CT	0.012051
NZ	N	-0.48779
CAN	C	0.529418
CAJ	CT	-0.05549
CAF	CT	0.19493
CAA	CT	-0.33896
HAA	HC	0.077688
HAAA	HC	0.077688
HAAB	HC	0.077688
HAF	HC	-0.00843
HAFA	HC	-0.00843
HAJ	HC	0.014412
HAJA	HC	0.014412
OAD	O	-0.55978
HNZ	H	0.316053
HE2	H1	0.07213
HE3	H1	0.07213
HD2	HC	0.035549
HD3	HC	0.035549
HG2	HC	0.044502
HG3	HC	0.044502
HB2	HC	0.072294
HB3	HC	0.072294
HA	H1	0.259731
C	C	0.886022
O	O	-0.63977

1.4.5 Proline Modifications

Table 28: Partial charges assigned to 3,4-hydroxyproline (DHP)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.08371
CD1	CT	-0.18024
HD11	H1	0.101469
HD12	H1	0.101469
CG1	CT	0.177548
HG1	H1	0.132692
OD2	OH	-0.62756
HD21	HO	0.429709
CB	CT	0.092582
HB1	H1	0.105269
OG2	OH	-0.57053
HG21	HO	0.411146
CA	CT	-0.3098
HA	H1	0.133011
C	C	0.685117
O	O	-0.59817

Table 29: Partial charges assigned to 3-hydroxyproline (HY3)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.1976
CD	CT	-0.03236
HD1	H1	0.063033
HD2	H1	0.063033
CG1	CT	-0.15073
HG11	HC	0.089019
HG12	HC	0.089019
CB	CT	0.207207
HB1	H1	0.043876
OG2	OH	-0.59443
HG21	HO	0.380575
CA	CT	-0.00277
HA	H1	0.078151
C	C	0.500171
O	O	-0.53619

Table 30: Partial charges assigned to 4-hydroxyproline (HYP)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.13142
CD	CT	-0.24411
HD21	H1	0.105581
HD22	H1	0.105581
CG	CT	0.319283
HG1	H1	0.066129
OD1	OH	-0.65864
HD11	HO	0.429398
CB	CT	-0.08497
HB1	HC	0.039646
HB2	HC	0.039646
CA	CT	-0.11104
HA	H1	0.102544
C	C	0.59469
O	O	-0.57232

1.4.6 Arginine Modifications

Table 31: Partial charges assigned to Dimethylarginine (Anti-symmetric) (DA2)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.376741
H	H	0.270465
CA	CT	-0.119858
CB	CT	-0.129377
CG	CT	-0.016346
CD	CT	-0.092919
NE	N2	-0.288262
CZ	CM	0.186646
NH1	N2	0.057692
C1	CT	-0.204521
HC11	H1	0.116462
HC12	H1	0.116462
HC13	H1	0.116462
C2	CT	-0.204521
HC21	H1	0.116462
HC22	H1	0.116462
HC23	H1	0.116462
NH2	N2	-0.468942
HN1	H	0.317322
HN2	H	0.317322
HNE	H	0.297160
HCD1	H1	0.113491
HCD2	H1	0.113491
HCG1	HC	0.055975
HCG2	HC	0.055975
HCB1	HC	0.078332
HCB2	HC	0.078332
HCA	H1	0.106985
C	C	0.770687
O	O	-0.617160

Table 32: New parameters assigned to Dimethylarginine (Anti-symmetric) (DA2)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
N2-CM	449	1.364	same as cc-nh	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
N2-CM-N2	72.93	115.96	same as nh-cc-nh	
CT-N2-CM	64.702	117.79	Calculated empirically	
H -N2-CM	49.11	119.38	same as c2-n3-hn	
CT-N2-CT	64.01	110.9	same as c3-n3-c3	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
N2-CM-N2-H	0.3	180	2	same as X -c2-n3-X
CT-N2-CM-N2	0.4566	2.705	-1	fitted,GK13
CT-N2-CM-N2	12.726	-175.79	-2	fitted,GK13
CT-N2-CM-N2	0.283	-173.945	-3	fitted,GK13
CT-N2-CM-N2	0.1875	-161.375	4	fitted,GK13

Table 33: Partial charges assigned to ω-Methylarginine (DA1)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.50278
H	H	0.314656
CA	CT	0.00531
CB	CT	-0.15979
CG	CT	-0.12551
CD	CT	0.109358
NE	N2	-0.57122
CZ	CM	0.476462
NH1	N2	-0.61243
HH11	H	0.364073
HH12	H	0.364073
NH2	N2	-0.30094
C1	CT	-0.19592
HC11	H1	0.129204
HC12	H1	0.129204
HC12	H1	0.129204
HH2	H	0.308468
HNE	H	0.406059
HCD1	H1	0.0707
HCD2	H1	0.0707
HCG1	HC	0.088214
HCG2	HC	0.088214
HCB1	HC	0.08218
HCB2	HC	0.08218
HCA	H1	0.110595
C	C	0.709342
O	O	-0.56962

Table 34: New parameters assigned to ω-Methylarginine (DA1)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
N2-CM	449	1.364	same as cc-nh	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
N2-CM-N2	72.93	115.96	same as nh-cc-nh	
CT-N2-CM	64.702	117.79	Calculated empirically	
H -N2-CM	49.11	119.38	same as c2-n3-hn	
TORSION				
Atom Types	$V_n/2$	γ	*n*	*Note*
N2-CM-N2-H	0.3	180	2	same as X -c2-n3-X
CT-N2-CM-N2	0.3	180	2	same as X -c2-n3-X

Table 35: Partial charges assigned to Dimethylarginine (Symmetric) (2MR)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.33626
H	H	0.255401
CA	CT	-0.03012
HA	H1	0.054492
CB	CT	-0.10153
HB2	HC	0.050864
HB3	HC	0.050864
CG	CT	0.032724
HG2	HC	0.033302
HG3	HC	0.033302
CD	CT	0.034682
HD2	H1	0.071285
HD3	H1	0.071285
NE	N2	-0.32964
HE	H	0.321023
CZ	CM	0.13247
NH1	N2	-0.11712
CQ1	CT	-0.32341
HQ11	H1	0.151691
HQ12	H1	0.151691
HQ13	H1	0.151691
HH1	H	0.270505
NH2	N2	-0.11712
CQ2	CT	-0.32341
HQ21	H1	0.151691
HQ22	H1	0.151691
HQ23	H1	0.151691
HH2	H	0.270505
C	C	0.650222
O	O	-0.56445

Table 36: New parameters assigned to Dimethylarginine (Symmetric) (2MR)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
CT-N2	320.6	1.47	same as c3-n3	
N2-CM	449	1.364	same as cc-nh	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CT-N2-CM	64.702	117.79	Calculated empirically	
N2-CM-N2	72.93	115.96	same as nh-cc-nh	
H -N2-CM	49.11	119.38	same as c2-n3-hn	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
N2-CM-N2-H	0.3	180	2	same as X -c2-n3-X
CT-N2-CM-N2	0.3	180	2	same as X -c2-n3-X

1.4.7 Serine Modifications

Table 37: Partial charges assigned to neutral Phosphoserine (SEN)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.22283
H	H	0.256767
CA	CT	-0.49904
CB	CT	-0.03007
OG	OS	-0.23226
P	P	0.94873
O1P	OH	-0.5705
HOP2	HO	0.45302
O2P	OH	-0.5705
HOP3	HO	0.45302
O3P	O2	-0.64688
HB2	H1	0.129004
HB2	H1	0.129004
HA	H1	0.224707
C	C	0.779535
O	O	-0.60172

Table 38: New parameters assigned to neutral Phosphoserine (SEN)

BOND			
Atom Types	K_r	r_{eq}	*Note*
OH - P	525	1.61	Adapted from [1]
OS - P	525	1.61	Adapted from [1]

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
CT - OS - P	100	120.5	Adapted from [1]
O2 - P - OH	140	108.23	Adapted from [1]
HO - OH - P	140	108.5	Adapted from [1]
OH-P -OH	72.99	95.71	same as oh-p4-oh

TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-OS-P-O2	1.004	-20.493	-1	fittedGK13
CT-OS-P-O2	0.664	3.177	-2	fittedGK13
CT-OS-P-O2	0.141	170.368	-3	fittedGK13
CT-OS-P-O2	0.032	101.83	4	fittedGK13

Table 39: Partial charges assigned to doubly negative Phosphoserine (SEP)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.46664
H	H	0.2967
CA	CT	0.031998
CB	CT	0.150957
OG	OS	-0.48887
P	P	1.159923
O1P	O2	-0.86142
O2P	O2	-0.86142
O3P	O2	-0.86142
HB2	H1	0.009656
HB3	H1	0.009656
HA	H1	0.036275
C	C	0.440306
O	O	-0.59571

Table 40: New parameters assigned to doubly negative Phosphoserine (SEP)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
OS - P	525	1.61	Adapted from [1]	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CT - OS - P	100	120.5	Adapted from [1]	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-OS-P-O2	0.623	-63.53	-1	fittedGK13
CT-OS-P-O2	0.213	-145.686	-2	fittedGK13
CT-OS-P-O2	2.246	-56.664	-3	fittedGK13
CT-OS-P-O2	0.176	148.989	-4	fittedGK13
CT-OS-P-O2	0.051	30.301	-5	fittedGK13
CT-OS-P-O2	0.188	-137.385	6	fittedGK13

1.4.8 Threonine Modifications

Table 41: Partial charges assigned to neutral Phosphothreonine (TON)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.35837
H	H	0.280143
CA	CT	-0.06651
CB	CT	0.045716
OG1	OS	-0.33801
P	P	0.957556
O1P	OH	-0.54678
HOP2	HO	0.442186
O2P	OH	-0.54678
HOP3	HO	0.442186
O3P	O2	-0.65038
CG2	CT	-0.26498
HG21	HC	0.099671
HG22	HC	0.099671
HG23	HC	0.099671
HB	H1	0.130699
HA	H1	0.142079
C	C	0.600977
O	O	-0.56874

Table 42: New parameters assigned to neutral Phosphothreonine (TON)

BOND			
Atom Types	K_r	r_{eq}	*Note*
P - OH	525	1.61	Adapted from[1]
P - OS	525	1.61	Adapted from[1]

ANGLE			
Atom Types	K_Θ	Θ_{eq}	*Note*
OH - P - OH	140	108.23	Adapted from[1]
CT - OS - P	100	120.5	Adapted from[1]
O2 - P - OH	140	108.23	Adapted from[1]
HO - OH - P	140	108.5	Adapted from[1]

Table 43: Partial charges assigned to doubly negative Phosphothreonine (TPO)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.55078
H	H	0.344635
CA	CT	-0.06056
CB	CT	0.36394
OG1	OS	-0.52192
P	P	1.21988
O1P	O2	-0.88414
O2P	O2	-0.88414
O3P	O2	-0.88414
CG2	CT	-0.48488
HG21	HC	0.106285
HG22	HC	0.106285
HG23	HC	0.106285
HB	H1	0.049717
HA	H1	0.075833
C	C	0.529578
O	O	-0.63189

Table 44: New parameters assigned to doubly negative Phosphothreonine (TPO)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
P - OS	525	1.61	Adapted from [1]	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CT - OS - P	100	120.5	Adapted from [1]	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CT-OS-P-O2	0.623	-63.53	-1	fittedGK13
CT-OS-P-O2	0.213	-145.686	-2	fittedGK13
CT-OS-P-O2	2.246	-56.664	-3	fittedGK13
CT-OS-P-O2	0.176	148.989	-4	fittedGK13
CT-OS-P-O2	0.051	30.301	-5	fittedGK13
CT-OS-P-O2	0.188	-137.385	6	fittedGK13

1.4.9 Tryptophan Modifications

Table 45: Partial charges assigned to 7-hydroxytryptophan (0AF)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.36882
H	H	0.271475
CA	CT	-0.05267
CB	CT	-0.22614
CG	C*	-0.01307
CD2	CB	0.005485
CE2	CN	0.210593
NE1	NA	-0.35741
CD1	CW	-0.18754
HD1	H4	0.204222
HE1	H	0.370713
CZ2	CA	0.179687
CH2	CA	-0.2855
CZ3	CA	-0.18205
CE3	CA	-0.28357
HE3	HA	0.170303
HZ3	HA	0.155944
HH2	HA	0.158382
OH2	OH	-0.56907
HH1	HO	0.448296
HB2	HC	0.110235
HB3	HC	0.110235
HA	H1	0.096844
C	C	0.603467
O	O	-0.57004

Table 46: New parameters assigned to 7-hydroxytryptophan (0AF)

ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CN-CA-OH	71.64	122.07	same as c2-c2-oh	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CN-CA-OH-HO	0.958	176.09	-1	fittedGK
CN-CA-OH-HO	0.824	178.125	-2	fittedGK
CN-CA-OH-HO	0.176	1.23795	3	fittedGK

Table 47: Partial charges assigned to 5-hydroxytryptophan (HTR)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.37972
H	H	0.256809
CA	CT	0.007369
CB	CT	-0.36861
CG	C*	-0.02678
CD2	CB	0.017597
CE2	CN	0.201751
NE1	NA	-0.40105
CD1	CW	-0.13154
HD11	H4	0.179129
HE11	H	0.369152
CZ2	CA	-0.28376
CH2	CA	-0.24059
CZ3	CA	0.220602
OH	OH	-0.57021
HOH1	HO	0.435498
CE3	CA	-0.21633
HE31	HA	0.17718
HH21	HA	0.152957
HZ21	HA	0.176569
HB1	HC	0.14975
HB2	HC	0.14975
HA	H1	0.090597
C	C	0.613083
O	O	-0.57921

1.4.10 Tyrosine Modifications

Table 48: Partial charges assigned to neutral Phosphotyrosine (PTN)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.3583
H	H	0.250246
CA	CT	-0.03242
CB	CT	-0.18335
CG	CA	0.097257
CD1	CA	-0.17109
CE1	CA	-0.2109
CZ	CA	0.229335
OH	OS	-0.28548
P	P	0.981548
O3P	OH	-0.55986
HO3P	HO	0.44556
O2P	OH	-0.55986
HO2P	HO	0.44556
O1P	O2	-0.66117
CE2	CA	-0.2109
CD2	CA	-0.17109
HD2	HA	0.142798
HE2	HA	0.174748
HE1	HA	0.174748
HD1	HA	0.142798
HB2	HC	0.088647
HB3	HC	0.088647
HA	H1	0.084605
C	C	0.634319
O	O	-0.57638

Table 49: New parameters assigned to neutral Phosphotyrosine (PTN)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
CA-OS	392.6	1.357	same as c2-os	
OS-P	525	1.61	Adapted from [1]	
OH-P	525	1.61	Adapted from [1]	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CA-OS-P	100	120.5	Adapted from [1]	
OS-P -OH	140	108.23	Adapted from [1]	
OS-P -O2	140	108.23	Adapted from [1]	
OH-P -O2	140	108.23	Adapted from [1]	
OH-P -OH	140	108.23	Adapted from [1]	
CA-CA-OS	71.04	121.89	same as c2-c2-os	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CA-CA-OS-P	1.05	180	2	same as X -c2-os-X
CA-OS-P-O2	1.1295	10.51	-1	fitted,GK
CA-OS-P-O2	0.2271	-15.788	-2	fitted,GK
CA-OS-P-O2	0.5013	-93.029	-3	fitted,GK
CA-OS-P-O2	0.2058	157.172	4	fitted,GK

Table 50: Partial charges assigned to doubly negative Phosphotyrosine (PTR)

ATOM NAME	ATOM TYPE	PARTIAL CHARGE
N	N	-0.32312
H	H	0.24754
CA	CT	-0.0162
CB	CT	-0.25366
CG	CA	-0.01881
CD1	CA	-0.19149
CE1	CA	-0.29064
CZ	CA	0.33299
OH	OS	-0.49686
P	P	1.204
O1P	O2	-0.84601
O2P	O2	-0.84601
O3P	O2	-0.84601
CE2	CA	-0.29064
CD2	CA	-0.19149
HD21	HA	0.12754
HE21	HA	0.16778
HE11	HA	0.16778
HD11	HA	0.12754
HB1	HC	0.09489
HB2	HC	0.09489
HA	H1	0.07288
C	C	0.68708
O	O	-0.71402

Table 51: New parameters assigned to doubly negative Phosphotyrosine (PTR)

BOND				
Atom Types	K_r	r_{eq}	*Note*	
CA-OS	392.6	1.357	same as c2-os	
OS-P	525	1.61	Adapted from[1]	
ANGLE				
Atom Types	K_Θ	Θ_{eq}	*Note*	
CA-OS-P	100	120.5	Adapted from[1]	
OS-P -O2	140	108.23	Adapted from[1]	
CA-CA-OS	71.04	121.89	same as c2-c2-os	
TORSION				
Atom Types	$V_n/2$	γ	n	*Note*
CA-CA-OS-P	1.05	180	2	same as X -c2-os-X
CA-OS-P-O2	0.4659	179.937	-1	fitted,GK13
CA-OS-P-O2	0.30383	121.386	-2	fitted,GK13
CA-OS-P-O2	0.6954	-18.001	-3	fitted,GK13
CA-OS-P-O2	0.1536	-32.4026	4	fitted,GK13

2 Supplementary Methods

2.1 Automated Curation of Pairs of Modified and Unmodified Protein Structures Contained in the PDB

It was necessary to establish a baseline of the background levels of secondary and tertiary structure dissimilarity as a result of post-translational modifications. Thus, a set of pairs of modified and unmodified experimental structures against which to test the parameters of the modified amino acids was needed. To establish this, all pairs of modified and unmodified structures had to be enumerated. According to the PSI-MOD protein modification browser previously available on the beta-staging webpage of the PDB, as of April 2012, 26,823 instances of modified amino acids were found. Some instances may be redundant across multiple modification types occurring on a single structure. To the best of our knowledge, this webpage is unfortunately not available anymore. From this set, 15,573 PDB structures contained disulfide bridges. The IDs containing disulfide bridges were subtracted from the global list. From this remaining set of structures, a Python script was written to go through each PDB ID containing one or more modifications and tabulate information contained in the lines denoted by *MODRES* in each structure.

A Python script was written to take as input all PDB structures with a single modification on them that is not a disulfide bridge. This initial set of structures at the time of the search contained 2136 PDB IDs. The script searches through each HTML page and stores the title, authors, number of chains, the number of modifications, and any related entries to the PDB. Then, for each related entry, the script automatically opens the related PDB's HTML form and searches through and populates the number of modifications, title, and author information; if the authors are the same, the title is $> 50\%$ similar, and there is one less modification than in the reference structure, then adds it to the candidate list of pairs found. The logical conditions were

converged upon through trial and error, eliminating as many false positives as possible while making sure true positives were always found. The initial list contained 170 pairs.

Upon populating the master list, each pair was manually confirmed to remove any spurious pairs caused by pairs of structures with a complex with a modified protein (for example, a methylated histone) and not in complex with the histone region, which were false positives. Similarly, false positives corresponding to sequences that had mutations other than the post-translational modification on the structure were removed. The final list of pairs after curating the entire database contained 40 pairs of structures. The approach was extended to find all pairs of multiply modified structures and their corresponding unmodified counterparts.

It is important to note that in the PDB there are structures that have been chemically modified post-translationally to enhance crystallizability. One such example is the dimethylation of surface exposed lysines. These are chemical modifications to the structure of the protein that also would yield a hit for a post-translational modification by name, albeit the modification may not have been done by an enzyme, but chemically by a method such as reductive methylation. Nevertheless, the protein structures are chemically modified in a way that produces the same side-chain modification as might be achieved by an enzyme; thus, these structures were included in analysis.

2.2 Structural Characterization of Baseline Secondary Structure and Tertiary Structure Features to Establish Background Control Levels

Each modified and unmodified structure was next assessed for local secondary structure for which to serve as the control to compare torsion angle distributions during simulation. It was observed that there was sometimes disagreement between STRIDE,[2] DSSP,[3] and the author's secondary structure assignment. Therefore, for each modified/unmodified pair, the consensus

of the three assignments was taken to assign the initial secondary structure. Next, each modified structure was aligned to its unmodified counterpart using the align command in PyMol[4] to calculate the C_α root-mean squared deviation (RMSD) between the structures. A cutoff of 100Å was used to ensure that no atoms were rejected during the RMSD calculation. We also populated information on whether it was known if the modification was the cause of the structural change based on literature search. This was done to establish the background level of structural dissimilarity between the modified and unmodified structures. Using functions available in the BioPython set of programs,[5] we constructed contact maps for each modified/unmodified pair to visually represent the interactions of every pair of amino acids i and j on the sequences. Contacts were calculated for residues that had a sequence distance of 5 or more apart.

2.3 Homology Modeling Missing Residues in Pairs of Modified/Unmodified Proteins Curated

It was observed that among the pairs of modified/unmodified structures, there existed structures with missing residues. For each structure with missing residues, the original PDB file was used as a template in SwissModel[6–8] to create a homology model for the purpose of generating initial coordinates of the missing residues. This approach generated a structure for the full length of the sequence contained in the PDB and removed any missing residues. This was done as simulation artifacts were observed in the energetics when performing simulations with the residues missing. Then, the structures with the missing residues filled in were aligned in PyMol to the original PDB structures and the C_α atom coordinates were extracted from the homology model and added to the PDB structure. When **tleap** was called subsequently to prepare these structures for all-atom molecular dynamics simulations, it filled in the missing atoms according to the AMBER ff03 forcefield. In Figure 1, an example is presented of the structure before and

after the homology model was built for the purpose of filling in the missing residues. For the structures simulated, no missing residues corresponded to a position containing a PTM.

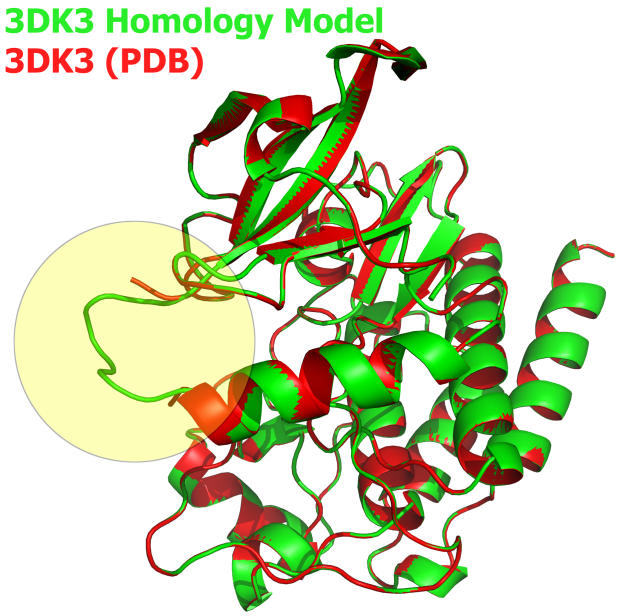


Figure 1: Comparison of PDB-deposited coordinates with missing residues and the homology model generated with the missing residues filled in based on the PDB-deposited structure using SwissModel. The coordinates from the PDB structure were preserved and only missing C_α atoms were filled in.

3 Derivation of Restrained Electrostatic Potential

As described by Bayly, Kollman, and coworkers,[9] to fit partial charges to atoms, first the quantum mechanically calculated electrostatic potential (ESP) must be determined for each point i in a grid in the solvent-accessible region of a molecule but beyond the vdw radius. The derivation follows the work of Bayly et al. for completion.[9] Charges q_j are fit to every atom center j by a least squares fitting procedure.

The ESP calculated in molecular mechanics following Bayly, Kollman, and coworkers[9] is

$$\widetilde{V}_{el}(r_i) = \sum_{j=1}^{Natoms} \frac{q_j}{r_{ij}} \tag{1}$$

where r_{ij} is the distance between grid point i to atom center j. The classically calculated ESP is denoted as $\widetilde{V}_{el}(r_i)$ and the QM calculated ESP are function evaluations in Gaussian09[10] as described in the methods and is denoted as $V_{el}(r_i)$. The objective function to be minimized is

$$\chi^2_{esp} = \sum_{i=1}^{M} \left(V_{el}(r_i) - \widetilde{V}_{el}(r_i) \right)^2 \tag{2}$$

where M is the total number of grid points. To fit the charges to the points, the objective is to minimize this function. Since the objective function is convex, the minimum will be the global minimum. The minimum is obtained where

$$\frac{\partial \chi^2_{esp}}{\partial q_j} = -2 \sum_{i}^{M} \frac{V_{el}(r_i) - \widetilde{V}_{el}(r_i)}{r_{ij}} = 0 \quad \forall\, j. \tag{3}$$

This can be solved using a system of equations in matrix form, with the constraint that the sum

of all charges equals the total molecular charge

$$\sum_{j}^{Natoms} q_j = q_{total}. \tag{4}$$

RESP introduces restraints in the form of a penalty function into the fitting process.[9] To utilize the RESP charge-fitting method, a penalty function must be added to χ^2_{esp} , making the objective function to be minimized

$$\chi^2 = \chi^2_{esp} + \chi^2_{rstr} \tag{5}$$

with a minimum at

$$\frac{\partial \chi^2}{dq_j} = \frac{\partial \chi^2_{esp}}{\partial q_j} + \frac{\partial \chi^2_{rstr}}{\partial q_j} = 0 \quad \forall\, j. \tag{6}$$

The penalty function is a hyperbolic function with form

$$\chi^2_{rstr} = a \sum_{j}^{Natoms} \left(\left(q_j^2 + b^2 \right)^{1/2} - b \right) \tag{7}$$

where scale factor a denotes the asymptotic limit of the strength of the restraint, and b is the "tightness" of the hyperbola around the minimum.[9] The derivative of the second term in Equation 7 becomes

$$\frac{\partial \chi^2_{rstr}}{dq_j} = a q_j \left(q_j^2 + b^2 \right)^{-1/2}. \tag{8}$$

Solving for the set of q_j after simplification becomes the solution of the matrix equation[11]

$$\mathbf{Aq} = \mathbf{B} \tag{9}$$

with the off-diagonal ESP-dependent elements

$$A_{jk} = \sum_{i=1}^{M} \frac{1}{r_{ij} r_{ik}} \quad , \tag{10}$$

the diagonal ESP-dependent elements

$$A_{jj} = \sum_{i=1}^{M} \frac{1}{r_{ij}^2} + \frac{\partial \chi_{rstr}^2}{dq_j} \quad , \tag{11}$$

and finally the ESP-dependent elements of $\mathbf{B}$[9]

$$B_j = \sum_{i=1}^{M} \frac{V_i}{r_{ij}^2}. \tag{12}$$

4 Supplementary Results

4.1 Energetic and structural statistics collected over the course of 5ns all-atom MD simulations for 17 pairs of modified/unmodified structures

Each modified and unmodified structure in the PDB that underwent MD simulation was first minimized to the nearest local minimum. A comparison of the structural deviation resulting from the unrestrained local minimization is shown in Table 52. The average C_α RMSDs of the minimized structures to their PDB modified/unmodified counterparts were tabulated. The results show virtually identical deviation for the modified and unmodified structures. Namely, the average C_α RMSD of the minimized modified structures to their PDB counterparts was 0.42 $\pm$ 0.15 Å, whereas for their analogous unmodified counterparts, this value was 0.43$\pm$0.19Å.

Table 52: Comparison of deviations caused by unrestrained local minimizations of starting structure to nearest local minimum.

Modified PDB	*Unmodified PDB*	*Modification Type*	*C_α RMSD of Start to Minimized Modified Structure*	*C_α RMSD of Start to Minimized Unmodified Structure*
2IVT	2IVS	PHOSPHOTYROSINE	0.40	0.42
3DK6	3DK3	PHOSPHOTYROSINE	0.18	0.22
1U54	1U46	PHOSPHOTYROSINE	0.50	0.35
1D4W	1D4T	PHOSPHOTYROSINE	0.22	0.47
2XKK	2XKJ	PHOSPHOTYROSINE	0.41	0.28
2JFM	2J51	PHOSPHOTHREONINE	0.35	0.22
3D5W	3D5U	PHOSPHOTHREONINE	0.43	0.33
3CKX	3CKW	PHOSPHOTHREONINE	0.39	0.45
1V50	1V4Z	PHOSPHOTHREONINE	0.55	0.37
2L5J	2L5I	PHOSPHOSERINE	0.74	0.84
3NAY	3NAX	PHOSPHOSERINE	0.23	0.32
1KKM	1KKL	PHOSPHOSERINE	0.44	0.41

Continued on next page

Table 52 – continued from previous page

Modified PDB	*Unmodified PDB*	*Modification Type*	*C_α RMSD of Start to Minimized Modified Structure*	*C_α RMSD of Start to Minimized Unmodified Structure*
3IAF	3IAE	PHOSPHOSERINE	0.31	0.28
1H4X	1H4Y	PHOSPHOSERINE	0.26	0.31
3F9X	3F9W	N-DIMETHYL-LYSINE	0.52	0.71
3S7D	3S7F	N-METHYL-LYSINE	0.53	0.53
2KWJ	2KWK	N(6)-ACETYLLYSINE	0.63	0.82
Total	**17 Pairs**		**0.42 ± 0.15**	**0.43 ± 0.19**

Table 53 presents all raw simulation data as presented in the main text.

Table 53: Energetic and structural statistics collected over the course of 5ns all-atom MD simulations for 17 pairs of modified/unmodified structures. Mean values are given for each metric ± one standard deviation.

Modified PDB	**Mean Total Energy (kcal/mol)**	**Mean C_α RMSD from Minimized PDB**	**Min/Max C_α RMSD**	**Unmodified PDB**	**Mean Total Energy (kcal/mol)**	**Mean C_α RMSD from Minimized PDB**	**Min/Max C_α RMSD**	**Modification Type**
2IVT	-1986.29±75.90	4.03±0.32	2.70/4.72	*2IVS*	-1741.71±70.04	2.55±0.30	1.38/3.75	O-PHOSPHOTYROSINE
3DK6	-1438.27±61.14	4.03±0.50	2.47/5.20	*3DK3*	-928.45±59.45	3.60±0.58	1.73/4.74	O-PHOSPHOTYROSINE
1U54	-2020.73±67.33	2.15±0.21	1.59/2.95	*1U46*	-1372.19±65.62	2.67±0.27	1.87/3.34	O-PHOSPHOTYROSINE
1D4W	-345.56±12.85	4.30±0.94	2.20/7.21	*1D4T*	-110.48±13.98	3.91±0.82	1.89/6.11	O-PHOSPHOTYROSINE
2XKK	-4823.84±90.02	4.37±0.70	2.32/5.89	*2XKJ*	-4272.45±89.35	4.00±0.25	3.15/4.64	O-PHOSPHOTYROSINE
2JFM	-1855.43±70.59	4.34±0.60	2.78/5.86	*2J51*	-1619.59±71.33	4.36±0.68	1.98/5.72	PHOSPHOTHREONINE
3D5W	-1905.28±68.65	4.85±1.57	1.67/7.01	*3D5U*	-1471.89±70.44	6.55±1.74	1.92/9.10	PHOSPHOTHREONINE
3CKX	-1651.22±70.97	3.84±0.50	1.91/4.89	*3CKW*	-1306.53±69.55	3.37±0.45	2.20/4.30	PHOSPHOTHREONINE
1V50	-561.89±20.22	4.35±0.65	1.70/5.74	*1V4Z*	-331.13±16.67	1.98±0.36	0.98/3.70	PHOSPHOTHREONINE
2L5J	-403.68±29.23	4.94*[1]±0.27	3.46/5.61	*2L5I*	-169.84±18.58	3.79±0.38	2.41/5.38	PHOSPHOSERINE

Continued on next page

[1]This value is the one from the 66 ns simulation as described in the Supplementary Material.

Table 53 – continued from previous page

Modified PDB	Mean Total Energy (kcal/mol)	Mean C_α RMSD from Minimized PDB	Min/Max C_α RMSD	Unmodified PDB	Mean Total Energy (kcal/mol)	Mean C_α RMSD from Minimized PDB	Min/Max C_α RMSD	Modification Type
3NAY	-1693.05±70.15	2.63±0.38	1.53/3.61	*3NAX*	-1278.23±67.66	3.38±0.80	1.53/5.00	PHOSPHOSERINE
1KKM	-797.96±35.81	3.59±0.91	0.90/5.19	*1KKL*	-551.04±35.66	3.71±0.75	1.01/5.40	PHOSPHOSERINE
3IAF	-1479.77±97.19	3.79±0.33	2.32/4.64	*3IAE*	-995.34±91.96	2.82±0.29	1.94/3.67	PHOSPHOSERINE
1H4X	-1525.55±43.18	2.38±0.36	1.13/3.55	*1H4Y*	-1242.64±42.68	1.39±0.20	0.88/2.03	PHOSPHOSERINE
3F9X	-311.02±15.57	4.92±1.96	0.74/7.05	*3F9W*	-331.17±13.17	3.97±0.92	1.59/6.73	N-DIMETHYL-LYSINE
3S7D	-21.06±9.67	2.43±0.07	2.02/2.74	*3S7F*	-35.32±10.07	2.27±0.38	0.46/2.87	N-METHYL-LYSINE
2KWJ	-388.89±19.35	5.77±0.65	3.57/7.99	*2KWK*	-385.13±18.17	5.80±1.12	2.85/8.47	N(6)-ACETYLLYSINE
Average		3.92				3.53		
Total	**34**							

4.2 Supplementary Discussion on 2L5J/2L5I Deviation in Simulation

The structural deviations in unrestrained simulations of the modified and unmodified structures were in general comparable on the aggregate, with the exception of 2L5J/2L5I. The deviation could be caused by several factors. First, all simulations performed in this study were performed at physiological pH. Thus, the protonation state simulated for the 21 amino acid 2L5J contained a -2 charge. The NMR structure was solved at a weakly acidic pH of 4.0,[12] and thus the protonation state of the structure possessed a -1 charge. The change in charge based on the protonation state may have destabilized the structure. It is possible that the phosphorylated structure may unfold at physiological pH, or may be less stably folded, which may be the reason Page et al.[12] found the phosphorylated structure to aggregate at moderate concentrations whereas the unmodified protein did not. During the simulation, the structure began at 3.48 Å from the native during the production stage. The structure within 1 ns unfolded to a near-extended conforma-

tion, 12 Å from the native which lead to the high average RMSD. It was interestingly observed that in the subsequent portions of the 5 ns simulation, the structure refolded itself within 4.67 Å of the native, which is a considerable gain after such a large initial deviation, without any restraints. The simulation was subsequently repeated identically with a different initial seed and for 66 ns instead of 5 ns, and the structure in the subsequent simulation remained stable and behaved similarly to the simulation of the unmodified structure, with an average C_α RMSD of 4.94 $\pm$ 0.27. At no point during the simulation did it exhibit unfolding and refolding, but rather remained stably folded near the NMR structure throughout. The phosphorylated serine in position 10, with -1 charge on the NMR structure of 2L5J forms a hydrogen bond with the backbone of the preceding arginine residue. In the simulation, where the phosphate group had a -2 charge, when the structure refolded, the same backbone interaction was preserved, but two simultaneous salt bridges were formed with Arg9 and Arg14.

The other phosphorylated structures simulated did not unfold in their physiological protonation state, and since they used the same generalized Born model, it is unlikely a possibility that it is an artifact of the generalized Born model. It is likely performing the simulations with restraints in the heating and equilibration stages prior to production would have lead to lesser deviations from the native, but, since the goal here was to strictly test the forcefields, we did not want to artificially keep them intact and allowed them to naturally sample space. Overall, with the exception of 2L5J, we observed that the parameters did not cause the structures tested to unfold or deviate more than their unmodified counterpart which only uses the parameters derived in ff03.

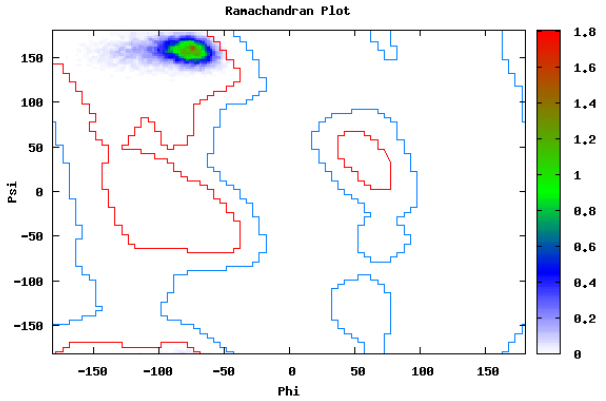
4.3 Plots of Energetics, Secondary Structure, and Tertiary Structure Space Sampled for Pairs of Modified/Unmodified Proteins

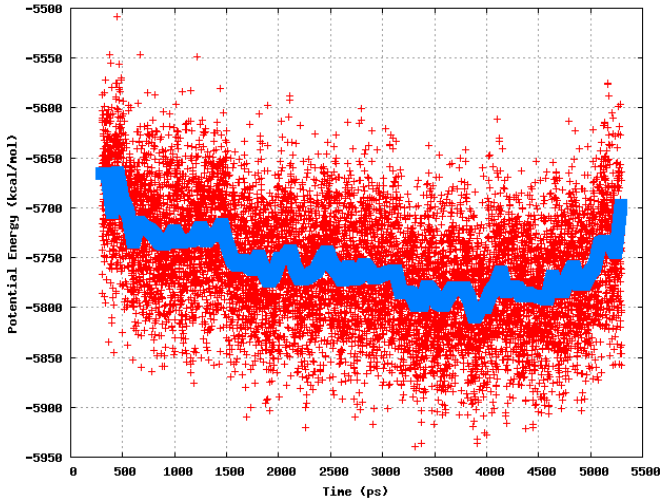
In this section, we supply plots corresponding to several of the parameters described in the paper for each pair of modified/unmodified structures simulated. In each figure, the top, middle, and bottom plots correspond to the same information, with the modified plots on the left, and the unmodified plots on the right. The top plot is a Ramachandran plot corresponding to the residue that becomes modified over the course of the simulation. The color bar represents the percent occupation per surface, which is proportional to the propensity of the torsion angle to be populated in that bin. In most cases, the secondary structure of the modified structure is preserved relative to its initial configuration, as well as its unmodified counterpart. The middle is a graph of Potential Energy vs. Time over the course of the simulation. All simulations were found to be stable. The bottom plots on each page were plots of the time evolution of the C_α RMSD from the native PDB structure for the modified and unmodified structures. This data is summarized in the main text via a mean and standard deviation. Additionally, several independent longer simulations using the same protocol described previously with a 25 ns production phase were also performed for 3 modified structures. The simulation results exhibited similar results to the short 5 ns simulations. The Ramachandran, potential energy, and time evolution of C_α RMSD plots as described are next presented as supporting material.

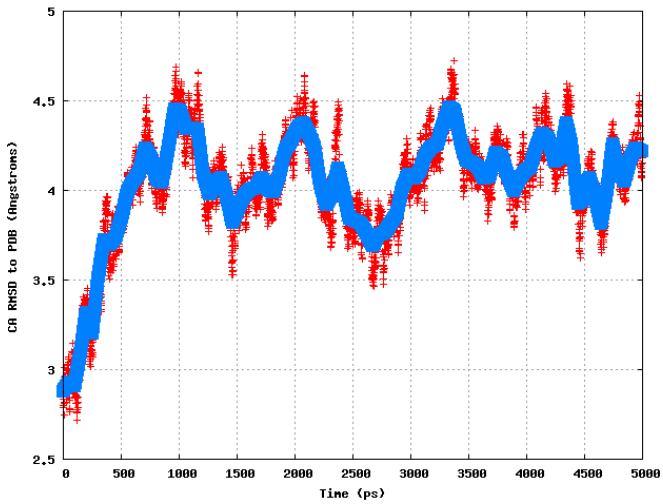
Modified PDB Simulation (Left)

Unmodified PDB Simulation (Right)

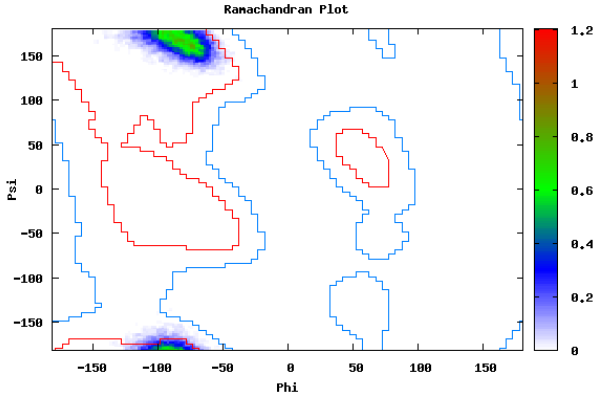
2IVT

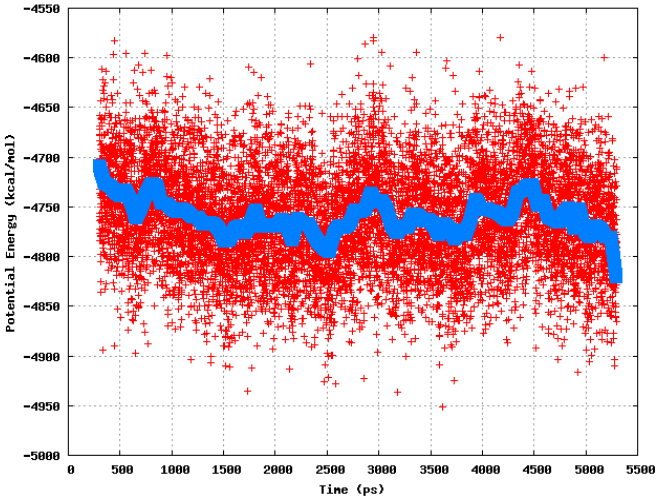


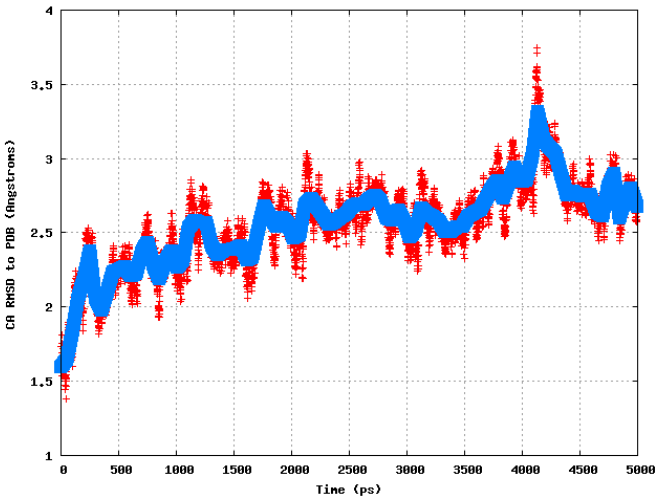




2IVS

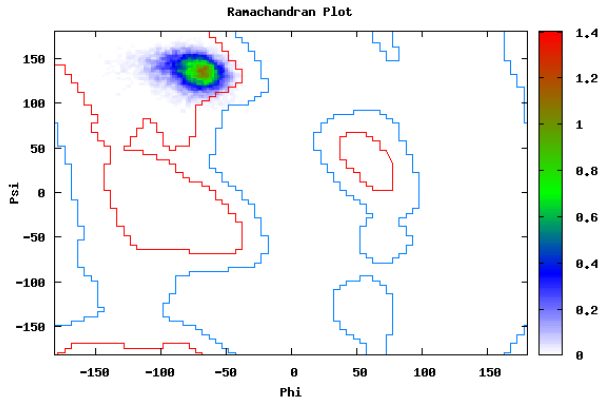


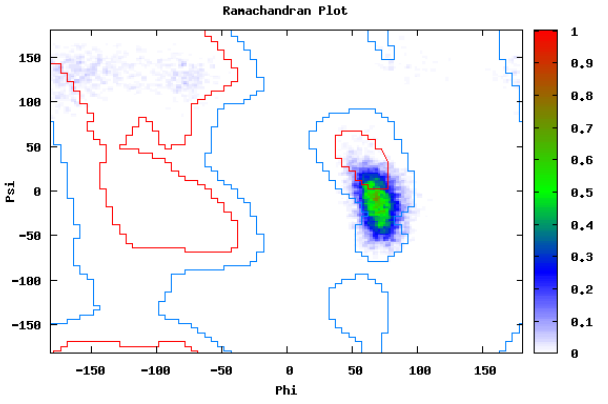


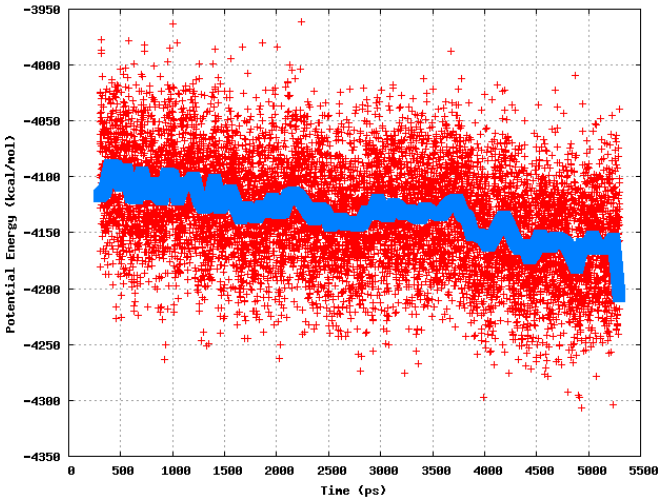


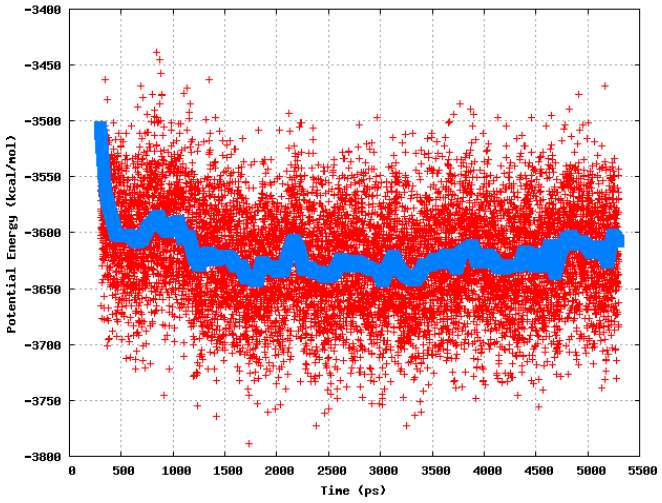
3DK6

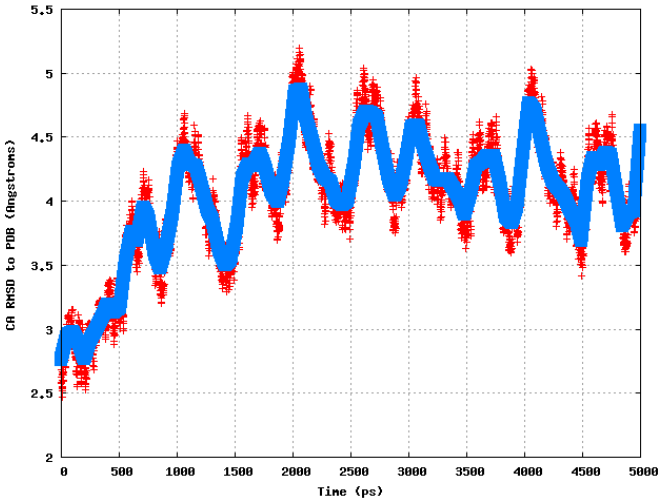
3DK3

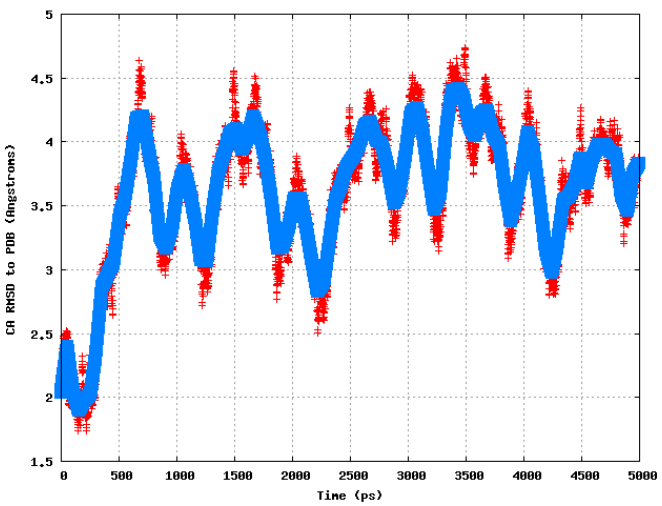

Ramachandran Plot
Psi
Phi


Ramachandran Plot
Psi
Phi

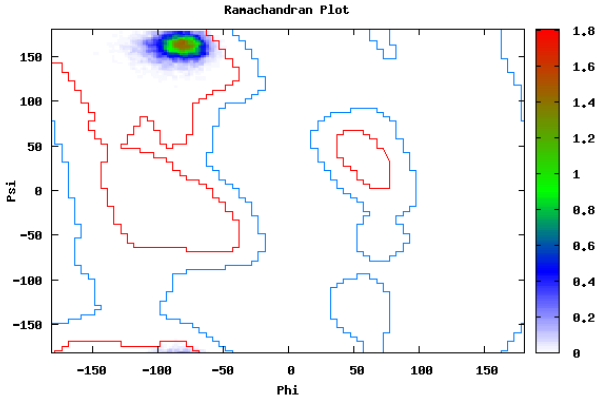

Potential Energy (kcal/mol)
Time (ps)

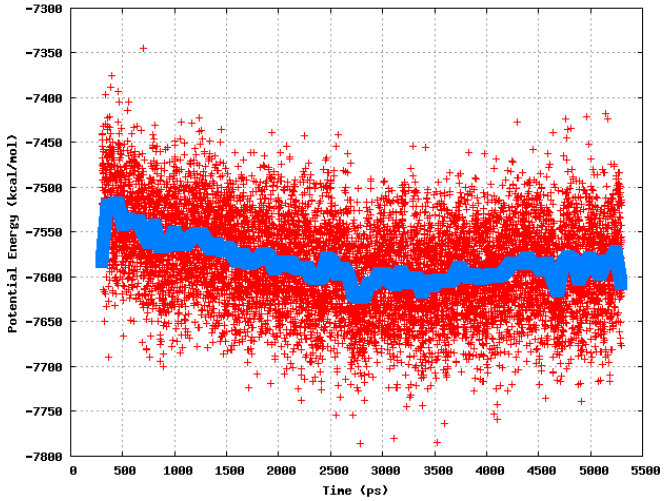

Potential Energy (kcal/mol)
Time (ps)

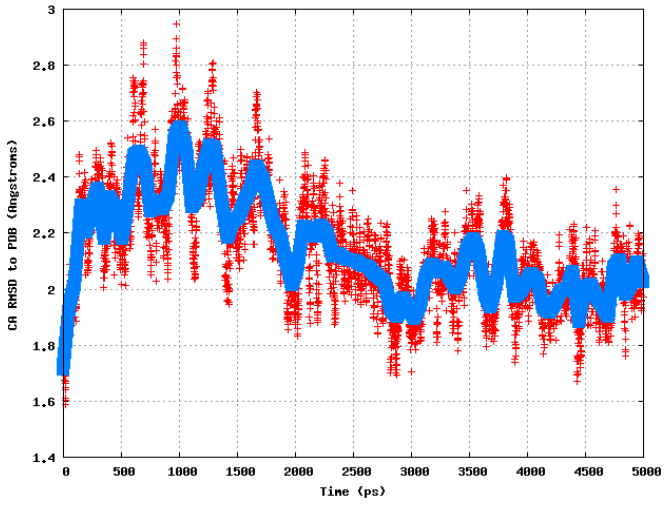

CA RMSD to PDB (Angstroms)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

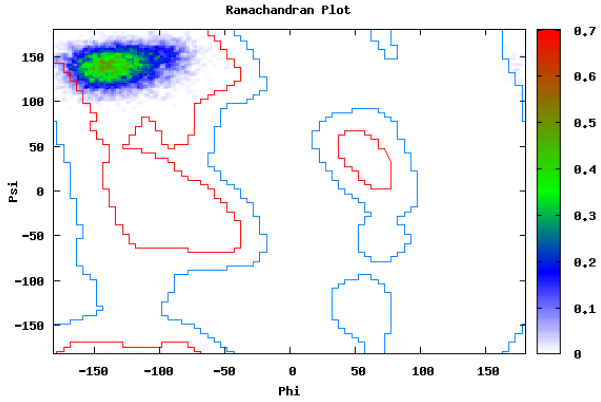
1U54

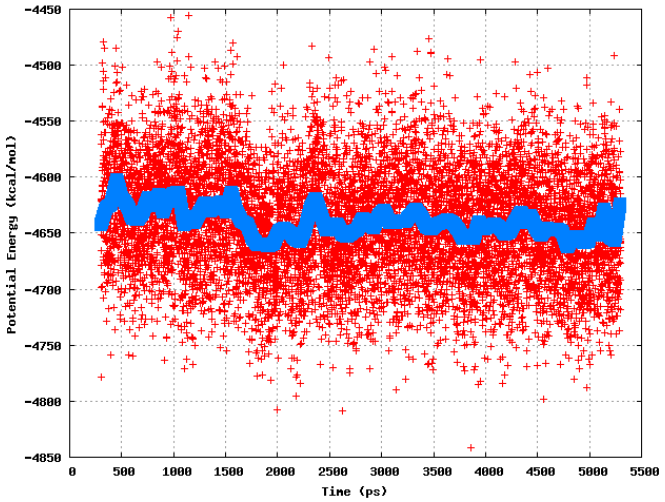

Ramachandran Plot
Psi
Phi

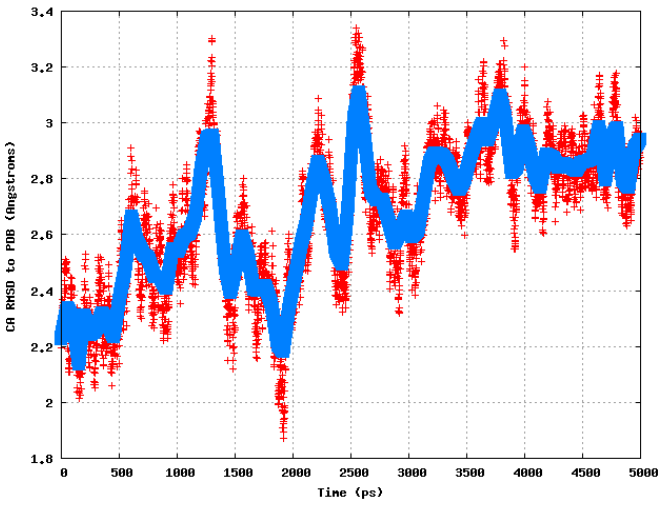

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

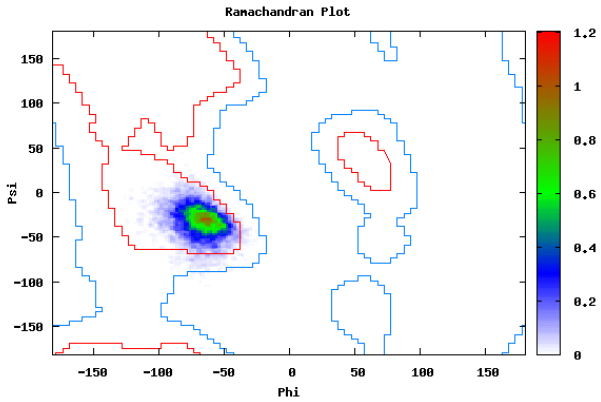
1U46

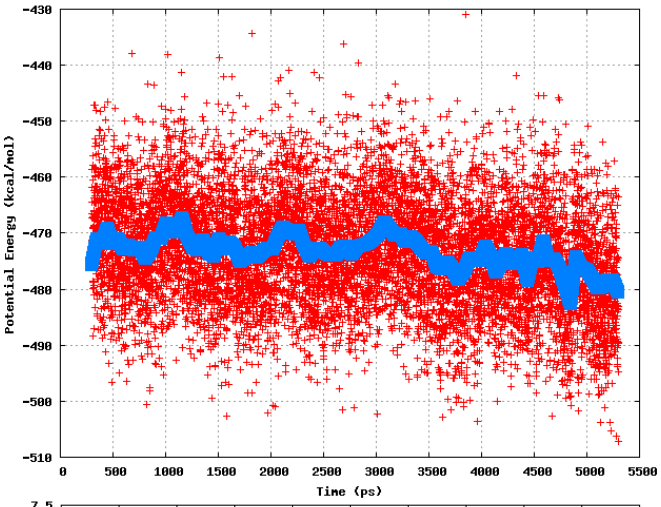

Ramachandran Plot
Psi
Phi

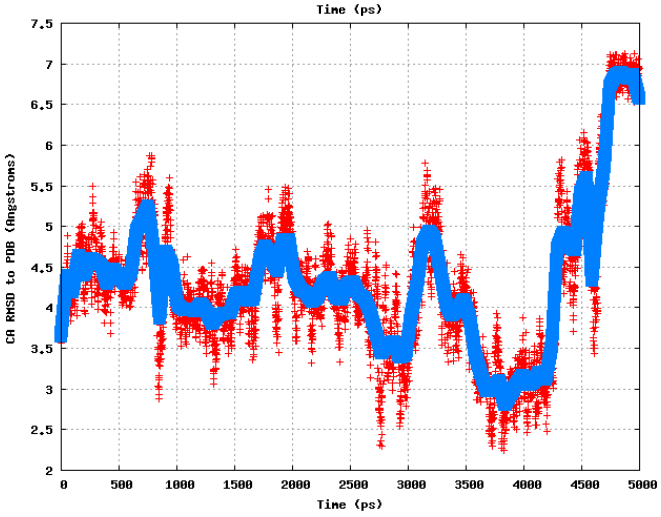

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

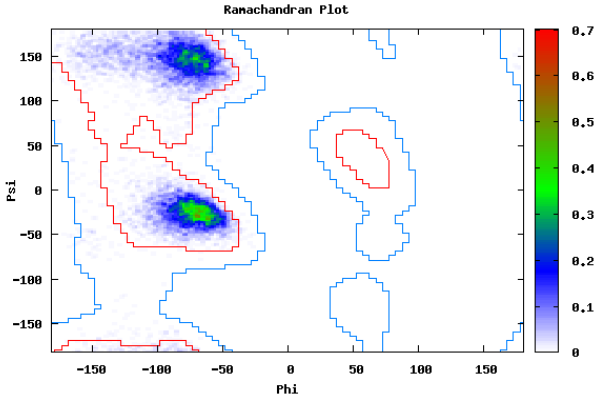
1D4W

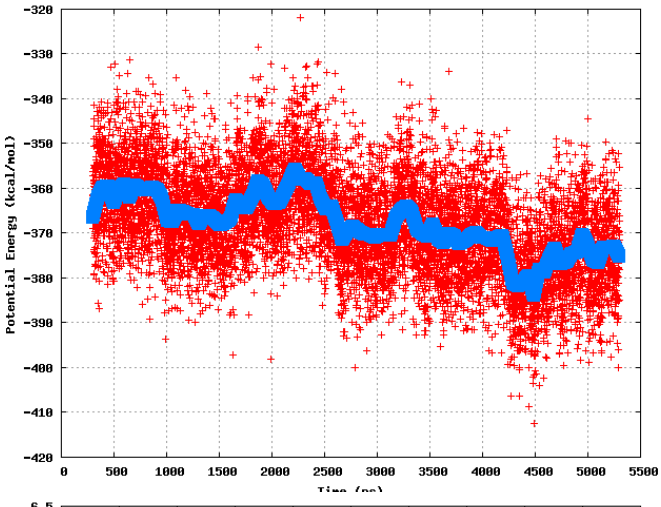

Ramachandran Plot
Psi
Phi

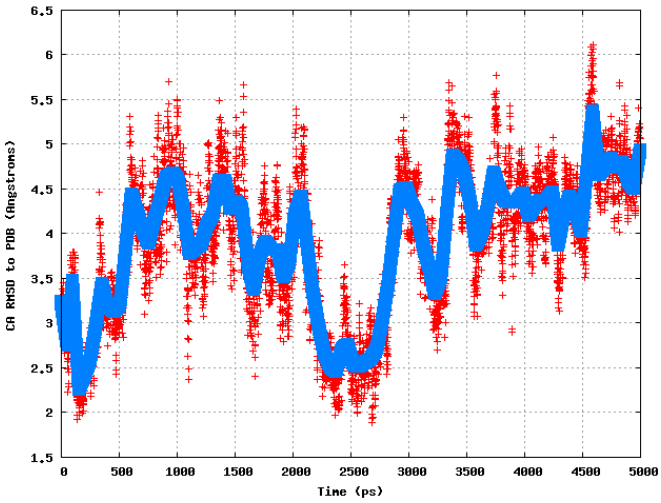

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

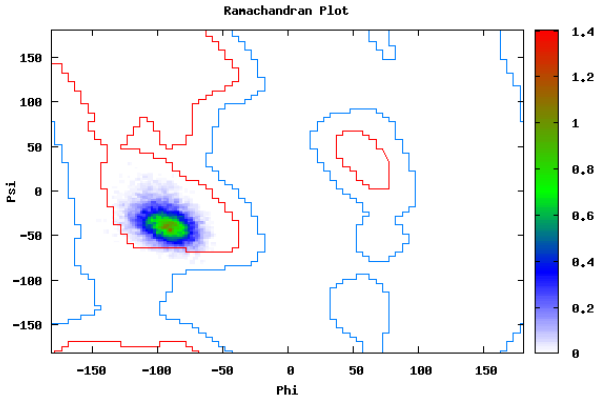
1D4T

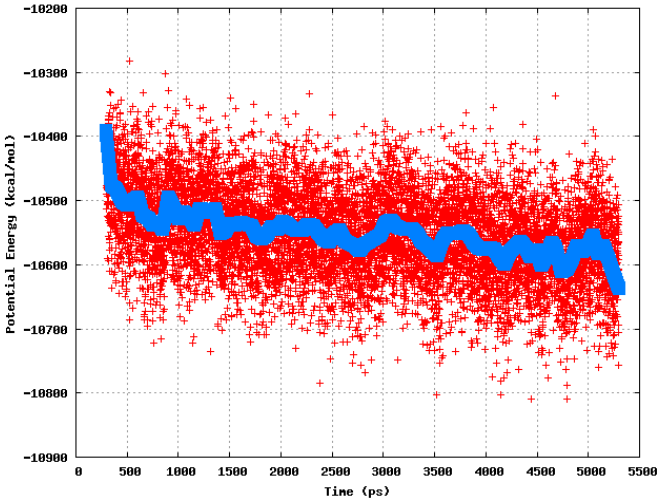

Ramachandran Plot
Psi
Phi

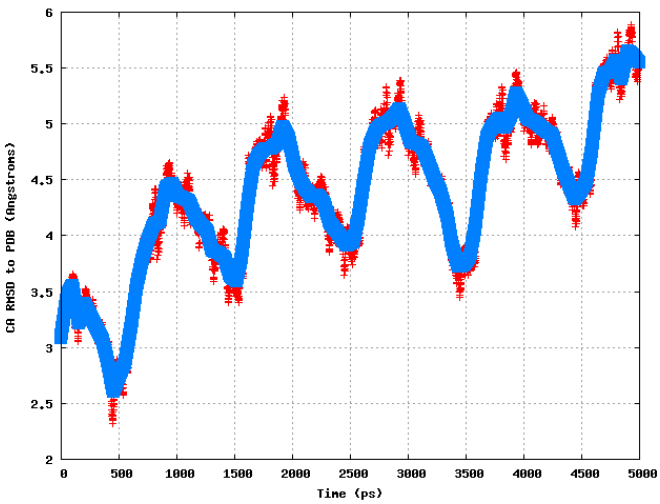

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

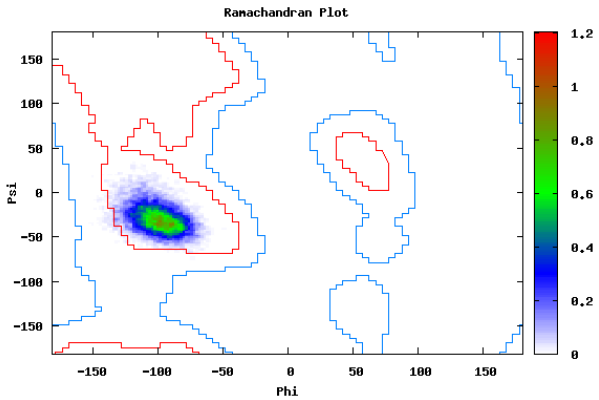
2XKK

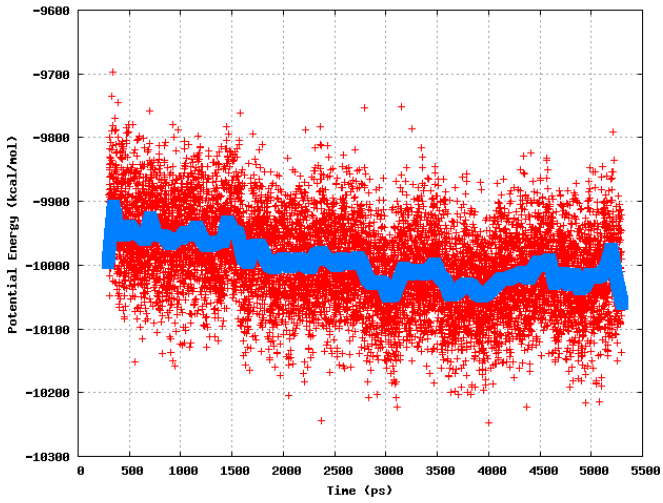

Ramachandran Plot
Psi
Phi

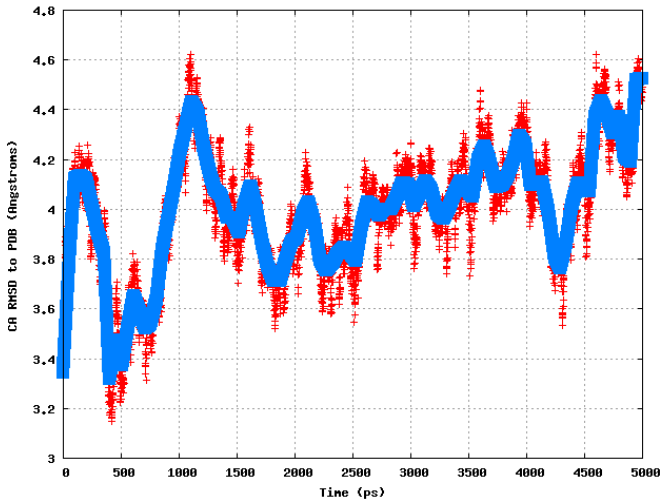

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

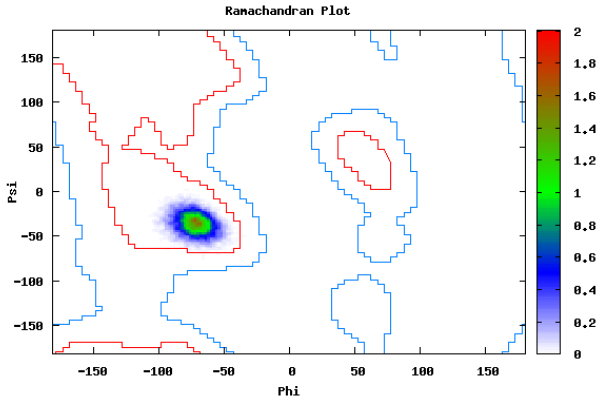
2XKJ

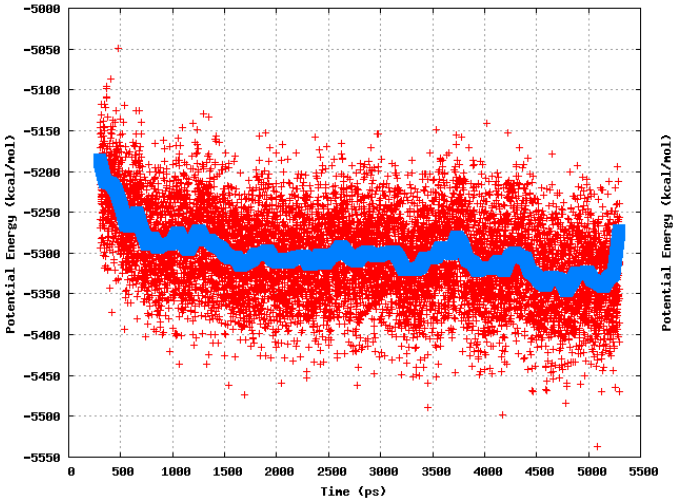

Ramachandran Plot
Psi
Phi

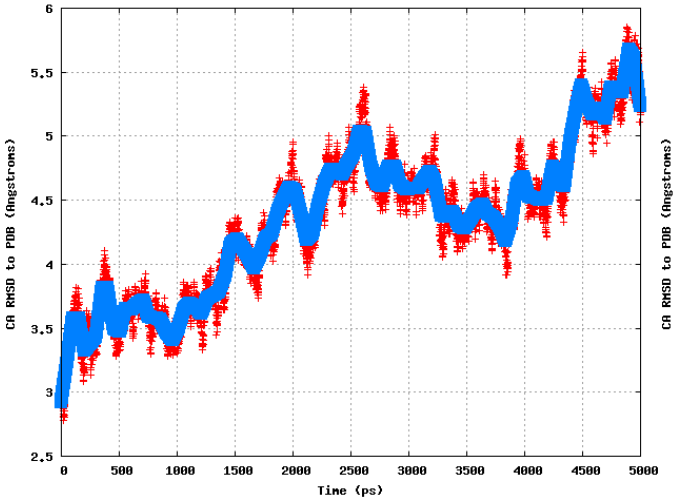

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

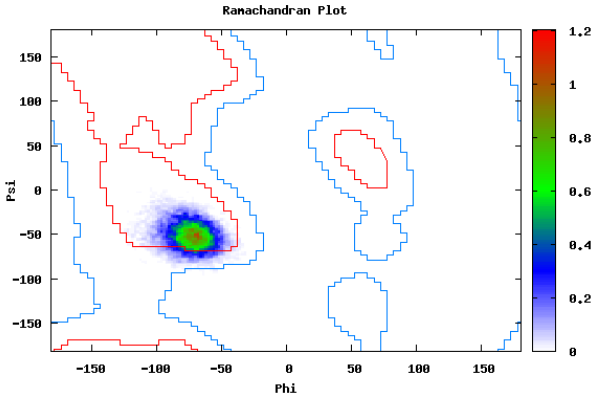
2JFM

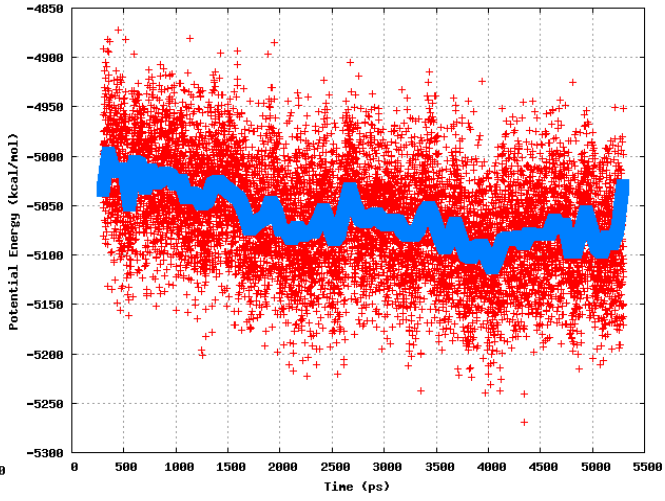

Ramachandran Plot
Psi
Phi

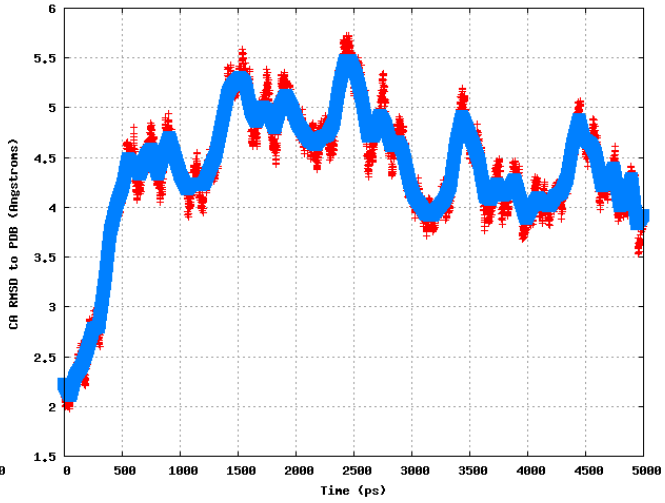

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

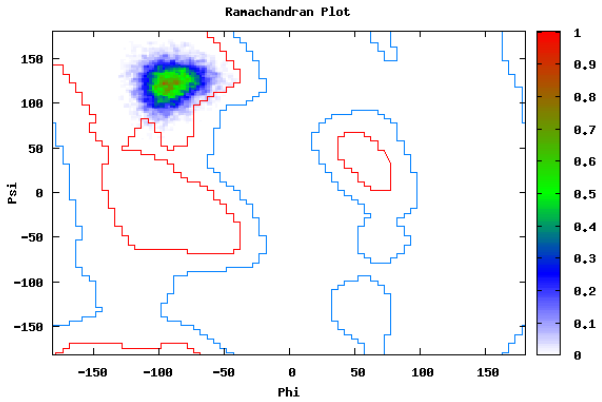
2J51

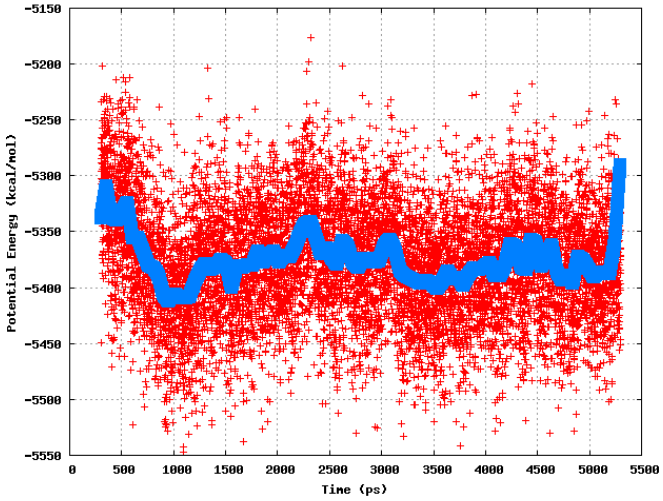

Ramachandran Plot
Psi
Phi

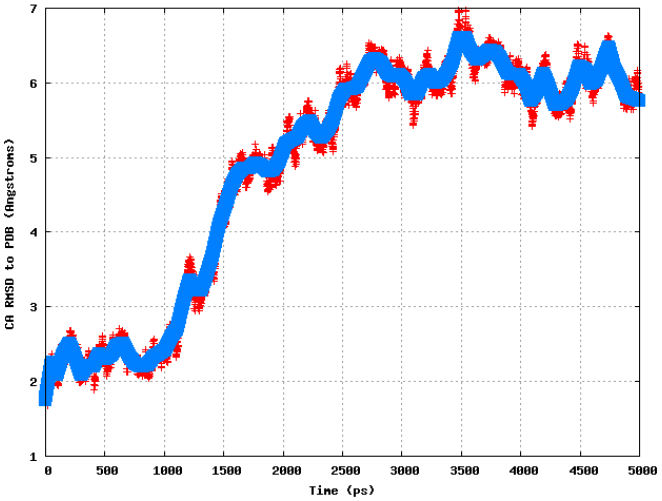

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

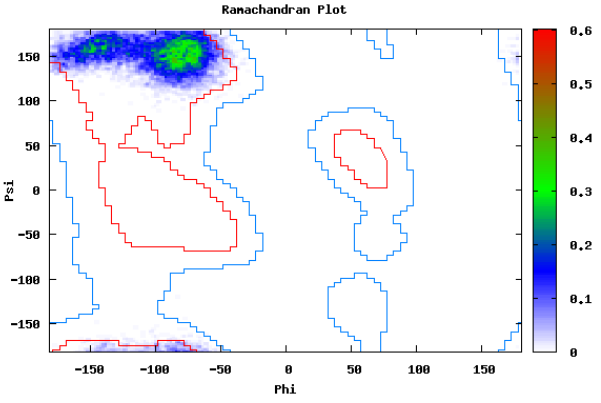
3D5W

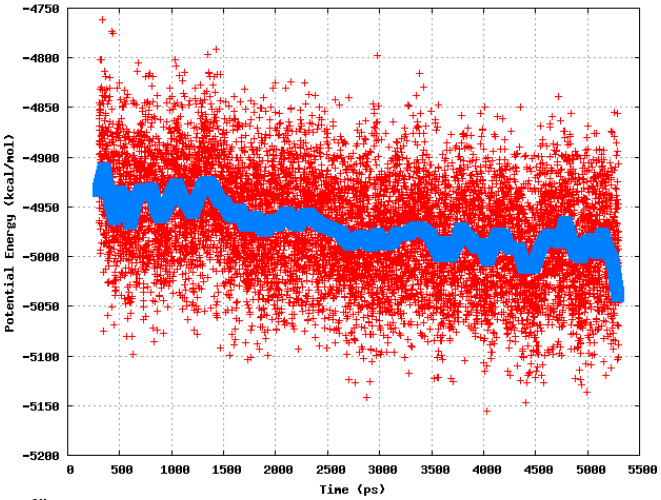

Ramachandran Plot
Psi
Phi

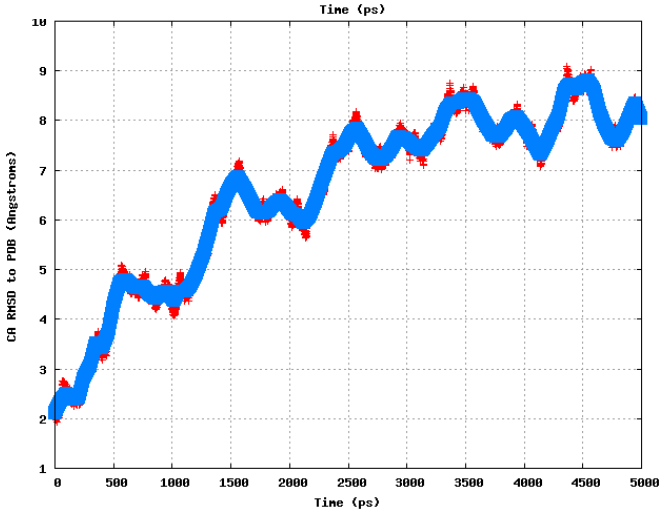

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

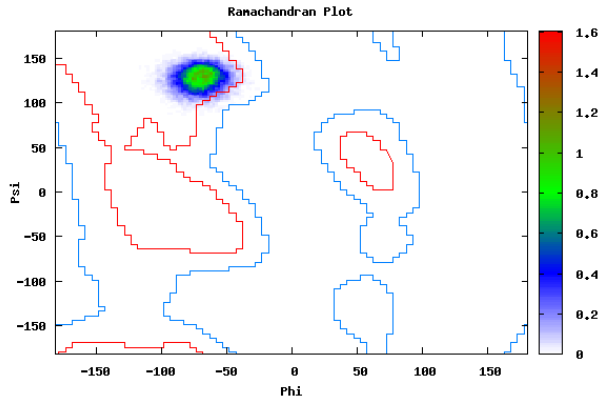
3D5U

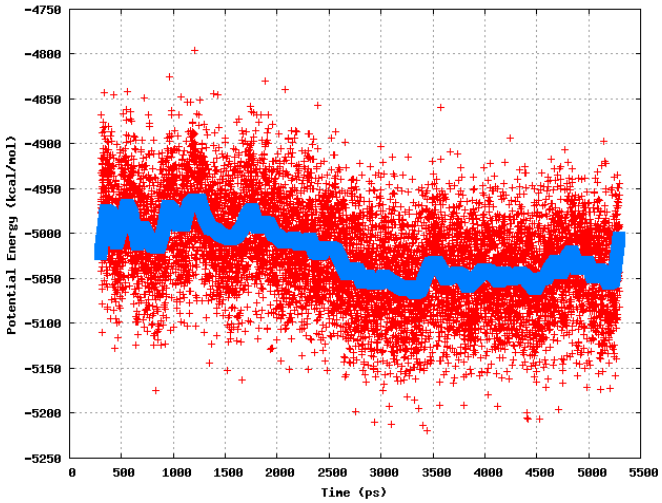

Ramachandran Plot
Psi
Phi

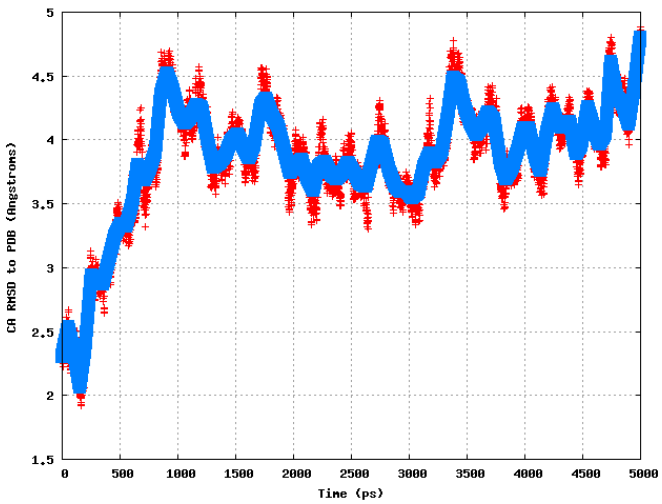

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

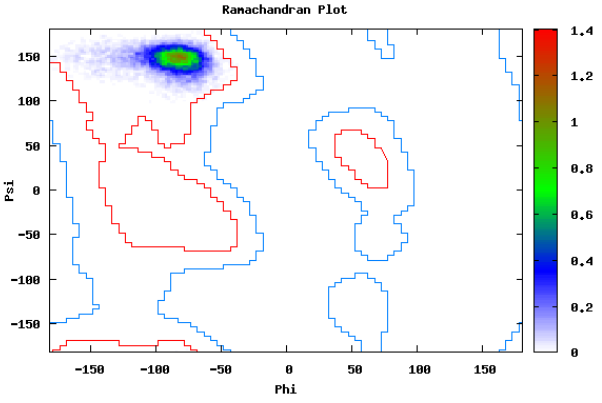
3CKX

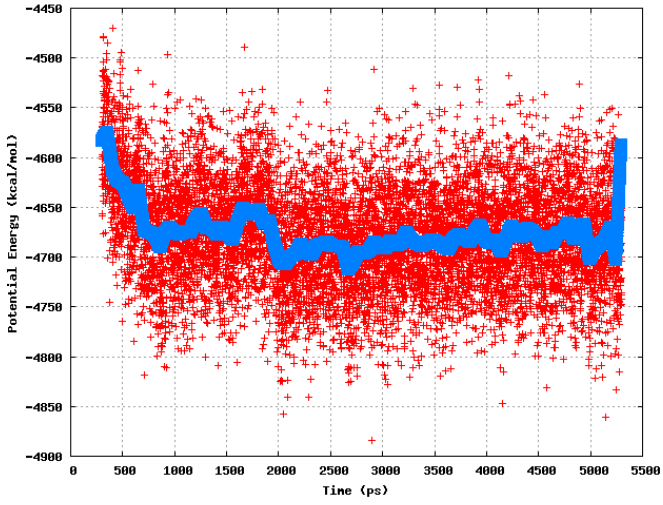

Ramachandran Plot
Psi
Phi

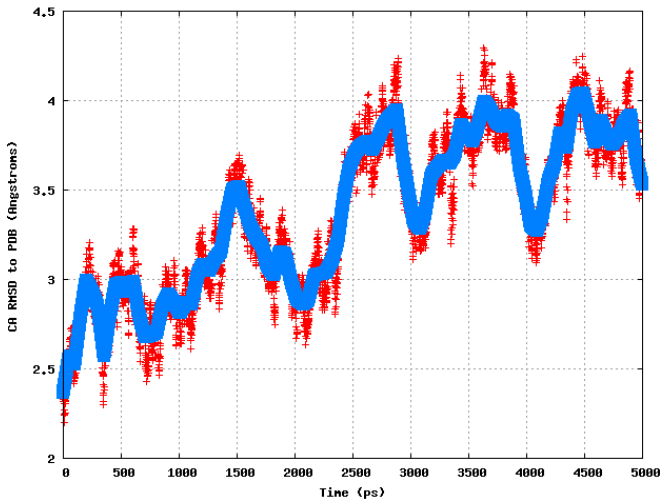

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

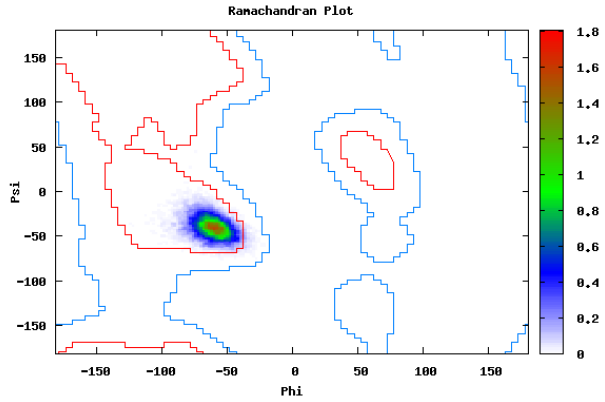
3CKW

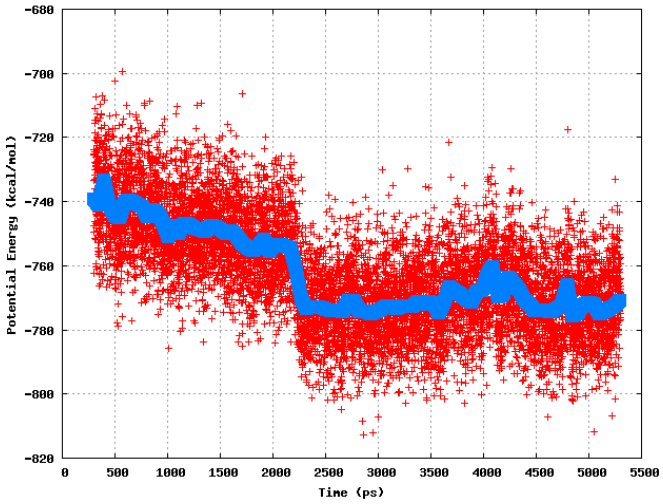

Ramachandran Plot
Psi
Phi

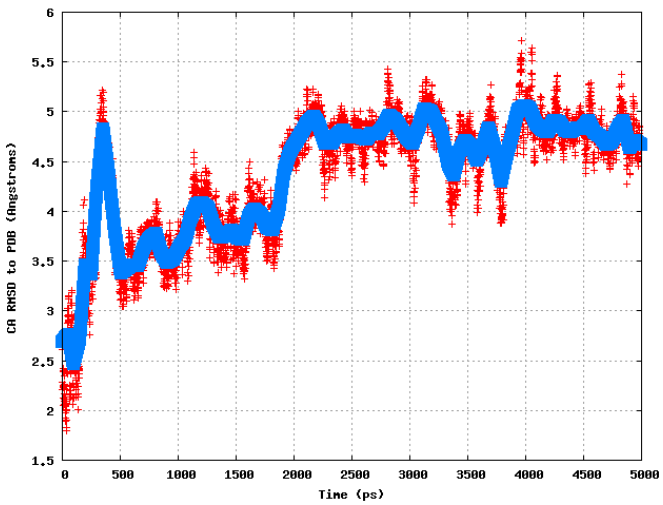

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

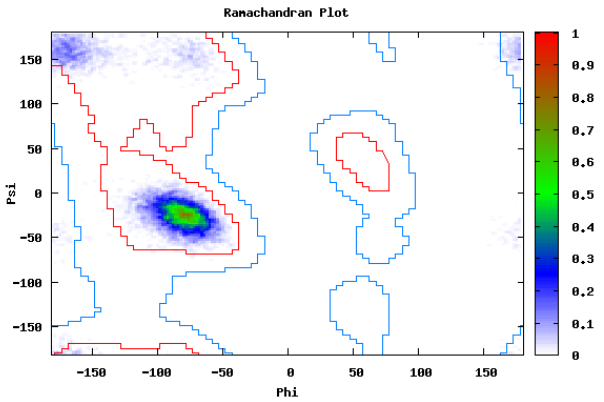
1V50

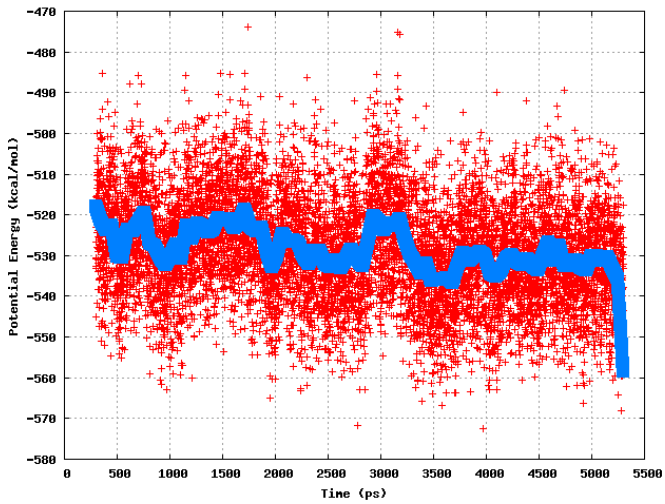

Ramachandran Plot
Psi
Phi

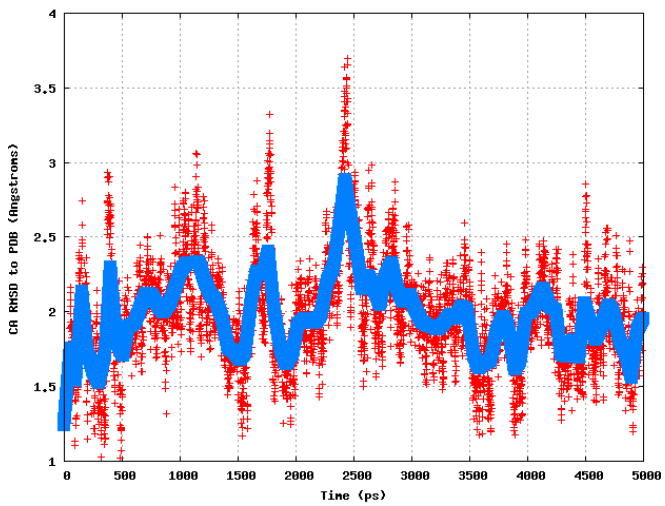

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

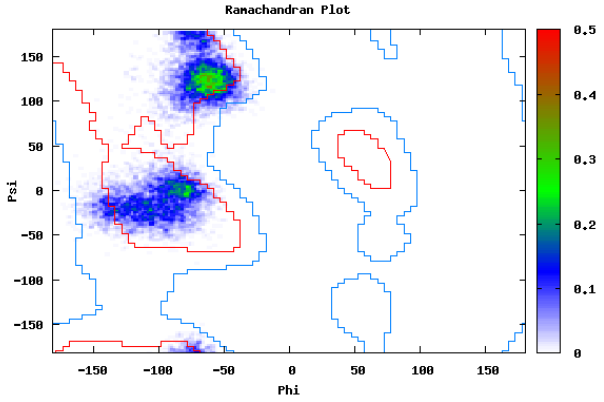
1V4Z

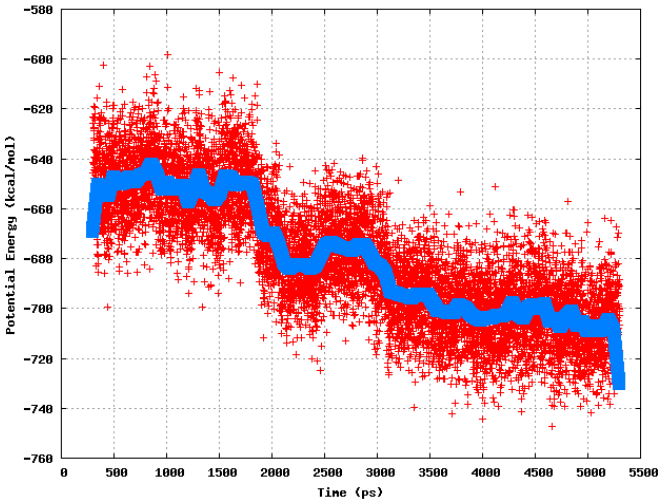

Ramachandran Plot
Psi
Phi

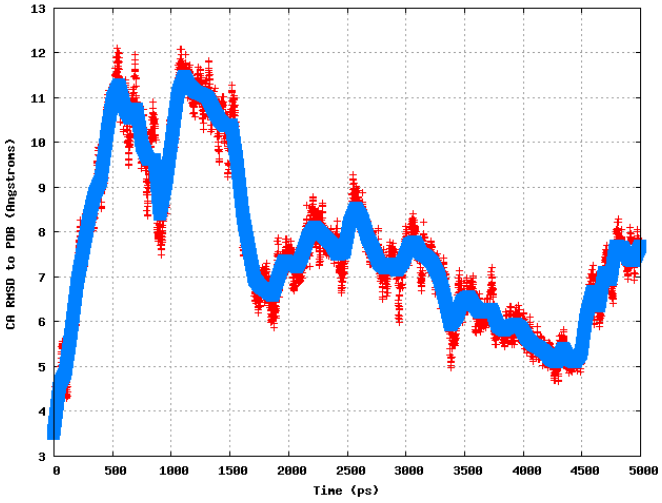

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

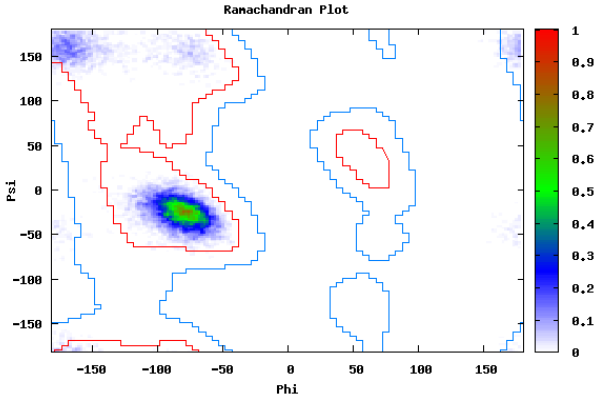
2L5J

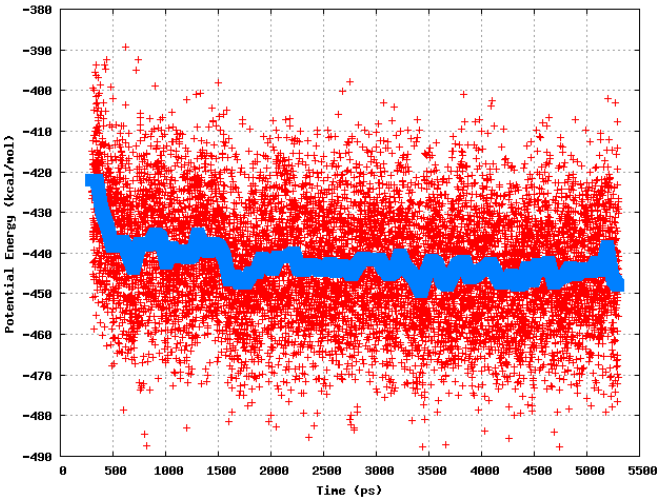

Ramachandran Plot
Psi
Phi

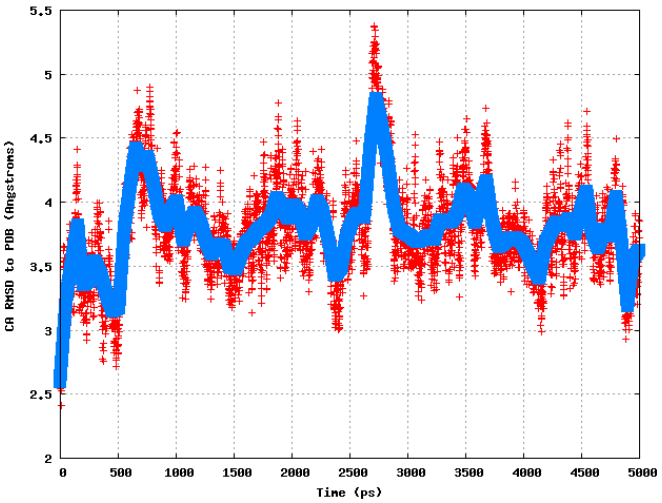

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

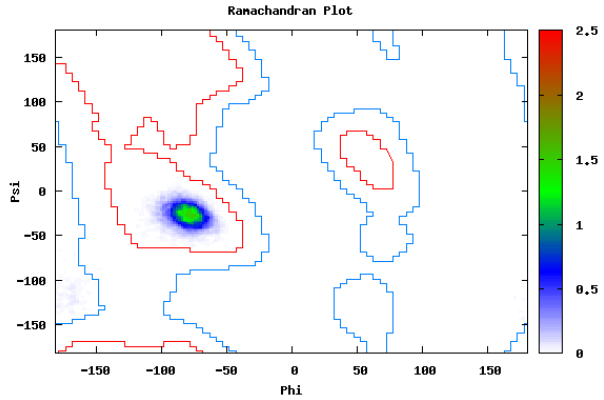
2L5I

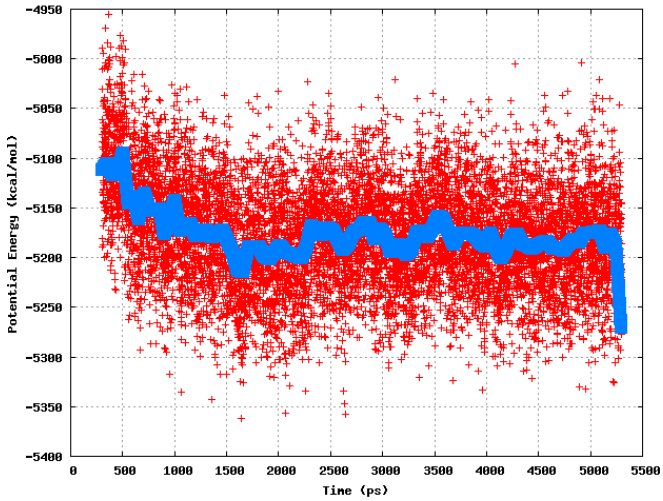

Ramachandran Plot
Psi
Phi

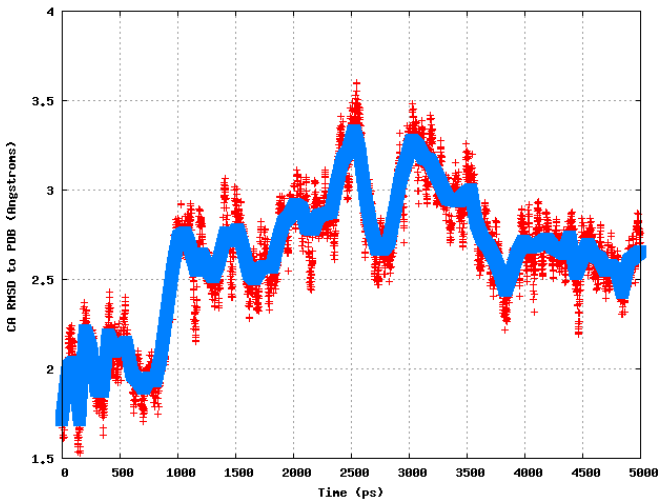

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

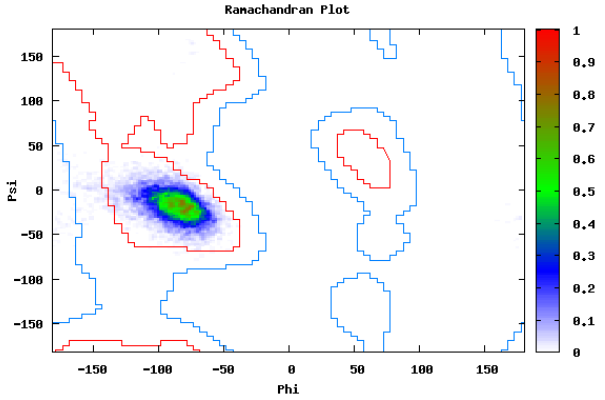
3NAY

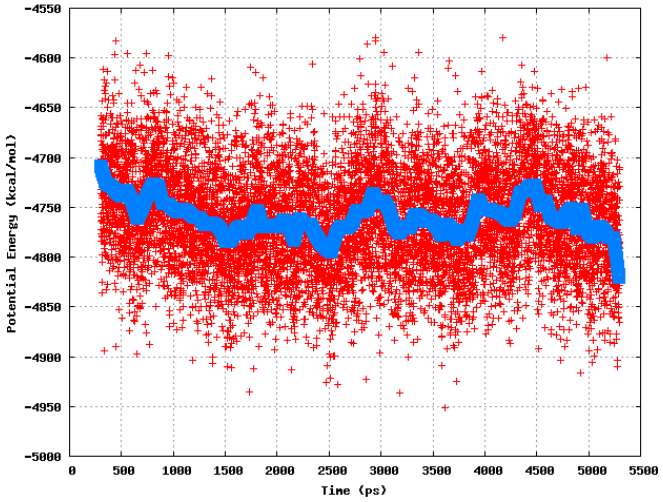

Ramachandran Plot
Psi
Phi

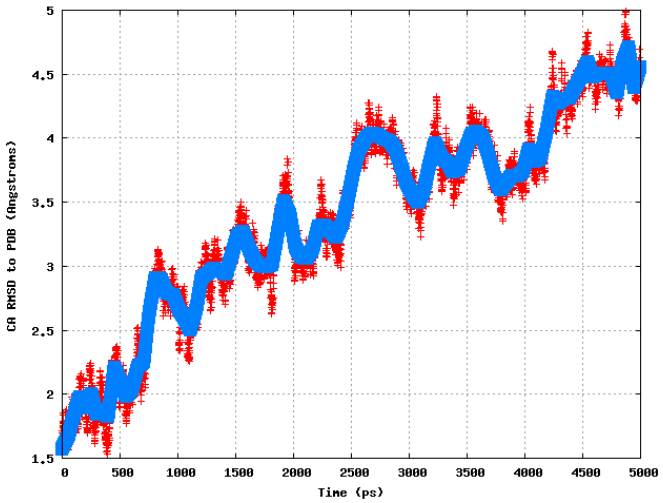

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstrons)
Time (ps)

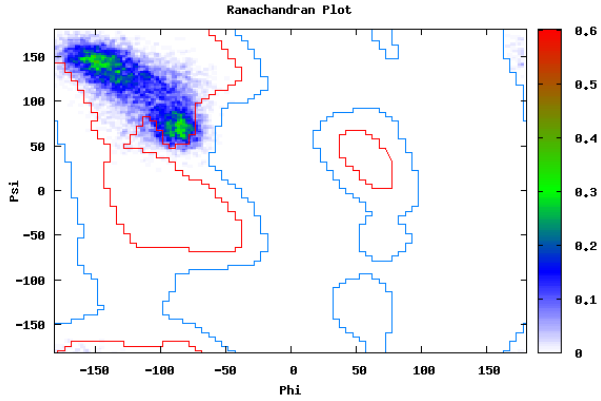
3NAX

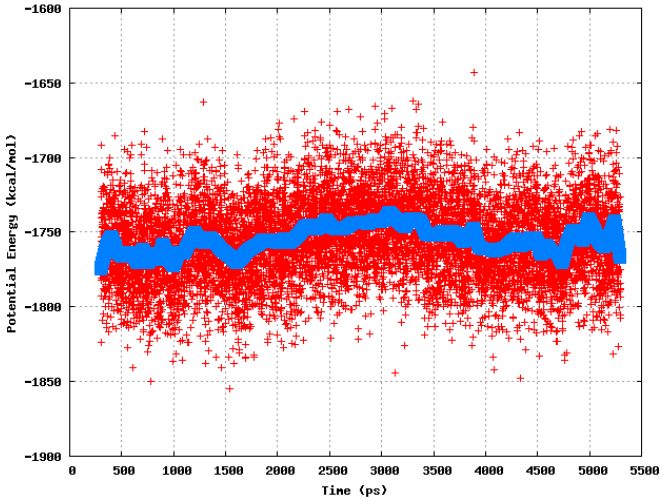

Ramachandran Plot
Psi
Phi

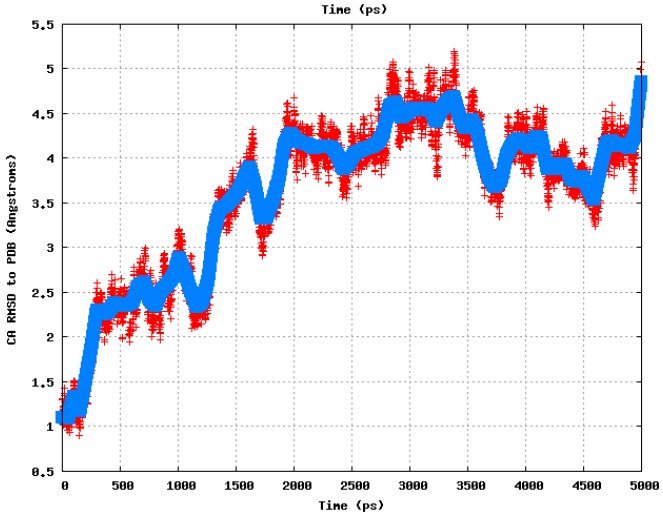

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstrons)
Time (ps)

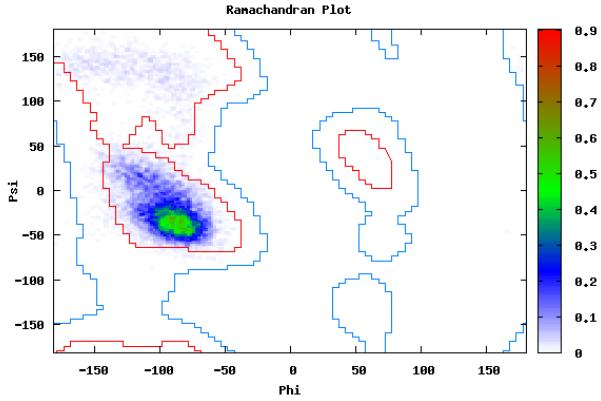
1KKM

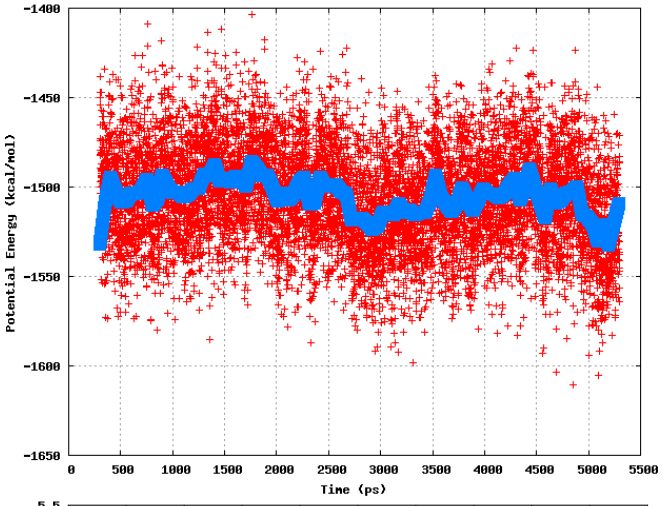

Ramachandran Plot
Psi
Phi

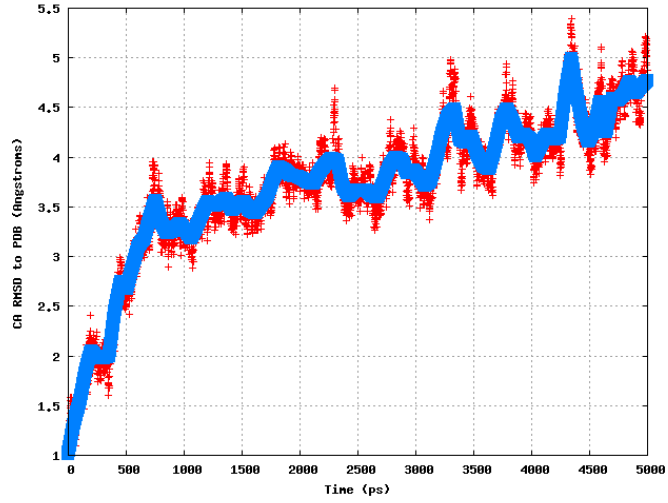

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

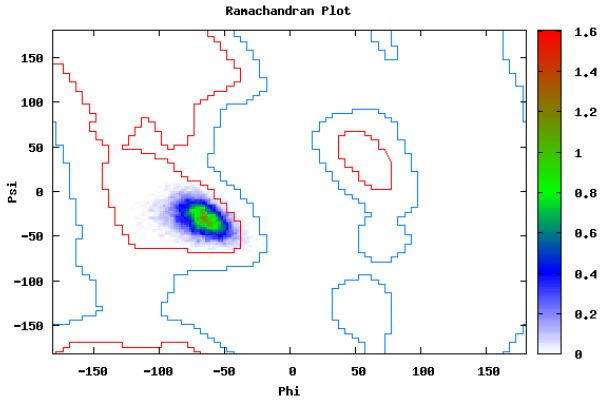
1KKL

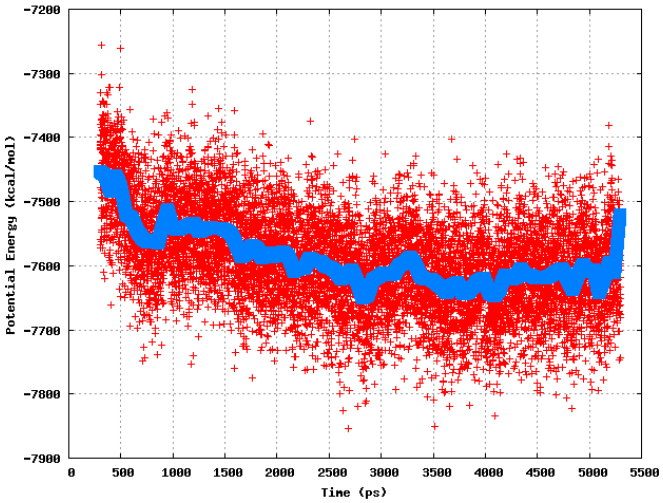

Ramachandran Plot
Psi
Phi

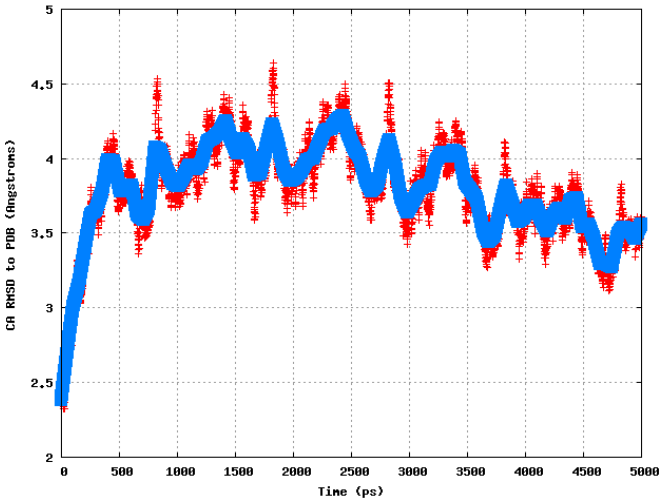

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

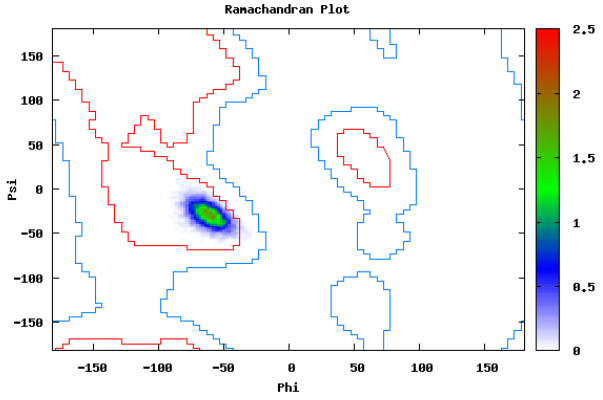
3IAF

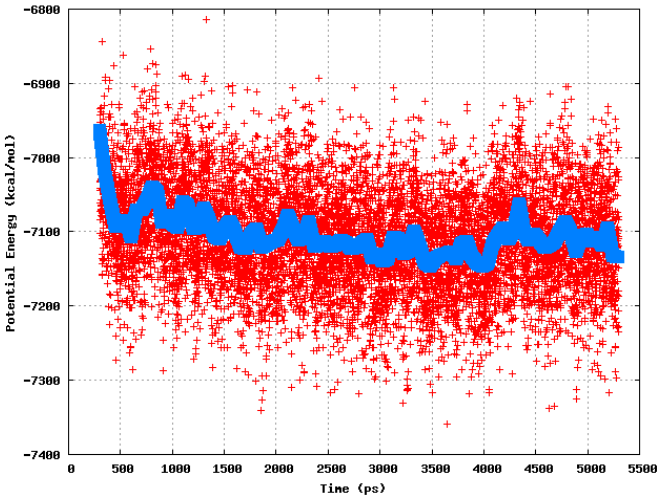

Ramachandran Plot
Psi
Phi

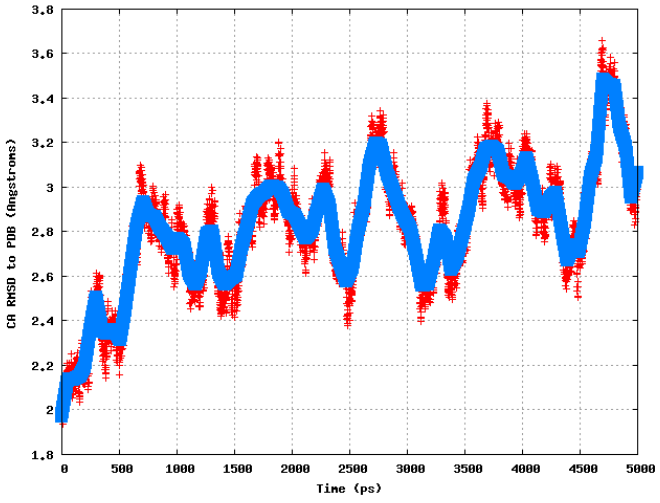

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

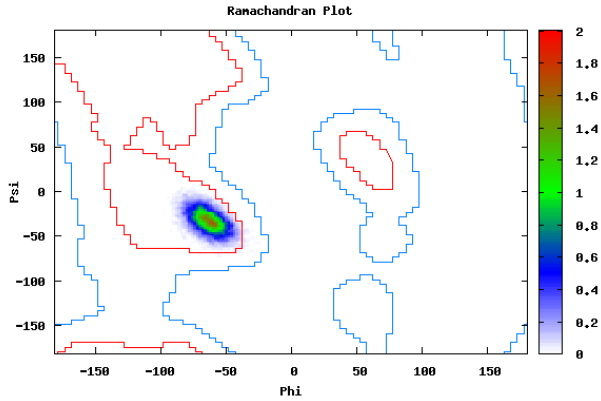
3IAE

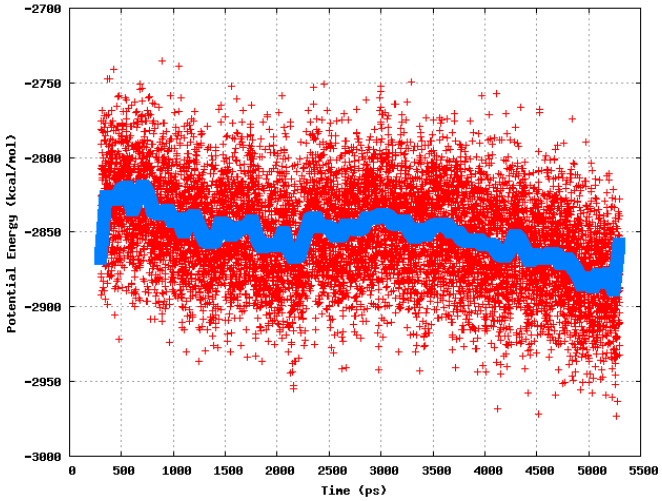

Ramachandran Plot
Psi
Phi

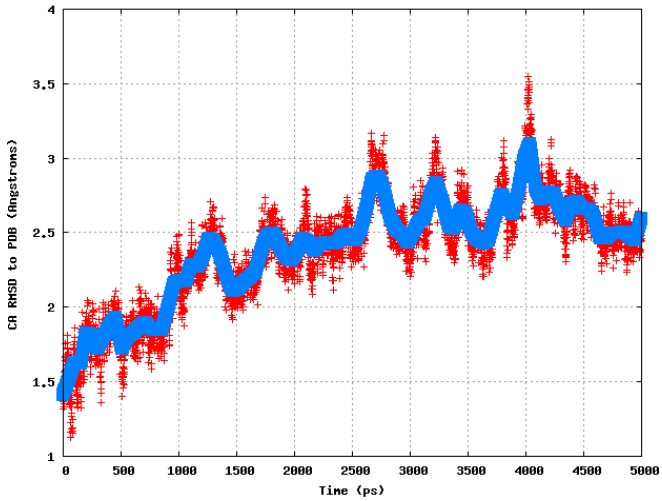

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

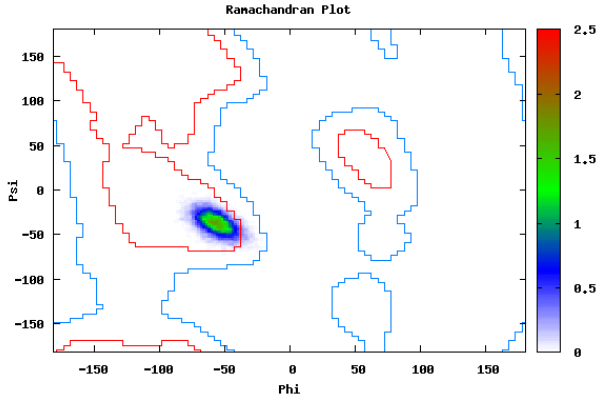
1H4X

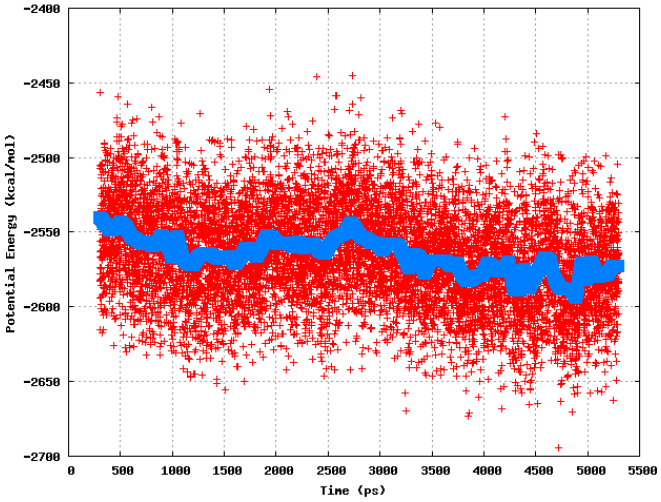

Ramachandran Plot
Psi
Phi

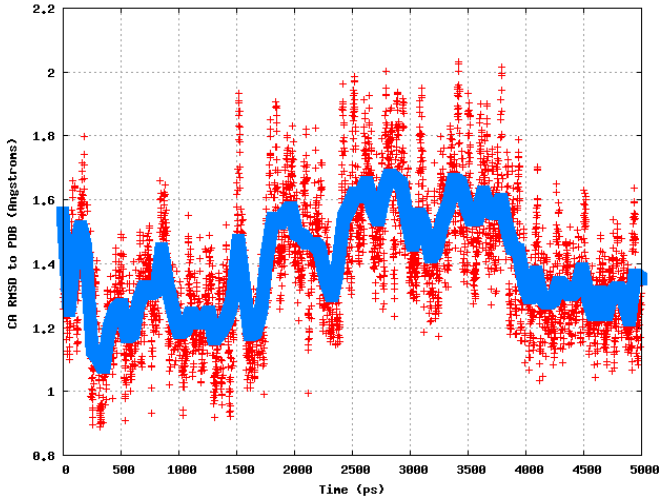

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

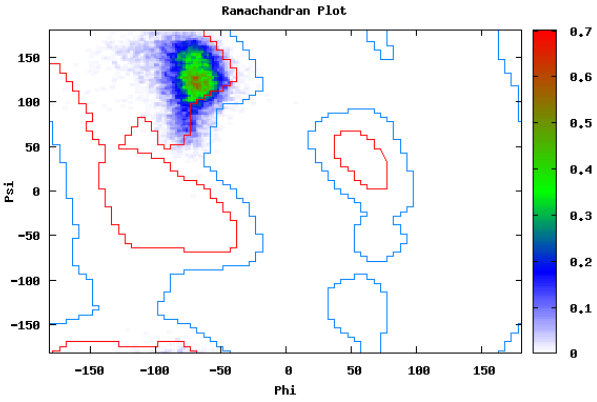
1H4Y

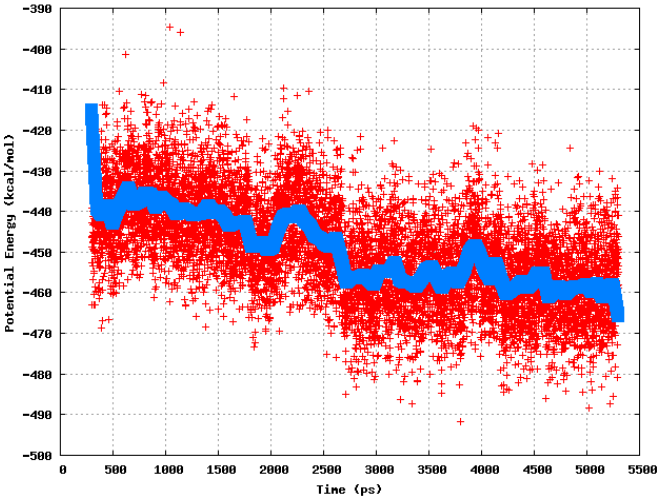

Ramachandran Plot
Psi
Phi

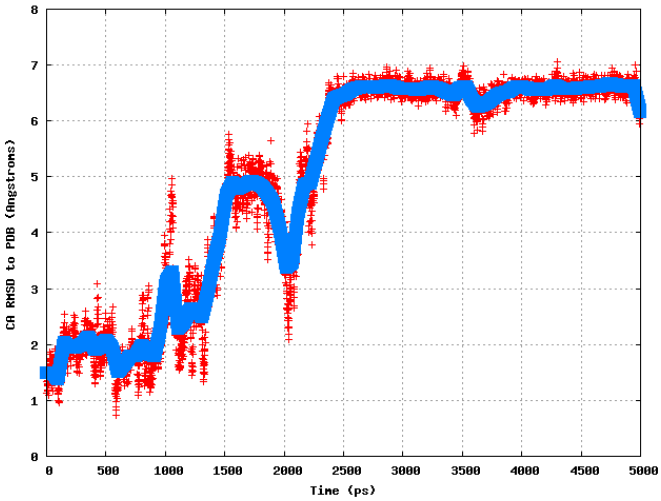

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

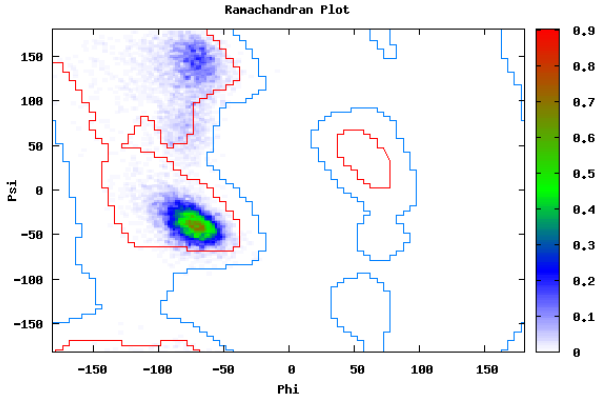
3F9X

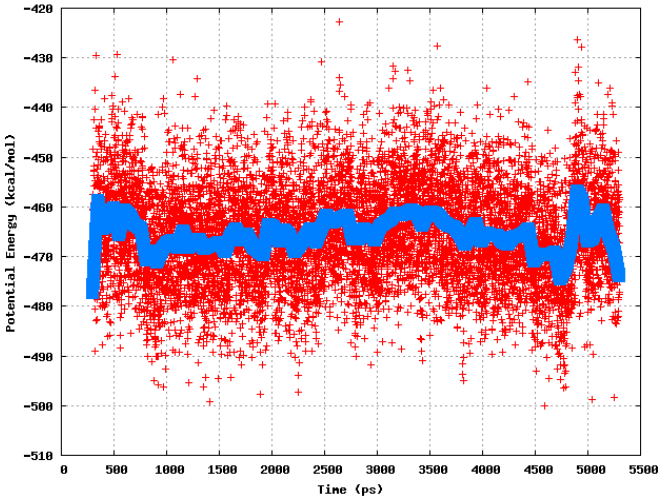

Ramachandran Plot
Psi
Phi

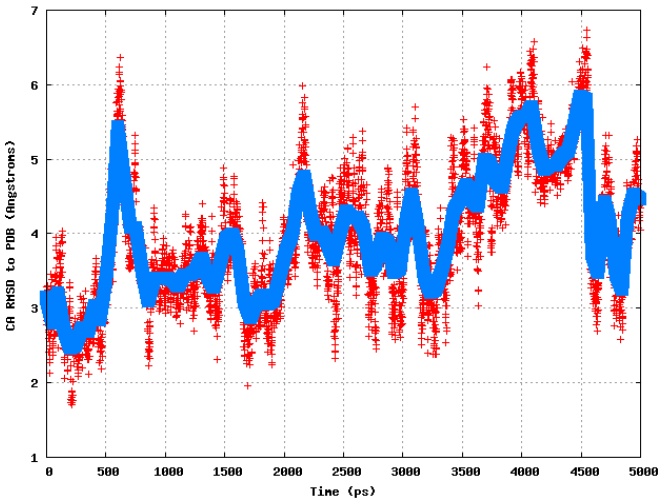

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

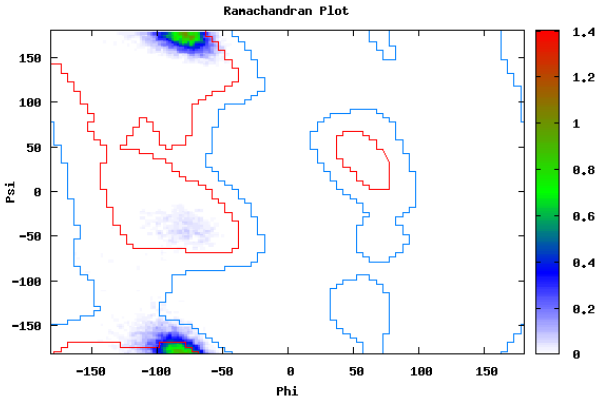
3F9W

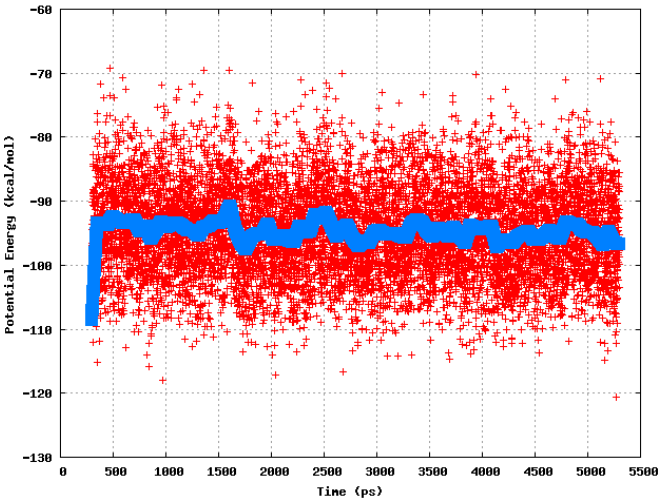

Ramachandran Plot
Psi
Phi

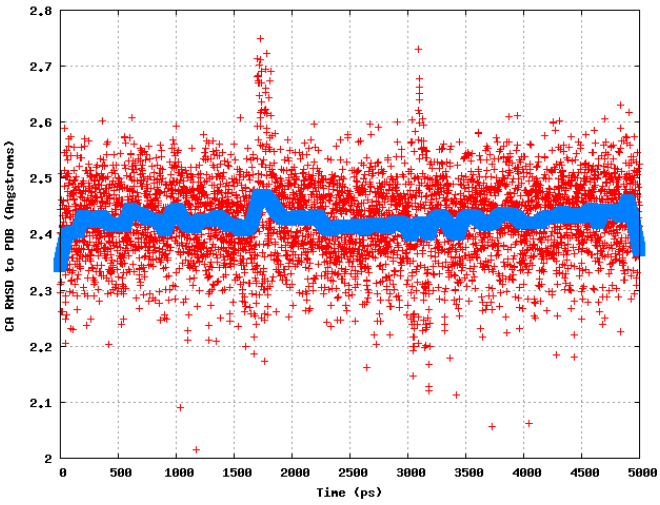

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

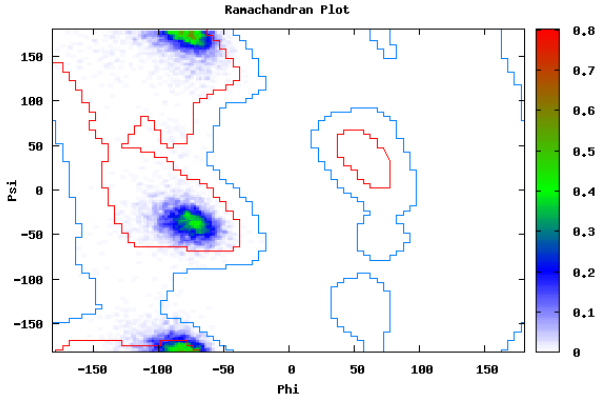
3S7D

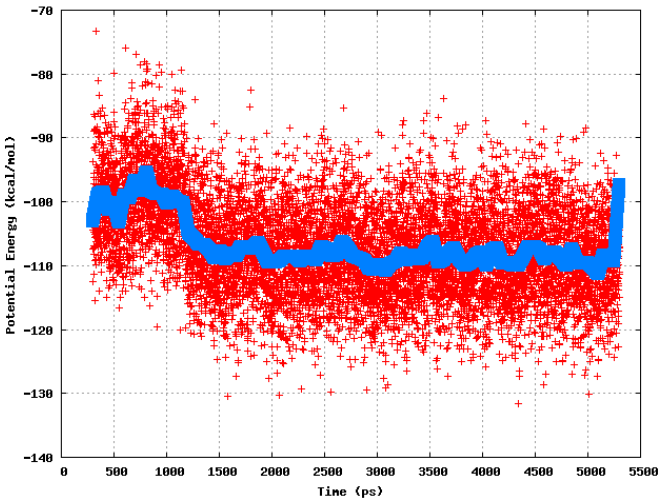

Ramachandran Plot
Psi
Phi

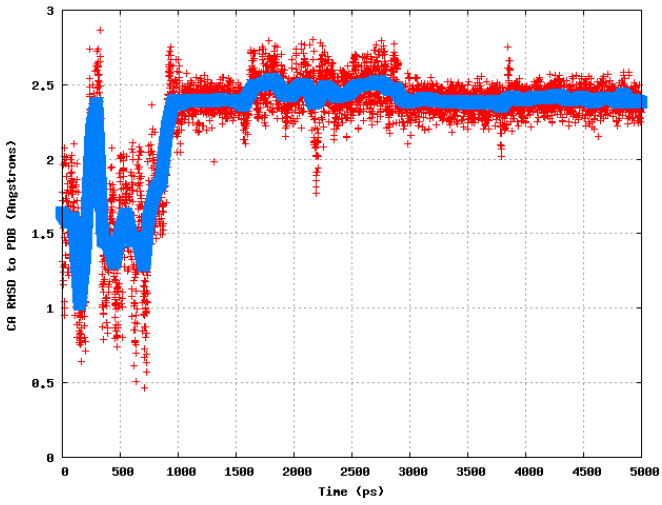

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

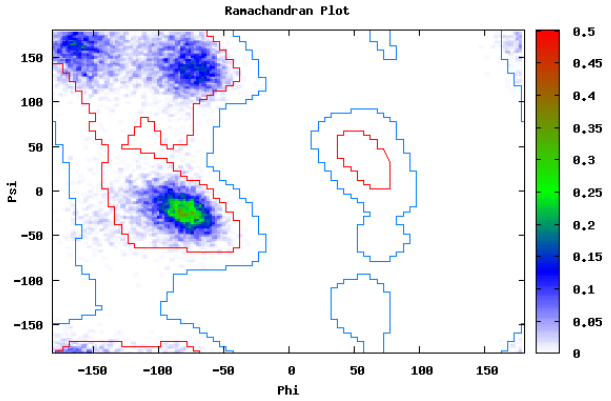
3S7F

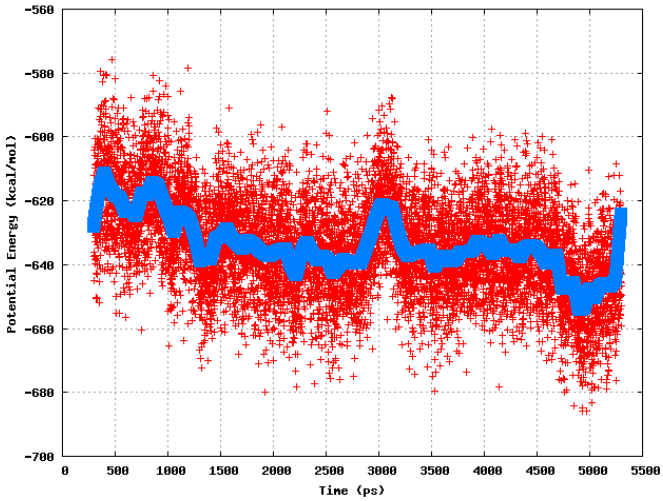

Ramachandran Plot
Psi
Phi

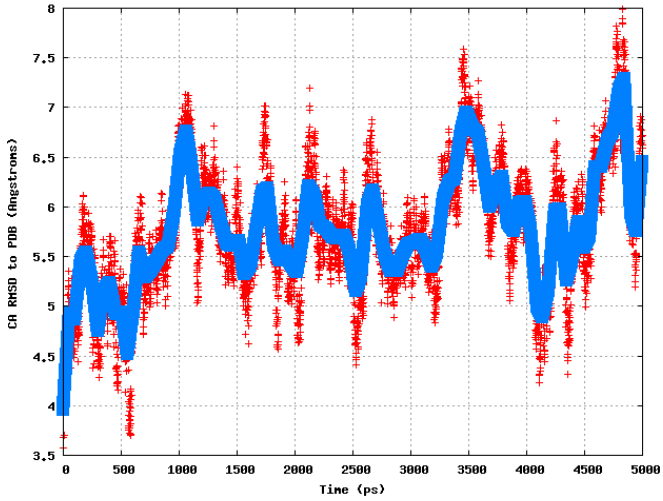

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

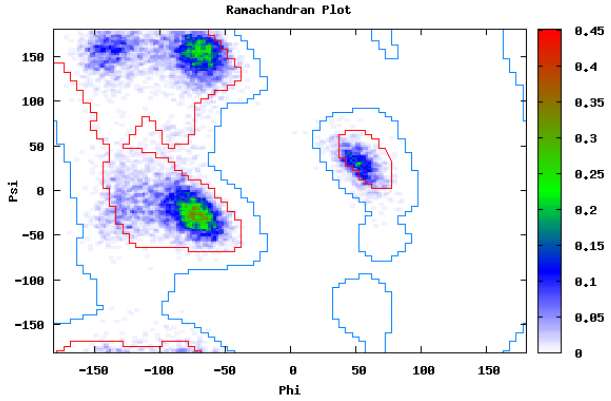
2KWJ

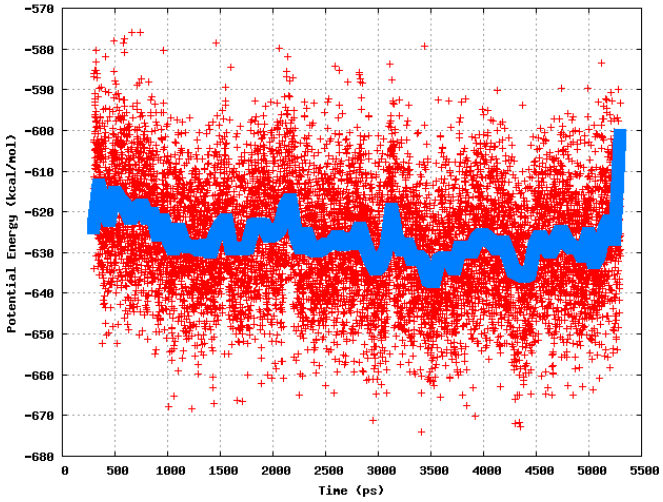

Ramachandran Plot
Psi
Phi

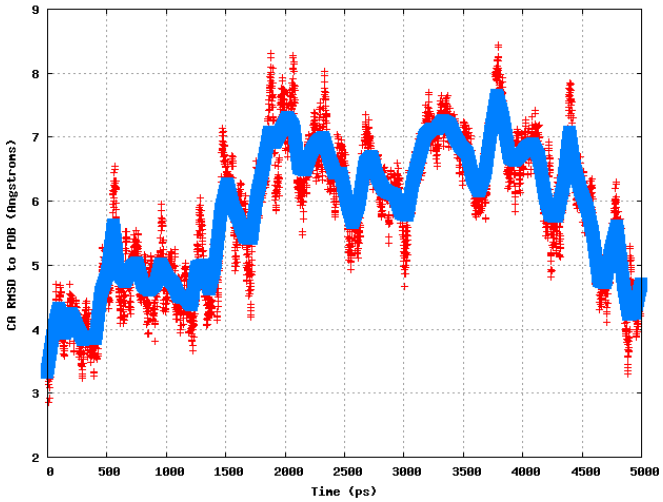

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

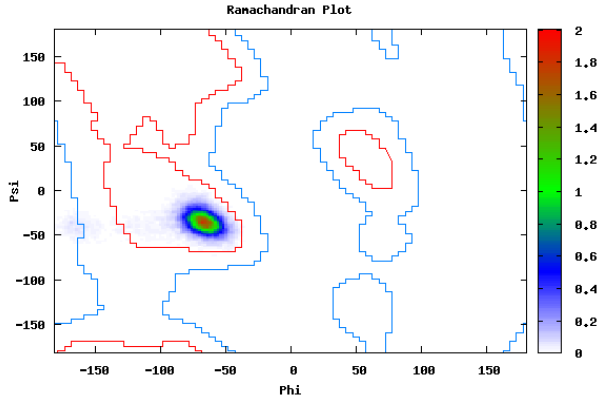
2KWK

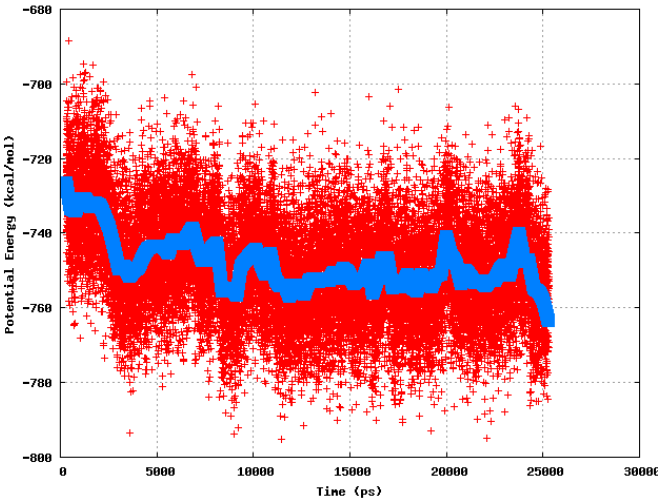

Ramachandran Plot
Psi
Phi

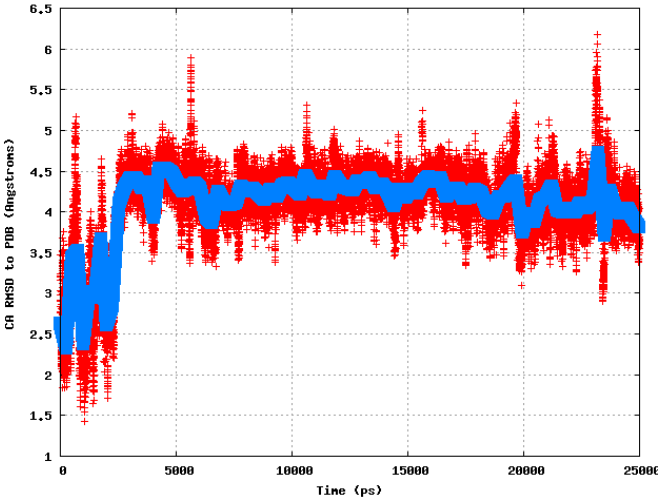

Potential Energy (kcal/mol)
Time (ps)


CA RMSD to PDB (Angstroms)
Time (ps)

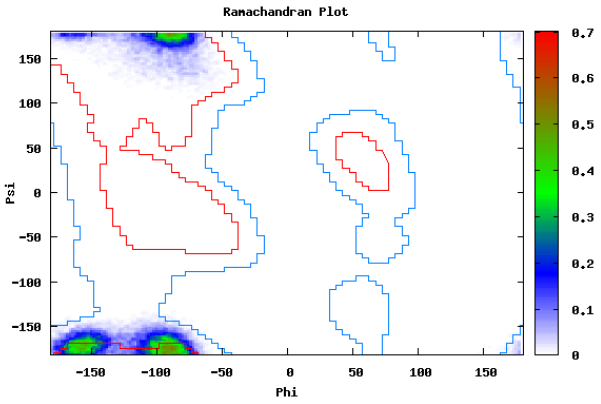
Independent 25ns simulation
1V50

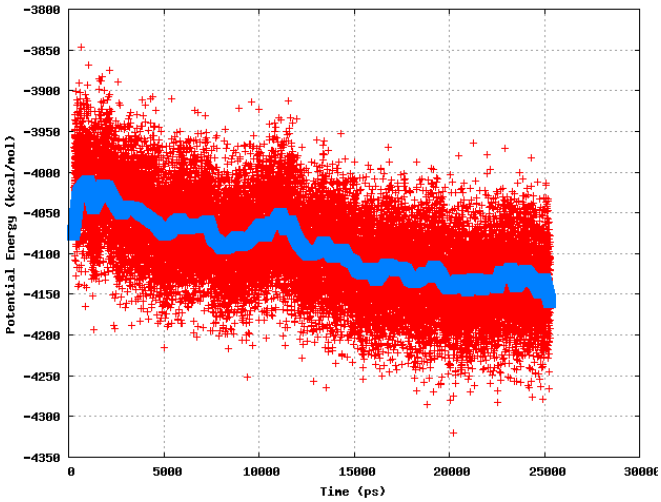


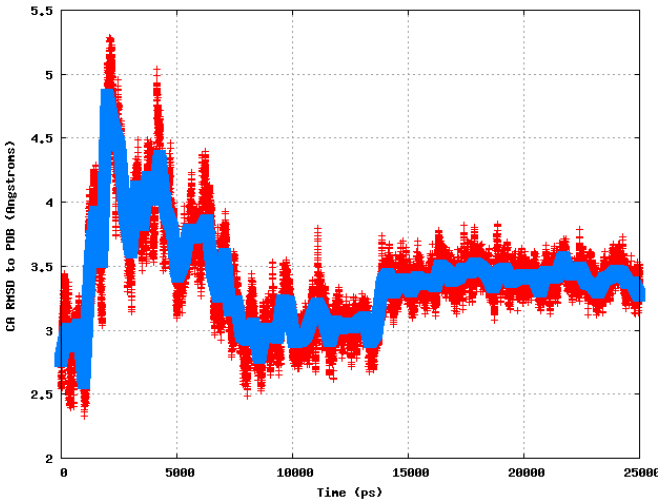




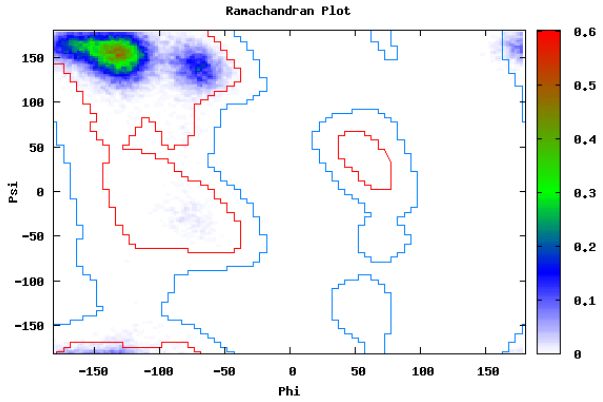
Independent 25 ns simulation
3DK6

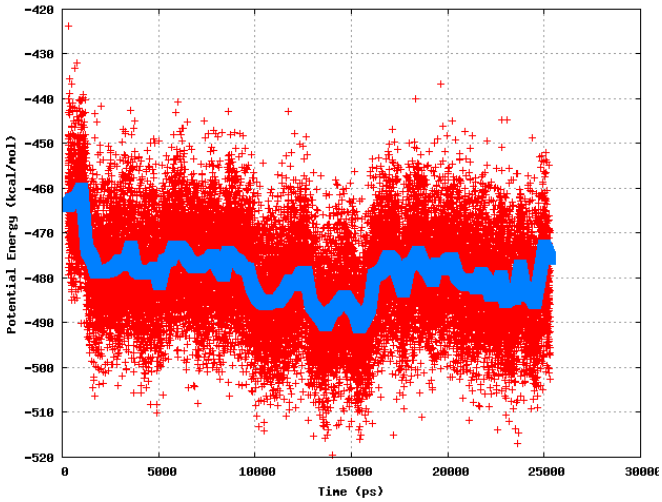


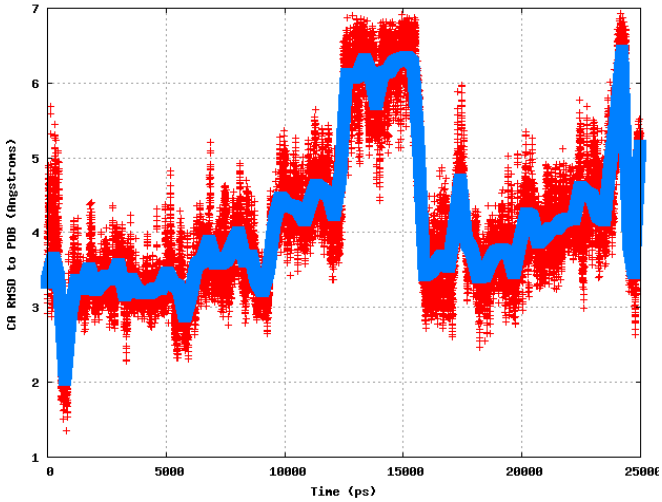




Independent 25 ns simulation
1D4W



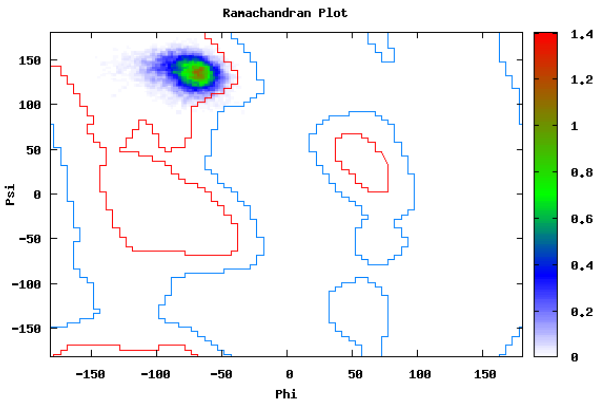


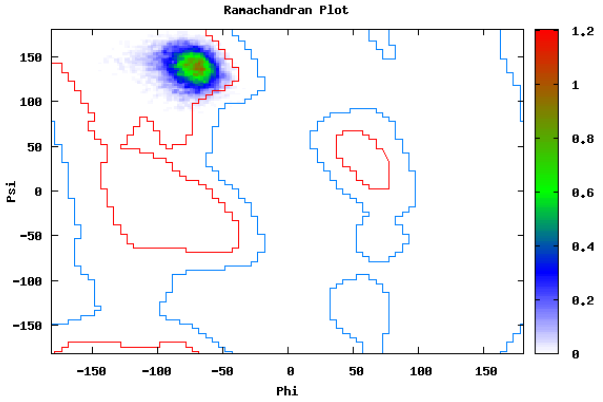


We present additional simulations of comparisons of the Unmodified Structure (U-PDB) Modified and Simulated, compared to the simulation of the Modified Structure and to the native modified structure (M-PDB). These are in addition to the 3S7F/3S7D example provided in the main text, denoted as Comparisons D and E.

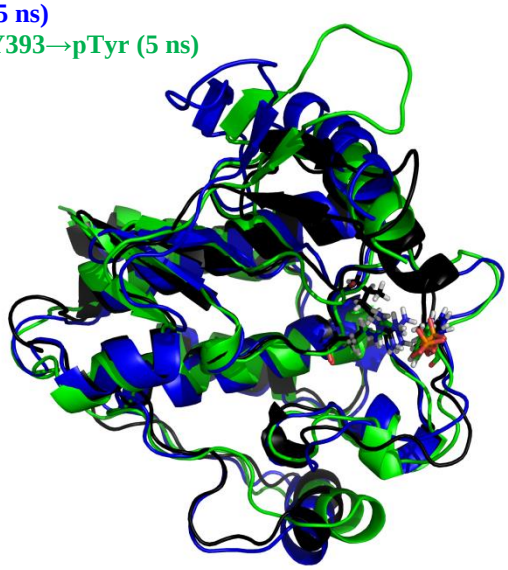
3DK6

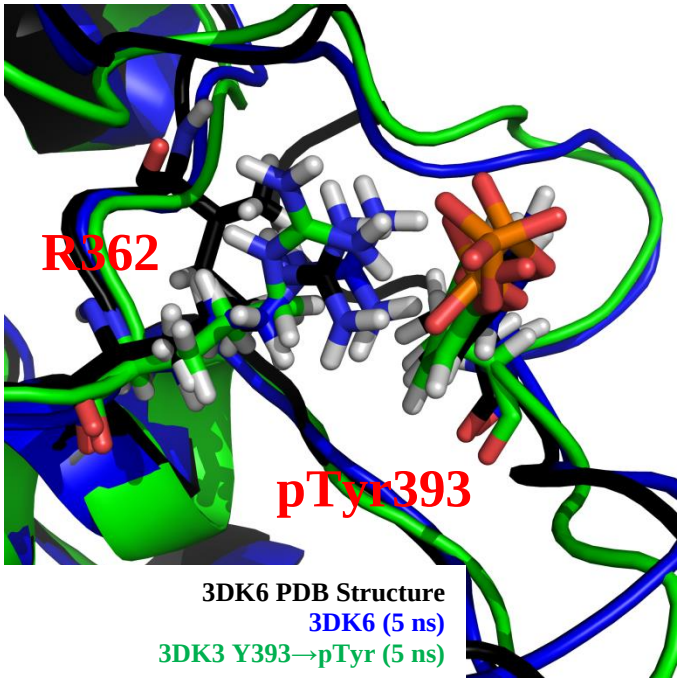
3DK3 Y393→pTyr



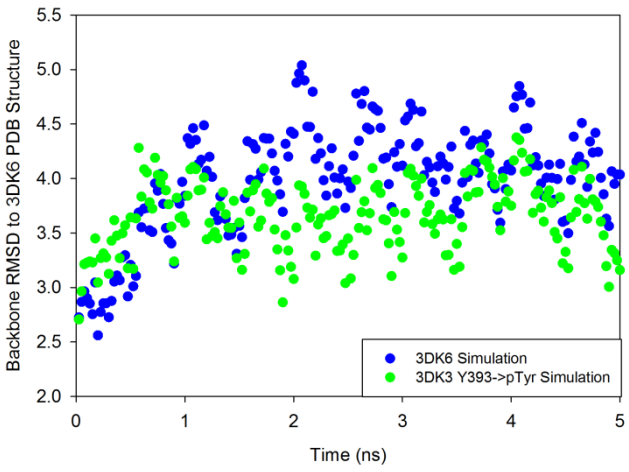


3DK6 PDB Structure
3DK6 (5 ns)
3DK3 Y393→pTyr (5 ns)



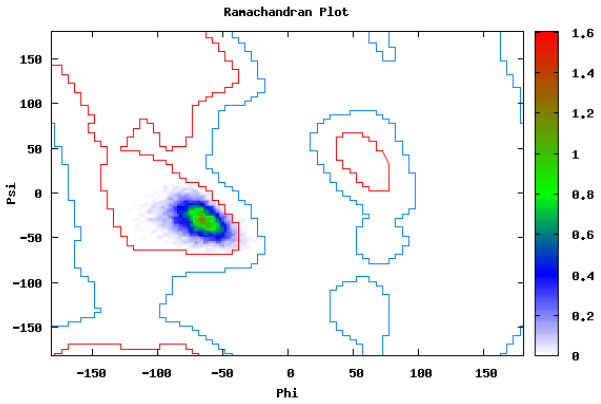


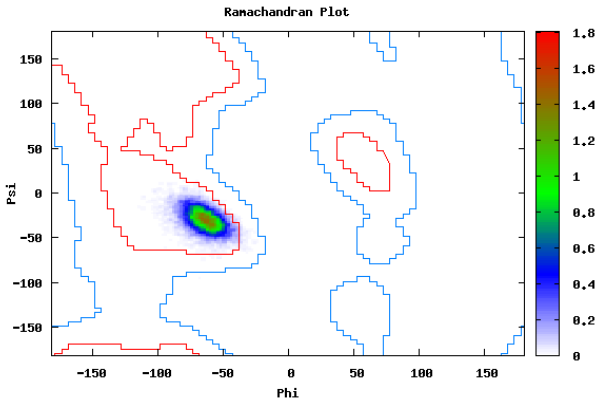
3DK6 PDB Structure
3DK6 (5 ns)
3DK3 Y393→pTyr (5 ns)

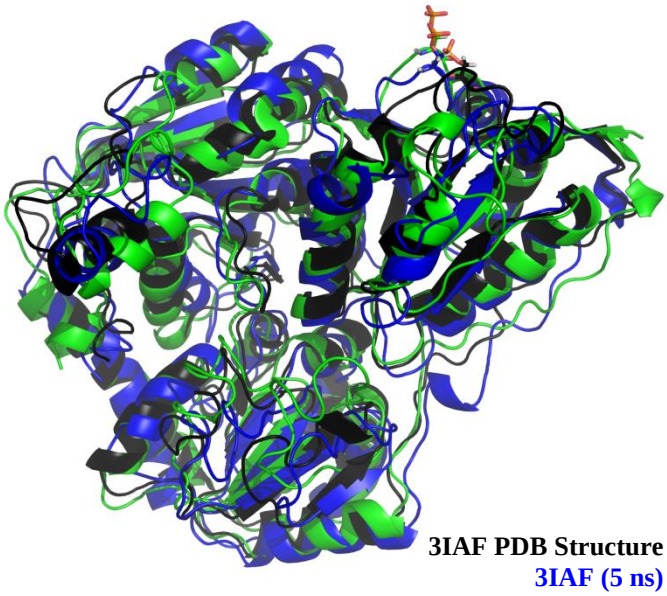


3IAF

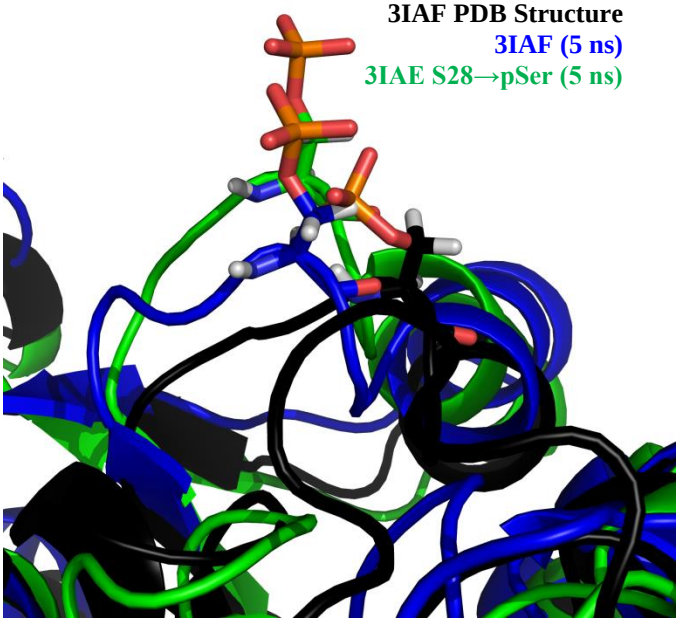
3IAE S28→pSer



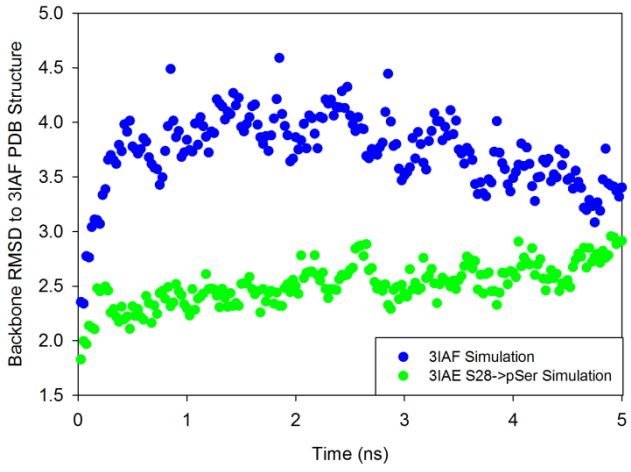




3IAF PDB Structure
3IAF (5 ns)
3IAE S28→pSer (5 ns)



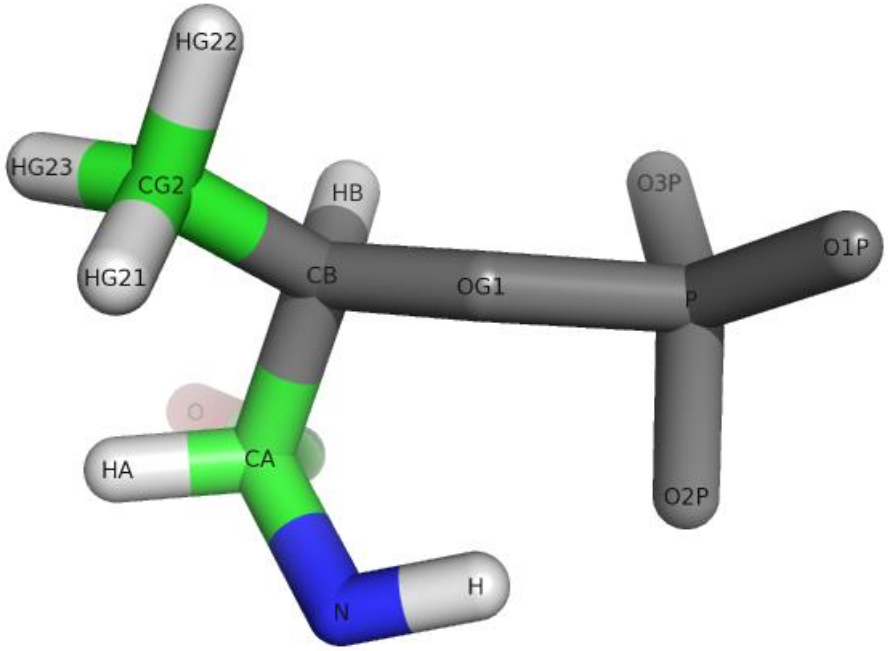
3IAF PDB Structure
3IAF (5 ns)
3IAE S28→pSer (5 ns)



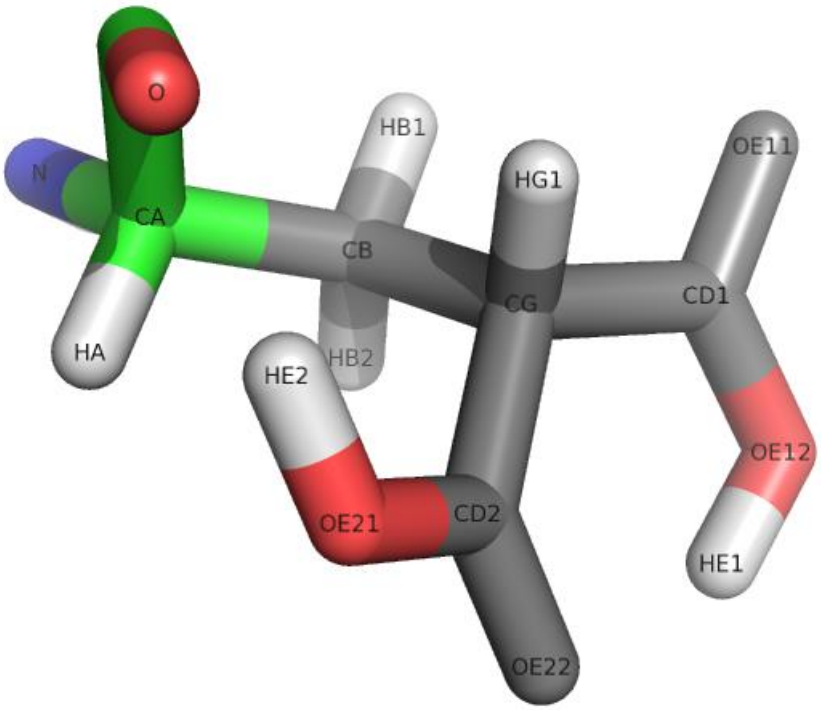
4.4 Images of Torsion Angles Assessed Using Ab Initio Rotational Profiles

Images of the specific torsions that were assessed for their *ab initio* rotational profiles are depicted below in grey.

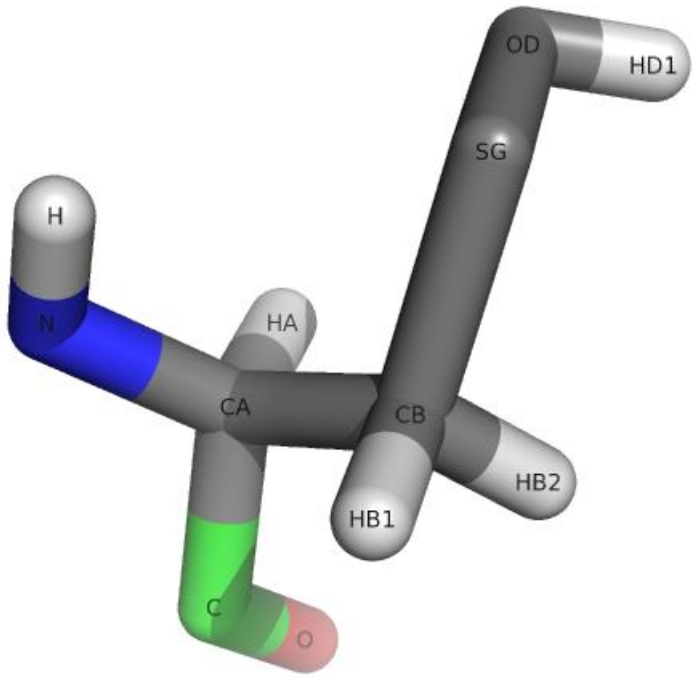
The torsion angles considered were CT-OS-P-OS from negatively charged phosphothreonine (TPO)



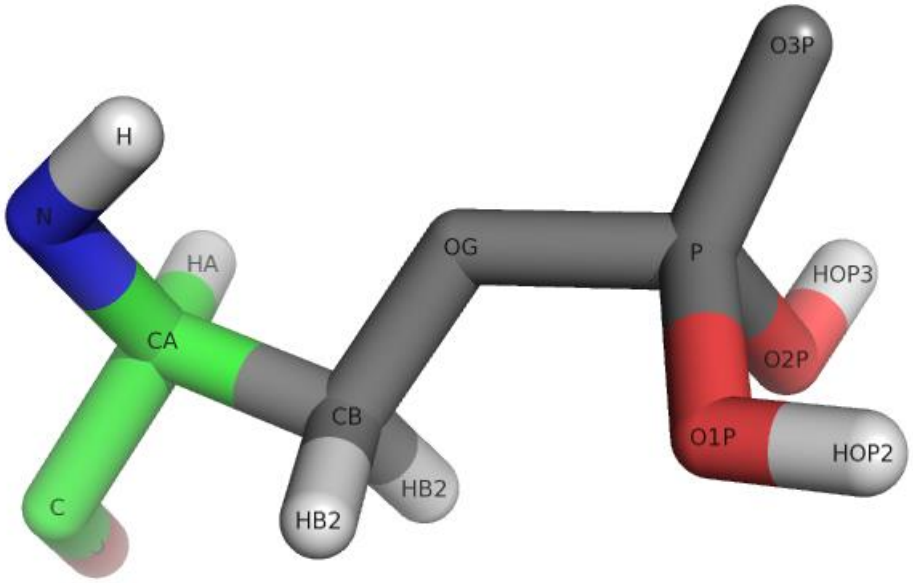
CT-CT-C-O from 1-carboxyglutamic acid (CGU)



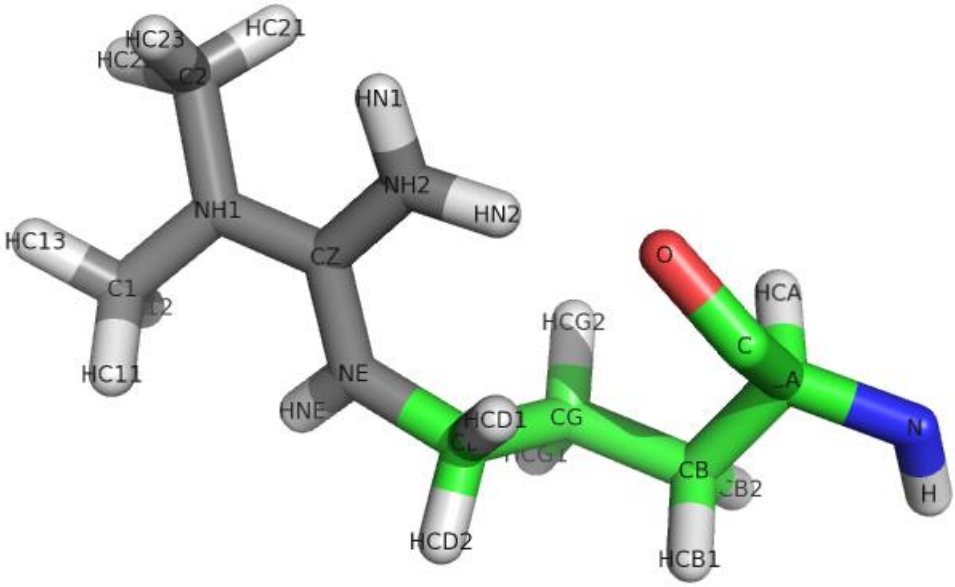
CT-CT-S-OH from cysteine sulfenic acid (CSO)



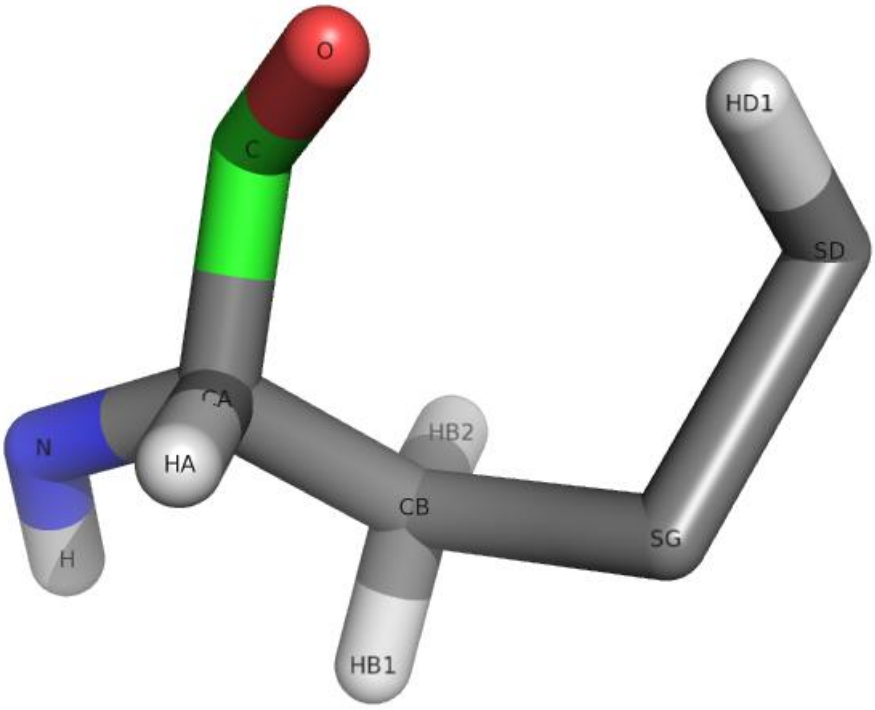
CT-OS-P-O2 from neutral phosphoserine (SEN)



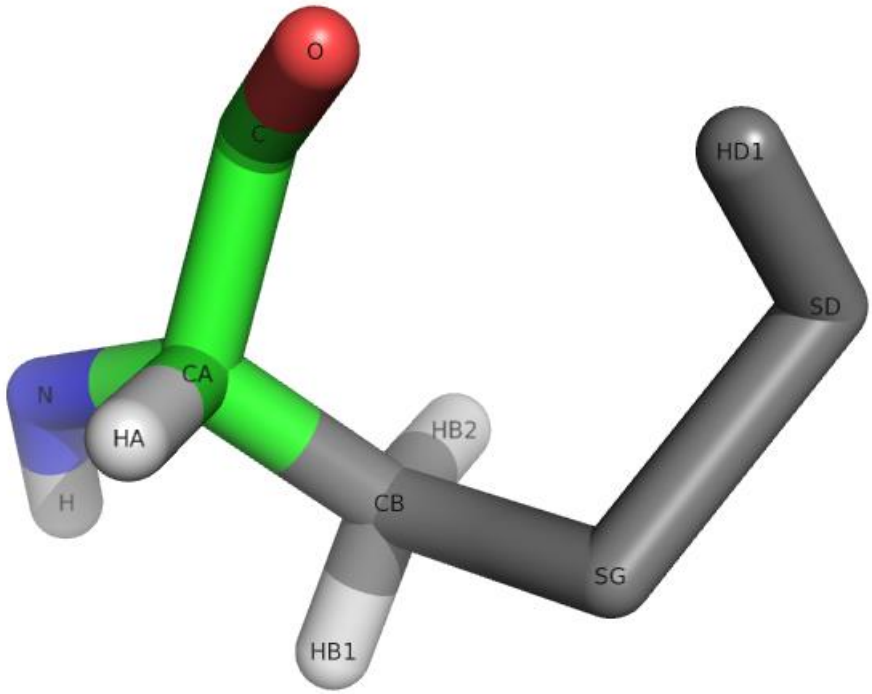
CT-N2-CM-N2 from antisymmetric dimethylarginine (DA2)



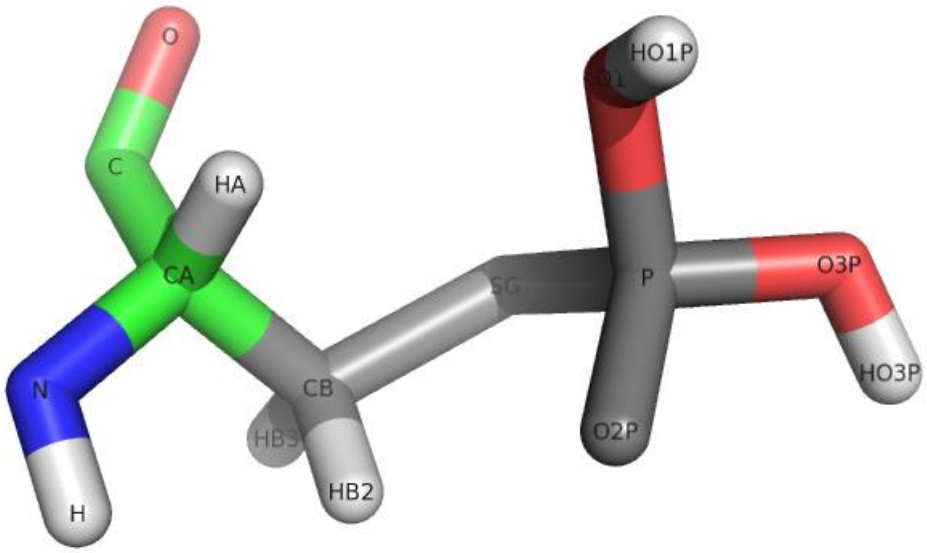
CT-CT-S-SH from cysteine persulfide (CSS)



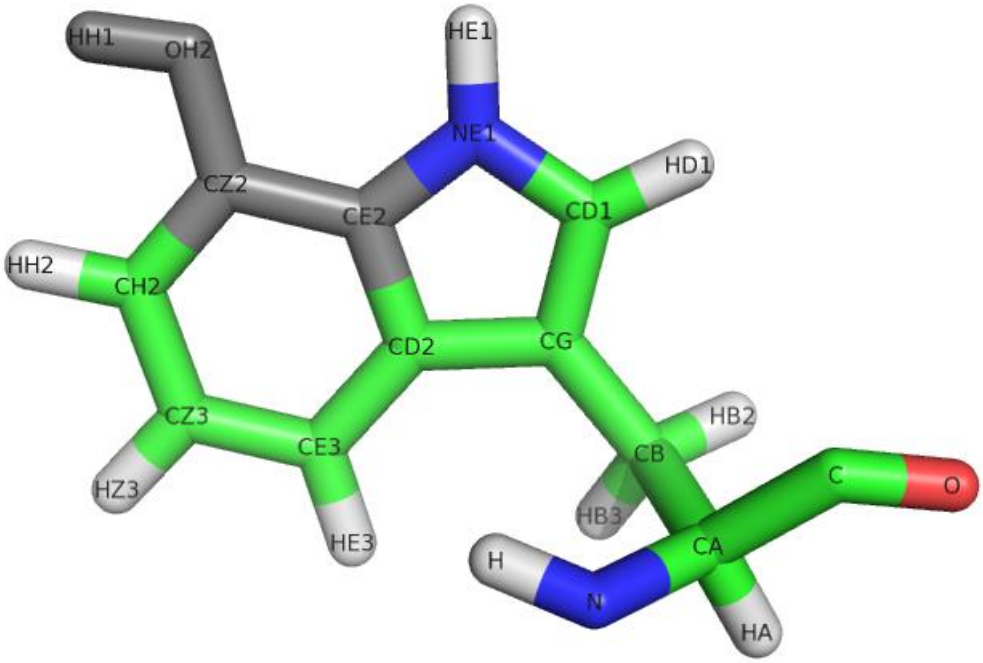
CT-S-SH-HS from cysteine persulfide (CSS)



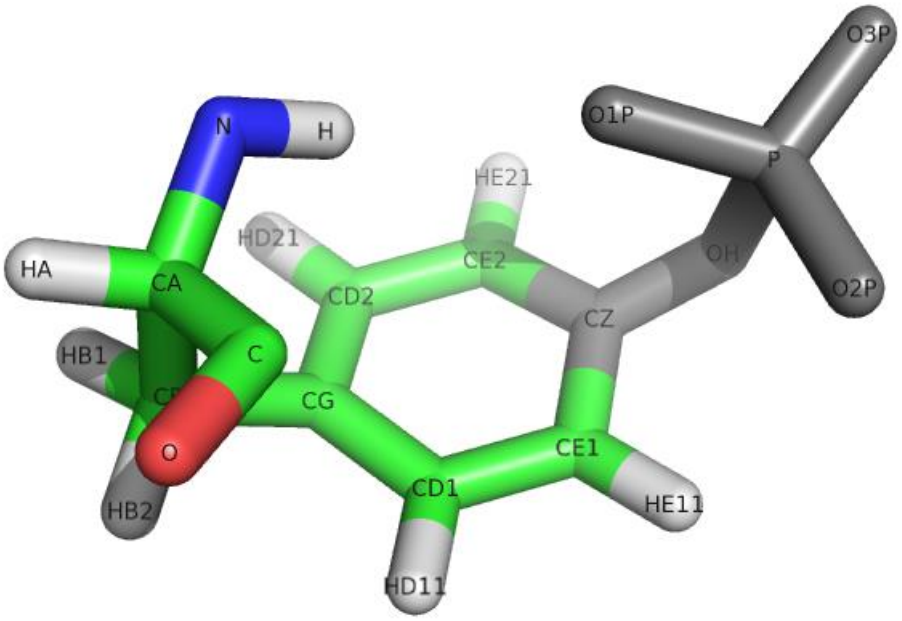
CT-S-P-O2 from neutral phosphocysteine (CSP)



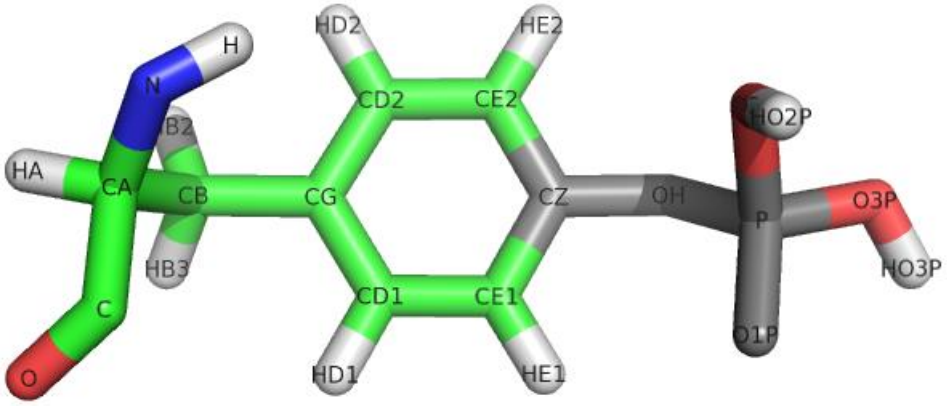
CN-CA-OH-HO from 7-hydroxytryptophan (0AF)



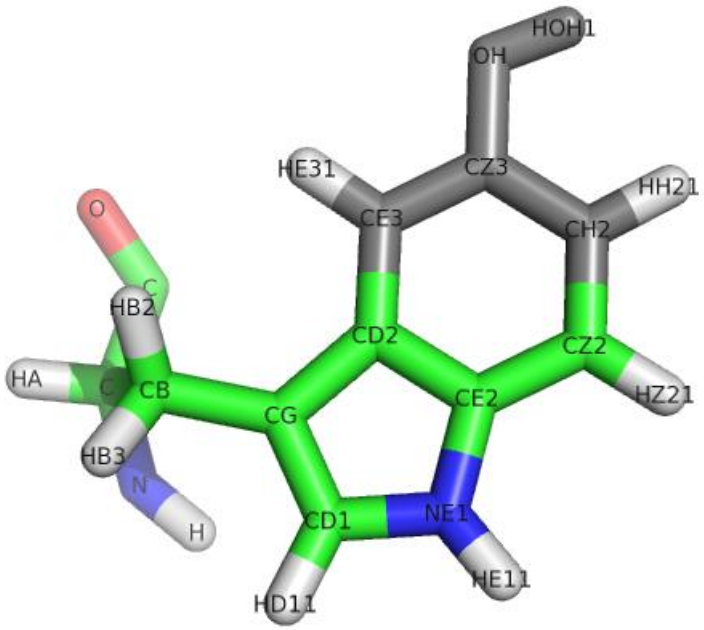
CA-OS-P-O2 from phosphotyrosine (PTR)



CA-OS-P-O2 from neutral phosphotyrosine (PTN)



and CA-CA-OH-HO from 5-hydroxytryptophan (HTR).



It should be noted in cases where there is a torsion with mixed atom types involved, such as in neutral phosphoserine (PTN), the entire phosphate group twists around and the correction is only applied to the single oxygen without the hydrogen (the double bonded one and not the hydroxyl hydrogens) since we observed corrections to these particular torsions will correct the rotation more significantly than the hydroxyl groups and the hydroxyl containing torsion will have little to no effect.

References

(1) Homeyer, N.; Horn, A.; Lanig, H.; Sticht, H. *J. Mol. Model.* **2006**, *12*, 281–289.

(2) Frishman, D.; Argos, P. *Proteins: Struct , Funct , Bioinf* **1995**, *23*, 566–579.

(3) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.

(4) DeLano, W. L. The PyMOL Molecular Graphics System. 2008.

(5) Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. *Bioinformatics* **2009**, *25*, 1422–1423.

(6) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714–2723.

(7) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. *Nucleic Acids Res* **2003**, *31*, 3381–3385.

(8) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. *Bioinformatics* **2006**, *22*, 195–201.

(9) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(10) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.;

Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, .; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

(11) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J Comput Chem* **1990**, *11*, 431–439.

(12) Page, N.; Schall, N.; Strub, J.-M.; Quinternet, M.; Chaloin, O.; Décossas, M.; Cung, M. T.; Van Dorsselaer, A.; Briand, J.-P.; Muller, S. *PloS one* **2009**, *4*, e5273.