Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.
This journal is © The Royal Society of Chemistry 2020

Supplementary Information

Development of Transferable Coarse-Grained Models of Amino Acids

Olivia Conway, Yaxin An, Karteek K. Bejagam, and Sanket A. Deshmukh*

Department of Chemical Engineering, Virginia Tech, Blacksburg, VA 24061, USA

*E-mail: sanketad@vt.edu, Phone: +1 540-231-8785.

S1. Implementation of Optimization Algorithms and Machine Learning in Force-Field Development:

Particle swarm optimization (PSO) is an algorithm that takes inspiration from the movements of a flock of birds or swarm of insects.[1] To begin with, a set of values (a "particle") for the force-field (FF) parameters to be optimized was randomly selected based on the supplied parameter ranges for use in an optimization simulation cycle. Several particles (typically 40 to 80) were generated for each simulation run or iteration in an optimization cycle. Each particle is analogous to a bird in a flock (or an insect in a swarm). In our case, we chose to generate 40 such particles for each iteration of the PSO for broad coverage of the parameter space to facilitate quicker identification of optimal FF parameters. At the end of each iteration, for each set of FF parameters (i.e. for each simulation from total 40 simulations in an iteration), heat of vaporization, surface tension, and density were calculated and were compared with the experimental values. Each FF parameter set was ranked based on the errors (difference between the calculated properties and experimental properties). The best set was defined as a set that corresponds to the lowest error. For the subsequent iteration, the best particle remains the same, and all other particles have their values adjusted slightly in the direction of the values of the best particle. A record of an individual particle's best set was kept, and updated when it changes so that the particle may be moved in that direction as well. PSO was implemented using **Equations S1** and **S2** in which V represents how quickly a given value changes (its velocity), x represents the value itself, the subscript n stands for the current step, and n+1 stands for the next step. The maximum amount by which each value can change was set to 0.02. The global best (the best set of values across all iterations) was represented by g_{best}, and p_{best} represented the personal best set of a particle. The value of rand() is a random

number between 0 and 1, *w* is an inertia factor, and c1 and c2 are swarm and personal constants, respectively. The first term in **Equation S1** is an inertial term that stabilizes the particle motion, and the constants c1 and c2 give the relative magnitude or pull in the direction of the global best and personal best, respectively. The position of a particle is updated based on the distance traveled over a discrete time interval ($\Delta t = 1$). In the case of this study, PSO was able to minimize the overall error within the first 40 or so iterations when using 40 particles.

$$V_{n+1} = w * V_n + c1 * rand() * (g_{best,n} - x_n) + c2 * rand() * (p_{best,n} - x_n) \quad \text{(S1)}$$

$$x_{n+1} = x_n + V_{m+1} * \Delta t \quad \text{(S2)}$$

In addition to the PSO, an artificial neural network (ANN) was incorporated after the first four iterations.[2] The data from those iterations and all subsequent iterations were given to the ANN which used the data to produce a parameter set for each following iteration that seemed likely, based on the ANN computations, to result in a reduced overall error. The ANN works in the following manner. A set of input values comprise the first "layer." The final layer is comprised of the output values, target values or predictions. Any layers between the input and output layers are called hidden layers, and they are comprised of "nodes" which are essentially processing centers, playing a role like that of biological neurons. Layers are connected to each other, receiving inputs from the nodes of the previous layer, and sending outputs to the next layer, except in the cases of the input layer, which has no inputs to itself and the output layer, which does not send output on to any further layers. The inputs to a node are individually weighted (with weights that can be tuned), and they are summed together and passed through an activation function to produce the output of the node. The activation function acts like an on/off switch such that the output is only passed on to the next layer if a certain threshold value is reached. For this research study, TensorFlow (a Python library created by Google developers) was utilized for implementation of the ANN with the exponential linear unit (ELU) activation function; an input layer, five hidden layers, and an output layer; and dropout to help prevent overfitting. The number of input nodes corresponded to the number of target properties for each system (e.g. targeting density, heat of vaporization, and surface tension corresponds to three input nodes), and the number of nodes in the output layer corresponded to the number of FF parameters to be fitted. Each hidden layer

contained fifty nodes. Incorporating the ANN into the PSO is not absolutely necessary for achieving good results, but it can lead to a satisfactorily minimal error sooner than PSO alone and is not in any way detrimental. 33 In some cases, it was even able to lead to more accurate predictions if the given ranges for the PSO did not encompass the best values, thus guiding the model development. **Figure S1** illustrates the process of ANN is incorporation into the optimization algorithm.

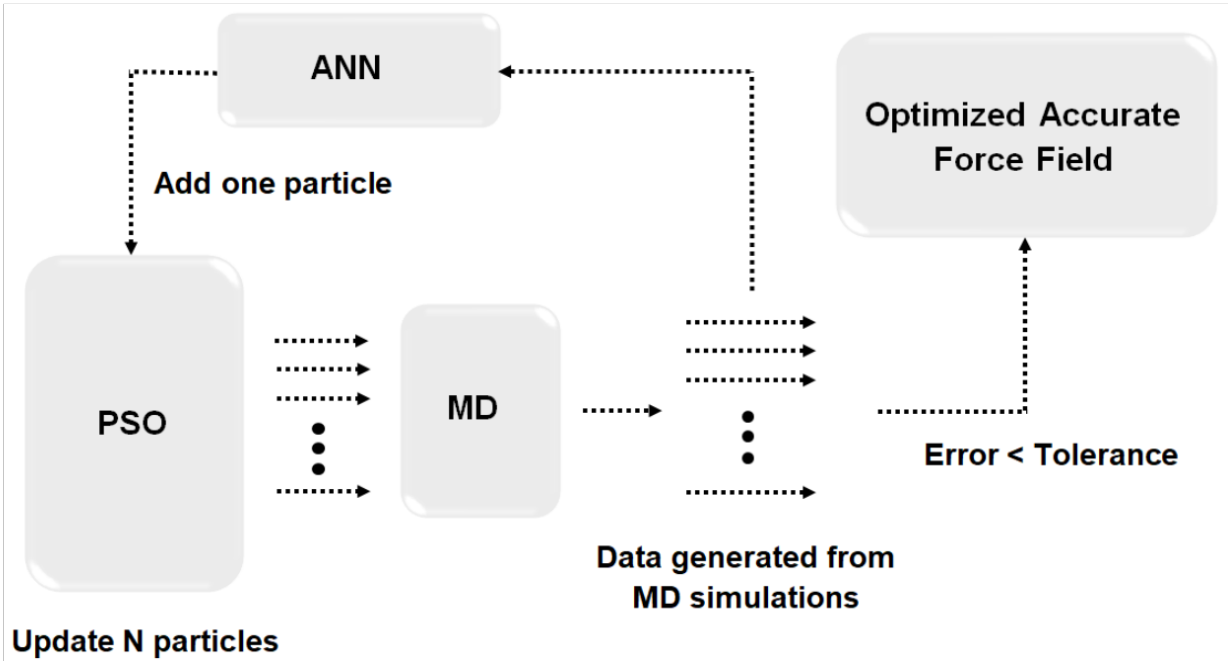
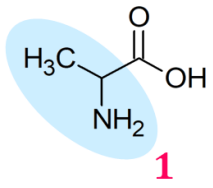
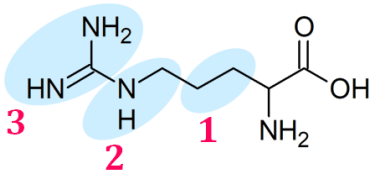
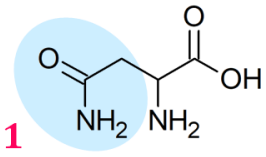
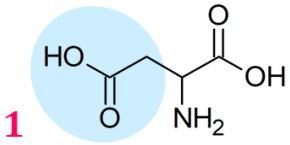
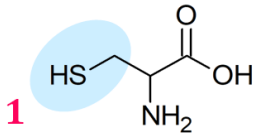
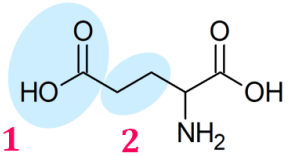
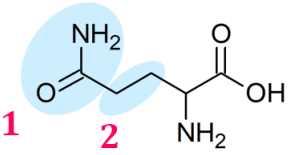
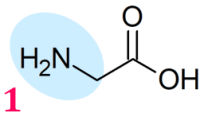
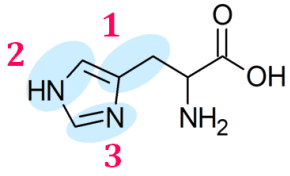
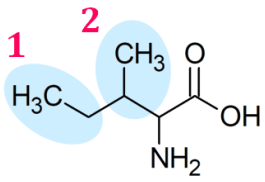
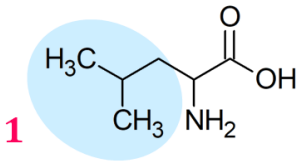
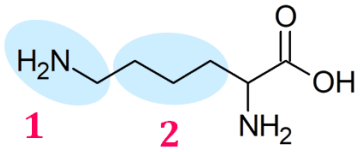
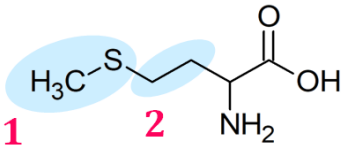
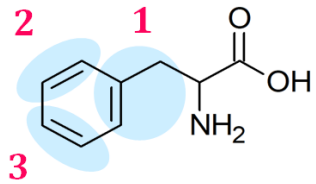
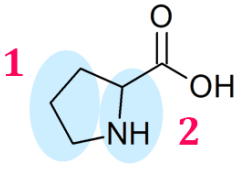


Figure S1: Schematic showing PSO and ANN incorporation with MD simulations to develop optimized force field parameters.

Table S.1: Amino acid mapping schemes, bead names, and bead types. The bead numbers in each amino acid is shown in red.

Name	Structure	Bead Name	Bead Type
Alanine		1) ANT (AB within backbone)	1) NCC1 (NCC2 within backbone)
Arginine		1) RS1 2) RS2 3) RS3	1) RS1 2) RS2 3) RS3
Asparagine		1) NS	1) CON2
Aspartic Acid		1) DS	1) COO2
Cysteine		1) CS	1) SC
Glutamic Acid		1) ES2 2) ES1	1) COO1 2) C2M
Glutamine		1) QS2 2) QS1	1) CON1 2) C2M

Glycine		1) GNT (GNC within backbone)	1) GNT (GNC within backbone)
Histidine		1) HS1 2) HS2 3) HS3	1) CCR1 2) NCR1 3) NCR2
Isoleucine		1) IS2 2) IS1	1) C2E 2) C2E2
Leucine		1) LS	1) C41
Lysine		1) KS2 2) KS1	1) NC2 2) C3M
Methionine		1) MS2 2) MS1	1) CS 2) C2M
Phenylalanine		1) FS1 2) FS21 3) FS22	1) TL2F 2) BZF1 3) BZF2
Proline		1) PS 2) PN	1) PS 2) PNC

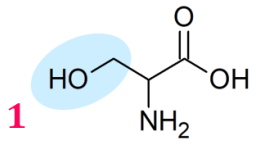
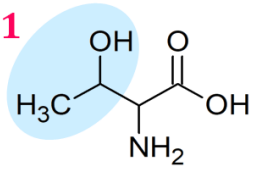
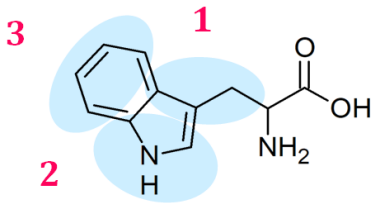
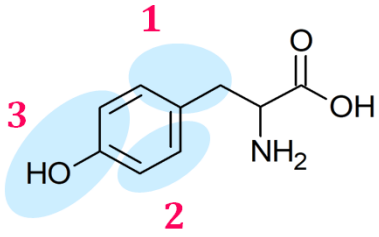
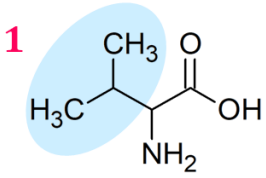
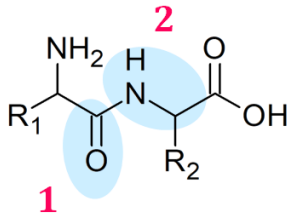
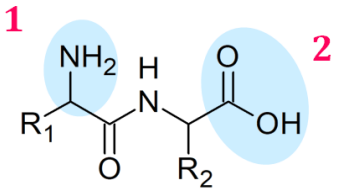
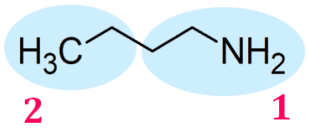
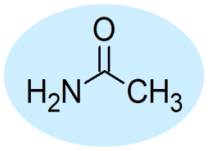
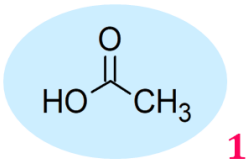
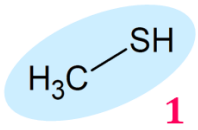
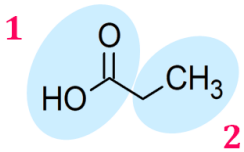
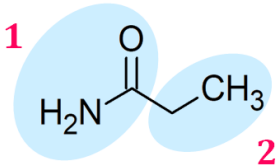
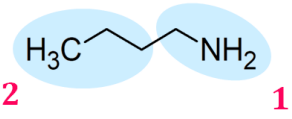
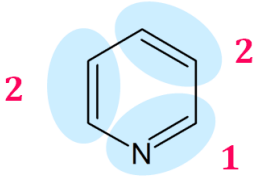
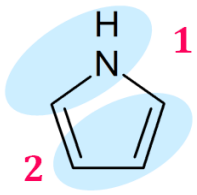
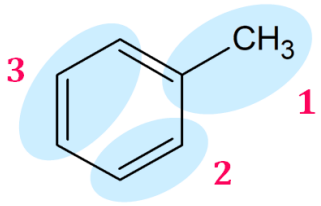
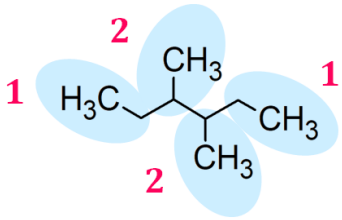
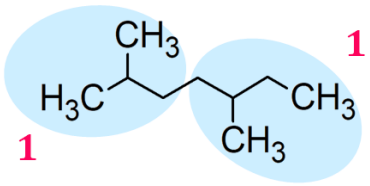
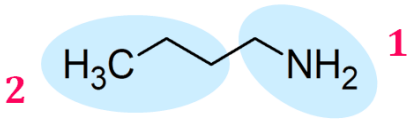
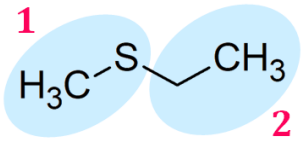
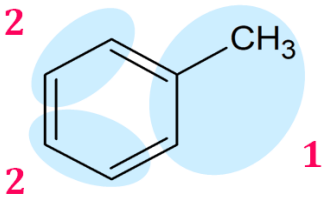
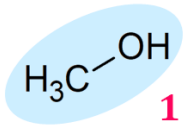
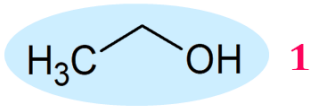
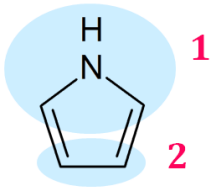
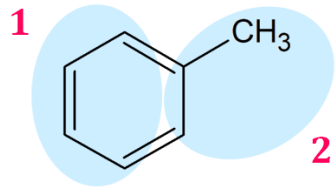
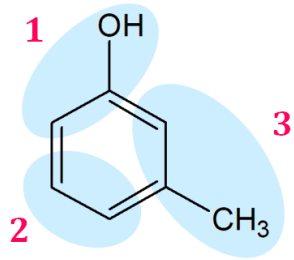
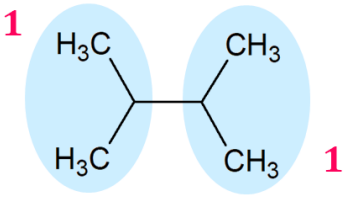
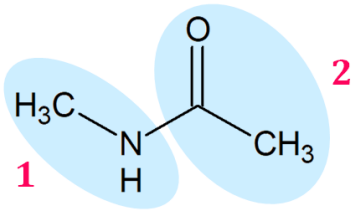
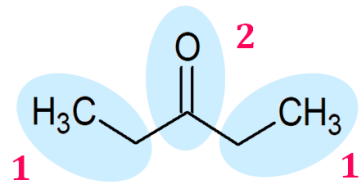
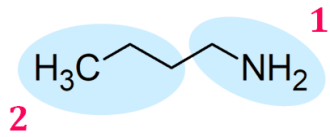
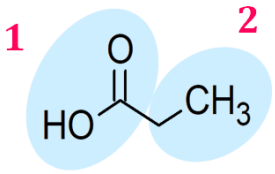
Serine		1) SS	1) COH1
Threonine		1) TS	1) CCOH
Tryptophan		1) WS1 2) WS2 3) WS3	1) TL2W 2) PL1 3) TOL4
Tyrosine		1) YS1 2) YS2 3) YS3	1) TL2Y 2) BZY 3) COH3
Valine		1) VS	1) C3E2
Backbone		1) BCO 2) BNC	1) CO 2) NC
Termini		1) NT 2) CT	1) NCT 2) COO1

Table S.2: Analogue mapping schemes and bead types

Amino Acid	Analogue	Structure	Bead Type
alanine	N-butylamine		1) NCC1 2) C2E
asparagine	acetamide		1) CON2
aspartic acid	acetic acid		1) COO2
cysteine	methanethiol		1) SC
glutamic acid	propionic acid		1) COO1 2) C2E
glutamine	propionamide		1) CON1 2) C2M
glycine	N-butylamine		1) NC2 2) C3M
histidine	pyridine		1) NCR2 2) BZ

histidine	pyrrole		1) NCR1 2) PL3
histidine	toluene		1) CCR1 2) BZ 3) TOL3
isoleucine	3,4-dimethylhexane		1) C2E 2) C2E2
leucine	2,5-dimethylhexane		1) C41
lysine	butylamine		1) NC2 2) C3E
methionine	ethyl methyl sulfide		1) CS 2) C2M
phenylalanine	toluene		1) TL2F 2) BZ (BZF1 and BZF2)

serine	methanol		1) COH1
threonine	ethanol		1) CCOH
tryptophan	pyrrole		1) PL1 2) PL2
tryptophan	toluene		1) TOL4 2) TOL2
tyrosine	m-cresol		1) COH3 2) BZY 3) TL2Y
valine	2,3-dimethylbutane		1) C3E2
backbone	methylacetamide		1) NC 2) CCO

backbone	diethylketone		1) CC 2) CO
N-terminus	N-butylamine		1) NCT 2) C3E
C-terminus	propionic acid		1) COO1 2) C2E

Topology File

!! Coarse-Grained Topology File for amino acids/peptides !!

```
MASS  -1  C2E          29.07
MASS  -1  C2M          28.06
MASS  -1  C3E          43.10
MASS  -1  C3M          42.09
MASS  -1  CCR1         26.04 ! His
MASS  -1  NCR1         28.04 ! His
MASS  -1  NCR2         27.03 ! His
MASS  -1  NCC1         44.09 ! Ala N-terminus
MASS  -1  NCC2         43.08 ! Ala backbone
MASS  -1  C41          57.13 ! Leu
MASS  -1  CON2         58.07 ! Asn
MASS  -1  COO2         59.05 ! Asp
MASS  -1  RS1          28.06 ! Arginine side chain bead
MASS  -1  RS2          29.05 ! Arginine side chain bead
MASS  -1  RS3          43.06 ! Arginine side chain bead
MASS  -1  SC           47.10 ! Cys
MASS  -1  CON1         44.04 ! Glutamine side chain outer
MASS  -1  GNT          30.06 ! Glycine N-terminus
MASS  -1  GNC          29.05 ! Glycine backbone
```

```
MASS  -1  C2E2          28.06 ! Ile side chain inner
MASS  -1  CS            47.10 ! Methionine side chain outer
MASS  -1  BZF2          26.04
MASS  -1  BZF1          26.04
MASS  -1  TL2F          39.06 ! Phenylalanine toluene bead
MASS  -1  TL2W          39.06 ! Tryptophan toluene bead
MASS  -1  TL2Y          39.06 ! Tyrosine toluene bead
MASS  -1  PNC           27.03 ! Proline N-C bead
MASS  -1  PS            42.09 ! Proline 3-C side chain bead
MASS  -1  COH1          31.04 ! Serine side chain bead
MASS  -1  CCOH          45.07 ! Threonine side chain
MASS  -1  TOL4          52.08 ! Trp
MASS  -1  PL1           40.05 ! Trp
MASS  -1  COH3          42.04
MASS  -1  BZY           26.04
MASS  -1  C3E2          43.10
MASS  -1  CO            28.01 ! backbone carbonyl
MASS  -1  NC            28.04 ! backbone N-C
MASS  -1  NC2           30.06
MASS  -1  COO1          45.02 ! neutral C-terminus
MASS  -1  NCT           29.05 ! neutral N-terminus
MASS  -1  H2O1          36.04

RESI HCT                0.00
ATOM  C1  C3E           0.00
ATOM  C2  C3M           0.00
ATOM  C3  C3M           0.00
ATOM  C4  C3M           0.00
ATOM  C5  C3M           0.00
ATOM  C6  CO            0.00

BOND  C1 C2  C2 C3  C3 C4  C4 C5  C5 C6

RESI ALA                0.00
ATOM  AB      NCC2      0.00
ATOM  BCO     CO        0.00

BOND  AB BCO

RESI ARG                0.00
ATOM  BNC     NC        0.00
ATOM  RS1     RS1       0.00
ATOM  RS2     RS2       0.00
```

```
ATOM  RS3      RS3      0.00
ATOM  BCO      CO       0.00

BOND BNC RS1  RS1 RS2  RS2 RS3  BNC BCO

RESI ASN                0.00
ATOM  BNC      NC       0.00
ATOM  NS       CON2     0.00
ATOM  BCO      CO       0.00

BOND  BNC NS  BNC CO

RESI ASP                0.00
ATOM  BNC      NC       0.00
ATOM  DS       COO2     0.00
ATOM  BCO      CO       0.00

BOND  BNC DS  BNC CO
 RESI CYS               0.00
ATOM  BNC      NC       0.00
ATOM  CS       SC       0.00
ATOM  BCO      CO       0.00

BOND  BNC CS  BNC CO

RESI GLU                0.00
ATOM  BNC      NC       0.00
ATOM  ES1      C2M      0.00
ATOM  ES2      COO1     0.00
ATOM  BCO      CO       0.00

BOND  BNC ES1  ES1 ES2  BNC BCO

RESI GLN                0.00
ATOM  BNC      NC       0.00
ATOM  QS1      C2M      0.00
ATOM  QS2      CON1     0.00
ATOM  BCO      CO       0.00

BOND  BNC QS1  QS1 QS2  BNC BCO

RESI GLY                0.00
ATOM  GNC      GNC      0.00
```

```
ATOM  BCO      CO       0.00

BOND GNC BCO

RESI HIS                0.00
ATOM  BNC      NC       0.00
ATOM  HS1      CCR1     0.00
ATOM  HS2      NCR1     0.00
ATOM  HS3      NCR2     0.00
ATOM  BCO      CO       0.00

BOND  BNC HS1  HS1 HS2  HS1 HS3  HS2 HS3  BNC BCO

RESI ILE                0.00
ATOM  BNC      NC       0.00
ATOM  IS1      C2E2     0.00
ATOM  IS2      C2E      0.00
ATOM  BCO      CO       0.00

BOND  BNC IS1  IS1 IS2  BNC BCO

RESI LEU                0.00
ATOM  BNC      NC       0.00
ATOM  LS       C41      0.00
ATOM  BCO      CO       0.00

BOND  BNC LS  BNC BCO

RESI LYS                0.00
ATOM  BNC      NC       0.00
ATOM  KS1      C3M      0.00
ATOM  KS2      NC2      0.00
ATOM  BCO      CO       0.00

BOND  BNC KS1  KS1 KS2  BNC BCO

RESI MET                0.00
ATOM  BNC      NC       0.00
ATOM  MS1      C2M      0.00
ATOM  MS2      CS       0.00

BOND  BNC MS1  MS1 MS2  BNC BCO
```

```
RESI PHE              0.00
ATOM  BNC     NC      0.00
ATOM  FS1     TL2F    0.00
ATOM  FS21    BZF1    0.00
ATOM  FS22    BZF2    0.00
ATOM  BCO     CO      0.00

BOND  BNC FS1  FS1 FS21  FS1 FS22  FS21 FS22  BNC BCO

RESI PRO              0.00
ATOM  PNC     PNC     0.00
ATOM  PS      PS      0.00
ATOM  BCO     CO      0.00

BOND  PNC PS  PNC BCO

RESI SER              0.00
ATOM  BNC     NC      0.00
ATOM  SS      COH1    0.00
ATOM  BCO     CO      0.00

BOND  BNC SS  BNC BCO

RESI THR              0.00
ATOM  BNC     NC      0.00
ATOM  TS      CCOH    0.00
ATOM  BCO     CO      0.00

BOND  BNC TS  BNC BCO

RESI TRP              0.00
ATOM  BNC     NC      0.00
ATOM  WS1     TL2W    0.00
ATOM  WS2     PL1     0.00
ATOM  WS3     TOL4    0.00

BOND  BNC WS1  WS1 WS2  WS1 WS3  WS2 WS3  BNC BCO

RESI TYR              0.00
ATOM  BNC     NC      0.00
ATOM  YS1     TL2Y    0.00
ATOM  YS2     BZY     0.00
ATOM  YS3     COH3    0.00
```

```
ATOM  BCO      CO       0.00

BOND  BNC YS1  YS1 YS2  YS1 YS3  YS2 YS3  BNC BCO

RESI VAL                0.00
ATOM  BNC      NC       0.00
ATOM  VS       C3E2     0.00
ATOM  BCO      CO       0.00

BOND  BNC VS  BNC BCO

PRES PEPB               ! Normal amino acid linkage
ATOM 1BCO     CO        0.00
ATOM 2BNC     NC        0.00

BOND 1BCO 2BNC


PRES PEPG               ! Amino acid to glycine linkage
ATOM 1BCO     CO        0.00
ATOM 2GNC     GNC       0.00

BOND 1BCO 2GNC

PRES PEPP               ! Amino acid to proline linkage
ATOM 1BCO     CO        0.00
ATOM 2PNC     PNC       0.00

BOND 1BCO 2PNC

PRES AT                 0.00 ! Neutral alanine N-terminus
ATOM ANT      NCC1      0.00
ATOM BCO      CO        0.00
DELETE ATOM AB
BOND ANT BCO

PRES NTER               0.00 ! Neutral standard N-terminus
ATOM NT       NCT       0.00
ATOM BCO      CO        0.00
DELETE ATOM BNC
BOND NT BCO

PRES GT                 0.00 ! Neutral glycine N-terminus
```

```
ATOM GNT      GNT     0.00
ATOM BCO      CO      0.00
DELETE ATOM GNC
BOND GNT BCO

PRES CTER             0.00 ! Neutral C-terminus
ATOM CT       COO1    0.00
ATOM BNC      NC      0.00
DELETE ATOM BCO
BOND  BNC CT

PRES HCTA             0.00
ATOM 1C6      CO      0.00
ATOM 2AB      NCC2    0.00
BOND 1C6 2AB

RESI WAT              0.00
ATOM H2O      H2O1    0.00
```

Parameter File

!! Coarse-Grained Parameter File for the amino acids and peptides !!
!! This is for use with "exclude 1-2" ONLY. !!
!! Only neutral amino acids and neutral termini are represented. !!

ATOMS
MASS -1 C2E 29.07
MASS -1 C2M 28.06
MASS -1 C3E 43.10
MASS -1 C3M 42.09
MASS -1 CCR1 26.04 ! His
MASS -1 NCR1 28.04 ! His
MASS -1 NCR2 27.03 ! His
MASS -1 NCC1 44.09 ! Ala N-terminus
MASS -1 NCC2 43.08 ! Ala backbone
MASS -1 C41 57.13 ! Leu
MASS -1 CON2 58.07 ! Asn
MASS -1 COO2 59.05 ! Asp
MASS -1 RS1 28.06 ! Arginine side chain bead
MASS -1 RS2 29.05 ! Arginine side chain bead
MASS -1 RS3 43.06 ! Arginine side chain bead
MASS -1 SC 47.10 ! Cys
MASS -1 CON1 44.04 ! Glutamine side chain outer
MASS -1 GNT 30.06 ! Glycine N-terminus
MASS -1 GNC 29.05 ! Glycine backbone
MASS -1 C2E2 28.06 ! Ile side chain inner
MASS -1 CS 47.10 ! Methionine side chain outer
MASS -1 BZF2 26.04
MASS -1 BZF1 26.04
MASS -1 TL2F 39.06 ! Phenylalanine toluene bead
MASS -1 TL2W 39.06 ! Tryptophan toluene bead
MASS -1 TL2Y 39.06 ! Tyrosine toluene bead
MASS -1 PNC 27.03 ! Proline N-C bead
MASS -1 PS 42.09 ! Proline 3-C side chain bead
MASS -1 COH1 31.04 ! Serine side chain bead
MASS -1 CCOH 45.07 ! Threonine side chain
MASS -1 TOL4 52.08 ! Trp
MASS -1 PL1 40.05 ! Trp
MASS -1 COH3 42.04
MASS -1 BZY 26.04
MASS -1 C3E2 43.10
MASS -1 CO 28.01 ! backbone carbonyl

```
MASS  -1  NC          28.04 ! backbone N-C
MASS  -1  NC2         30.06
MASS  -1  COO1        45.02 ! neutral C-terminus
MASS  -1  NCT         29.05 ! neutral N-terminus
MASS  -1  H2O1        36.04


BONDS
!
!V(bond) = Kb(r - r0)**2
!
!Kb: kcal/mole/A**2
!r0: A
!
!bead types           Kb      r0

! Alanine
NCC1  CO              15      2.57
CO      NCC2          10      2.63
NCC2  COO1            30      2.54

! Asparagine
CON2   NCT            8       3.00
NC      CON2          8       2.75

! Aspartic Acid
COO2   NCT            8       3.00
NC      COO2          55      2.71

! Arginine
NC      RS1           10      2.30 ! RS1=C2M
RS1     RS2           15      2.53
RS2     RS3           50      2.4

! Cysteine
SC      NCT           20      2.65
NC      SC            55      2.58

! Glutamic Acid
COO1   C2M            90.98   2.5
C2M     NCT           15      2.45 ! Same as Gln, Met
NC      C2M           34      2.30 ! Same as Gln, Met
```

```
! Glutamine
CON1   C2M          42      2.44
C2M    NCT          15      2.45 ! Same as Glu, Met
NC     C2M          34      2.30 ! Same as Glu, Met

! Glycine
GNT    CO           50      2.50
CO     GNC          70      2.15
GNC    COO1         60      2.50

! Histidine
NCR2   NCR1         50      1.86
NCR1   CCR1         30      2.45
NCR2   CCR1         50      2.53
CCR1   NCT          30      2.35
NC     CCR1         80      2.26

! Isoleucine
C2E    C2E2         44      2.41
C2E2   NCT          40      2.50
NC     C2E2         30      2.28

! Leucine
C41    NCT          12      3.18
NC     C41          10      2.95

! Lysine
NC2    C3M          20      3.07
C3M    NCT          5       2.90
NC     C3M          10      2.52

! Methionine
CS     C2M          32      2.5
C2M    NCT          15      2.45 ! Same as Glu, Gln
NC     C2M          34      2.30 ! Same as Glu, Gln

! Phenylalanine
BZF1   BZF2         62      2.23
BZF1   TL2F         55      2.875
BZF2   TL2F         53      2.45
TL2F   NCT          7       2.75
NC     TL2F         17      2.43
```

```
! Proline
CO      PNC             40      2.04
PNC     PS              45      1.77

! Serine
COH1    NCT             30      2.5
NC      COH1            35      2.28

! Threonine
CCOH    NCT             55      2.58
NC      CCOH            45      2.25

! Tryptophan
TOL4    PL1             60      3.11
TOL4    TL2W            40      3.358
TL2W    PL1             60      2.195
TL2W    NCT             20      2.90
NC      TL2W            10      2.50

! Tyrosine
TL2Y    BZY             50      2.45
BZY     COH3            50      2.45
COH3    TL2Y            50      3.52
TL2Y    NCT             15      2.83
NC      TL2Y            15      2.42

! Valine
C3E2    NCT             40      2.44
NC      C3E2            35      2.26

! Backbone
NCT     CO              35      2.35
CO      NC              60      2.08
NC      COO1            70      2.48

! Carbon-Carbon & Carbon-Peptide
C3E     C3M             37.695          3.561
C3M     C3M             37.695          3.561 ! from C3E-C3M
C3M     CO              38.537          2.914 ! from C2M-C3M
C2E     C2M             35.293          2.561
C2M     C2M             36.609          2.520
```

ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!

!bead types			Ktheta	Theta0	
! Alanine					
NCC1	CO	NCC2	20	97	
CO	NCC2	CO	5	91	
NCC2	CO	NCC2	17	96	
CO	NCC2	COO1	6	83	
! Asparagine					
CON2	NCT	CO	6	68	
CO	NC	CON2	3	138	
CON2	NC	COO1	50	114	
! Aspartic Acid					
COO2	NCT	CO	5	75	
CO	NC	COO2	2	138	
COO2	NC	COO1	27	114	
! Arginine					
CO	NC	RS1	10	140	
RS1	NC	COO1	20	83	
NC	RS1	RS2	10	153	
RS1	RS2	RS3	5	157	
! Cysteine					
SC	NCT	CO	10	71	
CO	NC	SC	30	166	
SC	NC	COO1	35	84	
! Glutamic Acid					
COO1	C2M	NCT	25	114	
C2M	NCT	CO	10	73	! Same in Gln, Met
CO	NC	C2M	12	173	! Same in Gln, Met
NC	C2M	COO1	25	162	
C2M	NC	COO1	20	83	! Same in Gln, Met

```
! Glutamine
CON1  C2M   NCT    30   100
C2M   NCT   CO     10   73    ! Same in Gln, Met
CO    NC    C2M    12   173   ! Same in Gln, Met
NC    C2M   CON1   30   160
C2M   NC    COO1   20   83    ! Same in Gln, Met

! Glycine
GNT   CO    GNC    30   90
CO    GNC   COO1   20   97

! Histidine
NCR2  CCR1  NCR1   40   44    ! This and next 3 must add to 180deg
CCR1  NCR1  NCR2   25   70
NCR1  NCR2  CCR1   25   66
NCR1  CCR1  NCT    5    140
CCR1  NCT   CO     40   68
CO    NC    CCR1   1    140
COO1  NC    CCR1   100  90
NC    CCR1  NCR2   7    124
NCR2  CCR1  NCT    10   135
NC    CCR1  NCR1   7    120

! Isoleucine
C2E   C2E2  NCT    10   70
C2E2  NCT   CO     30   70
CO    NC    C2E2   10   146
NC    C2E2  C2E    20   70
C2E2  NC    COO1   40   100

! Leucine
C41   NCT   CO     6    73
CO    NC    C41    2    144
C41   NC    COO1   6    110

! Lysine
NC2   C3M   NCT    3    140
C3M   NCT   CO     3    75
CO    NC    C3M    1    140
NC    C3M   NC2    5    140
C3M   NC    COO1   40   106
```

```
! Methionine
CS      C2M     NCT             2       120
C2M     NCT     CO              10      73      ! Same in Gln, Glu
CO      NC      C2M             12      173     ! Same in Gln, Glu
NC      C2M     CS              2       130
C2M     NC      COO1            20      83      ! Same in Gln, Glu

! Phenylalanine
BZF2    BZF1    TL2F            31      55.6
BZF1    BZF2    TL2F            32      75.75
BZF1    TL2F    BZF2            33      48.65
BZF2    TL2F    NCT             7       110
TL2F    NCT     CO              8       70
CO      NC      TL2F            5       135
COO1    NC      TL2F            40      108
NC      TL2F    BZF1            50      129.5
NC      TL2F    BZF2            5       110
BZF1    TL2F    NCT             10      145

! Proline
NCT     CO      PNC             5       111
CO      PNC     PS              8       144
CO      PNC     CO              3       106
PNC     CO      NC              10      103

! Serine
COH1    NCT     CO              15      74
CO      NC      COH1            30      173
COH1    NC      COO1            25      81

! Threonine
CCOH    NCT     CO              30      83
CO      NC      CCOH            7       157
CCOH    NC      COO1            50      93

! Tryptophan
TL2W    PL1     TOL4            40      76.4    ! this and next three = 180
PL1     TOL4    TL2W            40      39.4
TOL4    TL2W    PL1             40      64.2
TOL4    TL2W    NCT             10      140
PL1     TL2W    NCT             10      135
TL2W    NCT     CO              20      75
```

```
! Tyrosine
TL2Y   BZY    COH3        25      92
BZY    COH3   TL2Y        25      44
COH3   TL2Y   BZY         25      44
BZY    TL2Y   NCT         5       117
COH3   TL2Y   NCT         15      145
TL2Y   NCT    CO          15      67
TL2Y   NC     COO1        12      108
CO     NC     TL2Y        2       123
BZY    TL2Y   NC          2       110
COH3   TL2Y   NC          20      128

! Valine
C3E2   NCT    CO          20      75
CO     NC     C3E2        2       160
C3E2   NC     COO1        28      77

! Carbon-Carbon & Carbon-Peptide

C3E    C3M    C3M         3.4     145     ! based on C3E-C3M-C3E
C3M    C3M    C3M         3.4     145     ! based on C3E-C3M-C3E
C3M    C3M    CO          3.4     145     ! based on C3E-C3M-C3E
C3M    CO     NCC2        1       140     ! from Lys CO-NC-C3M

C2E    C2M    C2M         3.886   147.766
C2M    C2M    C2M         3.460   144.226

! Backbone
NCT    CO     NC          7       90
CO     NC     COO1        1       100
NC     CO     NC          8       103     ! from valylvalylvaline
CO     NC     CO          1       120     ! from valylvalylvaline
NC     CO     NCC2        20      97      ! from NCC1 CO NCC2 (alanine)


DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
```

! bead types				Kchi	n	delta
! Alanine						
NCC1	CO	NCC2	CO	1.0	1	90
CO	NCC2	CO	NCC2	0.6	1	0
NCC2	CO	NCC2	COO1	1.2	1	0
! Asparagine						
NS	NCT	CO	NC	0.3	2	0
NCT	CO	NC	NS	1.45	1	100
! Aspartic Acid						
DS	NCT	CO	NC	0.3	2	0
NCT	CO	NC	DS	1.7	1	102
! Arginine						
ANT	CO	NC	RS1	0.7	2	180
CO	NC	RS1	RS2	0.7	2	0
COO1	NC	RS1	RS2	0.7	1	180
NC	RS1	RS2	RS3	0.7	2	0
! Cysteine						
CS	NCT	CO	NC	1.0	1	0
NCT	CO	NC	CS	0.6	2	0
! Glutamic Acid						
COO1	C2M	NCT	CO	1.0	2	0
C2M	NCT	CO	NC	0.7	2	0
NCT	CO	NC	C2M	0.7	2	0
CO	NC	C2M	COO1	1.0	1	90
COO1	C2M	NC	COO1	1.0	1	90
! Glutamine						
CON1	C2M	NCT	CO	0.5	2	0
C2M	NCT	CO	NC	0.7	2	0
NCT	CO	NC	C2M	0.7	2	0
CO	NC	C2M	CON1	0.5	2	0
CON1	C2M	NC	COO1	0.5	1	180
! Glycine						
GNT	CO	GNC	COO1	0.5	2	0

! Histidine

NCR1	CCR1	NCT	CO	0.1	3	180
NCR2	CCR1	NCT	CO	0.1	1	180
CCR1	NCT	CO	NC	0.1	2	81
NCT	CO	NC	CCR1	0.5	1	60
CO	NC	CCR1	NCR2	0.5	1	180
CO	NC	CCR1	NCR1	0.5	1	180
COO1	NC	CCR1	NCR2	0.8	1	0
COO1	NC	CCR1	NCR1	0.5	1	0

! Isoleucine

C2E	C2E2	NCT	CO	0.5	1	0
C2E2	NCT	CO	NC	0.5	2	0
NCT	CO	NC	C2E2	0.5	1	90
CO	NC	C2E2	C2E	0.5	1	270
C2E	C2E2	NC	COO1	0.5	2	180

! Leucine

C41	NCT	CO	NC	0.5	1	0
NCT	CO	NC	C41	0.6	1	90

! Lysine

NC2	C3M	NCT	CO	0.2	2	76
C3M	NCT	CO	NC	0.7	1	90
NCT	CO	NC	C3M	2.0	1	120
CO	NC	C3M	NC2	0.2	3	180
NC2	C3M	NC	COO1	0.1	3	0

! Methionine

CS	C2M	NCT	CO	0.5	1	0
C2M	NCT	CO	NC	0.7	2	0
NCT	CO	NC	C2M	0.7	2	0
CO	NC	C2M	CS	0.5	1	0
CS	C2M	NC	COO1	0.4	1	60

! Phenylalanine

BZF1	TL2F	NCT	CO	0.3	2	0
BZF2	TL2F	NCT	CO	0.4	2	0
TL2F	NCT	CO	NC	0.3	2	76
NCT	CO	NC	TL2F	1	1	90
CO	NC	TL2F	BZF1	0.5	1	270
CO	NC	TL2F	BZF2	0.3	1	120
BZF1	TL2F	NC	COO1	0.3	1	270

```
BZF2   TL2F   NC     COO1      0.5    2      180

! Proline
C3E2   NCT    CO     PNC       1.0    2      0
NCT    CO     PNC    PS        0.3    1      60
NCT    CO     PNC    CO        0.3    1      60
CO     PNC    CO     NC        0.5    1      180
PS     PNC    CO     NC        0.3    2      0
PNC    CO     NC     C41       0.5    1      60
PNC    CO     NC     COO1      1.7    1      0

! Serine
COH1   NCT    CO     NC        0.5    2      0
NCT    CO     NC     COH1      0.8    1      0

! Threonine
CCOH   NCT    CO     NC        0.3    2      0
NCT    CO     NC     CCOH      1      1      90

! Tryptophan
TOL4   TL2W   NCT    CO        0.8    2      180
PL1    TL2W   NCT    CO        0.7    2      0
TL2W   NCT    CO     NC        1.0    2      0
NCT    CO     NC     TL2W      1.0    1      180
CO     NC     TL2W   PL1       1.0    1      180
CO     NC     TL2W   TOL4      1.0    1      180
COO1   NC     TL2W   PL1       1.2    1      0
COO1   NC     TL2W   TOL4      1.3    1      0

! Tyrosine
COH3   TL2Y   NCT    CO        0.5    1      180
BZY    TL2Y   NCT    CO        0.5    1      180
TL2Y   NCT    CO     NC        0.5    1      0
NCT    CO     NC     TL2Y      0.5    2      180
CO     NC     TL2Y   COH3      0.7    1      180
CO     NC     TL2Y   BZY       0.5    1      180
COH3   TL2Y   NC     COO1      1.0    1      0
BZY    TL2Y   NC     COO1      0.5    1      0

! Valine
C3E2   NCT    CO     NC        1.3    2      90
NCT    CO     NC     C3E2      2.0    1      43
```

```
! Backbone
NCT     CO      NC      COO1            0.5     1       0
NCT     CO      NC      CO              2.0     1       0       ! from valylvalylvaline (t8)
CO      NC      CO      NC              0.1     2       0       ! from valylvalylvaline (t8)
NC      CO      NC      COO1            2.5     1       0       ! from valylvalylvaline (t8)

X       X       X       X               0       0       0       ! sets all others not listed to 0
! So all dihedrals like LS1-NC-CO-NC and C3M-C3M-C3M are set to 0!


NONBONDED
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!bead            ignored         epsilon         Rmin/2

CCR1            0.000000        -0.3058         2.4725
NCR1            0.000000        -0.7506         1.8759
NCR2            0.000000        -0.4417         2.1546
NCT             0.000000        -0.6000         2.1650          ! same as NC2 (Lys bead)
CO              0.000000        -0.6690         2.4361
NC              0.000000        -0.6765         2.1436
COO1            0.000000        -1.2516         2.2088          ! from reference [3] propionic acid bead
NCC1            0.000000        -0.8699         2.4014
NCC2            0.000000        -0.8699         2.4014
C41             0.000000        -0.7120         2.8515
NC2             0.000000        -0.6000         2.1650
CON2            0.000000        -1.1332         2.4343
COO2            0.000000        -0.8714         2.3950
RS1             0.000000        -0.3420         2.4343          ! Arg side chain; same as C2M
RS2             0.000000        -0.6765         2.1436          ! Arg side chain; same as NC bead
RS3             0.000000        -1.0881         2.2635          ! Arg side chain
SC              0.000000        -0.8638         2.3606
CON1            0.000000        -1.0876         2.0609
GNT             0.000000        -0.6000         2.1650          ! same as NCT
GNC             0.000000        -0.6765         2.1436          ! same as NC
C2E2            0.000000        -0.3544         2.4146
CS              0.000000        -0.7001         2.3969
BZF1            0.000000        -0.3205         2.2332          ! from reference [4]
BZF2            0.000000        -0.3205         2.2332          ! from reference [4]
```

TOL2	0.000000	-0.5003	2.5318	
TL2F	0.000000	-0.5003	2.5318	
TL2W	0.000000	-0.5003	2.5318	
TL2Y	0.000000	-0.5003	2.5318	
PNC	0.000000	-0.6000	2.1650	! Same as NC2
PS	0.000000	-0.5545	2.6010	! Same as C3M
COH1	0.000000	-0.7152	2.0909	
CCOH	0.000000	-0.7999	2.3741	
TOL4	0.000000	-0.5839	2.6920	
PL1	0.000000	-1.0688	1.8769	
COH3	0.000000	-0.7359	2.0730	
BZY	0.000000	-0.3205	2.2332	! from reference [4]
C3E2	0.000000	-0.4599	2.7807	
C2E	0.000000	-0.3710	2.4343	! from reference [5]
C2M	0.000000	-0.3420	2.4343	! from reference [5]
C3M	0.000000	-0.5545	2.6010	! from reference [5]
C3E	0.000000	-0.5927	2.6010	! from reference [5]

NBFIX

! Bead Types		Emin (kcal/mol)	Rmin (Å)
! C2E	H2O1	-0.5130	4.2359
! C2M	H2O1	-0.4400	4.2359
! C3E	H2O1	-0.6202	4.5000
! C3M	H2O1	-0.5434	4.5000

END

References:

(1) Eberhart, R.; Kennedy, J. A New Optimizer Using Particle Swarm Theory. In *Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on*; 1995; pp 39–43.

(2) Bejagam, K. K.; Singh, S.; An, Y.; Deshmukh, S. A. Machine-Learned Coarse-Grained Models. *J. Phys. Chem. Lett.* **2018**, 4667–4672.

(3) An, Y.; Singh, S.; Bejagam, K. K.; Deshmukh, S. A. Development of an Accurate Coarse-Grained Model of Poly(acrylic Acid) in Explicit Solvents. *Macromolecules* **2019**, *52* (13), 4875–4887.

(4) Bejagam, K. K.; Singh, S.; An, Y.; Berry, C.; Deshmukh, S. A. PSO-Assisted Development of New Transferable Coarse-Grained Water Models. *J. Phys. Chem. B* **2018**, *122*, 1958–1971.

(5) An, Y.; Bejagam, K. K.; Deshmukh, S. A. Development of New Transferable Coarse-Grained Models of Hydrocarbons. *J. Phys. Chem. B* **2018**, *122* (28), 7143–7153.