Supplementary Information (SI) for Molecular Systems Design & Engineering.
This journal is © The Royal Society of Chemistry 2026

**Supplementary Information for**

# Self-assembly of Phenylalanine Oligopeptides: Development of Transferable Bottom-up Coarse-grained Potentials

Mason Hooten[1] and Meenakshi Dutt**[,2]

[1]*Department of Biomedical Engineering, Rutgers The State University of New Jersey, Piscataway, NJ 08854*

[2]*Department of Chemical and Biochemical Engineering, Rutgers The State University of New Jersey, Piscataway, NJ 08854*

** corresponding author: meenakshi.dutt@rutgers.edu

# Naming Conventions

## F2 / FF / F4 / FFFF

The models developed in this manuscript use the three-bead sidechain representation of the AR3 model described in the preceding paper. Throughout the manuscript and this document, FF and FFFF refer to peptide sequences, while F2 and F4 refer to the CG models of the respective peptides.

## Table S1. Bead names and descriptions

CG-modeled peptides contain beads of 7 bead types. Hydrogens are included in mappings with their associated heavy atoms.

| Bead Type | Represents |
|---|---|
| PHA | Beta carbon and 2 Ring carbons |
| PHB | 2 Ring carbons |
| PHC | 2 Ring carbons |
| CA | Alpha carbon |

| | |
|---|---|
| AMD | Amide group |
| NH3 | N terminal |
| COO | C terminal |

# Overview of Calculated Structural Distributions

In this section, we present structural calculations based on CG simulation and compare with the AA reference. AA reference trajectory data are typically mapped to the appropriate CG representation to facilitate comparison. Figures in the Multiple Peptides sections below are usually ordered by system size.

## F2 distributions

Figures S1-S4 show the conformational distributions for a single F2 peptide in water.
Figure S5 shows the interbead RDFs in a F2 system containing 25 peptides.
Figure S6 shows the distribution of Rg in individual peptides in a F2 system containing 25 peptides.

## F4 distributions

Figures S1-S4 show the conformational distributions for a single F2 peptide in water.
Figure S5 shows the interbead RDFs in a F2 system containing 25 peptides.
Figure S6 shows the distribution of Rg in individual peptides in a F2 system containing 25 peptides.

## Interpreting F2 conformational distributions

The values of the bonded DOFs simulated with the F2 model suggest good correspondence with representative measurements from the reference AA model, as shown in Supporting Information Figure S1. All bonds are in nearly perfect agreement with the reference, maintaining the peak locations and relative magnitudes of all distributions. All of the angles also sample the same ranges as the atomistic model. Three of the angles notably oversample specific modes relative to the atomistic reference, all of which are associated with the extension of side chain rings away from the backbone. Such sampling bias is a known phenomenon in bottom-up modeling, where the factors contributing to the undersampled modes may be inaccessible to the CG model. This may be due to biases in the CG potential energy surface resulting from the development process of the potential energy surface. It may also result from uncalibrated steric interference owing to the virtual size of the CG beads.

All parameterized dihedrals roughly sample the exact peak locations as the reference AA distributions, again with some observable bias in the configurations. Similar to the angles,

the most considerable divergences from the shape of the atomistic reference distributions are seen in DOFs associated with the orientation of side chains relative to the backbone. It is noted that all dihedrals incident on the side chain beads parameterize the orientation of the side chain. It is likely that a more judicious parameterization – such as directly specifying the COM position and the planar orientation of each ring – may reduce the model's complexity, thereby improving both computational performance and the accuracy of simulated orientations. Alternatively, developing dihedral potentials with a particular functional form may produce smoother functions, improving computational performance.

Backbone end-to-end distance of a single peptide agrees well with corresponding measurements from the atomistic reference, indicating a fairly rigid peptide backbone in both cases, as shown in Supporting Information Figures S3 and S4. The Rg is measured to confirm the overall shape of a peptide. The AA distribution of Rg shows two modes, one at 0.375 nm and one at 0.44 nm. The CG distribution overlaps substantially with the reference, but with only a single mode at 0.365 nm. The CG curve also achieves a lower minimum of approximately 0.29 nm, compared with the 0.30 nm observed in the AA curve. The slightly lower mass distribution on a correctly conformed backbone is explained by the shift of the first side chain away from the N-terminal and toward the amide bond at the center of the peptide, as expressed in the distributions of angles NH3-CA1-PHA1 and AMD1-CA1-PHA1, which describe the extension of the side chain, shown in Supporting Information Figure S1. It is noted that although these angles display bias relative to the reference, they nevertheless sample values and peak locations that match their AA counterparts and, thus, are not considered unphysical.

## Interpreting F4 conformational distributions

We refer to the main manuscript for additional information about the DOF results in the F4 model. Here, we interpret the dihedrals impacting the side chain presentation.

a) Backbone Twist

The dihedral angle distributions measured using the F4 model show several significant deviations affecting the twist of the peptide backbone and the positioning of the side chains relative to the main chain. Generally, agreement between the CG and AA dihedral distributions is not as smooth as in the other DOFs. This is mainly because refinement of the dihedrals is conditional on all of the different parameters in the model. Thus, the dihedrals must cope with the cumulation of concurrent constraints. With this in mind, the quality of CG dihedrals can be afforded some leniency, given that they sample the same ranges as the AA reference. As a result, most of the dihedrals sample correctly from the same conformation observed in the reference AA trajectories, with sampled regions roughly matching the peak locations of the reference distributions.

The two main chain dihedrals terminating on the third backbone alpha carbon bead (AMD1-CA2-AMD2-CA3 and CA3-AMD3-CA4-COO) show complementary deviations from their respective AA reference curves, each sampling a range of angles about 180 degrees out of phase with reference but in opposite directions (see Supporting Information Figure S7). Of these two, only the first inherited its initial parameterization from the AR3 model. It is interesting to note that the atomistic reference from FFF used in that model samples nearly the same region as the refined dihedral does here, although the fully refined AR3 model does not. Thus, for the

first dihedral, this is another case of initial parameterizations showing an outsized influence even after multiple rounds of IBI. The second dihedral is noted as an incident on the C-terminal bead, one of the two locations in the model subject to the fewest bonded potentials.

The remaining four main chain dihedrals, i.e., those not incident on any side chain bead, all agree well with the reference. If the two dihedral angles noted showing deviations were sampled simultaneously in a single realized conformation, it could be interpreted as a twisting of the third alpha carbon away from its correct backbone presentation. This twist would thus turn the 130-degree angle centered on that alpha carbon (AMD2-CA3-AMD3) opposite its appropriate direction, implying a linear displacement of the alpha carbon bead of about 2 Angstrom (2 Ang bond length * cos(130/2) * 2).

b) Side Chain Orientation

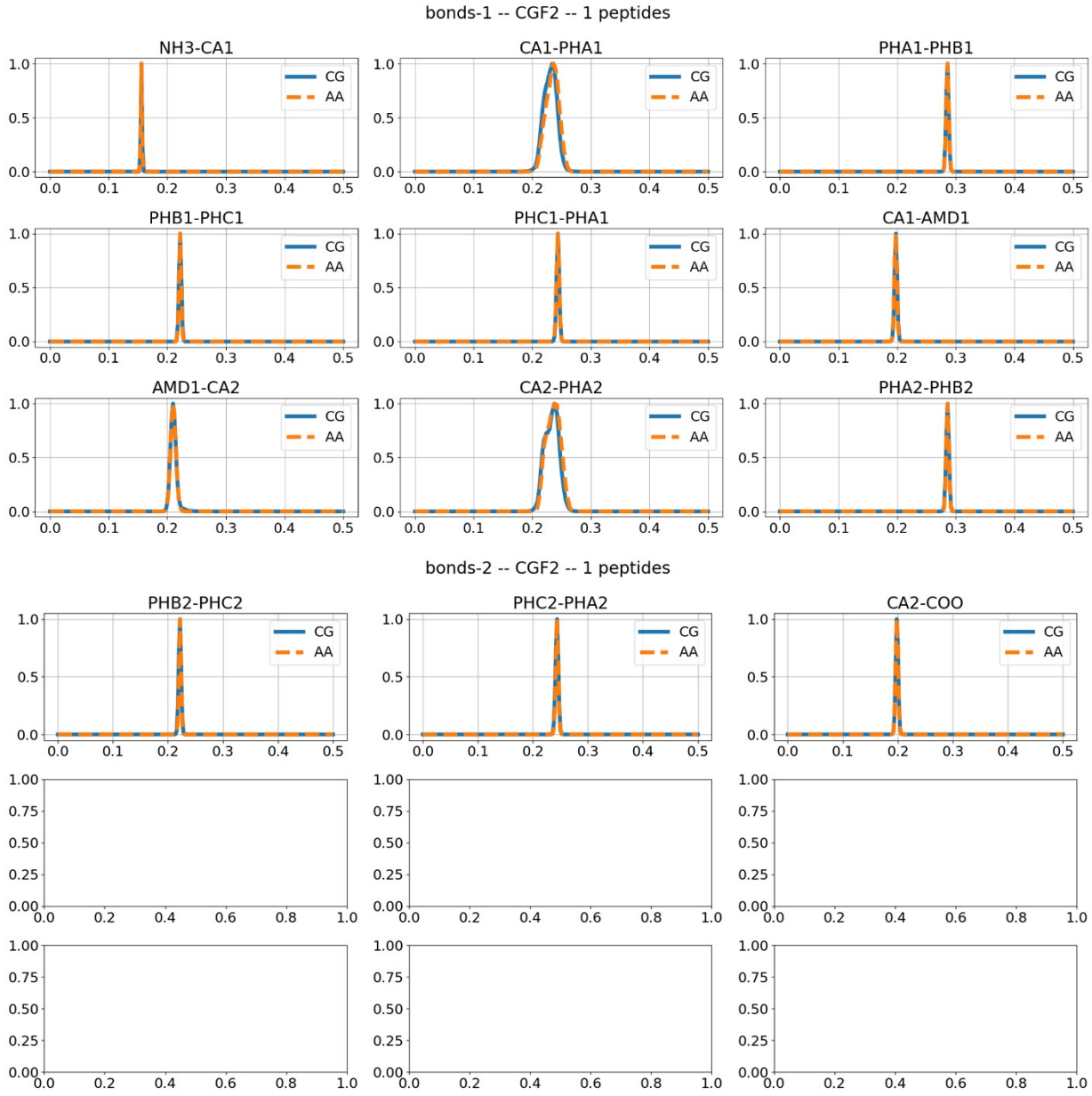
The distance by which side chains extend away from the main chain is mainly a function of the bond length between the alpha carbon bead and the innermost side chain bead. The angles incident on the innermost side chain bead and the main chain modulate this distance. The orientation of each side chain is a function of several parameterized dihedrals.

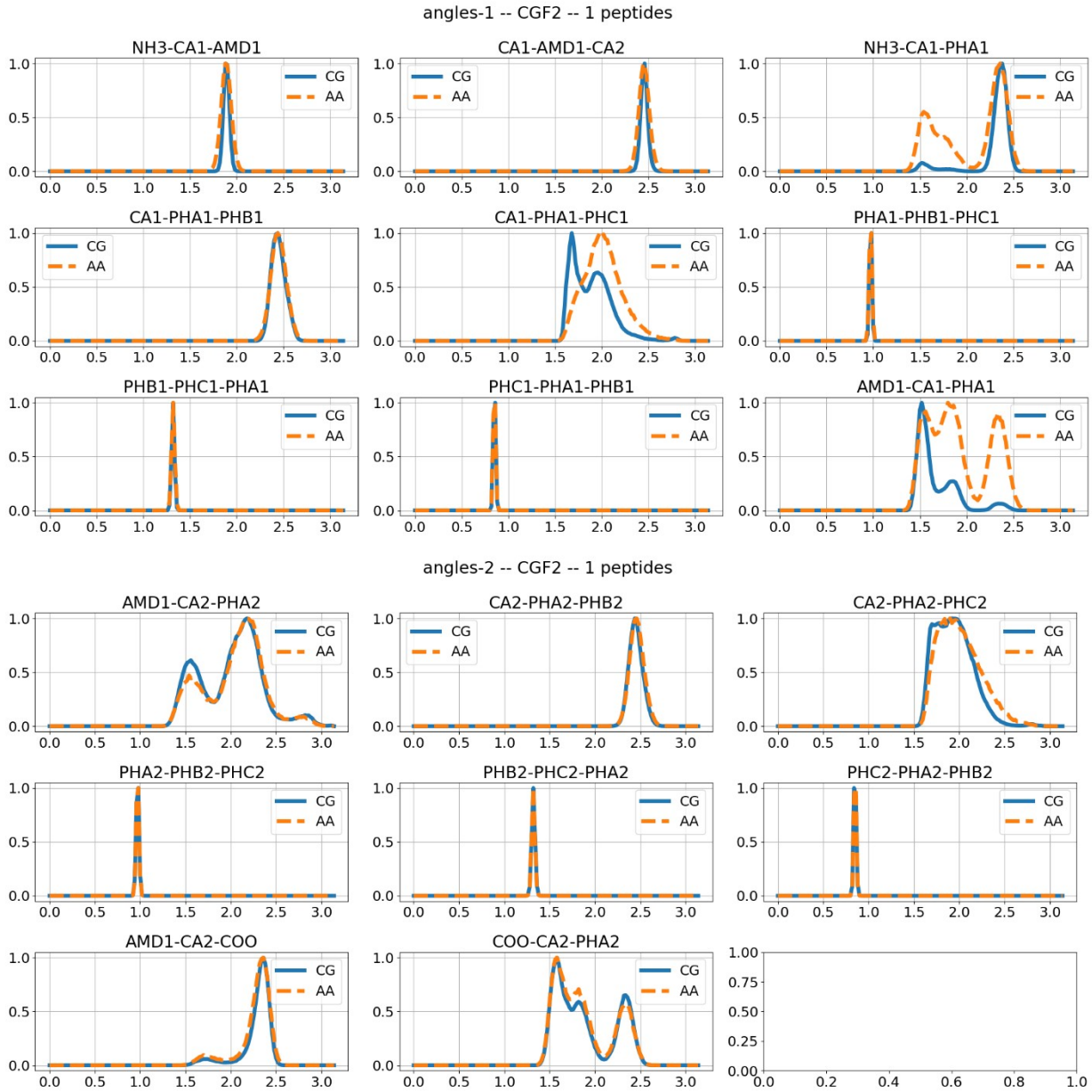
Eight dihedrals govern the orientations of the two interior side chains. In contrast, the side chains adjacent to the N- and C-termini are governed by seven dihedrals (partly described in the schematic in Figure 2). Typically, one or two dihedrals extending from the innermost side chain bead (type PHA) up or down the central chain control the rotameric character of the side chain presentation, the key feature in side chain isomerization. Another six dihedrals, terminating on the outer side chain beads (types PHB and PHC), and extending toward and onto the primary chain, control the latitude of the side chain ring or the direction it faces.

The second side chain shows a deviation in its rotamerization dihedral extending toward the N-terminal (CA1-AMD1-CA2-PHA2), where it is persistently about 90 degrees off the orientation seen in the atomistic reference, with a slightly bifurcated distribution. The complementary rotameric dihedral extending in the C-terminal direction (CA3-AMD2-CA2-PHA2) does not show the same deviation. Instead, it samples the same range as its atomistic reference, but develops a bimodal distribution with a trough at the location of the single peak of the AA reference. The bimodality of the two distributions, if not the DOF's mislocation, may be explained by the steric interference of adjacent side chains. The radial distribution functions (RDFs) among the side chain beads are shown in Supporting Information Figure S11, where the CG model demonstrates considerably more proximity among side chains than the AA reference.

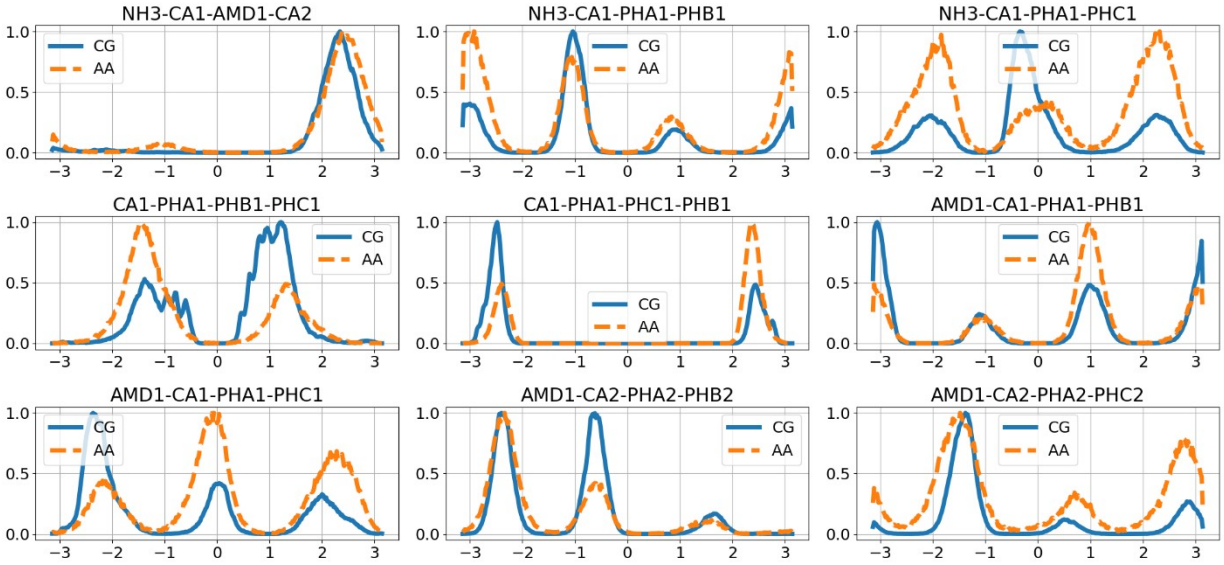
Deviations in latitude are observed for the first side chain, affecting rotation about the bond between the alpha carbon and the inner side-chain bead, as well as out-of-plane tilting of the ring. All four angles inherit their initial parameterization from the AR3 model. Examining the AA reference curves from that study,[34] it is again seen that the AA reference differs between FFF and FFFF, and that the values simulated by the F4 model would be a perfect match for the corresponding results from the AA FFF model.

# Figure S1. F2 Bonded DOF distributions – single peptide




angles-1 -- CGF2 -- 1 peptides
NH3-CA1-AMD1
CA1-AMD1-CA2
NH3-CA1-PHA1
CA1-PHA1-PHB1
CA1-PHA1-PHC1
PHA1-PHB1-PHC1
PHB1-PHC1-PHA1
PHC1-PHA1-PHB1
AMD1-CA1-PHA1
angles-2 -- CGF2 -- 1 peptides
AMD1-CA2-PHA2
CA2-PHA2-PHB2
CA2-PHA2-PHC2
PHA2-PHB2-PHC2
PHB2-PHC2-PHA2
PHC2-PHA2-PHB2
AMD1-CA2-COO
COO-CA2-PHA2
CG
AA

dihedrals-1 -- CGF2 -- 1 peptides




dihedrals-2 -- CGF2 -- 1 peptides

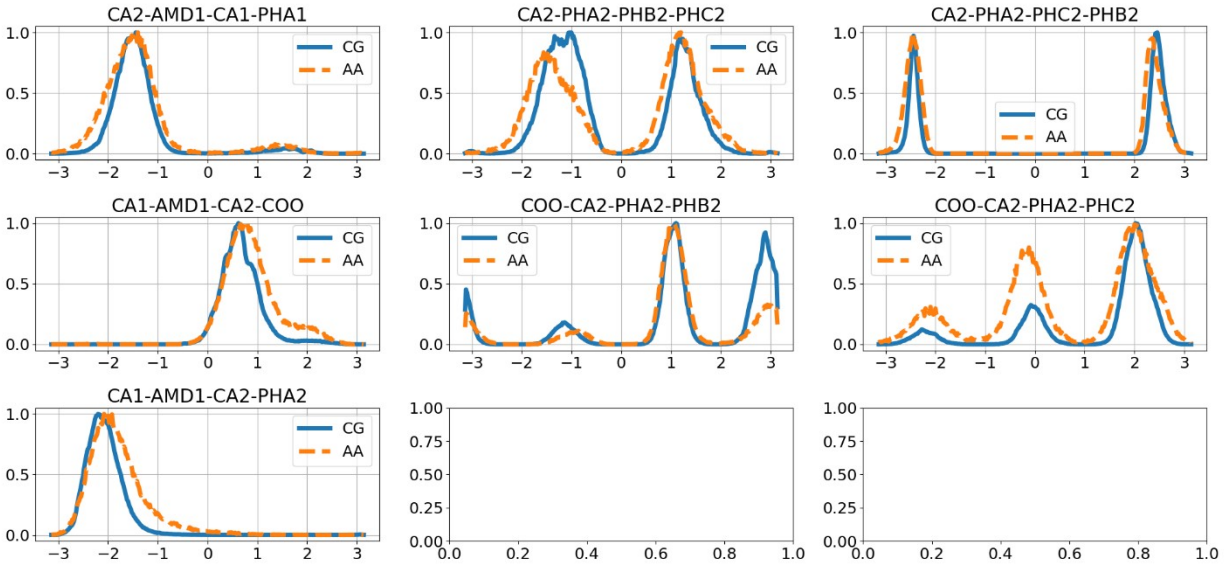


Figure S2. F2 Solvation RDFs – single peptide

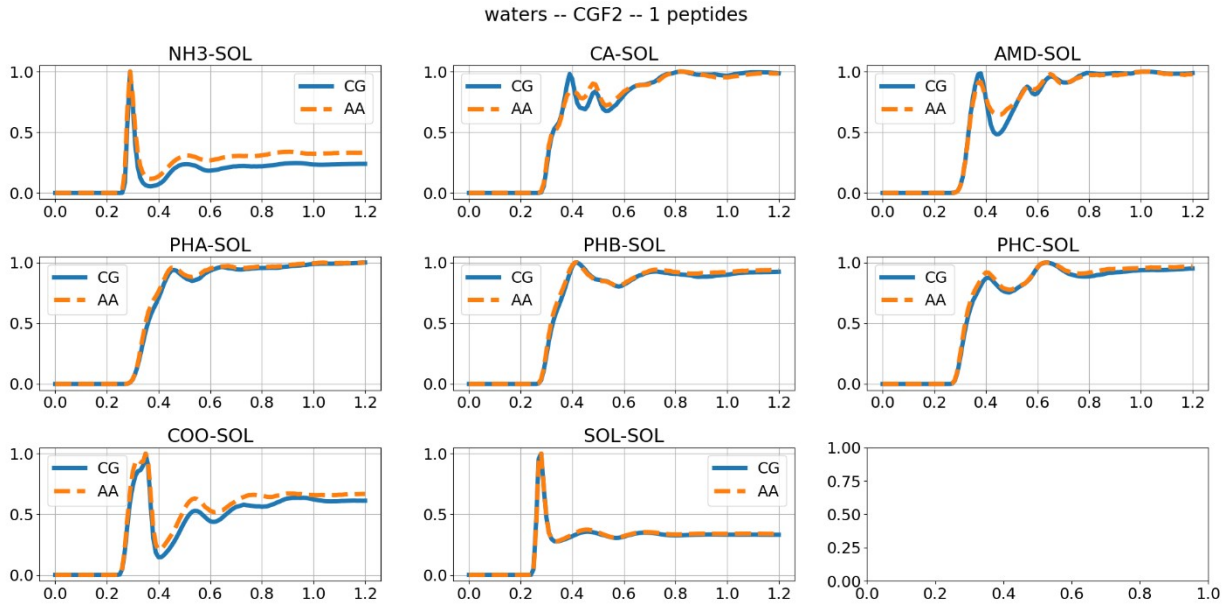



Figure S3. F2 End-to-end distance distributions – single peptide

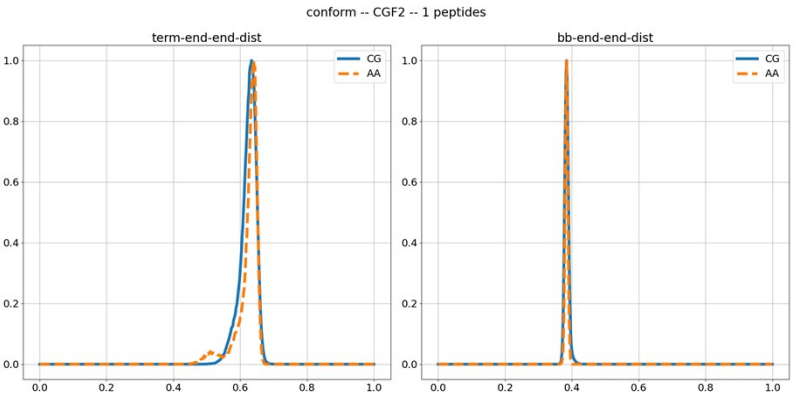
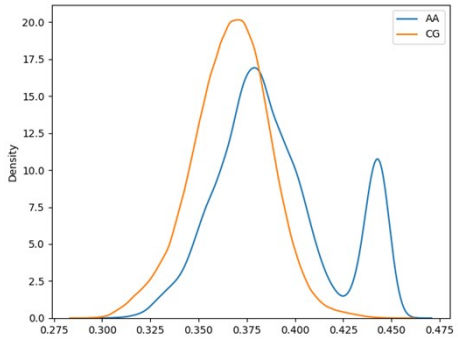
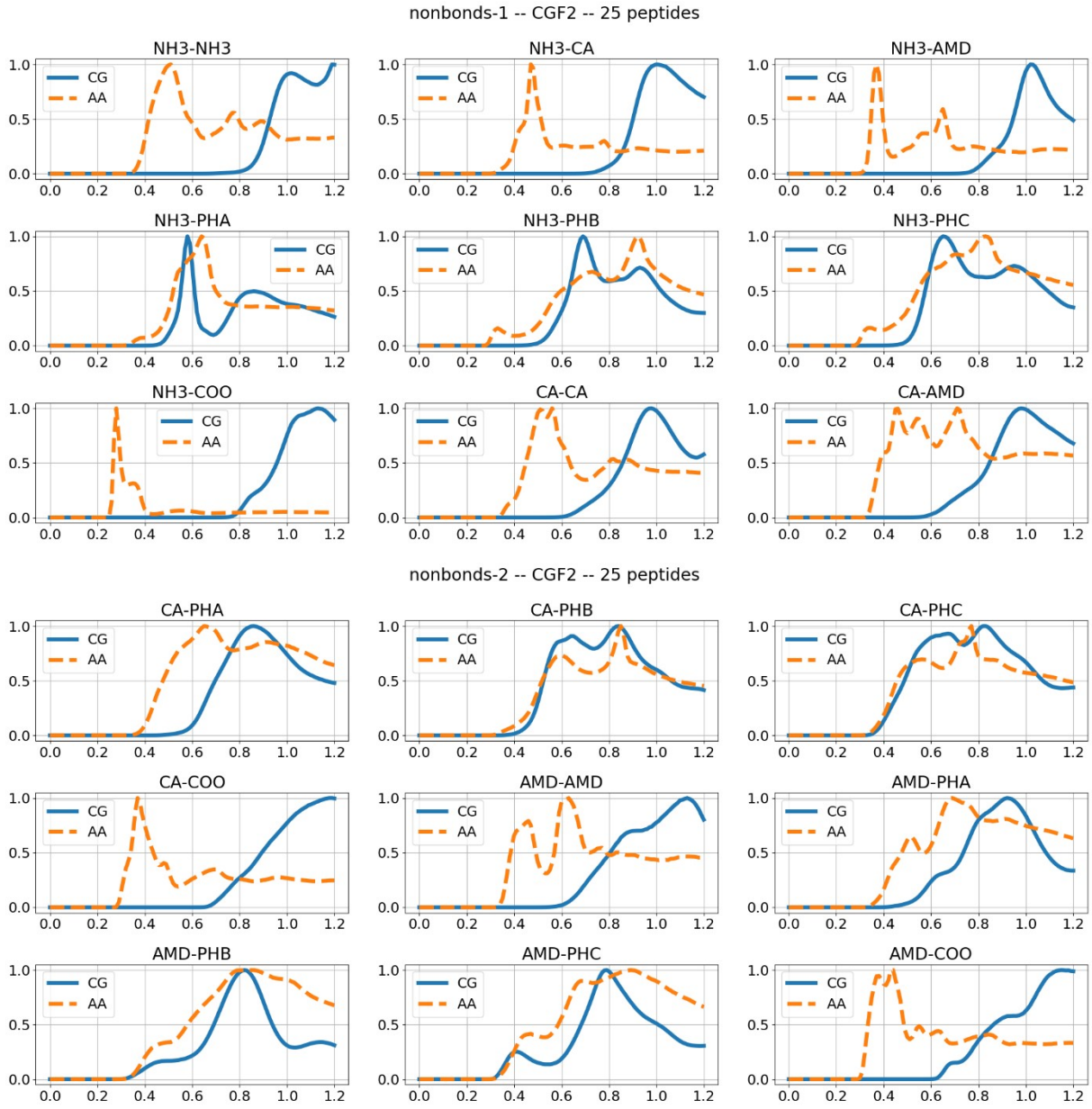


Figure S4. F2 Single-peptide Radius of Gyration



# Figure S5. F2 Bead RDFs – 25 peptides



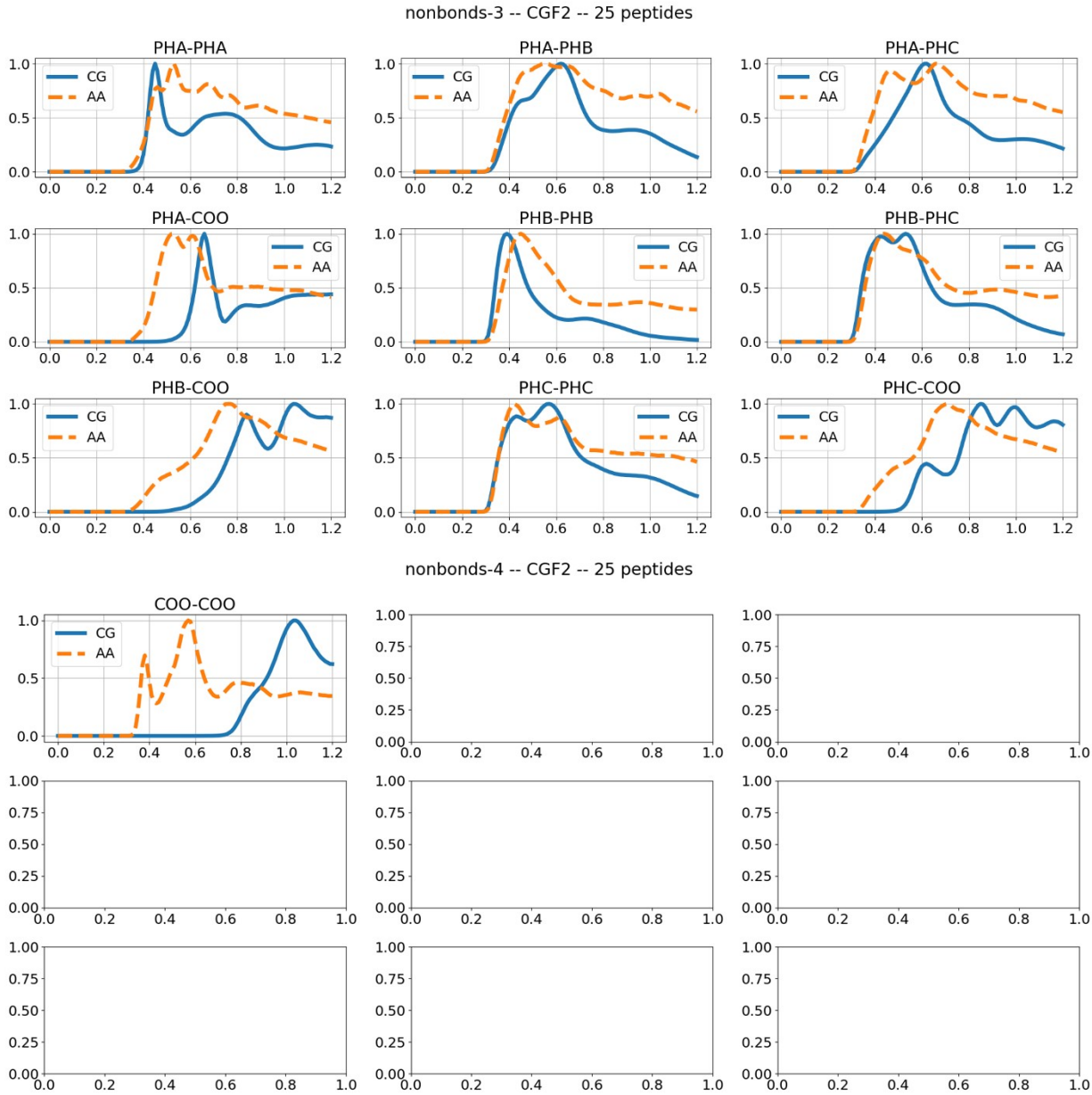

nonbonds-3 -- CGF2 -- 25 peptides
PHA-PHA
PHA-PHB
PHA-PHC
PHA-COO
PHB-PHB
PHB-PHC
PHB-COO
PHC-PHC
PHC-COO
CG
AA
nonbonds-4 -- CGF2 -- 25 peptides
COO-COO

Figure S6. F2 End-to-end distance distributions – 25 peptides

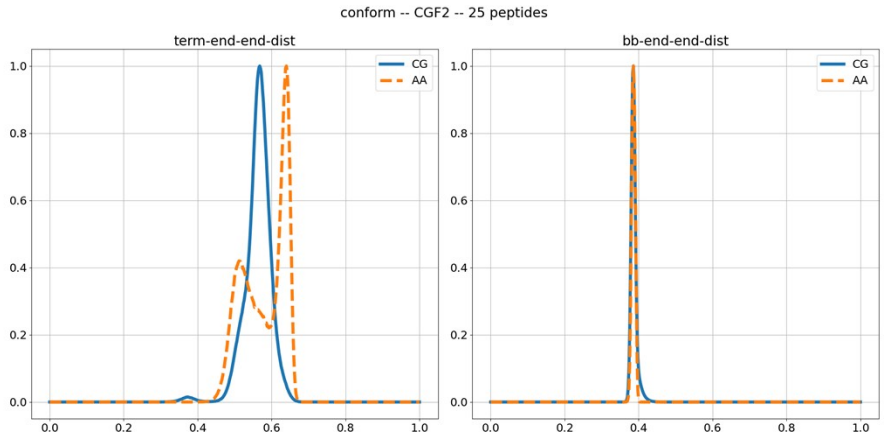
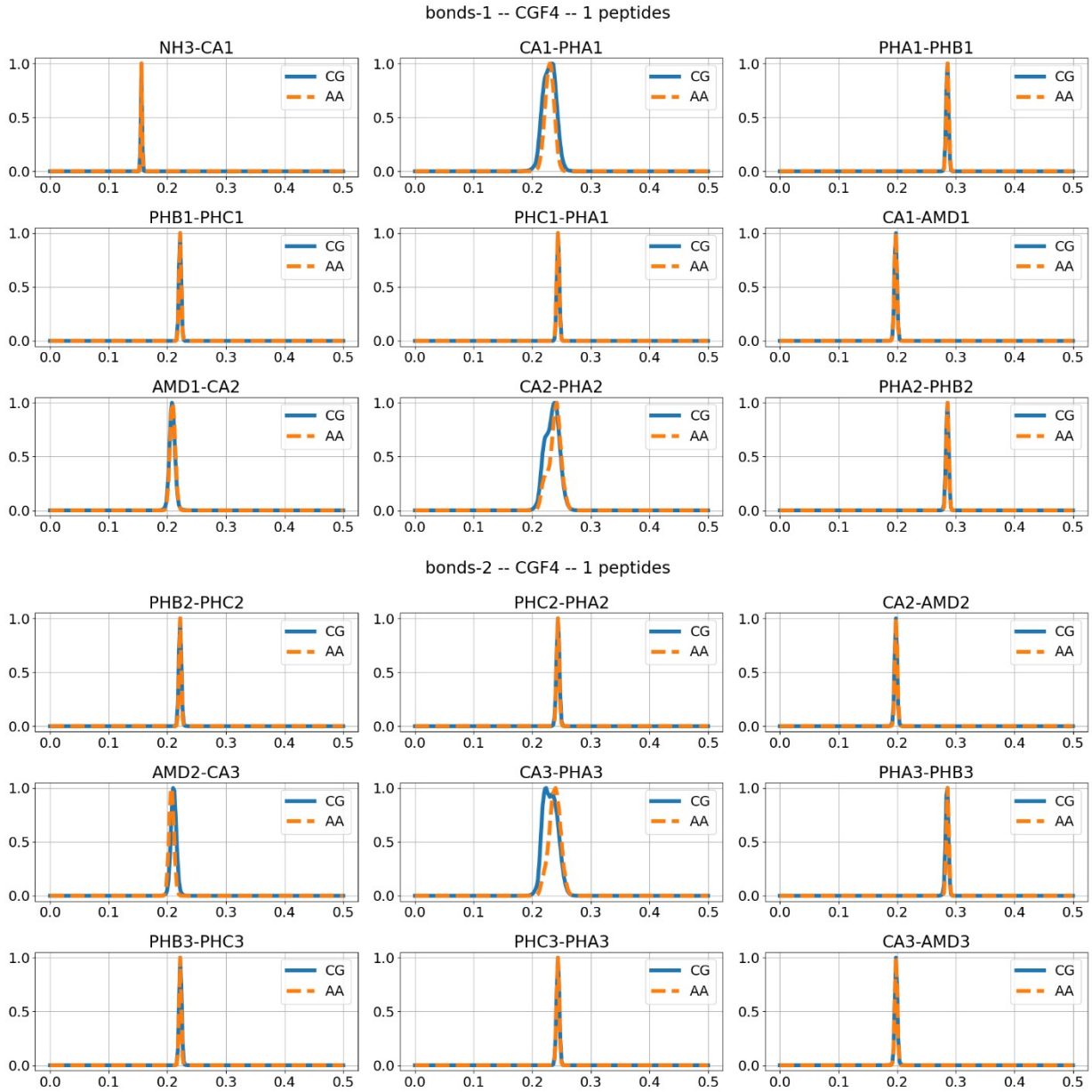
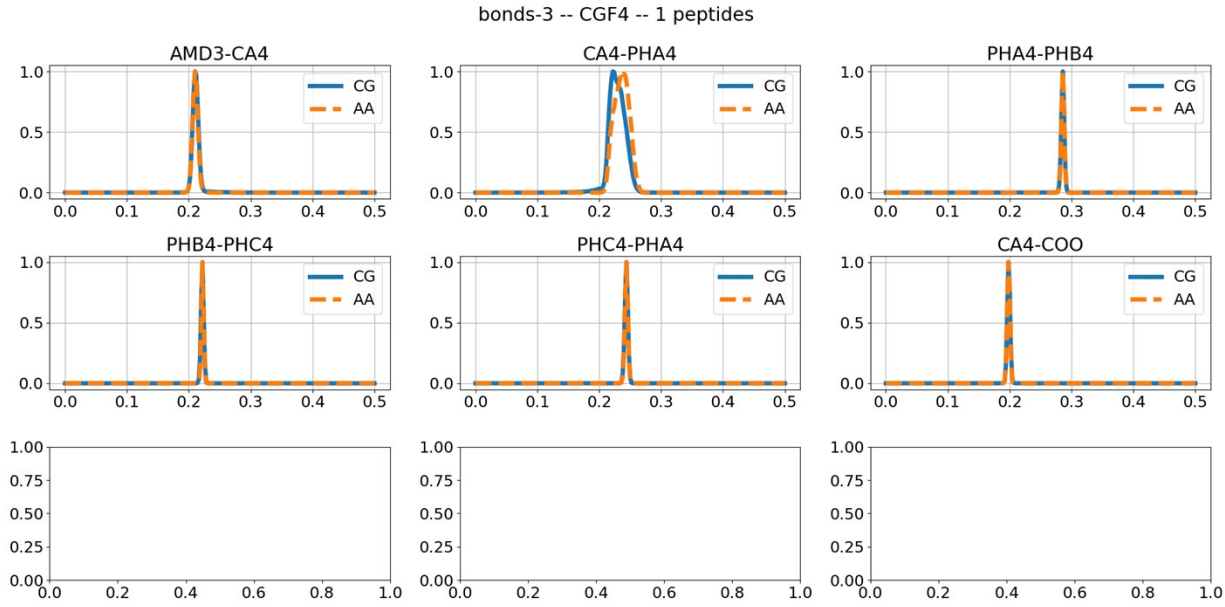
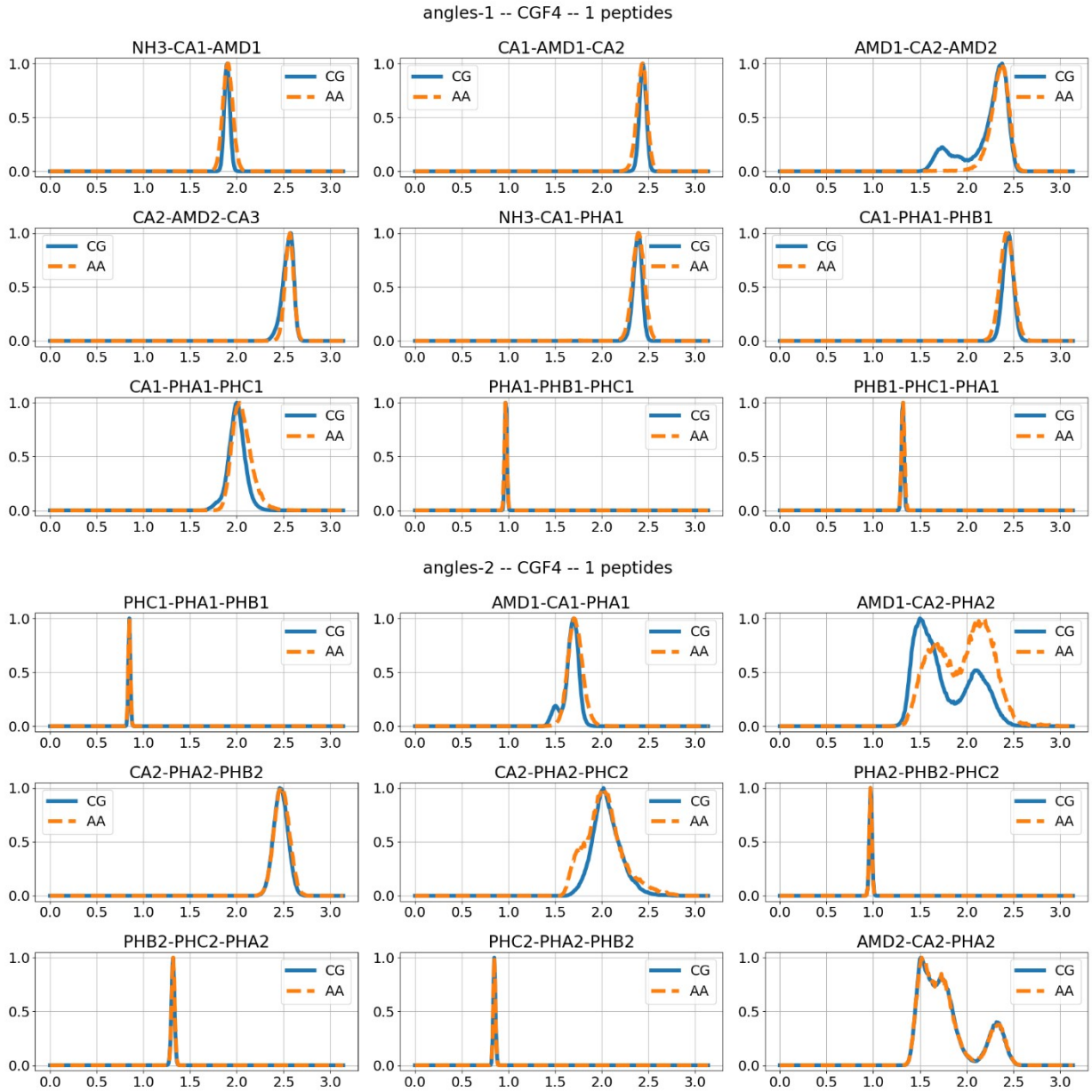


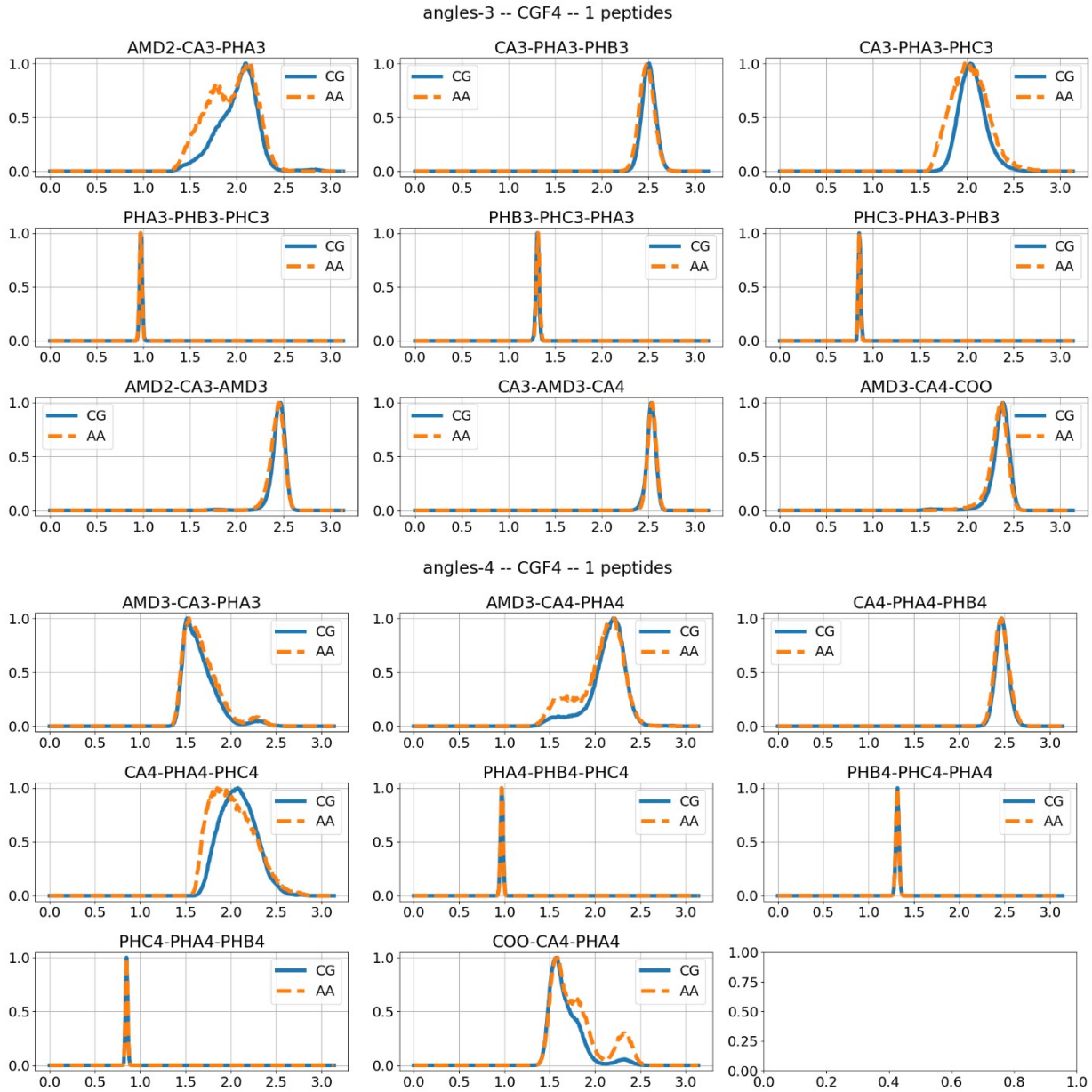
Figure S7. F4 Bonded DOF distributions – single peptide


bonds-1 -- CGF4 -- 1 peptides
NH3-CA1
CA1-PHA1
PHA1-PHB1
PHB1-PHC1
PHC1-PHA1
CA1-AMD1
AMD1-CA2
CA2-PHA2
PHA2-PHB2
bonds-2 -- CGF4 -- 1 peptides
PHB2-PHC2
PHC2-PHA2
CA2-AMD2
AMD2-CA3
CA3-PHA3
PHA3-PHB3
PHB3-PHC3
PHC3-PHA3
CA3-AMD3
CG
AA

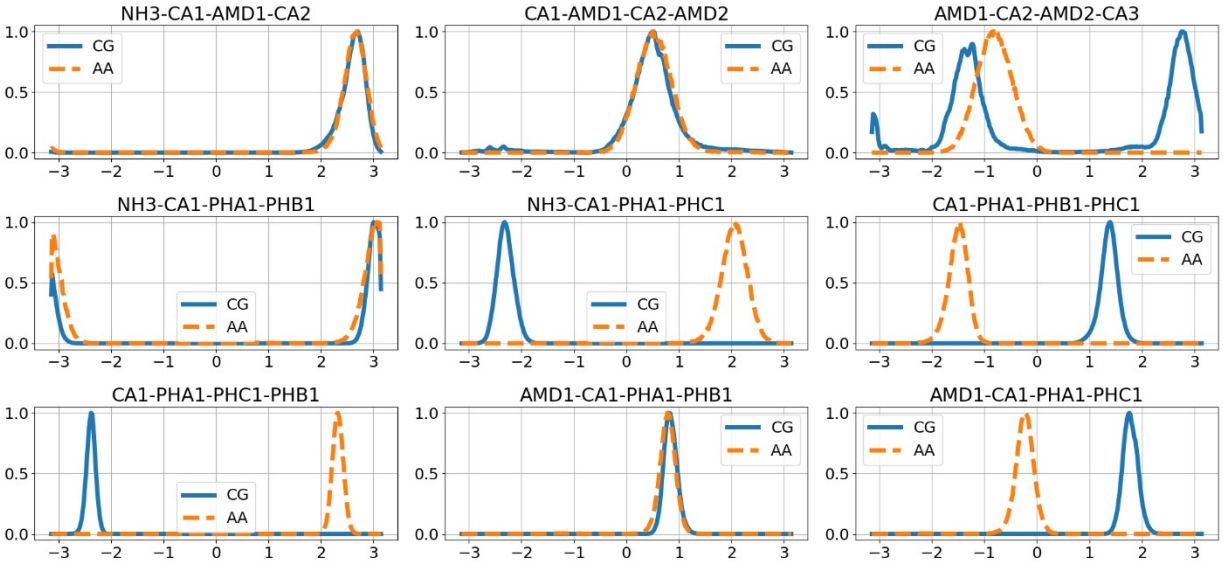

bonds-3 -- CGF4 -- 1 peptides
AMD3-CA4
CA4-PHA4
PHA4-PHB4
PHB4-PHC4
PHC4-PHA4
CA4-COO
CG
AA


angles-1 -- CGF4 -- 1 peptides
NH3-CA1-AMD1
CA1-AMD1-CA2
AMD1-CA2-AMD2
CA2-AMD2-CA3
NH3-CA1-PHA1
CA1-PHA1-PHB1
CA1-PHA1-PHC1
PHA1-PHB1-PHC1
PHB1-PHC1-PHA1
CG
AA

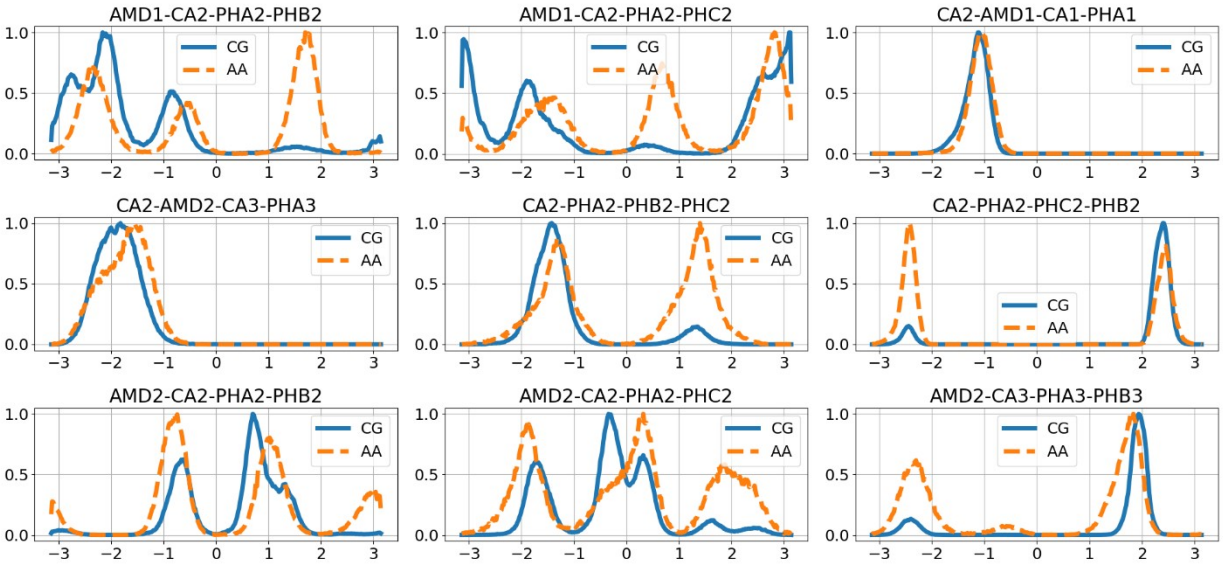

angles-2 -- CGF4 -- 1 peptides
PHC1-PHA1-PHB1
AMD1-CA1-PHA1
AMD1-CA2-PHA2
CA2-PHA2-PHB2
CA2-PHA2-PHC2
PHA2-PHB2-PHC2
PHB2-PHC2-PHA2
PHC2-PHA2-PHB2
AMD2-CA2-PHA2
CG
AA


angles-3 -- CGF4 -- 1 peptides
AMD2-CA3-PHA3
CA3-PHA3-PHB3
CA3-PHA3-PHC3
PHA3-PHB3-PHC3
PHB3-PHC3-PHA3
PHC3-PHA3-PHB3
AMD2-CA3-AMD3
CA3-AMD3-CA4
AMD3-CA4-COO
angles-4 -- CGF4 -- 1 peptides
AMD3-CA3-PHA3
AMD3-CA4-PHA4
CA4-PHA4-PHB4
CA4-PHA4-PHC4
PHA4-PHB4-PHC4
PHB4-PHC4-PHA4
PHC4-PHA4-PHB4
COO-CA4-PHA4
CG
AA

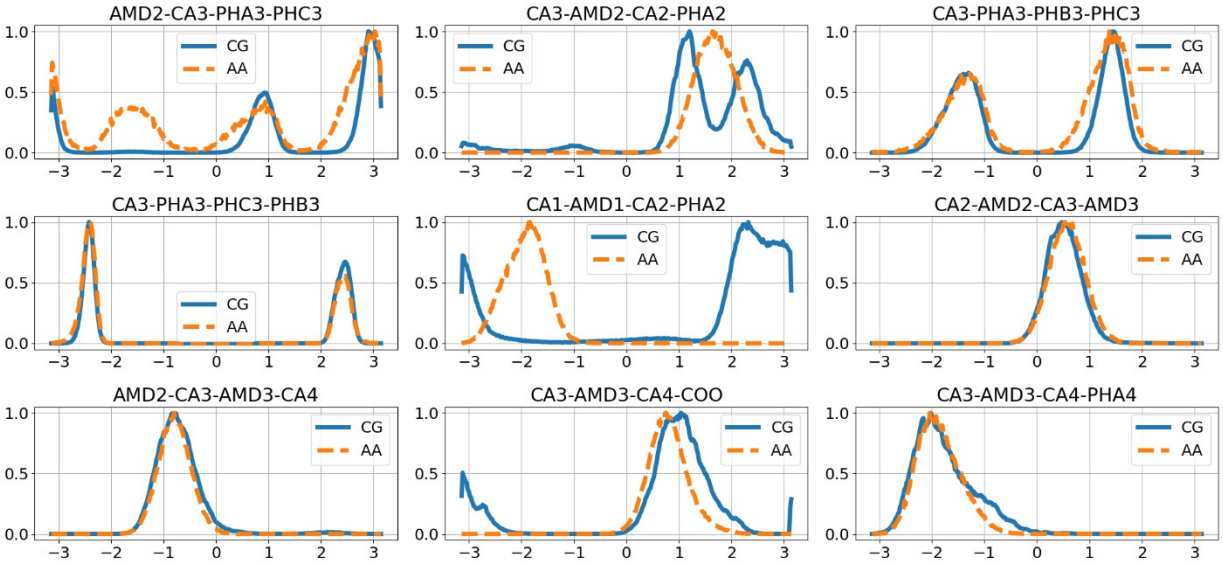
dihedrals-1 -- CGF4 -- 1 peptides




dihedrals-2 -- CGF4 -- 1 peptides



dihedrals-3 -- CGF4 -- 1 peptides




dihedrals-4 -- CGF4 -- 1 peptides

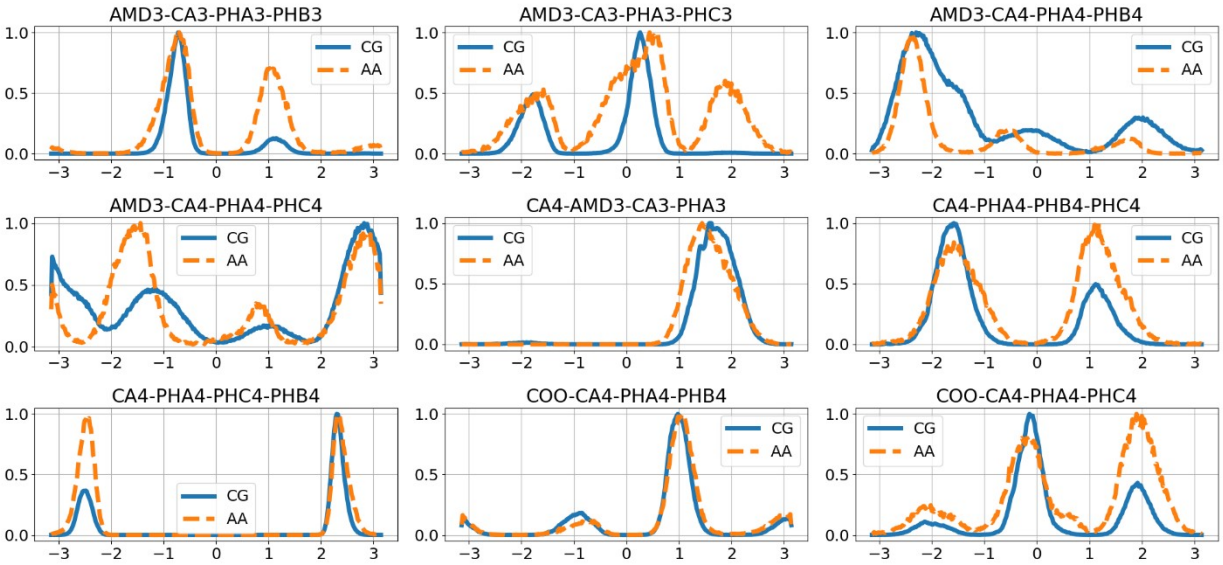


Figure S8. F4 Solvation RDFs – single peptide

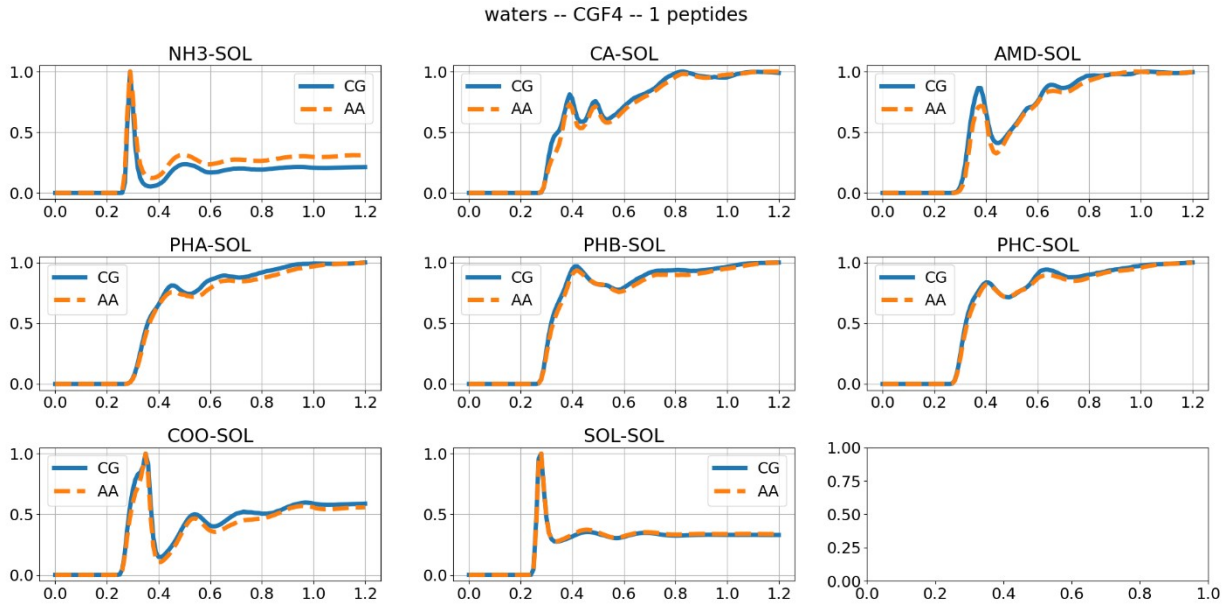



Figure S9. F4 End-to-end distance distributions – single peptide

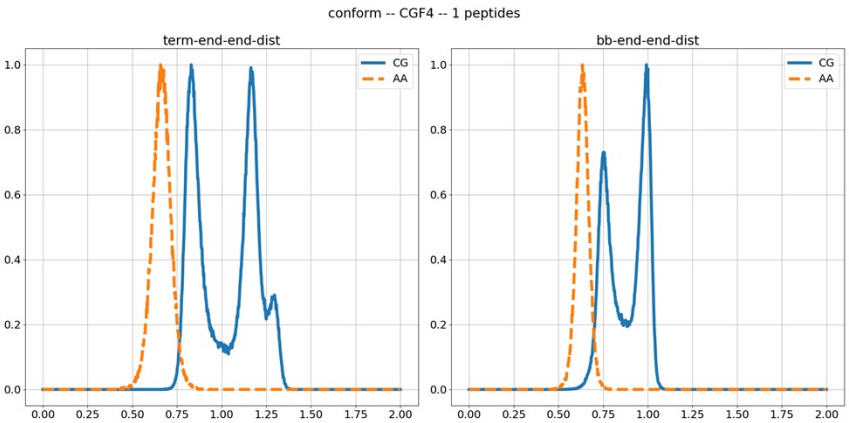
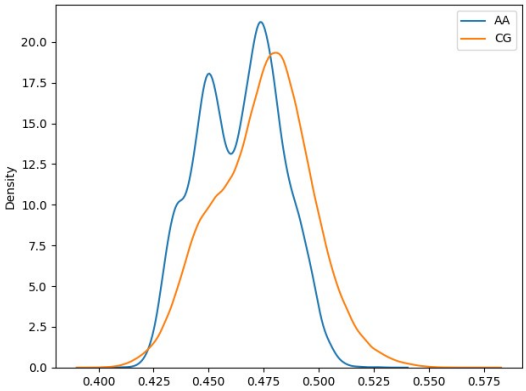
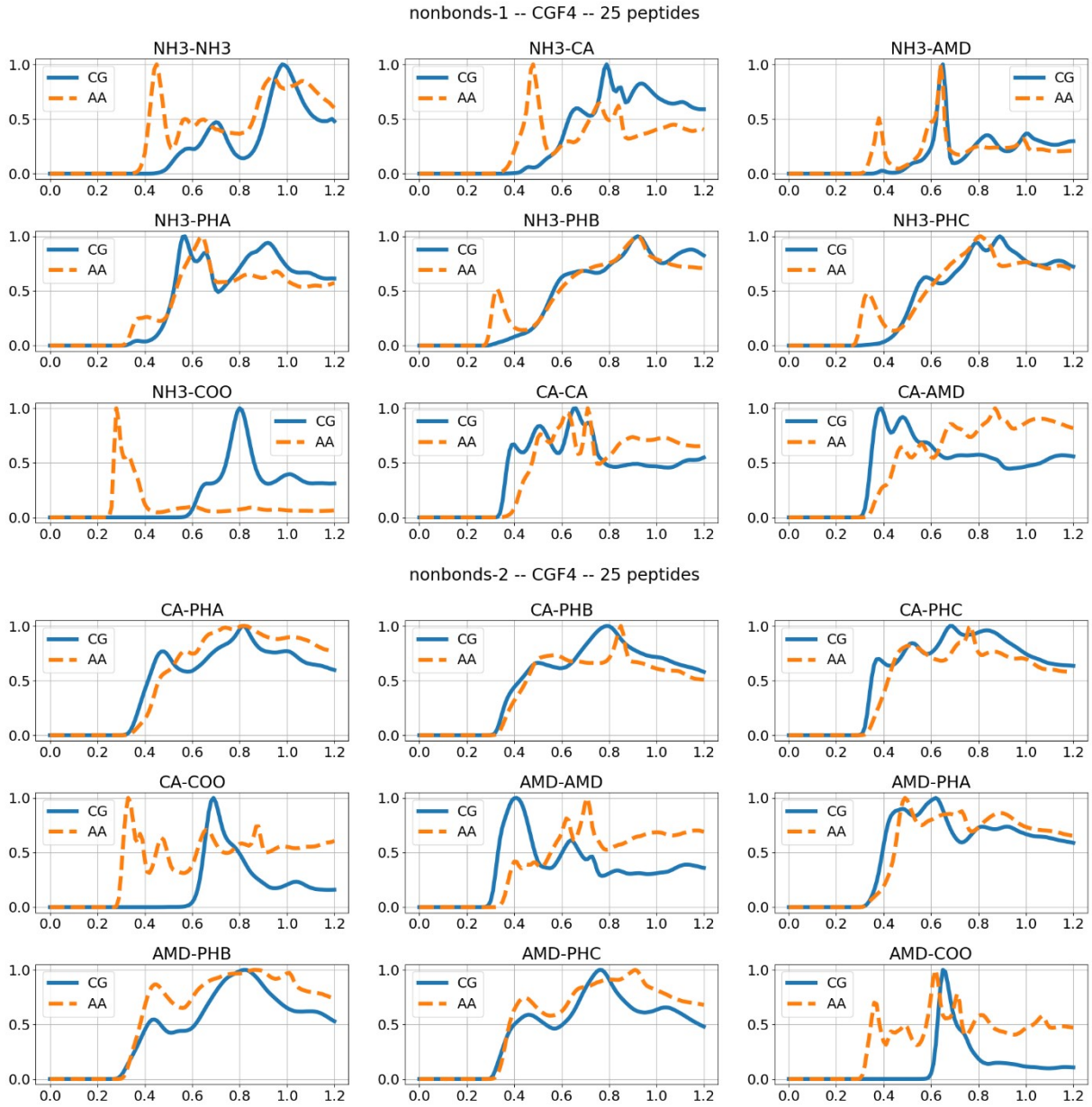


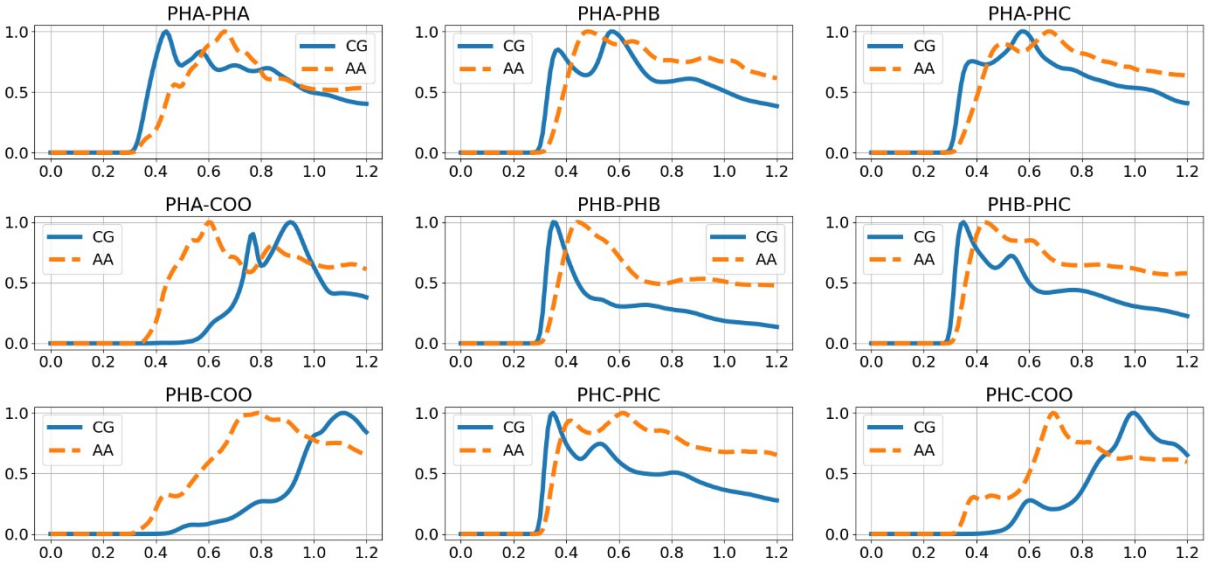
Figure S10. F4 Single-peptide Radius of Gyration



# Figure S11. F4 Bead RDFs – 25 peptides



nonbonds-3 -- CGF4 -- 25 peptides




nonbonds-4 -- CGF4 -- 25 peptides

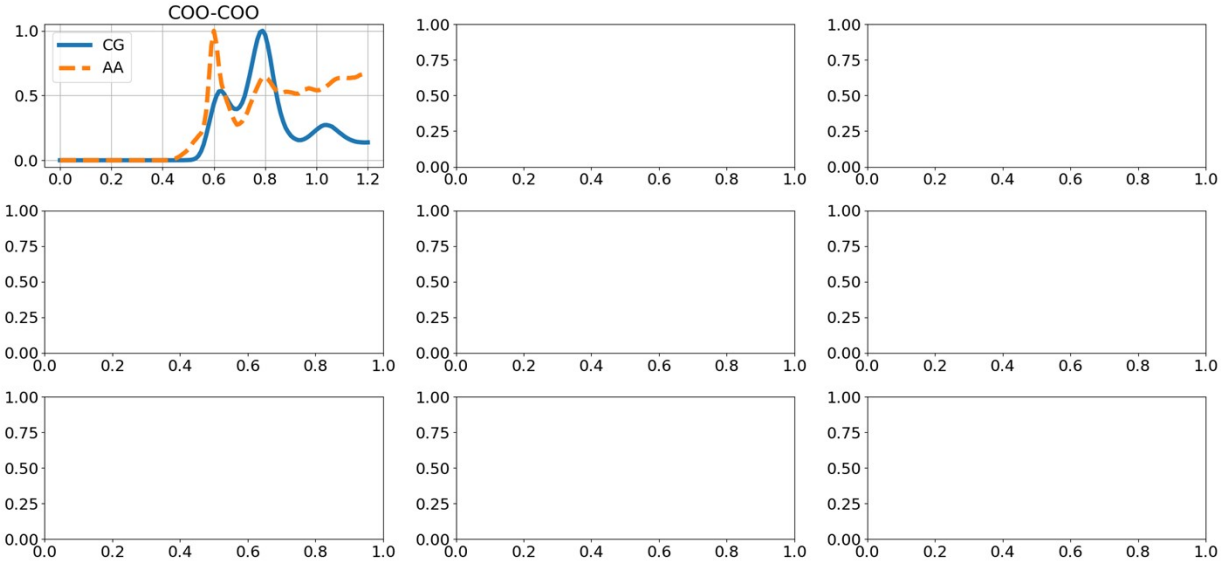
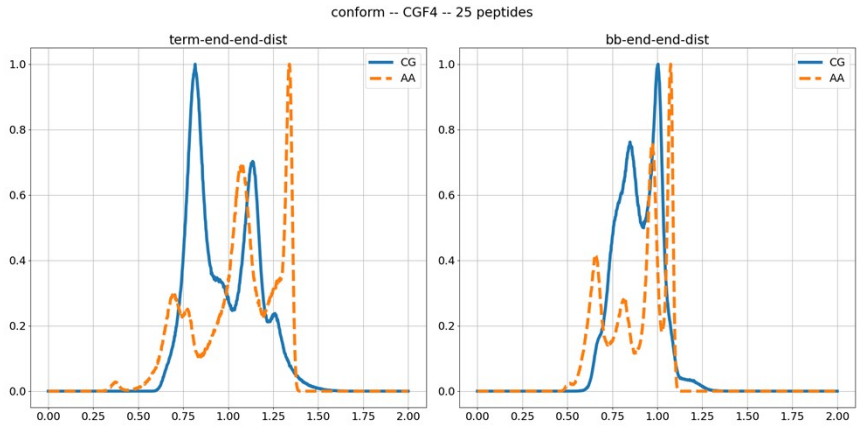


Figure S12. F4 End-to-end distance distributions – 25 peptides




# Comparing Performance between Bottom-Up Models and Martini

Earlier investigations[Banerjee 2022; Hooten 2024] into the performance of bottom-up CG models like those developed in this study have shown that such models effectively accelerate the dynamics of the underlying force fields and that the models exhibit strong scaling (namely, simulation performance falls in proportion to computing allocation) over a small number of cores. These conclusions support the feasibility of tuning the models for their application to larger chemical systems and to run on larger compute allocations. In this study, the practical performance of the F2 and F4 models in production simulations is compared with that of the Martini production simulations for FF and FFFF. A sample of production performance is summarized in Figure 16, where performance values are taken from the final stage of production MD simulations.

In bottom-up models, increasing production performance through judicious domain decomposition is possible, allowing production simulations to be run on allocations of up to 32 cores. However, as aggregates become densely packed, simulations can become unstable due to the many overlapping bonded structures and the rough potential energy surface of each. As such, extended simulation durations may be achieved by reducing the allocation to a small number of cores. However, by the time this is required, typically stable nanostructures are already formed as the system size increases from 32 to 64 peptides; the performance per core declines, although doubling again to 128 peptides does not result in a notable deterioration of performance.

The Martini force field has been optimized for high performance per core. All Martini systems were simulated on allocations of 16 cores, which also reflects an important practical consideration: modestly sized, microsecond-scale systems can be simulated using the Martini force field on a personal computer in reasonable durations of wall clock time. The increased number of particles in the FFFF model contributes to an efficiency gain, doubling or tripling the

simulation speed over that of the FF model. Performance per core varies more in the Martini simulations than in the bottom-up models, an effect visibly accentuated in the FFFF systems. This performance uncertainty is most likely due to the low constraints on the bonded structure[Marrink 2023].

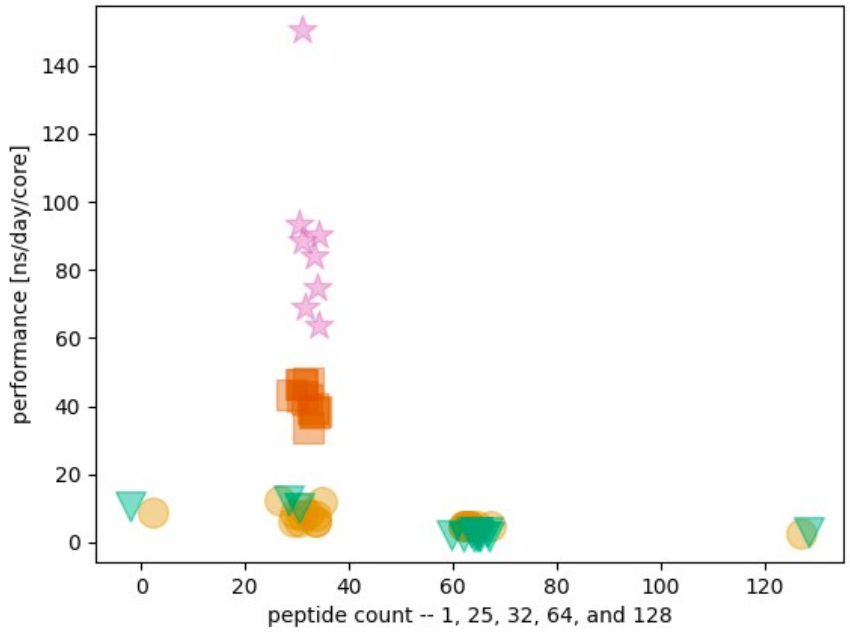



Figure S13. Production MD simulation performance per core. Orange circles: F2; Green triangles: F4; Red squares: Martini FF; Purple stars: Martini FFFF. Only systems of 32 peptides were simulated in Martini.

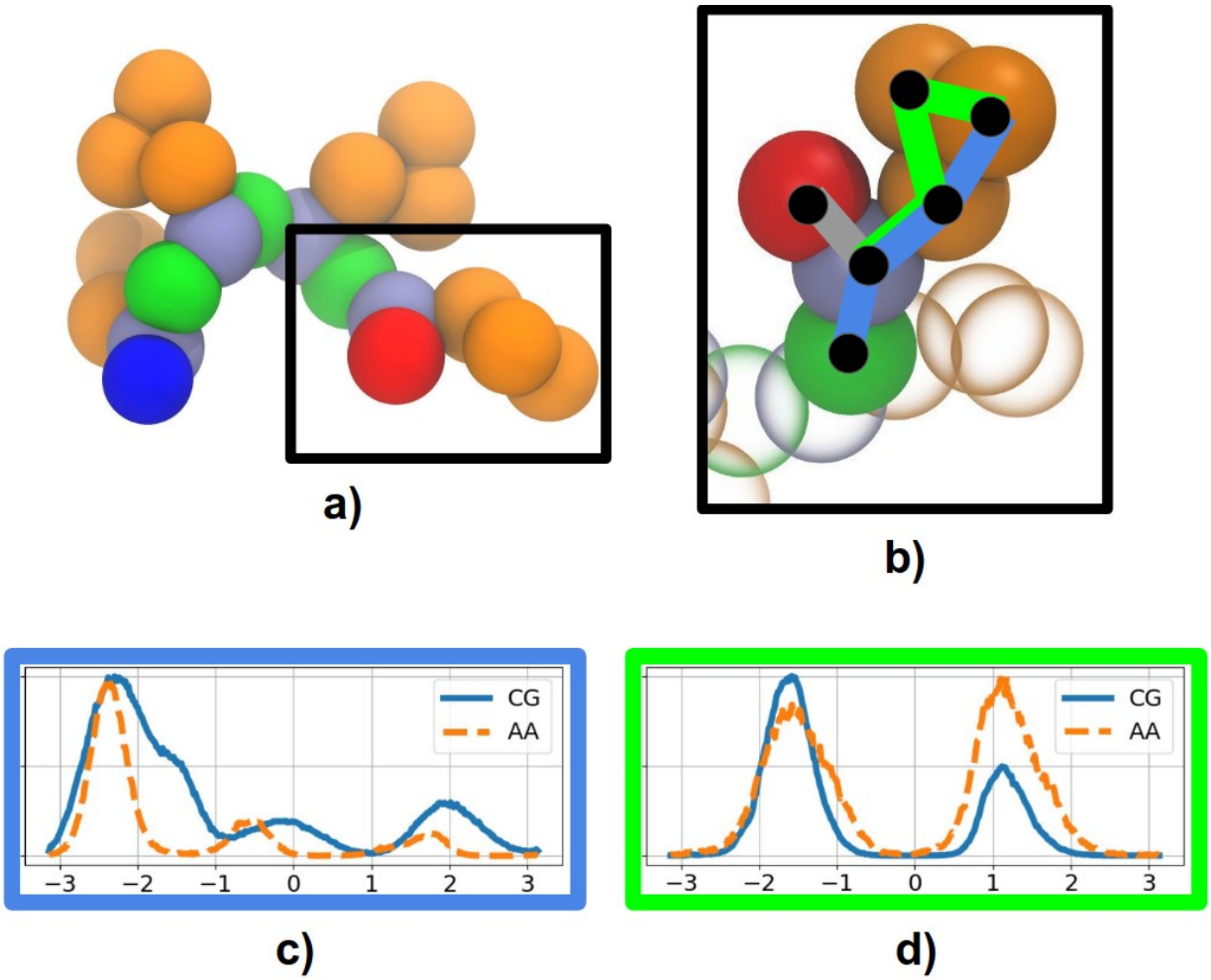



Figure S14. Schematic representing overlap in parameterized side chain dihedrals and differences in their performance relative to the atomistic reference. (a) F4 representation, with an inset box showing the location of (b). (b) Identifies two of the dihedrals impacting side chain orientations. One controlling the rotation of the Cα-Cβ bond is marked with blue lines, and one controlling the planar tilt of the aromatic ring is marked with green lines. Distributions of these two dihedrals at AA and CG resolution are shown in (c) and (d).

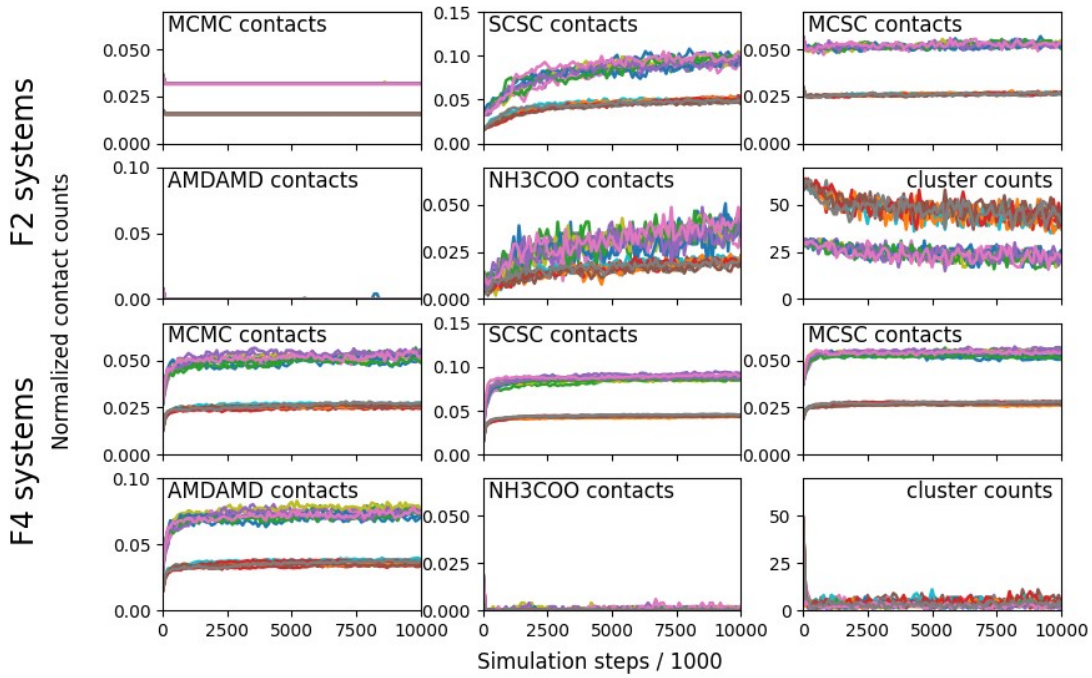



Figure S15. Contact and cluster count evolution over the first 10 million steps in a sample of F2 (upper six panels) and F4 (lower six panels) simulations of 32 or 64 peptides (9 systems of each configuration, 36 total). Contacts are calculated on Main Chain (MC) beads (type CA), Side Chain Beads (type PHA, PHB, PHC), amide group (AMD) beads (type AMD), and terminal (NH3COO) beads (type NH3, COO). Counts are normalized by dividing the actual count by the maximum number of contacts (i.e. the value if all beads of the appropriate type were overlapping). The limits of the y-axes are matched on each contact type, so that the difference in contact counts may clearly be distinguished between peptide types (F2 upper, F4 lower). Within each panel, the curves cluster into a lower and an upper group, reflecting the 32 and 64 peptide systems, respectively.

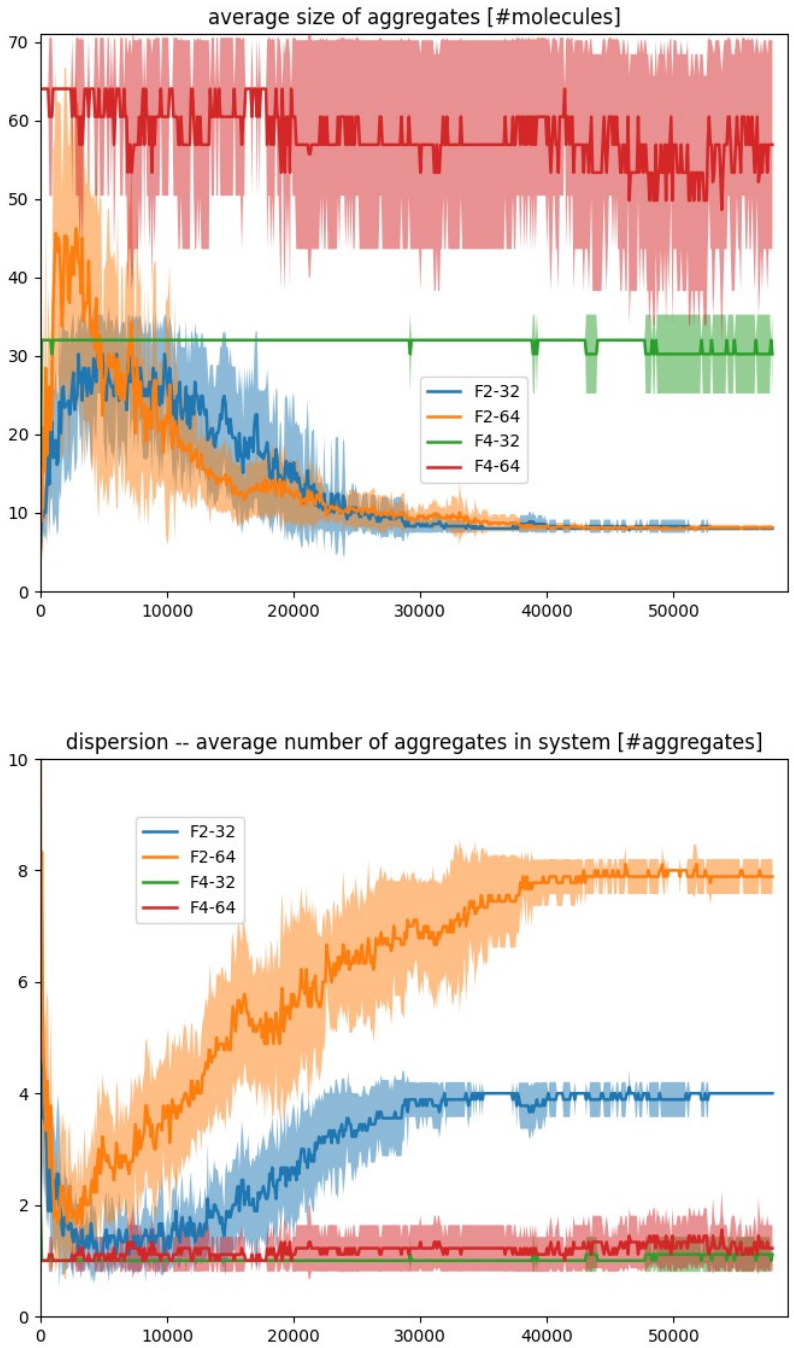



Figure S16. Ensemble average aggregate size over 50 million simulation steps. Upper panel shows the raw ensemble average size of aggregates. Lower panel divides the number of peptides in system by the average aggregate size, reflecting a measure of aggregate dispersity, where monolithic aggregation achieves value of 1 and a disperse arrangement of disconnected aggregates are higher than 1.

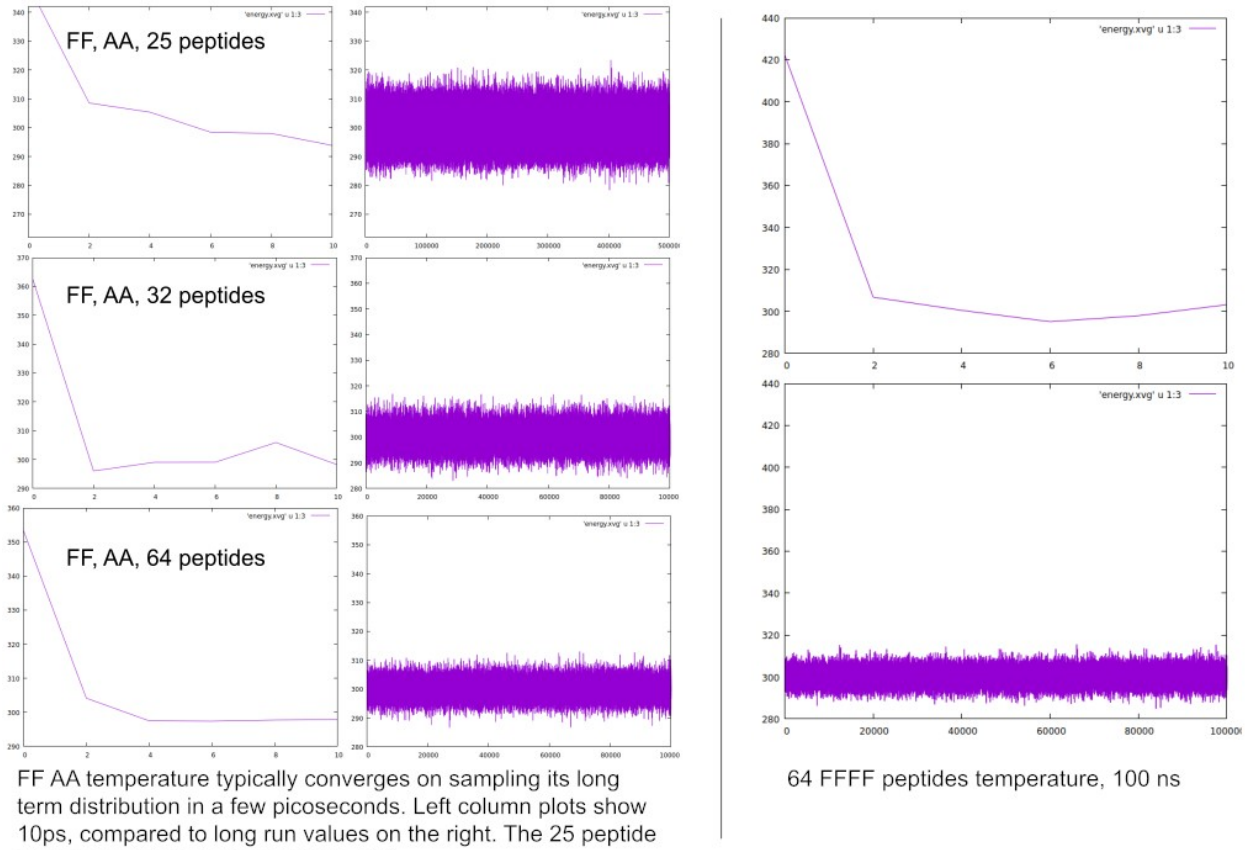



Figure S17. Temperature convergence of AA reference systems.

Banerjee, A.; Lu, C. Y.; Dutt, M. A Hybrid Coarse-Grained Model for Structure, Solvation and Assembly of Lipid-like Peptides. Phys. Chem. Chem. Phys. 2022, 24 (3), 1553–1568. https://doi.org/10.1039/D1CP04205J.

Hooten, M.; Banerjee, A.; Dutt, M. Multiscale, Multiresolution Coarse-Grained Model via a Hybrid Approach: Solvation, Structure, and Self-Assembly of Aromatic Tripeptides. J. Chem. Theory Comput. 2024, 20 (4), 1689–1703. https://doi.org/10.1021/acs.jctc.3c00458.

Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. T. Two Decades of Martini: Better Beads, Broader Scope. WIREs Computational Molecular Science 2023, 13 (1), e1620. https://doi.org/10.1002/wcms.1620.