Supplementary Information

Computational Studies of Intrinsically Disordered Proteins

Vy T. Duong,[1,2] Zihao Chen,[2] Mahendra T. Thapa,[3] Ray Luo[2,4]*

[1]Department of Chemistry, University of California, Irvine, California 92697-3900, USA
[2]Center of Complex Biological Systems, University of California, Irvine, California 92697-3900, USA
[3]Department of Physics, California State University, Chico, Chico, California 95929, USA
[4]Departments of Molecular Biology and Biochemistry, Chemical Engineering and Materials Science, and Biomedical Engineering, University of California, Irvine, California 92697-3900, USA

Table of Contents:

Cumulative Averaging of Observables:

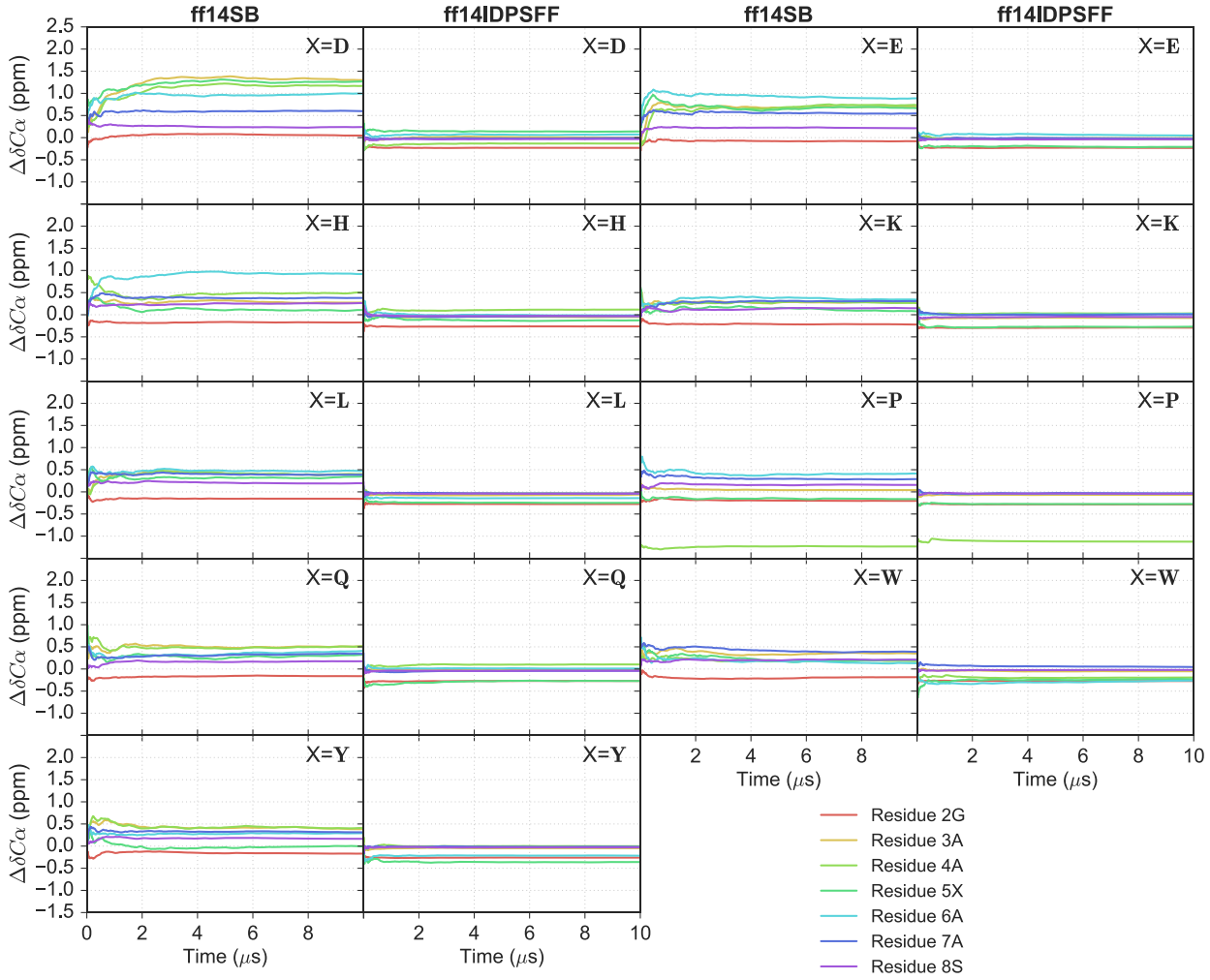

Figure S1. The $\Delta\delta C^{\alpha}$-derived cumulative averages per EGAAXAASS peptide and force field type were calculated and averaged between the 10 simulations. The first/third column is populated with short peptides simulated using the *ff*14SB and the second/fourth column is populated by the corresponding peptide simulated using the *ff*14IDPSFF. Each row represents an EGAAXAASS (X = D, E, H, K, L, P, Q, W, Y) peptide.

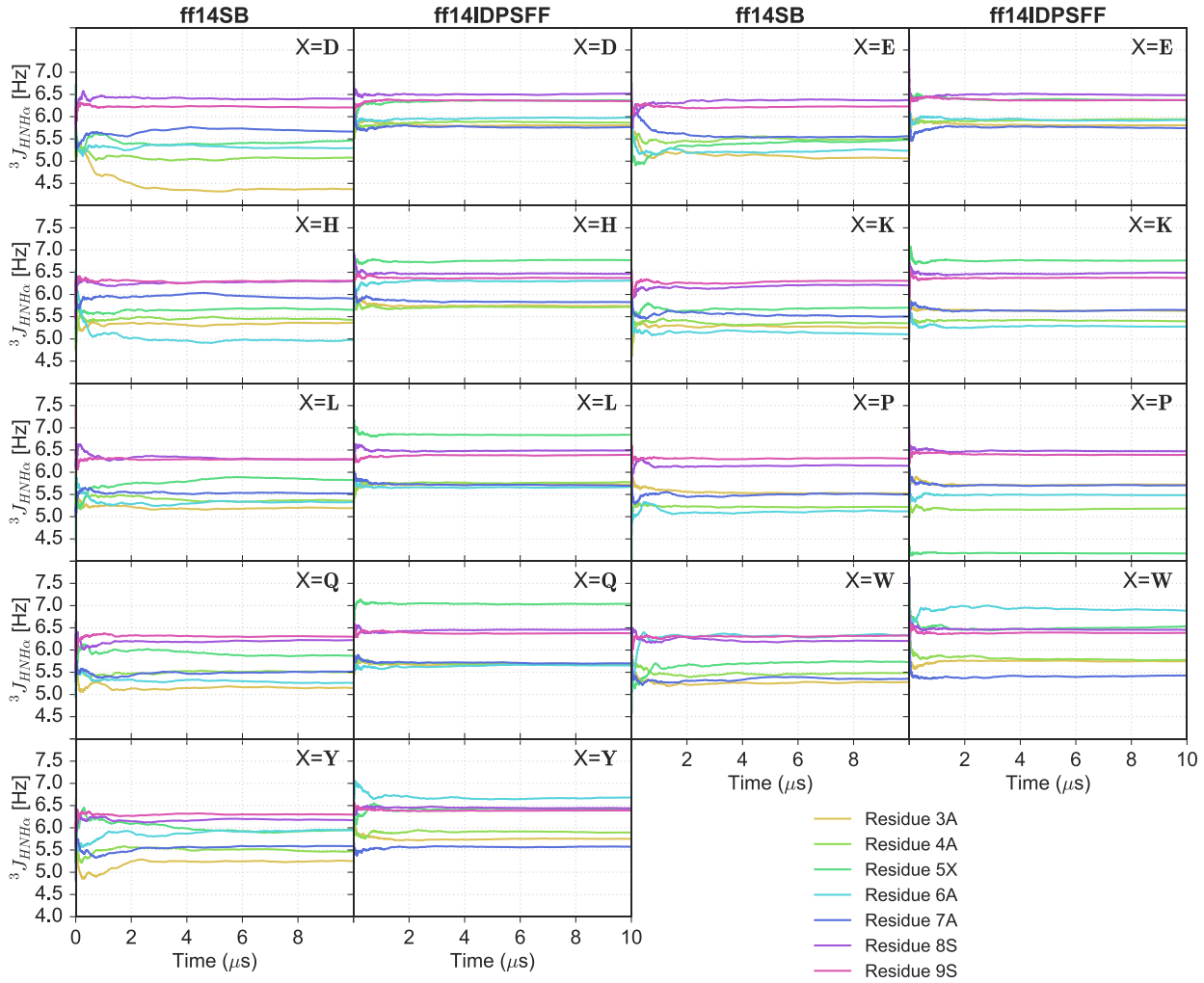

Figure S2. The $^3J_{HNH\alpha}$-derived cumulative averages per EGAAXAASS peptide and force field type were calculated and averaged between the 10 simulations. The first/third column is populated with short peptides simulated using the *ff*14SB and the second/fourth column is populated by the corresponding peptide simulated using the *ff*14IDPSFF. Each row represents an EGAAXAASS (X = D, E, H, K, L, P, Q, W, Y) peptide.

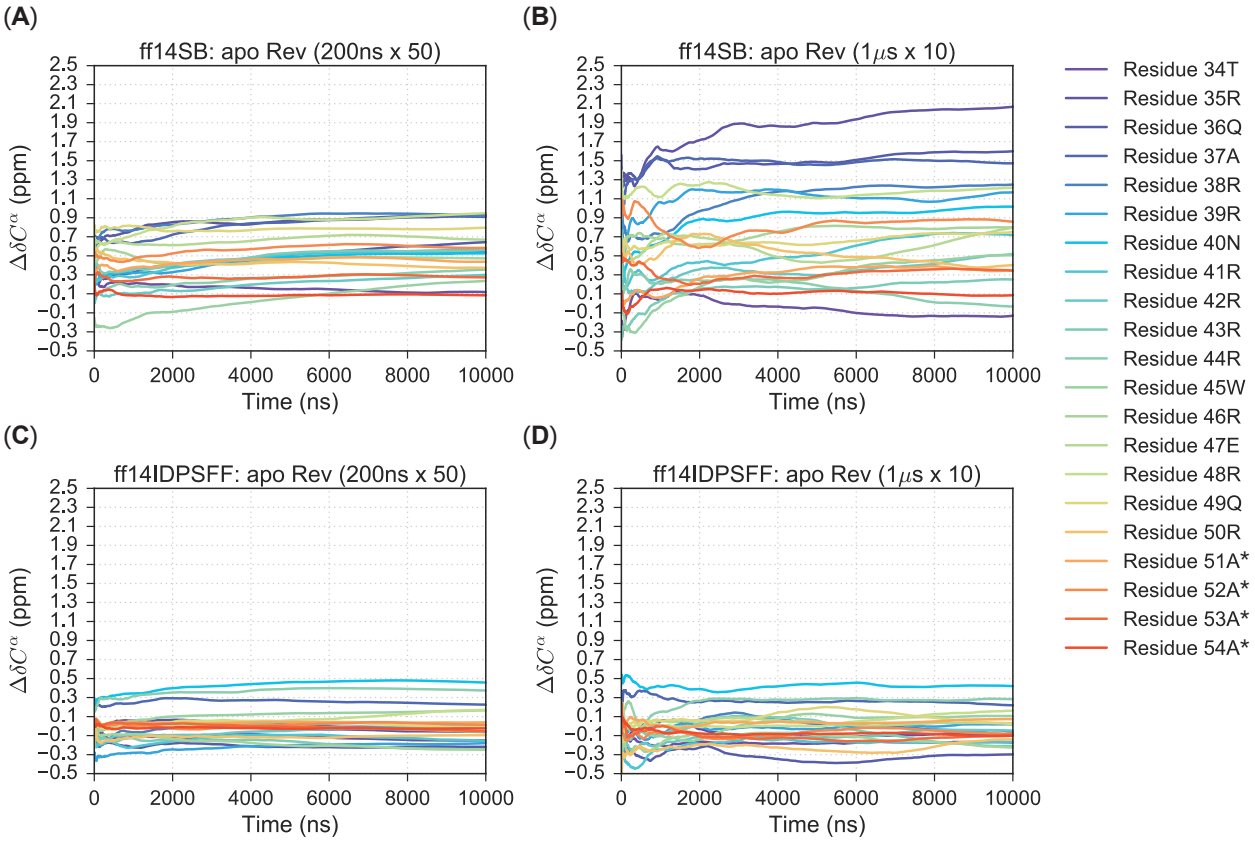

Figure S3. The $\Delta\delta C^{\alpha}$-derived cumulative averages per apo Rev peptide and force field type were calculated and averaged between the 10/50 simulations. Two simulations types were generated: fifty 200ns simulations using (A) *ff*14SB (B) and *ff*14IDPSFF, (C) and ten 1 µs simulations using *ff*14SB (D) and *ff*14IDPSFF. Residues are colored according to the legend with an asterisk (*) indicating non-native residues.

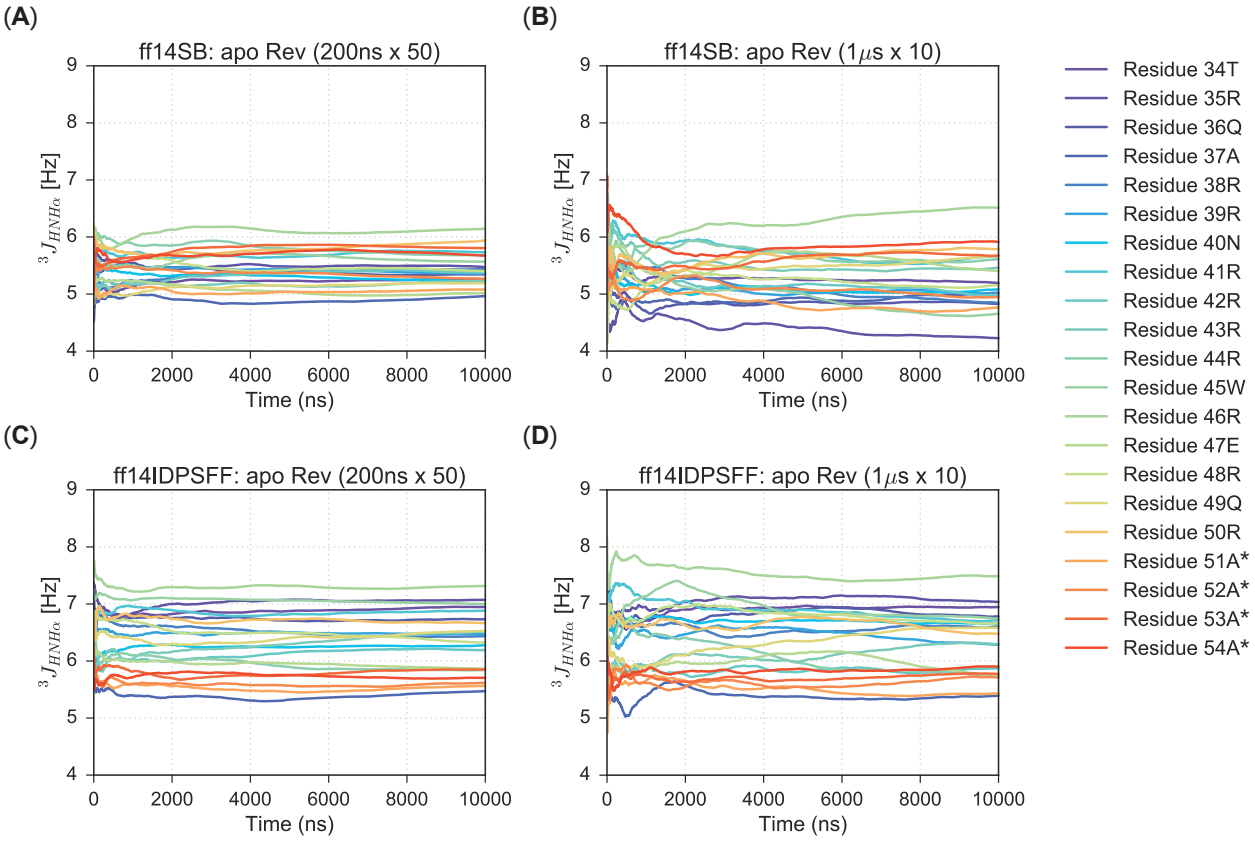

Figure S4. The $^3J_{HNH\alpha}$-derived cumulative averages per apo Rev peptide and force field type were calculated and averaged between the 10/50 simulations. Two simulations types were generated: fifty 200ns simulations using (A) *ff*14SB (B) and *ff*14IDPSFF, (C) and ten 1 µs simulations using *ff*14SB (D) and *ff*14IDPSFF. Residues are colored according to the legend with an asterisk (*) indicating non-native residues.

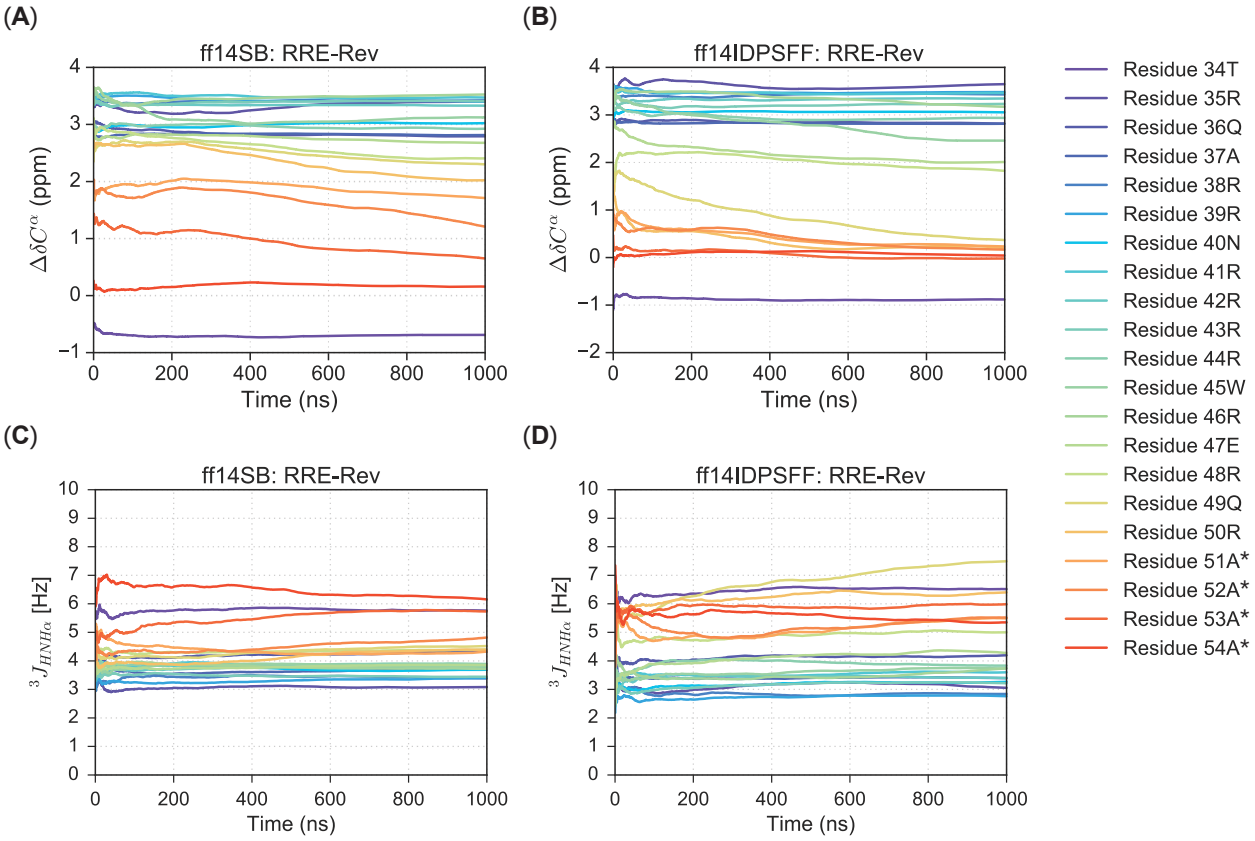

Figure S5. The $\Delta\delta C^{\alpha}$- and $^{3}J_{HNH\alpha}$-derived cumulative averages per RRE-Rev complex and force field type were calculated and averaged between the 5 simulations. Secondary chemical shifts occupy the first row from (A) *ff*14SB-generated simulations and (B) *ff*14IDPSFF-generated simulations. $^{3}J_{HNH\alpha}$-coupling constants occupy the second row from (C) *ff*14SB-generated simulations and (D) *ff*14IDPSFF-generated simulations. Residues are colored according to the legend with an asterisk (*) indicating non-native residues.

Biphasic Exponential Fitting of $\Delta\Delta\delta C^{\alpha}$ Datasets:

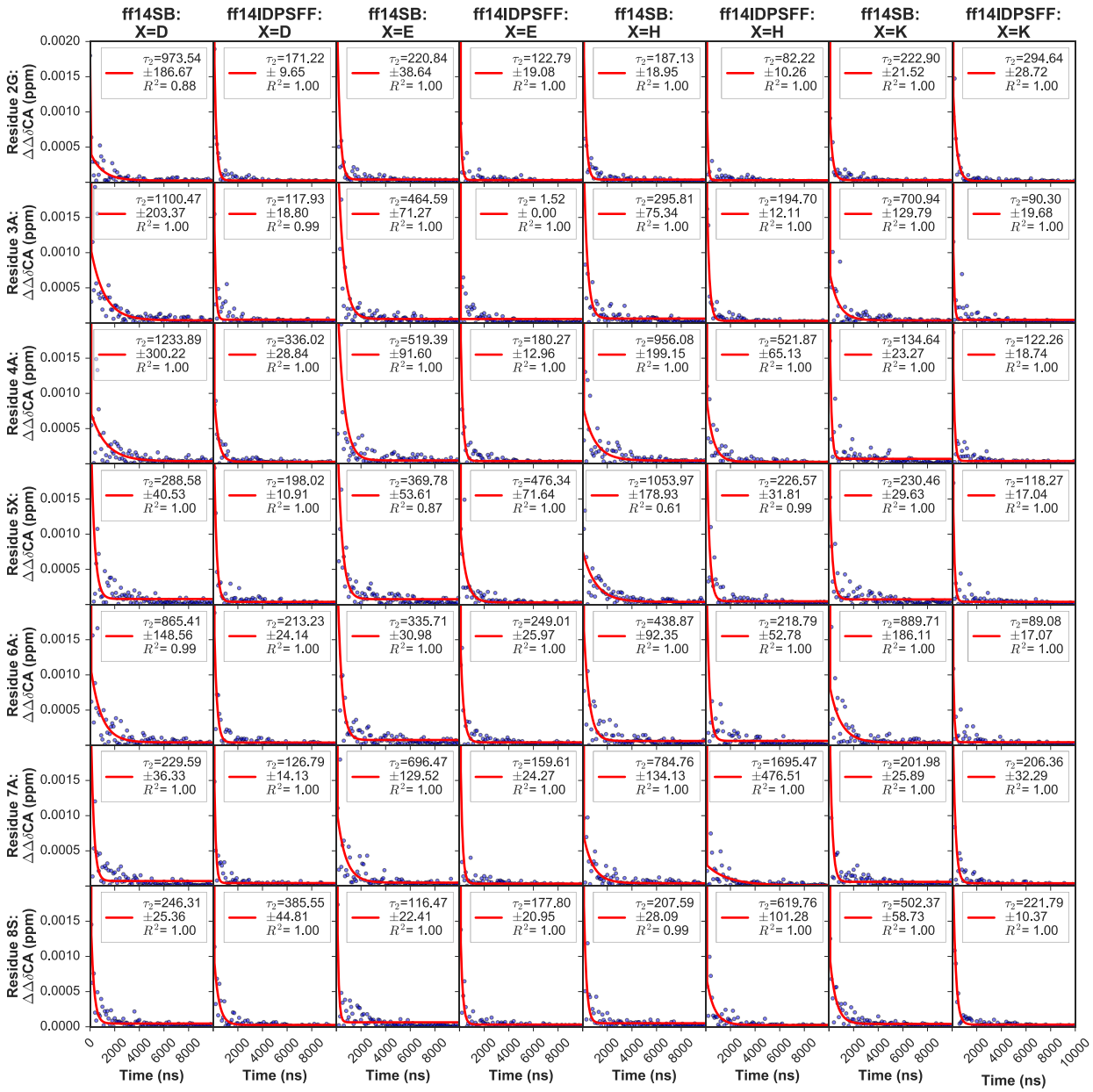

Figure S6. Biphasic exponential fittings were generated using $\Delta\Delta\delta C^{\alpha}$ from cumulative average data in Figure S1 for EGAAXAASS (X= D, E, H, K) peptides and force field types. Each average cumulative $\Delta\Delta\delta C^{\alpha}$ (blue dots) 100-ns increment was plotted per residue. Datasets were fitted to the following exponential decay function: $\Delta\Delta\delta C^{\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each column represents a peptide and force field, and each row represents a single residue. Only residues 2G-8S are fitted.

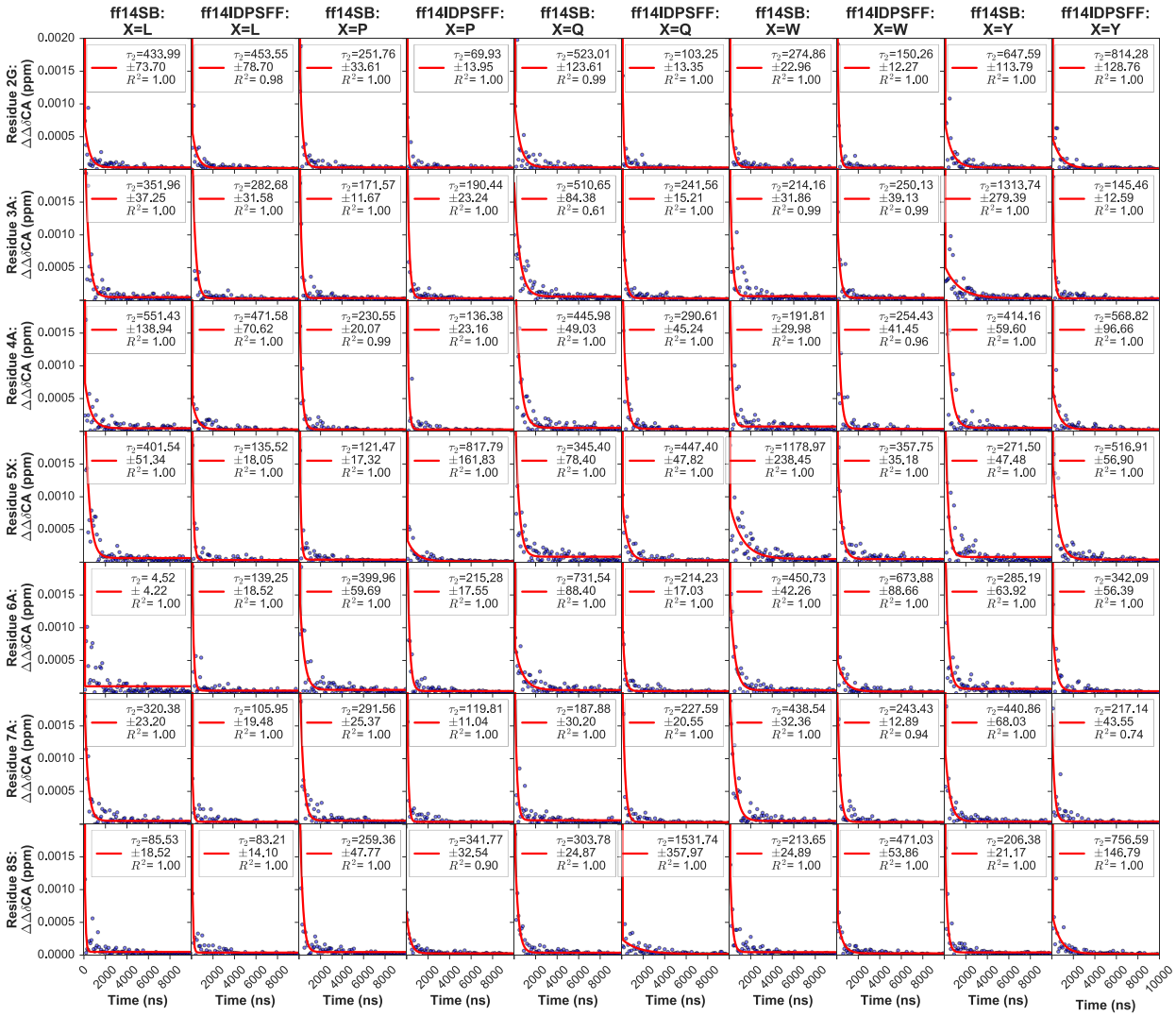

Figure S7. Biphasic exponential fittings were generated using $\Delta\Delta\delta C^{\alpha}$ from cumulative average data in Figure S1 for EGAAXAASS (X= L, P, Q, W, Y) peptides and force field types. Each average cumulative $\Delta\Delta\delta C^{\alpha}$ (blue dots) 100-ns increment was plotted per residue. Datasets were fitted to the following exponential decay function: $\Delta\Delta\delta C^{\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each column represents a peptide and force field, and each row represents a single residue. Only residues 2G-8S are fitted.

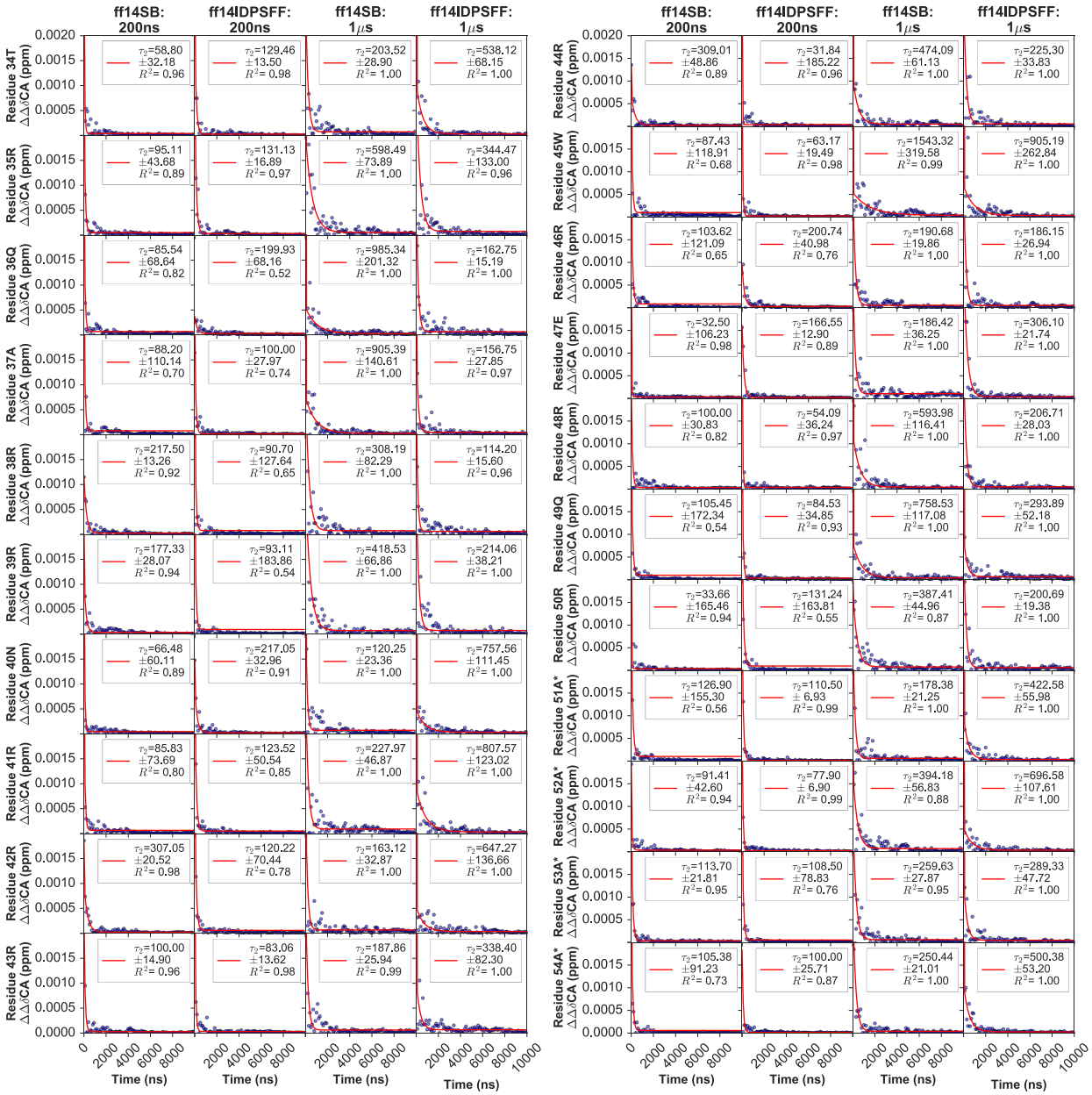

Figure S8. To evaluate cumulative average convergence of apo Rev simulations from Figure S3, a scatter plot of $\Delta\Delta\delta C^{\alpha}$ values (blue dots) and corresponding biphasic exponential fit were generated for each simulation (long, short) and force field (*ff*14SB, *ff*14IDPSFF) types. Datasets were fitted to the following exponential decay function: $\Delta\Delta\delta C^{\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). The above subplot columns are titled according to simulation and force field type and rows labeled according to residue, with non-native residues marked with an asterisk (*) on the y-axis.

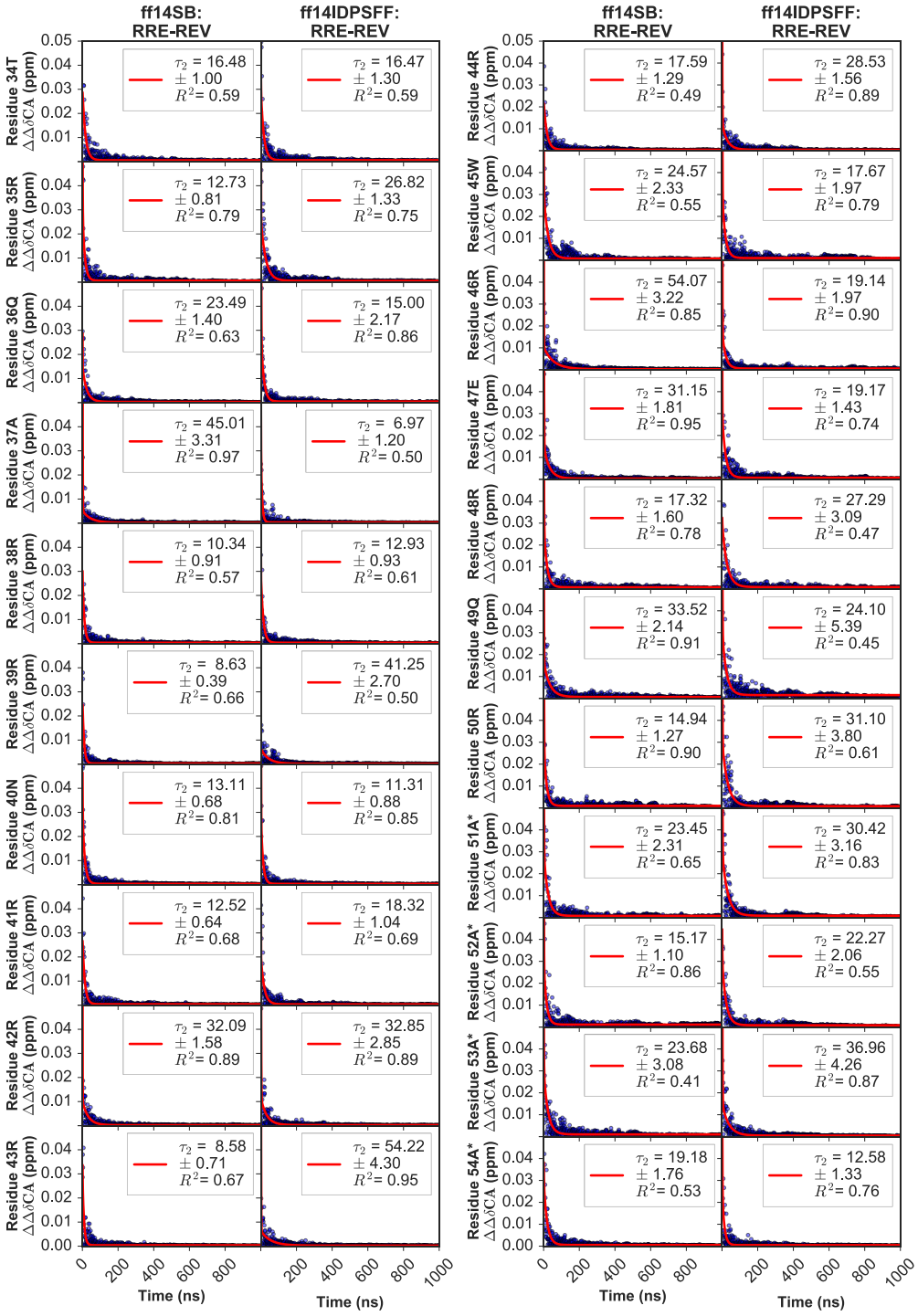

Figure S9. Biphasic exponential fittings were generated using $\Delta\Delta\delta C^{\alpha}$ from cumulative average data in Figure S5 for RRE-Rev complexes and force field types. We applied the same fitting to the following exponential decay function: $\Delta\Delta\delta C^{\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each average cumulative $\Delta\Delta\delta C^{\alpha}$ (blue dots) 1-ns increment was plotted per residue. Each column represents a peptide and force field, each row is labeled to its corresponding residue, and non-native residues marked with an asterisk (*).

Biphasic Exponential Fitting of $\Delta^3 J_{HNH\alpha}$ Datasets:

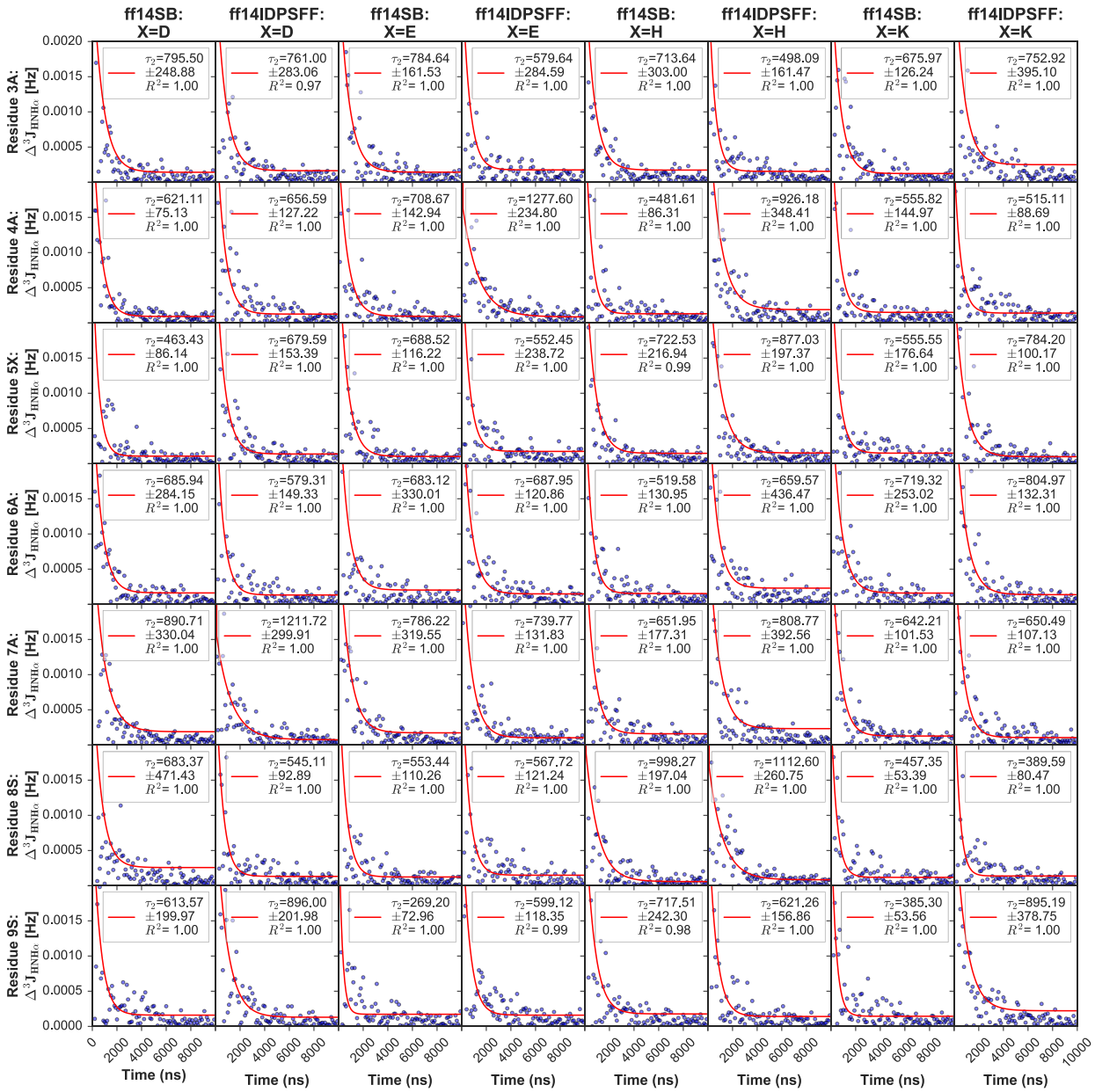

Figure S10. Biphasic exponential fittings were generated using $\Delta^3 J_{HNH\alpha}$ from cumulative average data in Figure S2 for EGAAXAASS (X= D, E, H, K) peptides and force field types. Each average cumulative $\Delta^3 J_{HNH\alpha}$ (blue dots) 100-ns increment was plotted per residue. Datasets were fitted to the following exponential decay function: $\Delta^3 J_{HNH\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each column represents a peptide and force field and each row represents individual residues. Only residues 3A-9S are fitted.

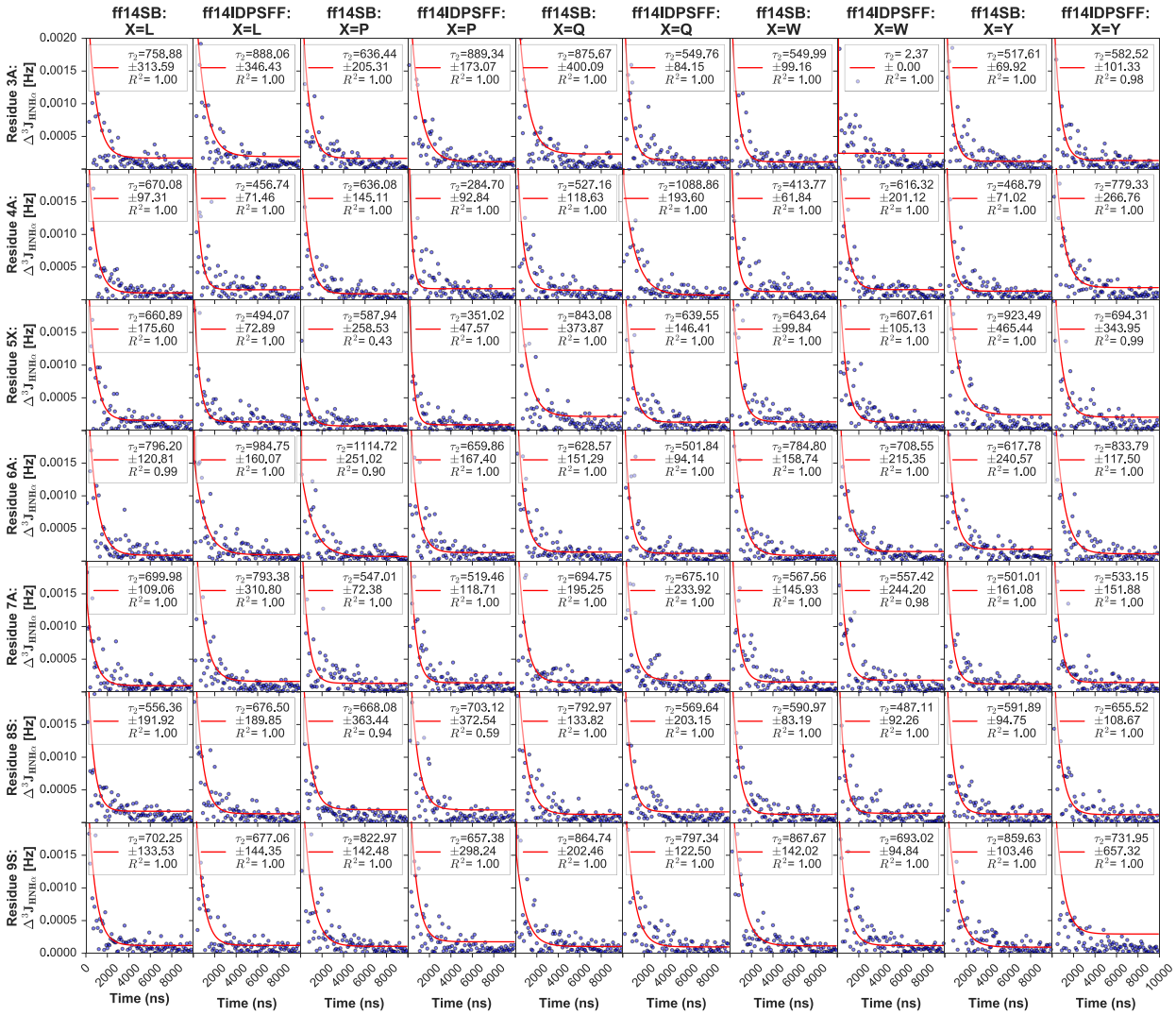

Figure S11. Biphasic exponential fittings were generated using $\Delta^3 J_{HNH\alpha}$ from cumulative average data in Figure S2 for EGAAXAASS (X= L, P, Q, W, Y) peptides and force field types. Each average cumulative $\Delta^3 J_{HNH\alpha}$ (blue dots) 100-ns increment was plotted per residue. Datasets were fitted to the following exponential decay function: $\Delta^3 J_{HNH\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each column represents a peptide and force field, and each row represents individual residues. Only residues 3A-9S are fitted.

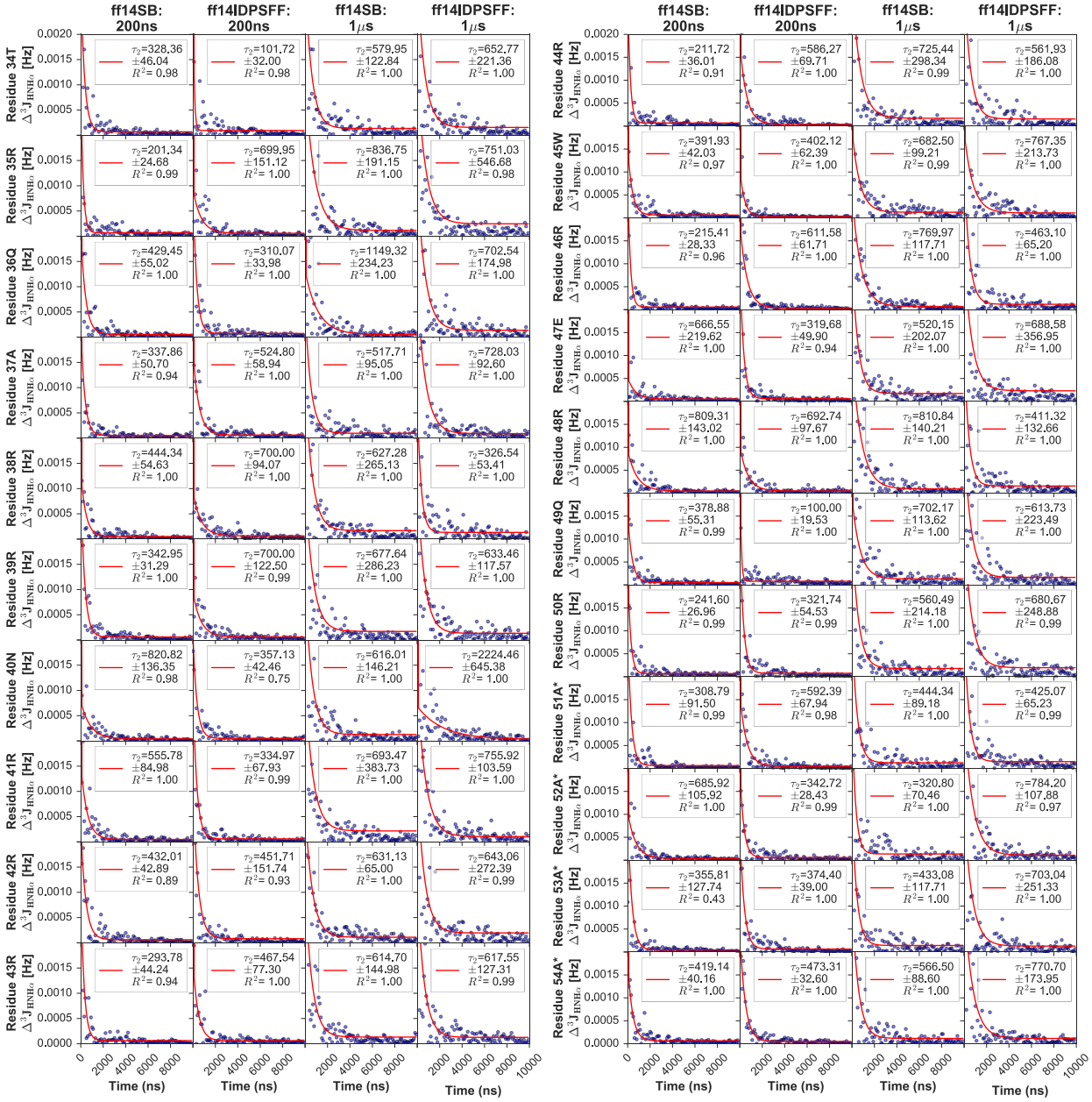

Figure S12. To evaluate cumulative average convergence of apo Rev simulations from Figure S4, a scatter plot of $\Delta^3 J_{HNH\alpha}$ values (blue dots) and corresponding biphasic exponential fit were generated for each simulation (long, short) and force field (*ff*14SB, *ff*14IDPSFF) types. Datasets were fitted to the following exponential decay function $\Delta^3 J_{HNH\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$. The above subplots are titled according to simulation and force field type. Each column represents a peptide and force field, each row is labeled to its corresponding residue, and non-native residues marked with an asterisk (*).

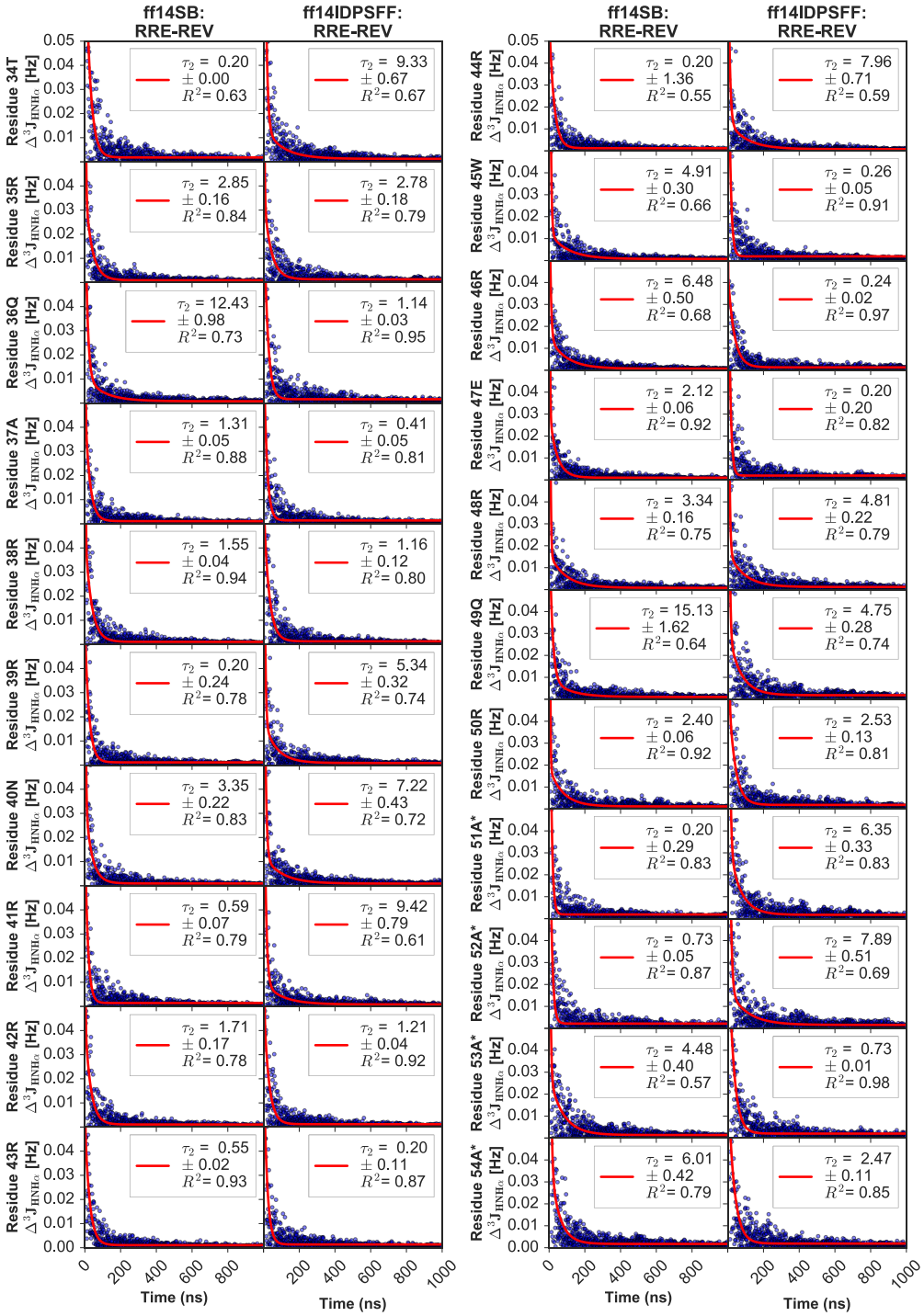

Figure S13. Biphasic exponential fittings were generated using $\Delta^3 J_{HNH\alpha}$ from cumulative average data in Figure S5 for RRE-Rev complexes and force field types. We applied the same fitting to the following exponential decay function: $\Delta^3 J_{HNH\alpha} = A_1 e^{-\frac{x}{\tau_1}} + A_2 e^{-\frac{x}{\tau_2}} + c$ (red line). Each average cumulative $\Delta^3 J_{HNH\alpha}$ (blue dots) 1-ns increment was plotted per residue. Each column represents a peptide and force field, each row is labeled to its corresponding residue, and non-native residues marked with an asterisk (*).

Clustering (apo Rev):

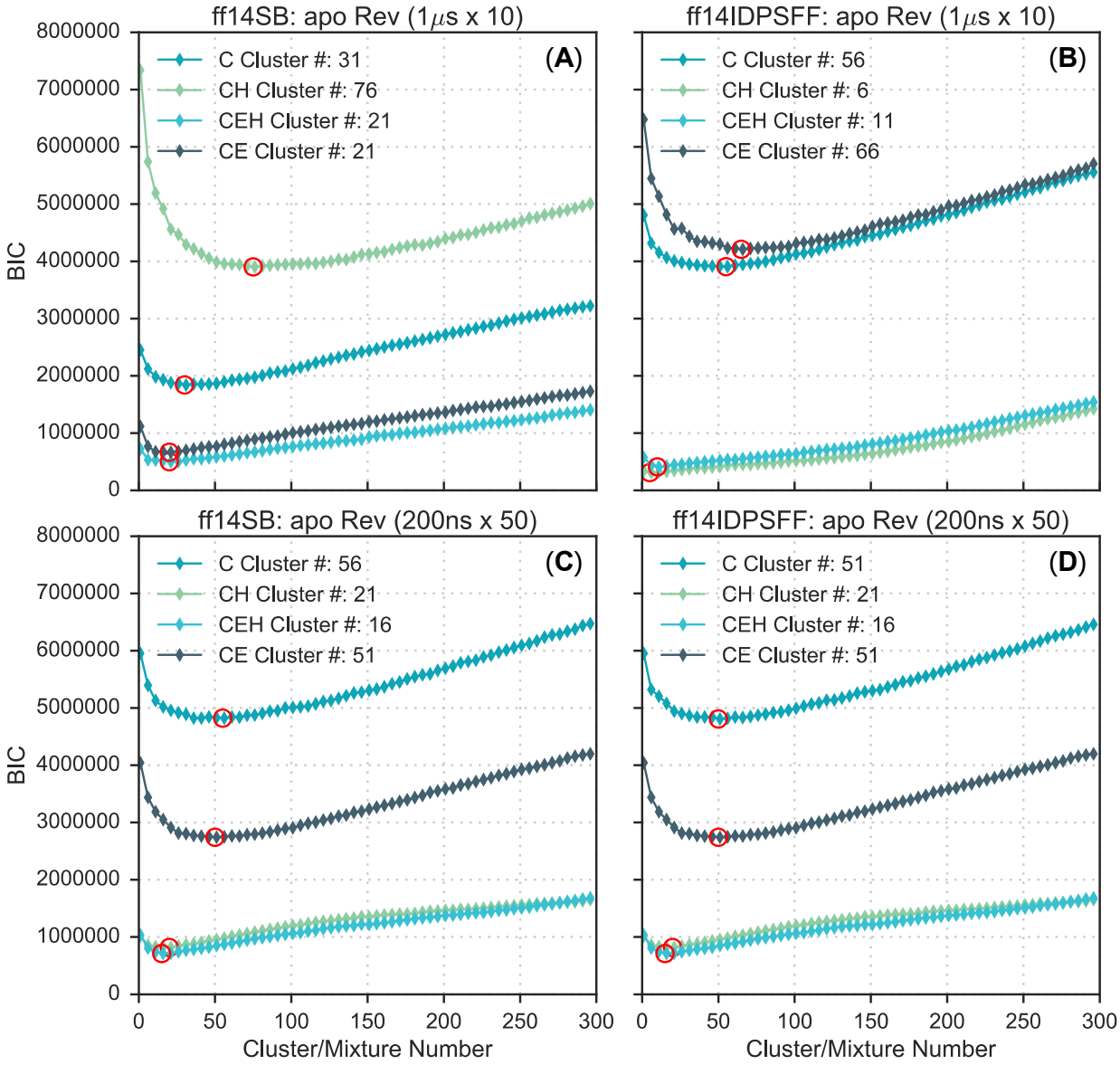

Figure S14. Determination of appropriate cluster/mixture number using the Bayesian information criterion (BIC) for apo Rev simulations. We calculated the BIC score between 1 to 300 mixtures, and the mixture/cluster number with the lowest BIC was selected for GMM generation. Chosen cluster numbers are indicated in the legend according to secondary structure categories from DSSP pre-clustering. (A) BIC plot of ten 1μs simulations using the *ff*14SB force field. (B) BIC of ten 1μs simulations using the *ff*14IDPSFF force field. (C) BIC plot of fifty 200ns simulations using the *ff*14SB force field. (D) BIC plot of fifty 200ns simulations using the *ff*14IDPSFF force field.

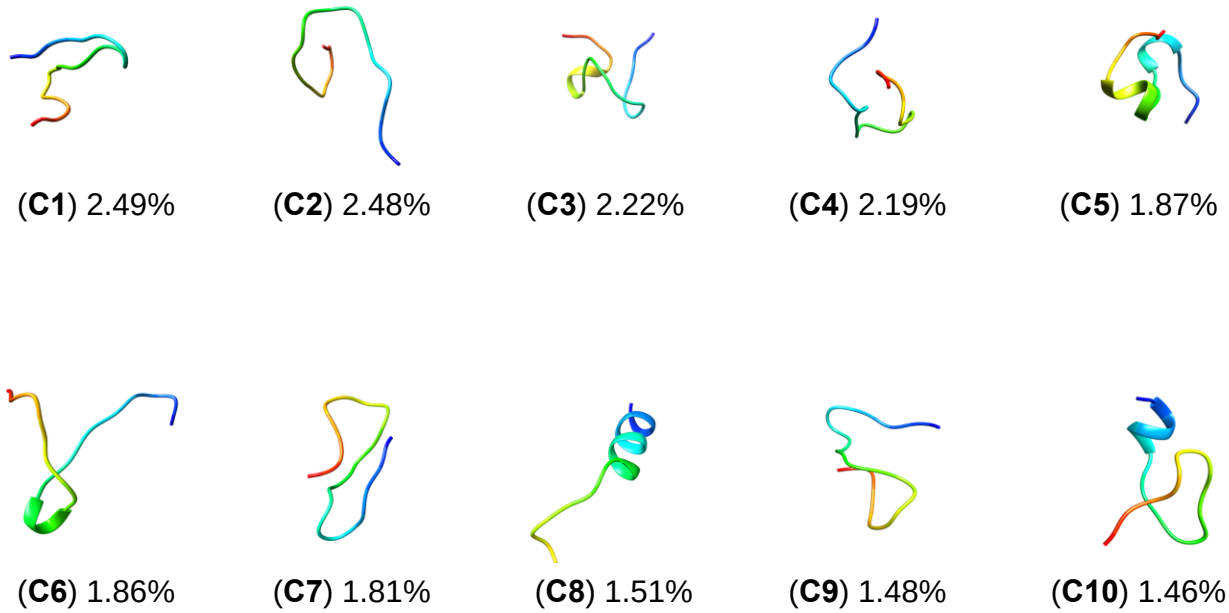

Figure S15. Top 10 clusters of *ff*14SB-parameterized simulations (200ns x 50) encompass 19.36% of all frames. Clusters are labeled C1-C10 and colored according to N- to C-termini sequence (red to blue).

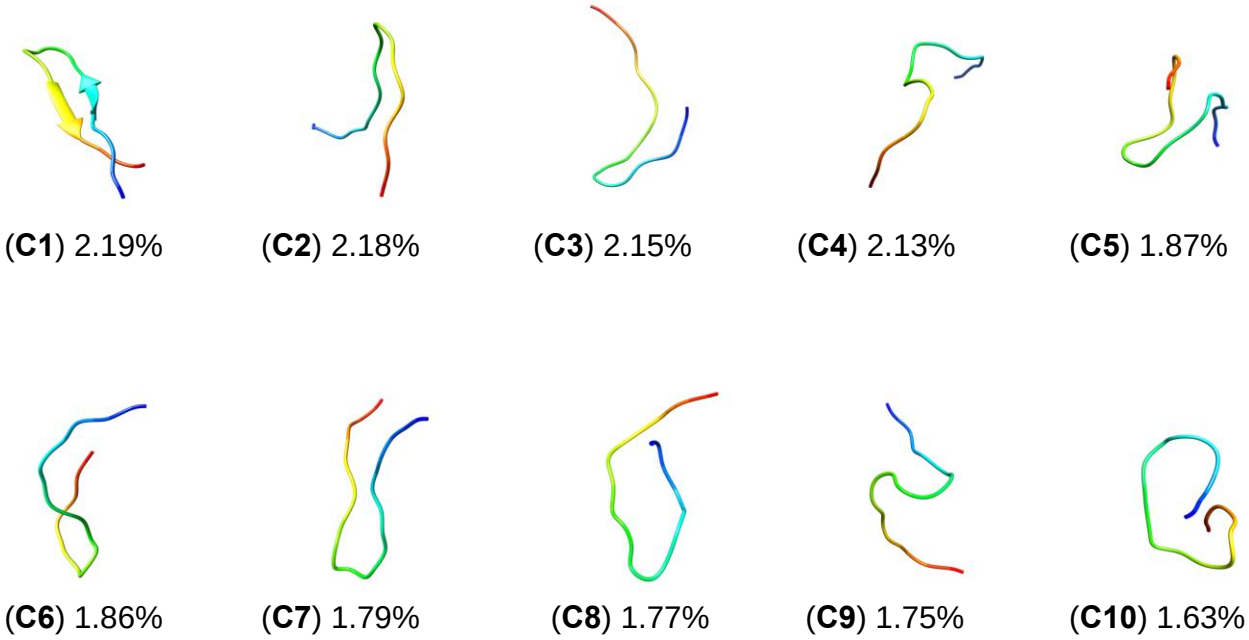

Figure S16. Top 10 clusters of *ff*14IDPSFF-parameterized simulations (200ns x 50) encompass 19.32% of all frames. Clusters are labeled C1-C10 and colored according to N- to C-termini sequence (red to blue).

DSSP:

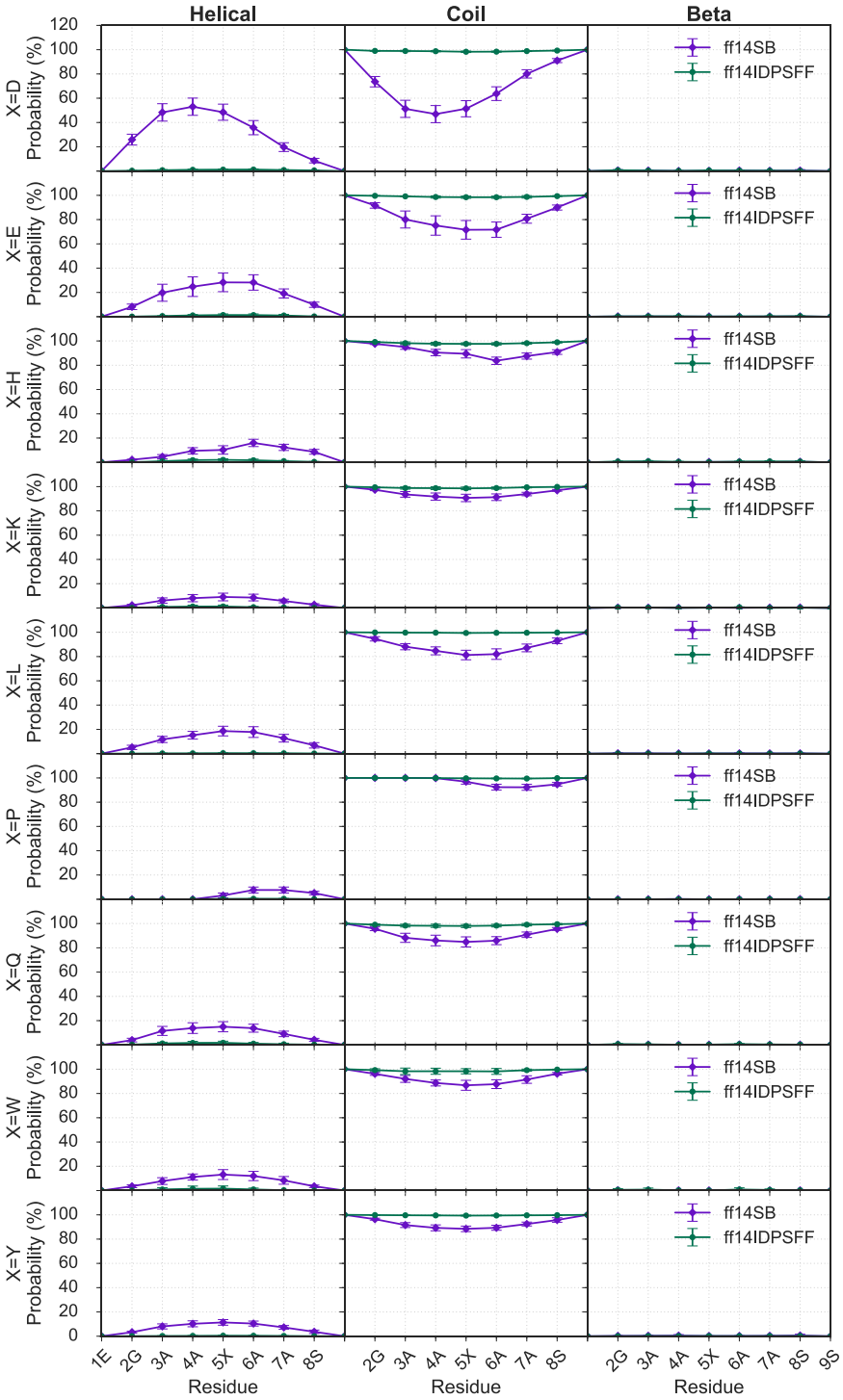

Figure S17. The average secondary structure propensity of each disordered short peptide. Colors correspond to force fields: purple – *ff*14SB, green – *ff*14IDPSFF. All values were calculated using the DSSP[1] program and MDtraj.[2] Rows indicate peptide (X = D, E, H, K, L, P, Q, W, Y) and columns indicate one of the three generalized secondary structures (helical, coiled, beta).

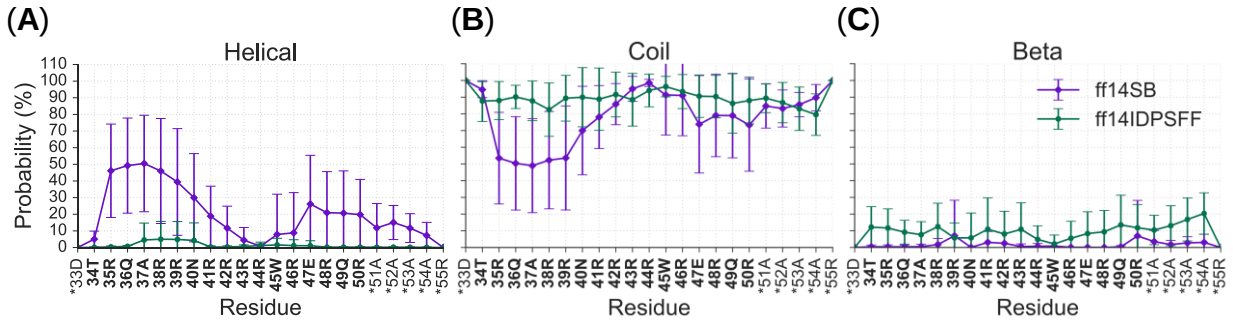

Figure S18. The average secondary structure propensity of each apo Rev residue was quantified from long simulation (1µs x 10) datasets. Colors correspond to force fields: purple – *ff*14SB, green – *ff*14IDPSFF. All values were calculated using the DSSP[1] program and MDtraj.[2] (A) The probability of a residue exhibiting helical content. (B) Probability of coil content per residue. (C) Displays the beta-sheet helical propensity per residue. Non-native residues are indicated with an asterisk (*).

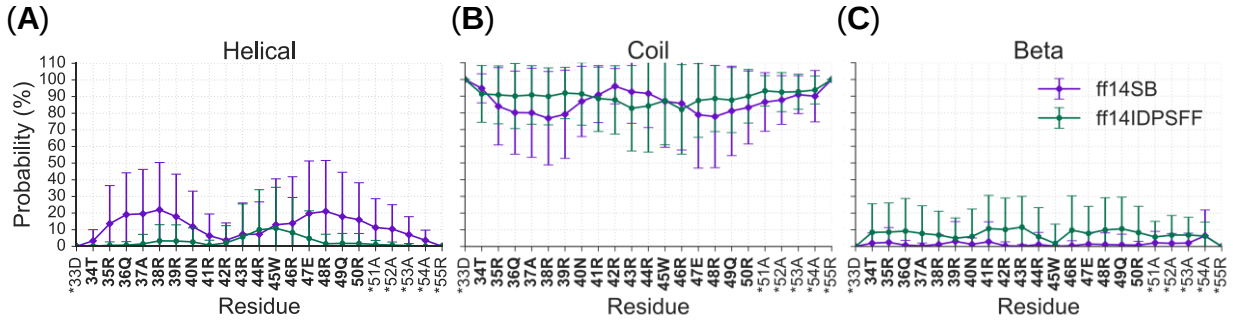

Figure S19. The average secondary structure propensity of each apo Rev residue was quantified from short simulation (200ns x 50) datasets. Colors correspond to force fields: purple – *ff*14SB, green – *ff*14IDPSFF. All values were calculated using the DSSP[1] program and MDtraj.[2] (A) The probability of a residue exhibiting helical content. (B) Probability of coil content per residue. (C) Displays the beta-sheet helical propensity per residue. Non-native residues are indicated with an asterisk (*).

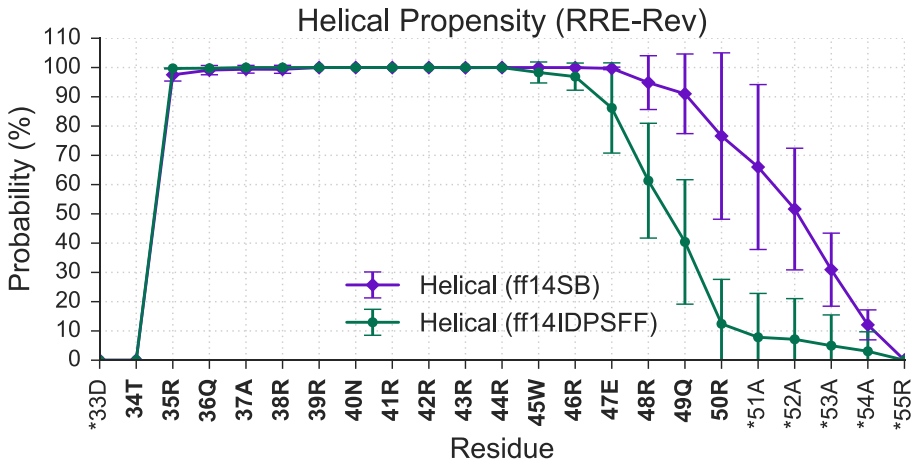

Figure S20. Average helical propensity of Rev from bound RRE-Rev simulations using the DSSP[1] program. Colors indicate force field: purple – *ff*14SB, green – *ff*14IDPSFF.

RMSF

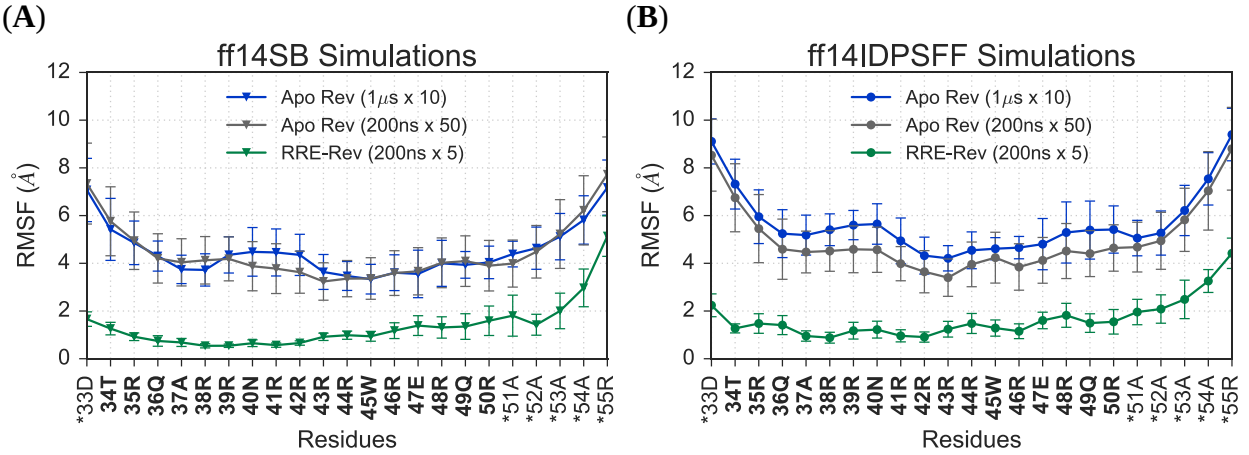

Figure S21. RMSF analyses of backbone Cα atoms Rev-related simulations. (A) Average RMSF of backbone atoms in apo and bound Rev *ff*14SB-parameterized simulations. (B) Average RMSF of backbone atoms in apo and bound Rev *ff*14IDPSFF-parameterized simulations. Non-native residues contain an asterisk (*).

References:

1. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983,** *22* (12), 2577-637.
2. McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernandez, C. X.; Schwantes, C. R.; Wang, L. P.; Lane, T. J.; Pande, V. S., MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **2015,** *109* (8), 1528-32.