Supporting Text

Convergence of molecular dynamics simulations

In an attempt to obtain reasonable conformational sampling of the 441 two-residue peptides studied here, we employed comparatively long simulation times (300 ns). Although this is, to our knowledge, much longer than has been used in other MD studies that have attempted to examine a very large number of peptide systems (e.g. refs. 2c and 54 of the main text), it is important to consider the extent to which such a simulation length allows sampling of the conformational behavior. To determine this, we divided each simulation trajectory into three 100 ns time intervals and calculated the standard deviations of the ϕ and ψ backbone dihedral angle distributions over these three time intervals (see Methods). The results, averaged over the N-terminal and C-terminal positions, are shown in Figure S1A as a 2D matrix with the x-axis indicating the identity of the N-terminal residue (i.e. residue Xaa in Ace–Xaa–Yaa–Nme) and the y-axis indicating the identity of the C-terminal residue (i.e. residue Yaa in Ace–Xaa–Yaa–Nme); separate plots for the N-terminal and C-terminal residues are shown in Figure S1B and S1C. Red squares throughout Figure S1 indicate those two-residue peptides for which standard deviations of the ϕ and ψ distributions are particularly high.

The largest standard deviations – and therefore the poorest degrees of sampling – are generally associated with two-residue peptides that contain Ile (column I or row I in Figure S1A) and/or Trp (column W or row W in Figure S1A); the effects of the former can be understood in terms of the effects of its β-branched sidechain limiting backbone flexibility, the latter may be a consequence of restrictions due to the steric effects of the large side-chain and/or the presence of

comparatively long-lived stacking interactions. To further examine sampling issues with the 'worst case scenario' peptides, we performed additional 300 ns MD simulations for the Ile-Pro, Ile-Tyr and Trp-Arg peptides. As shown in Figures S2A-C, the Ramachandran maps obtained from the 3 independent repeats of each simulation are quite similar to each other, suggesting that reasonable sampling of backbone conformations can probably be obtained in a 300 ns MD simulation even for those two-residue peptides for which conformational sampling is poorest. Interestingly, slightly poorer sampling at the N-terminal position appears to be a common result: a separate examination of the standard deviations for the N-terminal residue (Figure S1B) shows that they are, on average, higher than those for the C-terminal residue (Figure S1C). In general, however, the Ramachandran maps for most peptides appear to be quite well converged. For example, Figure S2D shows the Ramachandran maps obtained from three independent 300 ns MD simulations of the Asp-Val peptide (chosen because it has the median standard deviation of those displayed in Figure S1A). For this peptide, which is representative of the majority of peptides, the maps obtained from the simulations are effectively identical in the regions encompassing the α, α', β and PPII conformations, indicating that 300 ns is likely to allow reliable Ramachandran maps to be obtained for most of the peptides studied here.

Supporting Figure legends

Figure S1 **Convergence of Ramachandran maps in MD simulations**. The calculated standard deviations of the backbone ϕ and ψ distributions for 441 two-residue peptides at N-terminus (A), C-terminus (B) and average (C). The residue at the N-terminus of each peptide is plotted along the x-axis; the residue at the C-terminus is along the y-axis. Note that Hp corresponds to the protonated form of histidine.

Figure S2 **Ramachandran maps obtained from 3 replicate MD simulations of selected peptides**. **A**: N-terminal and C-terminal Ramachandran maps expressed in free energy form for simulations of Ile-Pro. **B.** Same as A but showing results for Ile-Tyr. **C.** Same as A but showing results for Arg-Trp. **D.** Same as A but showing results for Asp-Val. Energies are colored in descending order from blue to red.

Figure S3 **Residue-specific Karplus parameters improve agreement with experiment**. Same as Figure 1A in the main text, but showing simulation values calculated using Karplus equations independently adjusted for each type of amino acid in order to maximize agreement with the experimental data (ref. 12 in the main text). This has been achieved by implementing a simple Monte Carlo optimization scheme that allows each of the three coefficients used in the Karplus equation to vary independently for each type of amino acid; trial changes to a residue-type's Karplus coefficients were accepted on the basis of improvements to the combined absolute error of all $^3J_{HNH_\alpha}$ coupling constants of all peptides containing that type of amino acid. We note that attempts to use a *single* Karplus equation that was optimized for converting the MD distributions into experimental values did not produce results significantly better than those obtained using the experimentally-derived Karplus parameters of Hu and Bax (ref. 17 in the main text).

Figure S4 **Ramachandran maps for the N-terminal residues sampled in MD simulations of 441 two-residue peptides**. Rows identify the N-terminal residue in the two-residue peptide and columns identify the C-terminal residue. Protonated histidine is indicated by H_P. Energies are colored in descending order from blue to red.

Figure S5 **Ramachandran maps for the C-terminal residues sampled in MD simulations of 441 two-residue peptides**. Same as Figure S3 but showing results for the C-terminal residue of each peptide. Color scheme is the same as used in Figure S4.

Figure S6 **Comparison of average secondary structure populations observed in this study for two-residue peptides with those obtained by Zhou et al. (ref. 34 in the main text) for single-residue peptides**. Average fractional populations for this work are the average over both the N-terminal and C-terminal positions.

Figure S7 **Neighboring residue effects on α and β fractional populations. A.** Plot comparing the simulated effect of each type of amino acid, when present at the N-terminal position, on the average fractional α population at the C-terminal position with that obtained from analysis of a coil library (ref. 35 of the main text). **B.** Same as A but comparing the simulated effect of each type of amino acid, when present at the C-terminal position, on the average fractional α population at the N-terminal position. **C.** Same as A but showing results for the average fractional β population. **D.** Same as B but showing results for the average fractional β population.

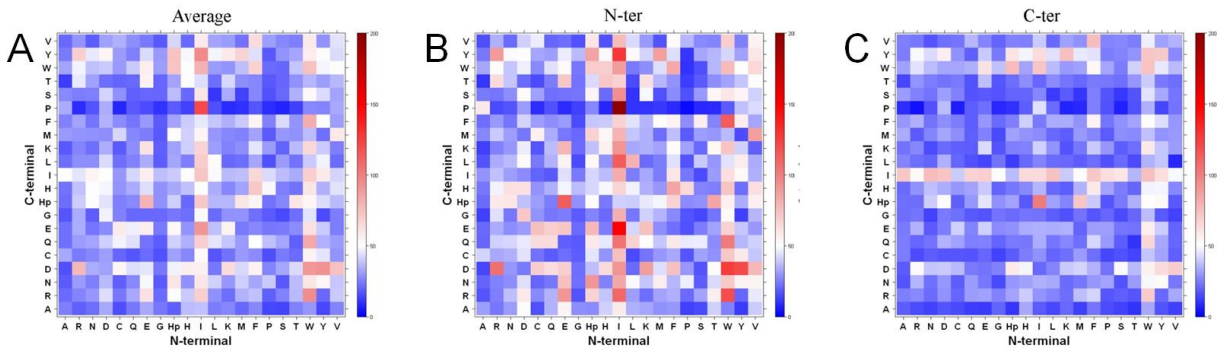
A
Average
B
N-ter
C
C-ter
C-terminal
N-terminal
V Y W T S P F M K L I H Hp G E Q C D N R A
A R N D C Q E G Hp H I L K M F P S T W Y V

Figure S1

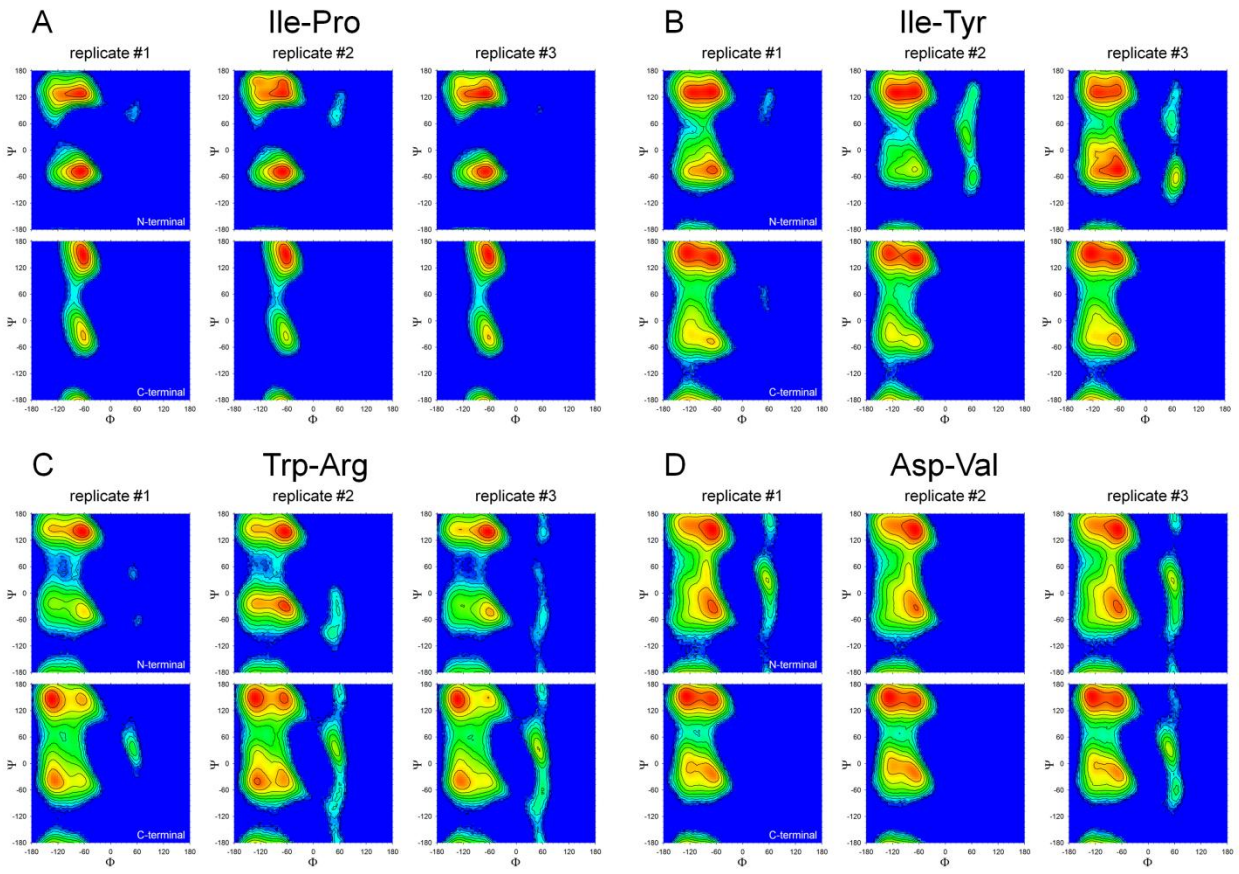
A
Ile-Pro
replicate #1
replicate #2
replicate #3
N-terminal
C-terminal
B
Ile-Tyr
replicate #1
replicate #2
replicate #3
N-terminal
C-terminal
C
Trp-Arg
replicate #1
replicate #2
replicate #3
N-terminal
C-terminal
D
Asp-Val
replicate #1
replicate #2
replicate #3
N-terminal
C-terminal
Ψ
Φ

Figure S2

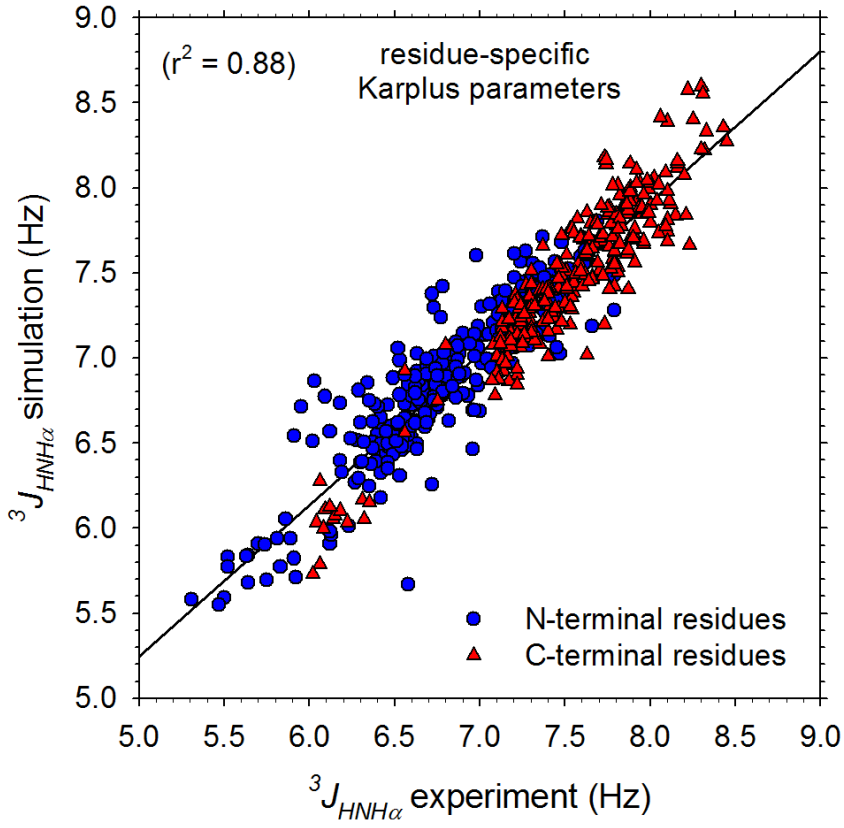
(r² = 0.88)
residue-specific
Karplus parameters
N-terminal residues
C-terminal residues
³J_HNHα simulation (Hz)
³J_HNHα experiment (Hz)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0

Figure S3

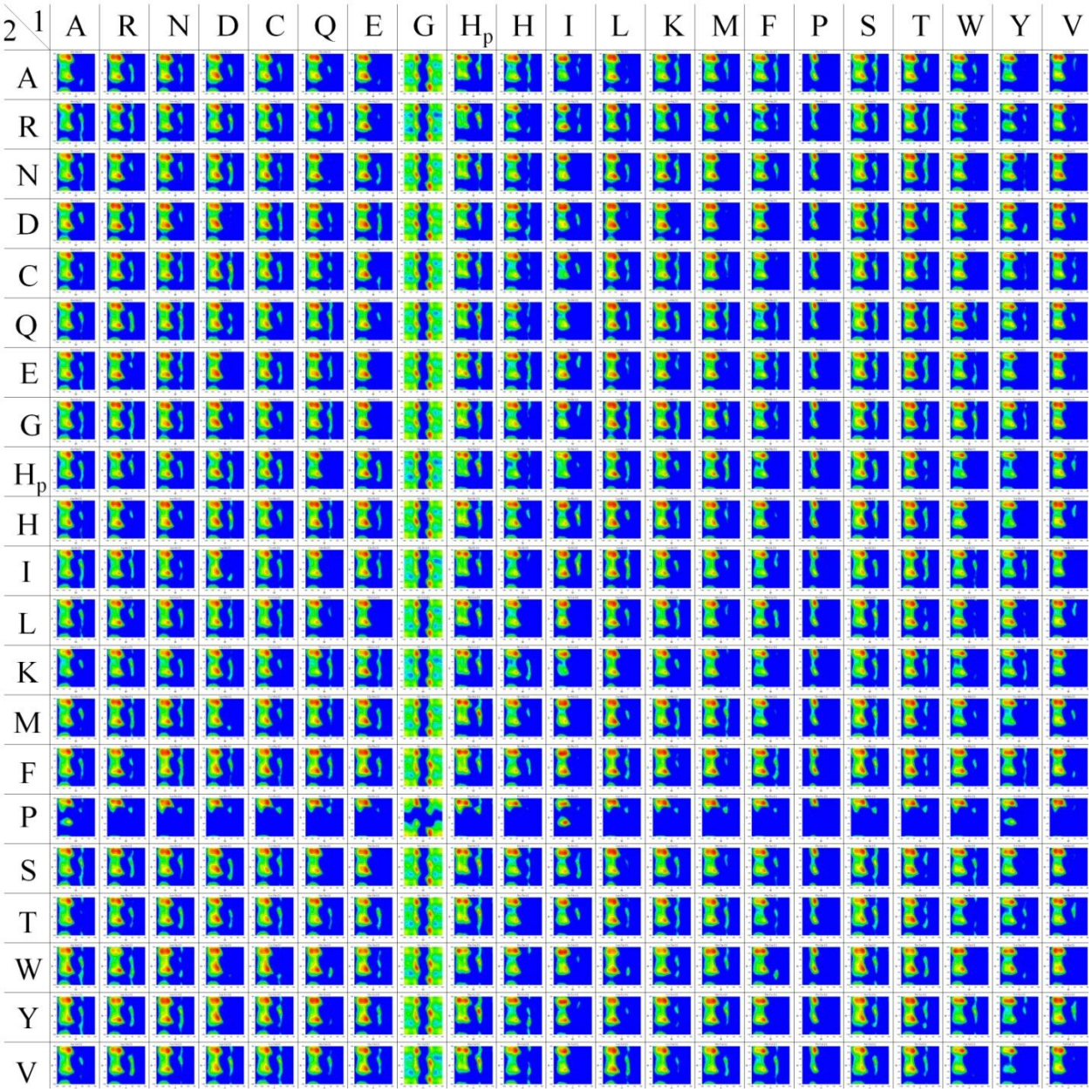

Figure S4

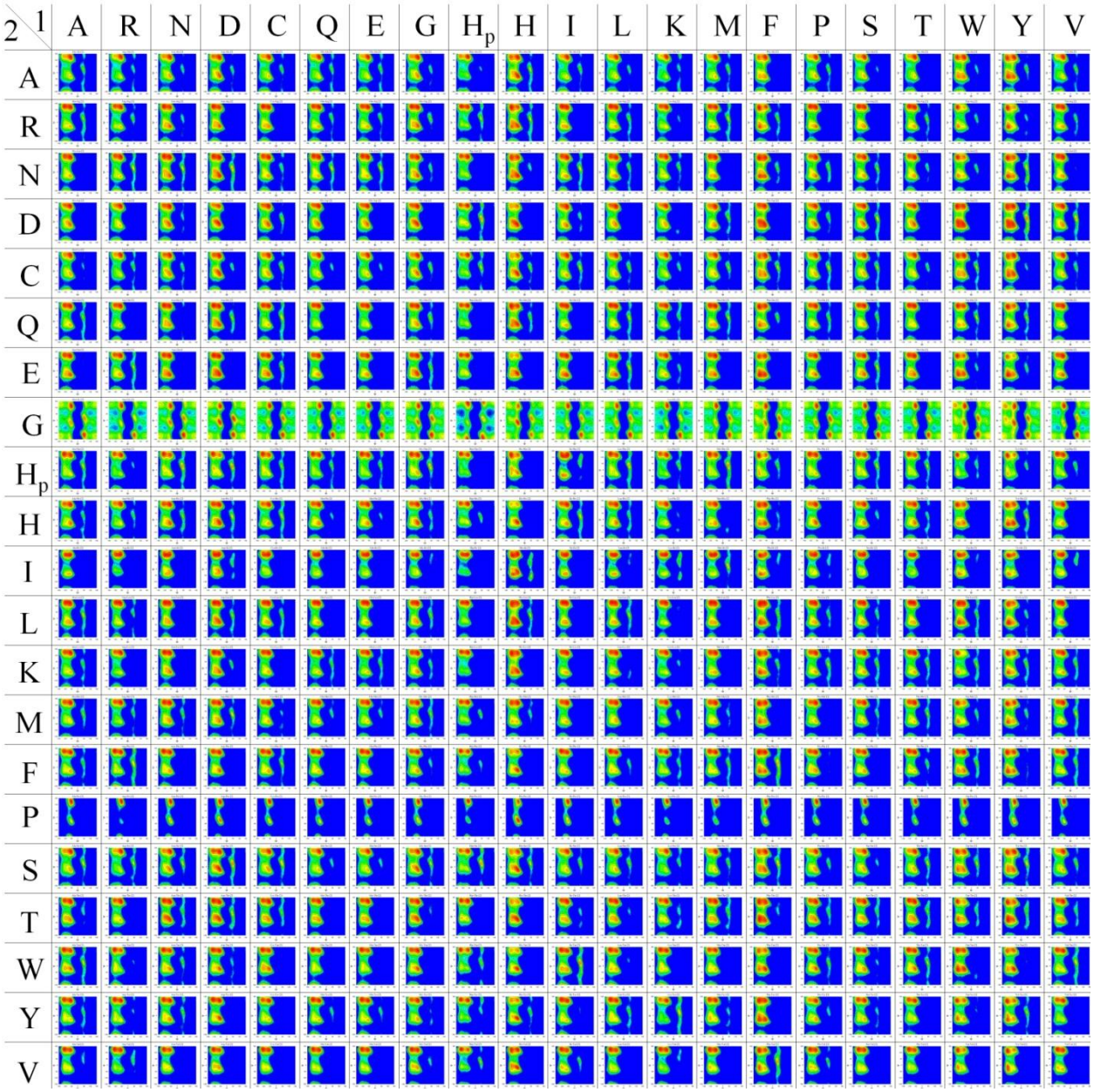

Figure S5

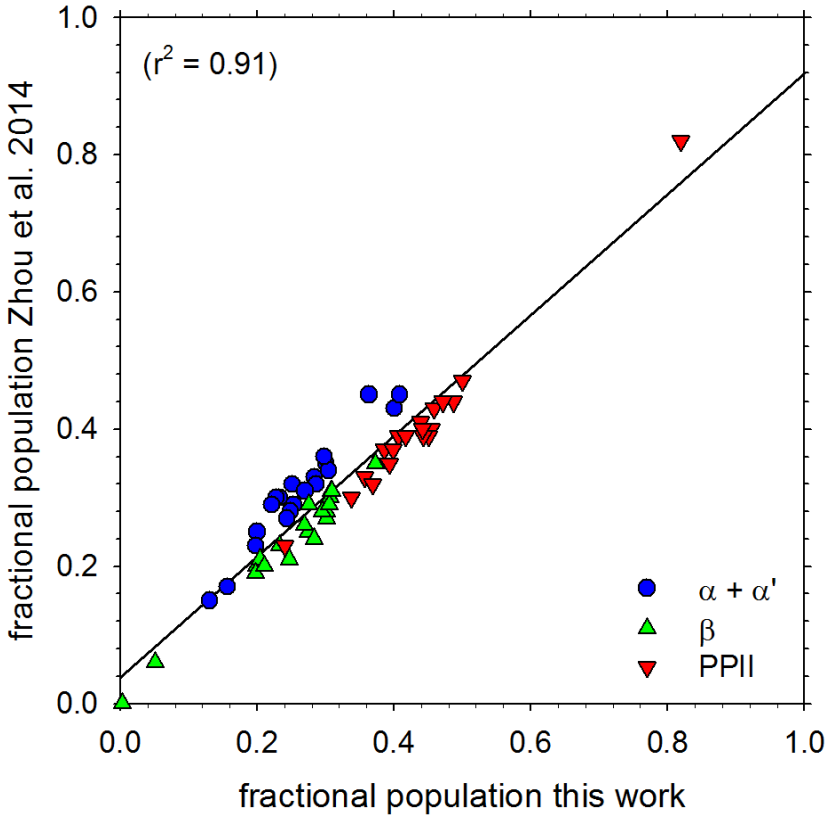
(r2 = 0.91)
fractional population Zhou et al. 2014
fractional population this work
α + α'
β
PPII
0.0
0.2
0.4
0.6
0.8
1.0

Figure S6

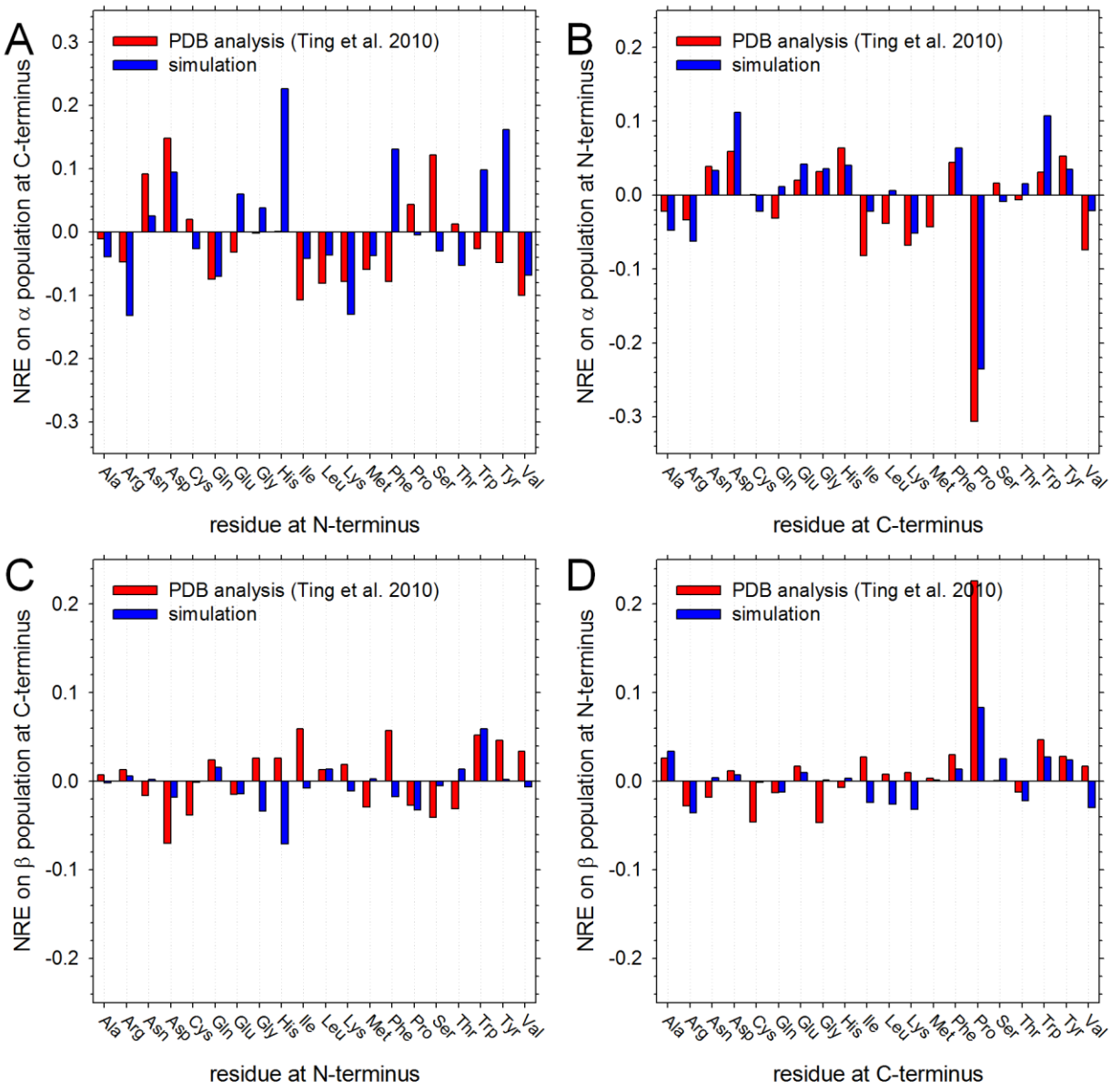
A
PDB analysis (Ting et al. 2010)
simulation
NRE on α population at C-terminus
residue at N-terminus
B
PDB analysis (Ting et al. 2010)
simulation
NRE on α population at N-terminus
residue at C-terminus
C
PDB analysis (Ting et al. 2010)
simulation
NRE on β population at C-terminus
residue at N-terminus
D
PDB analysis (Ting et al. 2010)
simulation
NRE on β population at N-terminus
residue at C-terminus
Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val

Figure S7

Table S1 **Residue-specific Karplus parameters derived by optimizing agreement between simulation and experiment for two-residue peptides.** Values are given for each of the parameters A, B, C found in the Karplus equation (see main text); θ in all cases was set to 60°. Standard deviations are from three independent repeats of the optimization procedure: each started with initial values of A, B, and C taken from ref. 17 of the main text. Note that no parameters are reported for Asp, Glu or Gly given the uncertainties in connecting simulation with experiment for these amino acids (see main text).

Amino acid	A (Hz)	B (Hz)	C (Hz)
Ala	8.09 ± 0.11	0.91 ± 0.13	1.97 ± 0.04
Arg	7.23 ± 0.02	-0.20 ± 0.01	2.13 ± 0.00
Asn	4.99 ± 0.12	-0.66 ± 0.17	3.77 ± 0.06
Cys	6.78 ± 0.02	-0.74 ± 0.02	2.69 ± 0.00
Gln	5.06 ± 0.01	-1.04 ± 0.00	2.89 ± 0.01
Hip	3.41 ± 0.10	-2.08 ± 0.00	3.71 ± 0.01
Ile	9.10 ± 0.01	0.07 ± 0.01	1.48 ± 0.01
Leu	12.14 ± 0.01	5.31 ± 0.42	2.93 ± 0.12
Lys	2.52 ± 0.31	-5.54 ± 0.02	1.17 ± 0.01
Met	7.35 ± 0.01	-0.48 ± 0.37	2.07 ± 0.14
Phe	7.87 ± 0.22	-0.65 ± 0.01	1.75 ± 0.00
Ser	4.17 ± 0.01	-3.21 ± 0.01	2.27 ± 0.01
Thr	5.44 ± 0.07	-2.05 ± 0.11	2.64 ± 0.04
Trp	8.43 ± 0.01	-0.32 ± 0.00	1.14 ± 0.01
Tyr	11.30 ± 0.03	3.38 ± 0.01	2.59 ± 0.01
Val	9.48 ± 0.01	0.33 ± 0.02	1.13 ± 0.02