Supporting Information: The acetylation landscape of the H4 histone tail: disentangling the interplay between the specific and cumulative effects.

David Winogradoff,[†,§] Ignacia Echeverria,[‡,§] Davit A. Potoyan,[¶] and Garegin A. Papoian[∗,‡]

Chemical Physics Program, University of Maryland, College Park, MD 20742, USA, Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA, and Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA

E-mail: gpapoian@umd.edu

Simulation details: Each H4 histone tail was solvated in a cubic water box of approximately $57 \times 57 \times 57$ Å^3 water box, ensuring a minimum buffer length of 12 Å to the box boundaries, equivalent to approximately three hydration shells between the protein and any edge of the cubic box. We introduced Na$^+$ and Cl$^-$ ions to neutralize the charge of the

---

[∗]To whom correspondence should be addressed
[†]Chemical Physics Program, University of Maryland, College Park, MD 20742, USA
[‡]Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, USA
[¶]Department of Chemistry and Center for Theoretical Biological Physics, Rice University, Houston, TX 77005, USA
[§]Contributed equally to this work.

peptide and represent the 0.150 M NaCl environment. The system was minimized using the steepest descent and conjugated gradient methods for 2500 steps each; first performed on the solvent and then on the entire system. Periodic boundary conditions were employed throughout all the simulations, and long-range electrostatics were treated with the Particle Mesh Ewald method, with a charge grid spacing of $\sim$1.0 Å, cubic spline interpolation, and a direct sum tolerance of 10^{-5}.[1] Non-bonded Coulomb and Lennard-Jones interactions were truncated at 12 Å, and all bonds involving hydrogen were constrained using the SHAKE algorithm.[2] After minimization, the systems were heated to 300 K by 200 ps of protein-restrained NVT MD simulation followed by 300 ps of unrestrained NVT MD simulation. In this simulations we used the Langevin thermostat with a collision frequency of 2.0 ps^{-1} Following, the systems were equilibrated at 300 K for 1.5 ns in the NPT ensemble. System pressures were regulated using isotropic position scaling and the Berendsen barostat with a relaxation time of 2 ps.

Our replica exchange protocol is included in the Methods section of the main text, and repeated here for completeness. After minimization and equilibration, we performed replica exchange dynamic (REMD)[3] simulation to characterize the conformational ensemble of the H4 histone tail at different levels of acetylation. Exchanges between replicas at different temperatures enhances the conformational sampling relative to standard MD simulation, creating an ensemble that includes both high and low energy configurations. First, each system was copied to generate a total of $\sim$ 60 replicas. The temperatures used in REMD simulations, ranging from 300 K to 450 K, were determined by T-REMD,[4] an REMD temperature online server, with a target exchange probability of 30%. Then, each replica was heated to the desired temperature over 500 ps in the NVT ensemble. REMD production runs were performed in the NVT ensemble, attempting exchanges every 5 ps with a 2 fs time-step, saving coordinates and energies every picosecond for further analysis. 100 ns of REMD simulations were performed for each system using the Langevin thermostat with a 2-ps time constant, totaling 6 μs of simulation each. The exchange probability observed for

each system was $\sim 50\%$. For analysis we only considered only the final 90 ns of trajectories set to 300 K. This allowed us to account for further thermal equilibration. To assess the convergence of the production simulations at 300 K, we analyzed the probability distributions of the peptide's radius of gyration and end-to-end distance (Figure S1).

Pairwise RMSD and Q analysis: To characterize the heterogeneity of the unfolded ensemble we considered the pairwise RMSD and Q between all structures in simulated trajectories. By this approach we built a histogram of the RMSD or Q for all the pairs of conformations sampled. The mean and standard deviation of those distributions provide a fingerprint of the conformational ensemble heterogeneity. For details on how to calculate the pairwise Q to characterize the heterogeneity of the conformational ensemble see.[5,6] Results are shown in figure S2.

Salt bridges analysis: We identified salt bridges between the one negatively charged aspartic acid residue and the positively charged arginines. An arginine and the aspartic acid were considered to form a salt bridge if the distance between the arginine center of positive charge, CZ, and one of the negatively charged oxygen atoms of aspartic acid, OD1 or OD2, was less than 4.0 Å. Results are shown in figure S9.

Table S1: Average R_g and helical propensity for all levels of acetylation.

| System	Ac level	$\langle R_g \rangle$ (Å)	Helical $(3_{10} + \alpha)\%$
H4-WT	0	9.7 ± 1.3	5.2 ± 0.4
H4-K16$_{ac}$	1	10.0 ± 0.8	5.0 ± 0.4
H4-K5$_{ac}$	1	9.4 ± 0.8	7.7 ± 1.5
H4-K8$_{ac}$K16$_{ac}$	2	9.5 ± 0.8	11.4 ± 2.2
H4-K5$_{ac}$K8$_{ac}$	2	9.4 ± 0.8	9.2 ± 0.4
H4-K8$_{ac}$K12$_{ac}$K16$_{ac}$	3	9.4 ± 0.8	9.7 ± 0.6
H4-K5$_{ac}$K8$_{ac}$K12$_{ac}$K16$_{ac}$	4	9.3 ± 0.8	10.5 ± 0.7
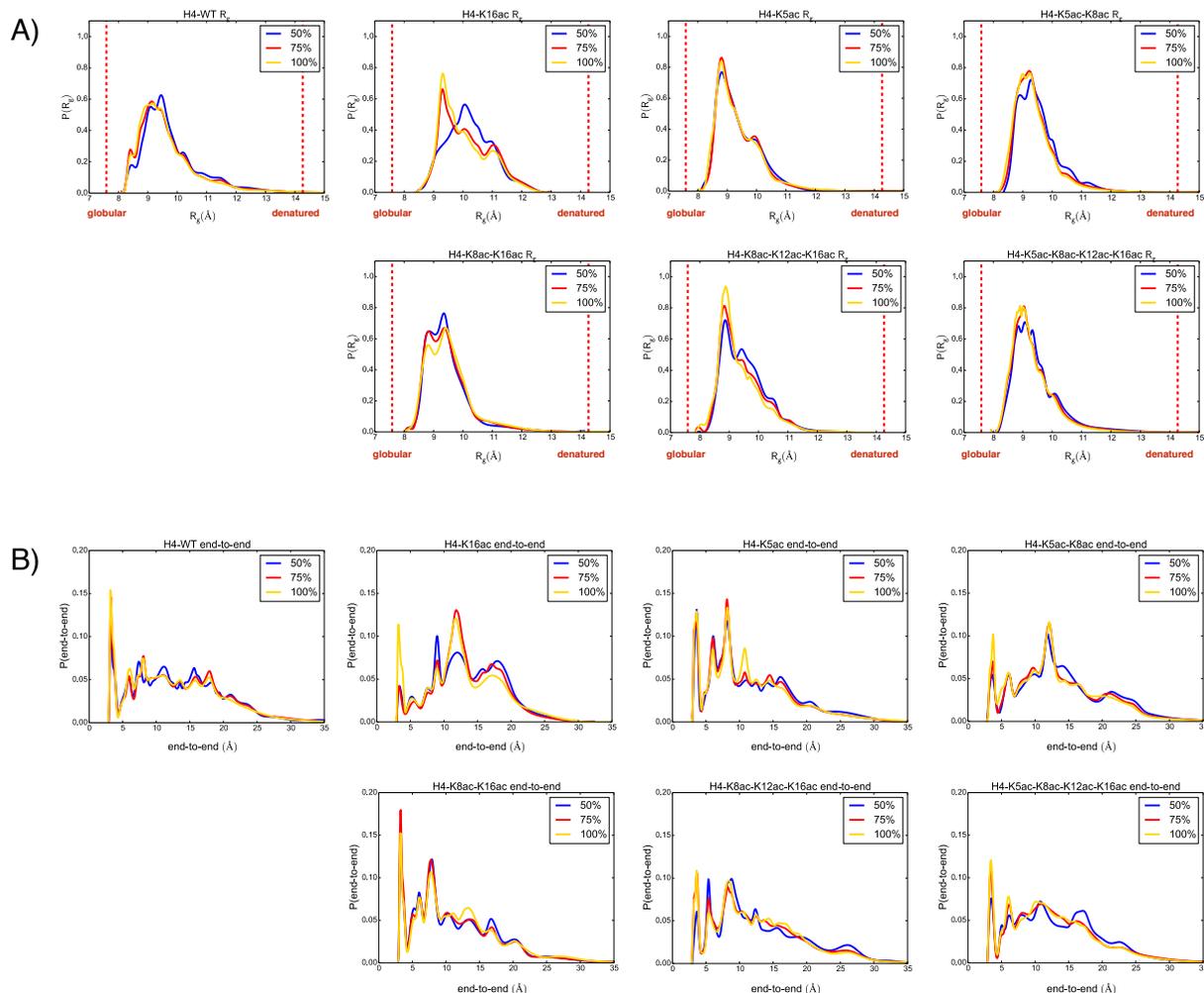
Figure S1: **Convergence of the replica exchange simulations.** A) Radius of gyration probability distributions obtained considering the full REMD at 300K trajectories (100%) and data sets with a fraction of the trajectories (75% and 50%). In all cases we considered the first X% of the trajectories. B) End-to-end distance probability distributions obtained considering the full REMD at 300K trajectories (100%) and data sets with a fraction of the trajectories (75% and 50%). We observe that for data sets where the trajectories longer than 75% of the trajectories, the differences in the sampled conformations are negligible.
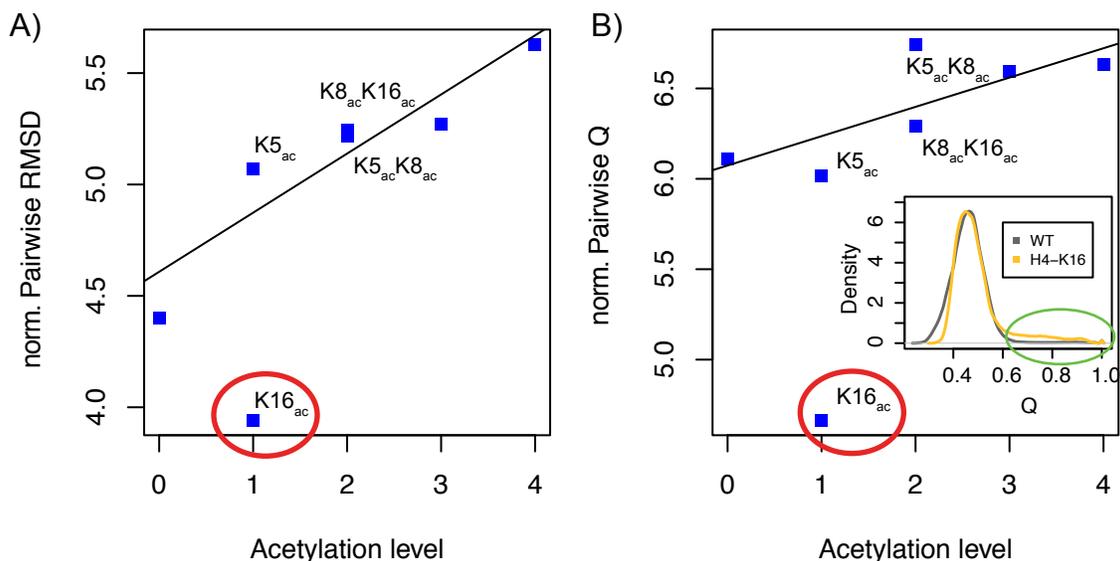
Figure S2: **Heterogeneity of the conformational ensemble** A) Mean pairwise RMSD among all sampled structures normalized by the standard deviation. B) Mean pairwise Q among all sampled structures normalized by the standard deviation. Linear fits where done for all cases, except the H4-K16$_{ac}$ model. Using both metric we determined that the H4-K16$_{ac}$ system exhibits the least conformational heterogeneity.
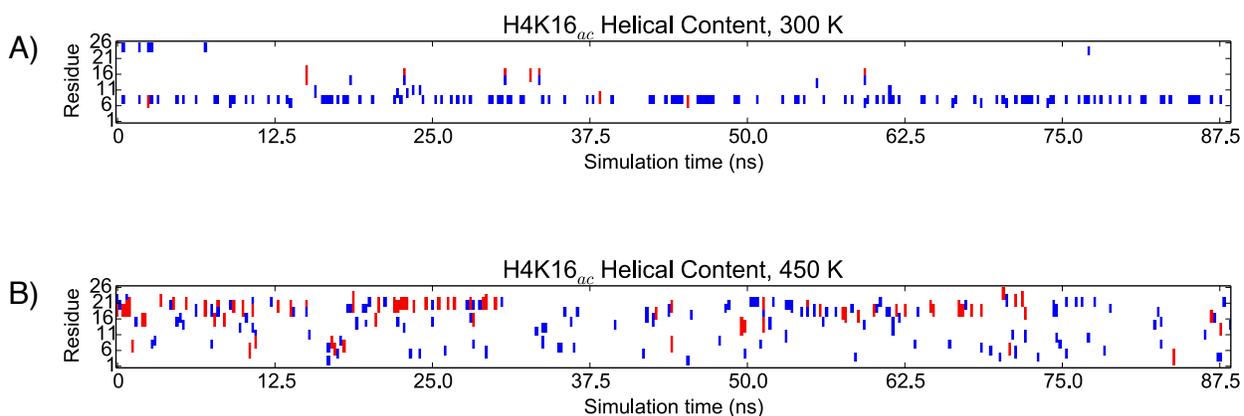

Figure S3: 3_{10} (blue) and α (red) helical content of the H4-K16$_{ac}$ tail by residue, as a function of simulation time for temperature trajectories at A) 300 K, and B) 450 K.
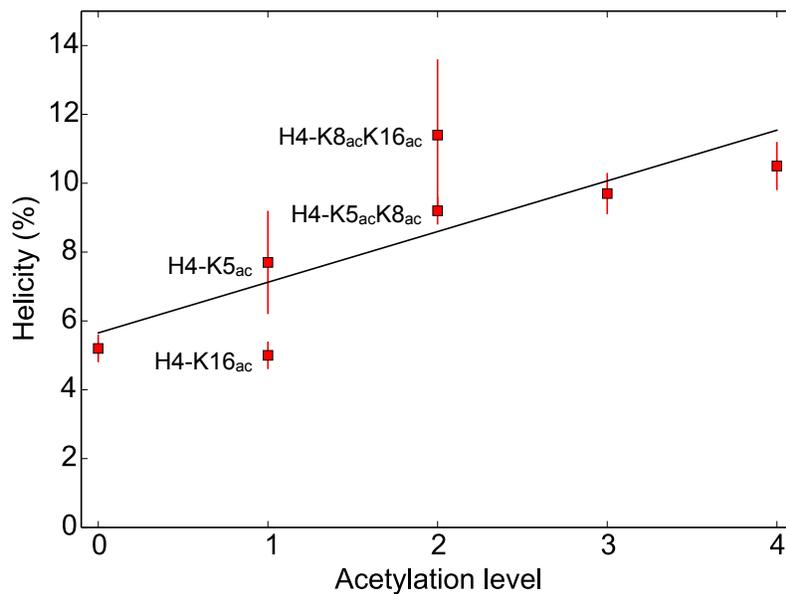
Figure S4: **Percentage of helical residues.** Helical content of the histone H4 tail as a function of the number of acetylated lysines.
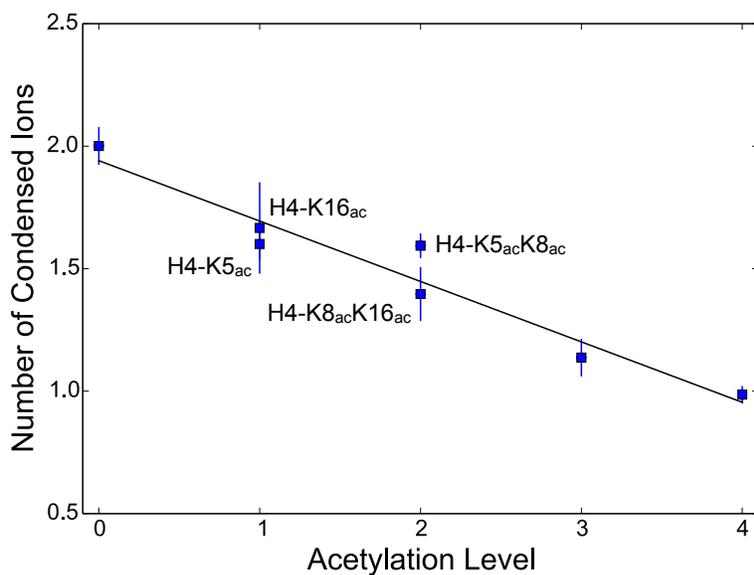

Figure S5: **Ion Association.** Average number of condensed anions (Cl^-) around the positively charged H4 tail as a function of acetylation. A chloride ion is considered to be associated with the H4 tail if it is within the Bjerrum length (7.5 Å) of the head group nitrogen of an arginine or lysine.
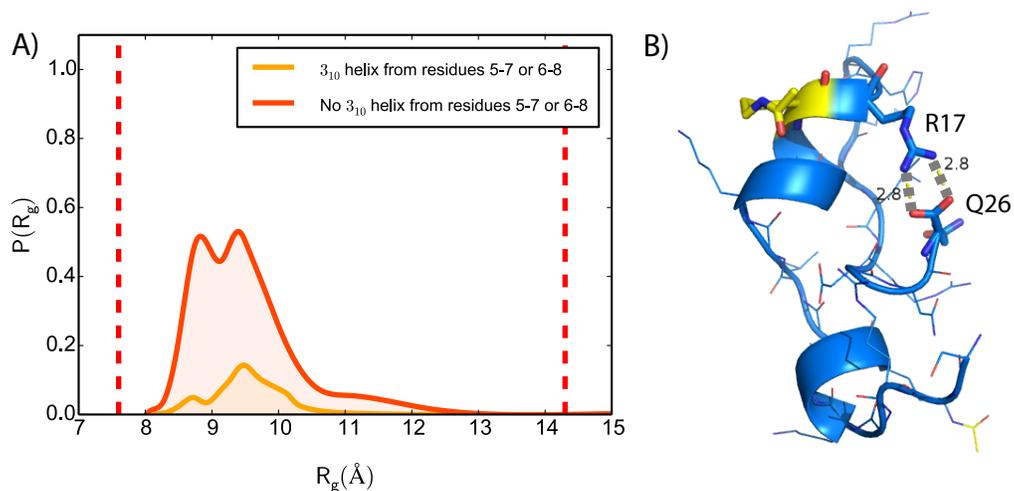
Figure S6: **The H4-K8$_{ac}$K16$_{ac}$ model.** A) The R_g of the H4-K8$_{ac}$K16$_{ac}$ di-acetylated model where the simulations frames were divided into two groups: (1) simulation frames with a 3$_{10}$ helix formed from residues 5 to 7 or 6 to 8, and (2) simulation frames without these specific helices. B) Characteristic structure obtained from the third most populated cluster exhibiting an elongated conformation. The acetylated K16 is shown in yellow.
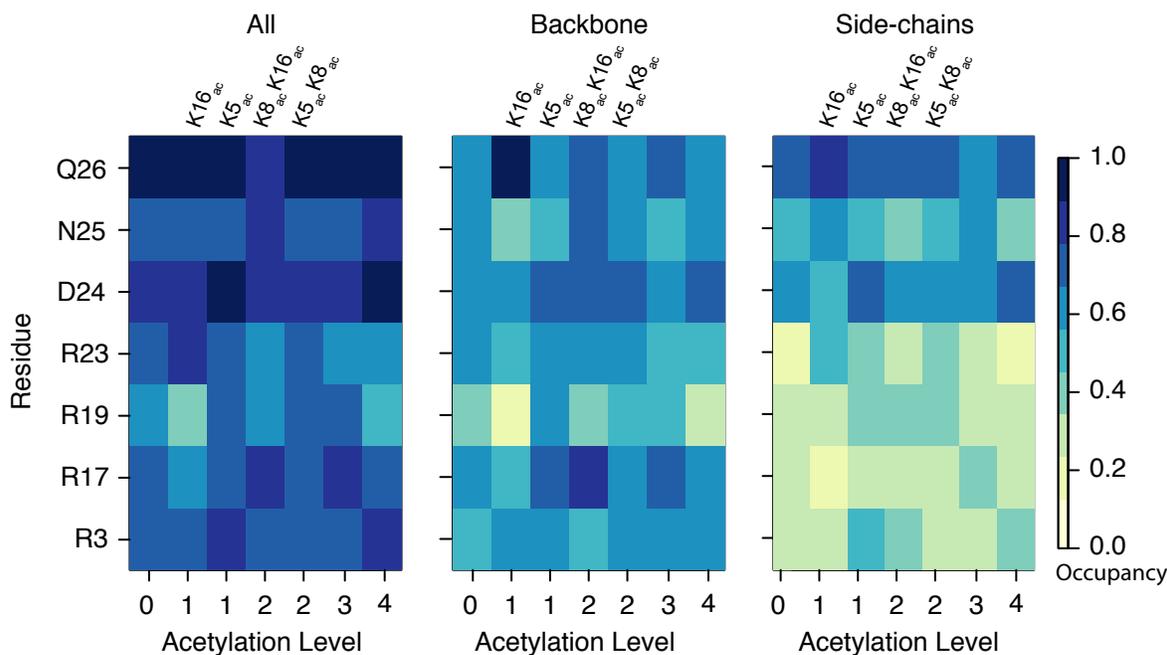

Figure S7: **Hydrogen bonds occupancies.**
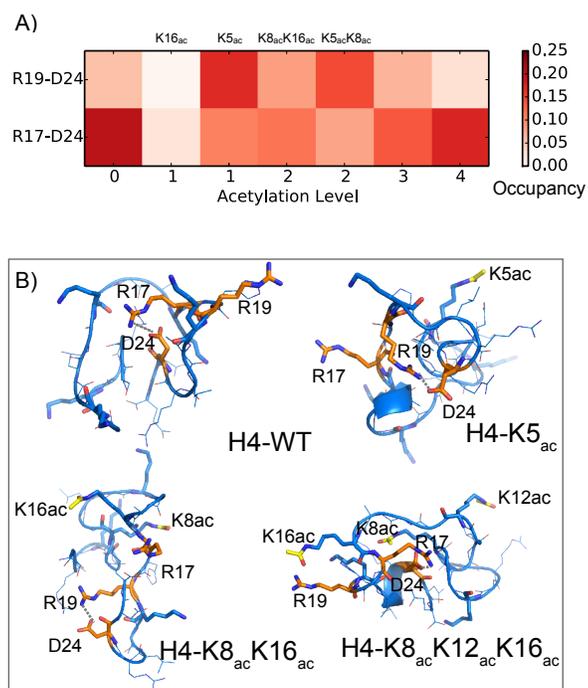
Figure S8: **Stabilizing interactions and salt bridge occupancies.** A) Salt bridge occupancies between the arginines (R17 or R19) and the aspartic acid (D24) at different levels of acetylation, measured as the percentage of time that the salt bridges are formed. B) Sample structures are different levels of acetylations where salt-bridges stabilize the structures. Residues R17, R19 and D24 are shown in orange. Acetylations are shown in yellow.
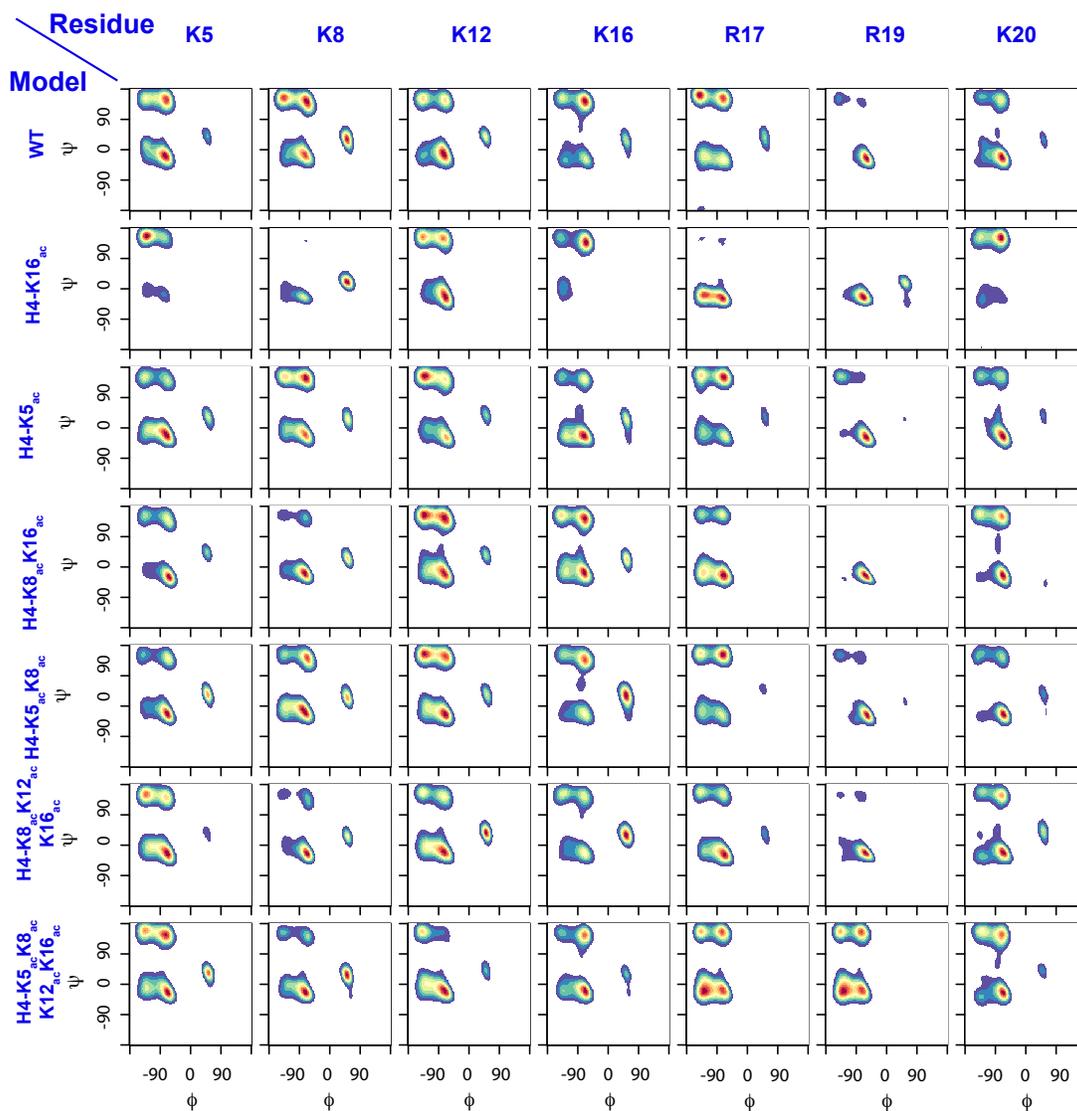
Figure S9: **Ramachandran plots for selected residues.**

References

(1) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(2) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. *J. Comput. Phys.* **1977**, *23*, 327–341.

(3) Sugita, Y.; Okamoto, Y. *Chemical Physics Letters* **1999**, *314*, 141–151.

(4) Patriksson, A.; van der Spoel, D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073–7.

(5) Potoyan, D. A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17857–62.

(6) Echeverria, I.; Papoian, G. A. *Israel Journal of Chemistry* **2014**, *54*, 1293–1301.