Figure S1

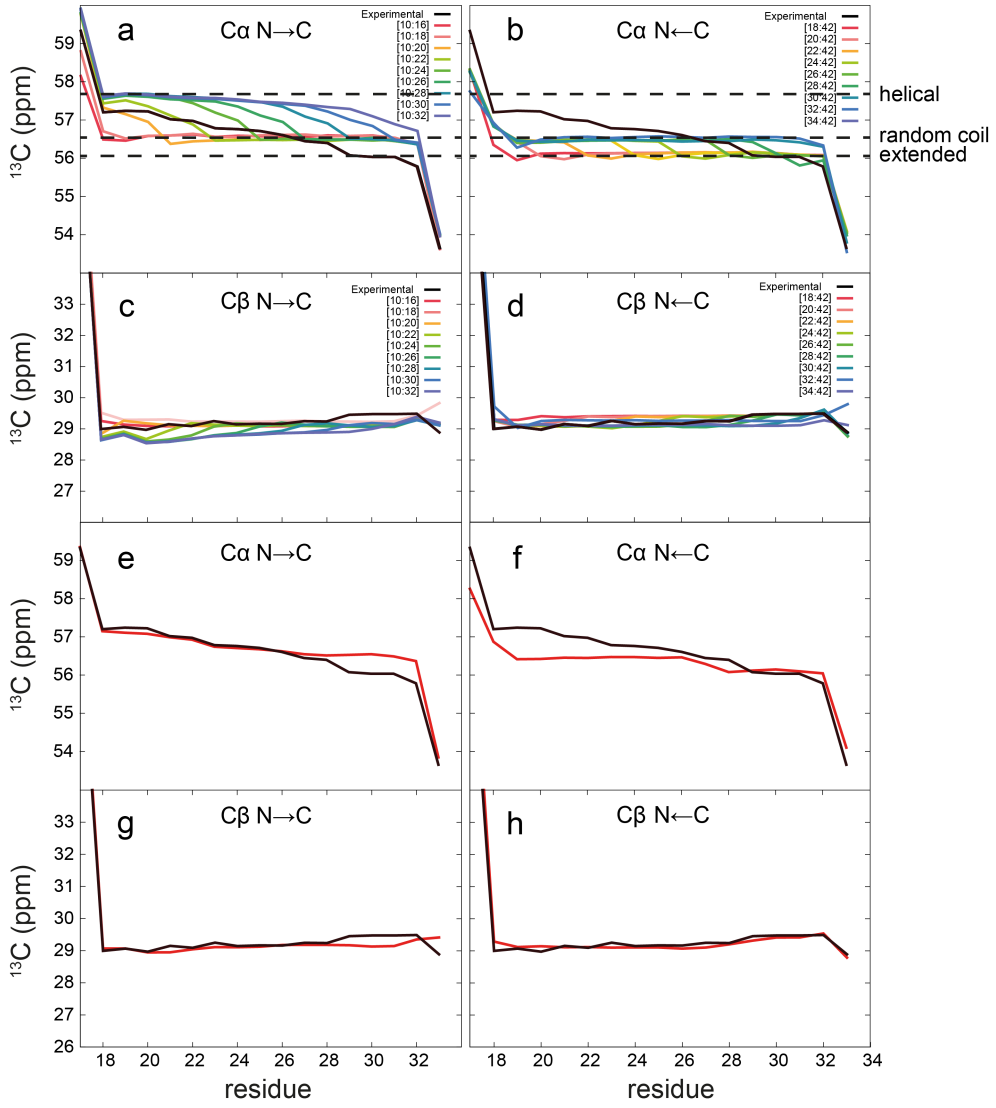

Figure S1. Chemical shift (CS) based ensemble refinement of H16. Related to Figure 3. (a-d) Overlay of the experimental (black) and theoretical (color) Cα **(a and b)** and Cβ **(c and d)** CSs of the poly-Q tract computed for the N→C **(a and c)** and the N←C **(b and d)** families of ensembles. The ensembles were built by incrementally incorporating glutamines in the partially structured region from Q18-Q33 (N→C) and from Q33 to Q18 (N←C). The boundaries chosen for the partially structured regions for each of the 5,000 conformation ensembles are indicated in the panel. Horizontal dashed lines indicate the three plateaus corresponding to helical, random coil and extended averaged conformations. Experimental (black) *vs.* ensemble-optimized (red) Cα **(e and f)** and Cβ **(g and h)** CSs exclusively fitted with the N→C **(e and g)** or the N←C **(f and h)** ensembles. For the N→C only Cα and Cβ experimental CSs from residues Q18-Q28 were used in the optimization. The reweighted model consisted of [10:17; 46.17%], [10:22; 26.93%], [10:23; 5.92%], and [10:26; 20.98%], where [X:Y] refers to the first and last residues considered as partially structured in the model. For the N←C only Cα and Cβ experimental CSs from residues Q29-Q33 were used in the optimization. The reweighted model consisted of [22:42; 46.05%], [28:42, 0.8%], [29:42, 42.86%], and [34:42, 10.28%].

Figure S2

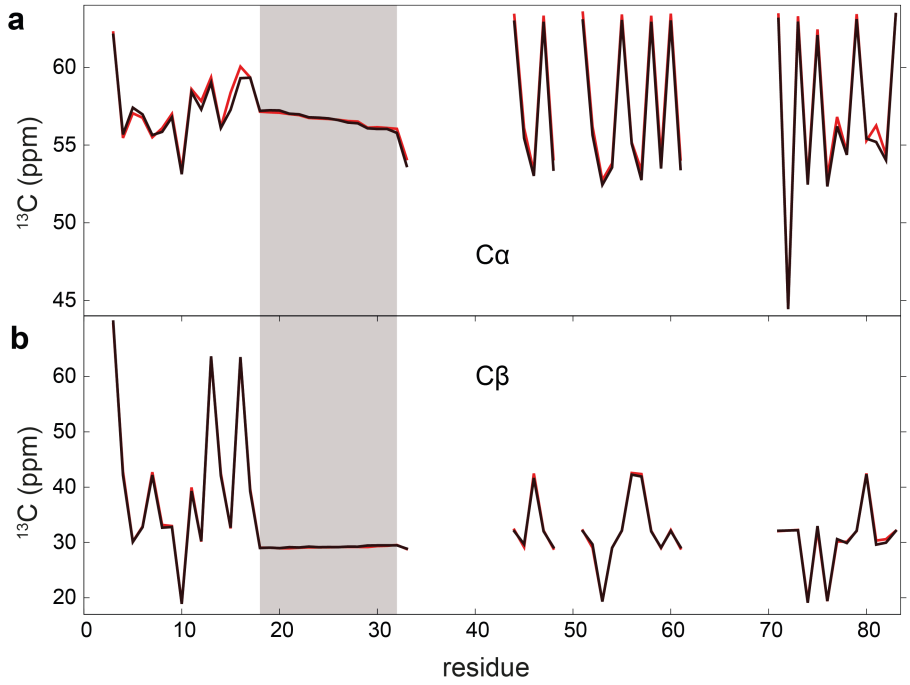

Figure S2. Chemical shift (CS) based ensemble refinement of H16. Related to Figure 3. Experimental (black) *vs.* ensemble-optimized (red) for all **(a)** Cα and **(b)** Cβ CSs measured for H16. The M1-Q28 and the Q29-P83 fragments of the optimized profile were built from those optimized using the N→C and N←C ensembles, respectively. The poly-Q tract is shaded in gray; gaps are due to proline residues for which we do not have experimental data.

Figure S3

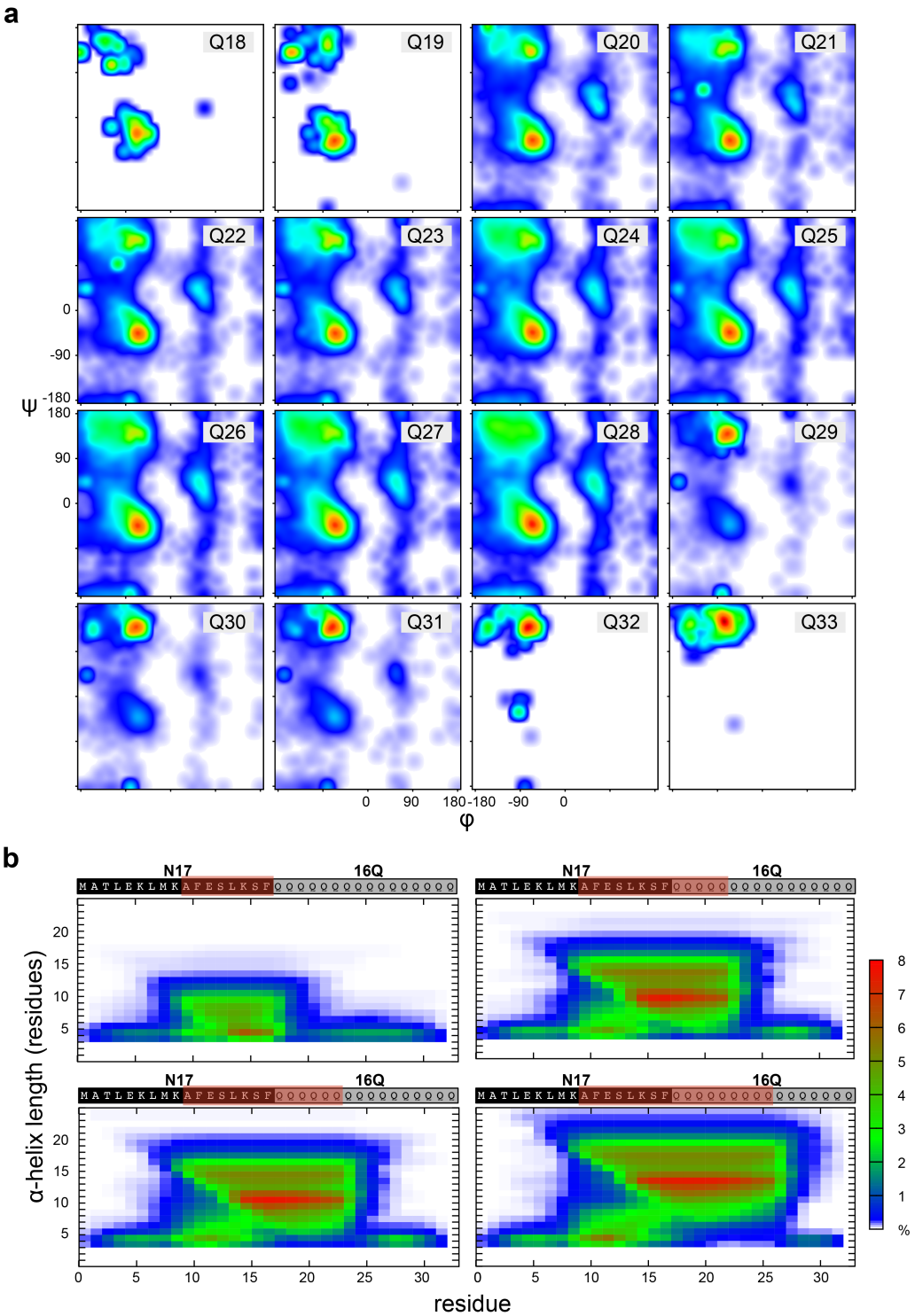

Figure S3. Ramachandran plots and SS-maps. Related to Figure 3. (a) Ramachandran distribution for the glutamine residues of the poly-Q tract of H16 derived from the CS-optimized ensemble of H16. The populations of α-helical ($0 > \varphi$; $50 > \psi > -120$), extended ($0 > \varphi$; $-120 > \psi > 50$), and *other* ($0 < \varphi$) conformations are reported in Figure 3c in the main text. The red color indicates a higher density of conformations. **(b)** SS-maps for the four ensembles selected using the experimental CSs. Fragments considered as partially structured are highlighted in red on the top of each panel. The percentage of the α-helical fragments is indicated in a color scale from high (red) to low (blue) populations.

Figure S4

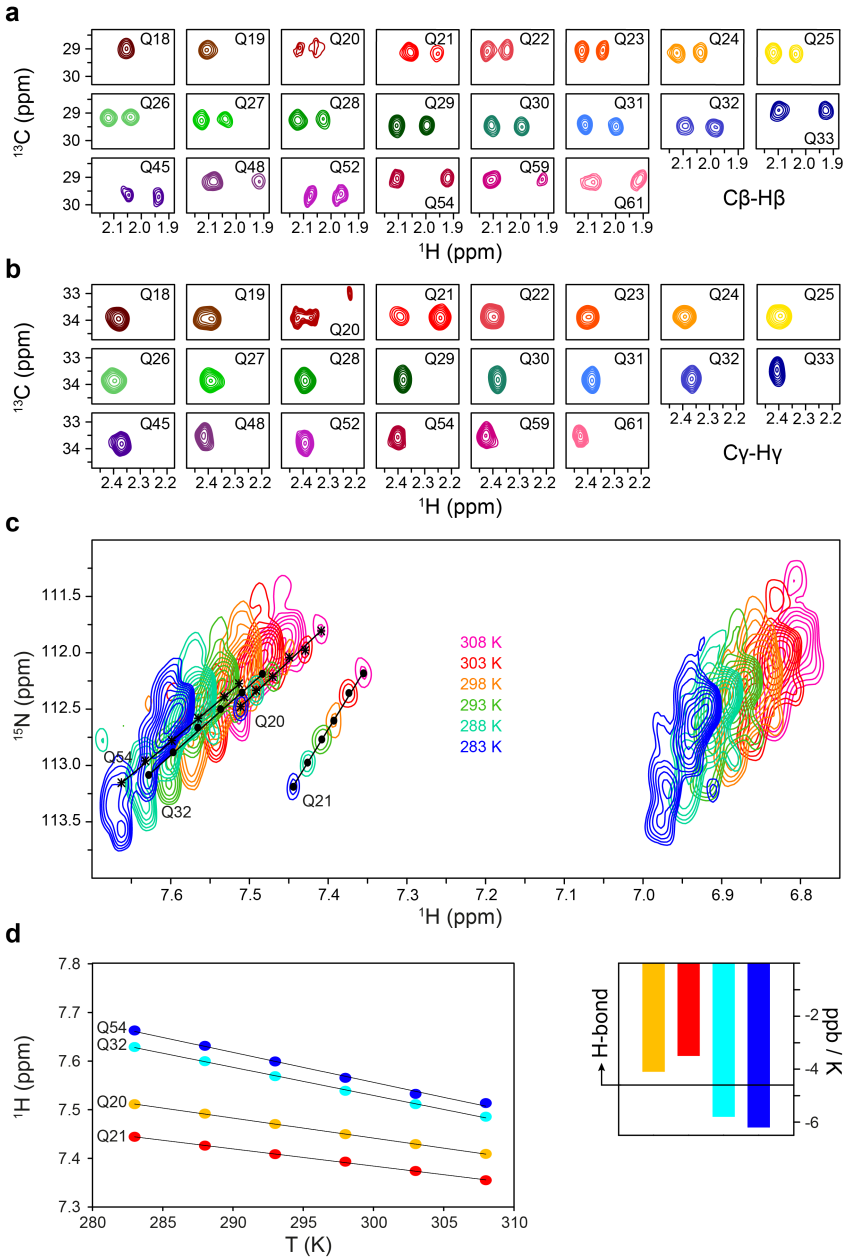

Figure S4. Side chain NMR scanning and temperature factors. Related to Figure 4. (a) Cβ and **(b)** Cγ ^{13}C-HSQC spectra for all glutamines in H16. With the exception of the first four glutamines (Q18-Q21), both families of spectra display a canonical behavior where Cβ-H_2 and Cγ-H_2 are doublets and singlets, respectively. The color code is equivalent to the one used in Figure 1 in the main text. **(c)** Zoom of the Nε-H_2 region of the ^{15}N-HSQC of fully labeled H16 measured at different temperatures. Solid lines connect the centers of the peaks for Q20, Q21, Q32 and Q54 at the different temperatures. Other peaks could not be unambiguously identified and therefore were not used in the analysis. **(d, left panel)** Linear fit of the ^{1}H frequency of the Nε-H_2 peaks of the four residues plotted against the temperature. The slope of the linear fit **(d, right panel)** reports on the temperature coefficient and the probability of the atom to be involved in a hydrogen bond. According to these slopes, Q20 and Q21 side chains form a hydrogen bond, while Q32 and Q54 side chains do not.

Figure S5

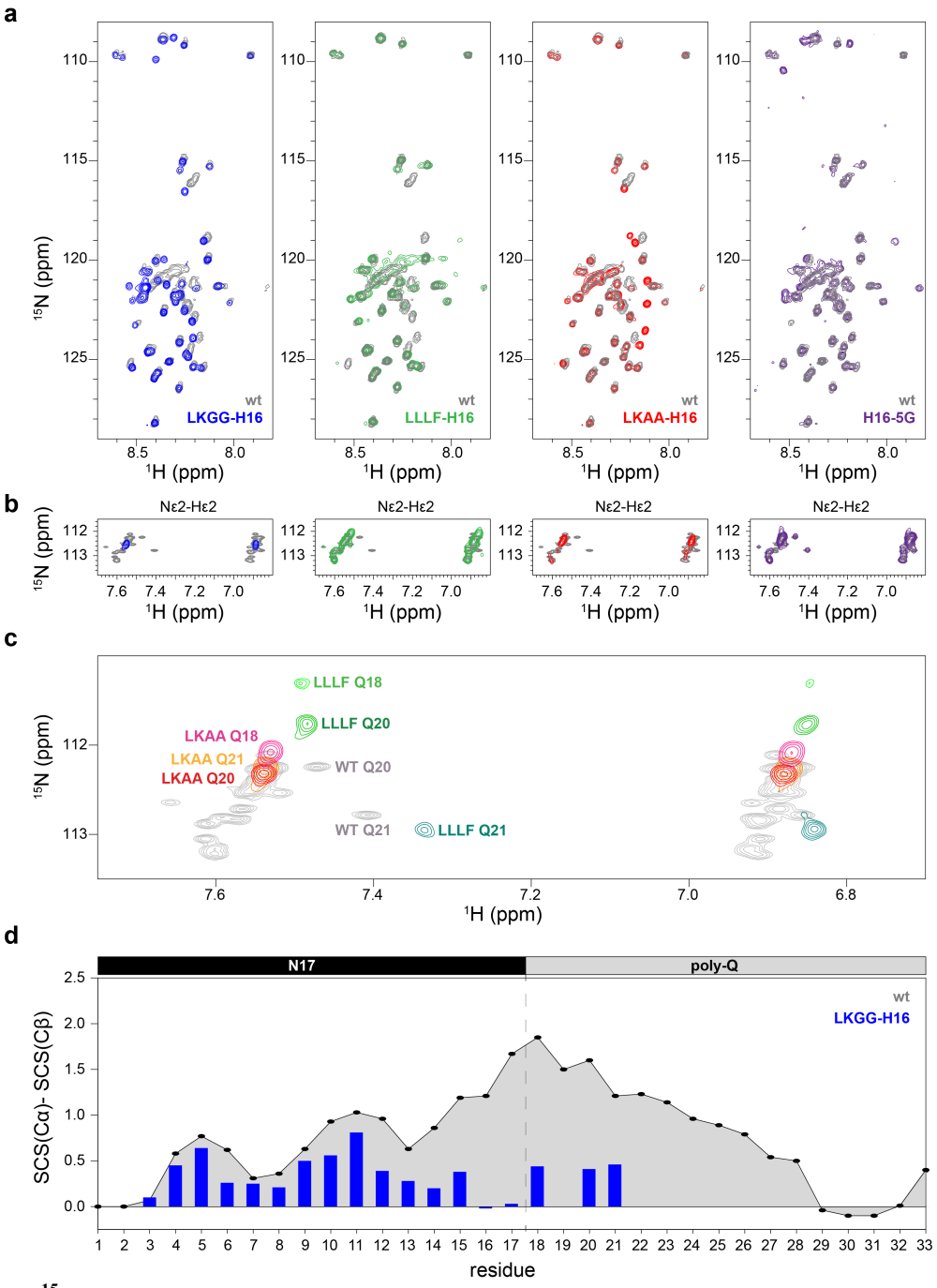

Figure S5. ^{15}N-HSQC spectra of the structural mutants and SCS analysis of LKGG-H16 in comparison with the wild-type. Related to Figure 5. (a) Backbone spectra of wild-type H16 (gray) overlaid with LKGG-H16 (blue), LLLF-H16 (green), LKAA-H16 (red) and H16-5G (purple). **(b)** Side chain spectra corresponding to the same mutants using the same color code. Note that peaks corresponding to the first glutamines for the LLLF-H16 and LKAA-H16 are not displayed due to the lower intensity (see below). **(c)** Zoom of the Nε-H_2 peaks for residues Q18, Q20 and Q21 of mutants LLLF-H16 and LKAA-H16 obtained from SSIL samples in comparison with the same region of the wild-type protein in gray. Their chemical shifts substantiate the hydrogen bond network involving F17 in httex1 (see main text). **(d)** Secondary chemical shift analysis using experimental Cα and Cβ chemical shifts and a neighbor-corrected random-coil library POTENCI for the wild-type (connected black points) and the LKGG-H16 mutant (blue histogram). Only residues belonging to N17 and the poly-Q tract are displayed. Data from Q18, Q20 and Q21 was obtained from SSIL samples; the other glutamines were not investigated.

Figure S6

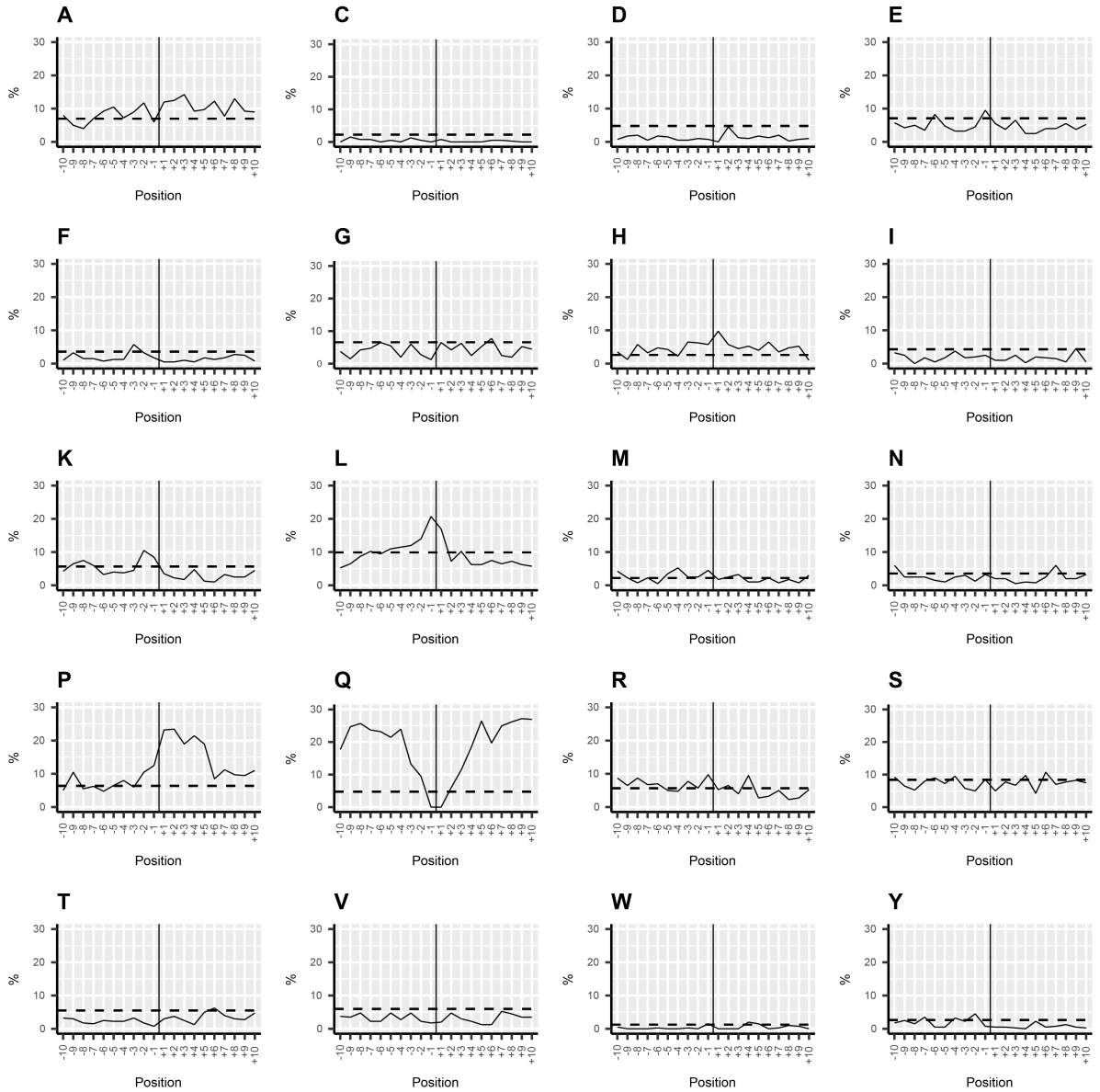

Figure S6. Compositional analysis of the poly-Q flanking regions in human proteins. Related to Figure 6. Percentage for each one of the 20 natural amino acids in the positions preceding (-10 to -1) and following (+1 to +10) the poly-Q tracts in human proteins. The solid vertical line corresponds to the position of the poly-Q tract. Poly-Q tracts were defined as having a maximum of 2 non-glutamine residues in fragments of 10 or more glutamine residues. Horizontal dashed lines define the percentage for each amino acid type in the human proteome.