Biophysical Journal, Volume 97

Supporting Material

Thermodynamics of Beta-Sheet Formation in Polyglutamine

Andreas Vitalis, Nicholas Lyle, and Rohit V. Pappu

SUPPLEMENTARY MATERIAL

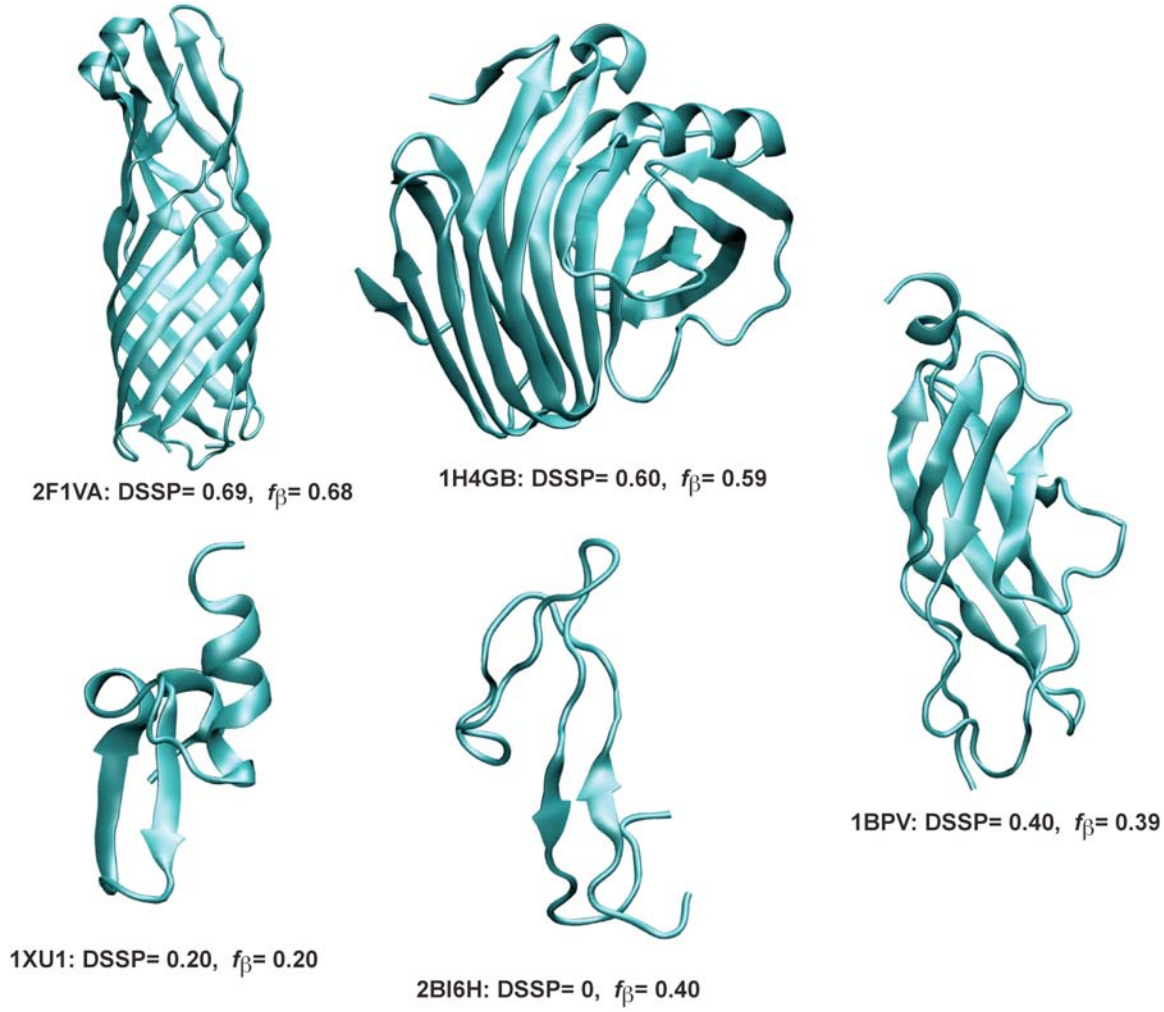

Figure S1: Representative structures from the PDBSelect database showing the correlation between f_β values and fractional DSSP-E scores. For each structure, the label denotes the PDB code with chain ID, fractional DSSP-E score, and f_β value, respectively. The correlation becomes weak when characteristic hydrogen-bonded patterns are absent within a structure as shown for 2BI6H.

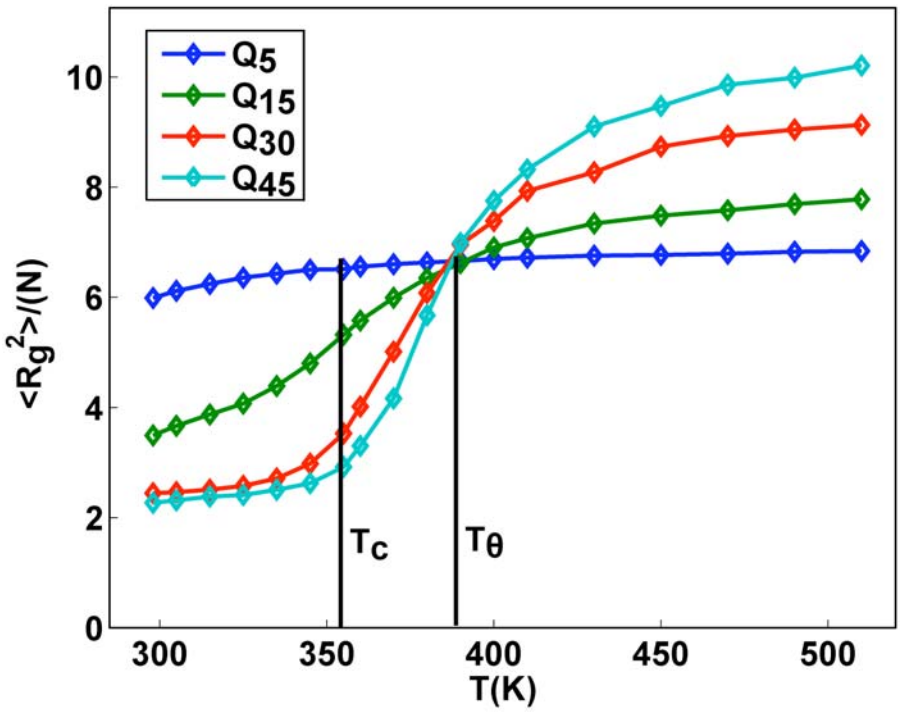

Figure S2: Coil-to-globule transition for monomeric polyglutamine. These data show the ensemble averaged mean square radius of gyration as a function of simulation temperature. Values on the ordinate are normalized by chain length because $\langle R_g^2 \rangle \propto N$ at the theta temperature, T_θ. Therefore, the plots for different chain lengths should coincide at $T=T_\theta$ for different chain lengths because polyglutamine molecules are homopolymers. This requirement is satisfied for $T=T_\theta \approx 390$K. The plots were made by analyzing data extracted from torsional space Metropolis Monte Carlo (MC) simulations using the forcefield and ABSINTH implicit solvation model described in the Materials and Methods section of the text. Details of the move sets used are provided in Table S1. Independent simulations were carried out for each temperature. For each temperature, the simulations involved 10^6 equilibration steps of MC followed by 4×10^7 steps of production. A complete analysis of the coil-to-globule transition for monomeric polyglutamine and an explanation for the shapes of these curves has been provided in previous work (1). $T=T_C \approx 355$K is the temperature below which collapsed states are favored for polyglutamine. Hence, for the forcefield used in this work, $T \leq T_C$ corresponds to poor solvent conditions, whereas the regime $T_C < T \leq T_\theta$ corresponds to the transition region.

Table S1: Overview of the frequency of the different Monte Carlo moves sets used in simulations of monomeric and pairs of polyglutamine molecules. To be able to probe multiple length scales simultaneously, Monte Carlo moves in ABSINTH either fully randomize a given degree of freedom, or perform a stepwise perturbation that has a maximum size. This is explained in detail in the footnotes. The frequencies for different moves are chosen to reflect the relevance of the various degrees of freedom to both the conformational equilibria and the association of these peptides. Additionally, these choices reflect the associated computational cost. As an example, ω-angles are sampled relatively infrequently, as their values are expected to remain close to the perfect *trans*-conformation. Note that a small number of moves for each simulation were swap attempts for replica exchange.

Move Type	**Frequency of move sets for simulations of monomeric polyglutamine**	**Frequency of move sets for simulations with pairs of polyglutamine molecules**
Rigid-body[1]	-	30% (50%, 10Å, 20°)
Omega[2]	7% (90%, 5°)	4.9% (90%, 5°)
Sidechain[3]	30% (4x, 60%, 30°)	21% (4x, 40%, 30°)
Pivot[4]	63% (70%, 10°)	44.1% (70%, 10°)

1. Rigid-body moves simultaneously change rotational and translational degrees of freedom of the whole molecule. The first value listed in parentheses is the fraction of moves assigned to finite perturbations, whereas the remaining attempts fully randomize the respective degrees of freedom. The second and third values are the maximum translational and rotational step-sizes associated with the finite perturbations.
2. Moves that perturb the ω-angles of peptide units. Due to the acetyl and methylamide capping groups there are N+1 ω-angles for a chain length of N. The two sets of values in parentheses are the fraction of ω-moves, which attempt a stepwise perturbation along with the maximum step-size.
3. Sidechain moves perturb the χ-angles of a given sidechain in the peptide. In each attempt to alter sidechain degrees of the freedom, two of the three χ-angles are randomly altered. Sidechain moves are inexpensive and therefore several sidechains are sampled during each "move" (first value in parentheses). The remaining two values in parentheses again give the fraction of χ-moves with a finite perturbation and the maximum value of that perturbation.

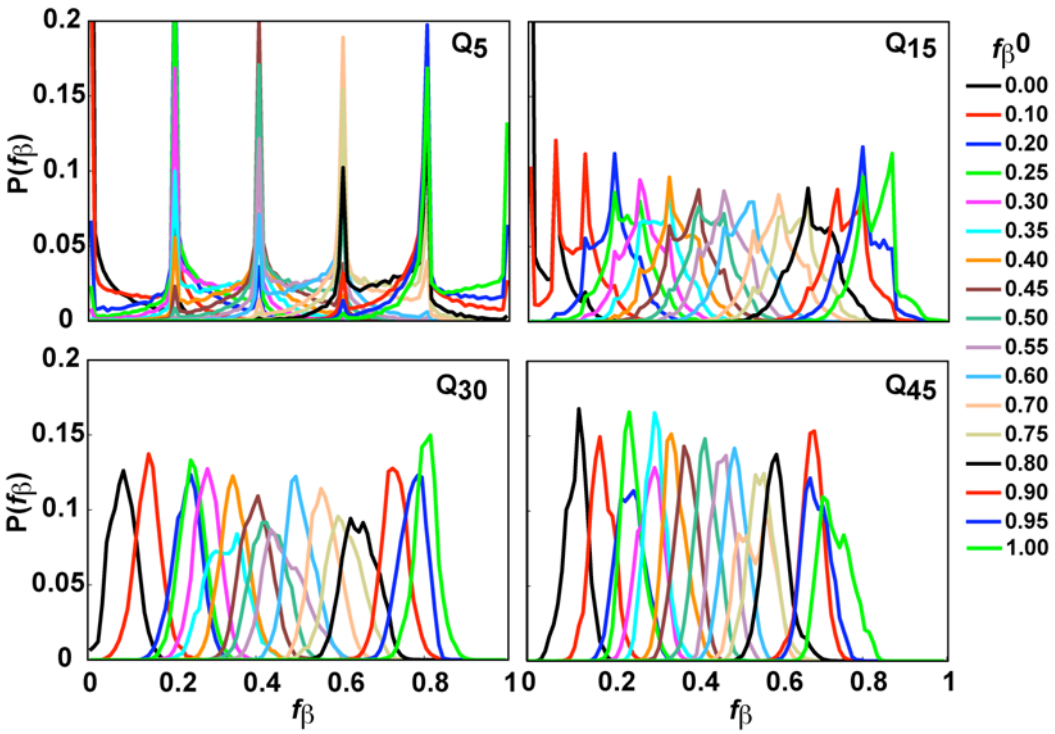

Figure S3: Distributions of f_β values obtained from the umbrella sampling simulations. The legend on the right identifies the target f_β^0 value for each simulation. Each histogram is the average from three independent simulations. The histograms indicate that the overlap between adjacent windows is finite and significant, thereby validating the f_β schedule and choice of restraints used in the umbrella sampling protocol. To further demonstrate the validity of our protocol, Figure S4 shows a summary of the overlap statistics extracted from the plots shown in this figure. As a reminder, we present data from umbrella sampling simulations that were performed using the following schedule of seventeen values for f_β^0: [0.0, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.7, 0.75, 0.8, 0.9, 0.95 1.0].

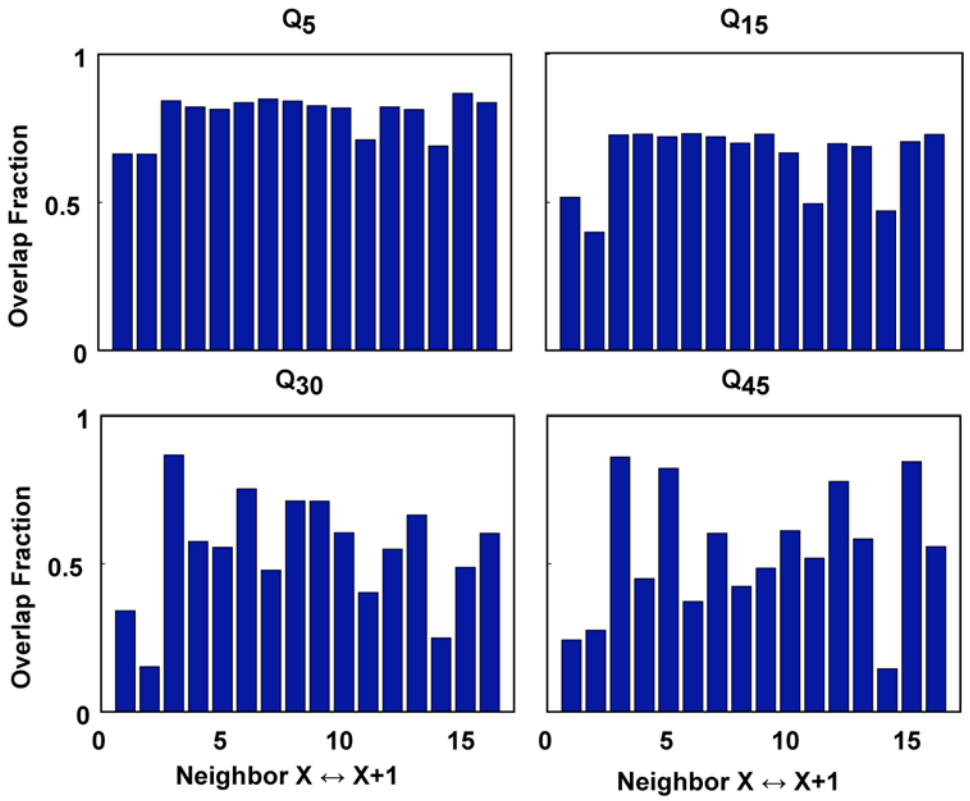

Figure S4: Quantification of overlap between adjacent f_β histograms shown in Figure S3. Sixteen bars are shown for overlap statistics computed between seventeen sets of adjacent windows. The f_β schedule is as shown in the legend to Figure S3. For a pair of windows X and X+1, $\text{Overlap Fraction} = \dfrac{2 - \int_{f_\beta=0}^{f_\beta=1} \left| P_X\left(f_\beta\right) - P_{X+1}\left(f_\beta\right) \right| df_\beta}{2}$. We computed each integral numerically. In this formula, P_X and P_{X+1} are the average f_β histograms for windows X and X+1, respectively and these histograms are shown in Figure S3. If adjacent histograms P_X and P_{X+1} overlap perfectly, then the value of the overlap fraction is unity. Conversely, if the histograms do not overlap at all, then the overlap fraction is zero. The smallest overlap fraction values are seen for two pairs of neighbors in the simulations for Q_{30} and Q_{45}. However, the reconstructed PMFs obtained using either WHAM or TI-WHAM are smooth even in regions where the overlap is the smallest, although the error bars (standard errors and bootstrap errors) are large in these high free energy regions. Overall, the quality of data obtained using the umbrella sampling protocol adopted in this work appears to be satisfactory and yields reliable PMFs thereby allowing us to draw the conclusions summarized in the main text.

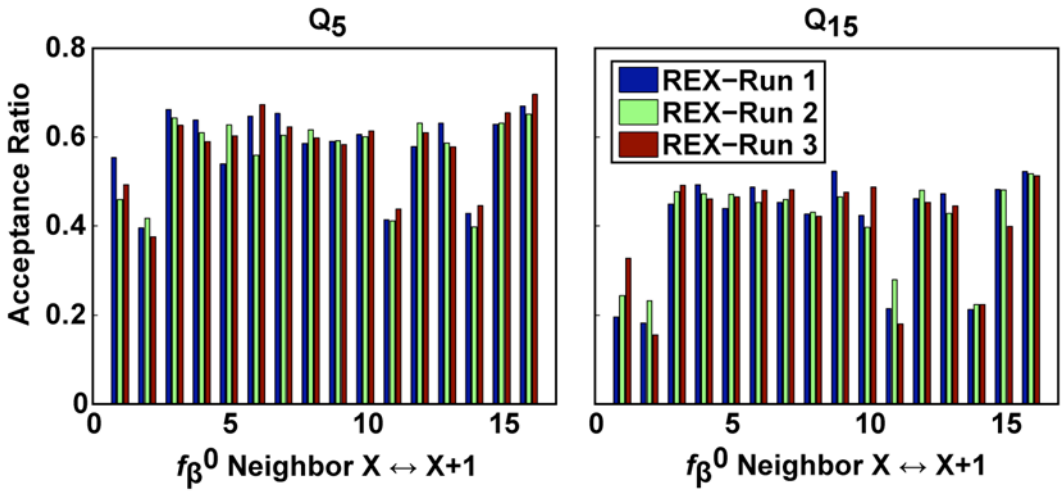

Figure S5: Plot of acceptance ratios for replica exchange swaps between nearest-neighbor windows in the umbrella sampling simulations that were carried out for monomeric polyglutamine. As noted in the text, we carried out three independent replica exchange simulations, which were combined with the umbrella sampling protocol. The f_β schedule used in these simulations is identical to that shown in the legend to Figure S3. The acceptance ratios are shown here for each of the three replica exchange (REX) runs and the data are shown for Q_5 and Q_{15}. For Q_{30} and Q_{45}, the replica exchange simulations were initially carried out using a coarse, 11-window f_β schedule. However, the final PMFs were computed using the full 17-window f_β schedule shown in the legend for Figure S3. This was accomplished by collecting data from independent umbrella sampling runs (three per additional window) for windows that were not present in the coarse schedule. The decision to use the two-tier approach for the longer chains was based on the availability of CPU resources. The coarse f_β schedule used for replica exchange plus umbrella sampling simulations for Q_{30} and Q_{45} is as follows: [0.0, 0.1, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8, 0.9, 1.0].

Calculation of PMFs for monomeric polyglutamine using WHAM with standard errors: As noted in the materials and methods section, we collected data from umbrella sampling simulations for each target f_β^0 value. For Q_5 and Q_{15}, we used umbrella sampling combined with replica exchange based on the 17-window f_β schedule shown in the main text. Data were collected from three independent replica exchange plus umbrella sampling simulations. For Q_{30} and Q_{45}, we combined umbrella sampling with replica exchange using a coarser 11-window f_β schedule and we carried out three independent simulations as well. The target f_β^0 values for each window in the coarse schedule were set to be: [0.0, 0.1, 0.25, 0.3, 0.4, 0.5, 0.6, 0.75, 0.8, 0.9, 1.0]. We subsequently added data from three independent simulations for each of the windows that were missing in the coarse schedule and were present in the finer schedule. In this way, we ended up with three independent f_β histograms (one from each run) for each of the seventeen windows shown in the Materials and Methods section. Each of three sets of 17 independent histograms were used in the WHAM analysis to construct three independent, unbiased PMFs. Panel A in Figure 3 shows the average PMF constructed from the three independent estimates. The error bars in this figure are standard errors across the three independent estimates. The naïve standard errors represent the upper limit on the errors in our WHAM estimates of the PMFs.

Calculation of PMFs for monomeric polyglutamine using WHAM with bootstrap errors: For each window (with 17 windows in total), we have three independent simulations and for each of these simulations we recorded the f_β value every 5,000 MC steps. There were 4×10^7 production steps for Q_5,Q_{15}, and Q_{30}; therefore, we collected 8,000 f_β values for each window and for each independent simulation. For Q_{45} each run consisted of 6×10^7production steps, and we collected 12,000 f_β values for this chain. The f_β values samples were pooled across all independent simulations for each window. Thus, for each window characterized by a target f_β^0 value, the dataset contained 24,000 f_β samples for Q_5, Q_{15}, and Q_{30} and 36,000 f_β samples for Q_{45}. From this pooled dataset, we selected 800 f_β samples at random with replacement. This procedure was carried out for each window. These f_β values were binned to create an f_β histogram for each of the 17 windows with target f_β^0 values. Each histogram was created using 100 bins with bin widths of 0.1. This yielded a histogram for each of the 17 f_β windows. The standard WHAM analysis was applied to these constructed histograms to generate a PMF. This protocol of generating 17 sets of histograms using random sampling with replacement combined with WHAM analysis was repeated 200 times to generate 200 independent PMFs. Panel B in Figure 3 shows the average PMF (averaged over 200 trials) and bootstrap error, which is simply the standard error computed over the 200 unbiased PMFs.

Calculation of PMFs for monomeric polyglutamine using TI-WHAM with standard errors: For each independent replica exchange plus umbrella sampling simulation, we calculated $\partial E_i/\partial f_\beta^0{}_i$. This is the partial derivative of the system energy (including the restraint potential) for window i with respect to f_β^0. These statistics were gathered every thousand steps during the production run (4×10^7 for Q_5,Q_{15},Q_{30} and 6×10^7 for Q_{45}). Values for $\langle \partial E_i/\partial f_{\beta i} \rangle$ are numerically integrated using a cumulative trapezoidal integrator. This yields estimates for the free energy differences between all pairs of adjacent windows. These relative free energies were then fed into the WHAM analysis to combine f_β histograms for each window into a single unbiased probability distribution. The use of the TI step circumvents the iterative step in WHAM and

allows us to analyze the PMF that results from an independent method for assessing relative free energies between windows. For the TI-WHAM procedure, we used data based on the coarse (as opposed to fine) f_β schedule. This allowed us to query the dependence of the reconstructed PMFs on the coarseness of the f_β schedule (in addition to the circumvention of the first, iterative step in WHAM). For each chain length, the TI-WHAM procedure was applied to each of the three independent datasets. The PMF shown in Panel C of Figure 3 is an average over the three runs and the error bars are standard errors.

Analysis of thermal replica exchange data for dimers of polyglutamine: Energy values are collected every 5,000 steps. From the production runs we obtained 7,900 energy values for Q_5, Q_{15}, and Q_{30} 12,000 samples for Q_{45}. We constructed histograms of energy values using placed 200 bins (with bin widths that vary with chain length). All error bars were standard errors computed across three independent runs.

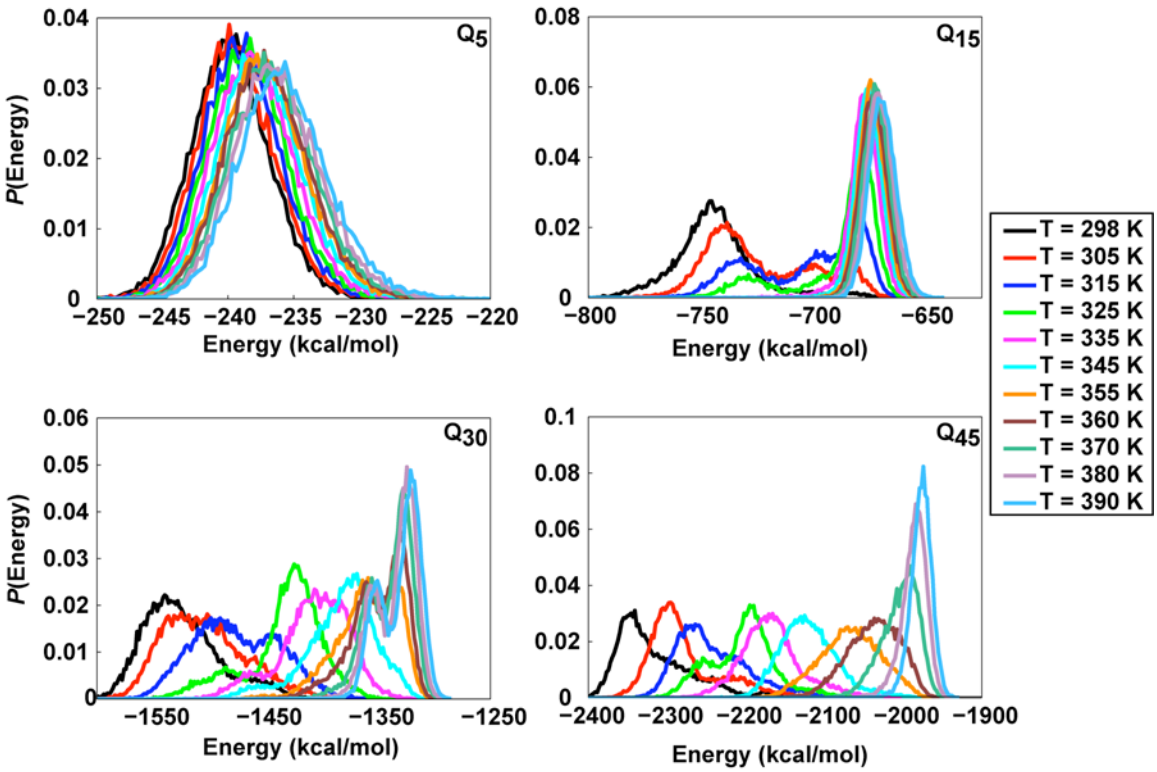

Figure S6: Temperature-dependent energy histograms for dimers of polyglutamine molecules where both chains are restrained to a target value of f_β^0=1. The histograms demonstrate the overlap between adjacent replicates and justify the temperature schedule used in all replica exchange simulations of homodimerization. The overlap between distal replicas decreases with increasing chain length and this is consistent with the differences in thermal stabilities for associated versus dissociated dimers for different chain lengths (longer chains form associated dimers over a wider temperature range). As noted in the text, we carried out three independent thermal replica exchange runs for each chain length and target f_β value. The figure shows average histograms for each temperature. Figure S7 summarizes the overlap statistics that were calculated using the histograms shown in this Figure.

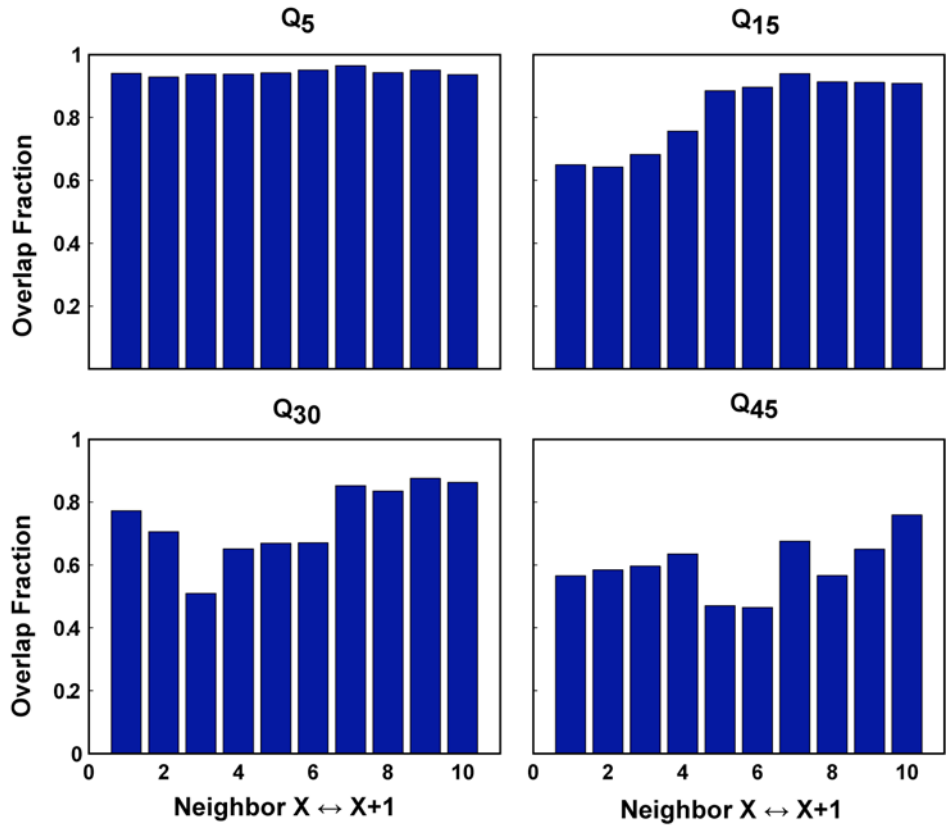

Figure S7: Quantification of the overlap between adjacent thermal replicas shown in Figure S6. Ten bars are shown for overlap statistics computed between eleven sets of adjacent windows. The temperature schedule is presented in the methods section. Computation of the overlap statistics is analogous to the method described in the caption for Figure S4. The overlap statistics appear to justify the adequacy of the temperature schedule used in the replica exchange runs.

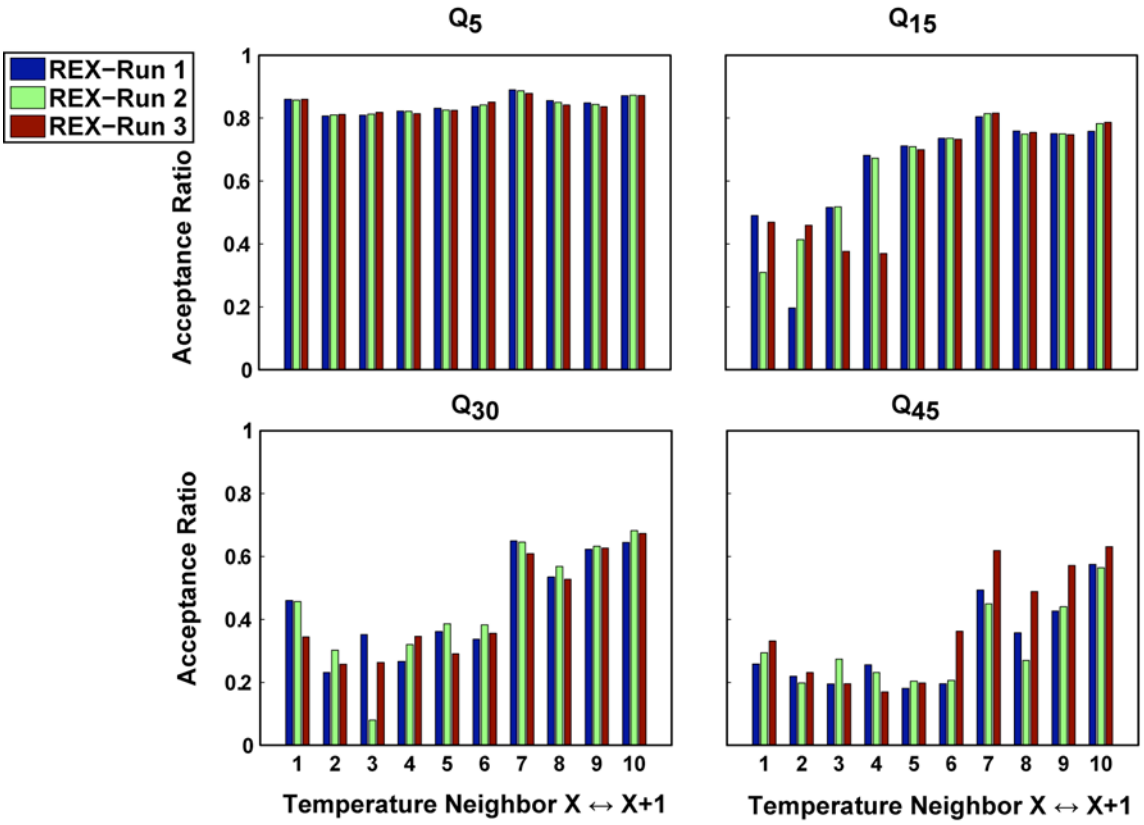

Figure S8: Plot of acceptance ratios for swaps between nearest-neighbor thermal replicas in simulations of homodimerization. Data are shown for the case where both chains are restrained to a target value of f_{β}^{0}=1. The acceptance ratios are shown here for each of the three replica exchange (REX) runs that were carried out for each peptide. Even for the longer chains, the smallest acceptance ratios are greater than 0.2 (on average) indicating that the sampling quality is not compromised.

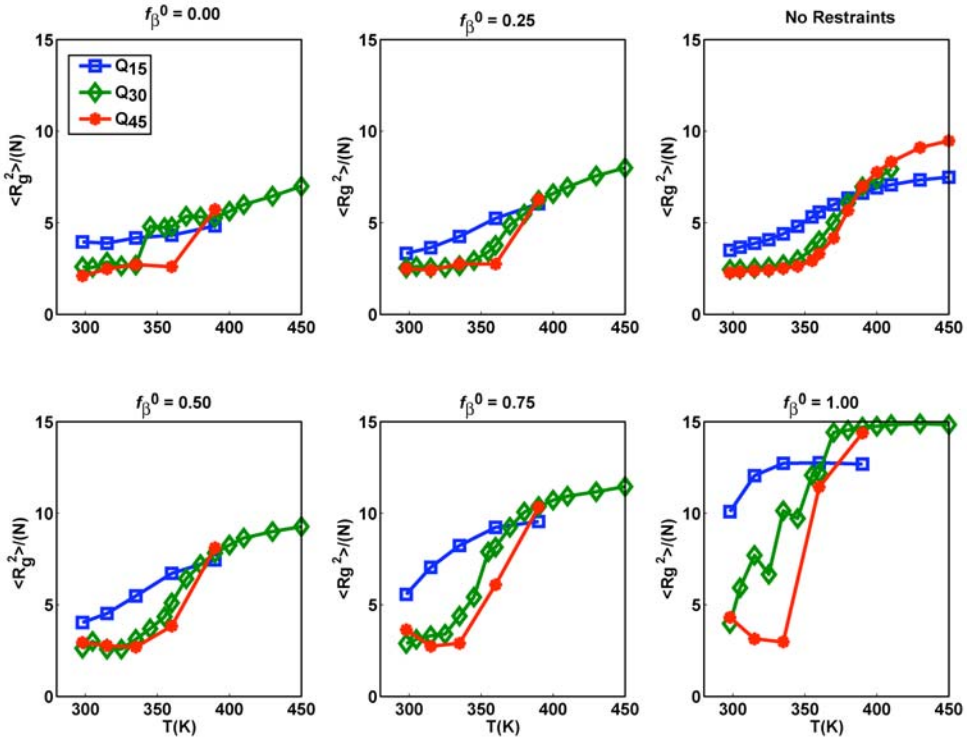

Figure S9: Quantitative analysis of the effects of conformational restraints on coil-to-globule transitions for monomeric polyglutamine of different lengths. Two trends are apparent in the data: First, the coil-to-globule transitions becomes sharper vis-à-vis the unrestrained case for target values of f_β greater than 0.5, whereas the opposite is true for $f_\beta \leq 0.25$. Second, the normalized value of $\langle R_g^2 \rangle$ remains in the vicinity of the value for the unrestrained chain for temperatures that are in the globule regime; conversely, for temperatures in the coil regime the normalized $\langle R_g^2 \rangle$ values are significantly larger than unrestrained values if $f_\beta \geq 0.75$, whereas the opposite is true for $f_\beta \leq 0.25$.

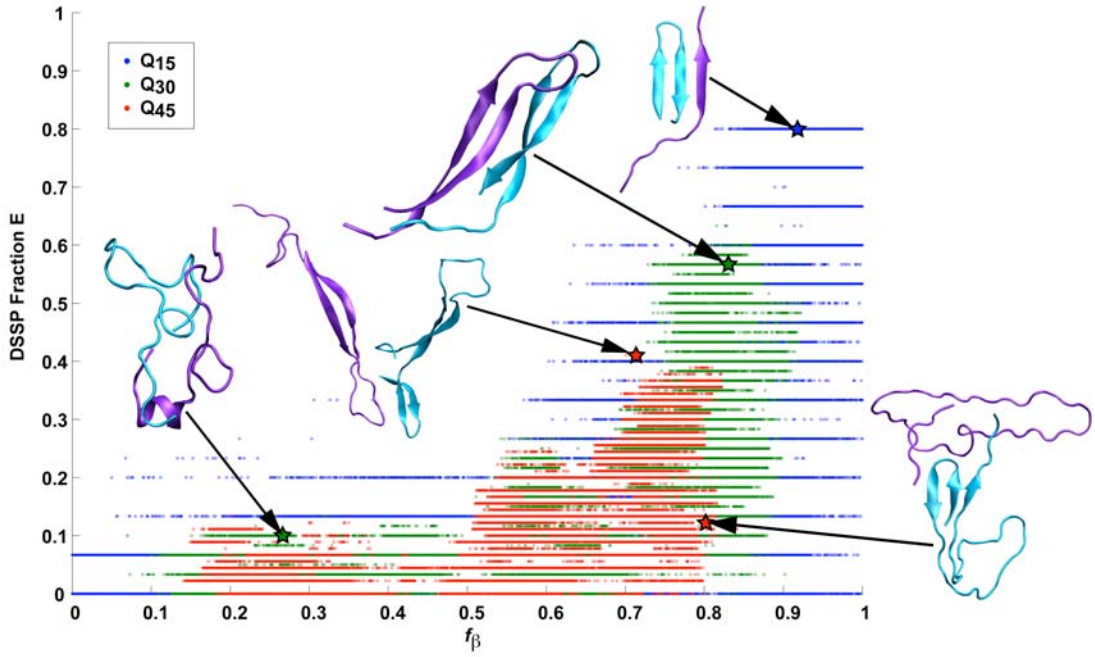

Figure S10: Scatter plot of all recorded snapshots in all dimer simulations for Q_{15}, Q_{30}, and Q_{45} correlating the fractional β-content according to DSSP with the values for f_β at 298 K. This plot is analogous to the one shown in Figure 4 in the main text.

References

1. Vitalis, A., X. Wang, and R. V. Pappu. 2008. Atomistic Simulations of the Effects of Polyglutamine Chain Length and Solvent Quality on Conformational Equilibria and Spontaneous Homodimerization. Journal of Molecular Biology 384:279-297.