Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.


# Disordered peptide chains in an $\alpha$-C-based coarse-grained model: Electronic Supplementary Information

Łukasz Mioduszewski,Marek Cieplak

## 1 Distance distributions and lack of angular dependence

Figs. S1 through S3 show the distance distributions in the ss overlap-determined contacts for all pairs of amino acids for which an ss contact is possible. The distributions are only for the situations in which the directional criteria to make a contact are satisfied. The bin size is 0.1 Å and the columns add up to 1. Fig. S4 shows that there is very little correlation between the $C_\alpha$-$C_\alpha$ distance and the angles used in ss directional criteria (Fig. S4). This is illustrated for VAL-VAL (small side chains), GLN-GLN (mid-sized side chains) and ARG-ARG (large side chains).

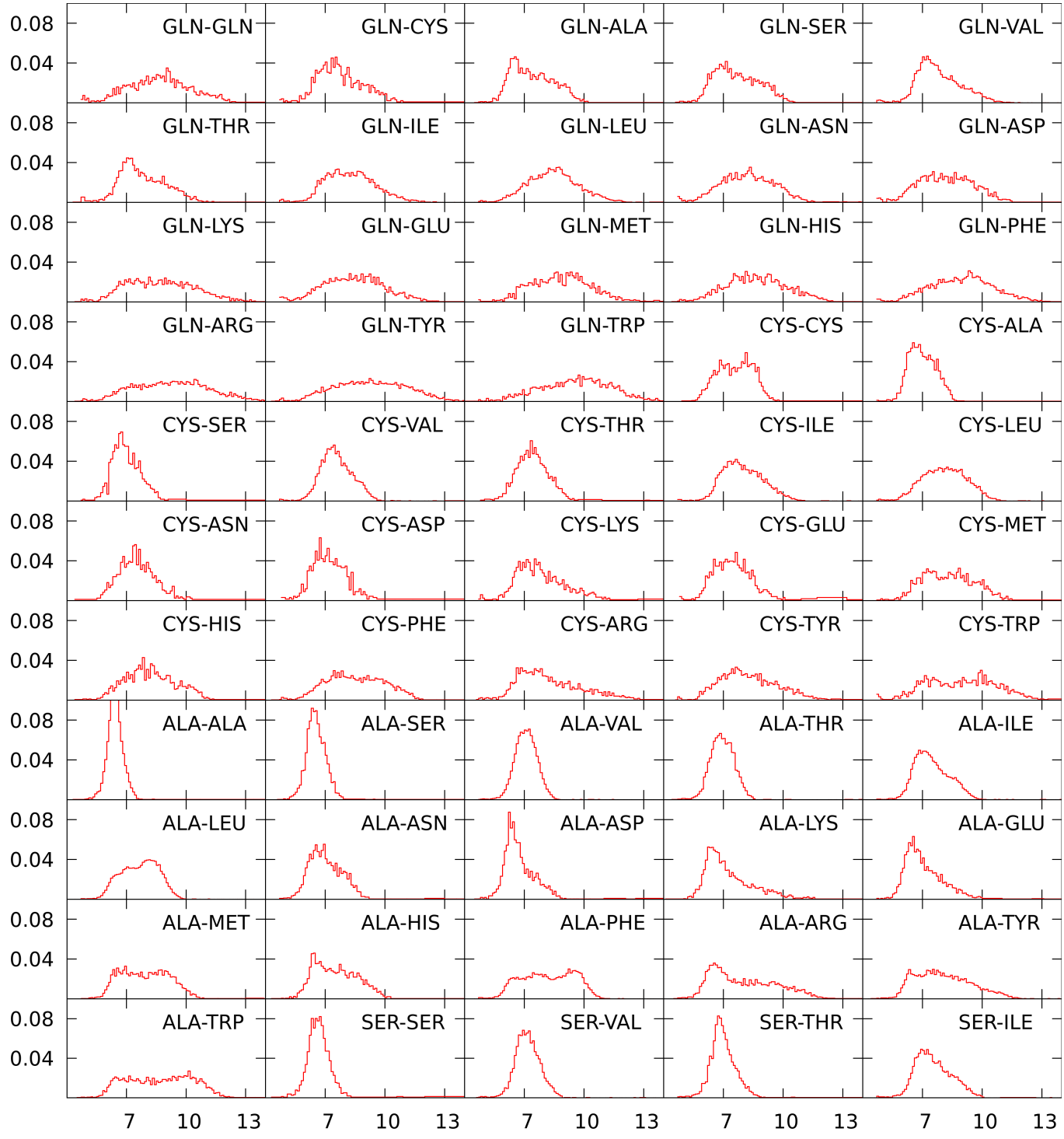


Figure S1: Distance distributions in the ss contacts for the indicated residues.

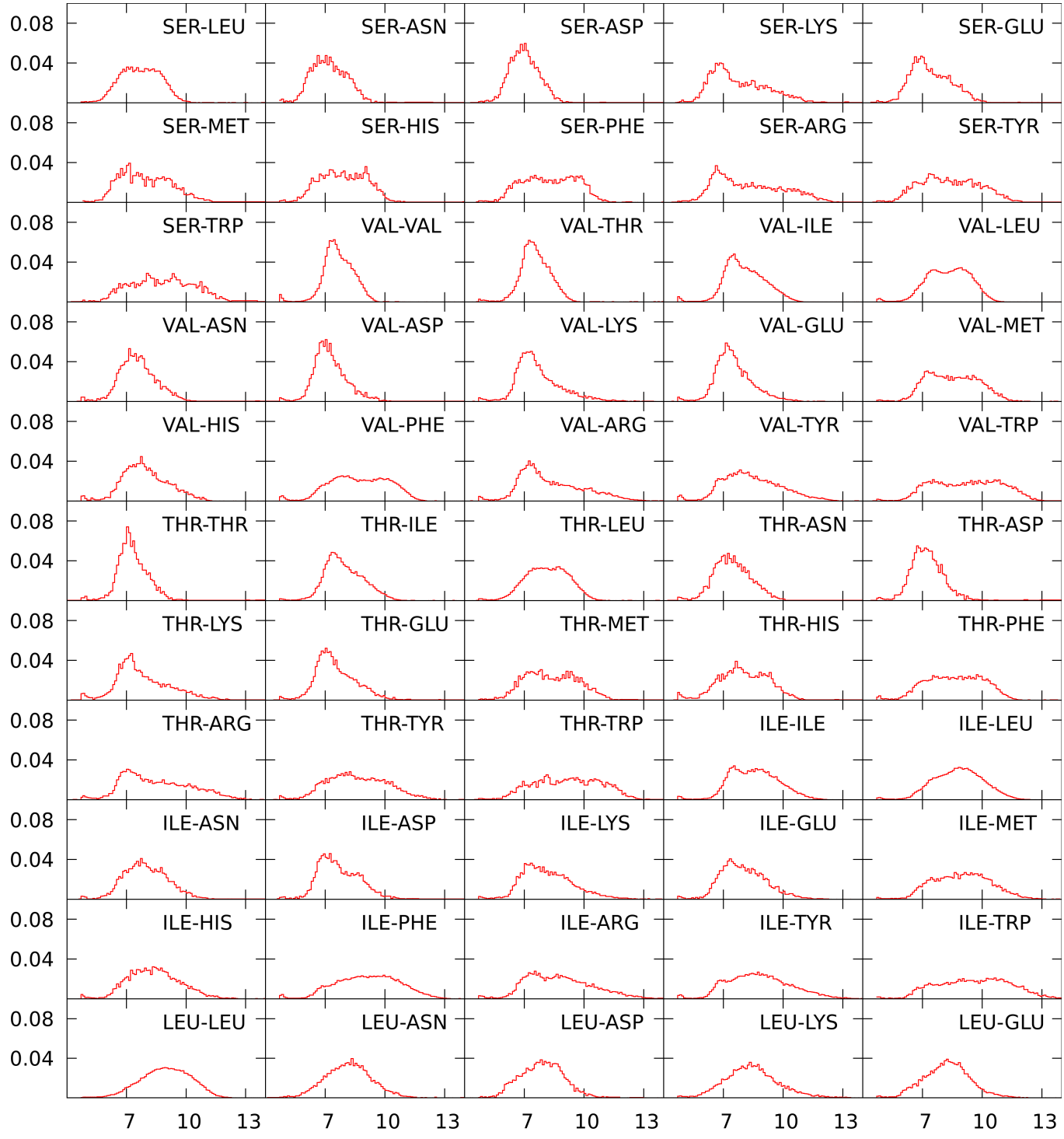


Figure S2: Distance distributions in the ss contacts for the indicated residues.

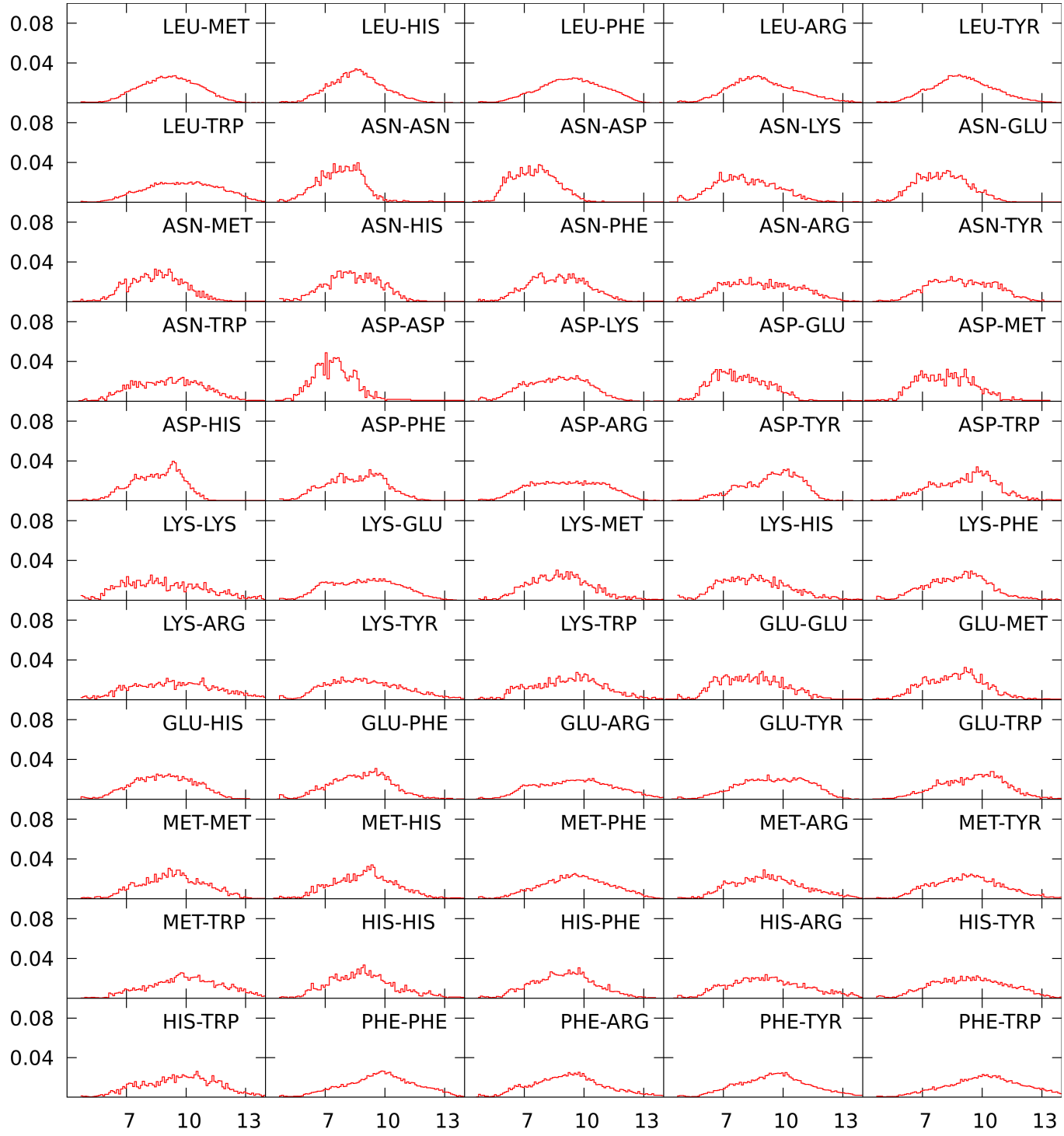


Figure S3: Distance distributions in the ss contacts for the indicated residues.

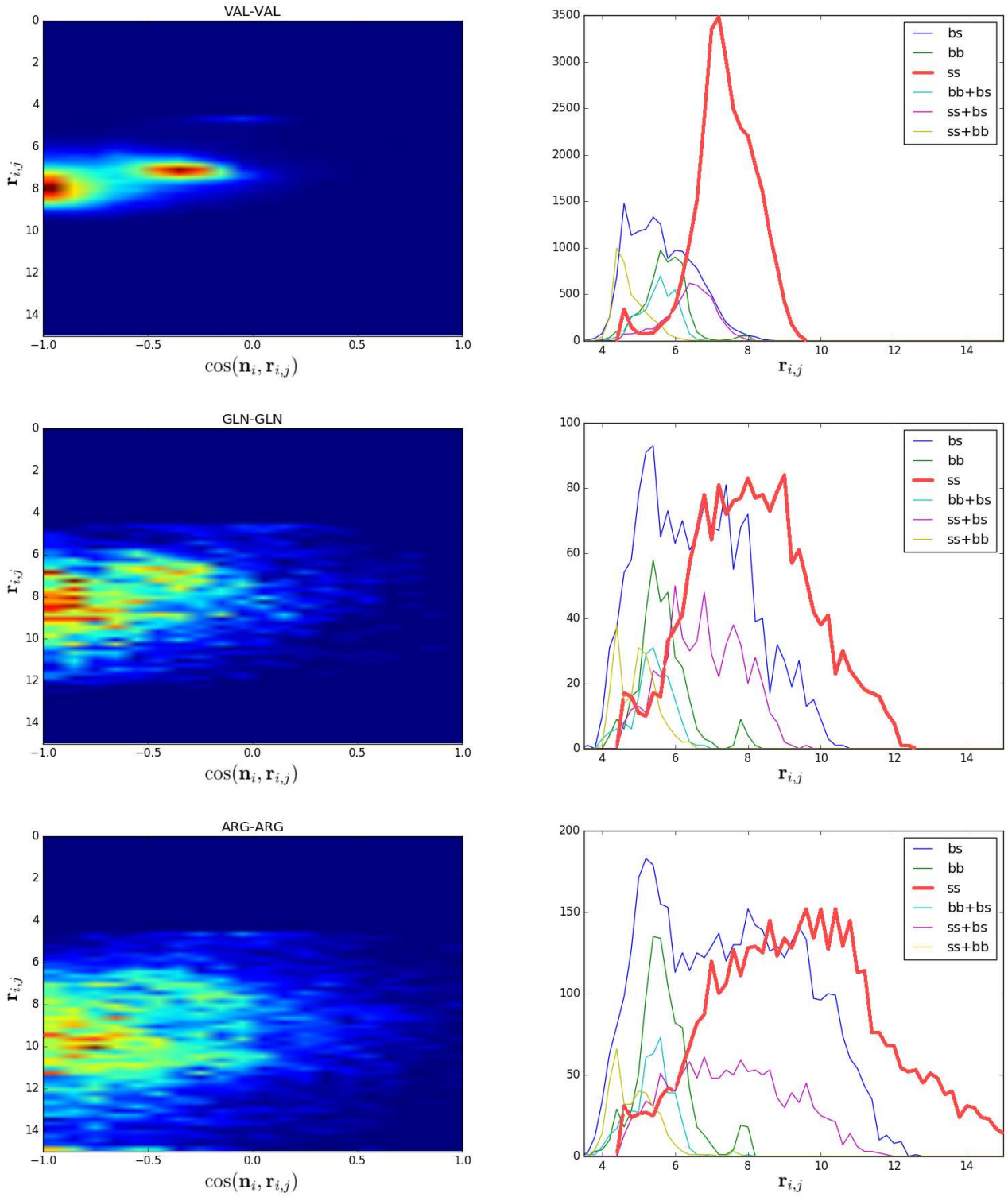


Figure S4: The left panels show 3D histograms of the $C_\alpha$-$C_\alpha$ distance and $\cos(\mathbf{n}_i, \mathbf{r}_{i,j})$ (the blue color means fewer contacts for that distance and cosine, the red color – more contacts) for the ss contacts. The right panels show 2D histograms of the $C_\alpha$-$C_\alpha$ distance. See main text for the definition of $\mathbf{n}_i, \mathbf{r}_{i,j}$.

# 2 The angle-dependent potentials

We use a level-2 angle potential [1], which means that its form depends on 2 subsequent amino acids. All coefficients given here in kJ/mol were converted to $\epsilon$ units assuming $\epsilon = 6.6$ kJ/mol$= 1.58$ kcal/mol.

## 2.1 Bending (bond) angle

We fit the statistical potentials to a sixth degree polynomial of the type $ax^6 + bx^5 + cx^4 + dx^3 + ex^2 + fx + g = 0$ ($x$ is bending angle in radians). The coefficients are listed in Table S1. Three examples of the resulting fits are presented (Fig. S5).

| Residue types | g | f | e | d | c | b | a |
|---|---|---|---|---|---|---|---|
| OGY | 137767.79 | -417519.49 | 523500.78 | -347689.12 | 129057.84 | -25394.62 | 2070.23 |
| OGP | 54278.92 | -166180.67 | 210155.26 | -140514.29 | 52413.91 | -10347.96 | 845.30 |
| OPY | 228674.80 | -725717.73 | 953197.76 | -663471.51 | 258240.30 | -53322.81 | 4566.15 |
| OPP | 70917.09 | -225383.01 | 295803.17 | -205330.47 | 79600.56 | -16366.17 | 1396.80 |
| OXY | 104836.85 | -322892.77 | 411580.60 | -277931.71 | 104885.76 | -20978.03 | 1737.72 |
| OXP | 111628.30 | -353562.64 | 462991.27 | -320775.91 | 124020.92 | -25374.95 | 2147.03 |

Table S1: Coefficients used in the polynomial fitting (in kJ/mol) for 6 different cases. O is any amino acid, Y any amino acid except for P, X any amino acid except for G and P.

## 2.2 Dihedral angle

The dihedral angle potential was fitted to a function with 5 coefficients: $a\sin(x)+b\cos(x)+c\sin^2(x)+d\cos^2(x)+e\sin(x)\cos(x)+f$. The coefficients used $(a, b, c, d, e)$ are listed in Table S2).

| Residue types | f | a | b | c | d | e |
|---|---|---|---|---|---|---|
| GG | 2.117 | -0.008 | 0.004 | -0.125 | 0.425 | -0.061 |
| GP | 2.639 | 0.929 | -0.185 | 0.016 | 0.286 | 0.073 |
| GX | 2.149 | -0.006 | 0.203 | -0.161 | 0.461 | 0.133 |
| PG | 2.165 | -0.102 | 0.109 | 0.149 | 0.152 | -0.742 |
| PP | 3.205 | 1.171 | 0.091 | -0.254 | 0.558 | -1.570 |
| PX | 2.304 | 0.115 | 0.429 | 0.201 | 0.100 | -0.803 |
| XG | 2.136 | 0.018 | -0.071 | 0.122 | 0.179 | -0.624 |
| XP | 2.740 | 0.739 | 0.686 | 0.219 | 0.083 | -0.791 |
| XX | 2.142 | 0.006 | 0.257 | 0.155 | 0.146 | -0.448 |

Table S2: Coefficients of fitting the dihedral potential (in kJ/mol) for 9 different combinations of the two middle amino acids. X can be any amino acid except for G and P.

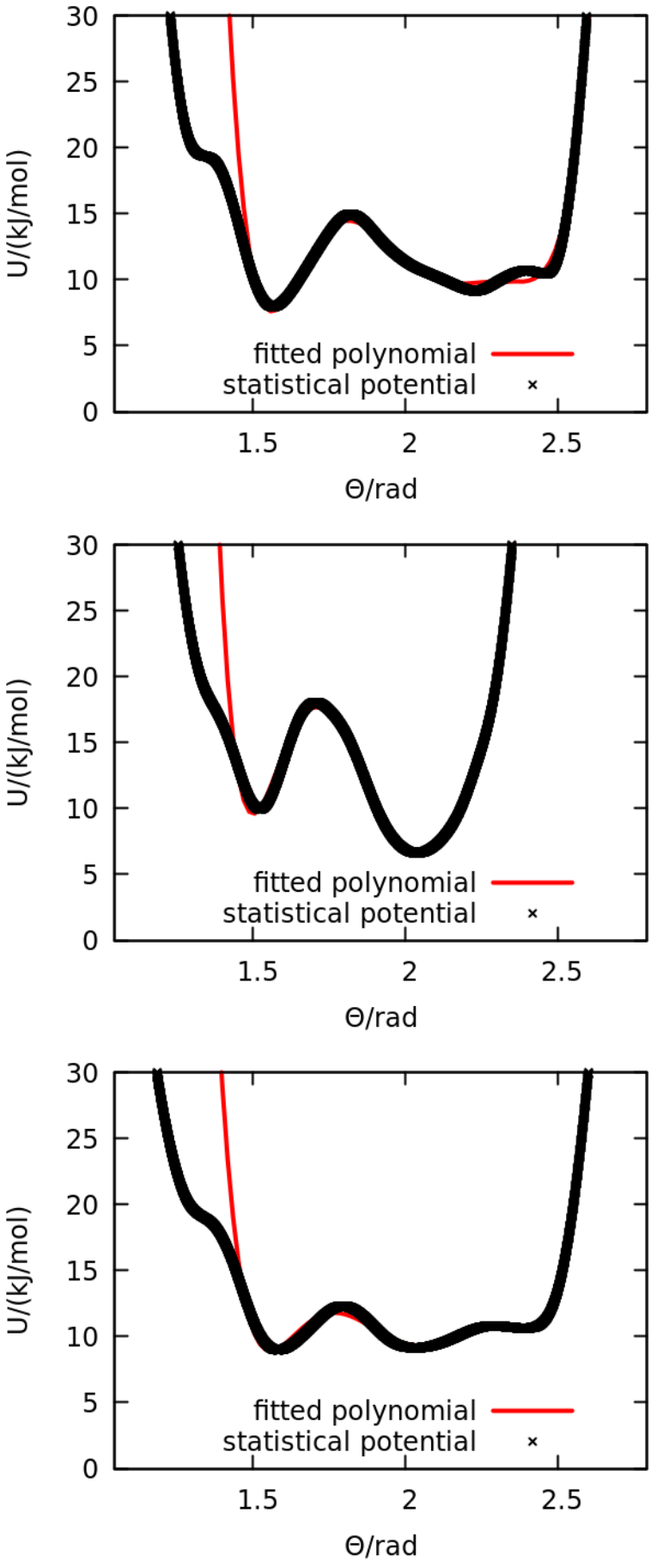


Figure S5: Statistical bending potential [1] (black) and fitted polynomial (red) for OGY, OPY and OXY (code as in [1] and in Table S1.

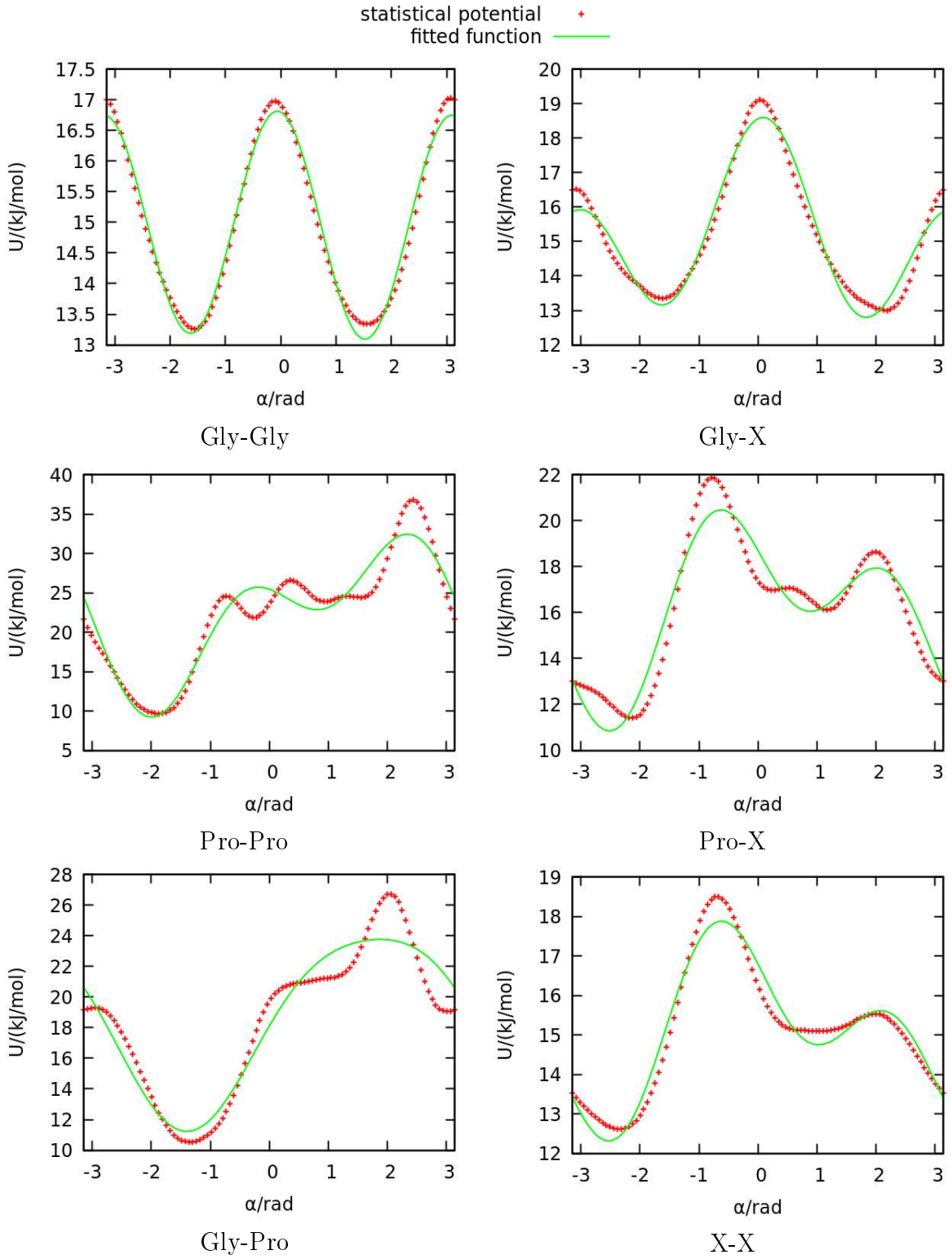


Figure S6: Statistical dihedral potential [1] (dots) and fitted polynomial (lines) for different cases (code as in [1] and in Table S1.

# 3 Turning contacts on and off adiabatically

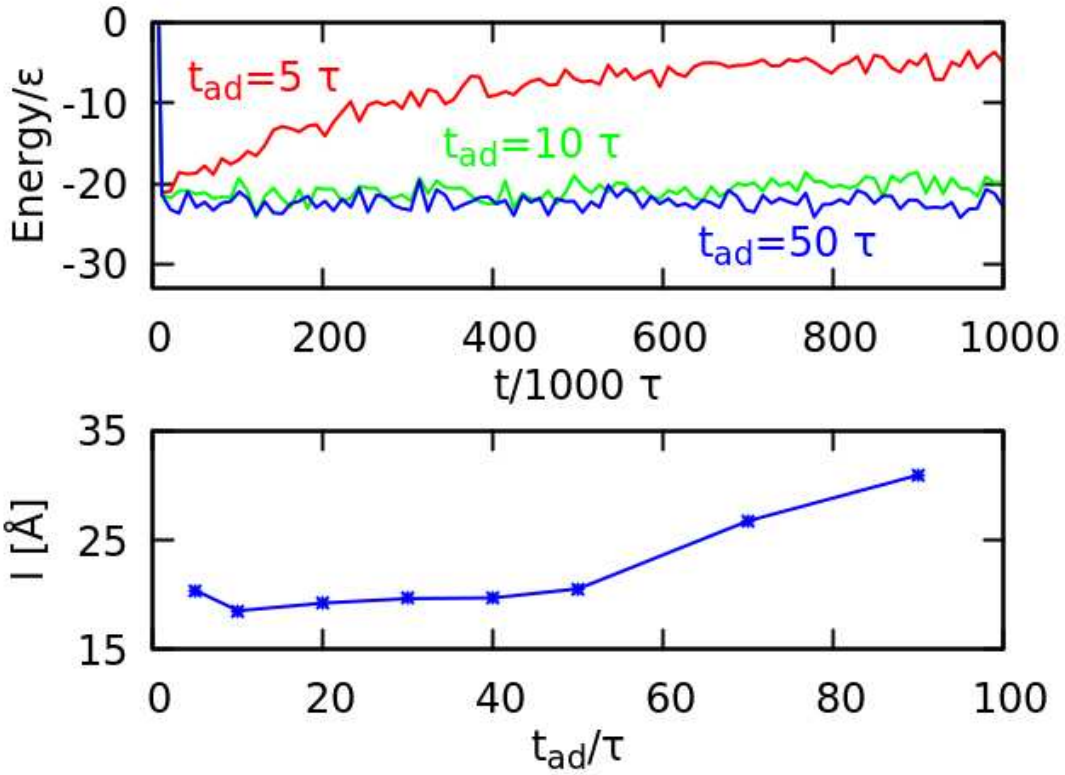


Figure S7: Data for the $Q_{30}$ system for the various durations, $t_{ad}$, of the adiabatic switching on and off of the contacts. The top panel shows the smoothed total energy, $E$, as a function of time in a single trajectory. The bottom panel shows the end-to-end distance, $l$, as a function of $t_{ad}$ as averaged over 100 trajectories. The error of the mean is of the order of the datapoint size. The variance, $\sigma$, also gets larger when $t_{ad}$ exceeds 50 $\tau$.

# 4 Temperature selection

Fig. S8 shows the equilibrium values of end-to-end distance for several temperatures for systems containing polyQ tracts. The best agreement with the all-atom and experimental data is obtained for $T = 0.3\ \epsilon/k_B$.

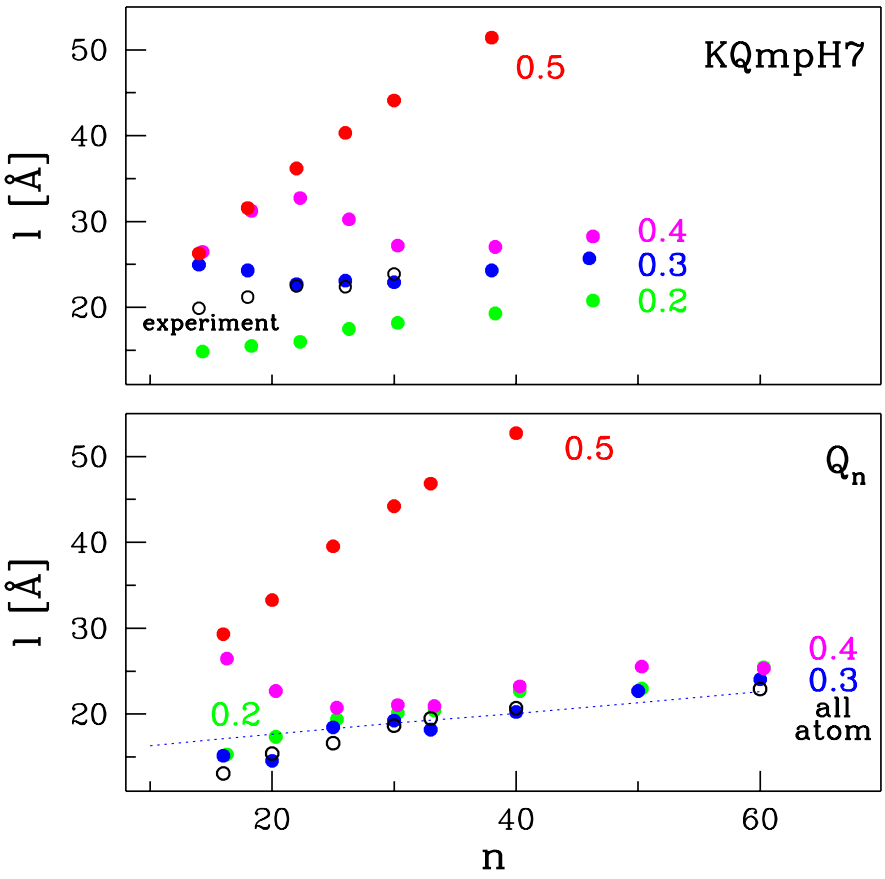


Figure S8: Average end-to-end distance as a function of the sequence length for polyQ (bottom panel) and for sequences of type $KKWQ_mAKK$ (top panel). The four temperatures indicated are in units of $\epsilon/k_B$. The solid data points are obtained by using our coarse-grained model. The open circles are based on the all-atom [3] and experimental [4] results. The error of the mean is lower than size of points.

# 5 Time evolution

Fig. S9 shows $f_{cc}$ matrices of 4 different homopolymers, differing in length and composition. The fraction of common contacts ($f_{cc}$) between two structures is computed as number of their common contacts divided by the total number of contacts in the structure that contains more contacts. Sharp changes in the $f_{cc}$ correspond to conformational transitions.

To quantify these transitions, the following clustering algorithm was developed: two structures are in one cluster if $f_{cc}$ between them is larger than a certain threshold (50 %) and they are next to each other in the timeline (the algorithm works by joining neighboring clusters and $f_{cc}$ between clusters is computed as an average $f_{cc}$ between structures). The red lines denote borders between clusters. The thresholds are system-dependent ($A_{30}$ is much more mobile and has too many conformational clusters in a given timescale, whereas $N_{60}$ is much more sluggish; in the latter case, some conformational changes in the first 150 000 $\tau$ are omitted). We observe that for some systems, large conformational changes take place on the scale of hundreds of microseconds. PolyA is found to keep forming a helix and then unravel it again, in distinction to all-atom results obtained for larger values of $n$. This behaviour of polyA can be due to the approximate nature of our model, but this can also be a matter of adopting relevant time scales.

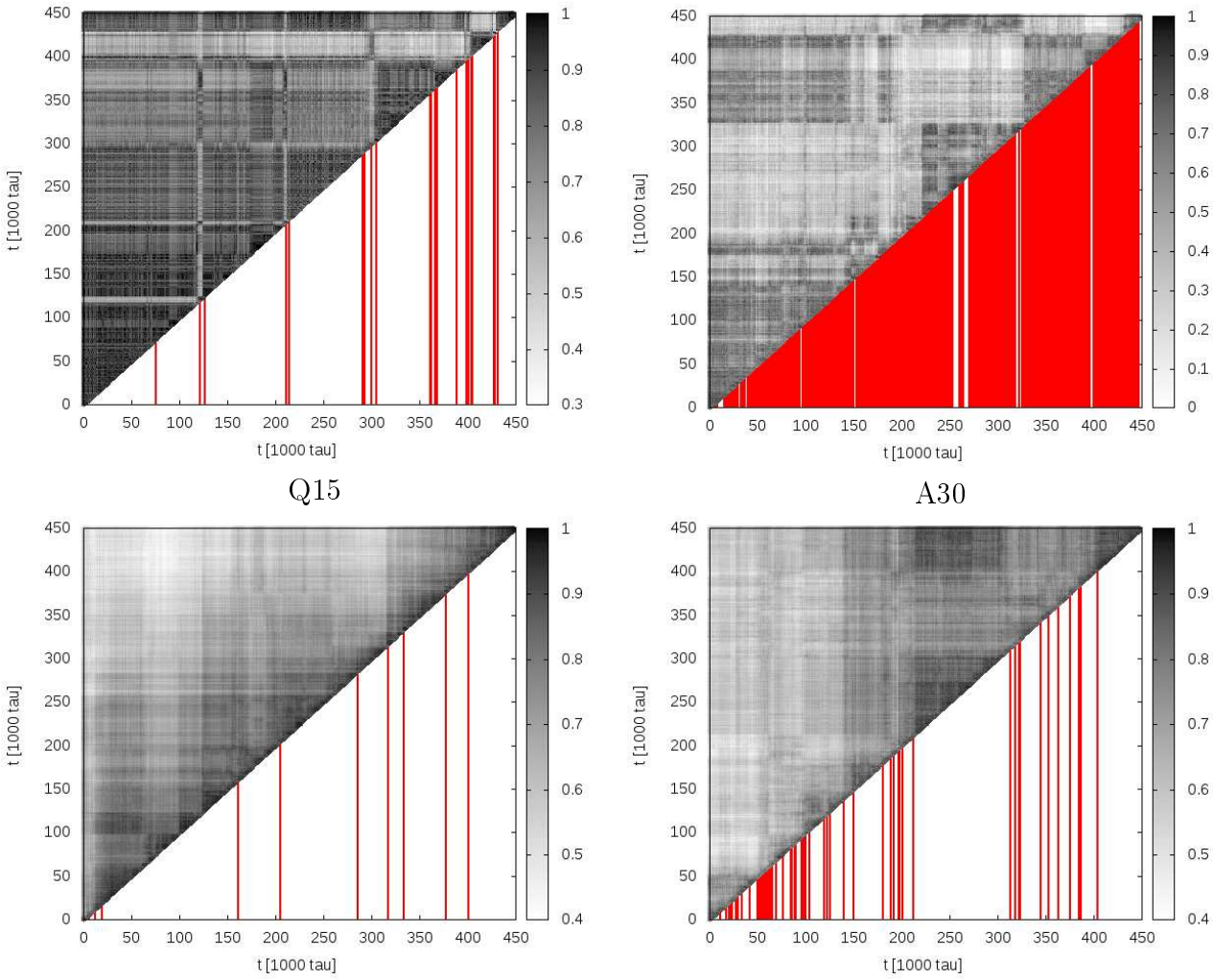


Figure S9: The $f_{cc}$ matrices (fraction of the same contacts between structures from different times, saved every 1000 $\tau$) for different homopolymers. The red lines are explained in the text.

# 6 Replacing LYS with ASN to mimic pH change

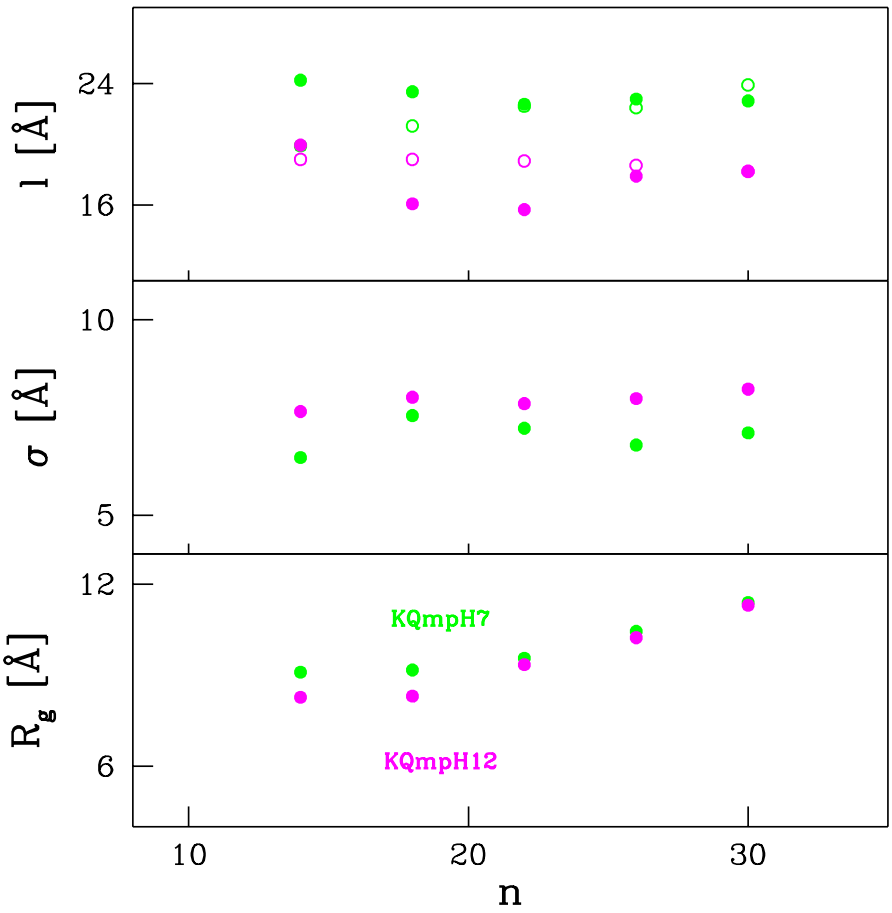


Figure S10: The results obtained by our IDP coarse-grained model for $Q_m$ flanked with lysine (solid green circles) and asparagine(solid purple circles) as a function of the number of residues, $n$. The top panel is for $l$, the middle for $\sigma$, and the bottom one for $R_g$. The open circles correspond to the results obtained by experiment as in ref. [4].

# References


[1] A. Ghavani, E. van der Giessen, and P. R. Onck, Coarse-grained potentials for local interactions in unfolded proteins, *J. Chem. Theor. Comp.*, 2013, **9**, 432-440.

[2] M. Chwastyk, M. Jaskólski and M. Cieplak, The volume of cavities in proteins and virus capsids, *Proteins*, 2016, **84(9)**, 1275-86.

[3] Á. Gómez-Sicilia, M. Sikora, M. Cieplak and M. Carrión-Vázquez, An exploration of the universe of polyglutamine structures *PLoS Comp. Biol.*, 2015, **11**, e1004541.

[4] R. H. Walters & R. M. Murphy, Examining Polyglutamine Peptide Length: A Connection between Collapsed Conformations and Increased Aggregation. *J. Mol. Biol.*, 2009, **393(4)**, 978–992.