Supplementary Information (SI) for Physical Chemistry Chemical Physics.

Supporting Information: Statistical accuracy of molecular dynamics-based methods for sampling conformational ensembles of disordered proteins

Adolfo Bastida,[*,†] José Zúñiga,[†] Federico Fogolari,[‡] and Miguel A. Soler[*,‡]

† *Departamento de Química Física. Universidad de Murcia, 30100 Murcia, Spain*

‡*Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Università di Udine, 33100 Udine, Italy*

E-mail: bastida@um.es; miguelangel.solerbastida@uniud.it

1 Supplementary Figures

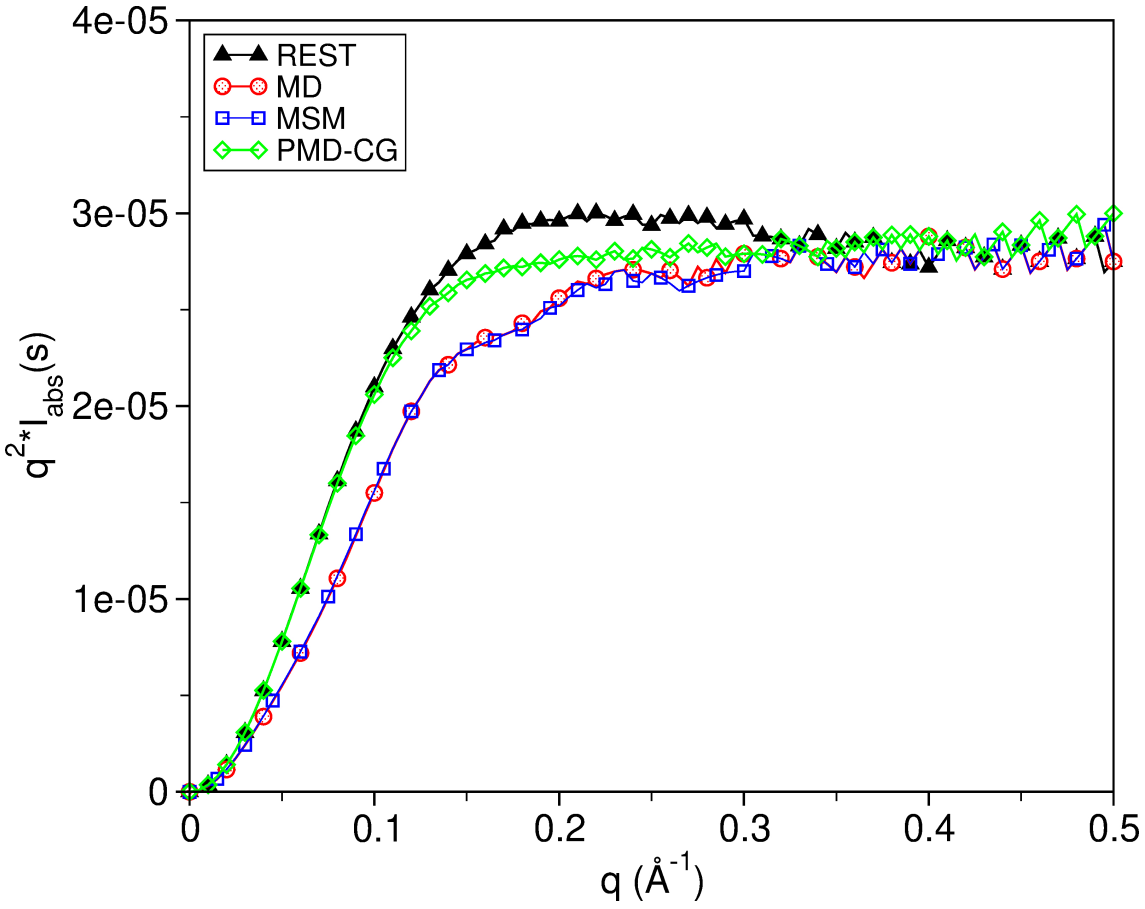

Figure S1: Kratky plot ($q^2I_{abs}(s)$ vs. q) obtained from REST, MD, MSM and PMD-CG conformational ensembles of p53-CTD.

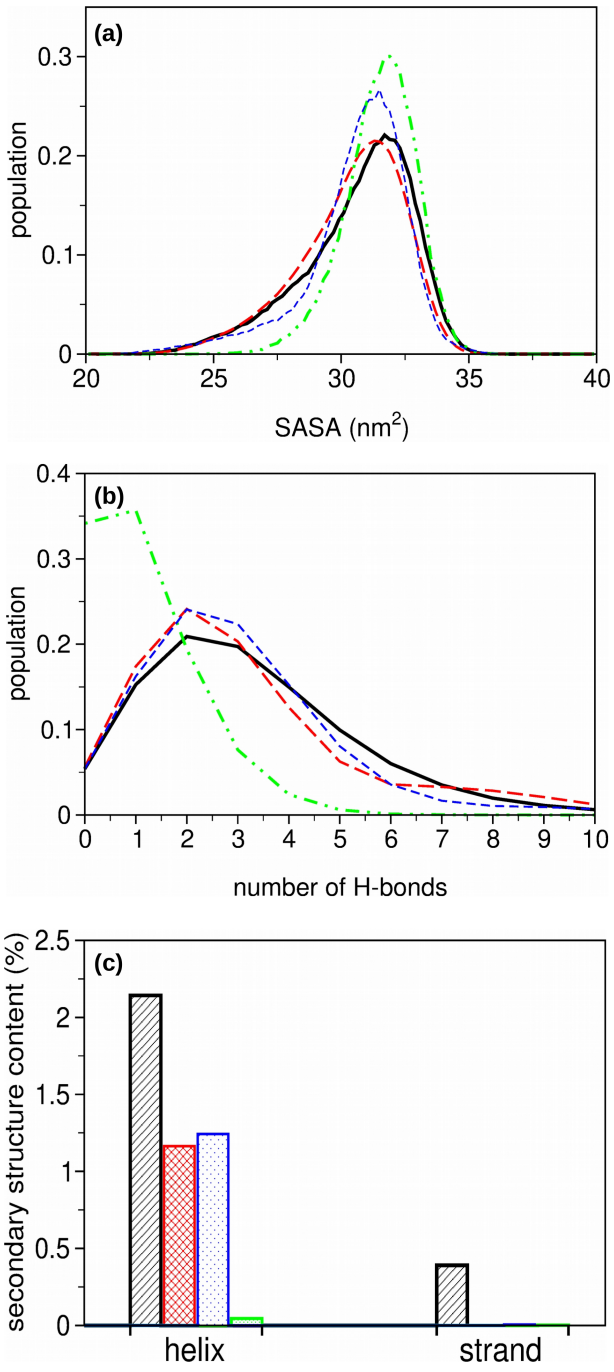

Figure S2: Structural variables obtained from REST (black), MD (red) and MSM (blue) conformational ensembles of p53-CTD. Histrogram profiles of (a) surface-accesible surface area (SASA) and (b) number of H-bonds. (c) Secondary structure content probability. SASA and the number of H-bonds were calculated using the modules *sasa* and *hbond* within GROMACS package v2021.2[S1,S2] and the secondary structural elements were evaluated using STRIDE[S3] in which helix conformation corresponds to the sum of alpha-helix and 3_10-helix.

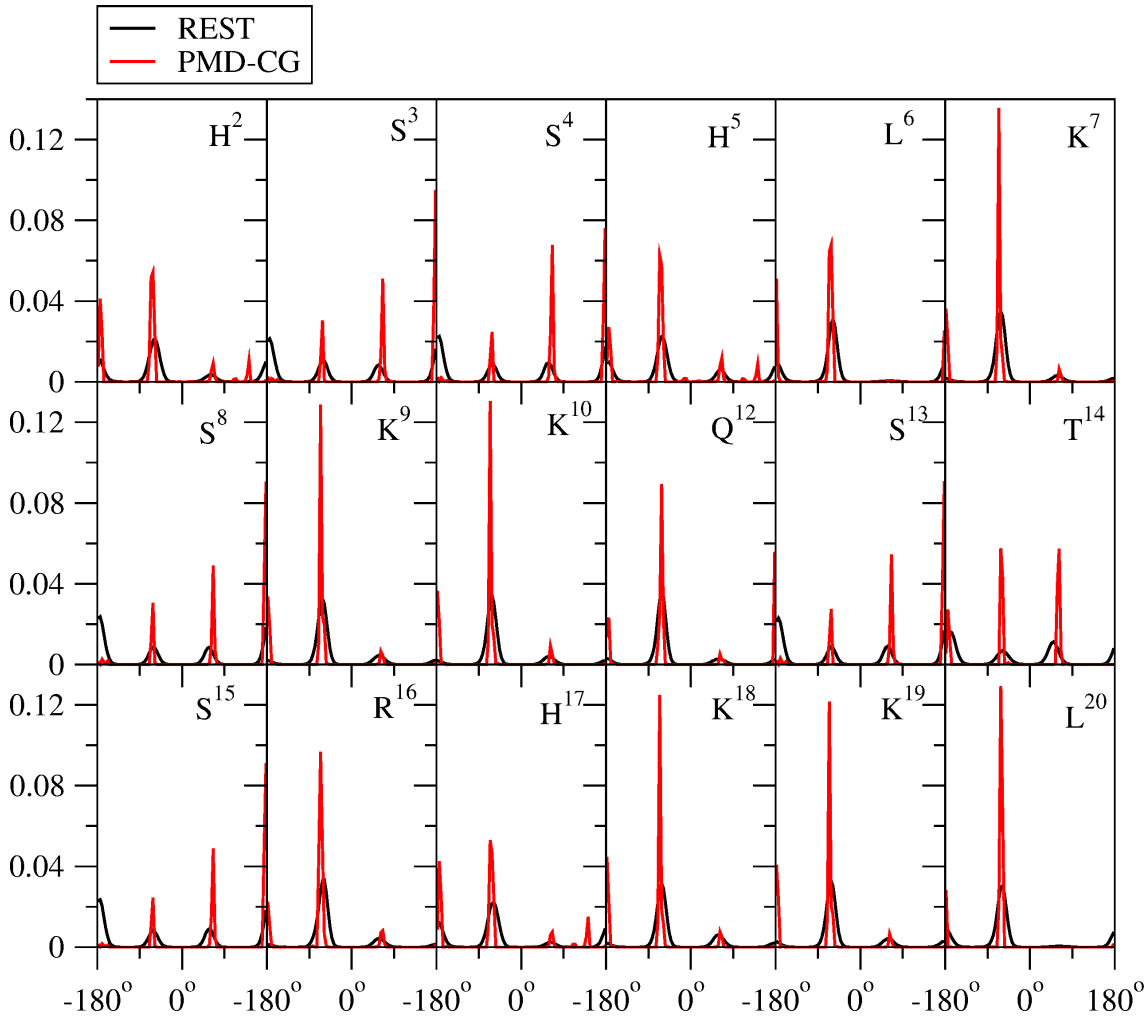

Figure S3: Histograms of χ_1 dihedral angles obtained from REST and PMD-CG conformational ensembles of p53-CTD.

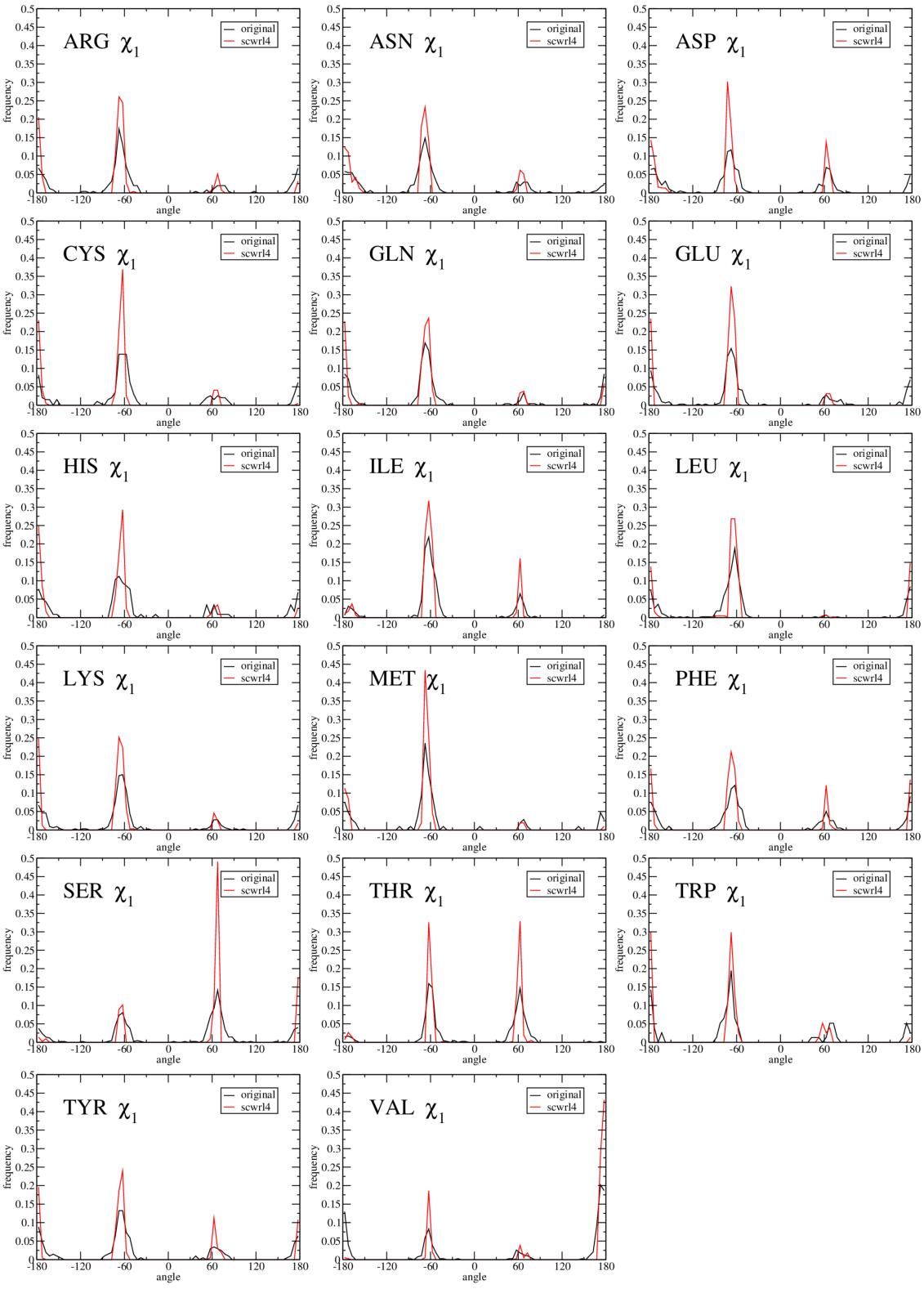

Figure S4: Histogram of χ_1 angles in 5° bins, for the different amino acids obtained from 55 experimental (see Ref. 78 in the main article) protein structures (black line) and after complete side-chain reconstruction of the same structures using Scwrl4 (red line).

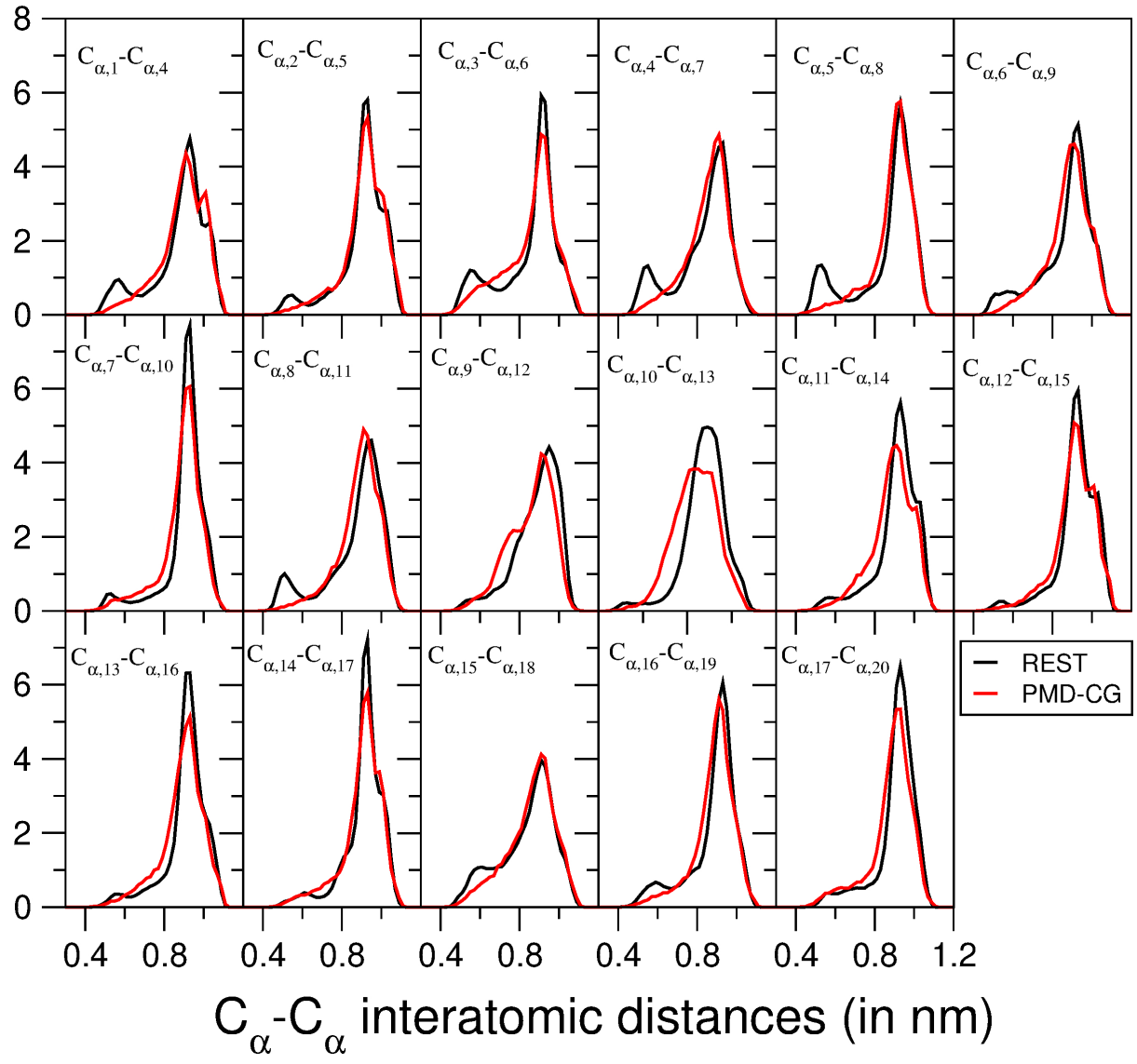

Figure S5: Distribution of C_α-C_α interatomic distances between residues separated by a gap of 2 amino acids obtained from REST and PMD-CG conformational ensembles of p53-CTD.

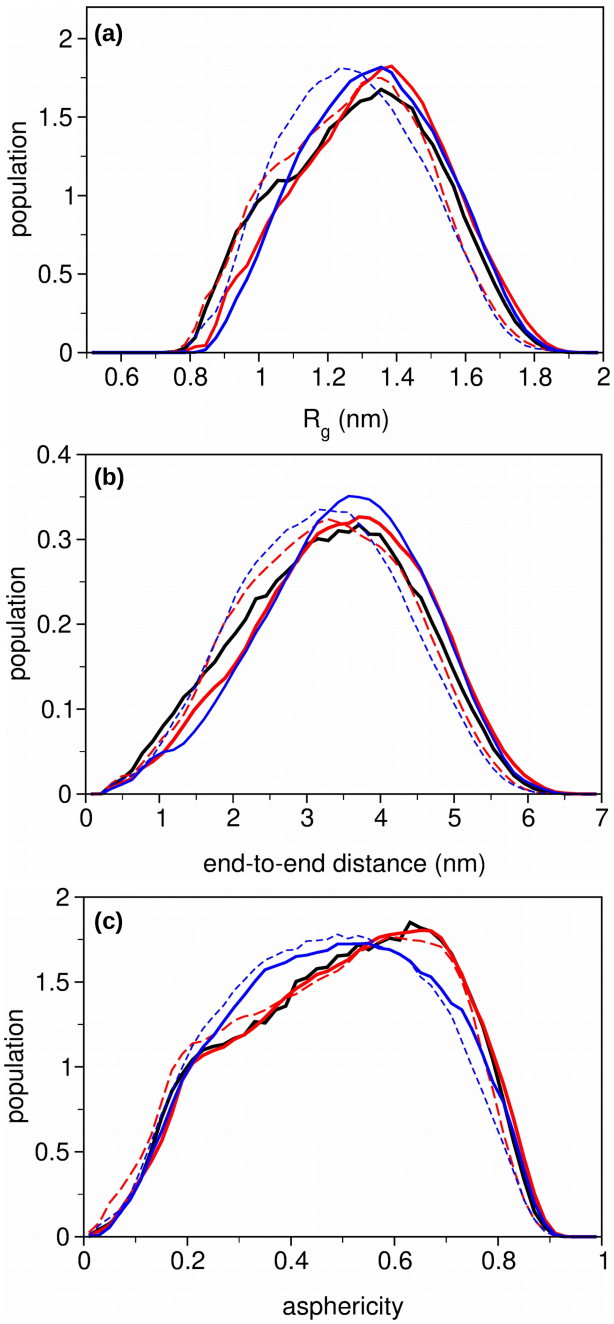

Figure S6: Structural variables obtained from REST (black), MD (red) and MSM (blue) conformational ensembles of p53-CTD. Histrogram profiles of (a) radius of gyration (Rg), (b) end-to-end distance, and (c) asphericity. The dashed lines of MD and MSM data correspond to the standard protocol trajectories while the continuous lines are from the trajectories starting from PMD-CG structures.

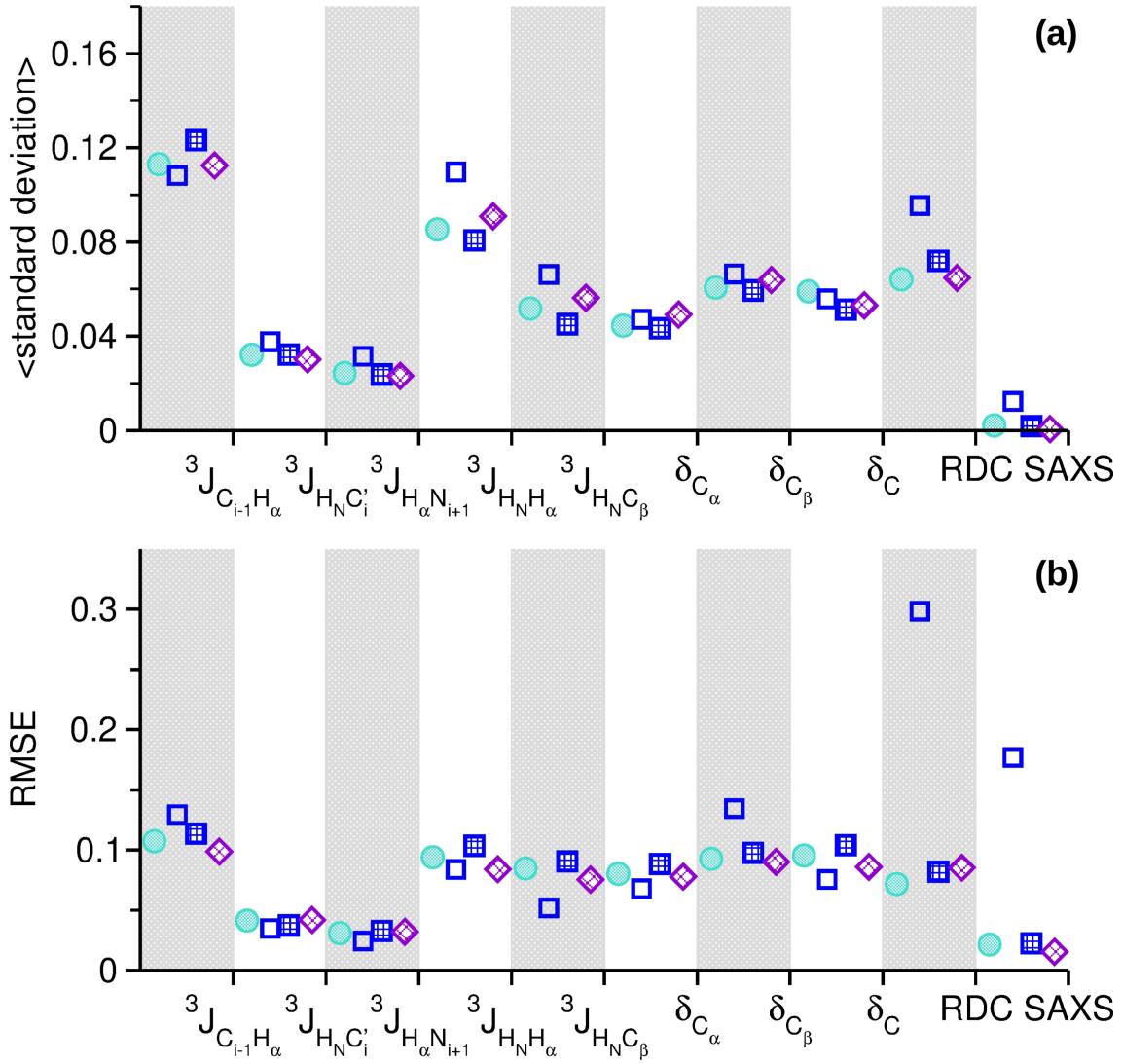

Figure S7: Comparison of (a) average standard deviations, and (b) RMSE of the NMR and SAXS variables obtained from MSM trajectories that employ different pairs of collective variables: $R_g - d_{ee}$ (circle turquoise), $R_g - \gamma$ (blue square) and $d_{ee} - \gamma$ (violet diamond). RMSE values have been calculated with respect to the REST variables. Standard deviation and RMSE units are the same as each variable unit. The corresponding values for RDC have been divided by 10 and those for SAXS have been multiplied by 500.

References

(S1) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theor. Comp.* **2008**, *4*, 435–447.

(S2) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. et al. GROMACS 4.5: A High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(S3) Frishman, D.; Argos, P. Knowledge-based protein secondary structure assignment. *Proteins-Struc. Func. and Genetics* **1995**, *23*, 566–579.