Supplementary Figures

Measuring Similarity Between Dynamic Ensembles of Biomolecules

Shan Yang[1], Loïc Salmon[2], and Hashim M. Al-Hashimi[3*]

1. Department of Chemistry, University of Michigan, Ann Arbor, MI, USA

2. Biophysics, University of Michigan, Ann Arbor, MI, USA

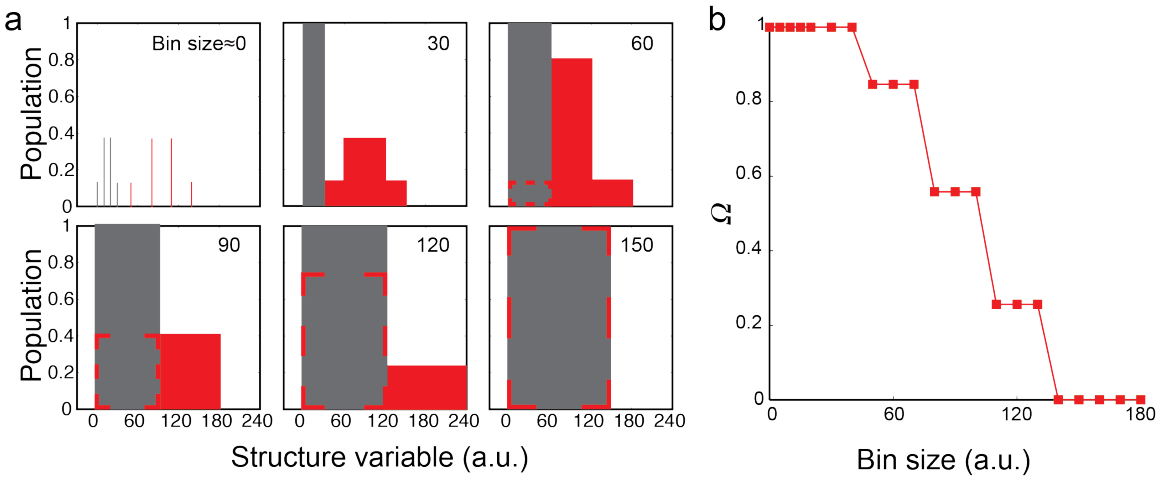
3. Department of Biochemistry and Chemistry, Duke University School of Medicine, Durham, NC, USA

**To whom correspondence should be addressed: hashim.al.hashimi@duke.edu*

Tel: 919-660-1113

Supplementary Figure 1

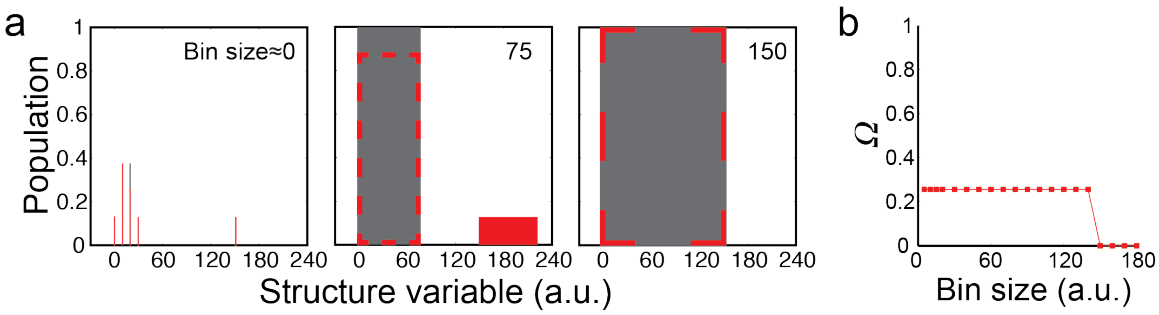
Measuring the nature and extent of similarity between ensembles using 2D Ω versus bin size.



Relative to **Figure 1a**, the gray ensemble is identical and the red ensemble is more scattered to represent a population distribution with four clusters. (**a**) Two discrete ensembles (gray and red) described in terms of an arbitrary structural variable are shown as a function of increasing bin size used to build the histogram. Red dashed box around gray ensemble indicates the portion of the red ensemble that is binned together with the gray ensemble. (**b**) 2D Ω versus bin size plots measuring the nature and extent of similarity between the two ensembles. As the bin size increases, the four structural clusters are gradually binned together with the gray ensemble and this is also reflected as four discrete reductions in the value of Ω as each of the four clusters are binned together with the gray ensemble, one by one.

Supplementary Figure 2

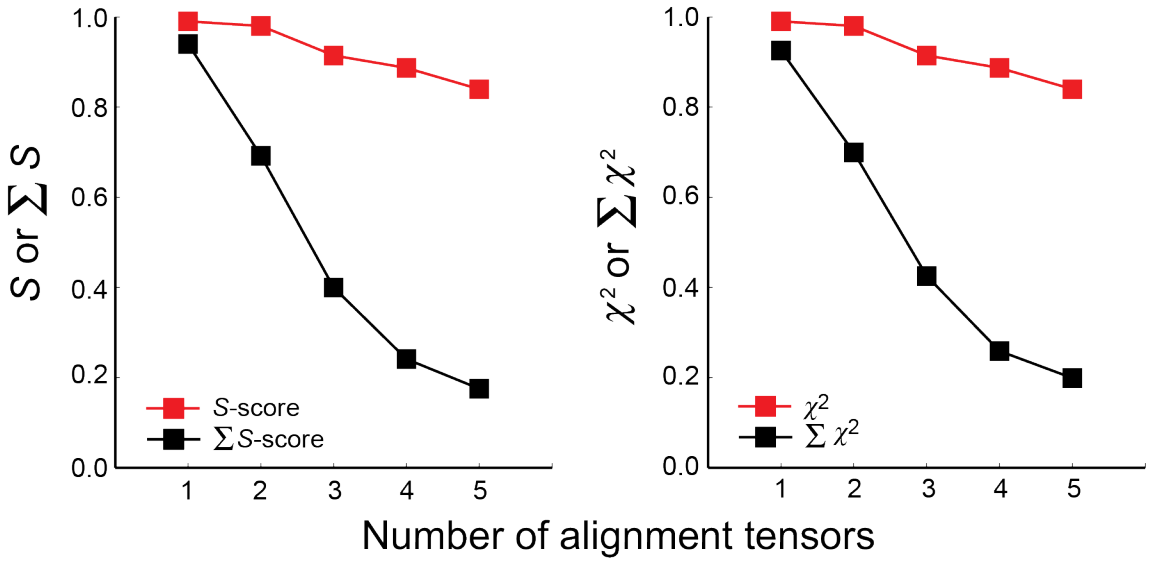
Measuring similarity between ensembles containing outliers using 2D Ω versus bin size plots.



(**a**) Binning of two identical ensembles (gray and red) with the exception of a single outlier. (**b**) 2D Ω versus bin size plots measuring the similarity between the two ensembles. The relatively low Ω values at very small bin sizes accurately capture sharp similarities within the ensemble, the long lasting plateau captures the outlier and its structural dissimilarity, while the sharp drop in the Ω value to Ω=0 at large bin size indicates that any outlier(s) are narrowly distributed.

Supplementary Figure 3

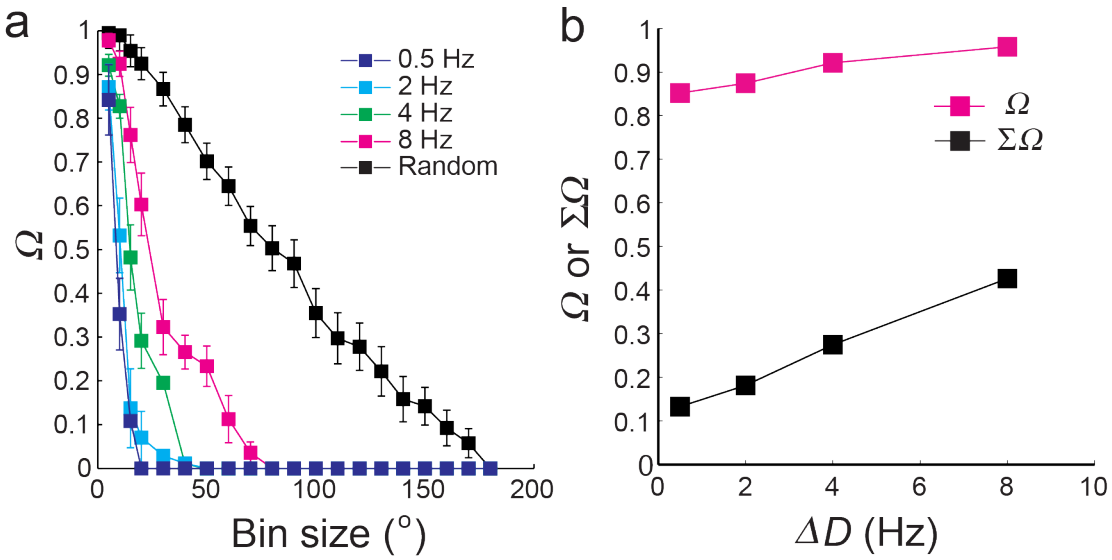
Measuring similarity between ensembles using *S*-score (*S*) and χ^2.



As in **Figure 1e** but using the normalized *S*-score as measure of similarity between the target and the reconstructed ensemble (left panel). Similar results are obtained when χ^2 is used as the measure of similarity between ensembles (right panel).

Supplementary Figure 4

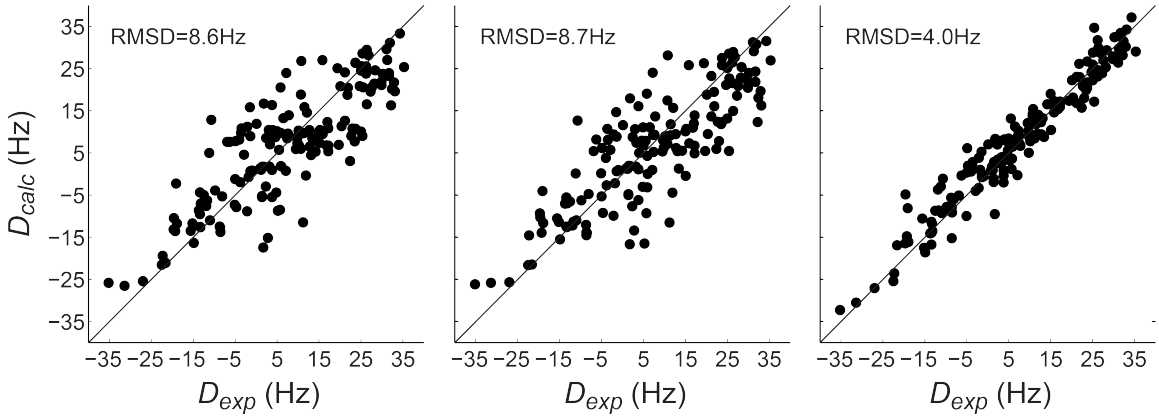
Investigating the accuracy of predicted ensemble as a function of RDC error (ΔD) used in SAS approach.



(**a**) Ω versus bin size comparing the inter-helical angle distributions about a trinucleotide bulge linker between a target ensemble (N=5) and ensembles (N=5) that are selected from the pool randomly (black) or using synthetic RDCs corrupted by increasing uncertainties in SAS selections (color-coded, see inset). Each prediction is repeated for 50 times and the standard deviations of Ω at each bin size are shown as error bars. (**b**) The value of Ω at bin size=5° (magenta squares) and $\Sigma\Omega$ (black squares) as a function RDC uncertainty used in ensemble reconstruction.

Supplementary Figure 5

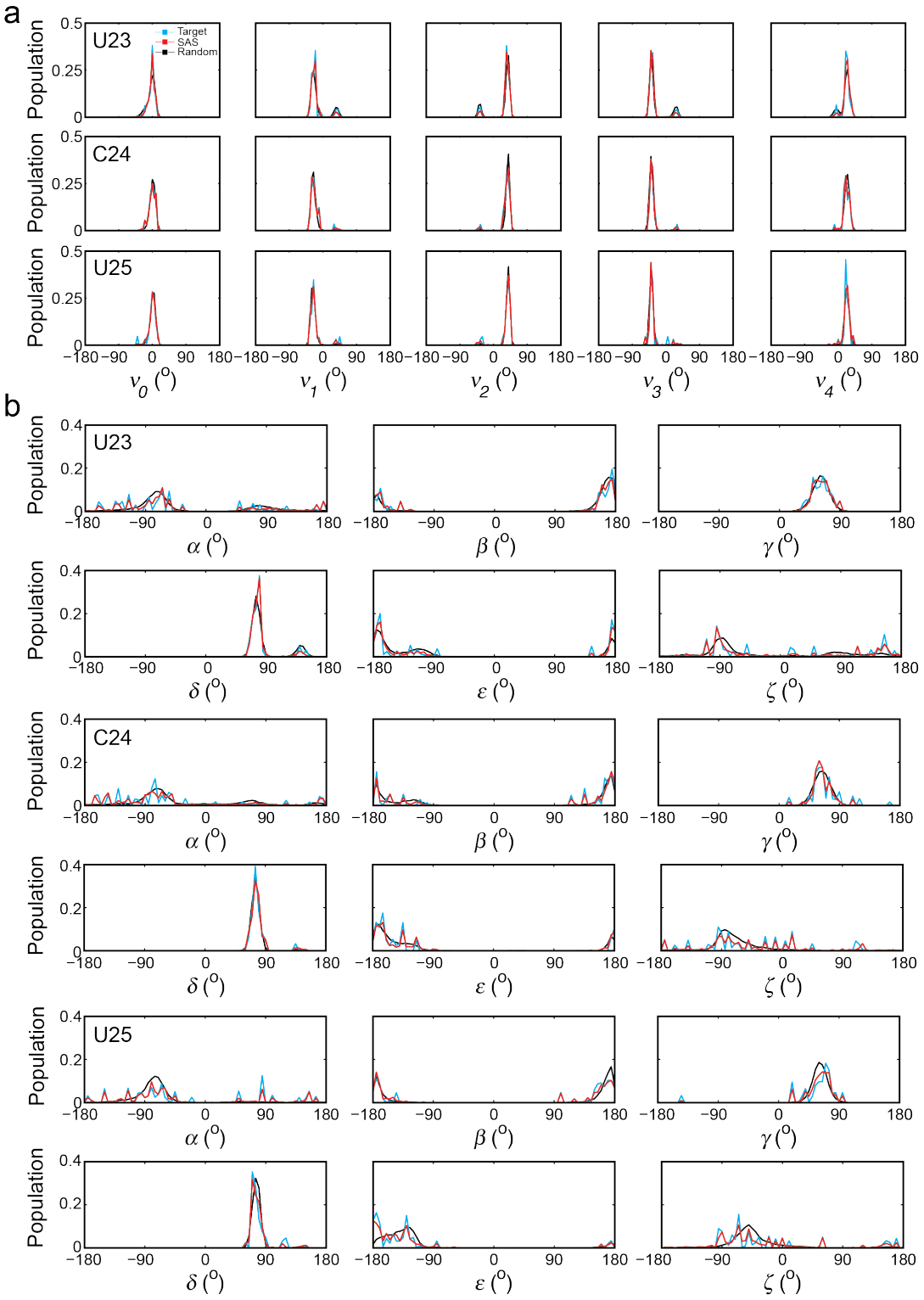
Comparison of experimentally measured and calculated RDCs using Anton MD trajectory, randomly selected ensemble, and SAS selected ensemble.



The RDCs are calculated from and averaged over the entire 8.2 μs Anton MD trajectory (left panel), by combining 10 sets of 20 random conformers (middle panel) and by selecting 20 conformers using SAS approach (right panel) and compared with the experimentally measured RDCs.

Supplementary Figure 6

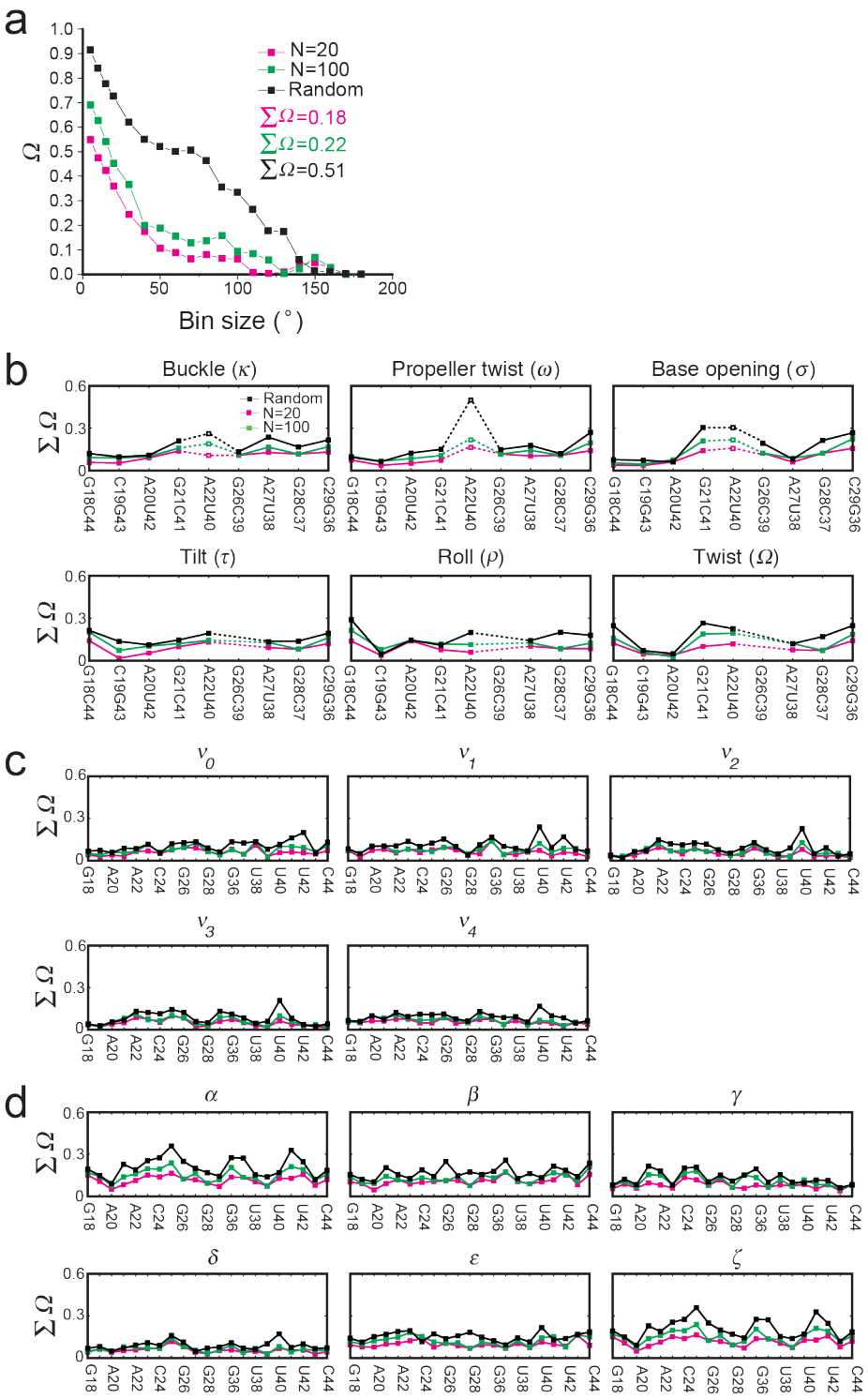
Distributions of sugar and backbone torsion angles of bulge residues in the different TAR conformational ensembles.



Distributions of (**a**) sugar torsion angles and (**b**) backbone torsion angles for bulge residues U23, C24 and U25. The distributions of target ensemble, SAS selected ensemble and randomly selected ensembles are shown in cyan, red and black, respectively.

Supplementary Figure 7

Investigating the selection power of the SAS approach.



A Monte-Carlo based scheme was used to investigate the selection power of SAS approach. The SAS selected TAR ensemble[1] is used as the target ensemble for which 100 independently noise corrupted RDC data sets are generated corresponding to the experimentally available RDC dataset. The SAS approach was implemented to predict the target ensemble using N=20 and N=100. A corresponding random selected ensemble is also presented. The comparison between the target versus RDC-selected (N=20 and N=100) and target versus randomly selected ensembles is shown in magenta, green and black respectively using Ω and $\Sigma\Omega$ for (**a**) inter-helical orientation and $\Sigma\Omega$ for (**b**) base-pair parameters, (**c**) sugar, and (**d**) backbone torsion angles. Substantial improvement in the prediction of inter-helical orientation is observed for the SAS selected ensemble (for both N=20 or 100), leading to corresponding $\Sigma\Omega$ values that indicate a good level of prediction (similar as local angles). The prediction of base-pair parameters (**b**), sugar (**c**) and backbone (**d**) torsion angles consistently show that the SAS approach provides better predictions than the randomly selected ensemble. The intra-base-pair parameters for the flexible junction A22-U40 base-pair are shown using open symbols and dashed lines and inter-base-pair parameters are not shown for the junction G26-C39 base-pair because they are ill-defined due to the presence of the bulge between G26-C39 and A22-U40.

References

1. Salmon, L., Bascom, G., Andricioaei, I. & Al-Hashimi, H.M. *J. Am. Chem. Soc.* **135**, 5457-5466 (2013).