S3 Text

Clustering and DASH analyses

To handle the large amount of data produced by the 4.1 µs simulation time, it was decided to reduce the data density by extracting one snapshot per nanosecond, leading to a total of 4,100 microstates for subsequent data analysis. This was achieved using the cluster analysis functionality of the PTRAJ module provided in the AMBER package. To group the configurations found in the MD simulations into distinct sets, a means algorithm based on the RMS deviation of the protein Cα-atoms was applied.1 Snapshots representing the geometrical center of each cluster obtained were identified and served as the structural basis for subsequent simulations and analyses. A cluster analysis performed at the end of the simulation (4.1 µs) led to 15 clusters including five generated by a preliminary cluster analysis performed after 1.8 µs. The cluster centers obtained from the two analyses were subjected to cavity identification and comparison.

As an alternative to a clustering based on the RMS deviation of Cα-atoms given in Cartesian coordinates, the DASH2 procedure was employed using the φ and ψ dihedral angles of the protein backbone of each residue and snapshot as data source. The analysis tool Simulaid was used to extract the dihedral angles for each snapshot from the MD trajectory.3

9 References and Notes

