Supporting Information

Molecular Dynamics of a DNA Holliday Junction: The Inverted Repeat Sequence d(CCGGTACCGG)$_4$

Elizabeth G. Wheatley[*], Susan N. Pieniazek[*], Ishita Mukerji[†], D. L. Beveridge[*‡]

[*]Department of Chemistry, [†]Department of Molecular Biology and Biochemistry, and Molecular Biophysics Program
Wesleyan University, Middletown, Connecticut

[‡]Correspondance: dbeveridge@wesleyan.edu

Supporting Materials and Methods

S1 MD Simulations

MD trajectories for each OPN-HS, OPN-MS, STX-HS, STX-MS, CBD-HS, and CBD-MS DNA complex were obtained for 100 ns. The OPN-MS simulation was repeated five times with different initial velocities. Details for each simulation are shown in Table S1.

Table S1. Setup for each MD simulation with type of DNA structure, salt concentration, duration, size of water box, and number of atoms.

MD Name	DNA	NaCl [mM]	Time (ns)	Number of waters	Total Number of atoms
OPN-HS	**4WJ**	**200**	**100**	**9667**	**30371**
OPN-MS*+	**4WJ**	**–**	**100**	**9741**	**30519**
STX-HS	**4WJ**	**200**	**100**	**6983**	**22297**
STX-MS*	**4WJ**	**–**	**100**	**7034**	**22398**
CBD-HS	**B DNA**	**200**	**100**	**5048**	**15830**
CBD-MS*	**B DNA**	**–**	**100**	**5086**	**15906**

Notes: *Na^+ only

+Plus five duplicates at 50ns each

S2 MD Set-up

The crystal structure 1dcw served as a reference 4WJ for all simulations. The junction is composed of four strands of CCGGTACCGG, with ApC steps connecting the junction crossover points. The actual crystal structure coordinates were used as a starting structure for the STX simulations. For the OPN simulations, a canonical open model built using 3DNA was obtained from W. Olson of Rutgers and was mutated to match the 1dcw sequence by removing base atoms and renaming the "residue type" for remaining backbone atoms. The missing nucleotide atoms were automatically replaced by tLeap [AMBER 9; AmberTools 1.2] and the skewed bases were gently repositioned in Pymol [Delano Scientific, 2006] to facilitate Watson-Crick base pairing. The simulations were set up using tLeap to create the coordinate and parameter files. Each system was first checked for errors, aligned, and then solvated with an octahedral TIP3P water box with a maximum 12 Å cutoff separating the edge of the solute to the edge of the box. Next each solvated system was neutralized with Na+, and an additional 200 mM NaCl was added for high salt (HS) conditions based on the size of the water box. The ions were then randomized using the randomizeions command in the ptraj(1) module of AmberTools 1.2. The ramdomizeions command was executed with a 5 Å minimum ion-ion and ion-solute distance cutoff. For example, for the OPN-HS system, the command was set up as follows:

```
$ptraj *.prmtop <<EOF
> trajin *_ur.inpcrd
> randomizeions :41-150 around :1-40 by 5.0 overlap 5.0
> trajout *.inpcrd restrt
> EOF
```

with 41-76 referring to the resid numbers of the ions in the system and 1-40 the bases of the DNA 4WJ. The new randomized inpcrd file was used for the MD simulation. The size of each system and number of water and ions are shown in Table S1.

S3 Ion and Ordered Water Density, and Occupancy Analysis

Average ion and ordered water density histograms were calculated using the grid command in the ptraj(1)module of AmberTools 1.2. The calculation was performed with a bin spacing of 0.5 Å over a total area of 100 $Å^3$, large enough to encompass the entire 4WJ structure. The 4WJ was centered and the solvent was re-imaged and RMSD fit before the grid calculation was obtained. Corresponding average 4WJ structures were built from the same snapshots that contributed to the solvent and ion grids. The resulting grids were visualized using Chimera. An example of how the grid command was used for the 4WJ is shown below:

```
trajin *.mdcrd 1 X 50

center :1-10 mass origin
image origin center familiar

center :11-20 mass origin mass
image origin center familiar

center :21-30 origin mass
image origin center familiar

center :31-40 origin mass
image origin center familiar

center :1-40 origin mass
image origin center familiar

rms first mass out *.dat :2-9@O5',C5',C4',C3',O3',P,:12-
19@O5',C5',C4',C3',O3',P,:22-29@O5',C5',C4',C3',O3',P,:32-
39@O5',C5',C4',C3',O3',P

grid WAT-*.grid 100 0.5 100 0.5 100 0.5 :WAT@O
grid Na-*.grid 100 0.5 100 0.5 100 0.5 @Na+

translate x -0.25 y -0.25 z -0.25

average *.pdb pdb :1-40
```

The hbond command in ptraj(1) was used to calculate average Na+ ion occupancies around the junction base phosphate oxygen atoms. The resulting average occupancy percentages were used to determine the approximate charge neutralization of the junction.

A 5 Å cutoff was employed to minimize overcompensation of charge neutralization that could occur if the cutoff were too large and extraneous phosphates and their corresponding counterions were included for a single phosphate group of interest.

S4 PCA on the STX-MS, STX-HS, OPN-MS, and OPN-HS Covariance Matrices

We carried out principal component analysis (PCA) on the covariance matrix for each simulation by fitting the DNA backbone atoms of each configuration to the average structure of the full-length trajectories (Fig S9). A covariance matrix—or, a $3N$ x $3N$ matrix, where N is the number of DNA atoms in each molecule—was calculated from the new fitted trajectory using the ptraj(1) module of AmberTools 1.2. The elements of the matrix consist of covariances of atomic displacements relative to their average positions. We diagonalized the matrix to obtain a set of eigenvectors and eigenvalues, which represent components and variance along each principal component, respectively. Each trajectory was projected onto its respective principal component; from these projections, two—dimensional plots along the first two principal components were generated to represent the sampled distribution and corresponding populations in configurational space.

Supporting Information Results

S5 Dynamical Stability of Simulations

S.5.1 RMSD plots for OPN-MS, OPN-HS, STX-MS, STX-HS, CBD-MS, and CBD-HS

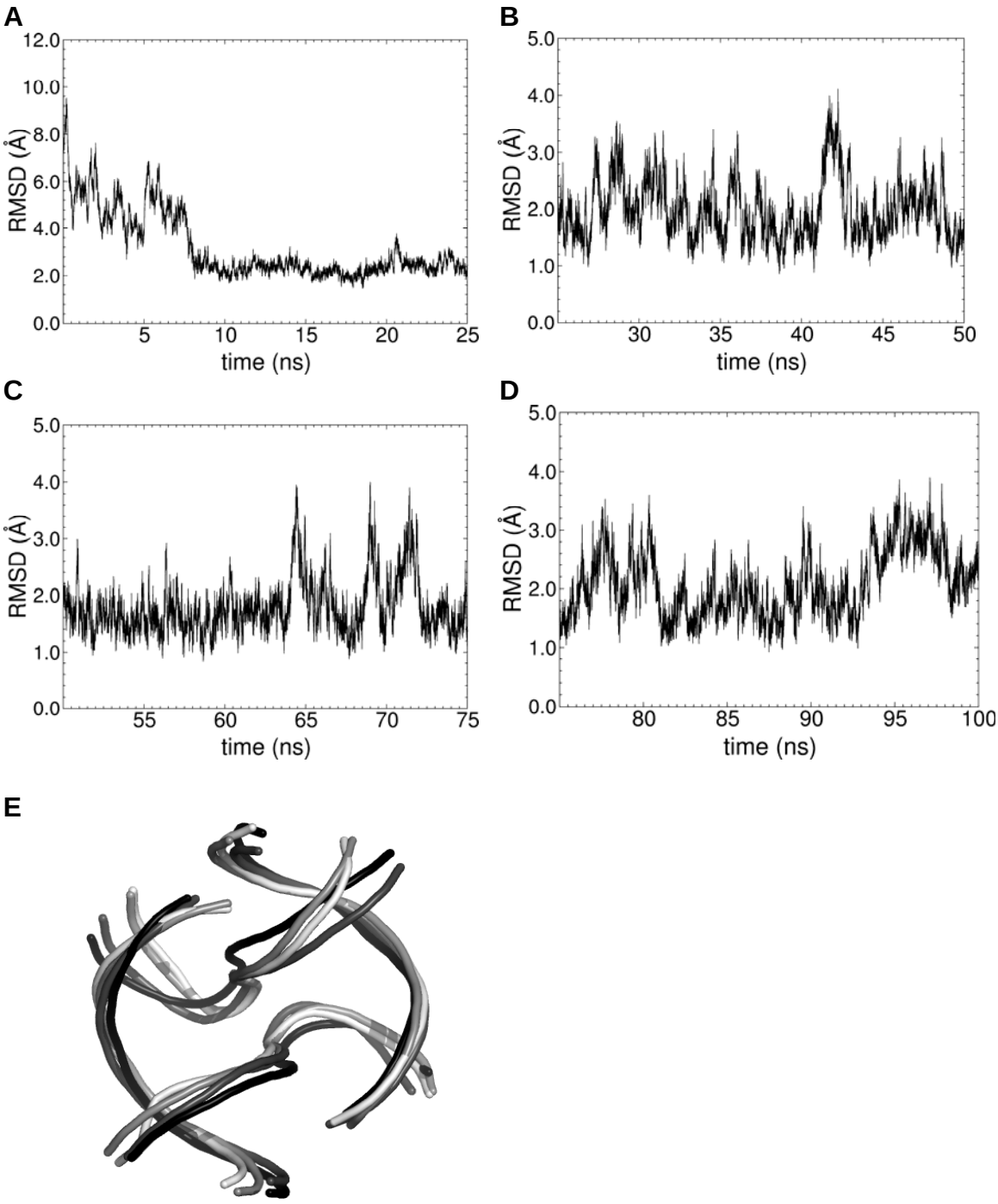

FIGURE S1. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for OPN-MS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four structures is 1.2 Å.

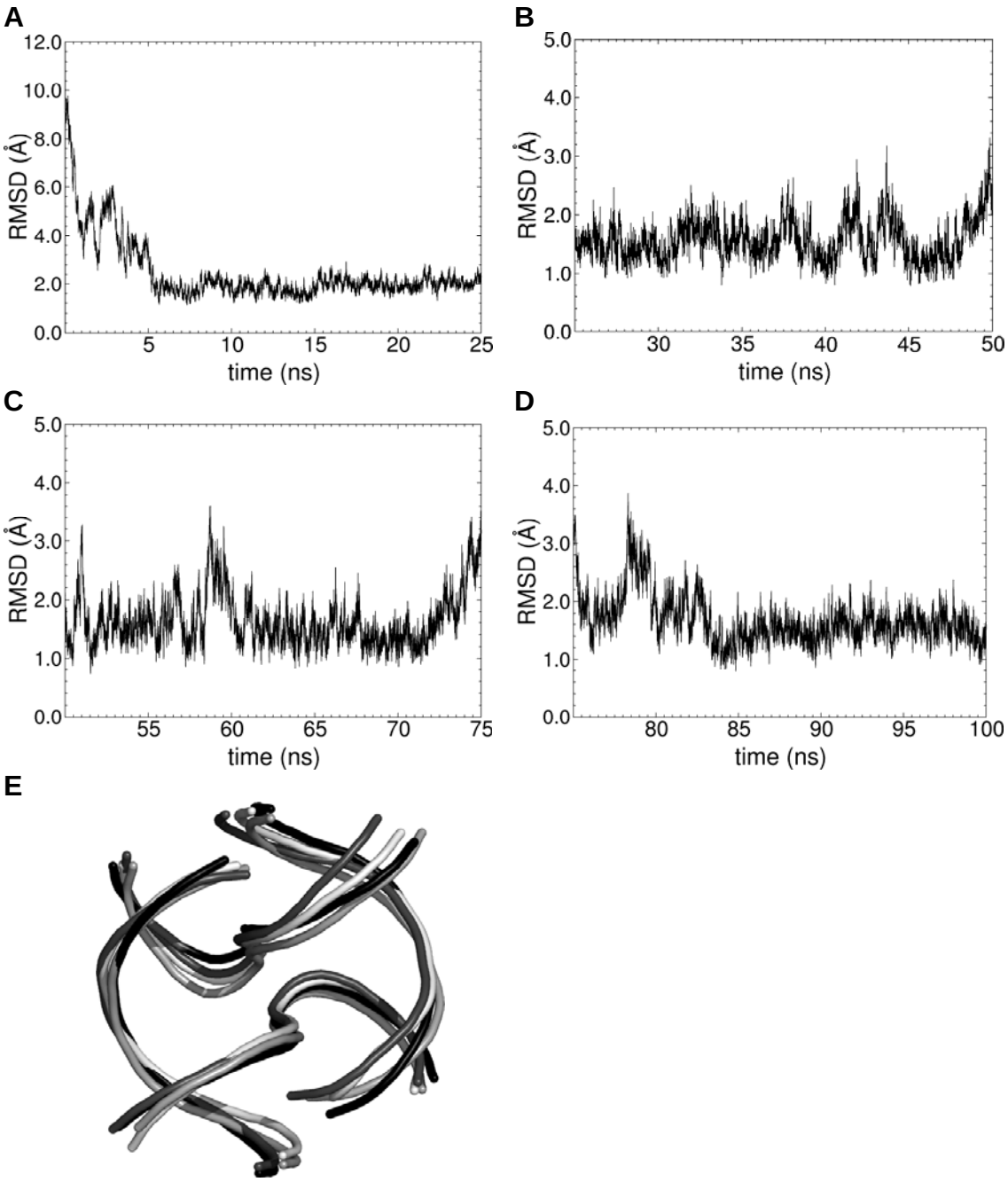

FIGURE S2: Backbone atom (P, O3', O5', C3', C4', C5') RMSD for OPN-HS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four average structures is 0.8 Å.

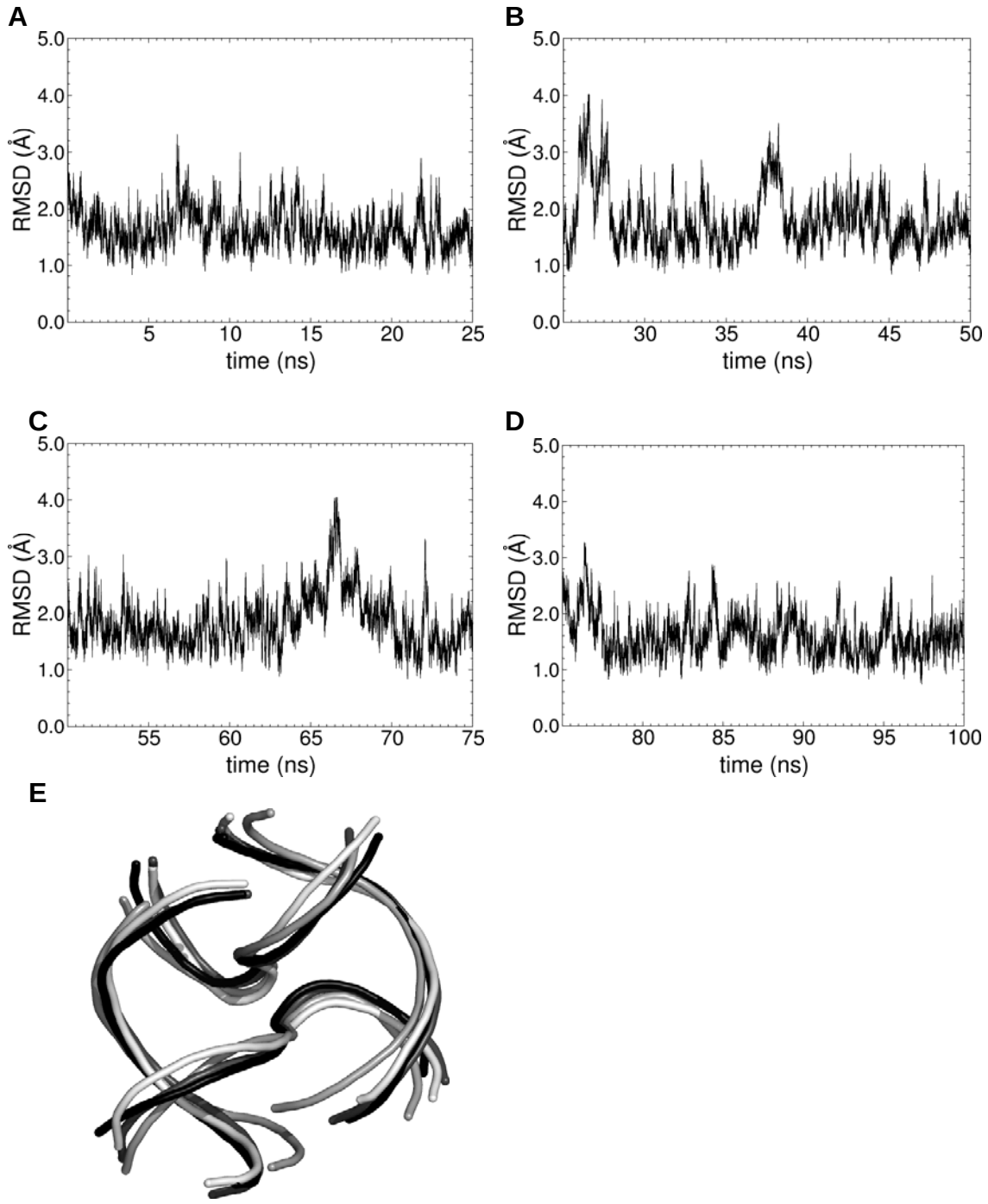

FIGURE S3. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for STX-MS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four average structures is 0.5 Å.

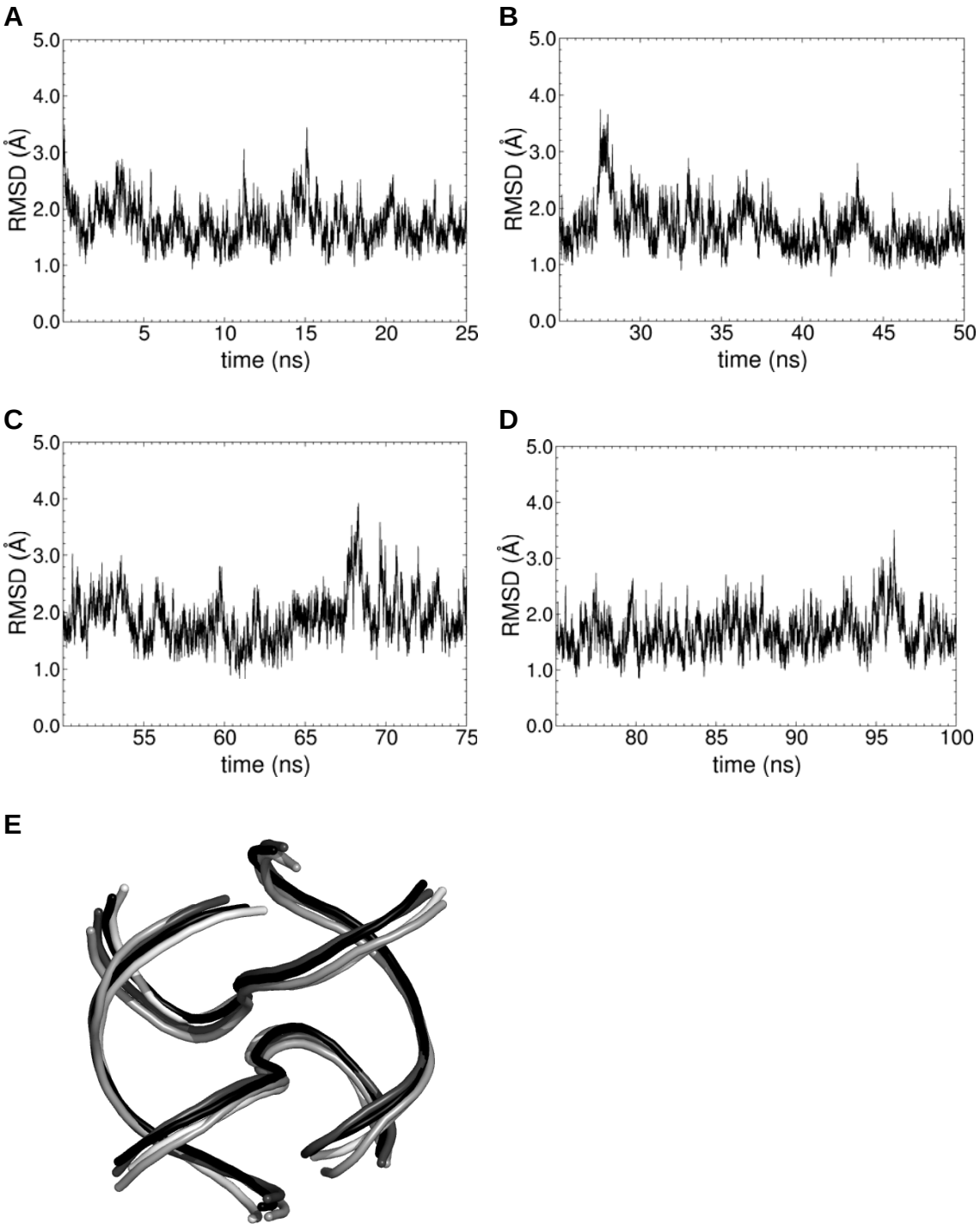

FIGURE S4. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for STX-HS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four average structures is 0.8 Å.

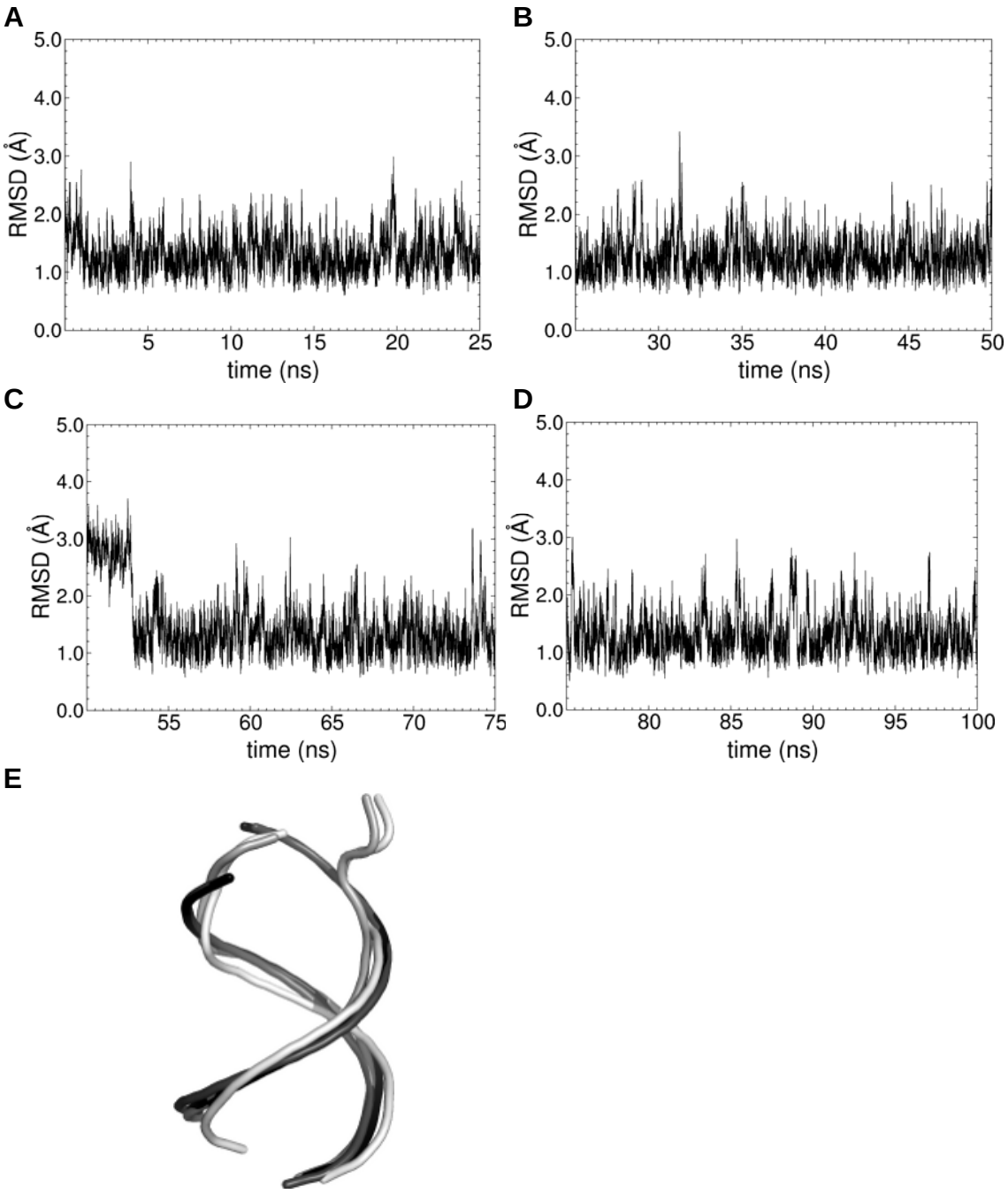

FIGURE S5. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for CBD-MS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four average structures is 1.8 Å.

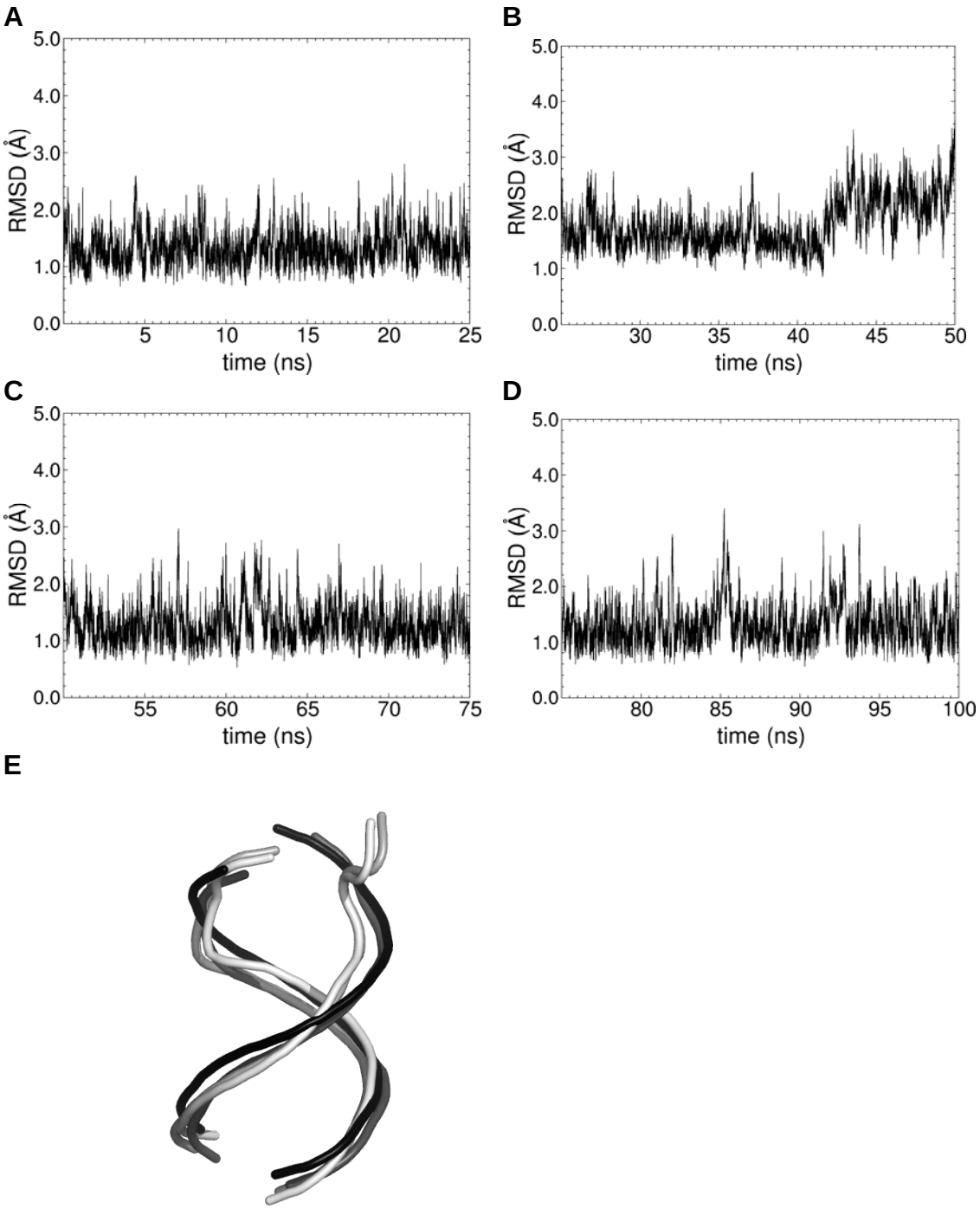

FIGURE S6. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for CBD-HS from (A) 0 – 25 ns, (B) 26 – 50 ns, (C) 51 – 75 ns, and (D) 76 – 100 ns. All snaps are relative to the average structure from the corresponding block of 25 ns. End base pairs are omitted. (E) Overlay of the backbones of the four average structures from each 25 ns block. The black structure is representative of 1 – 25 ns, with progressively lighter tones representative of 26 – 50 ns, 51 – 75, and 76 – 100 ns respectively. Average backbone RMSD between the four average structures is 1.6 Å.

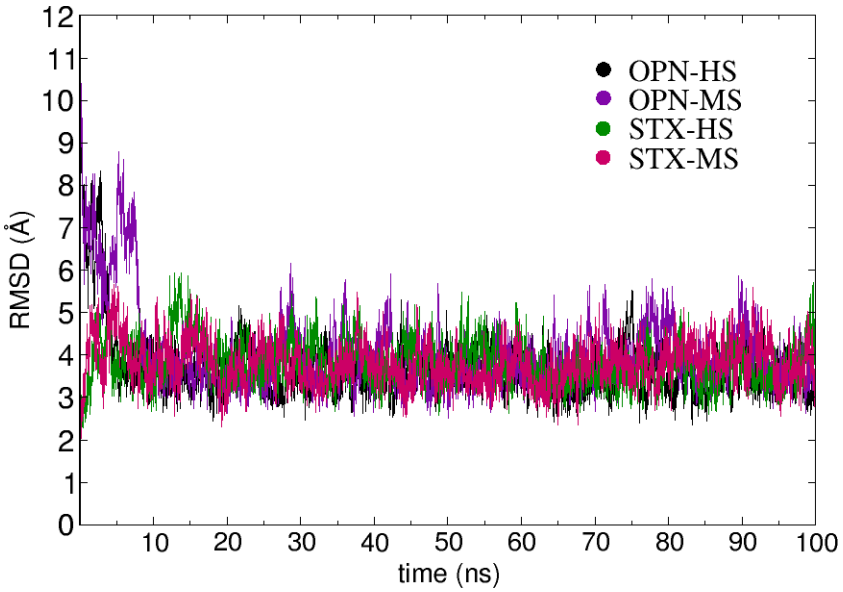

FIGURE S7. Backbone atom (P, O3', O5', C3', C4', C5') RMSD for OPN-HS, OPN-MS, STX-HS, and STX-MS for 100 ns of MD. All snaps are relative to the 1dcw crystal structure. End base pairs are omitted.

S.5.2 2D-RMSD plots of OPN-HS, OPN-MS, STX-HS, and STX-MS

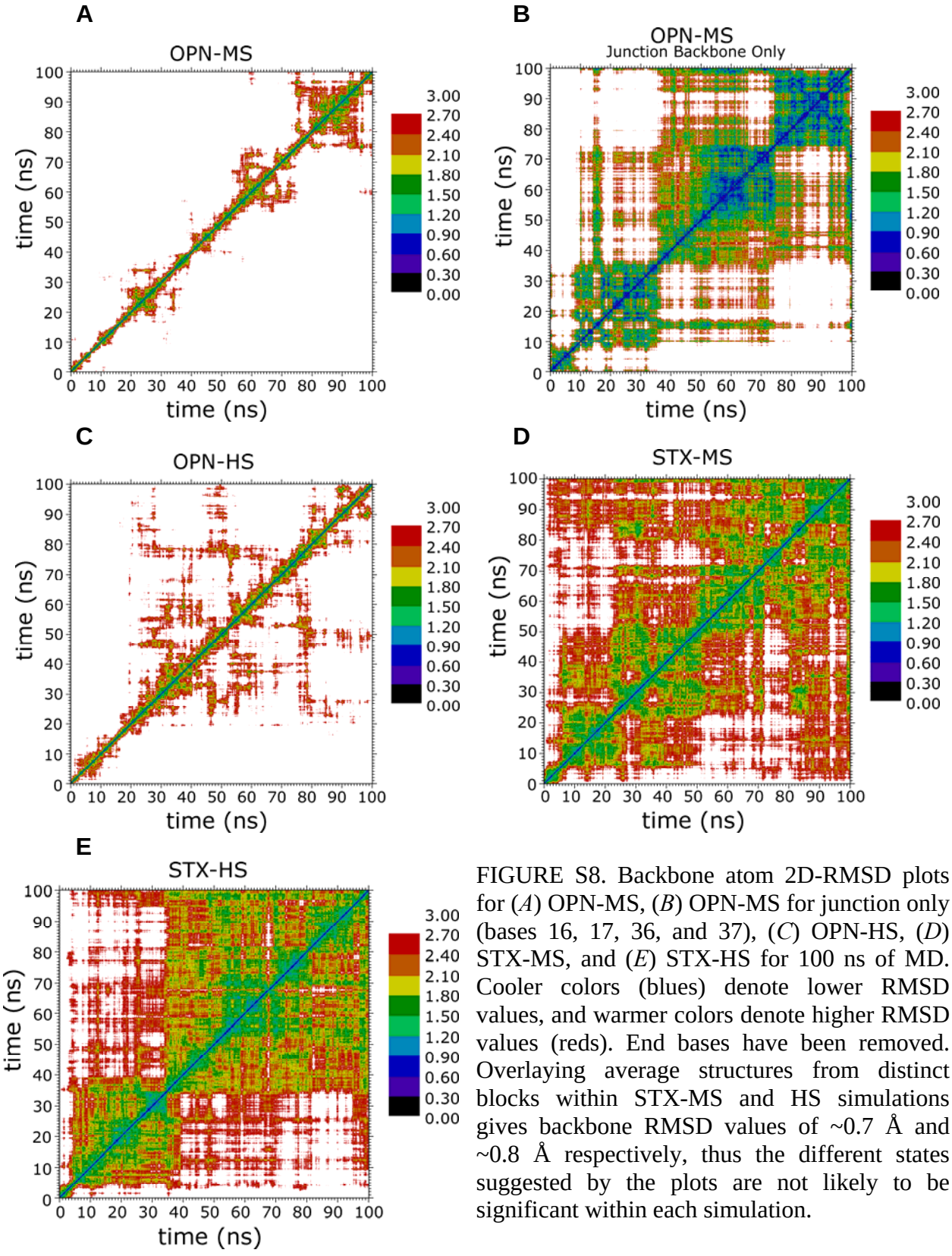

FIGURE S8. Backbone atom 2D-RMSD plots for (*A*) OPN-MS, (*B*) OPN-MS for junction only (bases 16, 17, 36, and 37), (*C*) OPN-HS, (*D*) STX-MS, and (*E*) STX-HS for 100 ns of MD. Cooler colors (blues) denote lower RMSD values, and warmer colors denote higher RMSD values (reds). End bases have been removed. Overlaying average structures from distinct blocks within STX-MS and HS simulations gives backbone RMSD values of ~0.7 Å and ~0.8 Å respectively, thus the different states suggested by the plots are not likely to be significant within each simulation.

S.5.3 Free-Energy Landscapes Derived from 2D-RMSD: Projection of 3D-RMSD Space into 2D

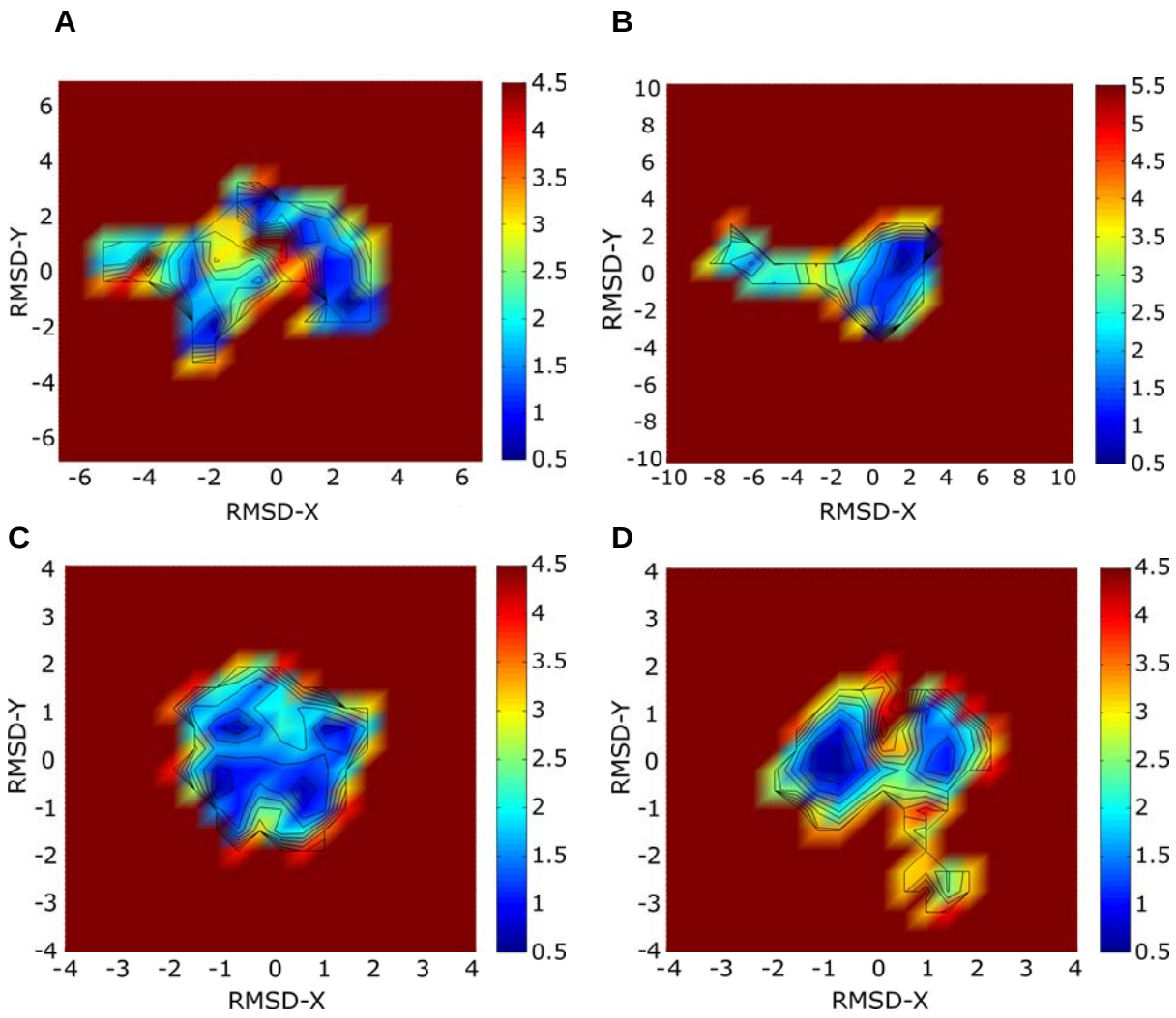

FIGURE S9 Free energy contour maps of 100ns MD simulations derived from RMSD in the x-y direction for (*A*) OPN-MS, (*B*) OPN-HS, (*C*) STX-MS, and (*D*) STX-HS. The scale is shown in kcal/mol, with energetic minima shown in dark blue, and maxima in dark red.

The plots above are energy landscapes for the various simulations in RMSD space. To make these maps, a vector representation of each MD trajectory in 3D RMSD space is obtained. What is presented here are 2D slices of this grid which is sufficient for qualitative interpretation. The population of vector termini P_{ij} associated with each 2D grid element is calculated. From this, normalized probabilities P_{ij}/P_0 are obtained, where P_0 is the maximum population. A free energy is then calculated as $\Delta F_{ij} = - RT \ln (P_{ij}/P_0)$ and displayed as a contour map. The figures were generated using MATLAB and the x and y axes are shown in units of Å.

The two OPN simulations in (*A*) and (*B*) show two minima, one corresponding to the initial OPN form (on the left of both plots) and one lower energy minimum

corresponding to the STX form (on the right of both plots). The minima are separated energetically in these plots by a barrier of ~3 kcal/mol, and spatially by ~6 Å in (*A*) and ~10 Å in (*B*). The OPN-MS simulation took longer to equilibrate, and thus sampled a larger area of 3D-RMSD space than the HS condition during the transition to the STX form.

The STX-MS in (*C*) shows one relatively compact minima spanning ~2 Å spatially, and the STX-HS in (*D*) shows two closely located minima separated by ~2 Å as well, with a slight barrier of 2-2.5 kcal/mol in between. The two minima in (*D*) most likely correspond to the two seemingly distinct conformations shown as blocks in the 2D-RMSD plot for STX-HS (Fig. S8 *E*) at 5-35 ns and 35-100 ns. Backbone RMSD between the average structures from each of these is ~0.8 Å, indicating the two structures are not significantly distinct. Overall, the 2D-RMSD-derived free energy contour maps indicate that the STX forms are occupying a local minimum corresponding to the STX form, and the OPN simulations sample two distinct minima, corresponding to the initial OPN form and the final stabilized STX form. These results are consistent with the progression of the simulation observed in Principal Component Analysis (PCA) as shown in the following section.

S.5.4 PCA plots for STX-MS, STX-HS, OPN-MS, and OPN-HS

We show the results of the first two principal components (q_1 and q_2) of STX-HS, STX-MS, OPN-HS, and OPN-MS trajectories, which account for the majority of the variance, plotted against one another to elucidate the substate sampling of each simulation as a function of time (Fig S9). Alongside each principal component axis is a population distribution plot in red, showing how the population of each principal component, $P_1(q_1)$ and $P_2(q_2)$, is distributed throughout possible substates. The time progression of the simulation is denoted by progression through the visible spectrum with red indicating the initial portion of the simulation and purple the end portion. A comparison of STX-HS, STX-MS, OPN-HS, and OPN-MS indicate that for the ApC junction the salt concentration affects equilibration time and the transitional pathway, but the equilibrated prediction structure for all four systems was the STX form.

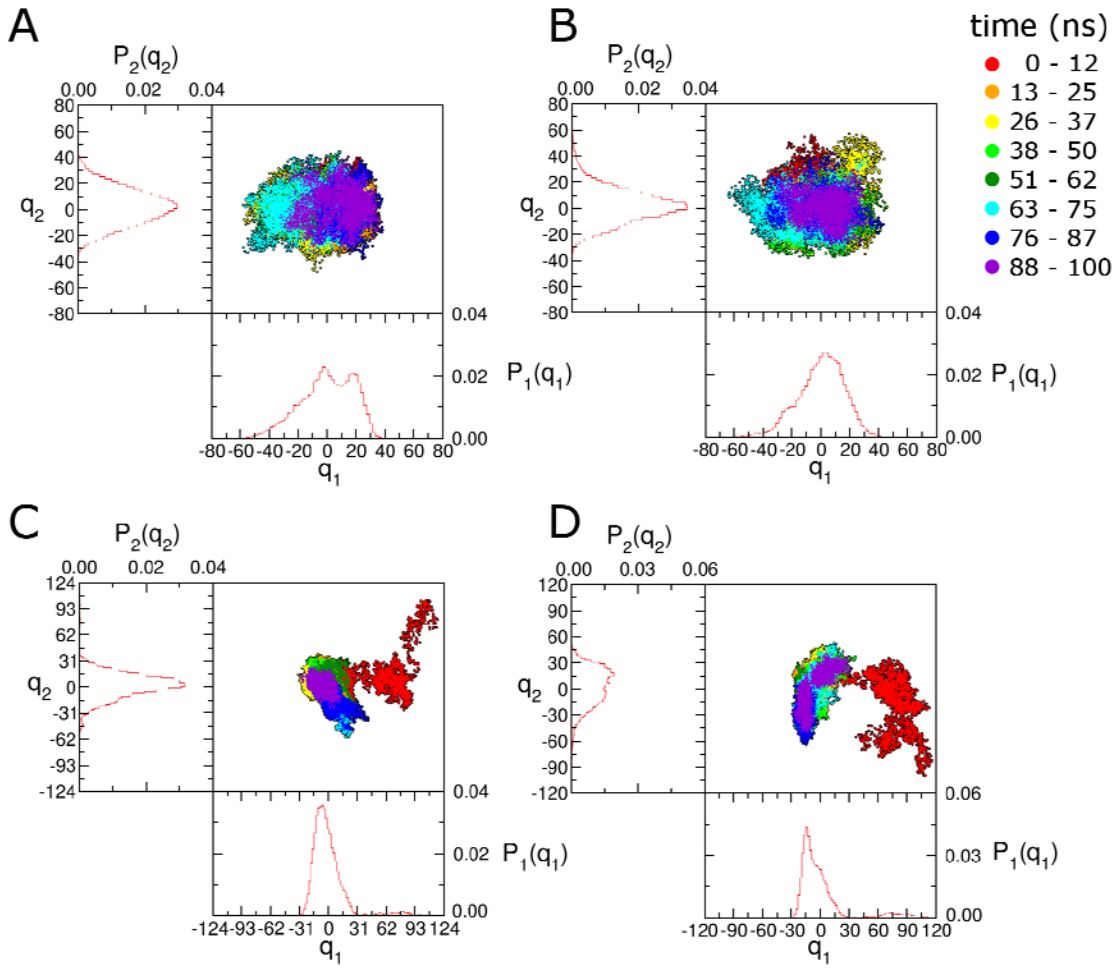

FIGURE S10. Principal component analysis (PCA) of the covariance of atomic fluctuations for MD on (*A*) STX-HS, (*B*) STX-MS, (*C*) OPN-HS and (*D*) OPN-MS simulations. The time progression of the simulation is denoted by progression through the visible spectrum with red indicating the initial portion of the simulation and purple the end portion. The corresponding population distribution plots are shown in red.

S6 Geometrical Comparison of STX-HS and CBD-HS

S.6.1 Geometrical Parameters

Geometrical DNA parameters were calculated using Curves+ and Canal(2) for the first pseudo-duplex of the STX-HS simulation, and the CBD-HS duplex DNA simulation. A comparison was made between the two to elucidate any differences between the MD on junction arms compared with MD on duplex DNA in solution. Base pair parameters x displacement (XDP), inclination, (INC), and base pair step parameters roll (ROL), tilt (TLT), twist (TWS), and rise (RIS) were calculated for the equilibrated portion of each MD. XDP and INC are shown in the main text.

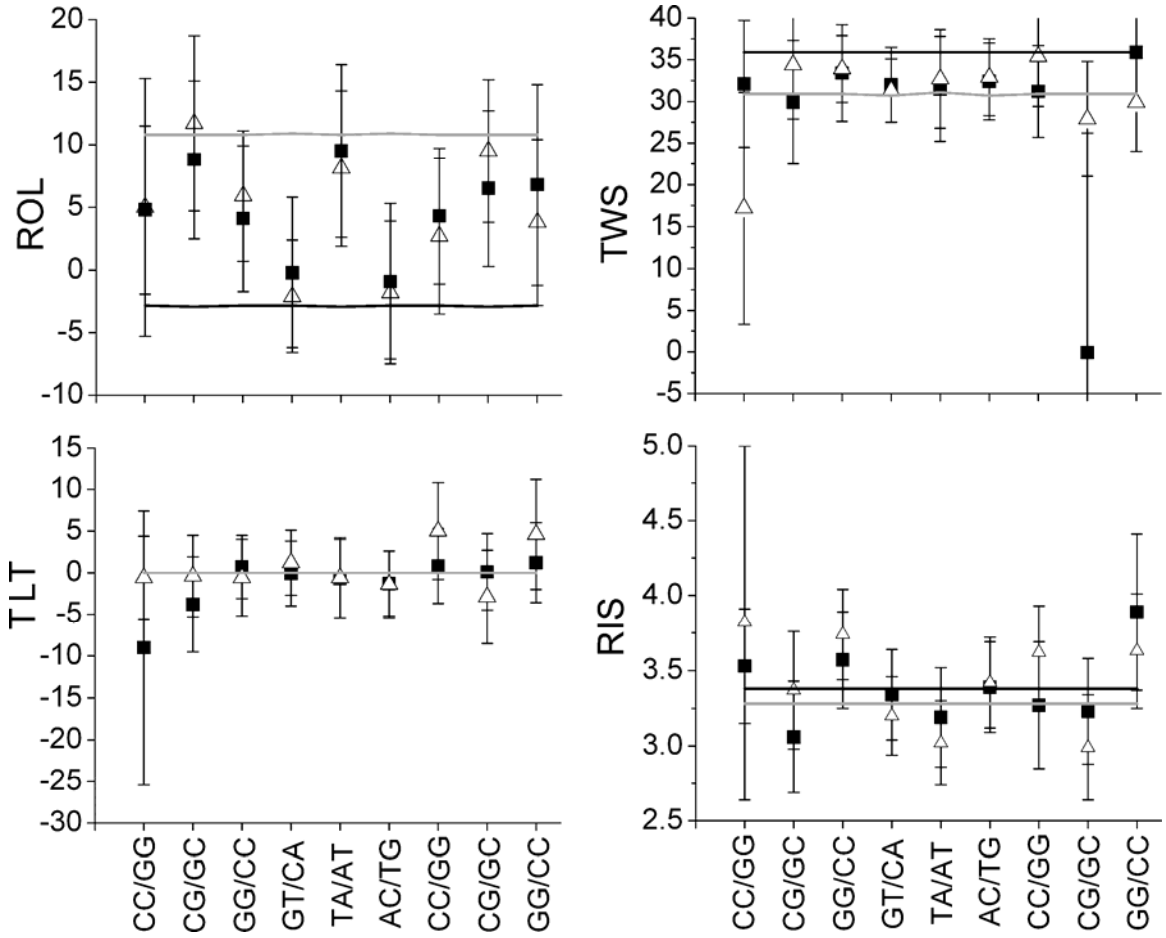

FIGURE S11. Base pair step parameters roll (ROL), tilt (TLT), twist (TWS) and rise (RIS), generated using Curves+ and Canal(2). The black line indicates canonical B DNA values; the gray line indicates canonical A DNA values. Black squares are CBD-HS values, and open triangles are STX-HS values, averaged over 80ns of equilibrated MD from each simulation. Error bars indicate one standard deviation.

S.6.2 Statistical Analysis of Geometrical Parameters

In order to assess the statistical significance of any differences between STX-HS and CBD-HS geometrical parameters, a t-test was performed between the two simulations for each bp or bp step for each parameter. The t values for each parameter and the corresponding percent probability of being statistically different is shown in Table S2. A comprehensive plot of the probabilities of values being statistically significant is shown in Fig. S11. Conventionally, the cutoff for definite statistical significance is 95% probability of being different, and none of the values for any of the six parameters obtained from comparing the STX-HS and CBD-HS exceed 70% confidence. Thus, with respect to finer geometry, the pseudo duplex of the junction is behaving like duplex DNA in solution.

Table S2. T values and corresponding probabilities of being statistically different for geometrical differences between STX-HS and CBD-HS. The level is either the bp or bp step value for each geometrical parameter.

	XDP		**INC**		**ROL**		**TLT**		**TWS**		**RIS**	
Lv.	T val.	% P	T val.	% P	T val.	%P	T val.	%P	T val.	%P	T val.	%P
1	0.80	57.63	0.63	47.13	0.02	1.6	0.49	37.59	0.94	65.28	0.23	18.19
2	0.22	17.41	0.52	39.69	0.31	24.34	0.45	34.73	0.46	35.45	0.58	43.81
3	0.07	5.58	0.3	23.58	0.23	18.19	0.22	17.41	0.07	5.58	0.39	30.35
4	0.31	24.34	0.42	32.55	0.25	19.74	0.24	18.97	0.12	9.55	0.35	27.37
5	0.20	15.85	0.06	4.78	0.15	11.92	0.01	0.80	0.14	11.13	0.39	30.35
6	0.20	15.85	0.54	41.08	0.11	8.76	0.02	1.60	0.08	6.38	0.07	5.58
7	0.16	12.71	0.74	54.07	0.19	15.07	0.57	43.13	0.52	39.69	0.67	49.71
8	0.45	34.73	0.89	62.65	0.36	28.12	0.41	31.82	1.03	69.70	0.48	36.88
9	0.66	49.07	0.79	57.05	0.29	22.82	0.42	32.55	0.70	51.61	0.40	31.08
10	0.58	43.81	0.77	55.87								

The standard cutoff for confidence that two numbers are statistically different is 95%, and all values observed are < 70%, indicating there are no statistically significant differences in geometrical parameters between the pseudo duplex in STX-HS and the duplex CBD-HS DNA.

S7 Independent MD Simulations for OPN-MS

Five independent MD simulations of the OPN-MS system were performed in addition to the original simulation discussed in the text. Each repeat simulation was assigned a different random seed number for the starting velocities. Of the six total simulations, all six equilibrated to the antiparallel STX conformer, with five stacked in the *Iso II* conformation and one stacked in the *Iso I* conformation. Although the sequences of the four ApC junction strands are identical, the junction itself is not symmetric because there are two 4 bp arms and two 6 bp arms, and it has a major groove face and a minor groove face. The *Iso I* and *Iso II* naming convention is used to describe different conformations of the antiparallel STX structure of a Holliday junction, shown schematically in Fig. S12. The naming convention is derived from a junction of a different sequence, and the *Iso II*

conformation is more commonly observed, so we have designated the more commonly observed conformation of our simulations on 1dcw as such.

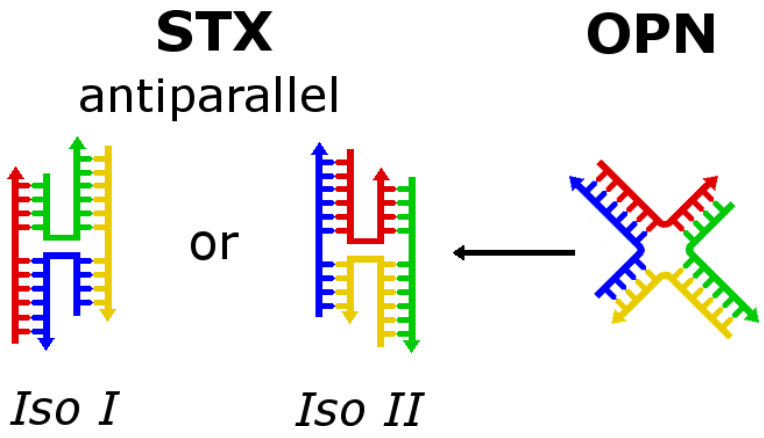

FIGURE S12. Schematic representation of the two different antiparallel STX conformations that the ApC junction can adopt. The schematic assumes the minor groove "face" of the junction is facing the reader. The *Iso II* conformation is the conformation adopted in the 1dcw crystal structure, and was used as the starting configuration for the STX-MS and STX-HS systems (which remained stacked in this conformation).

Of the six total OPN-MS simulations, five (the original OPN-MS simulation plus independent MD simulations 1-4) folded into the *Iso II* conformation. Of these five simulations, four equilibrated to folded structures with similar J_{twist} values of ~40° and an average backbone RMSD value of ~2 Å, while one folded with a relative J_{twist} value of ~ -70° and an average backbone RMSD value of ~9 Å from the other four simulations. Independent MD simulation 5 folded in the less commonly observed *Iso I* conformation, and also had a relative J_{twist} value of ~ -70°. See Fig. S13 for details. J_{twist} values were approximated by using the dihedral angle measuring function in VMD to measure torsions between axes calculated using Curves+(2).

S.7.1 RMSD of Independent MD Simulations for OPN-MS

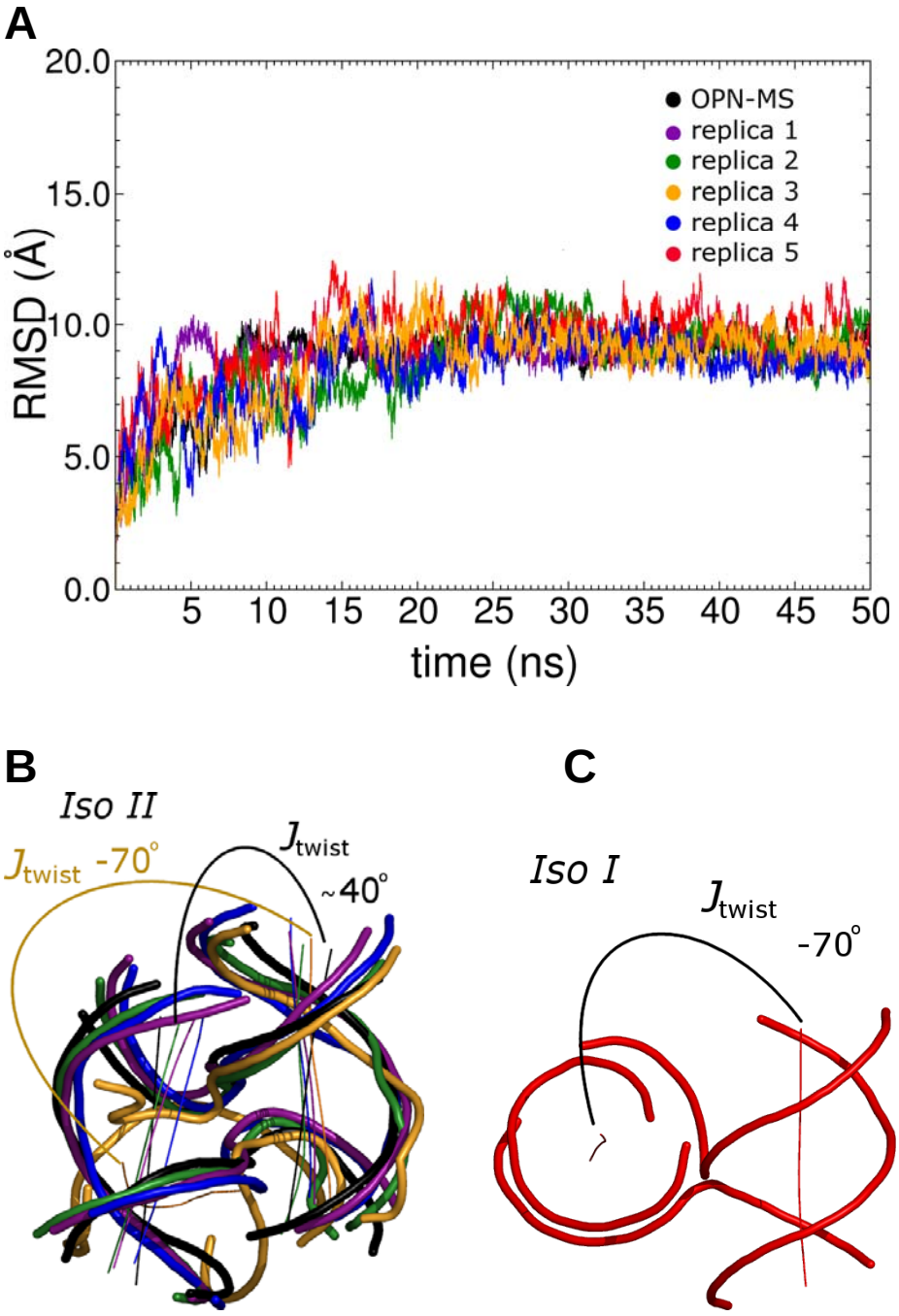

FIGURE S13. (*A*) Backbone atom (P, O3', O5', C3', C4', C5') RMSD for OPN-MS and the 5 independent MD simulations from 0 – 50 ns. All snaps are relative to the first and end base pairs are omitted. (*B*) Overlay of average structures for OPN-MS and independent MD simulations 1-4 from 35 – 50 ns of MD. Colors match the legend in (*A*) and J_{twist} values are shown for the four similar simulations and for the slightly different independent simulation 3 (yellow). (*C*) Average structure for the outlier independent simulation 5 from 35 – 50 ns MD, which folded into the *Iso II* conformation.

S.7.2 PCA of Independent MD Simulations for OPN-MS

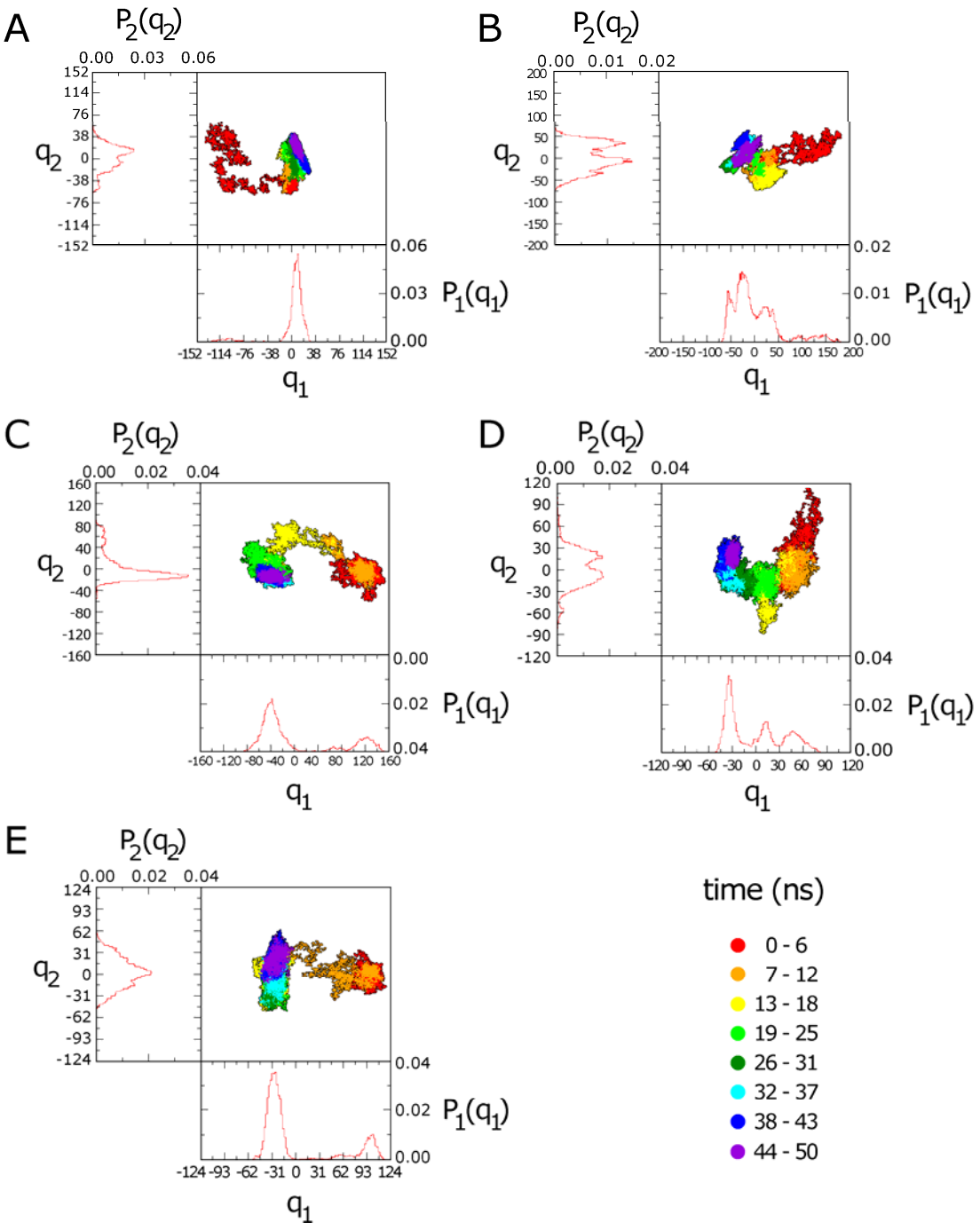

FIGURE S14. Principal Component Analysis (PCA) for independent MD simulations 1(*A*), 2(*B*), 3(*C*), 4(*D*), and 5(*E*) of OPN-MS for 50ns of MD. Progression through the simulation is denoted by progression through the color spectrum, with red indicating the initial portion and purple indicating the end part of the simulation. The population distributions of principal component one, $P_1(q_1)$, and two, $P_2(q_2)$, are shown in red below and to the left of each colored plot, respectively.

S8 Information about Figures in Main Text and SI

The portions of figures containing DNA structures were generated using the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

Ion density plots were visualized along with each corresponding average structure using the graphics program Chimera from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081).

All PCA and RMSD plots were created using Xmgrace (P. J. Turner and Grace development team), and Figs. 3, 4, and S10 were created using Origin (OriginLab). Figure S9 was generated using MATLAB.

Figures were edited using Gimp ver. 2.6.

SUPPORTING REFERENCES

1. Cheatham, T. E. 3rd, Case, David A., T.A. Darden, C.L. Simmerling, J. Wang, R.E. Duke, R., K. M. M. Luo, D.A. Pearlman, M. Crowley, R.C. Walker, W. Zhang, B. Wang, S., A. R. Hayik, G. Seabra, K.F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H., L. Y. Gohlke, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D.H. Mathews, C., and W. S. R. Schafmeister, and P.A. Kollman. 2006. AMBER 9. University of California, San Francisco.
2. Lavery, R., M. Moakher, J. H. Maddocks, D. Petkeviciute, and K. Zakrzewska. 2009. Conformational analysis of nucleic acids revisited: Curves+. Nucleic acids research 37:5917-5929.