

# Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent

Hai Nguyen, James Maier, He Huang, Victoria Perrone and Carlos Simmerling

## SUPPORTING INFORMATION

### Methods

#### ff14SBonlysc

Our ff14SBonlysc force field, freely available as part of AmberTools 14 from the Amber web site at ambermd.org, used the backbone dihedral corrections of ff99SB<sup>1</sup> with updated dihedral side chain corrections fit to MP2<sup>2</sup>/6-31+G\*\*3//HF/6-31G\*<sup>3</sup> ab initio side chain energy surfaces of dipeptides at  $\alpha$  (-60°, -45°) and  $\beta$  (-135°, 135°) backbone conformations; all other parameters were from ff94<sup>4</sup>. To limit variability to predominantly side chain motions and to limit backbone-side chain hydrogen bonding that may be incorrectly modeled by fixed charges in vacuo, all backbone dihedrals were restrained. Valine was fit to 10°  $\chi$ 1 scans. Aspartate (ionic and neutral), asparagine, cysteine, isoleucine, leucine, serine, threonine, phenylalanine, tyrosine, tryptophan, and histidine ( $\delta$ -,  $\epsilon$ -, and doubly-protonated) were fit to 20°  $\chi$ 1 and  $\chi$ 2 two-dimensional scans. Glutamate (ionic and neutral), glutamine, and methionine were fit to randomly distributed conformations extracted from high temperature simulations. Quantum calculations of the one- and two-dimensional scans employed GAMESS (US) (1 MAY 2012 (R1))<sup>5</sup> whereas quantum calculations of the structures from high temperature simulation employed Gaussian 98<sup>6</sup>. Molecular mechanics calculations were performed using Amber 11 and 12<sup>7,8</sup>. Fitting was performed by a genetic algorithm<sup>9</sup> using GALib<sup>10</sup>, with parameters restrained to phase shifts of 0 or  $\pi$  to permit simulation of different enantiomers. A complete description of the parameter development will be published elsewhere.

#### MD

All MD simulations were carried out using the GPU implementation<sup>11</sup> of the pmemd program in Amber14<sup>12</sup> with the combination of GB-Neck2 model,<sup>13</sup> mbondi3 radii,<sup>13</sup> and ff14SBonlysc. We did not use the modified backbone dihedral parameters from ff14SB, since they involve empirical adjustments to ff99SB aimed at improving agreement between experiment and simulations in TIP3P explicit water.

There are many potential limitations of simple implicit solvent models, such as lack of structured water and ions. In addition, nonpolar solvation contributions were not included in this work, as methods for their accurate treatment are less well developed, and their treatment *via* surface area (as done in Amber) is overly simplistic, significantly slows the calculations, and has been reported to bias nonpolar interactions.<sup>14</sup> Although the hydrophobic effect plays a major role in protein folding, we note that neglecting nonpolar solvation also omits the attractive dispersion interaction with solvent, partially compensating for the hydrophobic effect<sup>15,16</sup>. On the whole, it seems reasonable to test our model without the nonpolar term, since we showed previously that simulations without the nonpolar term performed well on smaller peptide systems.<sup>30</sup> In the section for each system below, we provide figures showing the SASA as a function of RMSD for each system, which provides a qualitative indication of the potential impact of including the SASA term in the simulations.

Initial structures were built using the LEaP module of AmberTools<sup>17</sup> then minimized and equilibrated in three 250 ps stages: heating from 100 K to the production temperature with heavy atom positional restraints of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup>, reducing force constant from 10.0 to 1.0 and then to 0.1 kcal mol<sup>-1</sup> Å<sup>-2</sup>. A time step of 4 fs was used with hydrogen mass repartitioning.<sup>18,19</sup> Bonds involving hydrogen were constrained by the SHAKE algorithm<sup>20</sup> with a tolerance of 0.00001. Temperature was controlled with a Langevin thermostat with collision frequency  $\gamma$  = 1.0 ps<sup>-1</sup>. We used 300 K except as follows. We initially used 300 K for Fip35; as the native structure was stable for 10  $\mu$ s, however, the temperature was raised to 325 K to aid folding. We used the same temperature for GTT, which is a variant of Fip35. HP36 and BBL unfolded within tens of ns at 300 K, so these systems were simulated at 290 K. Our subsequent use of REMD avoids the need for selecting a single optimal folding temperature.

## Clustering

Means algorithm was used with distance defined by Ca RMSD to generate 50 clusters using default settings in ptraj<sup>21</sup>. The clustering for REMD was performed for the lowest temperature trajectory. Snapshots were used from 5 ns intervals, but this interval was adjusted to ensure between 4000 and 7000 frames.

## Native Contacts Analysis

To provide an alternate measure of folded structure quality, we determined the fraction native contacts at each timestep during our folding trajectories. Contacts were evaluated to occur whenever Ca atoms were 6 Å apart or less, using the residue masks defined above. First, we determined the native contacts by examining the experimental structure. Then, we obtained a baseline of the unfolded state by counting the native contacts for our completely extended initial conformations. Then, for each frame in a trajectory, we counted the number of native contacts, subtracted the baseline number of contacts in the extended state, and then divided by the difference between the number of native contacts and the number of native contacts formed in the extended state. Thus we analyzed the degree of progress from completely extended to completely folded by the number of native contacts. Native contacts were counted via the 'contacts' command of cpptraj<sup>22</sup>.

## Nonpolar Solvation Analysis

Structures were extracted every 1 ns from extended and native MD simulations. The combined set was postprocessed in SANDER to calculate the cavity contribution to nonpolar solvation (gbsa=2 in Amber), which is proportional to the solvent-accessible surface area determined by recursively optimizing spheres around each atom starting from icosahedra<sup>23</sup>. We then generated population histograms of surface area contribution versus RMSD to the native structure with grid spacing of 0.5 kcal mol<sup>-1</sup> in nonpolar solvation energy and 0.5 Å in RMSD.

## Protein folding events (Fip35)

Folded and unfolded conformational cutoffs were assigned by visual inspection of two-dimensional RMSD population histograms (RMSD values for hairpin 1 and for hairpin 2). The Fip35 folded cutoffs were 2.7 and 1.2 Å RMSD for hairpins 1 and 2, respectively. The Fip35 unfolded cutoffs were 5.0 and 4.5 Å RMSD for hairpins 1 and 2, respectively. These numbers were empirically selected to reflect visual boundaries in population around the two states. Whenever a structure went above the two unfolded cutoffs (both hairpins unfolded), it was considered unfolded. Whenever a structure went below the two folded cutoffs, it was considered folded. The total simulation period between an unfolded conformation and a folded conformation was considered a folding path. Each path was plotted in two-dimensional RMSD, with lines colored by time through red, yellow, green, cyan, and blue. The sequence of folding was determined by manually evaluating which RMSD dropped first—visually, whether the folded state was reached from the top (metric on x-axis folded first), the side (metric on y-axis folded first), or diagonally from a conformation where neither was pre-formed.

## Order parameter calculations

Lipari-Szabo NH librational order parameters  $S^{2,24}$  were calculated using the cpptraj<sup>21</sup> implementation of iRED<sup>25</sup>, which does not require separation of internal and external motions, over 8 ns windows for lysozyme, as done elsewhere<sup>26</sup>, and 5 ns for cold shock protein, consistent with its tumbling time<sup>27</sup>, in each trajectory. Uncertainties were determined by standard errors in the average  $S^2$  for each trajectory. Simulations for order parameter calculations were performed with a 1 fs timestep. GB-Neck2 simulations used Langevin dynamics with a constant of 91 ps<sup>-1</sup> to mimic water viscosity<sup>28,29</sup>. TIP3P simulations did not use barostat or thermostat following equilibration. Lysozyme was extended to 96 ns simulation time, and cold shock protein to 60 ns, yielding 12 windows per simulation.

TIP3P<sup>30</sup> simulations used the particle-mesh Ewald approximation<sup>31</sup> with a direct non-bonded cutoff of 8 Å. Equilibration proceeded by minimization of the experimental structure with 100 kcal/mol/Å<sup>2</sup> restraints on protein heavy atoms, followed by 100 ps of restrained heating at constant volume from 100 K to 300 K using the weak-coupling (Berendsen) thermostat<sup>32</sup>. Following 100 ps at 300 K and constant volume, the pressure was equilibrated to 1 bar with isotropic position scaling, for 100 and 250 ps with time constants of 100 fs and then 500 fs and restraints of 100 kcal/mol/Å<sup>2</sup> and then 10 kcal/mol/Å<sup>2</sup>. Then the N, Ca, and C were restrained during minimization, followed by three 100 ps simulations with temperature and pressure time constants of 500 fs, reducing restraints from 10 kcal/mol/Å<sup>2</sup> to 1 kcal/mol/Å<sup>2</sup>, and then 0.1 kcal/mol/Å<sup>2</sup>. Finally, the volume of the unrestrained system was equilibrated with time constants of 1 ps with a 2 fs time step, removing center-of-mass translation every ps, for 1 ns.

NVE production simulations used a direct sum tolerance of 10<sup>-6</sup> and SHAKE<sup>33</sup> applied to bonds to hydrogen with a tolerance of 10<sup>-6</sup> Å.

## General REMD setup

Temperature ranges for REMD were chosen for an acceptance ratio of ~0.25. The number of replicas ranged from 6 to 24, depending on system size (Table S2). Exchanges were attempted every 1 ps. Snapshots were saved every 100 ps.

## Extended REMD (extREMD)

Extended REMD refers to REMD simulations initiated from fully extended initial structures (built by LEaP in Amber) for all replicas. We performed 17 REMD runs for 17 systems starting from extended conformations. Temperatures are indicated in Table S3.

We performed additional REMD calculations for hypothetical protein 1WHZ using final snapshots from extended MD run (named extMDREMD).

### Seeded REMD

The goal of running seeded REMD was to indicate which structure (folded or unfolded) is preferred at low temperatures, under conditions in which all of the structures of interest are present in the replica set at the same time. Even though all clusters may have been sampled in the REMD run, they may not have been sampled at the same time, and thus the temperature of the replicas (or the population of the clusters at various temperatures) does not indicate relative favorability. Although the seeding procedure does provide REMD with the opportunity to rank the structures, the resulting “melting” behavior is artificial, since it depends on the numbers of structures of each type used in seeding. We performed seeded REMD for NuG2 variant, CspA, Lambda Repressor, 1WHZ, and Top7. Temperatures are indicated in Table S4.

We performed two seeded REMD simulations with NuG2 variant. In the first, we continued the extREMD calculation, but adding 2 native structures (from an MD run of the crystal structure) at 2 new temperatures in the middle of the previous temperature ladder: 309.0 K and 334.0 K. In the second seeded REMD, we alternated the most populated cluster from extREMD (11.4 Å RMSD) and a native-like structure (1.0 Å RMSD) through twelve temperatures beginning at 250.0 K.

For CspA, we alternated misfolded (10.0 Å), near-native (4.7 Å), and native-like (1.2 Å) cluster structures through twelve temperatures beginning at 250.0 K.

For  $\lambda$ -repressor, we alternated misfolded (12.0 Å), the lowest RMSD from extREMD, and native structures through twelve temperatures beginning at 250.0 K.

For hypothetical protein 1WHZ, we alternated 2 unfolded (10.6 Å, 10.0 Å), 2 partly folded (3.1 Å, 4.2 Å), and 1 native-like (1.5 Å) replicas through twenty temperatures beginning at 242.0 K.

For Top7, we alternated partly folded (2.7 Å), unfolded (11.2 Å), and native-like (1.5 Å) replicas through eighteen temperatures beginning with 240.0 K.

Seeding REMD was run for ~40ns for all cases. This was determined to be adequate to sort the replicas such that all replicas starting from the same structure were grouped in a continuous temperature range, as compared to the alternation that was used at the start. The simulation length was also short enough that they generally did not sample large structure changes, since this is not the goal in these calculations.

### RMSD calculation

RMSD calculations and cluster analysis were performed with ptraj<sup>21</sup> in AmberTools<sup>17</sup>. RMSD calculations excluded flexible termini or other regions, such as loops, that were not well defined in the crystal structure or family of NMR structures (as described below and tabulated in Table S2). The reference structure was the experimentally derived structure or, where none was available, the structure of a homologue as described.

#### CLN025

We simulated full-length CLN025. All the C $\alpha$  atoms in the x-ray structure<sup>34</sup> were used to calculate RMSD.

#### Trp-cage

We simulated full-length Trp-cage tc5b. We calculated RMSD against the first model of the NMR ensemble<sup>35</sup>, excluding residues 1 to 2 and 19 to 20 as flexible termini.

#### BBA

We simulated full-length BBA. We calculated RMSD against the first model of the NMR ensemble<sup>36</sup>, excluding residues 1 to 3 and 27 to 28 as flexible termini.

#### Pin1 WW domain mutants: Fip35 and GTT

2F21 is a fast-folding Pin1 WW domain mutant<sup>37</sup>. Fip35 is a faster folding (13  $\mu$ s) mutant based on residues 6–38 of 2F21<sup>38</sup>. GTT is an even faster folding (4  $\mu$ s) mutant based on Fip35 plus the two prior residues in 2F21<sup>39</sup>. We simulated full-length Fip35 and GTT. We calculated RMSDs against 2F21, using residues 10 to 32—the first residue of the first  $\beta$ -strand to the final residue in the last  $\beta$ -strand.

#### HP36

We simulated the thermostable C-terminal fragment (residues 41 to 76) of the chicken villin headpiece (HP36). Shaw and colleagues<sup>40</sup> used the HP35 variant of villin peptide with norleucine double mutant<sup>41</sup> to accelerate folding. We chose the HP36 variant<sup>41</sup>, which includes only standard amino acids. We calculated RMSDs against an averaged NMR structure (PDB ID: 1VII)<sup>41</sup>, using residues 43 to 72 regions to exclude flexible termini of the NMR ensemble of a G34L mutant<sup>42</sup>.

### **NTL9 (39 AA)**

NTL9 (39) is an N-terminal truncation (residues 1 to 39) of N-terminal domain of 50S ribosomal protein L9 (NTL9). We simulated the K12M mutant. We calculated RMSDs against residues 1 to 39 of the crystal structure of the full-length K12M sequence (PDB ID: 2HBA)<sup>43</sup>.

### **BBL**

We simulated the H142W mutant of BBL, residues 124 to 170—the residues in a solution structure ensemble<sup>44</sup> (PDB ID: 2WXC). We calculated RMSDs against the experimental structure skipping the flexible N-terminal residues and the flexible loop from residue 152 to 158. Our mask thus included residues 133–151 and 159–170.

### **Protein B**

We simulated a K51/K39V double mutant of truncated Protein B (residues 7–53) of the NMR structure (PDB ID: 1PRB), as done previously<sup>7</sup>. We calculated RMSDs against the NMR structure using residues 8–50, including the start of the first helix to the end of the last helix.

### **Homeodomain**

We simulated a computationally re-designed variant of *Drosophila Melanogaster* Engrailed homeodomain<sup>45</sup>. We calculated RMSDs against the first model of the NMR ensemble (PDB ID: 2P6J)<sup>45</sup>, using residues 5 to 48 to exclude flexible termini.

### **NTL9 (52 AA)**

In addition to the 39 residue NTL9 described above, we also simulated the full length N-terminal domain of the 50S ribosomal protein L9 (NTL9) K12M<sup>43</sup>, denoted as NTL9 (52 AA). We calculated RMSDs against the first monomer in the crystal structure (PDB ID: 2HBA).

### **NuG2 variant**

We simulated residues 6 to 61 of a N37A/A46D/D47A mutant of NuG2 (PDB ID: 1MI0<sup>46</sup>), as done previously<sup>40</sup>. We calculated RMSDs against the crystal structure of unmodified NuG2, including all simulated residues from 6 to 61.

### **CspA**

We simulated the major cold shock protein of *Escherichia coli*, CspA, excluding the first residue missing from the x-ray structure (PDB ID: 1MJC<sup>47</sup>). We calculated RMSD against all structured regions in the experimental structure (residues 4–14, 16–23, 29–36, 48–56, and 62–70), as the loops are flexible in an NMR ensemble of the same sequence (PDB ID: 3MEF<sup>27</sup>).

### **Hyp protein 1WHZ**

We simulated full-length hypothetical protein from *Thermus thermophilus* HB8. We calculated RMSDs against all residues in the crystal structure (PDB ID: 1WHZ<sup>48</sup>).

### **$\alpha$ 3D**

We simulated full-length  $\alpha$ 3D, a de novo designed three-helix bundle<sup>49</sup>. We calculated RMSDs against all residues in the solution structure (PDB ID: 2A3D<sup>49</sup>).

### **$\lambda$ -repressor**

We simulated truncated (residues 6–85) monomeric D14A/Y22W/Q33Y/G46A/G48A mutant of  $\lambda$ -repressor studied previously<sup>40</sup>. We calculated RMSD against the unmutated homologue (PDB ID: 1LMB<sup>50</sup>), however. In the x-ray structure, dimeric  $\lambda$ -repressor binds DNA, along with multivalent ions. We calculated RMSDs against all simulated residues (6–85) in the first protein chain in this complex.

### **Top7**

We simulated residues 3 to 94 of Top7, which was computationally designed with a novel fold<sup>51</sup>. We calculated RMSDs against the x-ray structure (PDB ID: 1QYS<sup>51</sup>), using residues 3 to 94.

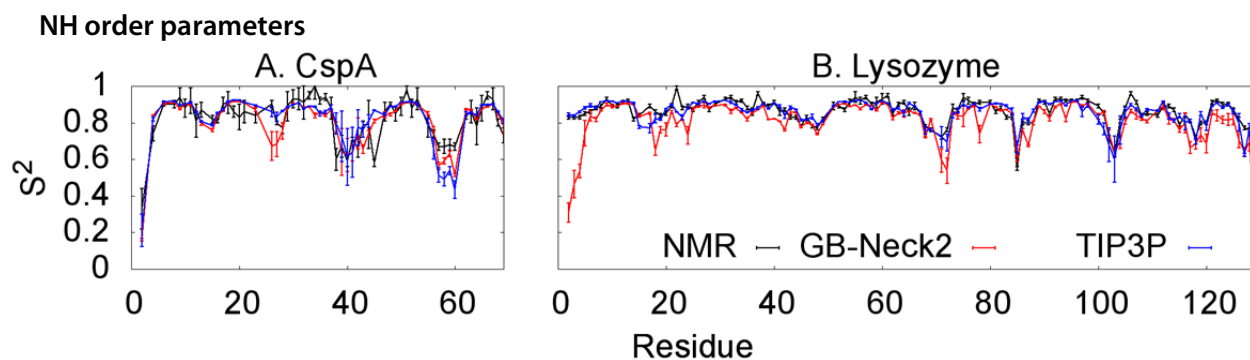


Figure S1. Order parameters measuring the NH librational motions of (A) CspA and (B) lysozyme according to NMR<sup>27,52</sup> (black), GB-Neck2 (red), and TIP3P (blue). All simulation data used force field 14SbOnlysc with order parameters backcalculated by iRED. Error bars reflect the standard deviation of the averages from windows in the simulation.

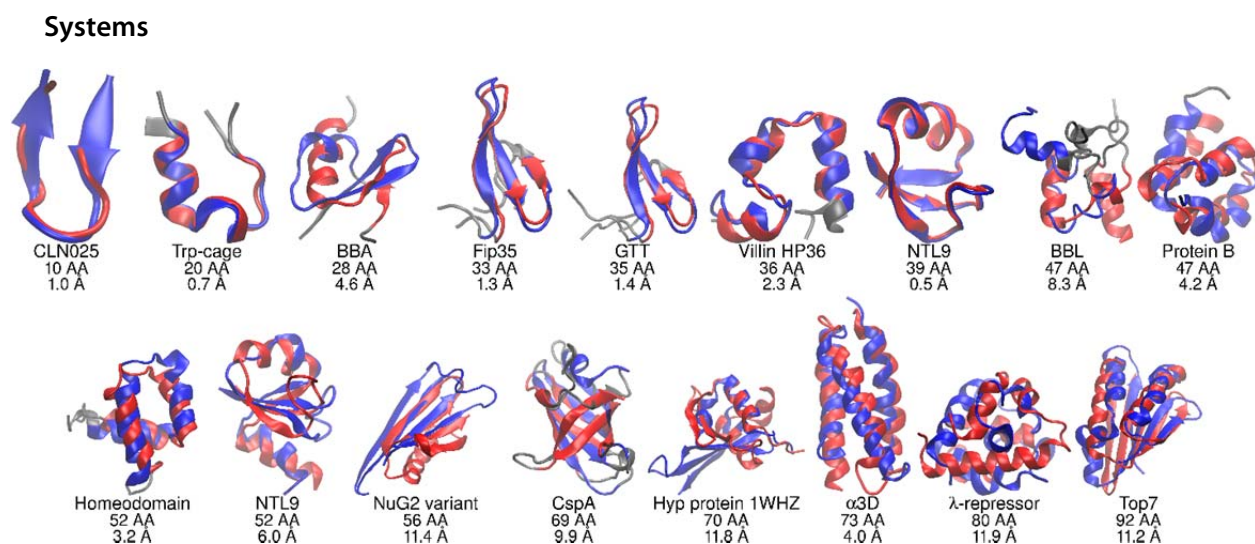


Figure S2. The most populated cluster of each protein starting from extended REMD simulations, in blue, aligned to the experimental structure, in red. In the cases of BBA, BBL, NuG2 variant, CspA, Hyp protein 1WHZ, and Top7, the alignment above reflects the parts of the structure best reproduced by the simulations, rather than the alignment yielding the lowest RMSD. Below each rendering is the system name, the number of amino acids (AA), and the RMSD between the two structures (neglecting the flexible gray regions, as in Figure 1).

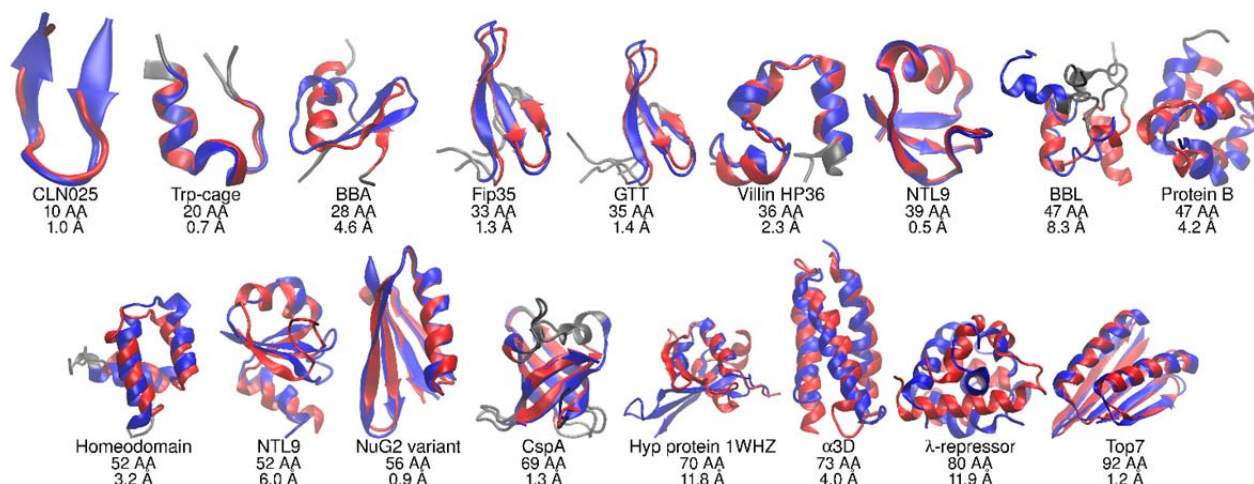


Figure S3. The structure of each protein preferred by the force field, either: the centroid of the most populated cluster from extended REMD; or, as in NuG2 variant, CspA, and Top7, the preferred cluster in seeded REMD (see main text for details). The color code follows Figure S2.

System name	Sequence
CLN025	YYDPETGTWY
Trp-cage	NLYIQWLKDGPPSSGRPPPS
BBA	EQYTAKYKGRTRNEKELRDFIEKFKGR
WW domain Fip35	KLPPGW EK RMSRDGRVYFNH <sup>δ</sup> ITNASQFERPSG
WW domain GTT	GSKLPPGW EK RMSRDGRVYFNH <sup>ε</sup> ITGTTQFERPSG
Villin HP36	MLSDEDFKAVFGMTRSAFANLPLWKQQLKKEKGLF
NTL9 (39 AA)	MKVIFLKDVKGMGKKGEIKNVADGYANNFLFKQGLAIEA
BBL	GSQNNDALSPAIRLLAEWNLDASAIKGTGVGGRLTREDVEKH <sup>δε</sup> LAKA
Protein B	LKNAIEDAIAELKKAGITSDFYFNAINKAKTVEEVNALVNEILKAH <sup>ε</sup> A
Homeodomain	MKQWSENVEEKLKEFVKRH <sup>δ</sup> QRITQEELH <sup>δ</sup> QYAQRLGLNEEAIRQFFEEFEQRK
NTL9 (52 AA)	MKVIFLKDVKGMGKKGEIKNVADGYANNFLFKQGLAIEATPANLKALEAQKQ
NuG2 variant of Protein G	DTYKLVIVLNGTTFYTTTEAVDAATAEKVFKQYANDAGVDGEWYDAATKTFTVTE
CspA (1MJC)	SGKMTGIVKWFNADKGFGITPDDGSKDVFVHPSAIQNDGYKSLDEGQKVSFTIESGAKGPAAGNVTSL
Hyp protein 1WHZ	MWMPPRPEEVARKLRRLGFVERMAKGGHRLYTHPDGRIVVVPFHSGELPKGTFKRILRDAGLTEEEFHN L
α3D	MGSWAEFKQRLAAIKTRLQALGGSEAEAAFEKEIAAFESQLQAYKKGKNPEVEALRKEAAAIRDELQAYR H <sup>δ</sup> N
λ-repressor	PLTQEQLAARRLKAIWEKKKNEGLSYESVADKMGMGQS AVAALFNGINALNAYNAALLAKILKVSVEEFSPSIAREIY
Top7	DIQVQVNIDDNGKNFDYTYTVTTESELQVLNLMYIKKQGAKRVRISITARTKKEAEKFAAILIKVFAE LGYNDINVTFDGDTVTEGQL

**Table S1.** Sequence of peptides and proteins simulated in this work. H<sup>δ</sup>, H<sup>ε</sup>, and H<sup>δε</sup> stand for Histidine that is protonated at N<sup>δ</sup>, N<sup>ε</sup> or both N<sup>δ</sup> and N<sup>ε</sup>, respectively. All His protonation states were used as indicated in the experimental studies.

Protein	PDB ID	# AA	Secondary Structure type	RMSD region (amino acid numbers)	$\mu\text{s}/\text{day}$	MD T, K	MD-native length, $\mu\text{s}$	MD-extended length, $\mu\text{s}$	Lowest RMSD, Å	Largest cluster RMSD, Å	REMD-extended length, $\mu\text{s}$	Lowest RMSD, Å	Largest cluster RMSD, Å	Experimental folding time ( $\mu\text{s}$ )
CLN025	Honda et al. <sup>34</sup>	10	beta	1-10	1.4	300 K	1.2	2.4	0.5	1.0	0.8	0.3	1.0	$\sim 0.1$ (300 K) <sup>53</sup>
Trp-cage	1L2Y <sup>35</sup>	20	alpha	3-18	1.3	300 K	1.7	1.0	0.5	0.6	0.4	0.3	0.7	$\sim 4$ (298 K) <sup>54</sup>
BBA	1FME <sup>36</sup>	28	mixed	4-26	1.4	300 K	5.2	7.8	1.0	1.9	9.1	0.9	4.6	N/A (low stability) <sup>36</sup>
Fip35	Freddolino et al. <sup>37</sup>	33	beta	5-27	1.4	325 K	29.0	25.6	0.5	1.6	3.0	0.4	1.3	13 (337 K) <sup>38</sup>
GTT	2F21 <sup>*55</sup>	35	beta	10-32	1.4	325 K	12.4	21.6	0.6	1.5	10.5	0.5	1.3	$\sim 4$ (353 K) <sup>39</sup>
Villin HP36	1VII <sup>41</sup>	36	alpha	43-72	1.4	290 K	22.1	26.7	1.1	2.4	4.2	1.1	2.3	$\sim 10$ (330–350 K) <sup>56</sup>
NTL9 (39 AA)	2HBA <sup>*57</sup>	39	mixed	1-39	1.4	300 K	47.6	65.8	1.9	6.4	11.9	0.4	0.5	$\sim 700$ (298 K) <sup>43</sup>
BBL	2WXC <sup>58</sup>	47	mixed	133-151,159-170	1.2	290 K	14.1	17.1	3.2	8.5	2.2	2.1	8.3	$\sim 14$ (283 K) <sup>44</sup>
Protein B	1PRB <sup>*59</sup>	47	alpha	8-50	1.0	300 K	4.6	10.3	1.6	4.2	1.9	1.6	3.3 (4.2)	$\sim 1$ (298 K) <sup>60</sup>

Homeodomain	2P6J <sup>45</sup>	52	alpha	5-48	1.0	300 K	7.2	17.3	1.9	3.0	3.5	1.6	3.2	~ 13 (308 K) <sup>61</sup>
NTL9 (52 AA)	2HBA <sup>57</sup>	52	mixed	1-52	1.0	300 K	11.8	10.2	6.1	11.4	21.2	1.6	6.0	~1400 (298 K) <sup>43</sup>
NuG2 variant	1MI0 <sup>*46</sup>	56	mixed	6-61	1.0	300 K	51.3	54.7	7.5	9.6	28.8	4.8	11.4	~ 60 (298 K) <sup>62</sup>
CspA	1MJC <sup>47</sup>	69	beta	4-14,16-23,29-36,48-56,62-70	0.8	300 K	2.7	6.9	8.7	10.1	29.4	2.5	9.9	~5000 (298 K) <sup>63</sup>
Hypothetical protein from <i>Thermus thermophilus</i> 1WHZ	1WHZ <sup>48</sup>	70	mixed	6-70	0.8	300 K	14.3	22.5	5.9	9.7	9.0	1.9	11.8	not available
$\alpha$ 3D	2A3D <sup>49</sup>	73	alpha	1-73	0.8	300 K	6.6	20.5	2.5	3.7	1.2	2.9	4.0	> 3.2 (323 K) <sup>64</sup>
$\lambda$ -repressor	1LMB <sup>*50</sup>	80	alpha	Chain3 6-85	0.7	300 K	26.6	39.3	4.4	10.5	24.0	2.9	11.9	~ 10 (350 K) <sup>65</sup>
Top7	1QYS <sup>51</sup>	92	mixed	3-94	0.6	300 K	8.0	5.4	12.0	14.7	18.2	2.6	11.2	> 10 <sup>5</sup> (295 K) <sup>66</sup>

**Table S2. System details.** The protein name, PDB ID, number of amino acids of simulated system, overall topology, residues in RMSD mask, MD speed ( $\mu$ s/day), MD temperature, native MD length ( $\mu$ s), extended MD length ( $\mu$ s), lowest RMSD in extended MD ( $\text{\AA}$ ), RMSD of extended MD largest cluster centroid ( $\text{\AA}$ ), extended REMD simulation time ( $\mu$ s), lowest RMSD in extended REMD ( $\text{\AA}$ ), RMSD of extended REMD largest cluster centroid ( $\text{\AA}$ ), experimental folding time. Asterisks following PDB IDs indicate differences between the system in the crystal and in the simulation. RMSD region is listed as amino acid residue IDs in PDB from RCSB<sup>67</sup>.



System	Extended REMD temperatures (K)
CLN025	275.1 · 300.0 · 327.2 · 356.8 · 389.1 · 424.3
Trp-cage	264.0 · 281.4 · 300.0 · 319.8 · 340.9 · 363.3 · 387.3 · 412.9
BBA	243.8 · 256.8 · 270.4 · 284.8 · 300.0 · 316.0
Fip35	285.4 · 300.0 · 315.4 · 331.5 · 348.5 · 366.3 · 385.0 · 404.7 · 425.5 · 447.2 · 470.1 · 494.2 · 519.5 · 546.0 · 574.0 · 603.4
GTT	250.0 · 262.6 · 275.8 · 289.7 · 304.3 · 319.6 · 335.6 · 352.5 · 370.3 · 388.9
Villin HP36	250.0 · 262.2 · 275.0 · 288.4 · 300.0 · 317.3 · 332.8 · 349.0
NTL9 (39)	250.0 · 261.9 · 274.5 · 287.6 · 301.3 · 315.7 · 330.8 · 346.6 · 363.1 · 380.5 · 398.7 · 417.7
BBL	274.9 · 287.2 · 300.0 · 313.4 · 327.4 · 342.0 · 357.2 · 373.2 · 389.8 · 407.2 · 425.3 · 444.3 · 464.1 · 484.8 · 506.5 · 529.1
Protein B	290.0 · 300.0 · 316.0 · 329.8 · 344.2 · 359.3 · 375.0 · 391.5 · 408.6 · 426.5 · 445.2 · 464.7 · 485.0 · 506.3
Homeodomain	288.7 · 300.0 · 311.7 · 323.9 · 336.6 · 349.8 · 363.5 · 377.7 · 392.4 · 407.8 · 423.8 · 440.3 · 457.6 · 475.5 · 494.1 · 513.4
NTL9 (52)	280.0 · 291.6 · 300.0 · 316.2 · 329.2 · 342.9 · 357.0 · 371.8 · 387.2 · 403.2
NuG2 variant	280.0 · 291.4 · 303.3 · 315.7 · 328.6 · 342.0 · 355.9 · 370.5 · 385.6 · 401.3
CspA	290.0 · 300.0 · 311.8 · 323.3 · 335.3 · 347.7 · 360.5 · 373.8 · 387.6 · 401.9
Hypothetical protein 1WHZ	280.0 · 289.6 · 300.0 · 309.9 · 320.5 · 331.6 · 343.0 · 354.7 · 366.9 · 379.6 · 392.6 · 406.1
$\alpha$ 3D	289.9 · 300.0 · 310.4 · 321.2 · 332.3 · 343.9 · 355.8 · 368.1 · 380.9 · 394.1 · 407.8 · 422.0 · 436.6 · 451.7 · 467.4 · 483.6 · 500.4 · 517.8 · 535.8 · 554.3 · 573.6 · 593.5 · 614.1 · 635.4
$\lambda$ -repressor	290.4 · 300.0 · 309.9 · 320.1 · 330.7 · 341.6 · 352.9 · 364.5 · 376.6 · 389.0
Top7	280.0 · 288.5 · 300.0 · 306.3 · 315.7 · 325.3 · 335.1 · 345.3 · 355.8 · 366.7 · 377.8 · 389.3

**Table S3. Temperatures used for extended REMD simulations**

System	Seeded REMD temperatures (K)
NuG2 variant (1)	250.0 · 260.2 · 270.8 · 281.9 · 293.4 · 305.3 · 317.8 · 330.8 · 344.3 · 358.3 · 372.9 · 388.2
NuG2 variant (2)	280.0 · 291.4 · 303.3 · 315.7 · 328.6 · 342.0 · 355.9 · 370.5 · 385.6 · 401.3 · 309.0 · 334.0
CspA	250.0 · 259.2 · 268.8 · 278.7 · 289.0 · 299.7 · 310.8 · 322.2 · 334.1 · 346.5 · 359.3 · 372.6
Hypothetical protein 1WHZ	242.0 · 250.0 · 258.6 · 267.5 · 271.0 · 276.7 · 286.2 · 296.0 · 306.2 · 311.0 · 316.7 · 327.6 · 338.9 · 350.5 · 356.0 · 362.6 · 375.1 · 388.0 · 401.3 · 410.0
$\lambda$ -repressor (1)	250.0 · 260.0 · 270.0 · 280.0 · 290.4 · 300.0 · 309.9 · 320.1 · 330.7 · 341.6 · 352.9 · 364.5 · 376.6 · 389.0
$\lambda$ -repressor (2)	258.3 · 266.8 · 275.6 · 284.7 · 294.1 · 303.8 · 313.8 · 324.2 · 334.9 · 345.9 · 357.3 · 369.1
Top7	240.0 · 247.3 · 254.8 · 262.6 · 270.6 · 278.8 · 287.3 · 296.0 · 305.0 · 314.3 · 323.8 · 333.7 · 343.8 · 354.3 · 365.1 · 376.2 · 387.6 · 399.4

**Table S4. Temperatures used for seeded REMD simulations**

### System data

CLN025

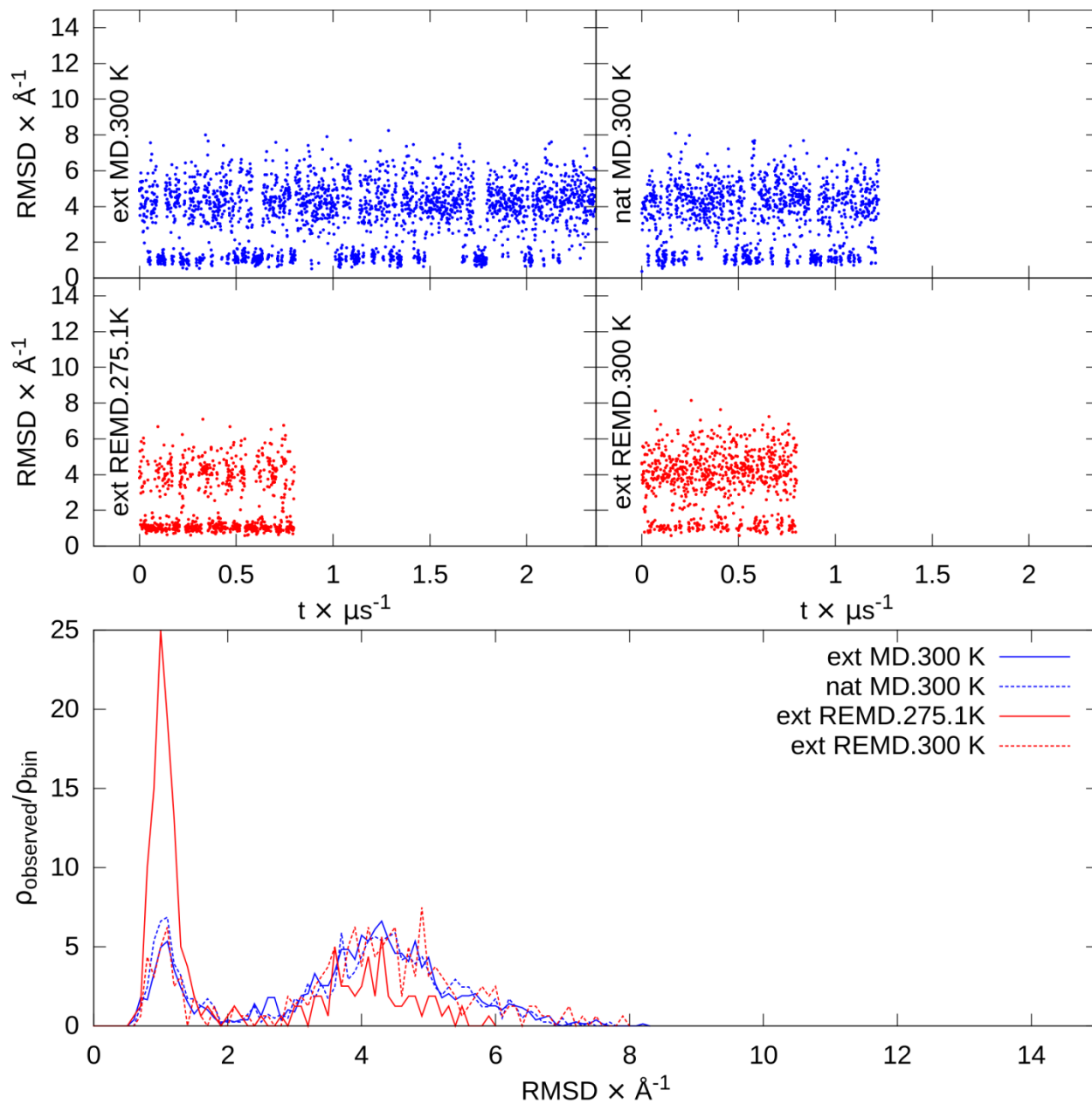


Figure S4. CLN025 RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

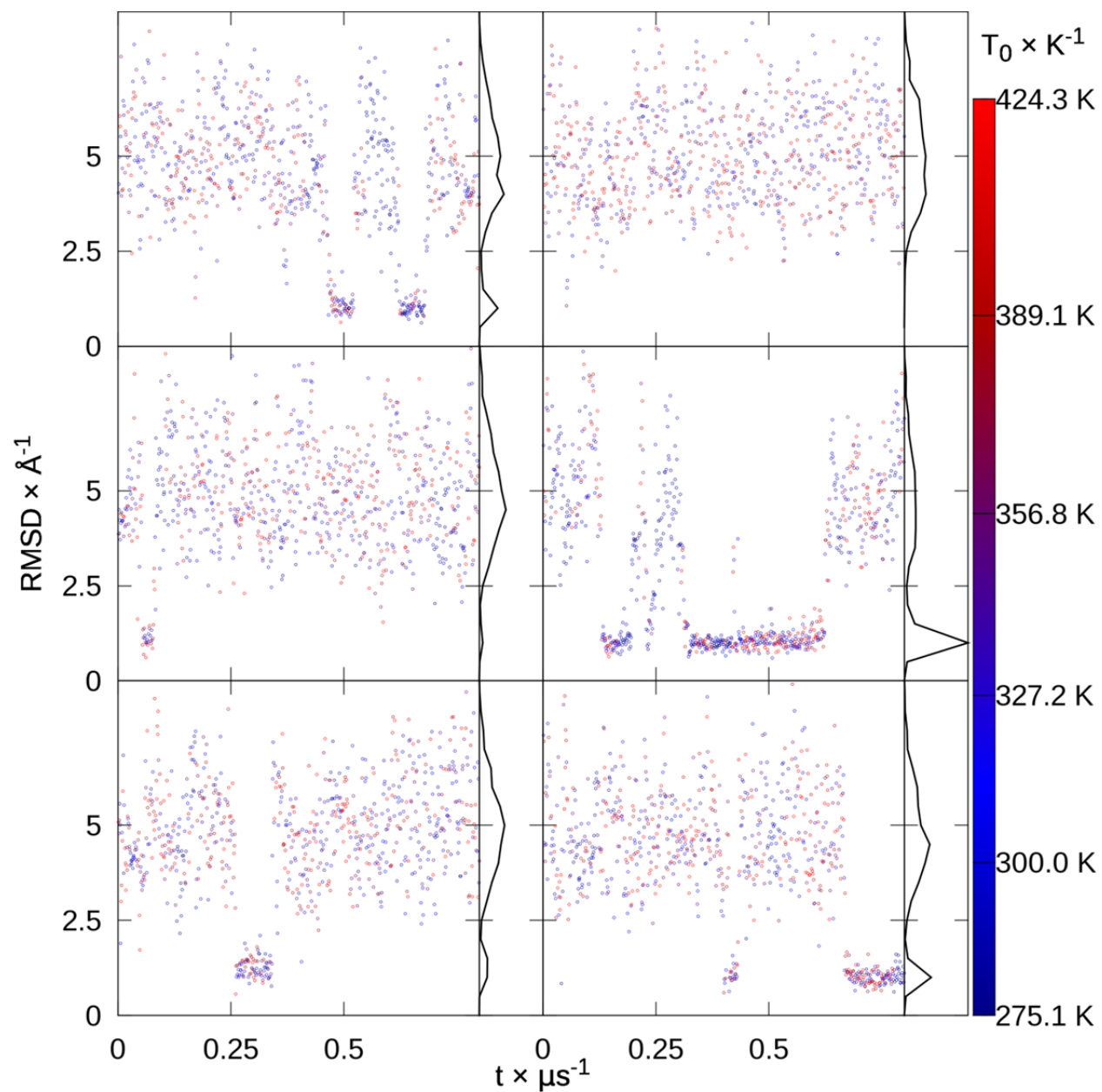


Figure S5. CLN025 replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms shown on the right.

Cluster population	57.6	7.8	2.9	2.0	1.9
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	1.0	4.2	3.7	5.1	3.4

Table S5. CLN025 top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

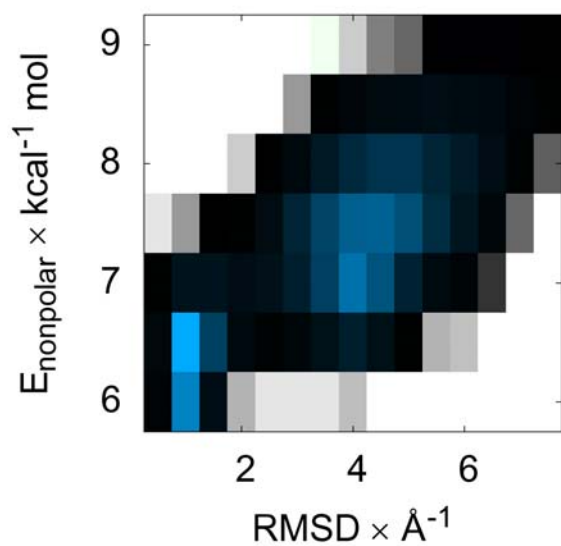


Figure S6. CLN025 surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is modestly more favorable at low (around  $1 \text{ \AA}$ ) than medium (around  $4 \text{ \AA}$ ) RMSDs.

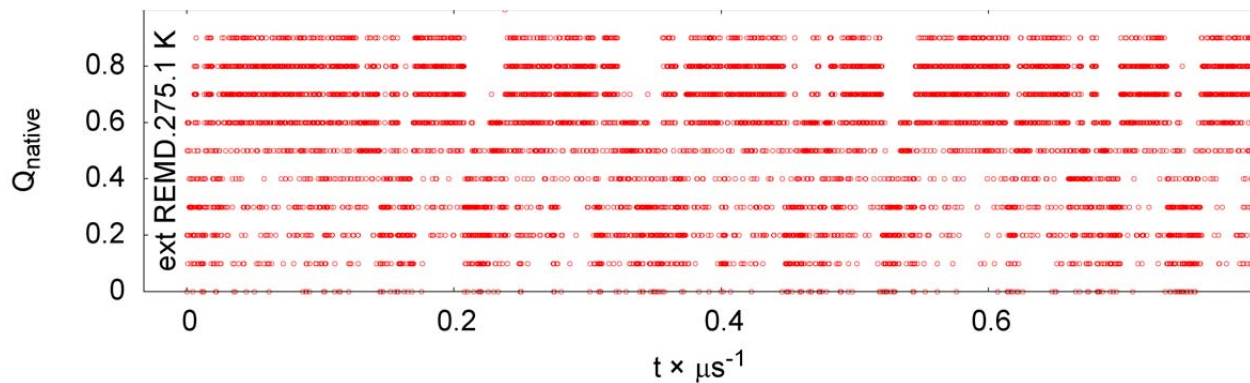


Figure S7. CLN025 native contacts versus time.

Trp-cage tc5b

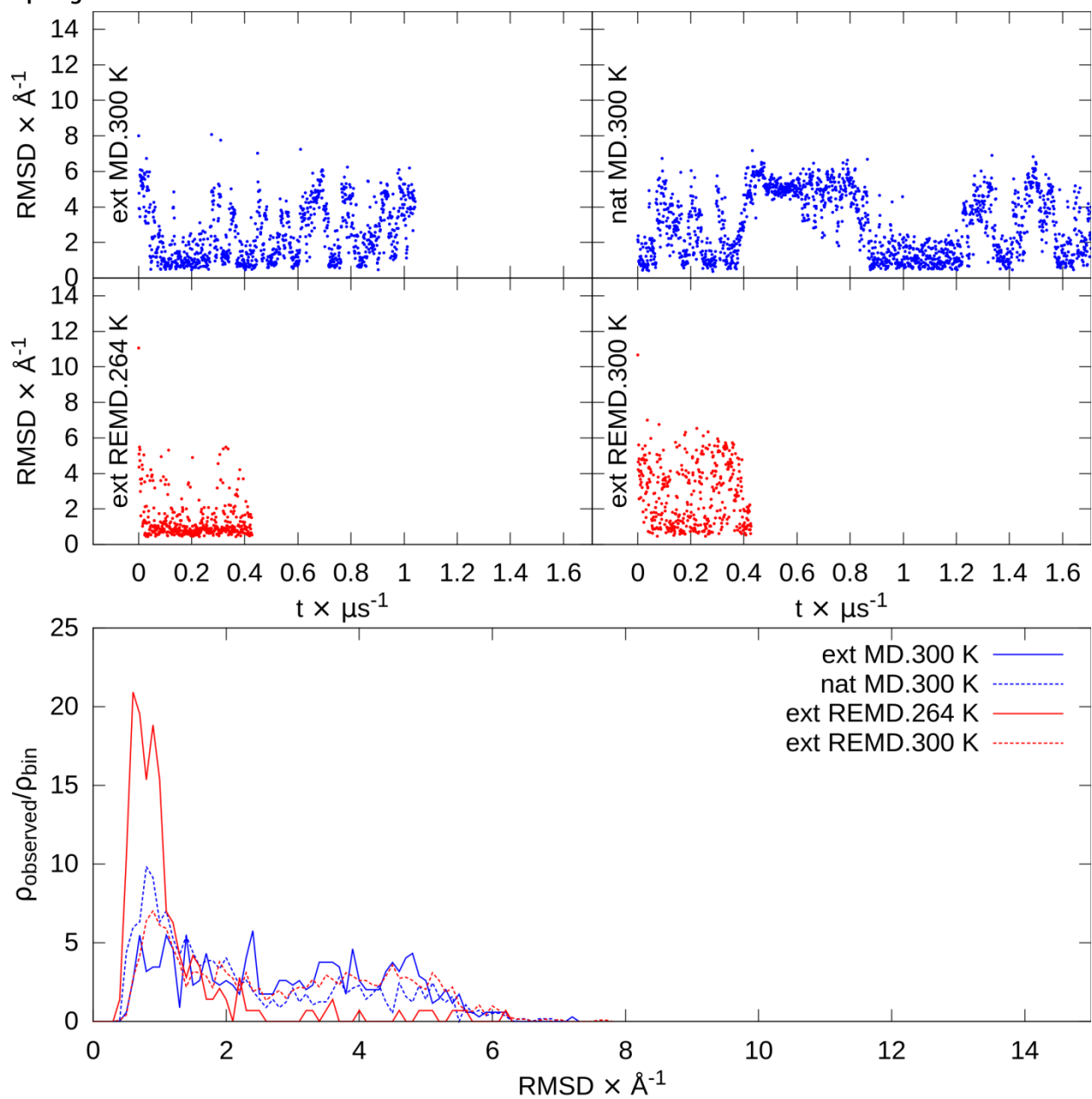


Figure S8. Trp-cage tc5b RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

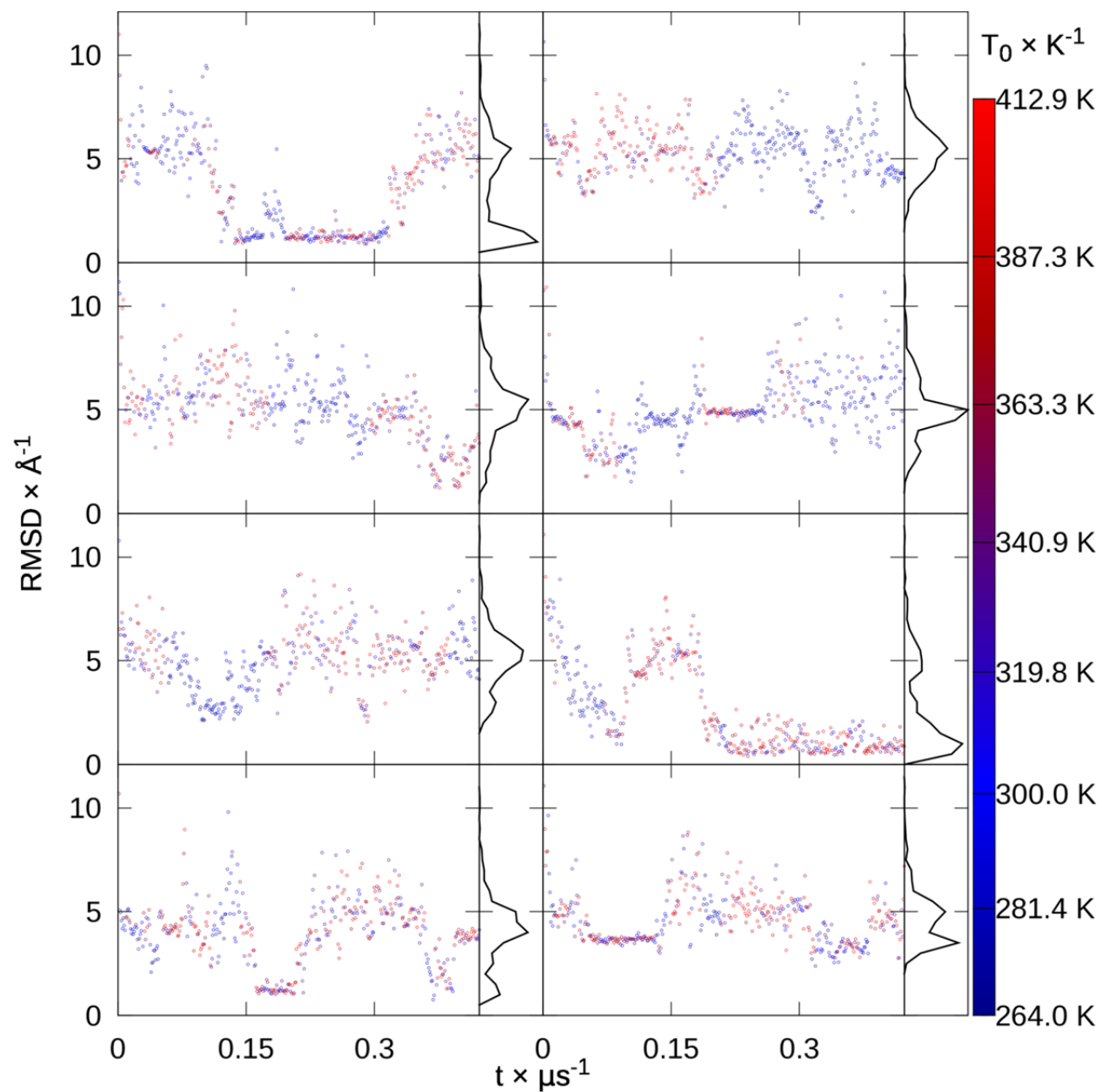


Figure S9. Trp-cage replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	28.7	25.0	14.9	13.8	8.1
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	0.7	0.5	0.7	1.6	0.9

Table S6. Trp-cage top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

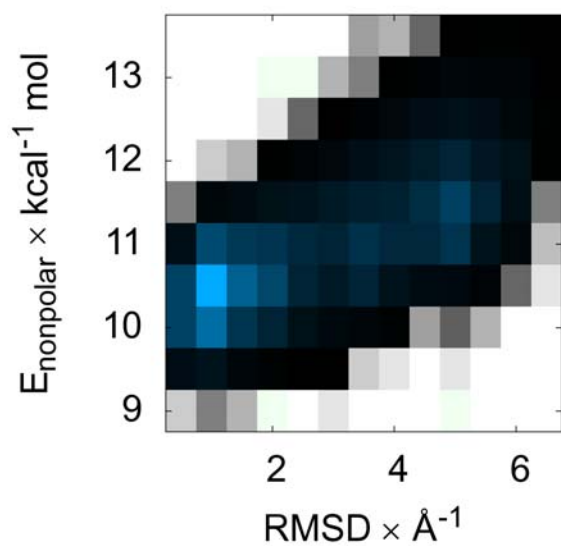


Figure S10. Trp-cage surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly to slightly more favorable at low (around  $1 \text{ \AA}$ ) than medium ( $3\text{-}5 \text{ \AA}$ ) RMSDs.

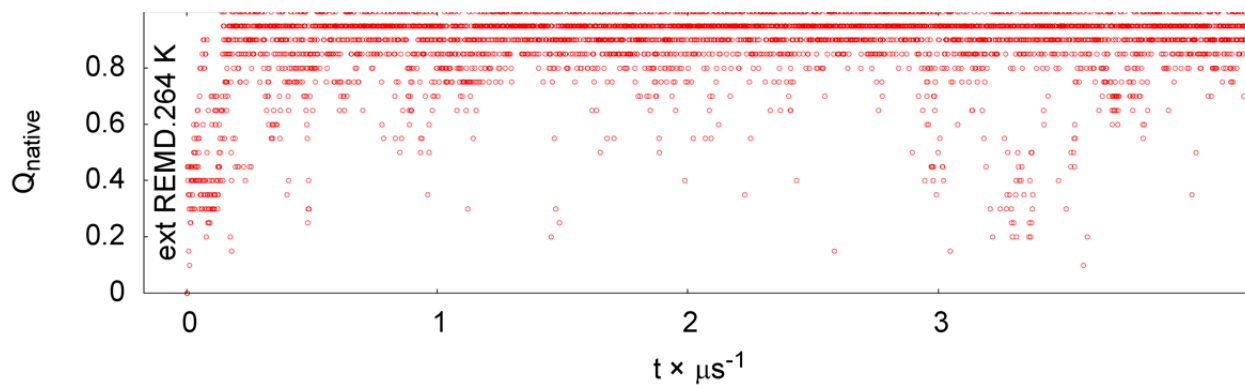


Figure S11. Trp-cage native contacts versus time.

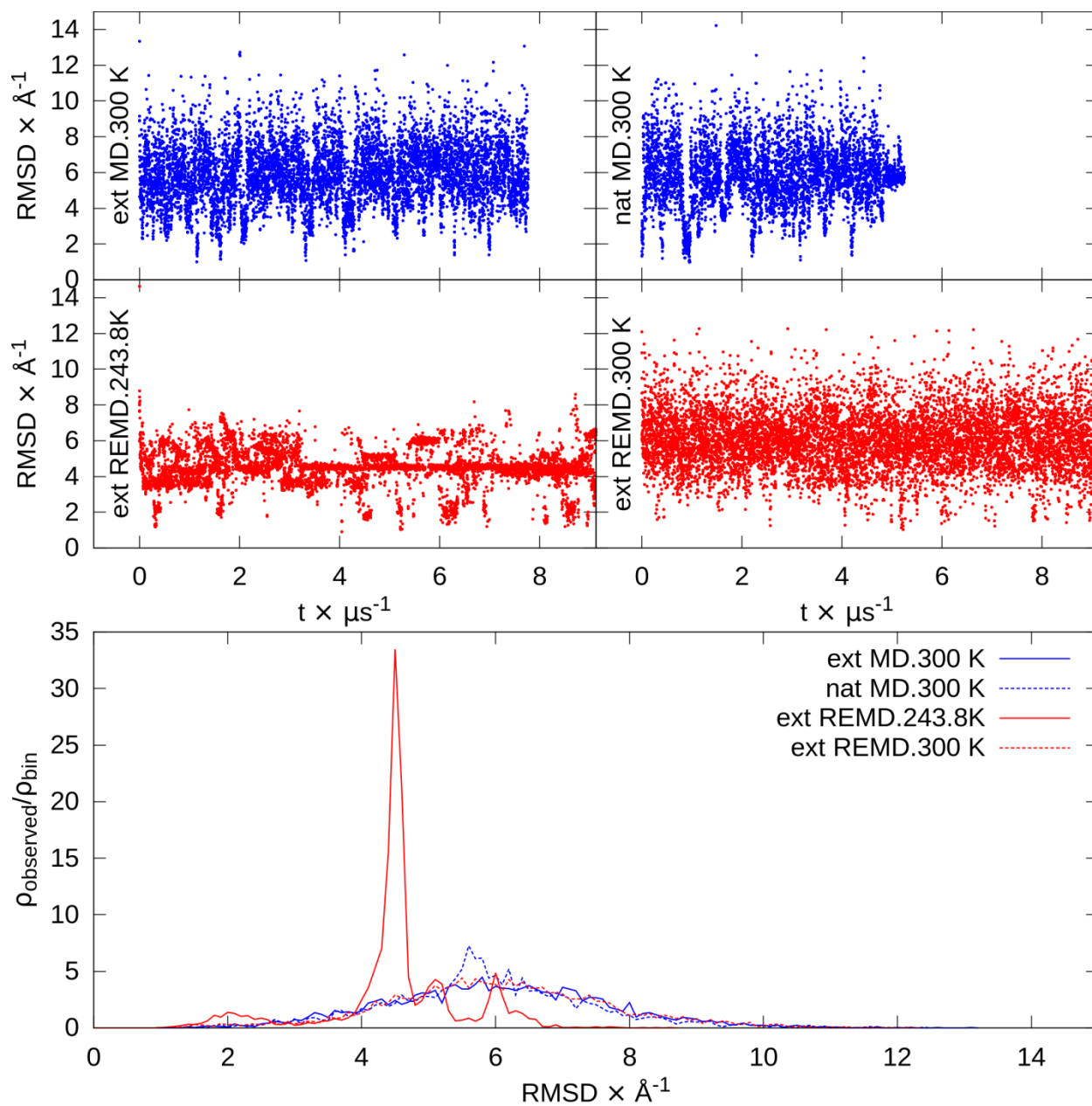
**BBA**

Figure S12. BBA RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).



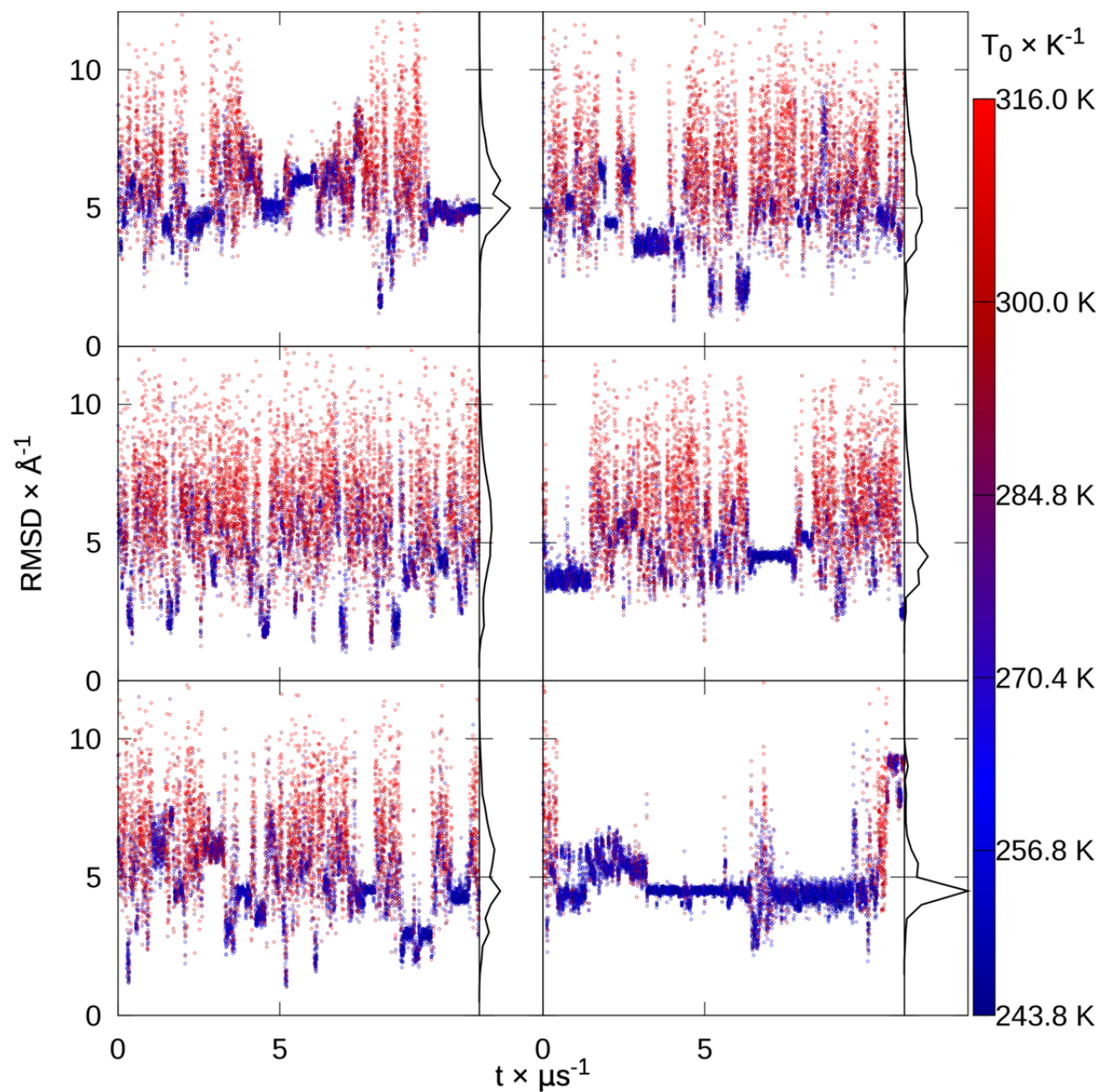


Figure S13. BBA replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	34.8	11.0	7.4	4.8	4.3
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	4.6	3.4	4.1	4.4	5.9

Table S7. BBA top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

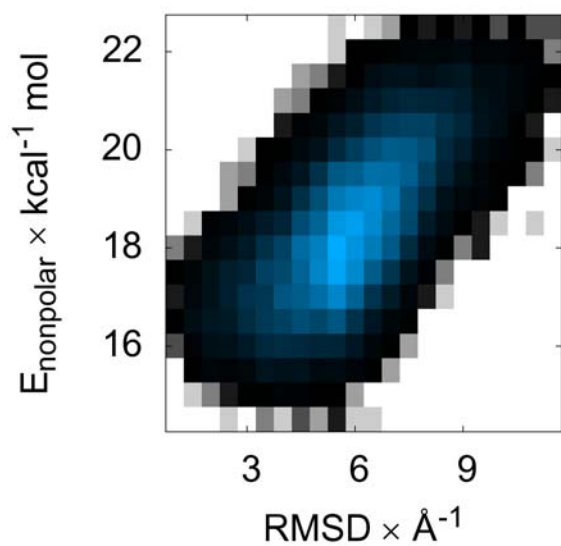


Figure S14. BBA surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly favorable at low ( $2 \text{ \AA}$ ) to mid ( $6 \text{ \AA}$ ) RMSDs, with no strong bias favoring low RMSD structures.

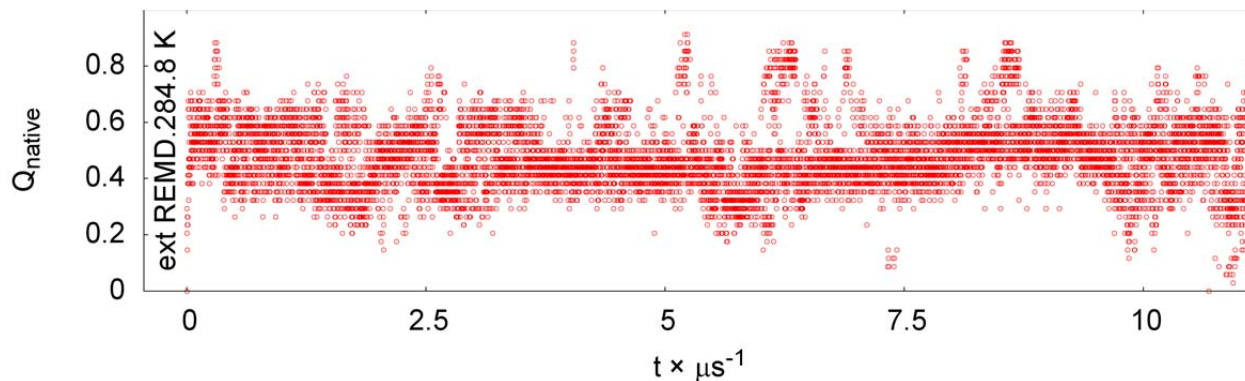


Figure S15. BBA native contacts versus time.

Fip35

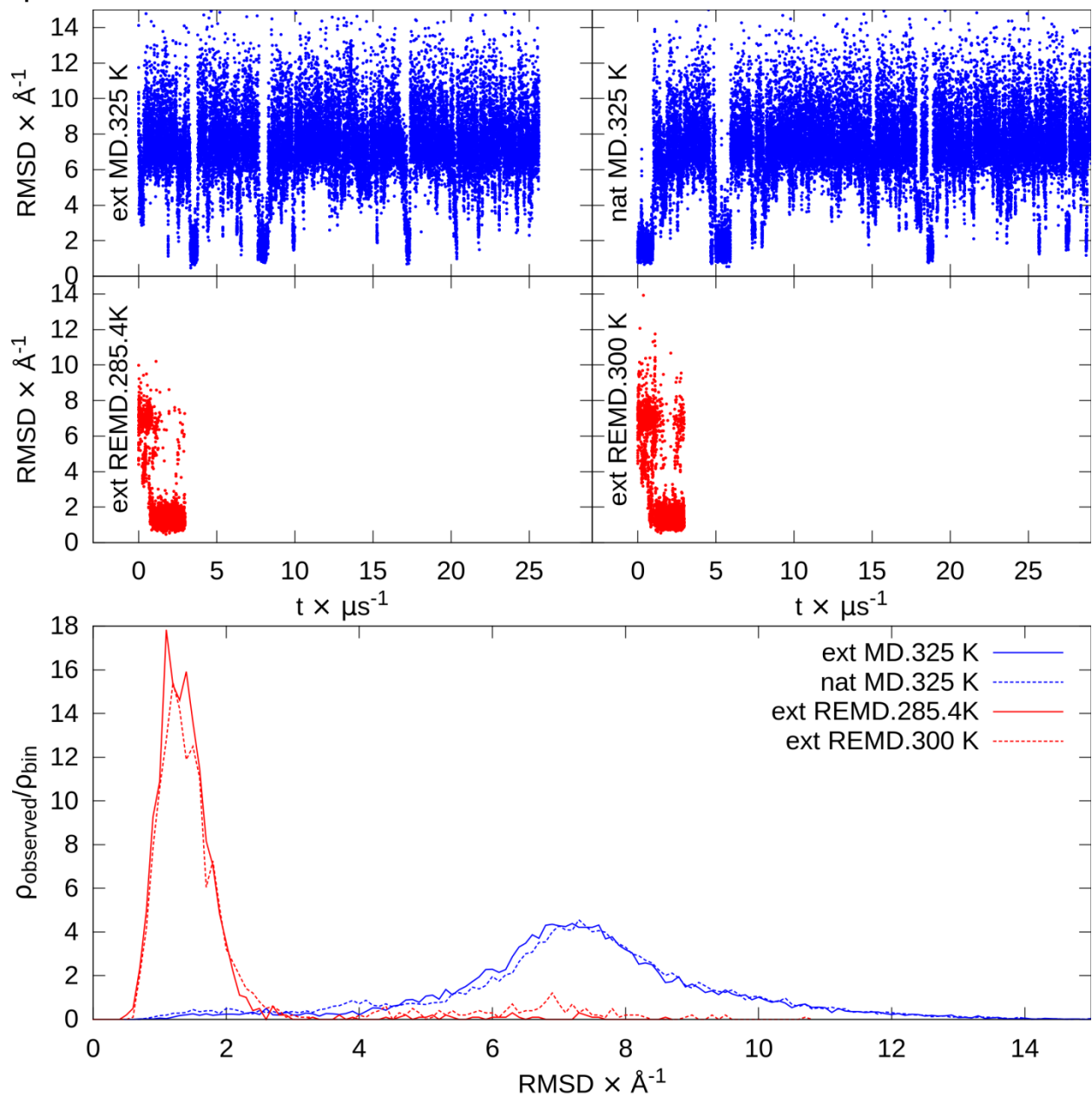


Figure S16. Fip35 RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

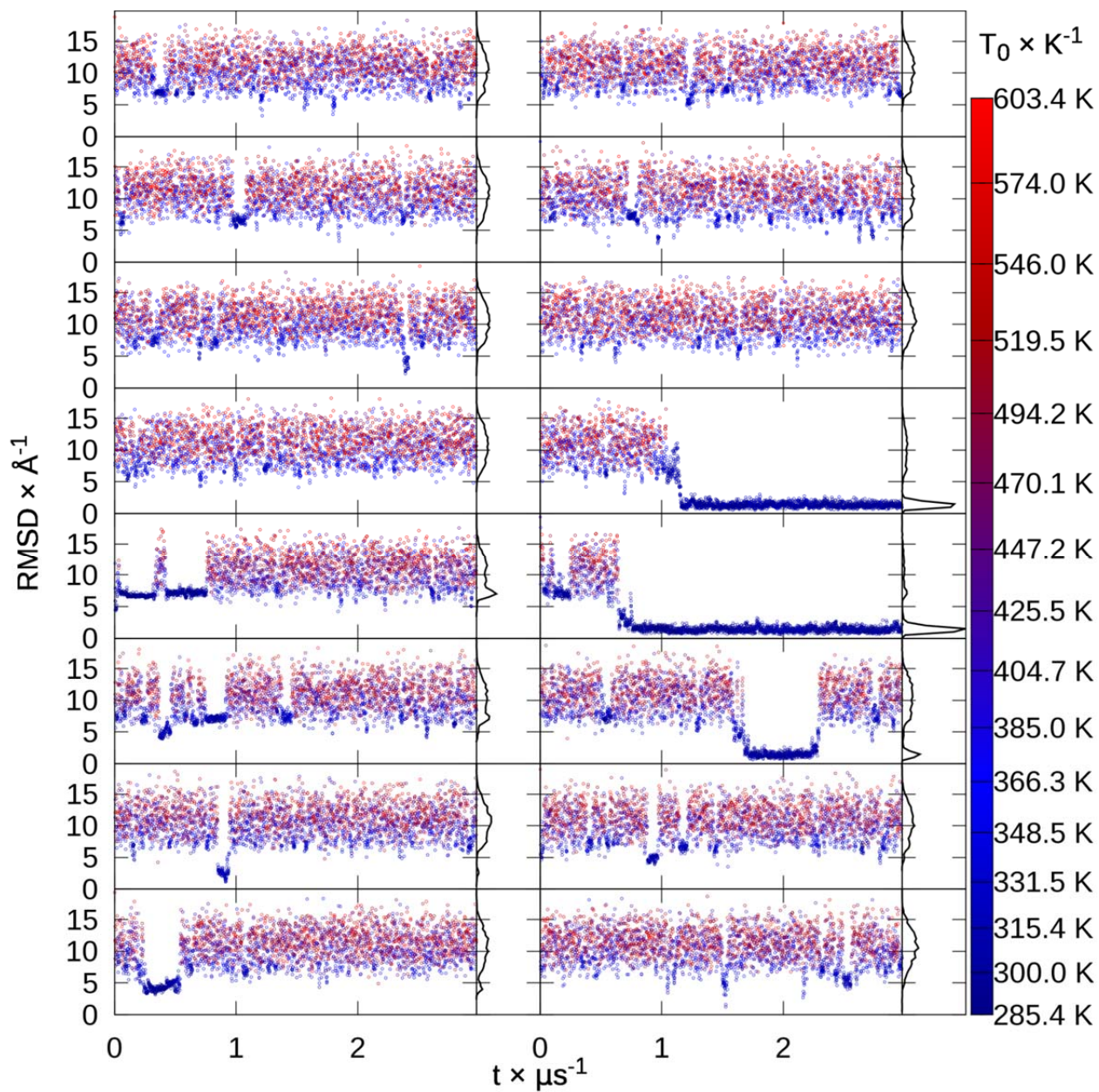


Figure S17. Fip35 replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	70.7	7.5	4.6	4.3	2.4
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	1.3	7.1	4.2	6.6	7.1

Table S8. Fip35 top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

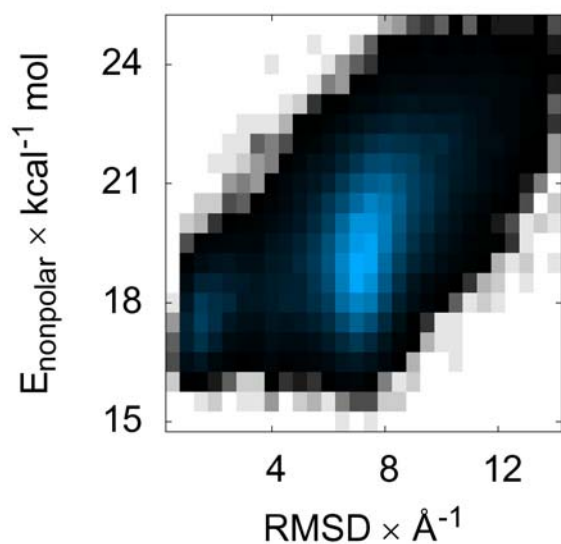


Figure S18. Fip35 surface area energy versus RMSD. Color indicates the histogrammed population in each 0.5 Å by 0.5 kcal mol<sup>-1</sup> bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly favorable from low (1 Å) to medium (8 Å) RMSDs, indicating no strong driving force toward low RMSD values.

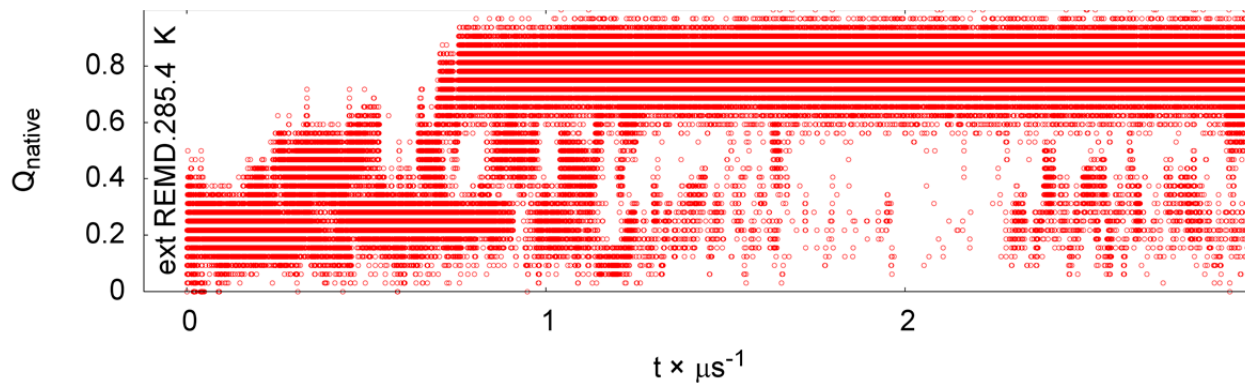


Figure S19. Fip35 native contacts versus time.

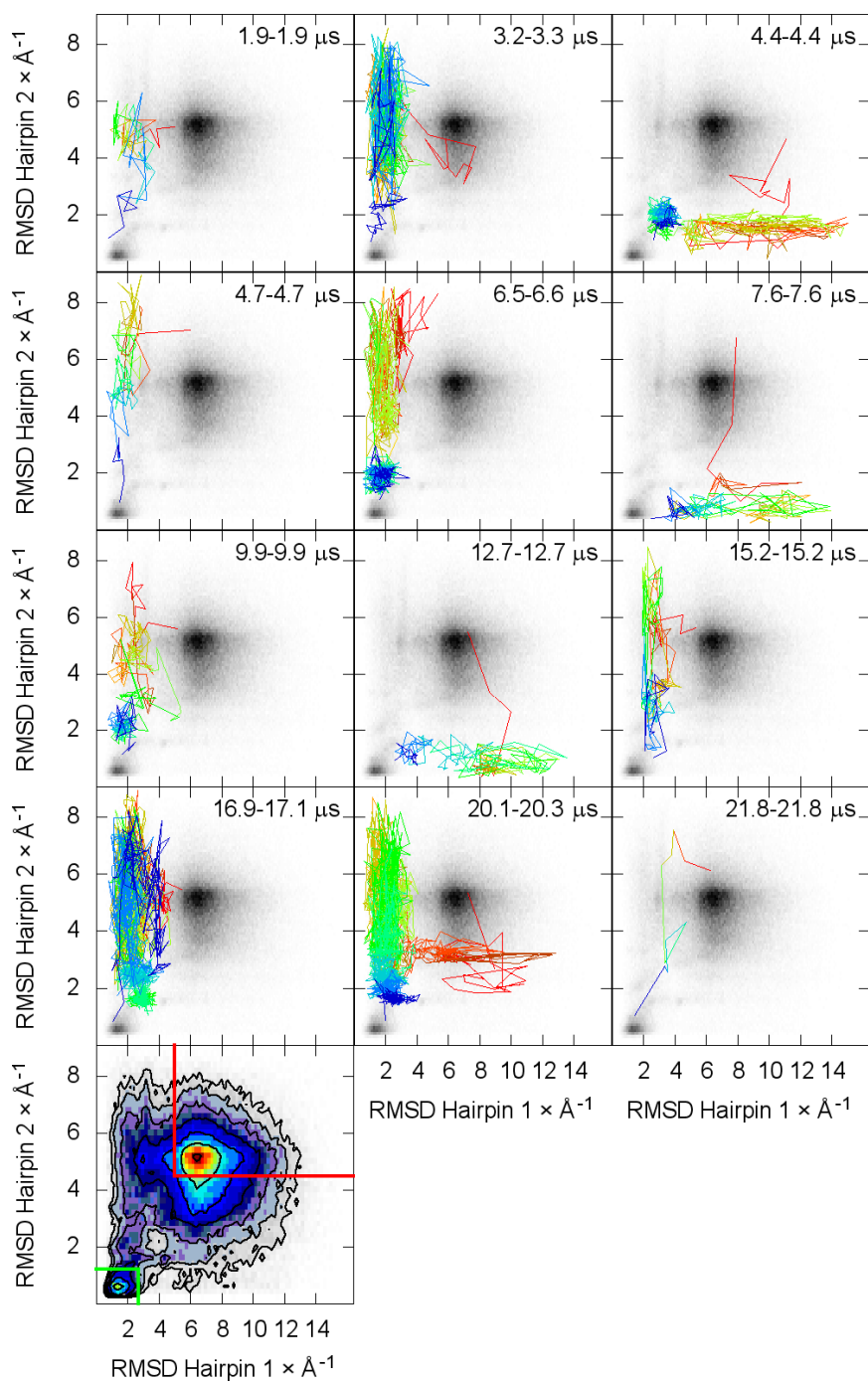


Figure S20. Fip35 folding pathways and population histogram. The twelve unique folding pathways from fully unfolded to fully folded, as defined in Methods, are colored from red to yellow to green to cyan to blue. Eight proceed by folding of hairpin 1 first, two by folding of hairpin 2 first, and two by both simultaneously. At bottom, histogram with contours defining exponents of 2 shows two states in hairpin 1-hairpin 2 RMSD space, with unfolded boxed in red and folded in green. Trajectories are from the extended MD run shown in the top left corner of Figure S16.

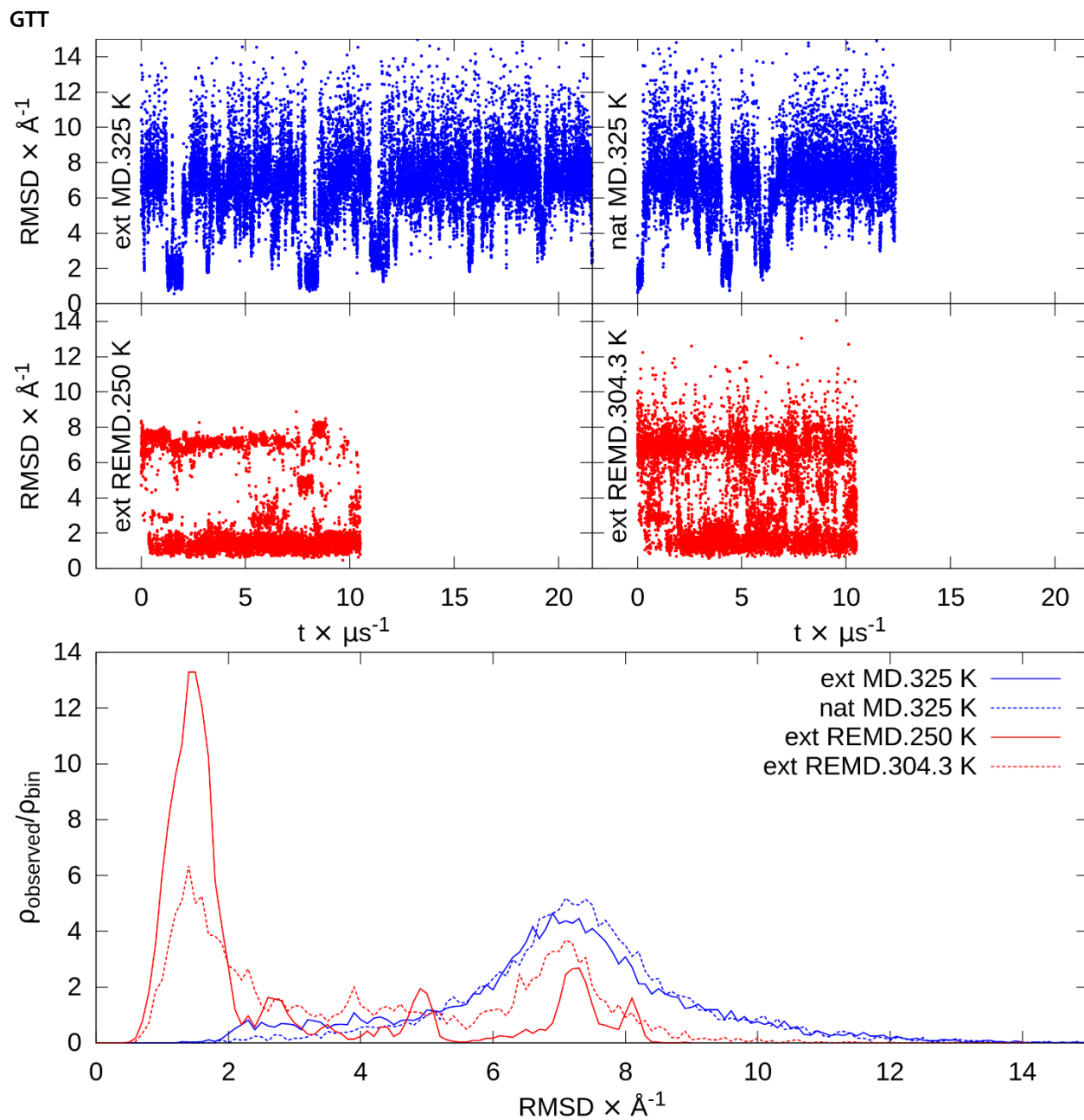


Figure S21. GTT RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

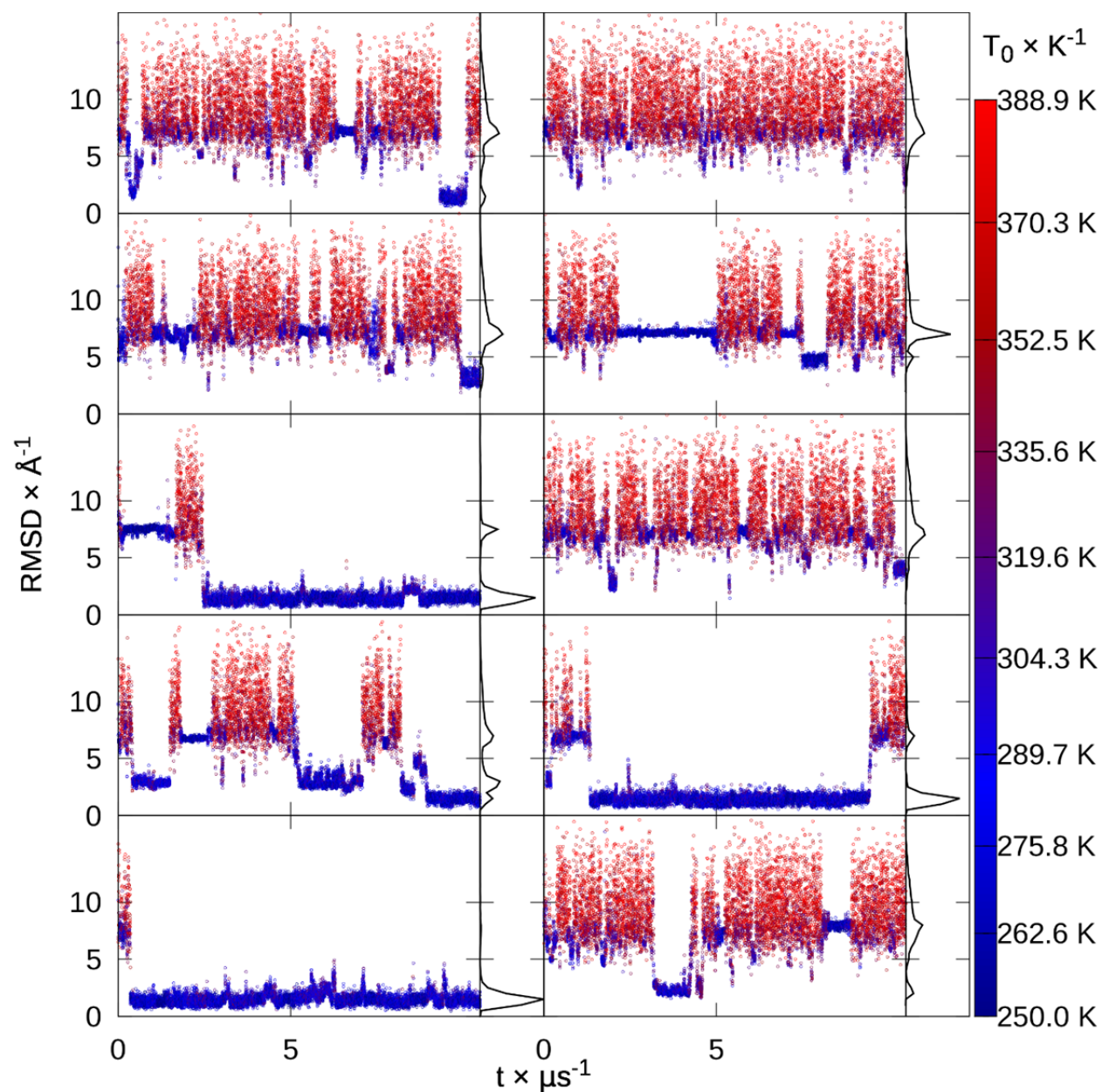


Figure S22. GTT replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	53.9	10.8	8.0	4.8	4.3
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	1.3	7.3	7.5	7.1	6.7

Table S9. GTT top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.



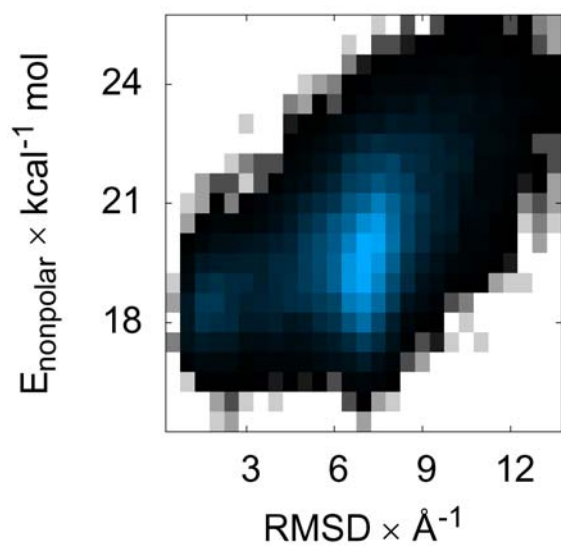


Figure S23. GTT surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is flat from mid ( $6\text{--}8 \text{ \AA}$ ) to low ( $2\text{--}3 \text{ \AA}$ ) RMSD.

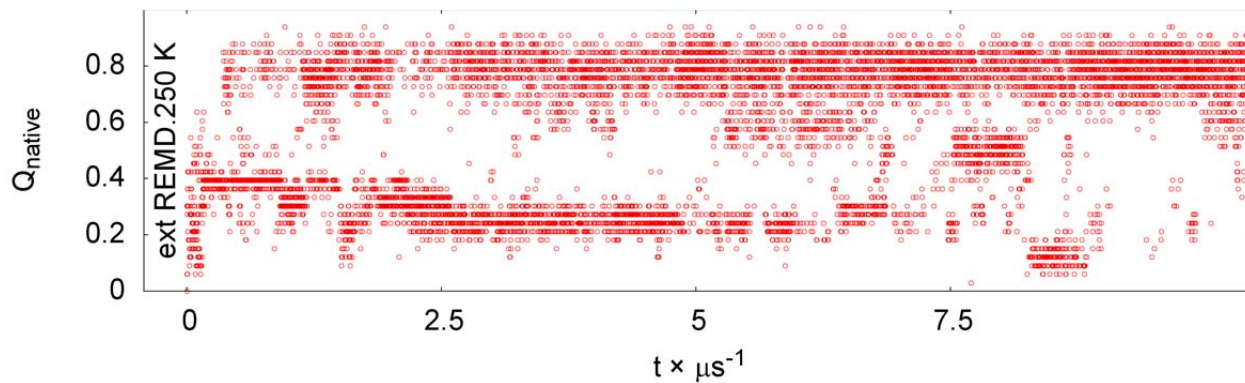


Figure S24. GTT native contacts versus time.

Villin HP36

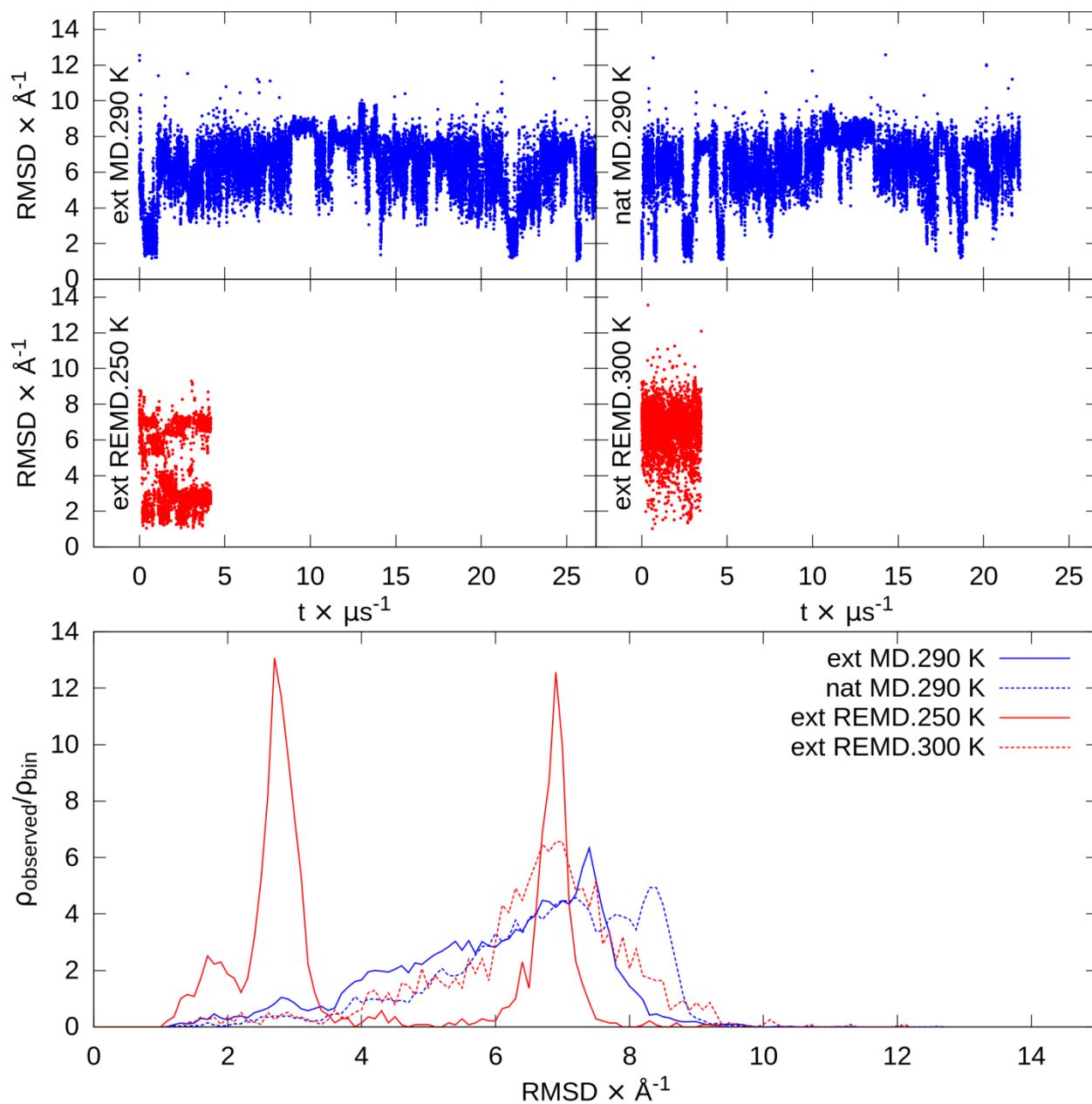


Figure S25. HP36 RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

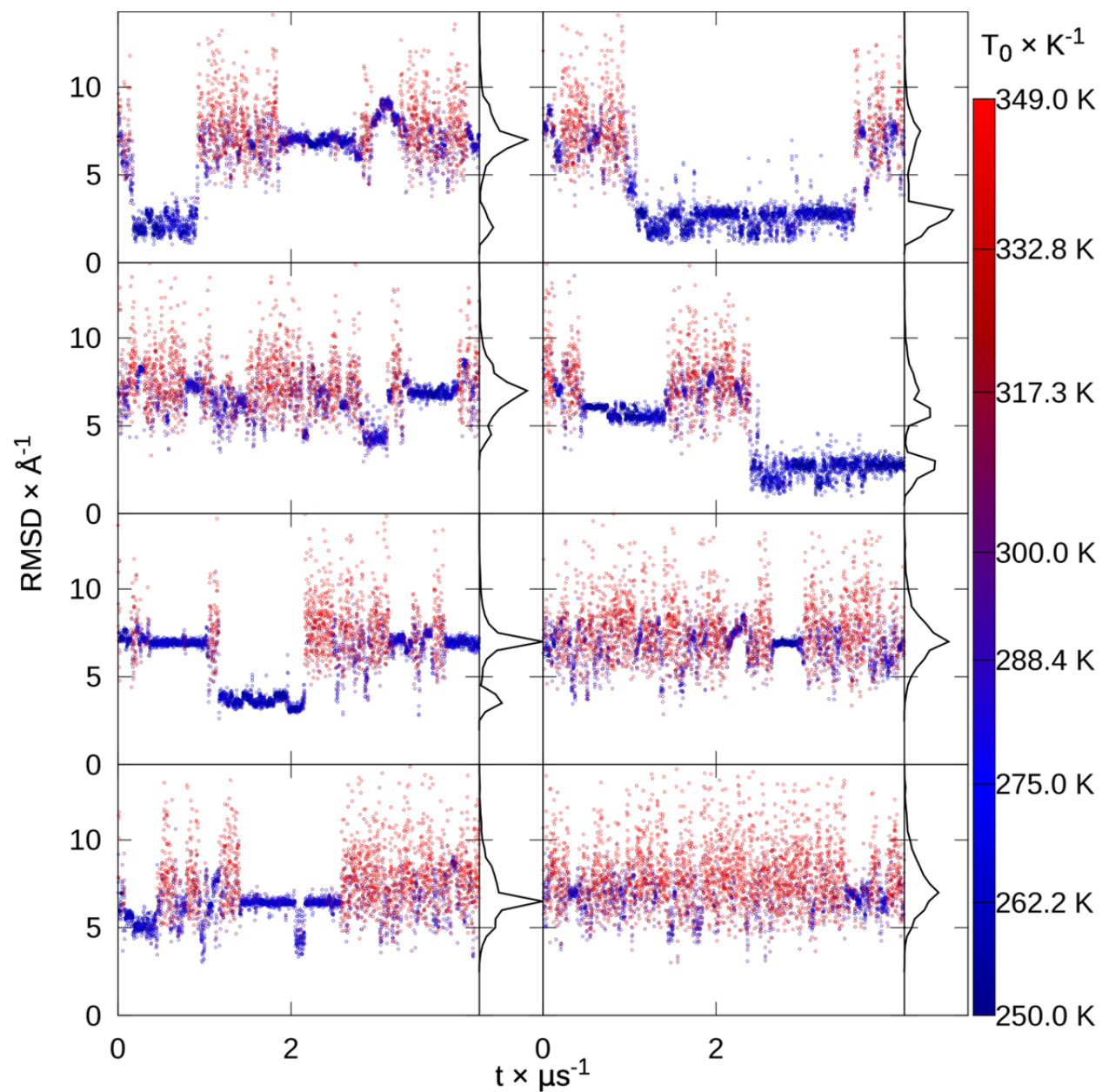


Figure S26. Villin HP36 replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	43.3	10.8	10.3	10.2	4.5
Centroid $C_\alpha$ RMSD (Å)	2.3	5.5	3.5	6.9	6.9

Table S10. Villin HP36 top 5 extended REMD cluster populations and centroid  $C_\alpha$  RMSDs.

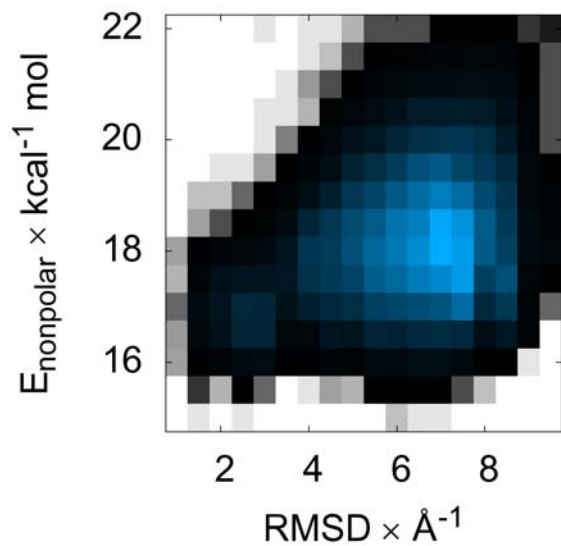


Figure S27. Villin HP36 surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is flat from mid ( $6\text{--}8 \text{ \AA}$ ) to low ( $1\text{--}3 \text{ \AA}$ ) RMSD.

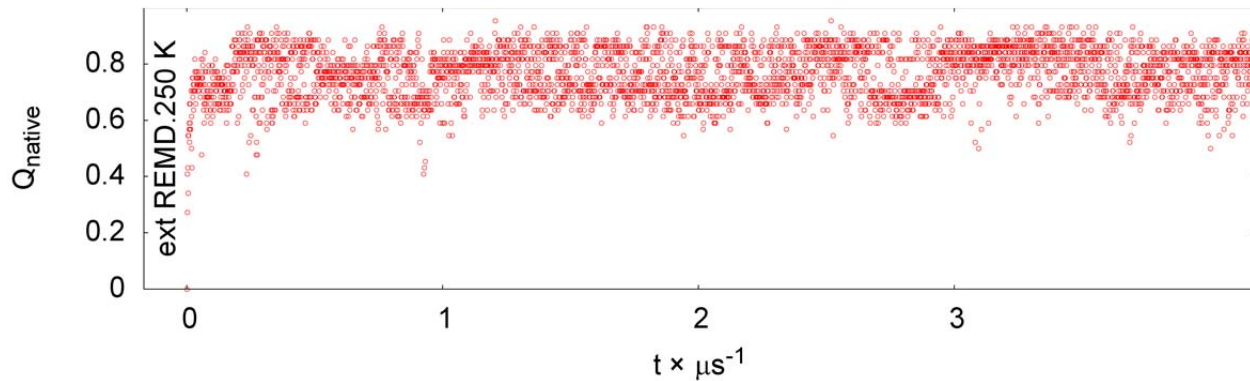


Figure S28. Villin HP36 native contacts versus time.

NTL9 (39 AA)

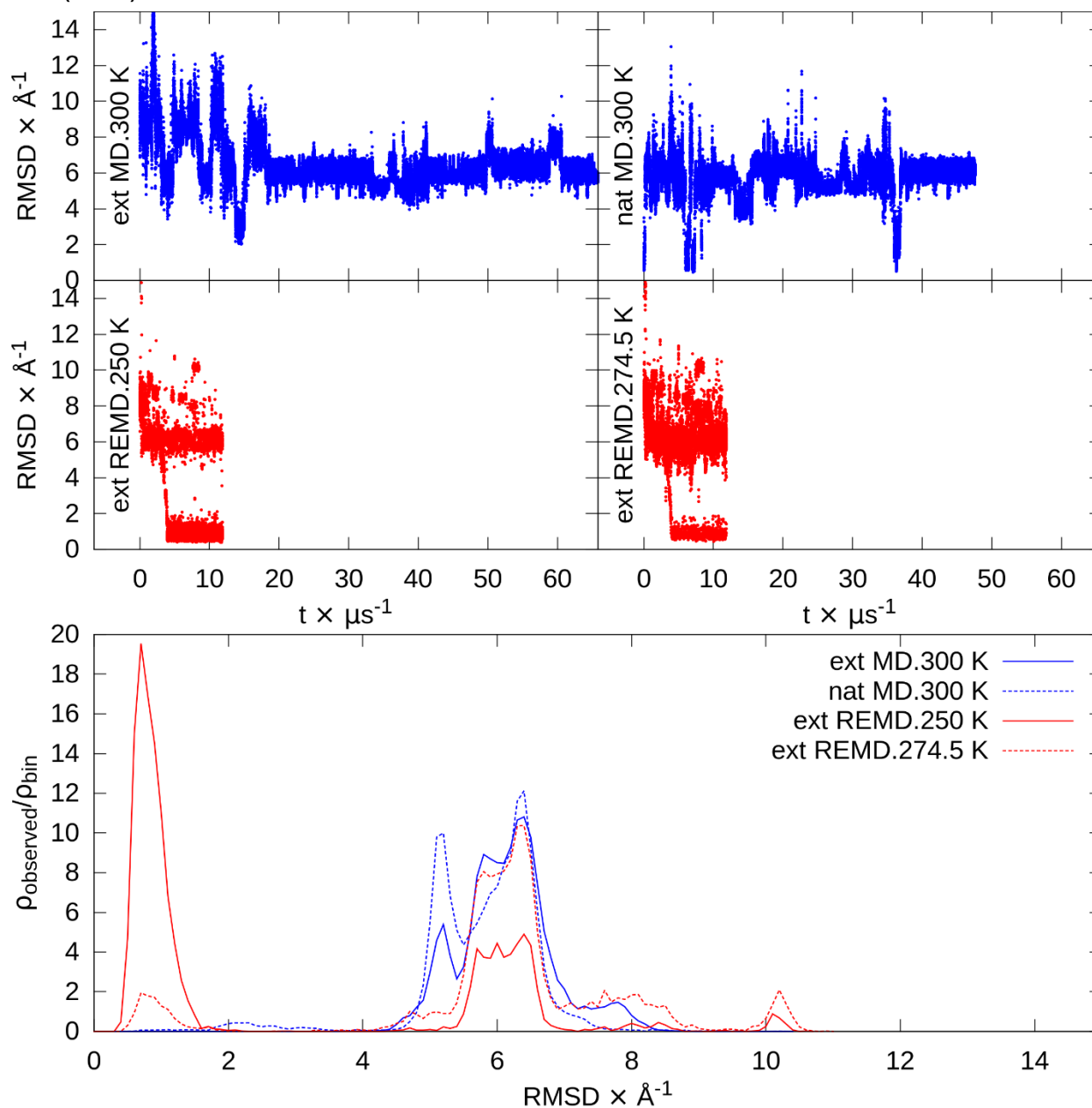


Figure S29. NTL (39 AA) RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

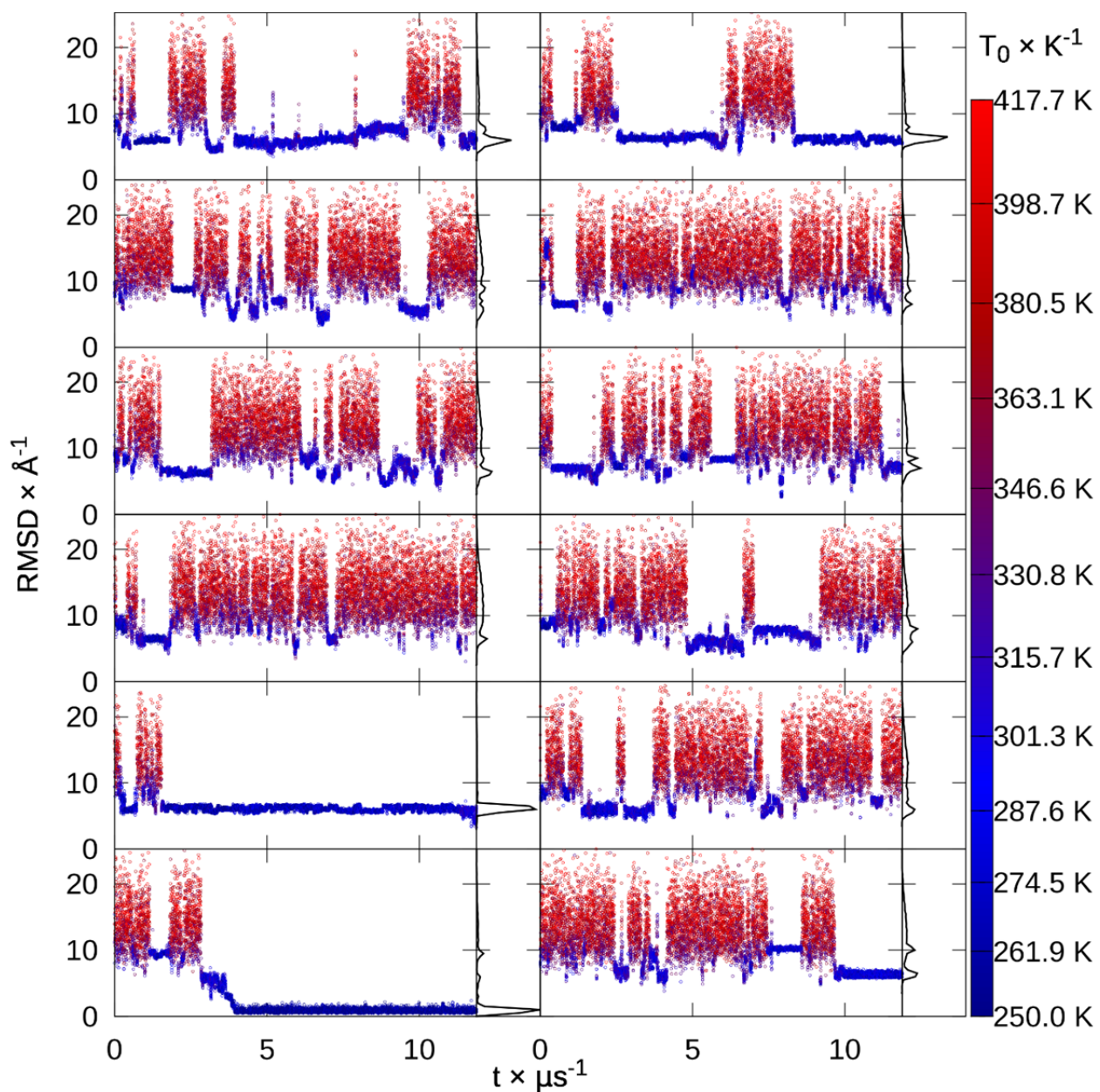


Figure S30. NTL9 (39 AA) replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	68.1	8.9	6.8	5.5	2.2
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	6.1	5.9	6.0	4.6	5.4

Table S11. NTL9 (39 AA) top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

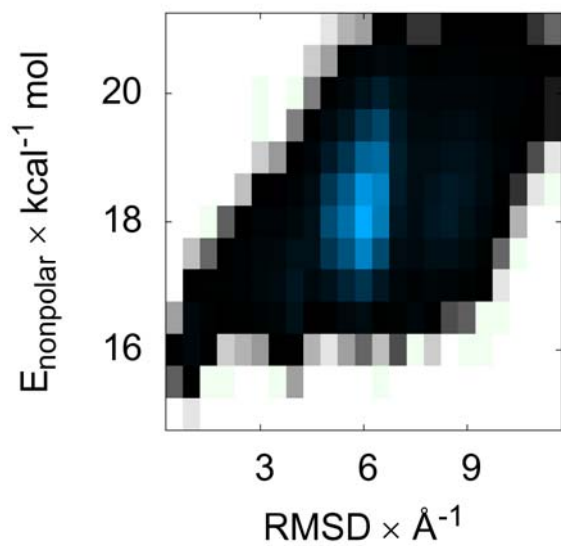


Figure S31. NTL9 (39 AA) surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is more favorable at low ( $\sim 1 \text{ \AA}$ ) RMSD.

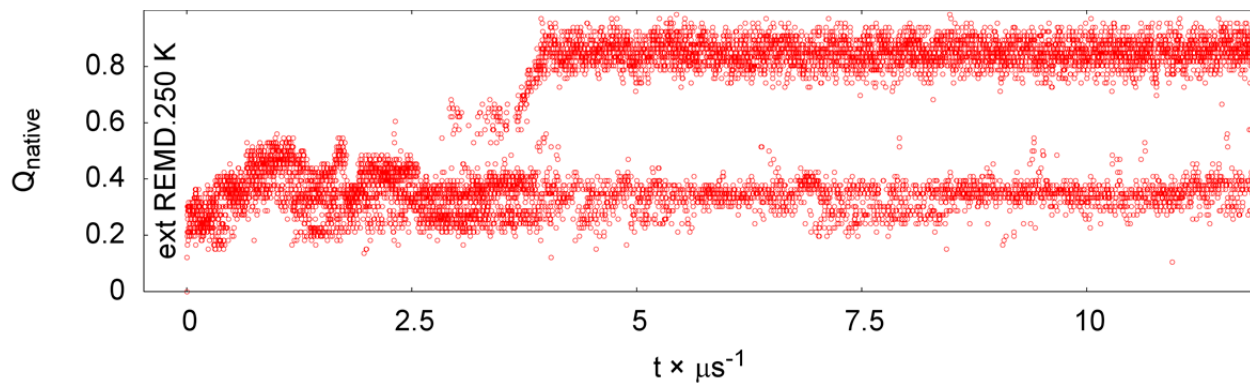


Figure S32. NTL9 (39 AA) native contacts versus time.

BBL

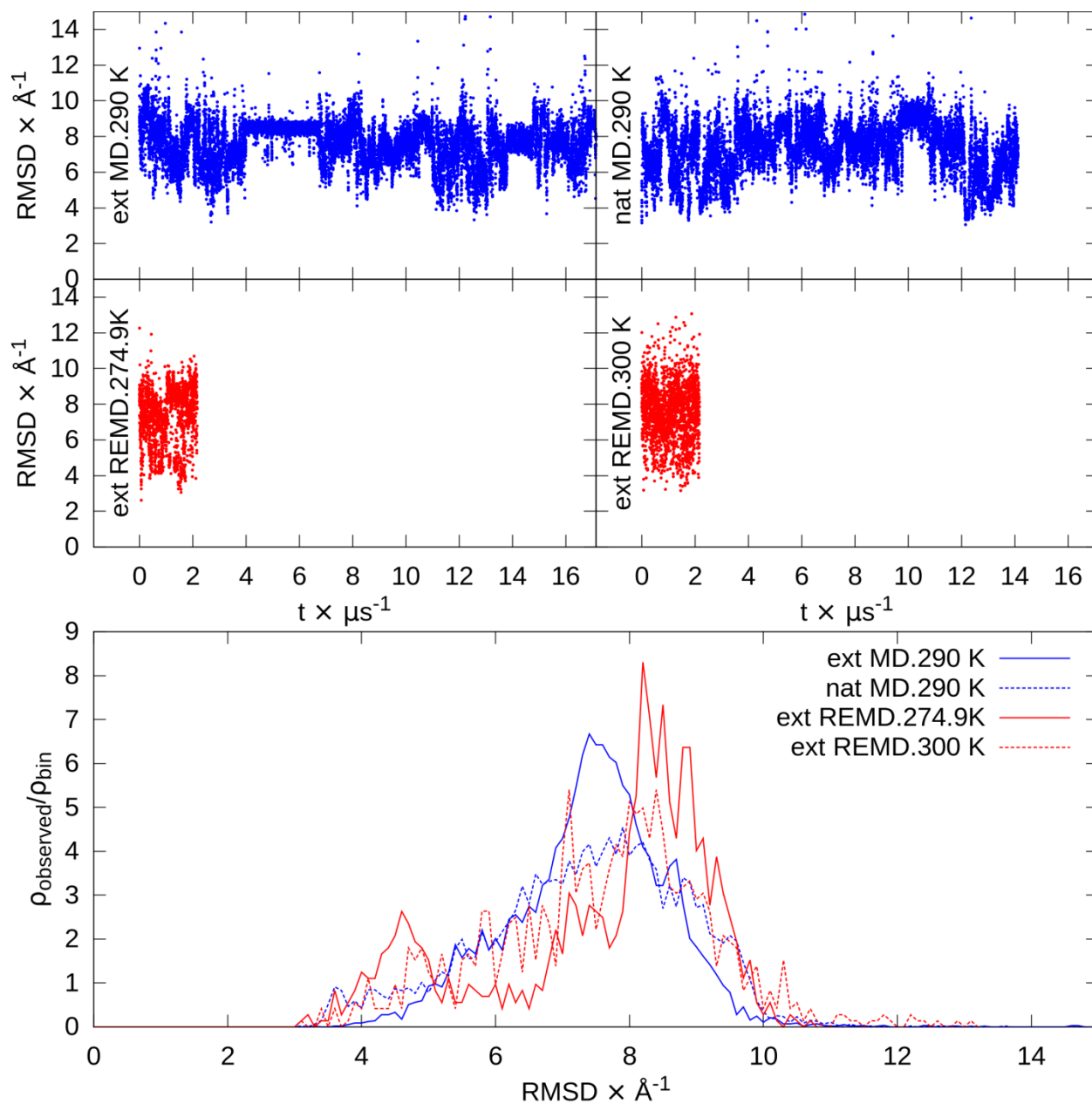


Figure S33. BBL RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).



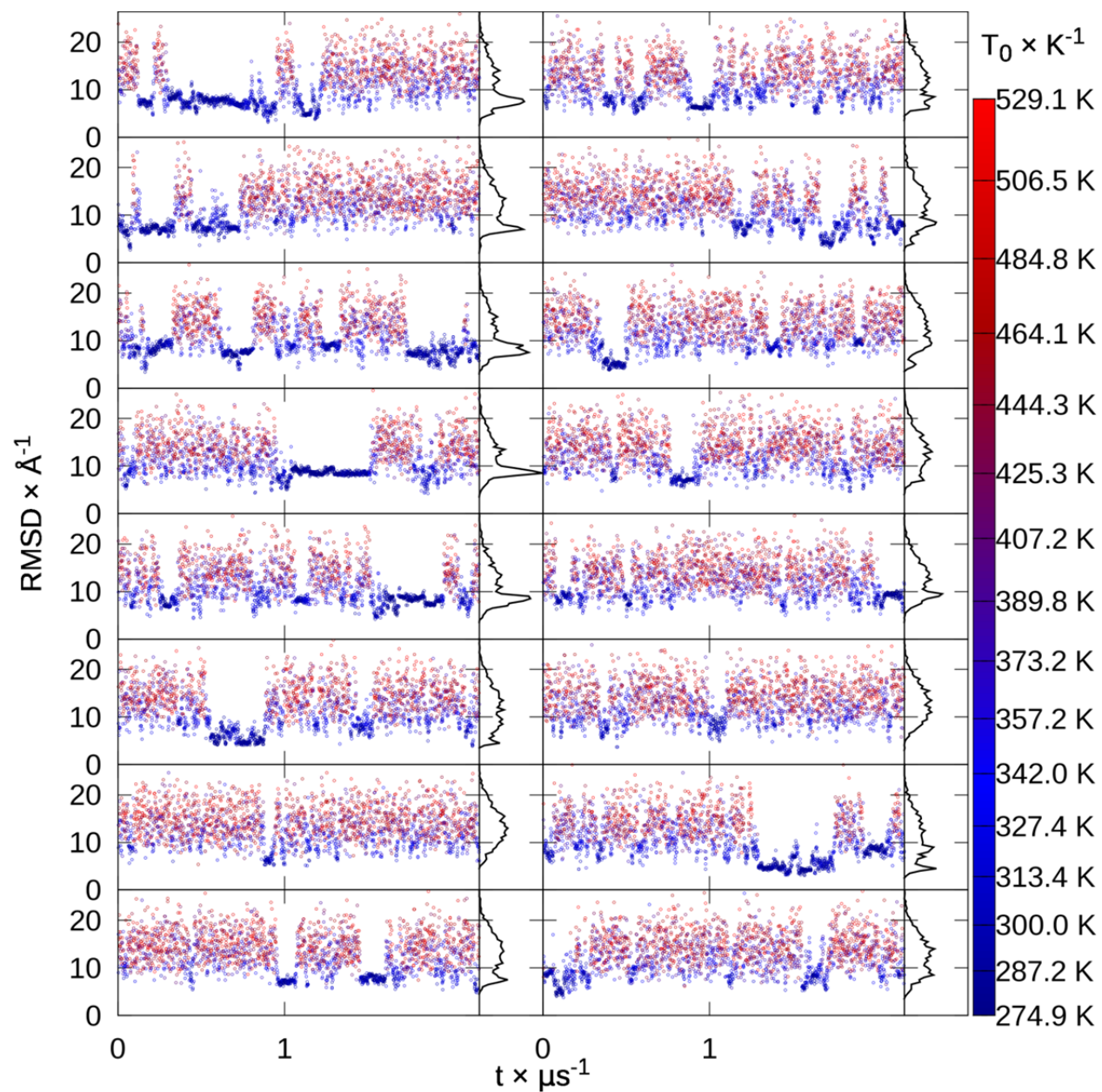


Figure S34. BBL replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	8.4	6.9	4.9	4.8	4.4
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	8.3	4.3	4.8	8.2	9.3

**Table S12.** BBL top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

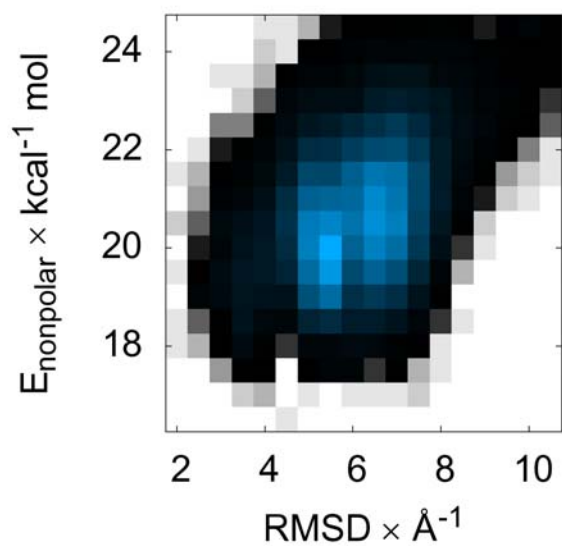


Figure S35. BBL surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is flat from mid ( $6\text{--}7 \text{ \AA}$ ) to mid-low ( $3\text{--}4 \text{ \AA}$ ) RMSD.

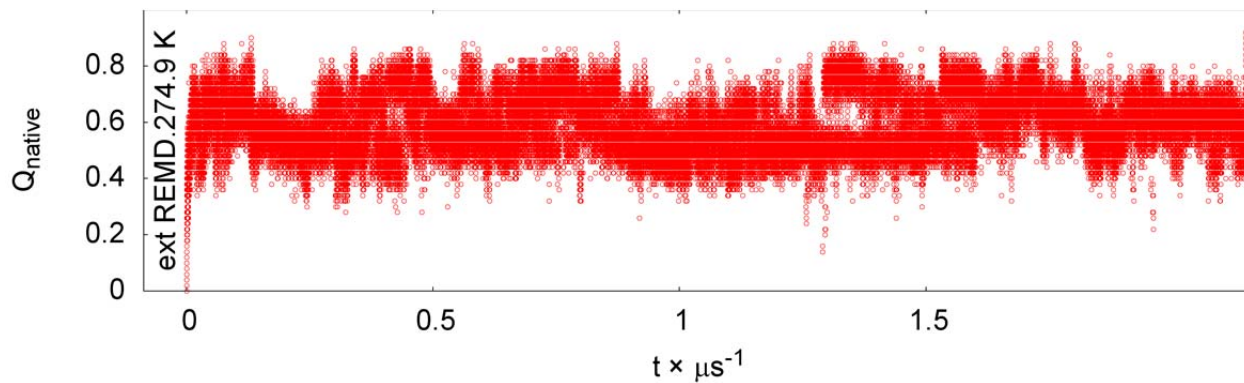


Figure S36. BBL native contacts versus time.

Protein B

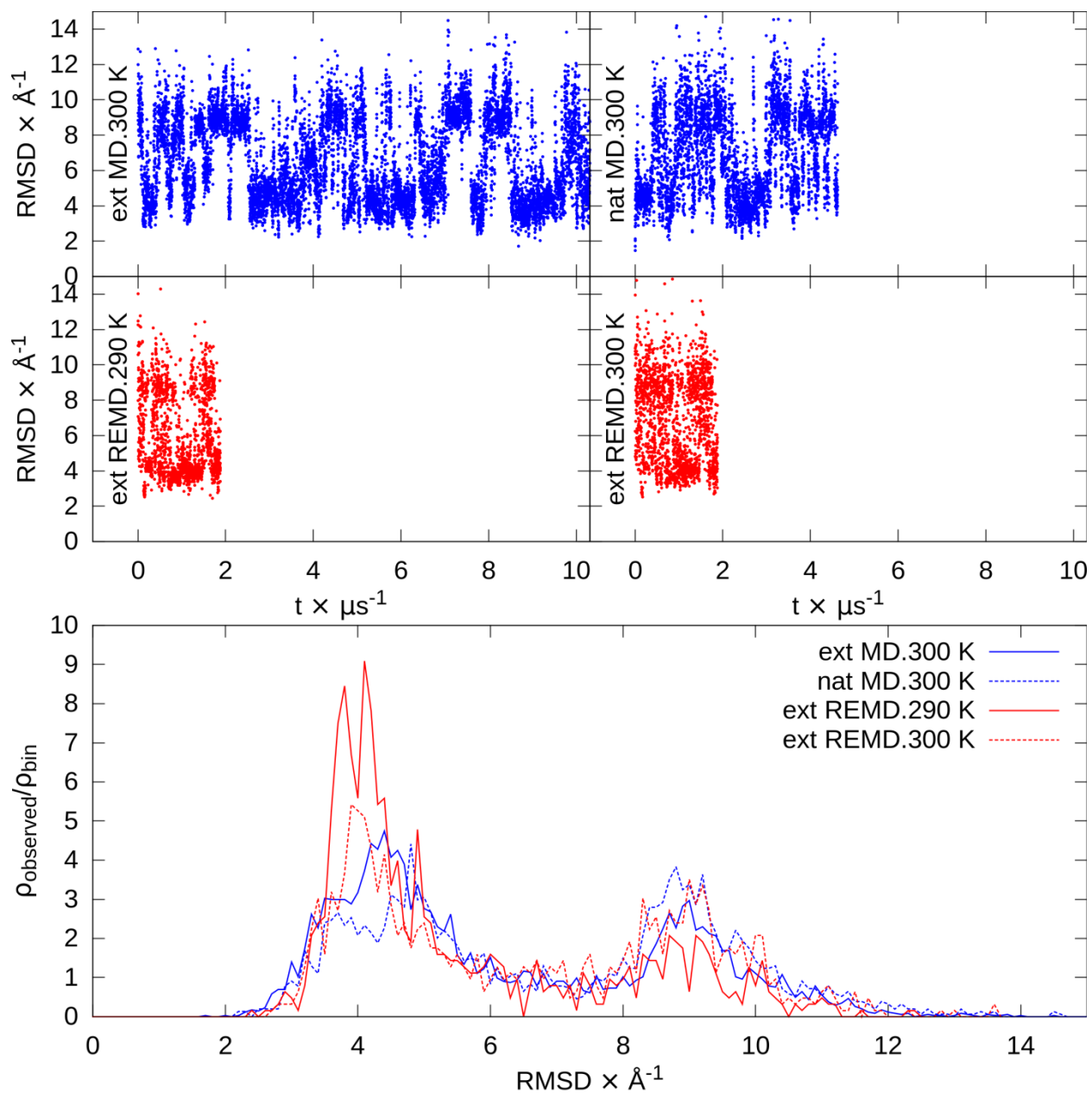


Figure S37. Protein B RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

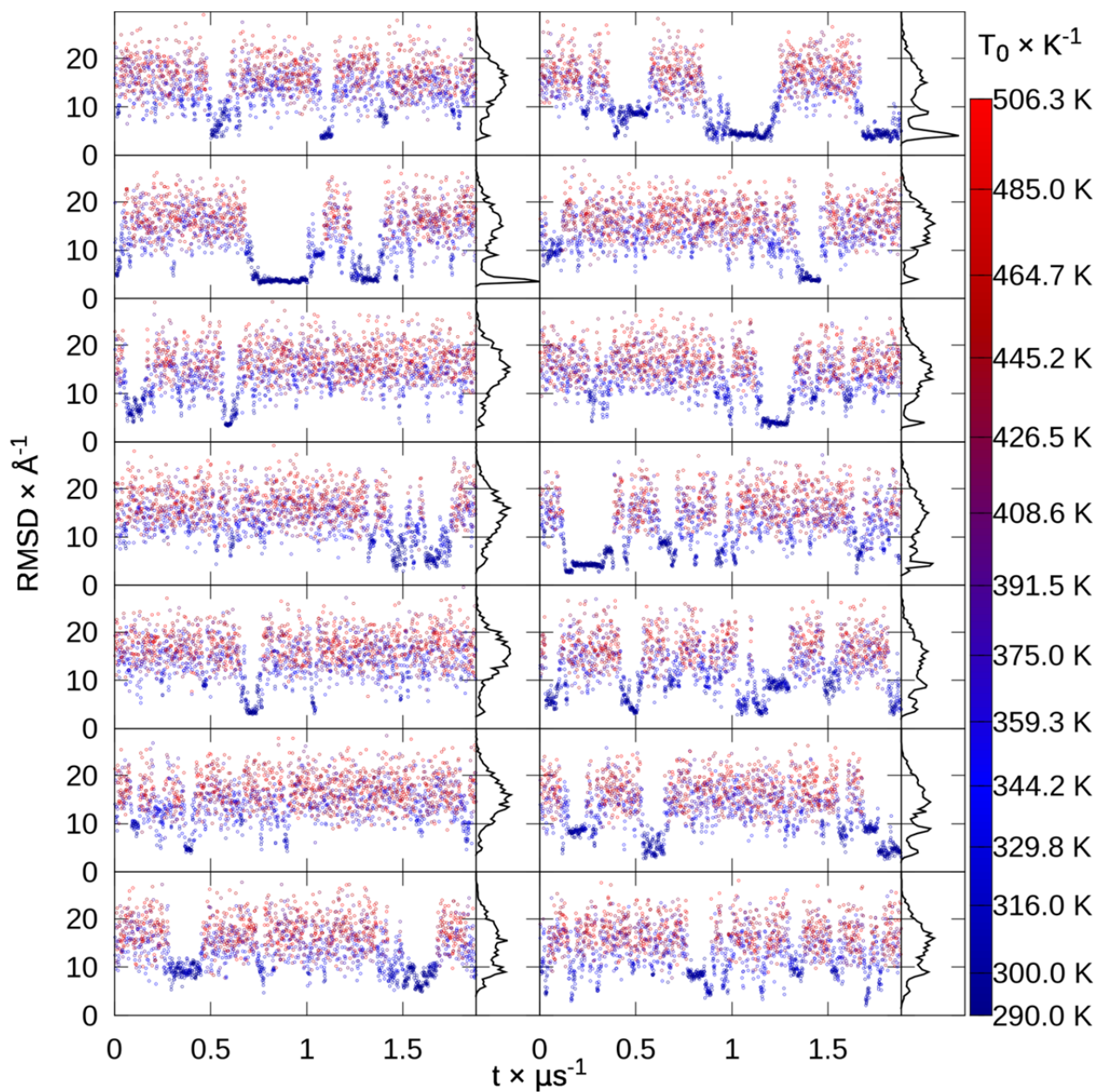


Figure S38. Protein B replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	18.6	13.9	9.1	4.6	4.6
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	4.2	3.4	2.7	3.8	3.4

Table S13. Protein B top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

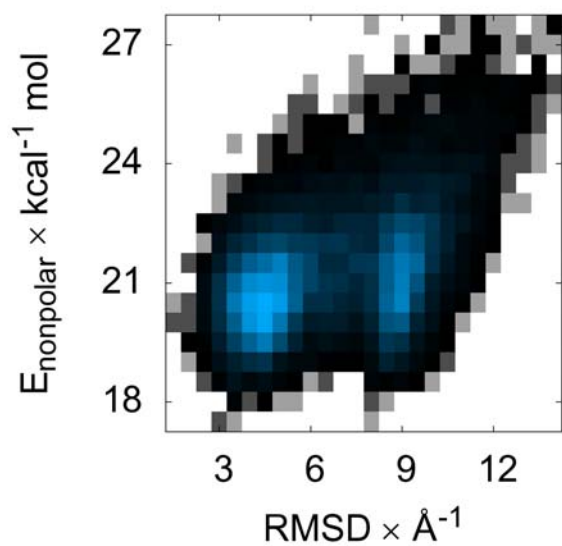


Figure S39. Protein B surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly favorable at low ( $2\text{--}4 \text{ \AA}$ ) and mid-high ( $8\text{--}9 \text{ \AA}$ ) RMSD.

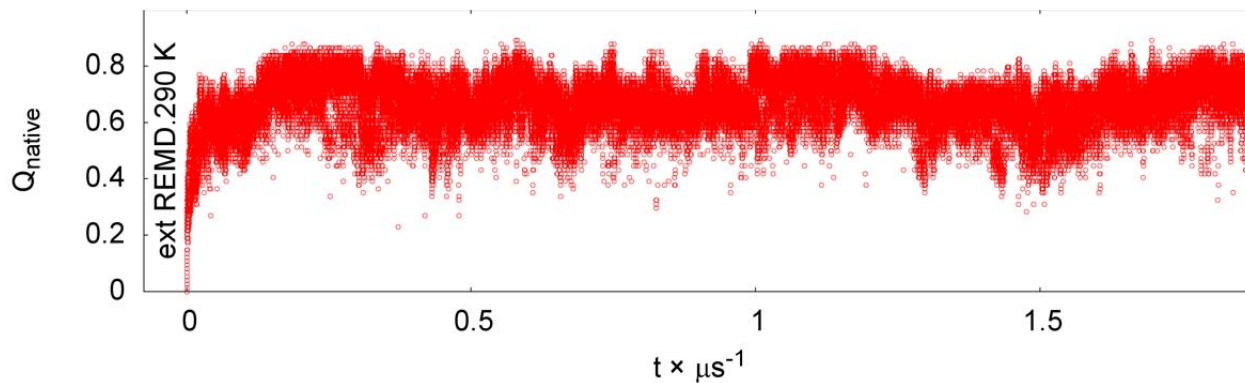


Figure S40. Protein B native contacts versus time.

### Engrailed homeodomain

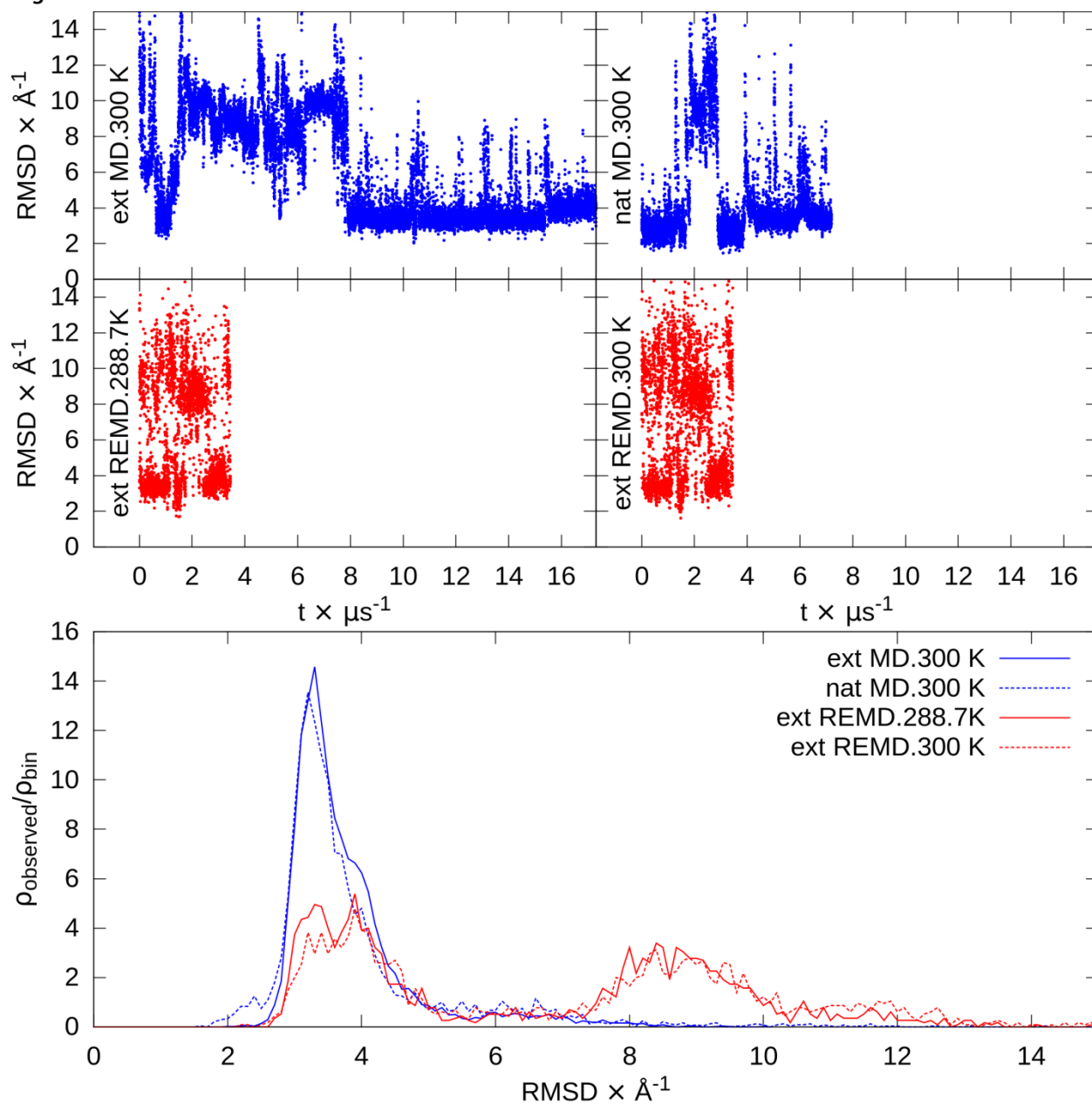


Figure S41. Homeodomain RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

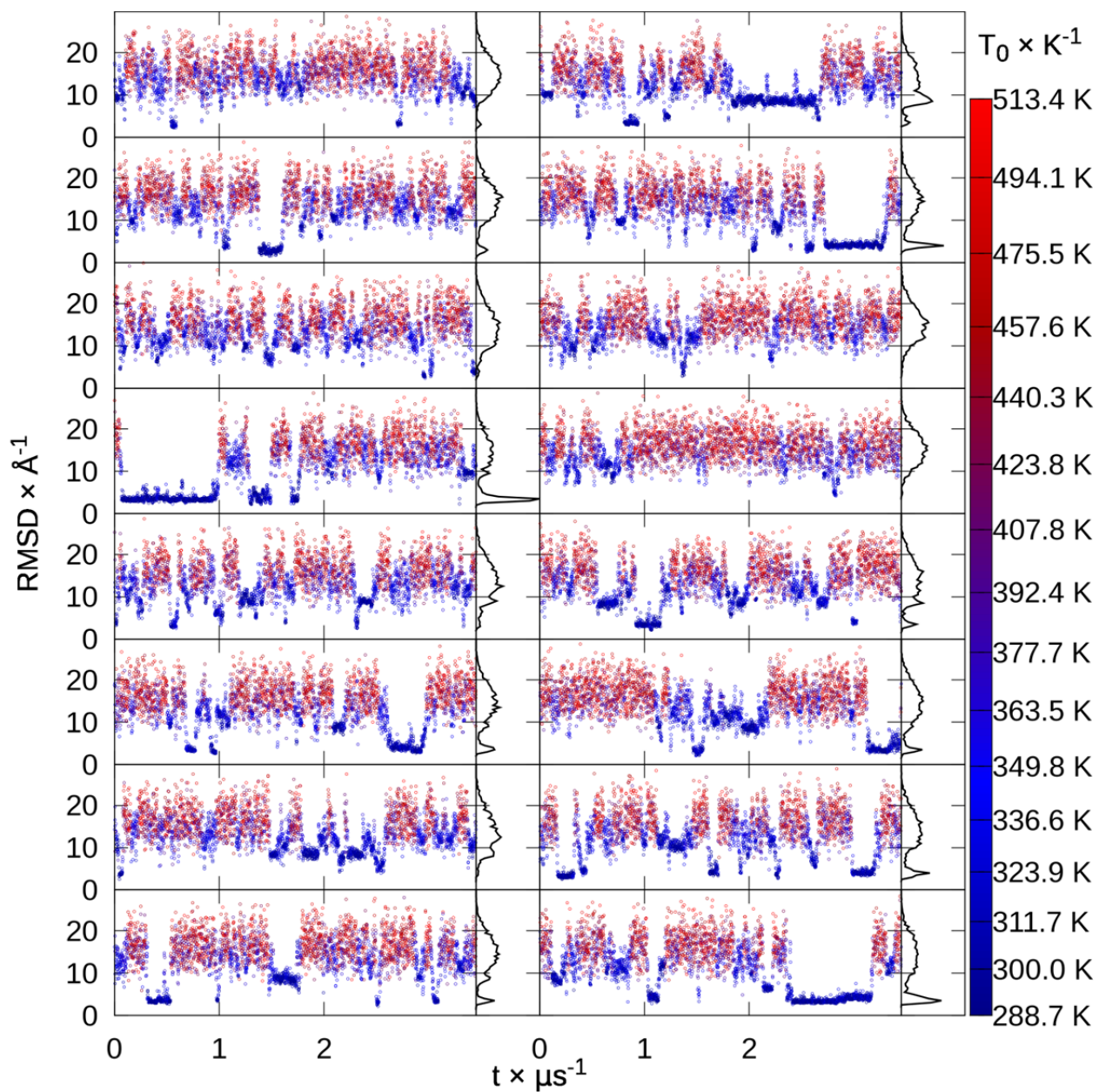


Figure S42. Engrailed homedomain replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	22.5	7.9	7.7	6.7	5.6
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	3.2	2.3	3.9	3.1	7.8

Table S14. Engrailed homeodomain top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

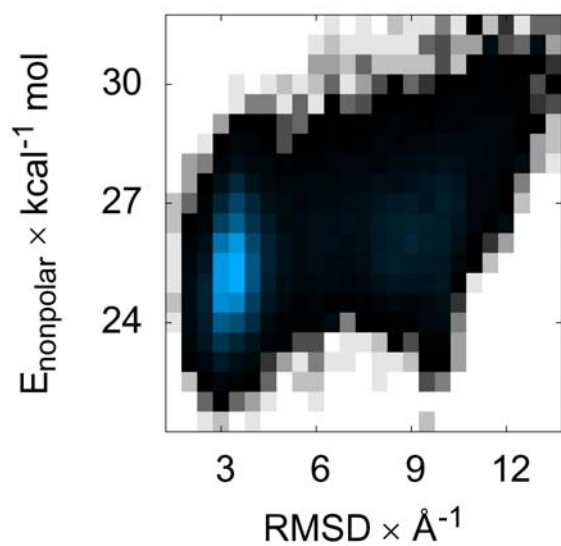


Figure S43. Engrailed homeodomain surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly favorable at low ( $2-4 \text{ \AA}$ ) and high ( $9-11 \text{ \AA}$ ) RMSDs.

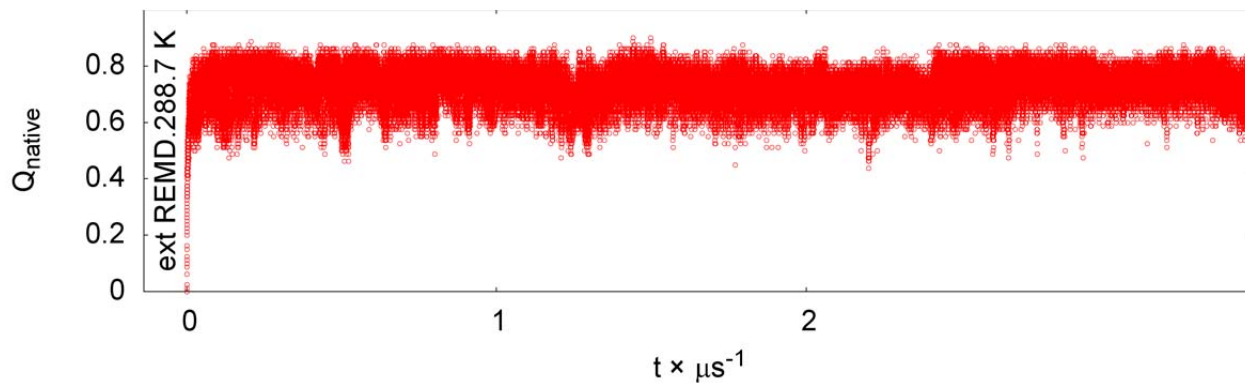


Figure S44. Engrailed homeodomain native contacts versus time.



NTL9 (52 AA)

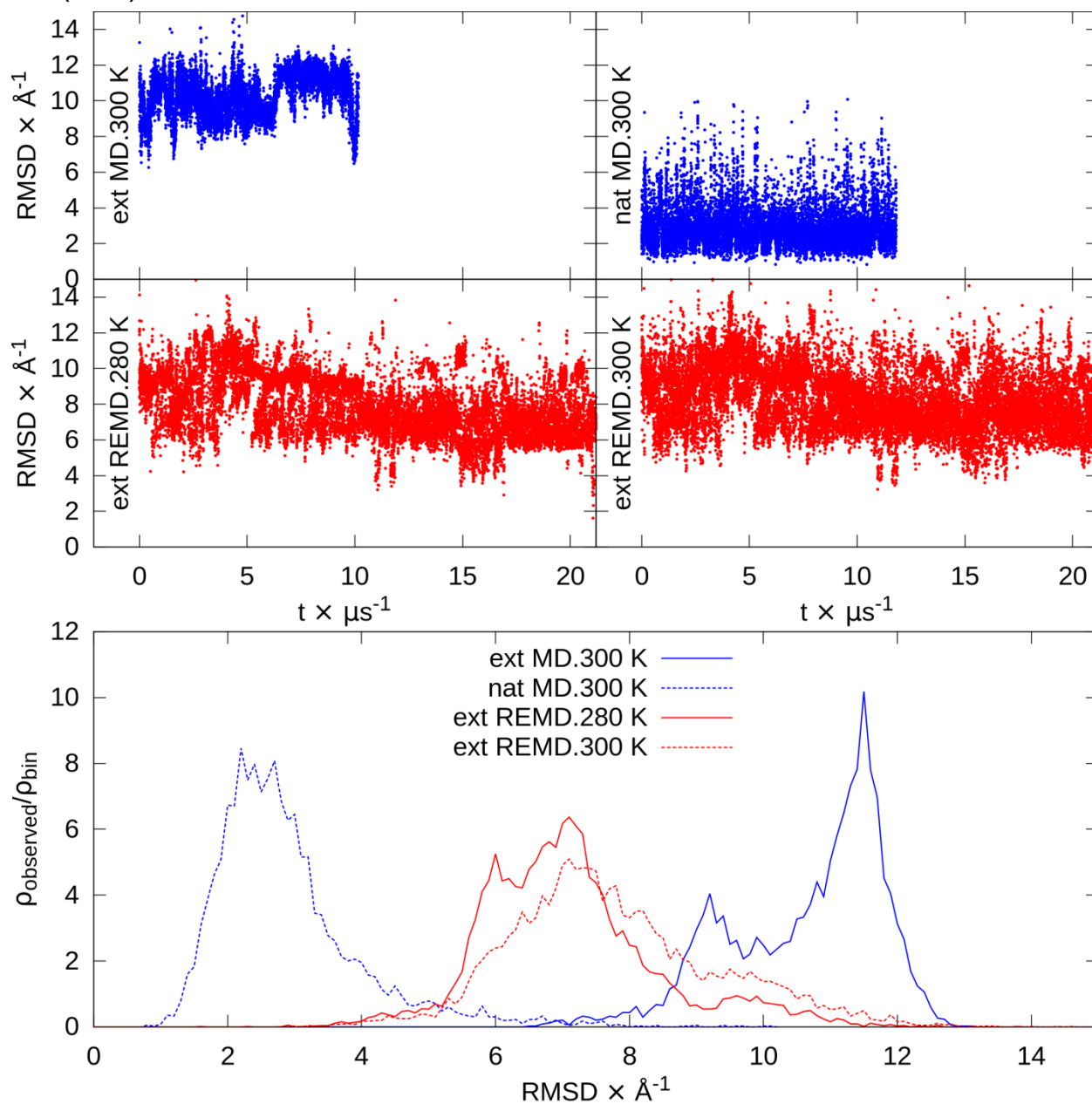


Figure S45. NTL9 (52 AA) RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

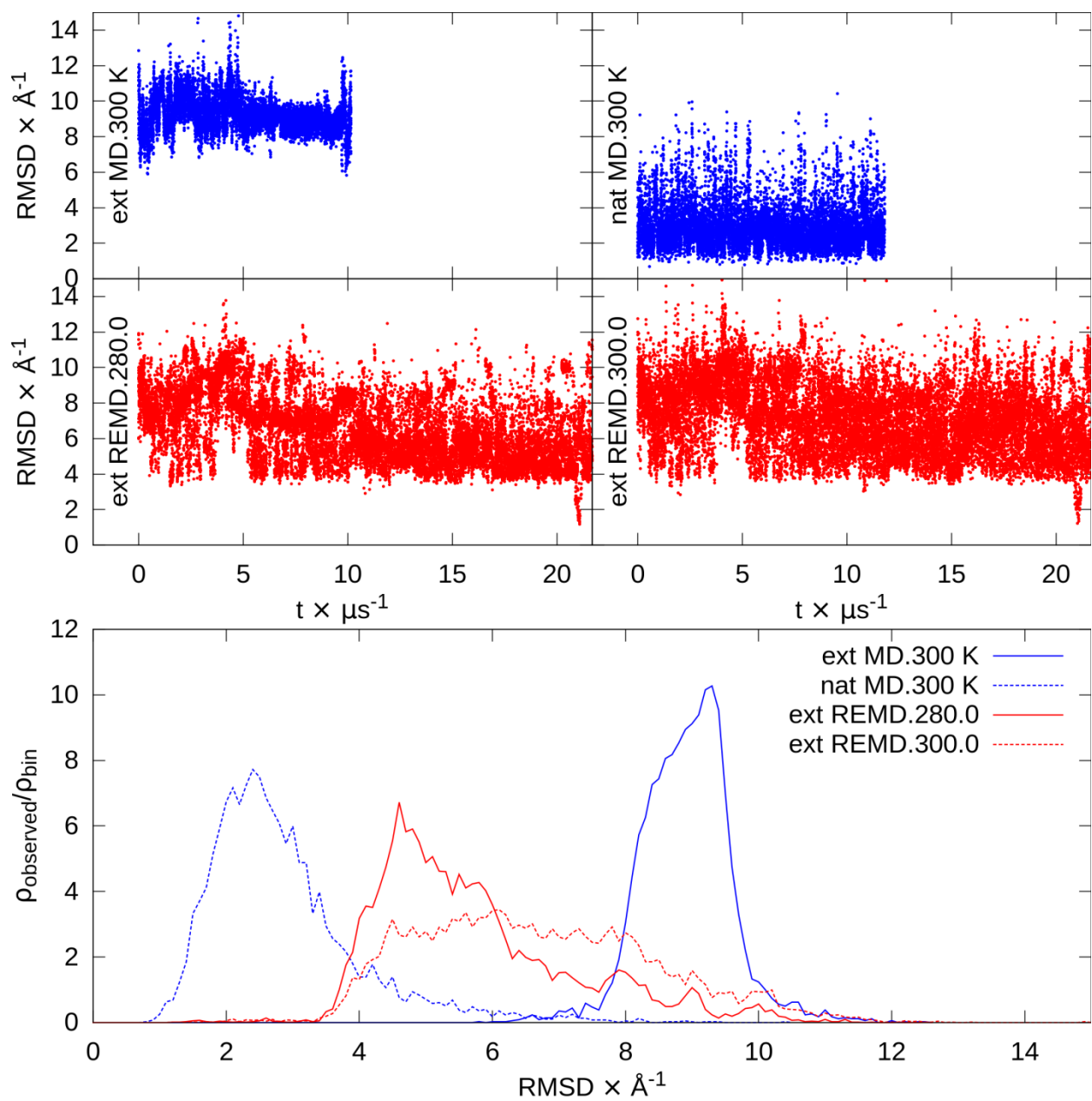


Figure S46. NTL9 (52 AA) RMSDs, excluding the 7-16 loop. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the bin size ( $\rho_{\text{bin}}$ ).

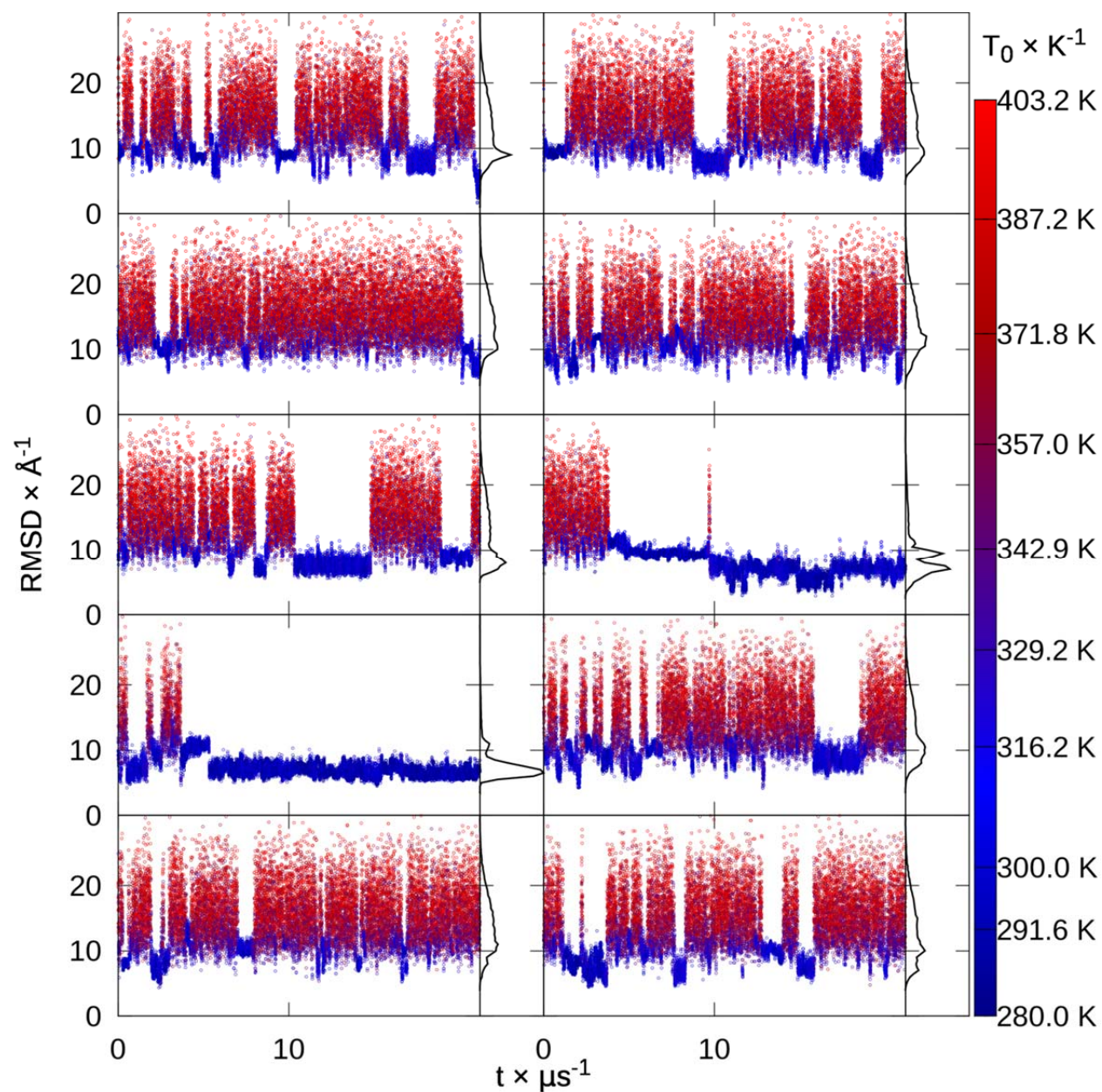


Figure S47. NTL9 (52 AA) replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	25.0	9.5	8.1	6.6	6.0
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	6.0	9.6	6.7	6.1	7.2

Table S15. NTL9 (52 AA) top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

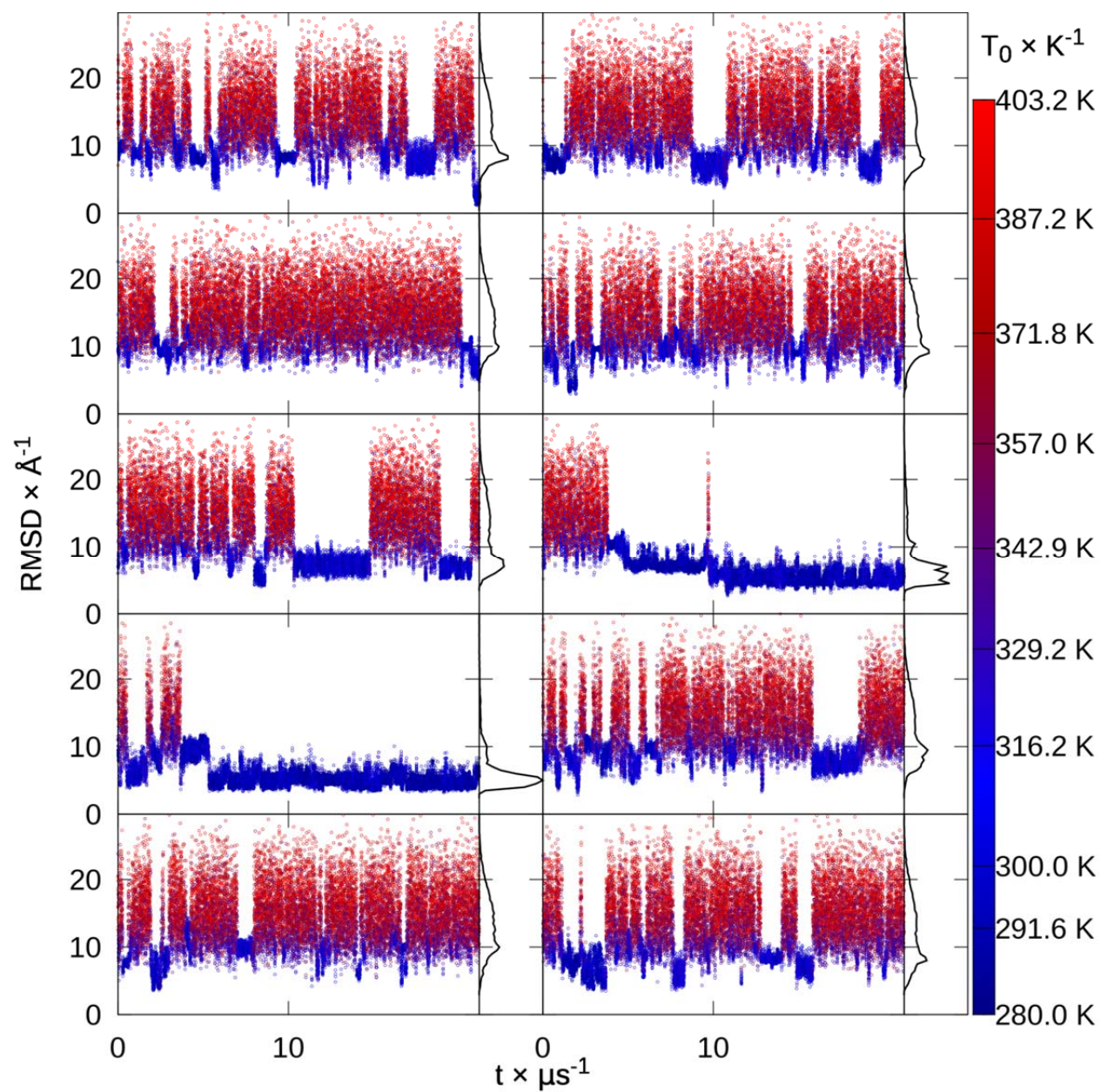


Figure S48. NTL9 (52 AA) replica RMSDs, excluding loop. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

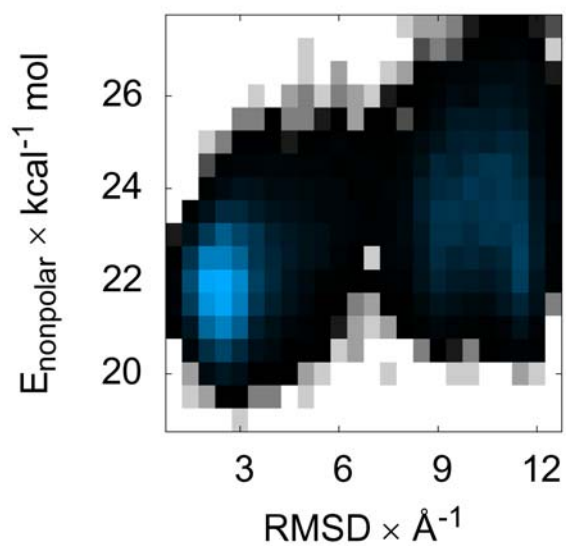


Figure S49. NTL9 (52 AA) surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is more favorable at low (1–3  $\text{\AA}$ ) than high (9–12  $\text{\AA}$ ) RMSDs.

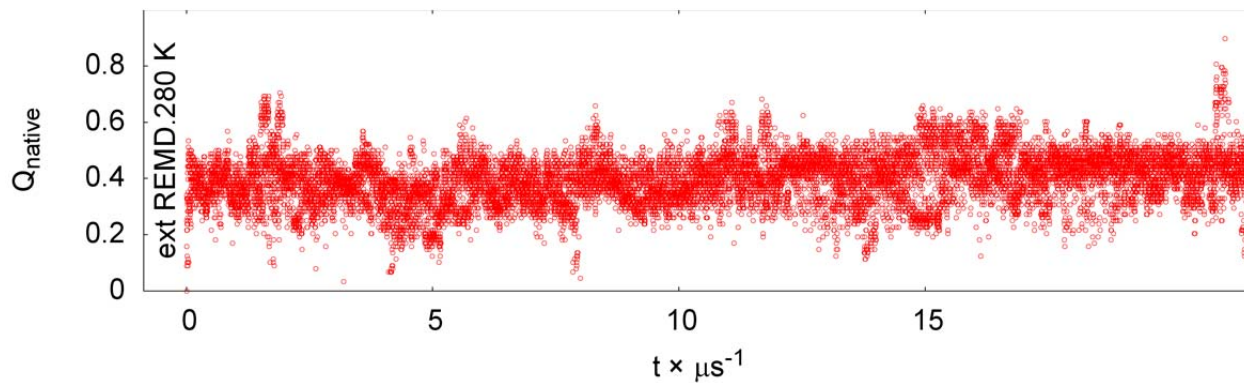


Figure S50. NTL9 (52 AA) native contacts versus time.

NuG2 variant

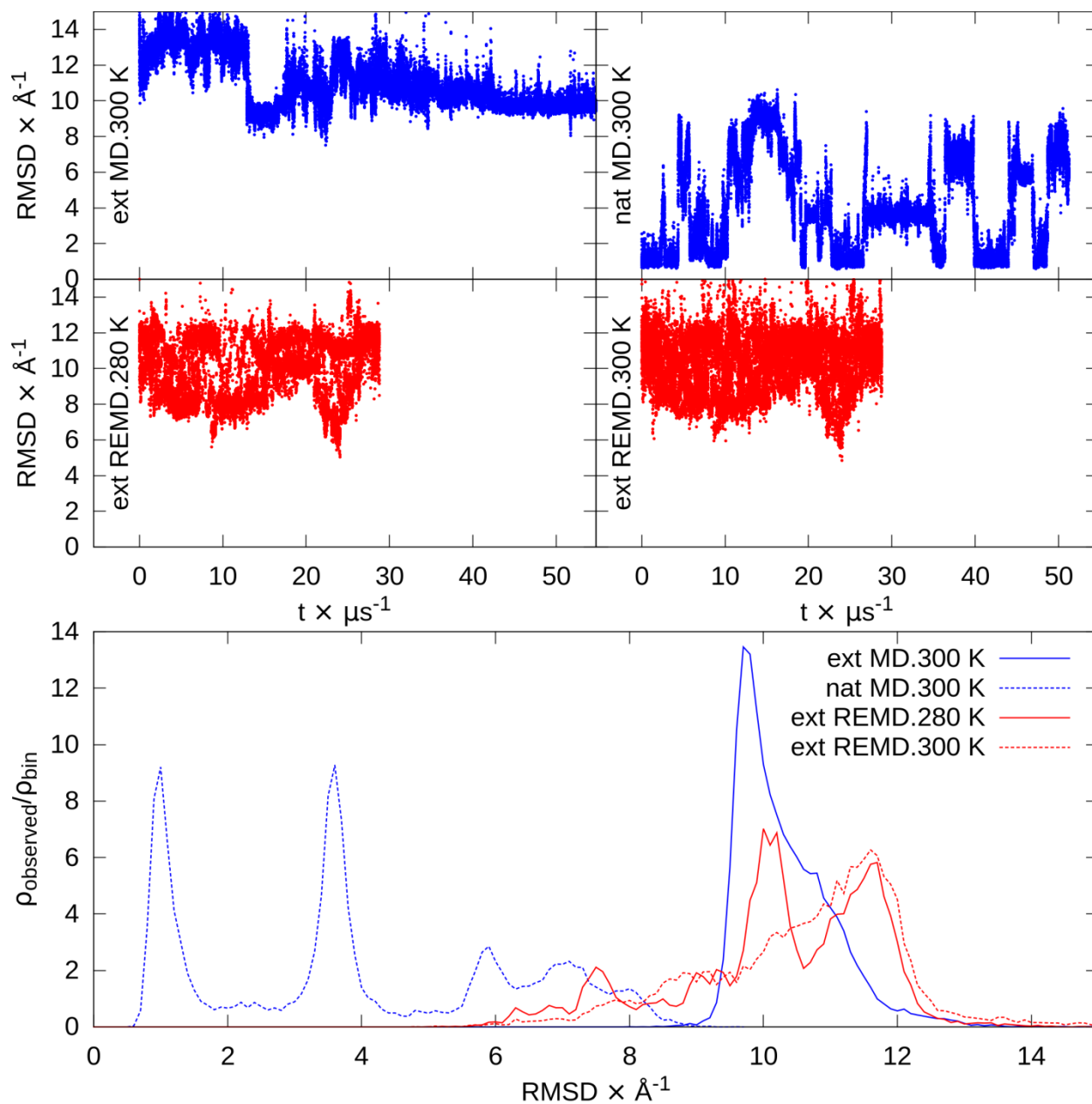


Figure S51. NuG2 variant RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

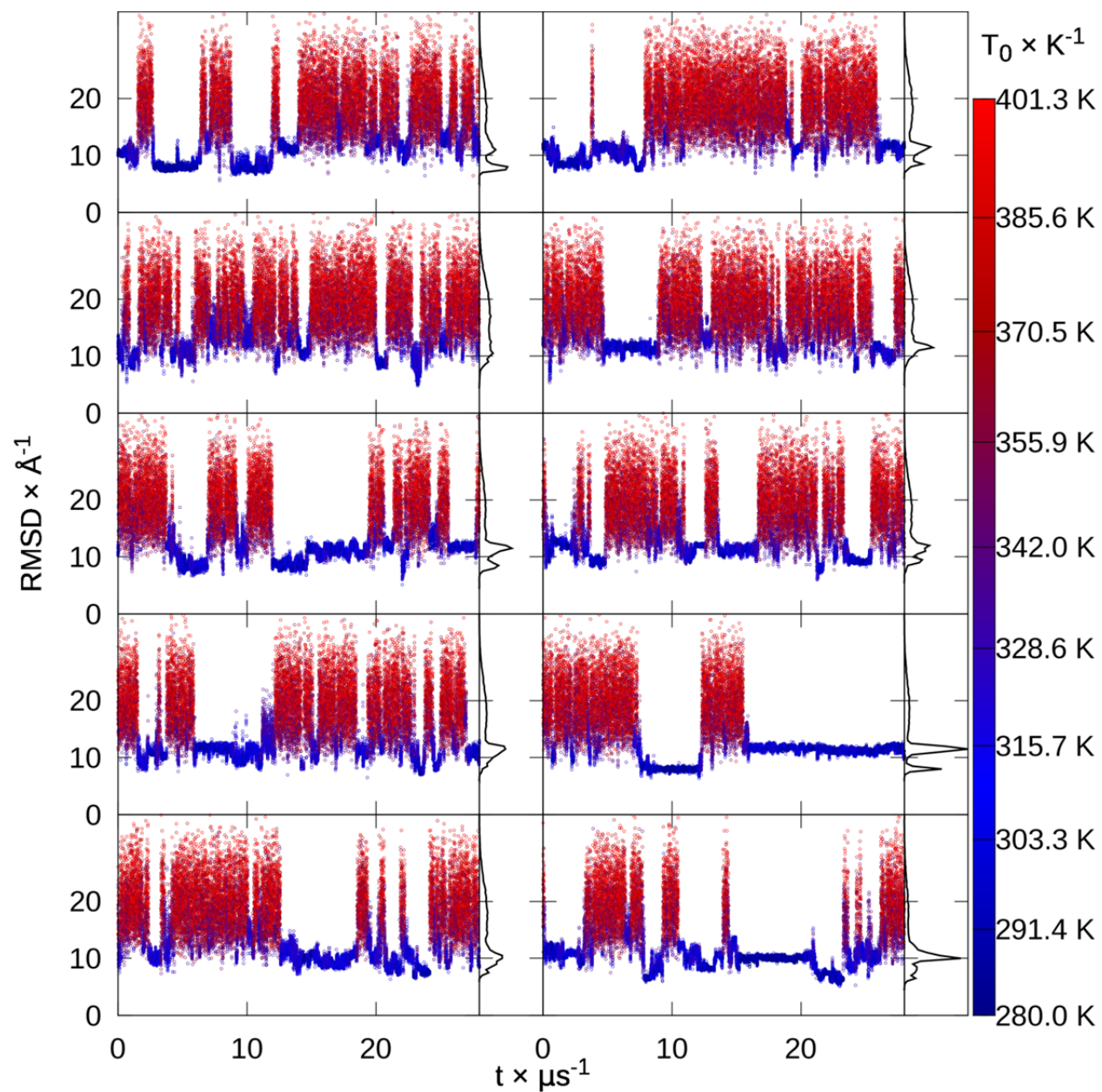


Figure S52. NuG2 variant replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	23.3	18.1	13.8	6.9	6.6
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	11.4	7.9	9.8	7.5	8.1

Table S16. NuG2 variant top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

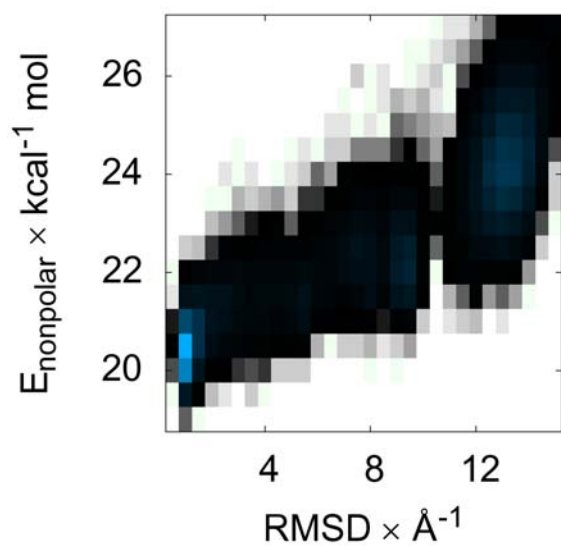


Figure S53. NuG2 variant surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is more favorable at low ( $0\text{--}2 \text{ \AA}$ ) RMSD.

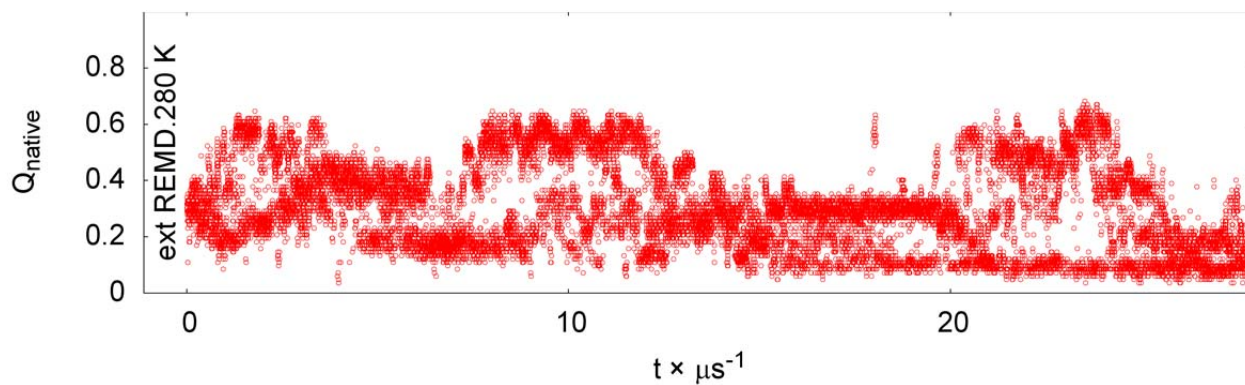


Figure S54. NuG2 variant native contacts versus time.



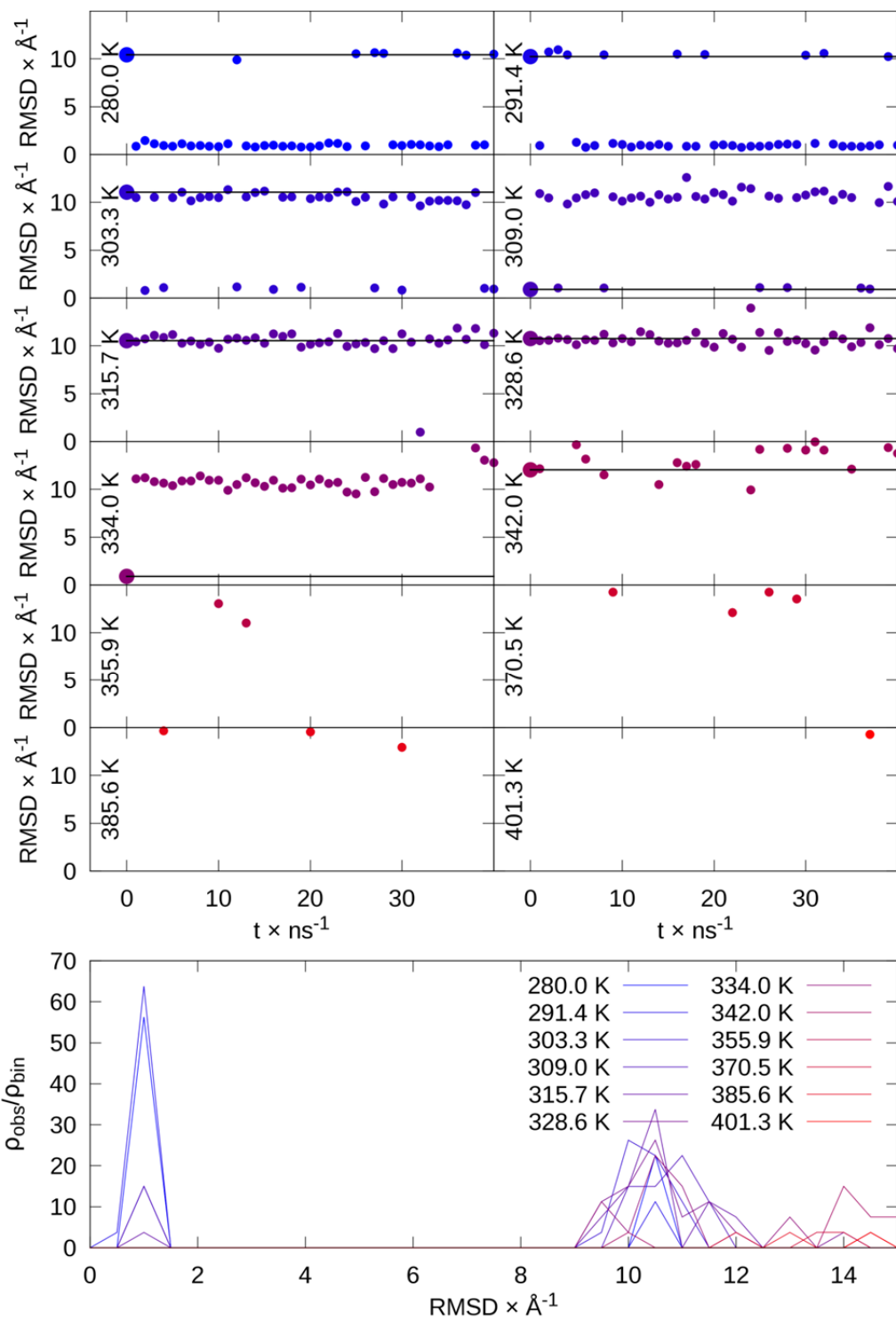


Figure S55. Seeded REMD sorting of NuG2 conformations:  $2 \times$  native-like ( $0.9 \text{ \AA}$ ) added to 10 conformations from extended REMD from  $10.4$  to  $30.3 \text{ \AA}$  RMSD. Lines indicate initial RMSD value that each temperature. At top, RMSD vs time for each temperature shows sorting of low RMSD conformations to low temperatures. At bottom, histogram shows preference of native-like conformations at low temperatures.

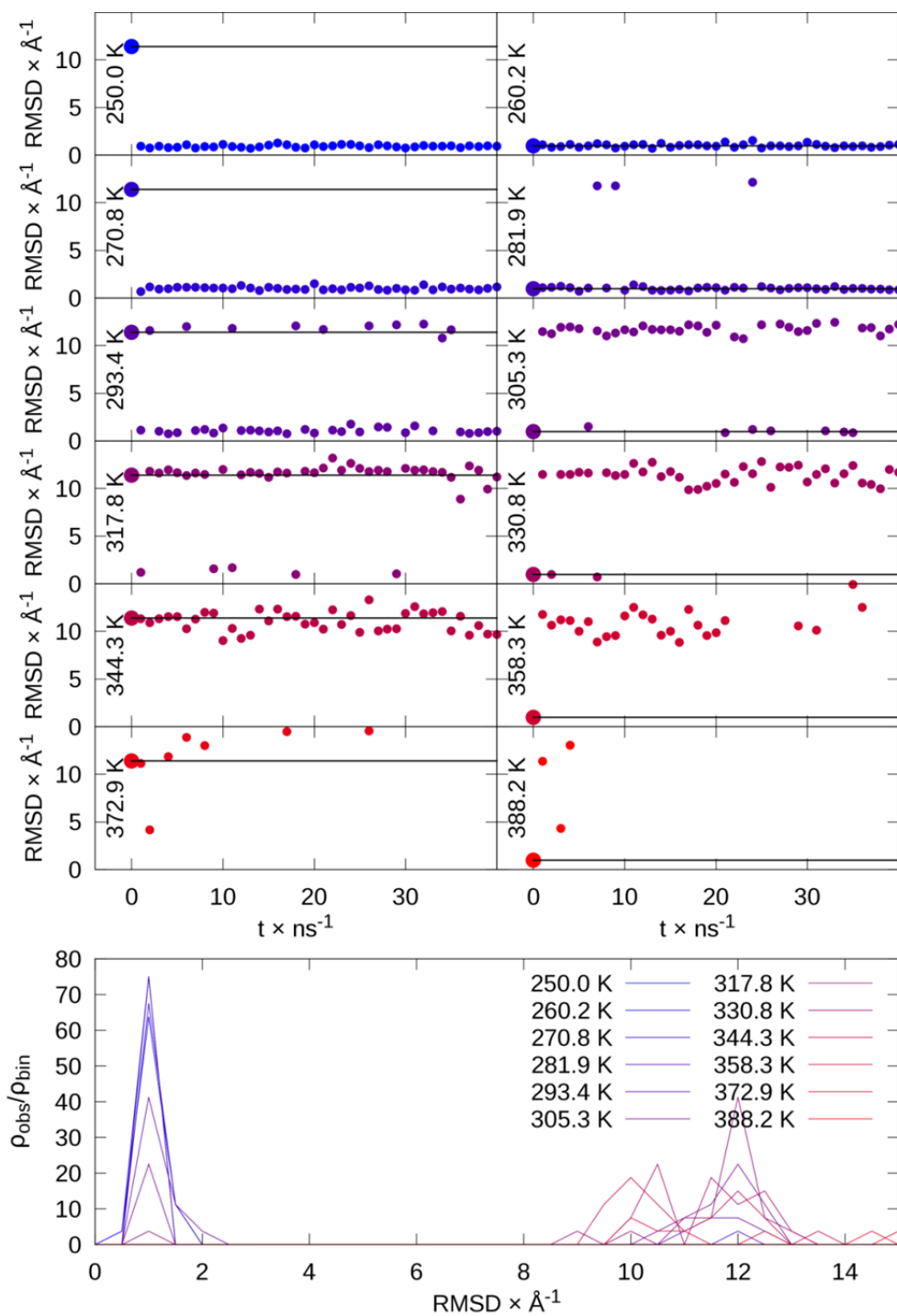


Figure S56. REMD sorting of NuG2 conformations: unfolded (11.4  $\text{\AA}$ ) and native-like (1.0  $\text{\AA}$ ), repeated for 12 replicas. Lines indicate initial RMSD value that each temperature. At top, RMSD vs time for each temperature shows sorting of low RMSD conformations to low temperatures, except for the native conformation at 388.2 K that unfolded. At bottom, histogram shows preference of native-like conformations at low temperatures.

## CspA

For the  $\beta$ -barrel CspA, the most populated conformation forms a barrel with correct strands 1-3, but the long flexible loop from positions 35-47 adopts a  $\beta$ -hairpin and displaces strand 5, which moves to where strand 4 should be, and the displaced strand 4 adopts a helical conformation (Figure S2). The population is comparable to that of another cluster with near-native fold at 4.8 Å (Table S17).

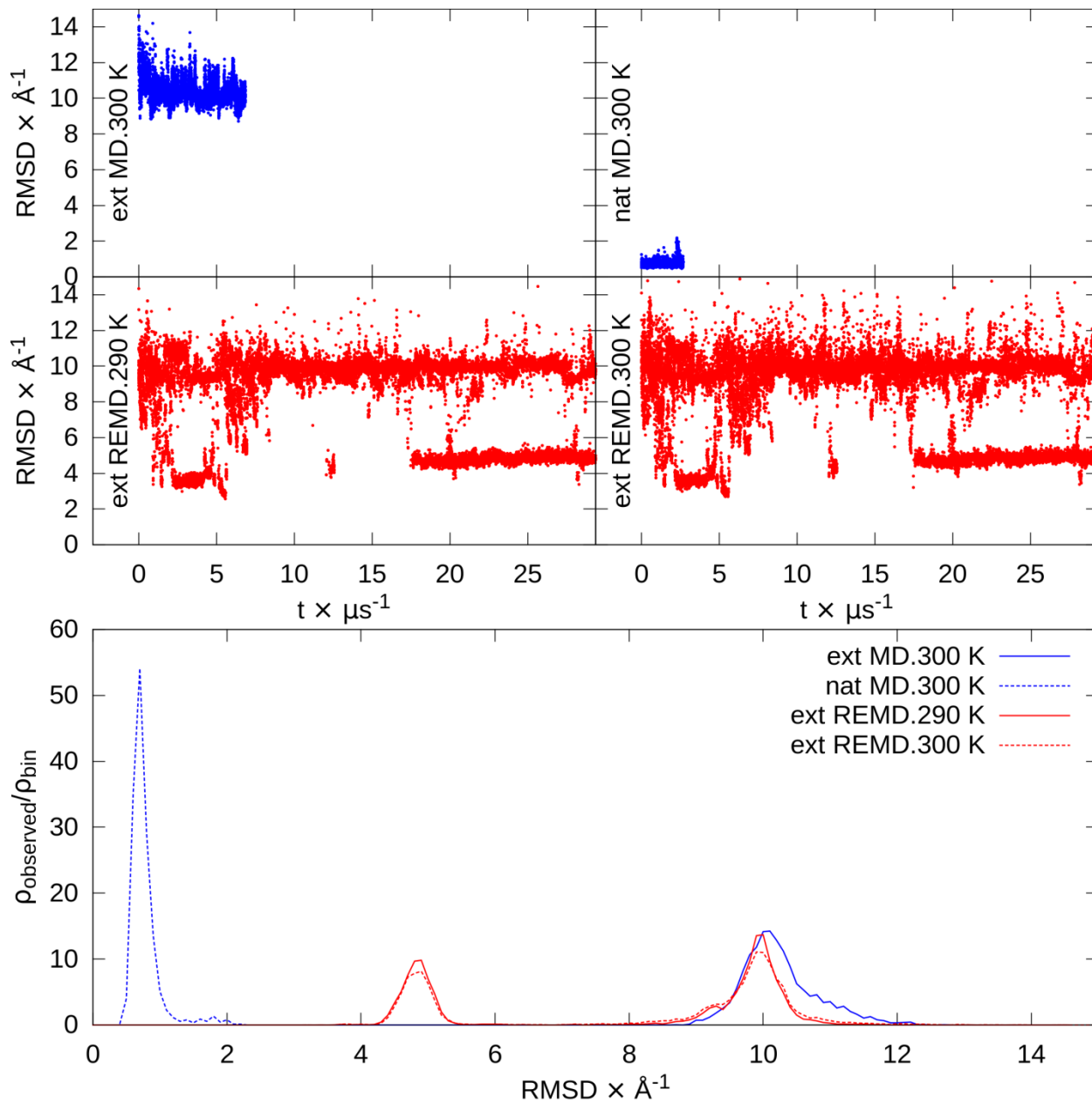


Figure S57. CspA RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

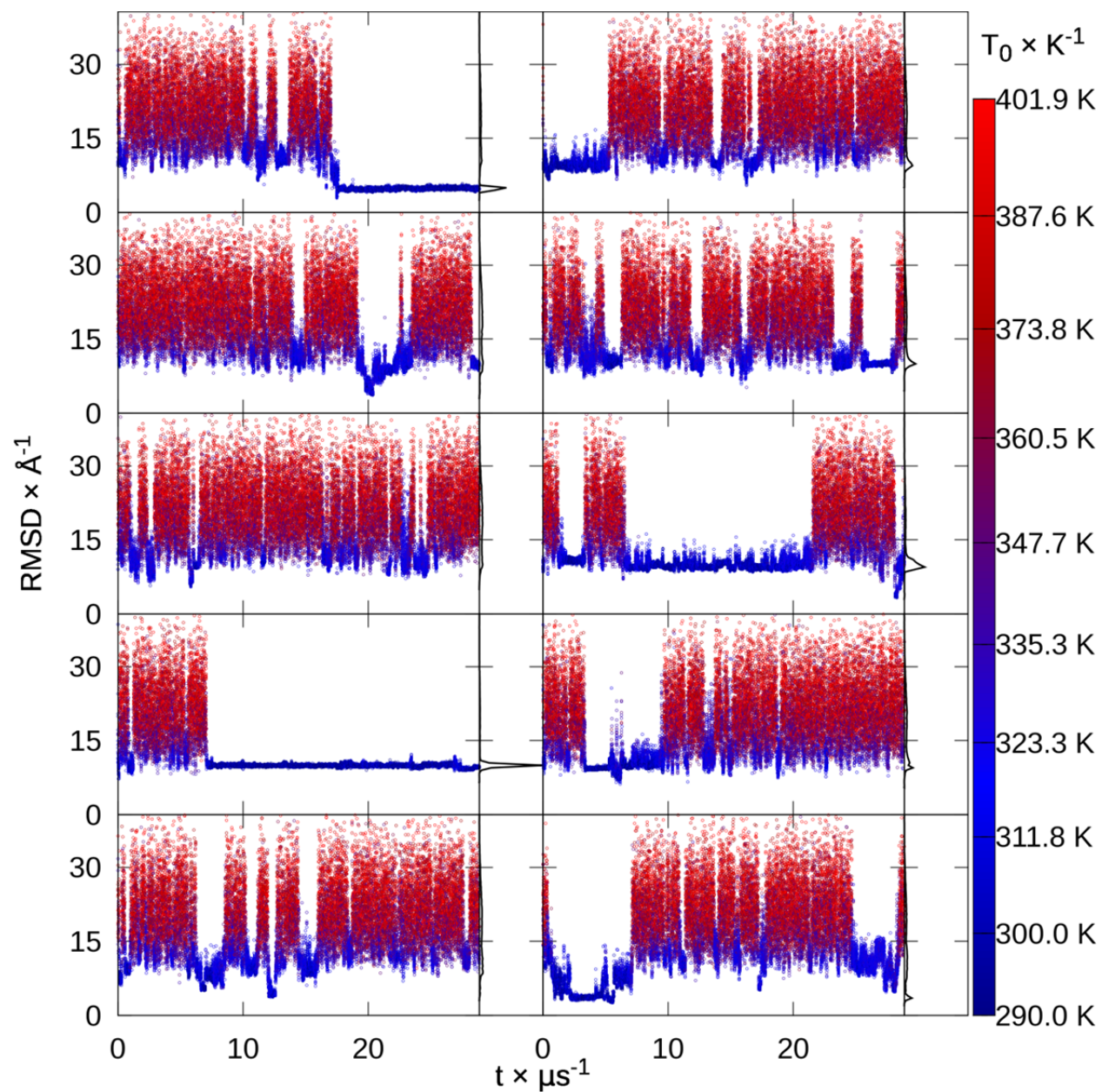


Figure S58. CspA replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	33.0	17.6	11.5	6.0	5.2
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	9.9	4.8	9.5	3.3	10.0

Table S17. CspA top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

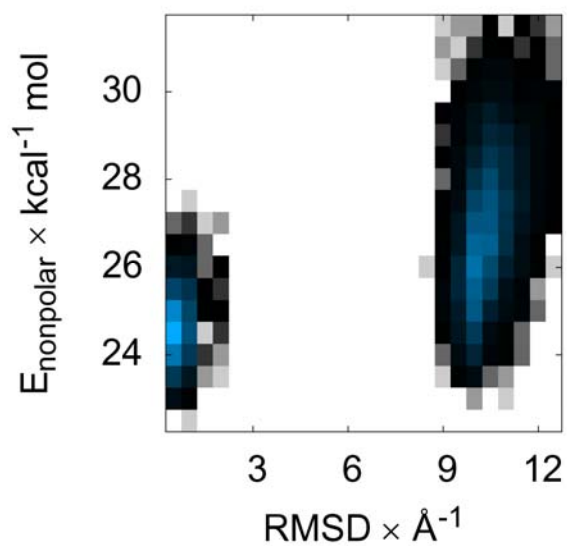


Figure S59. CspA surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is more favorable at low ( $0\text{--}2 \text{ \AA}$ ) than high ( $9\text{--}12 \text{ \AA}$ ) RMSDs.

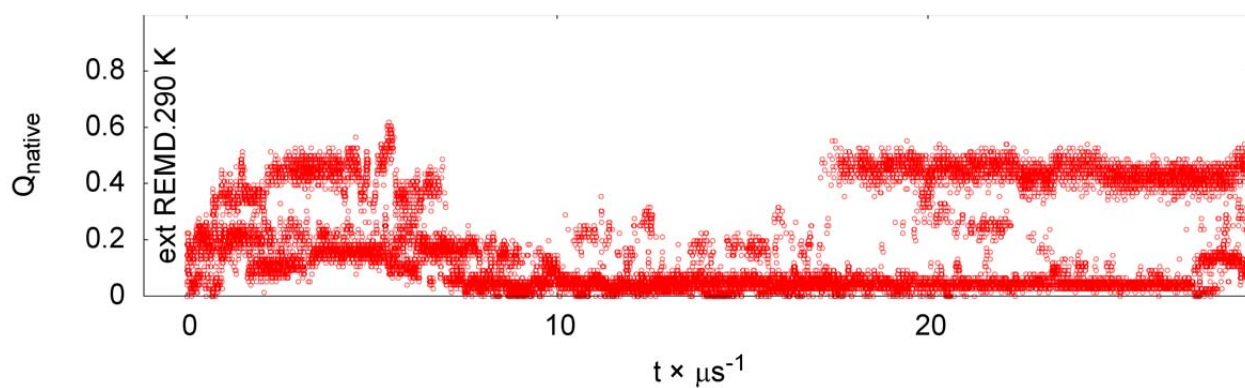


Figure S60. CspA native contacts versus time.

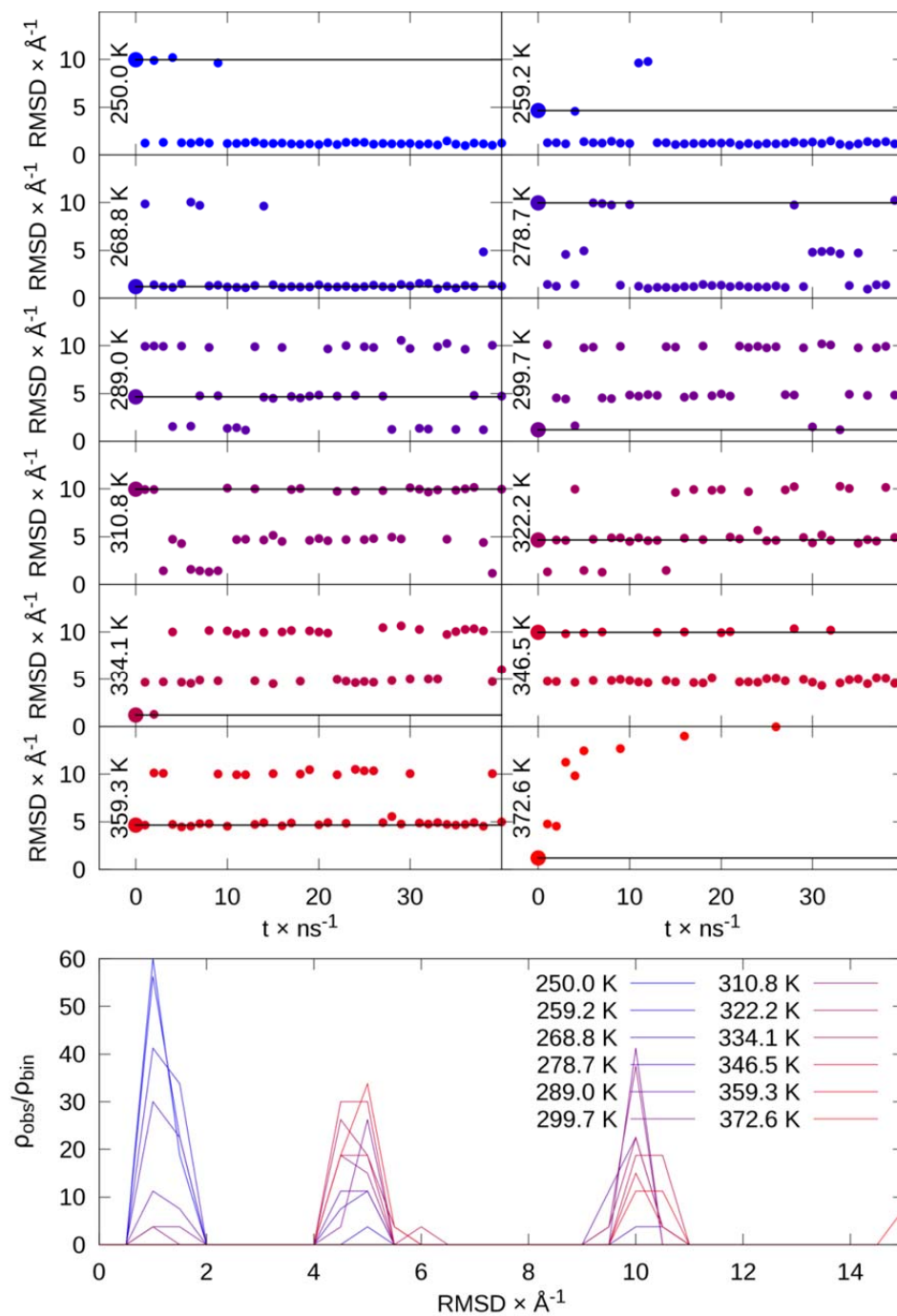


Figure S61. REMD sorting of CspA conformations: unfolded (10.0 Å), partly unfolded (4.7 Å), and native-like (1.2 Å), repeated for 12 replicas. At top, RMSD vs time for each temperature shows sorting of low RMSD conformations to low temperatures. Lines indicate initial RMSD value that each temperature. At bottom, histogram shows preference of low RMSD conformations at low temperatures.

## Hyp protein 1WHZ

The 70 amino acid hypothetical protein 1WHZ folds to the correct structure with a 3-stranded  $\beta$ -sheet and 3 helices, but the most populated structure replaces the first  $\beta$ -strand with a helix and the last two helices with two  $\beta$ -strands. Otherwise, the RMSDs of the first helix and N-terminus (residues 1 to 18) and the second and third  $\beta$ -strands (residues 28 to 44) are both 1.8 Å. Examining the RMSD evolution of individual replicas in the  $\sim 10$   $\mu$ sec REMD run indicates that multiple misfolded structures are sampled, typically stable for several  $\mu$ sec, adopting a variety of mixed  $\alpha/\beta$  topologies.

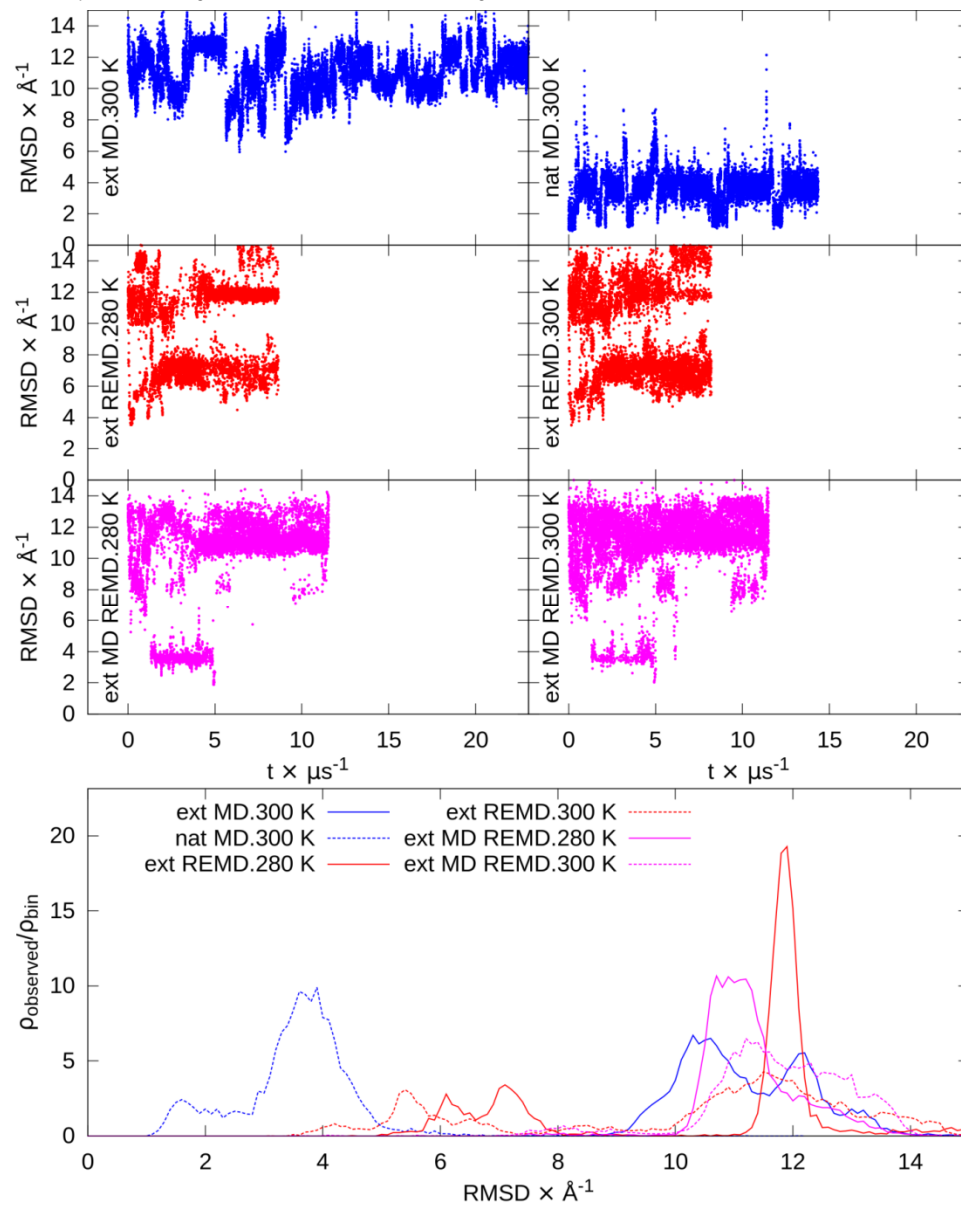


Figure S62. Hypothetical protein 1WHZ RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

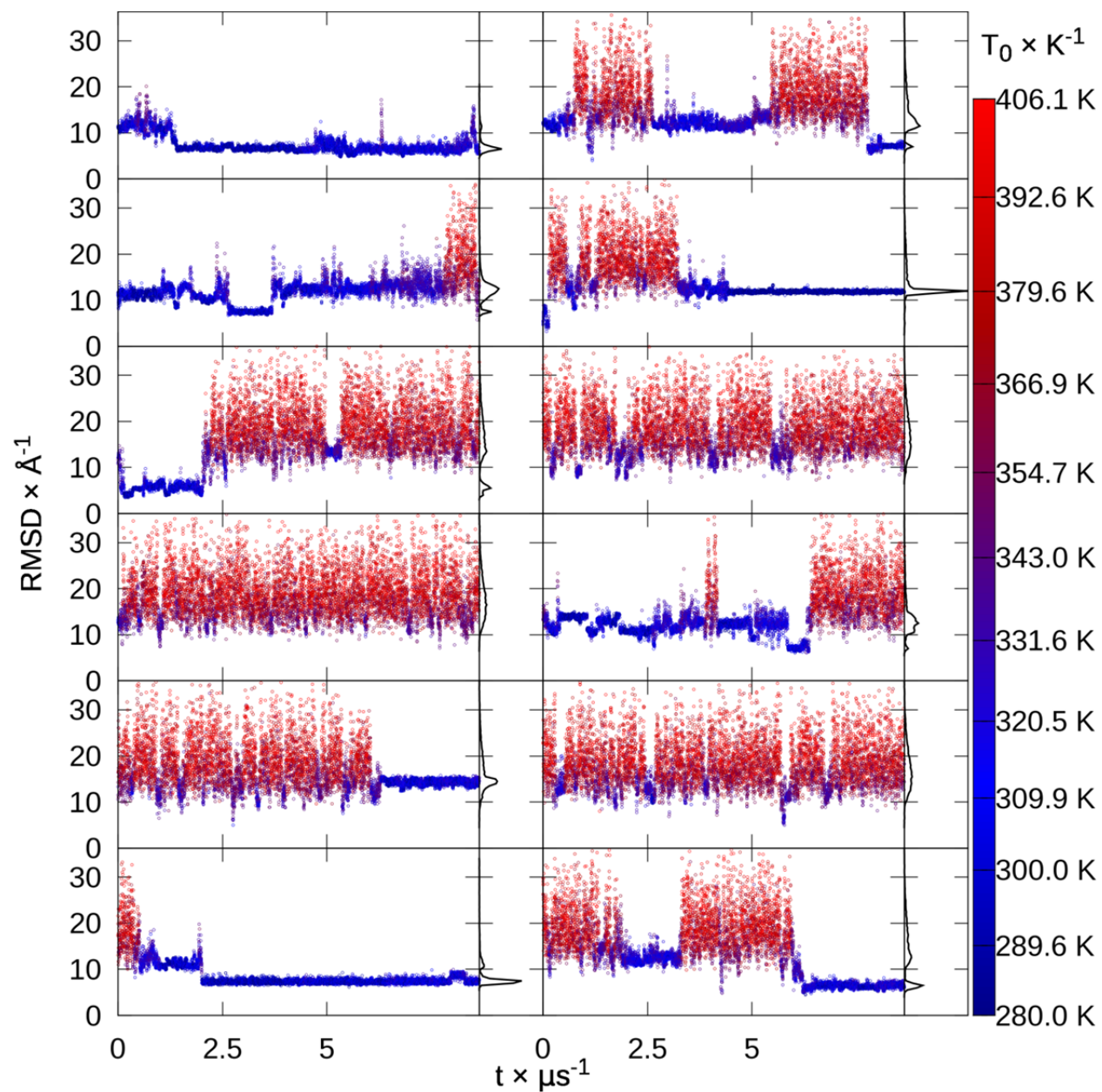


Figure S63. Hypothetical protein 1WHZ replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	36.1	16.7	14.8	5.5	4.6
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	11.8	7.2	6.6	11.8	13.9

Table S18. Hypothetical protein 1WHZ top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.



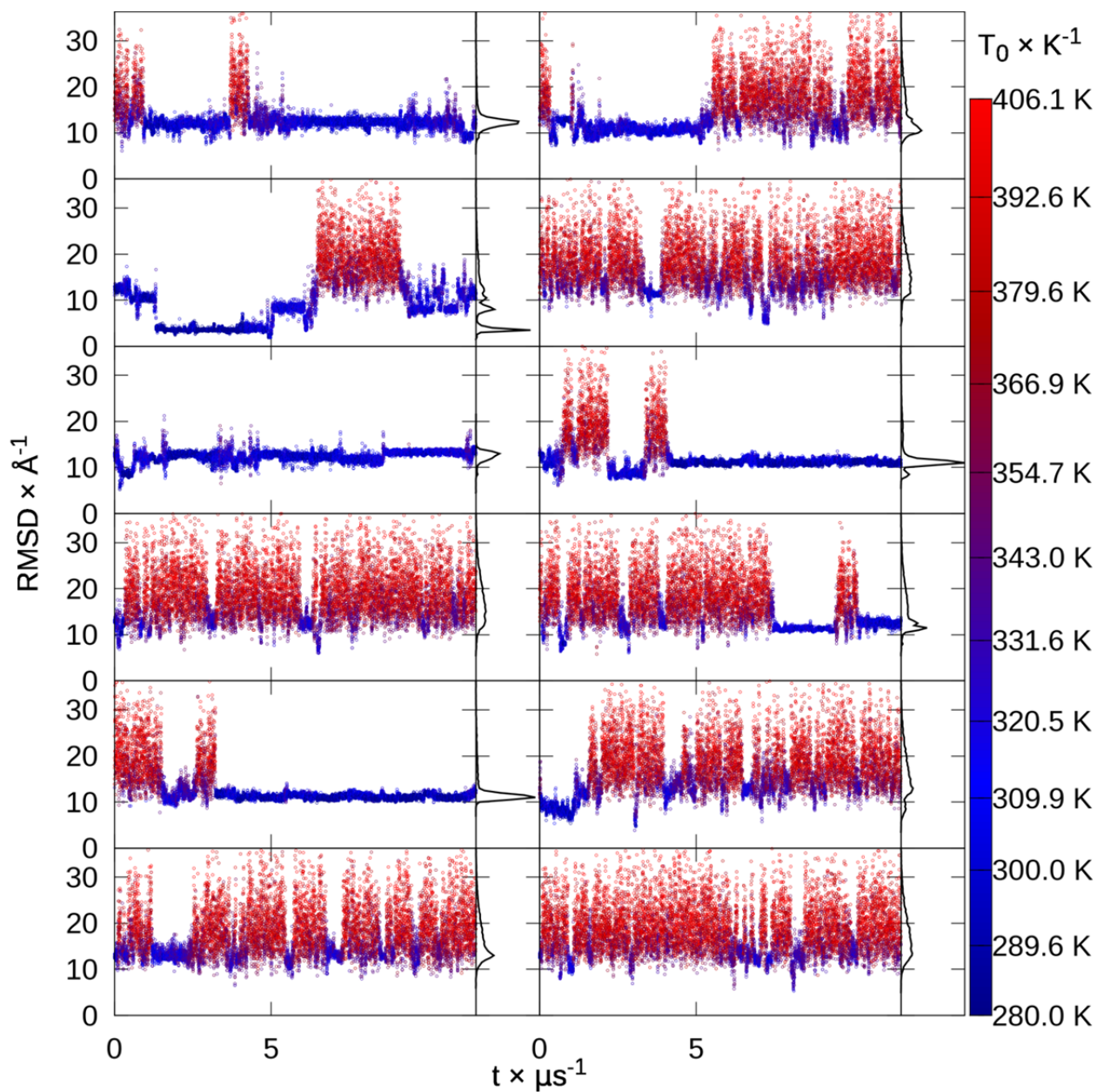


Figure S64. Hypothetical protein 1WHZ replica RMSDs. RMSD to native of each replica from replica exchange initiated with extended MD structures versus time, colored by snapshot temperature from blue to red, with histograms. This differs from the former hypothetical protein replica RMSDs by the starting structures of the REMD.

Cluster population (%)	43.9	18.1	9.9	5.1	3.3
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	11.0	3.4	12.5	12.7	11.8

Table S19. Hypothetical protein 1WHZ top 5 extended MD REMD cluster populations and centroid  $C\alpha$  RMSDs.

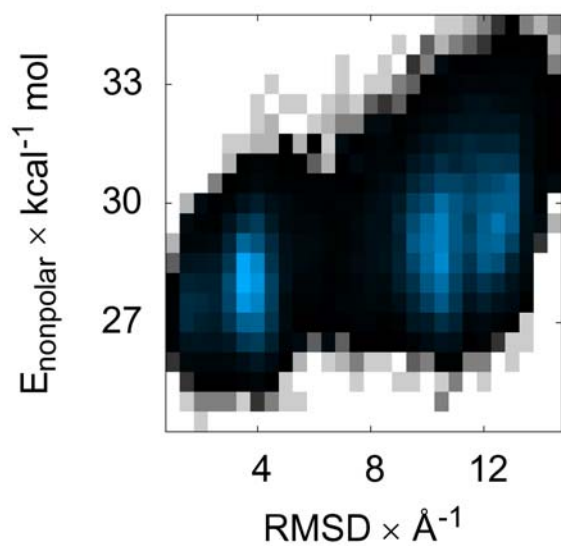


Figure S65. Hyp protein 1WHZ surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is slightly more favorable at low ( $1\text{--}3 \text{ \AA}$ ) than high ( $9\text{--}12 \text{ \AA}$ ) RMSDs.

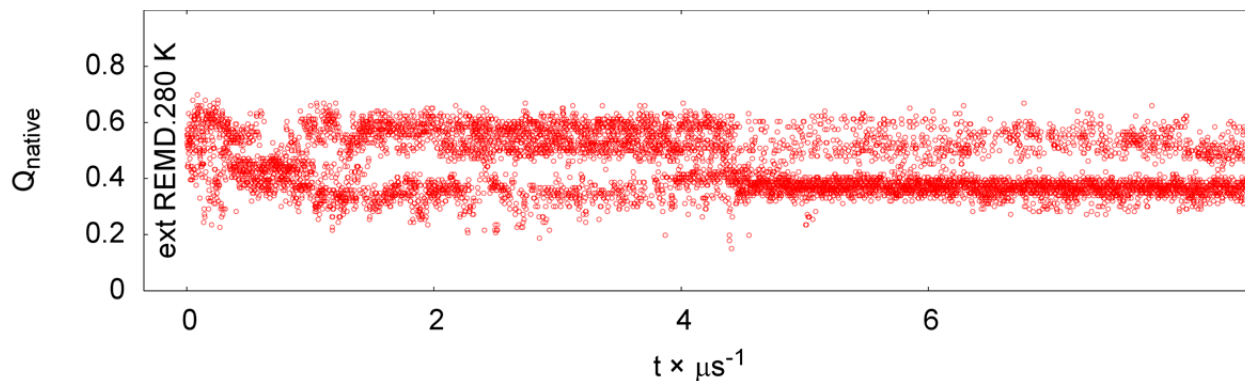


Figure S66. Hyp protein 1WHZ native contacts versus time.

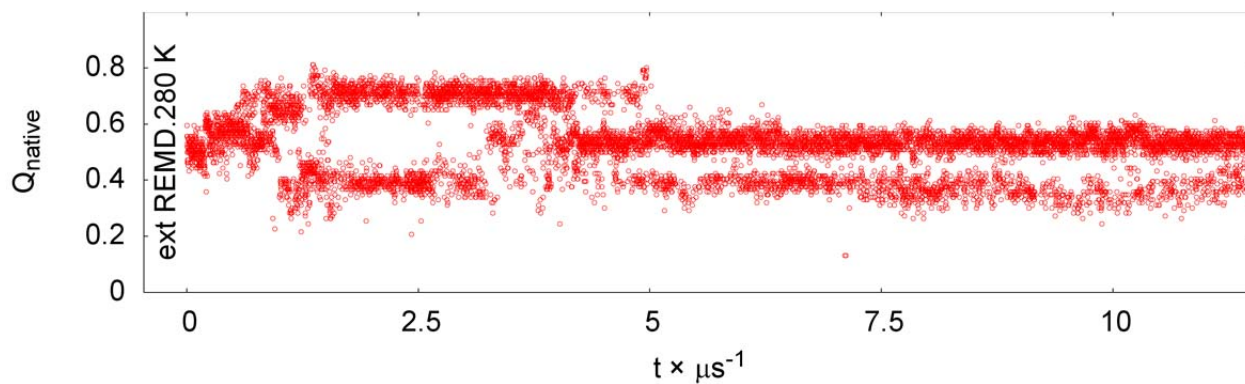


Figure S67. Hyp protein 1WHZ extended MD REMD native contacts versus time.

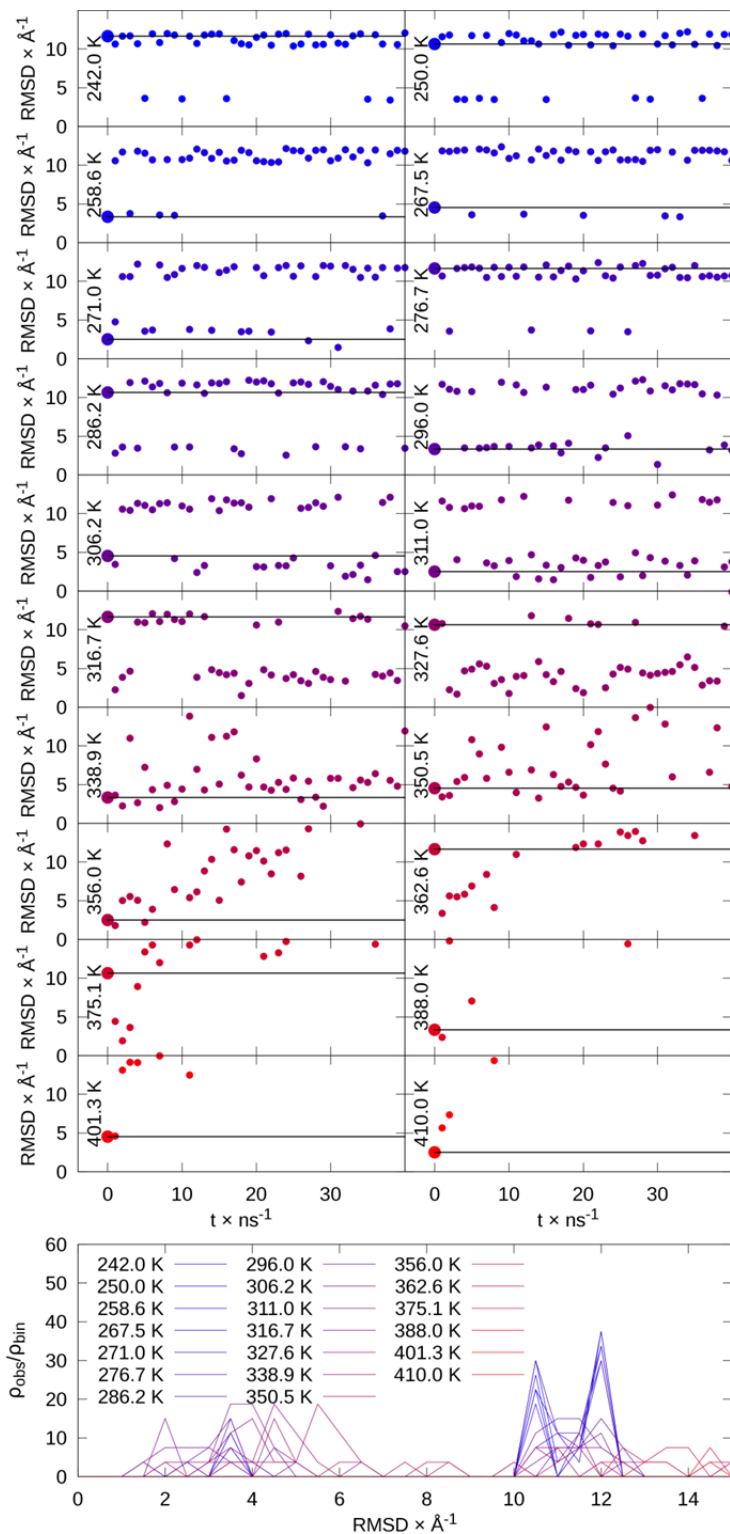


Figure S68. Seeded REMD sorting of hypothetical protein 1WHZ conformations: 2 unfolded (11.7  $\text{\AA}$ , 10.7  $\text{\AA}$ ), 2 partly folded (3.3  $\text{\AA}$ , 4.5  $\text{\AA}$ ), and 1 native-like (2.5  $\text{\AA}$ ), repeated for 20 replicas. At top, RMSD vs time for each temperature shows sorting of high RMSD conformations to low temperatures, followed by partly followed conformations. Lines indicate initial RMSD value that each temperature. At bottom, histogram shows preference of high and then intermediate RMSD conformations at low temperatures.

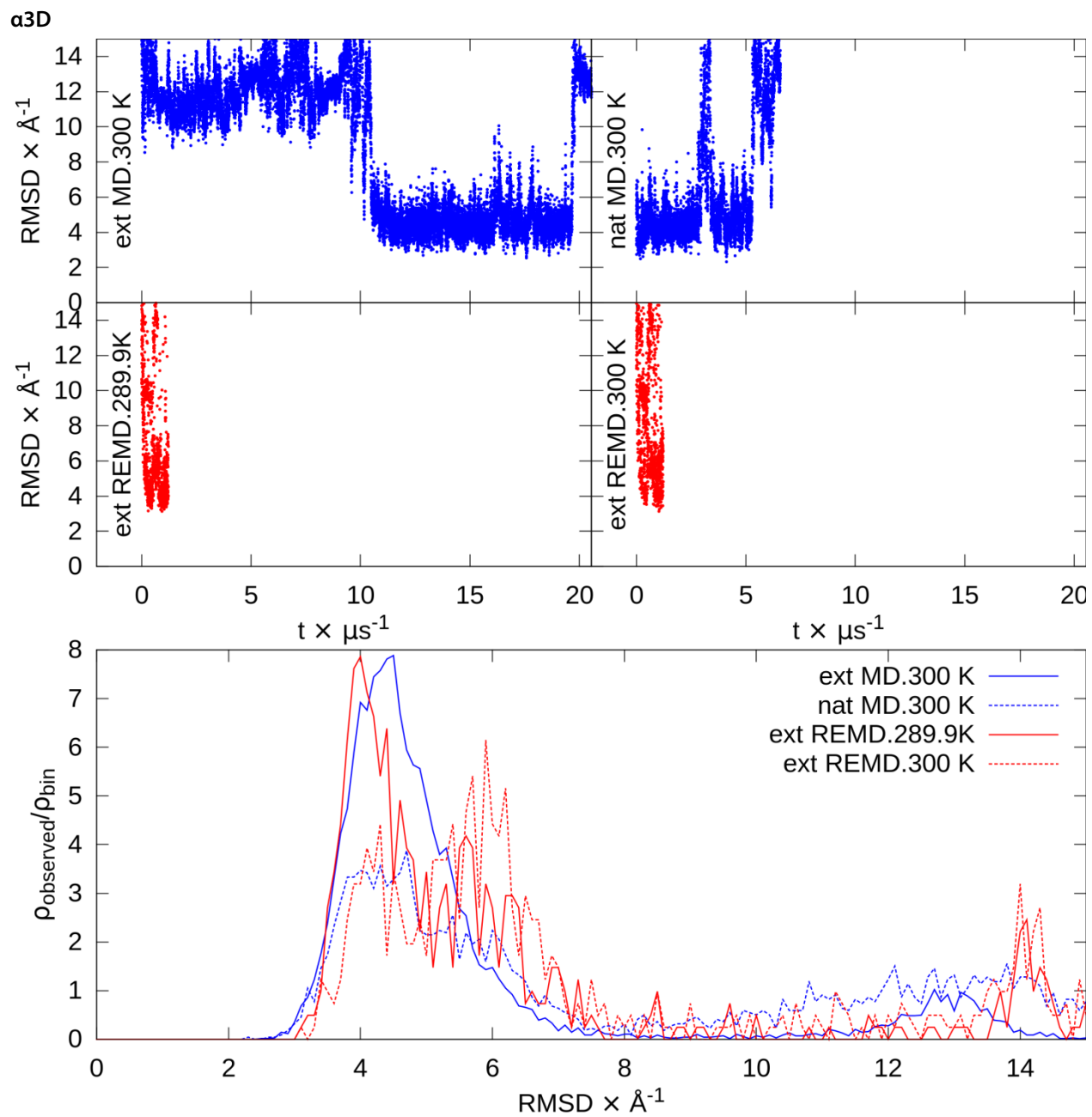


Figure S69.  $\alpha$ 3D RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

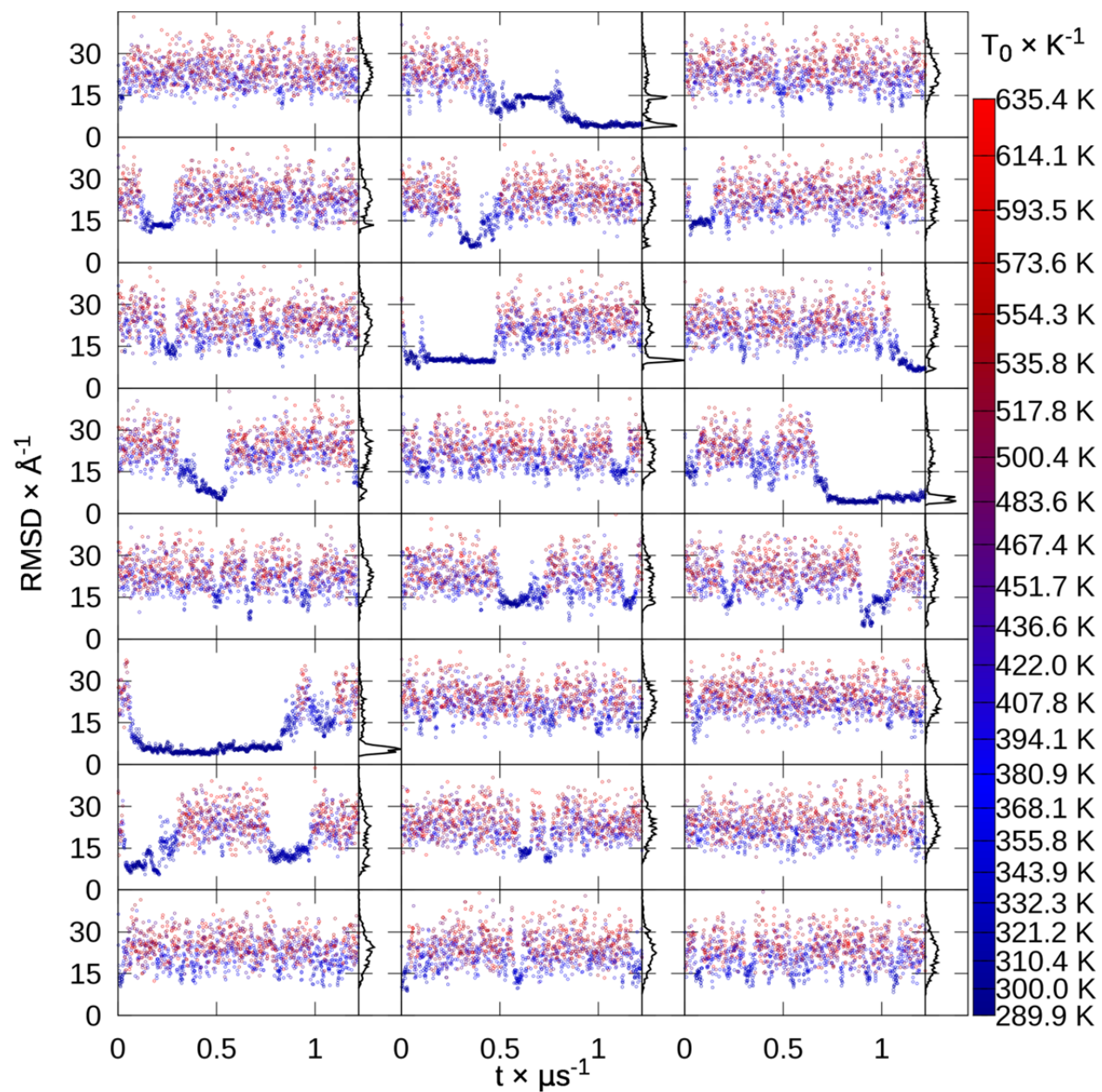


Figure S70.  $\alpha$ 3D replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	32.7	18.4	9.2	7.1	6.8
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	4.0	4.1	5.5	10.1	6.0

Table S20.  $\alpha$ 3D top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

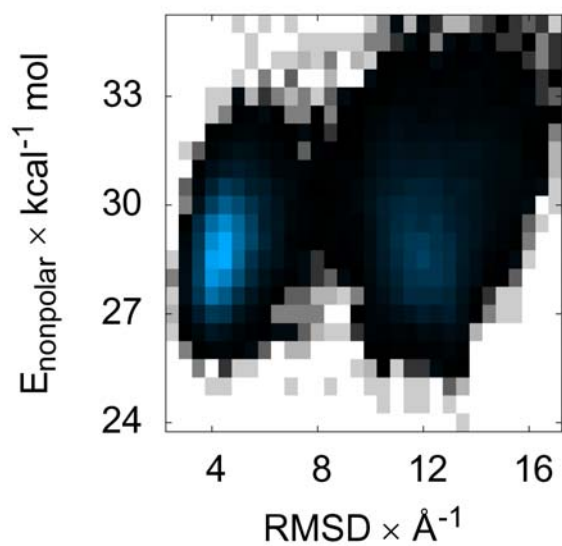


Figure S71.  $\alpha$ 3D surface area energy versus RMSD. Color indicates the histogrammed population in each 0.5 Å by 0.5 kcal mol<sup>-1</sup> bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is similarly to slightly more favorable at high (10–13 Å) than low (2–4 Å) RMSDs.

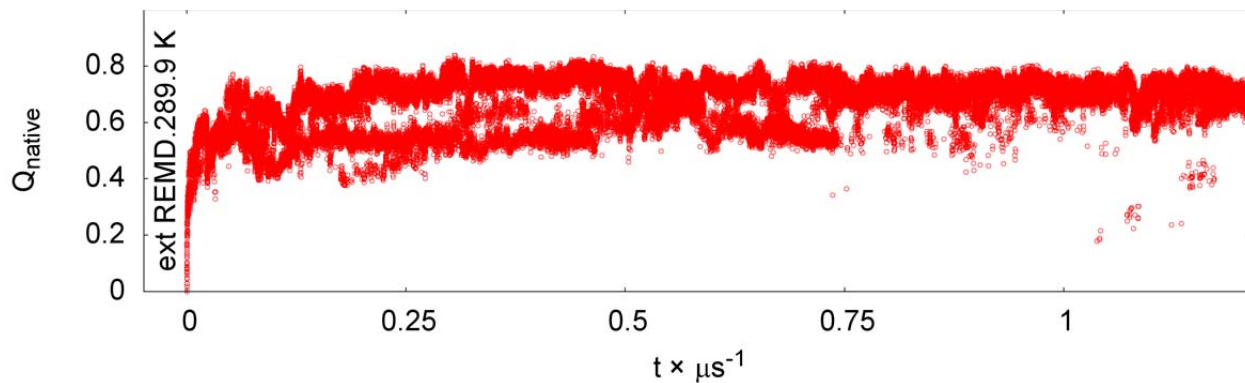


Figure S72.  $\alpha$ 3D native contacts versus time.

### $\lambda$ -repressor

The 80 amino acid  $\lambda$ -repressor shows transient folding to the native structure in REMD, but the majority of the population adopts a misfolded structure with RMSD of 12 Å (Figure S73). In this case, the 5  $\alpha$ -helices are largely present, but they pack against the first helix in a clockwise fashion, rather than counterclockwise as seen in the native fold (Figure S87). Using the coordinate-seeded REMD approach, we combined structures of the misfolded topology, the lowest RMSD from REMD, and native structures. Similar to 1WHZ, the results indicated our model prefers the structure with the helices formed but incorrectly arranged around the first helix (Figure S77).

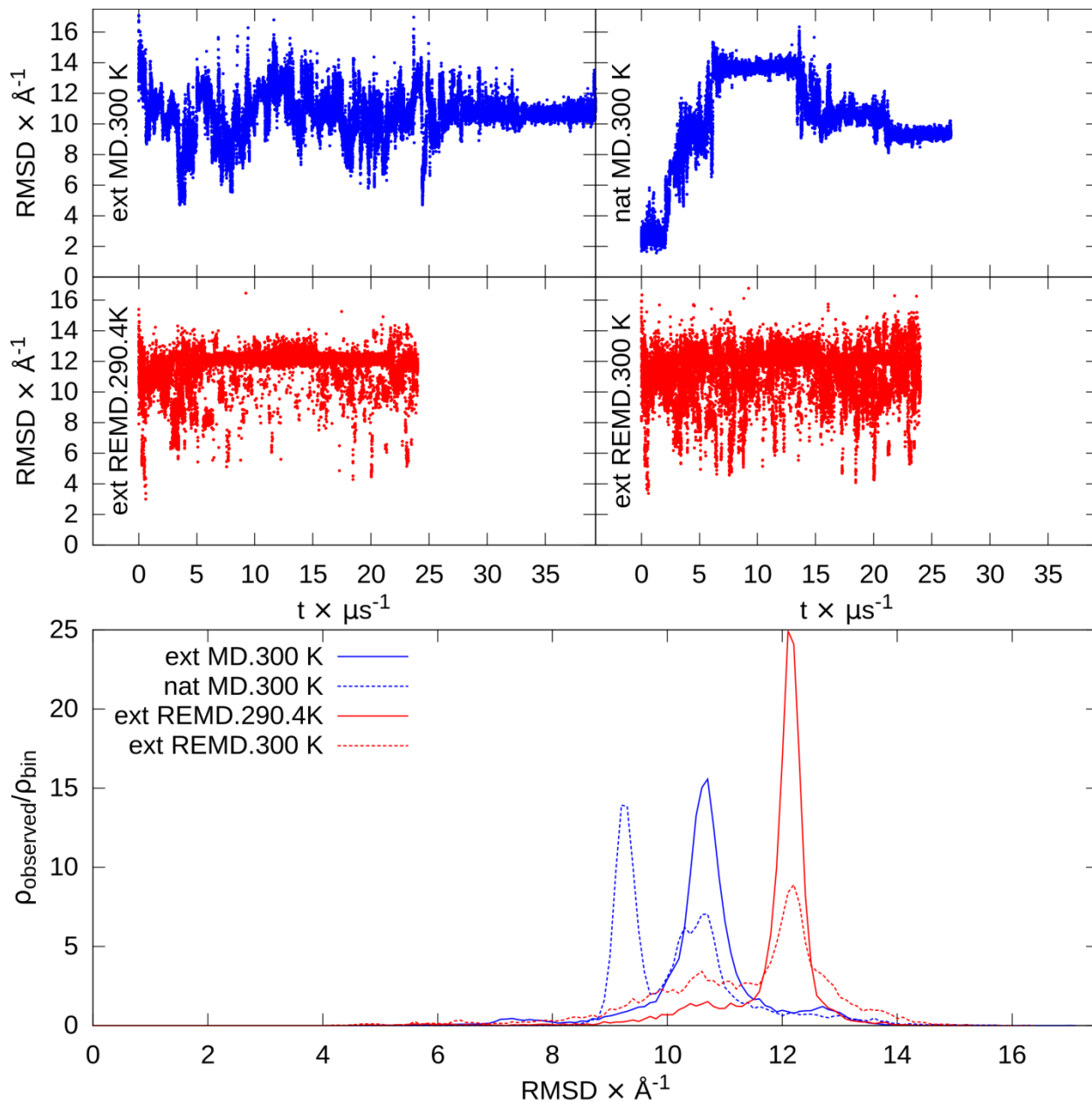


Figure S73.  $\lambda$ -repressor RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

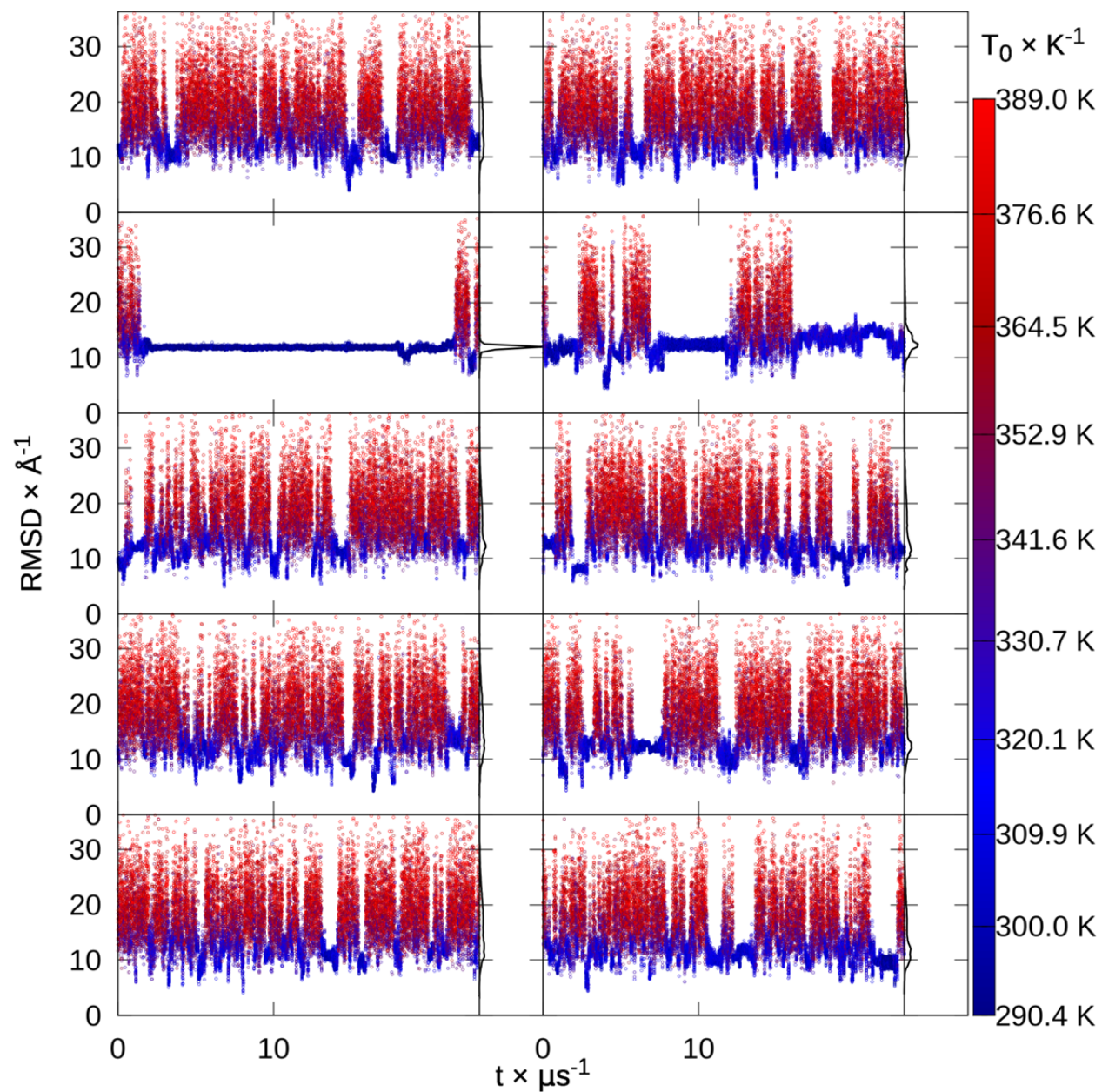


Figure S74.  $\lambda$ -repressor replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	53.9	4.4	3.8	3.4	3.3
Centroid $C\alpha$ RMSD ( $\text{\AA}$ )	11.9	10.9	11.2	12.1	9.8

Table S21.  $\lambda$ -repressor top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.



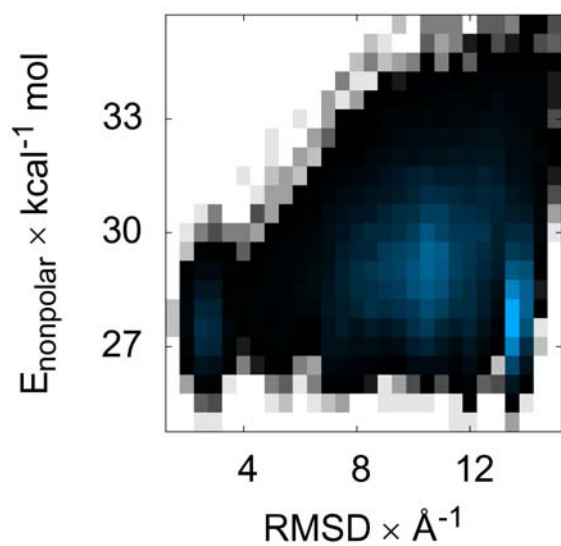


Figure S75.  $\lambda$ -repressor surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is flat across low ( $2\text{-}4 \text{ \AA}$ ) to high ( $12\text{-}14 \text{ \AA}$ ) RMSDs.

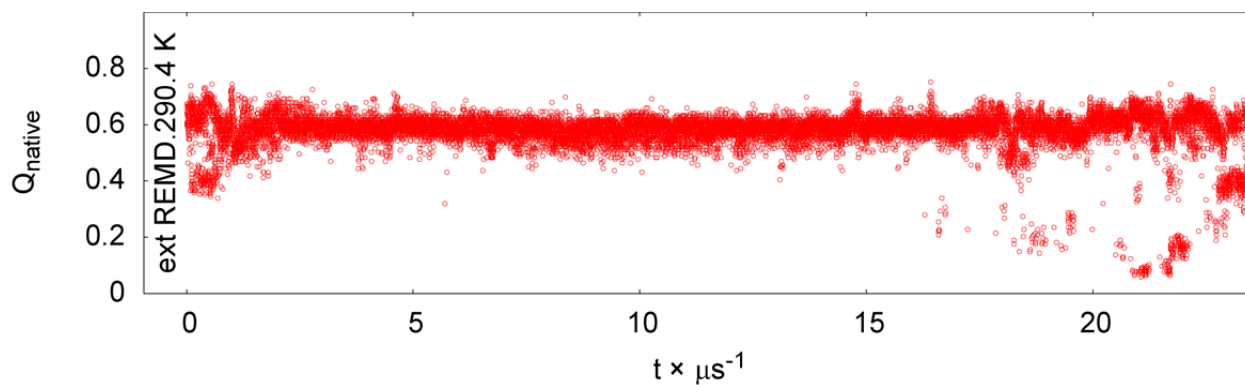


Figure S76.  $\lambda$ -repressor native contacts versus time.

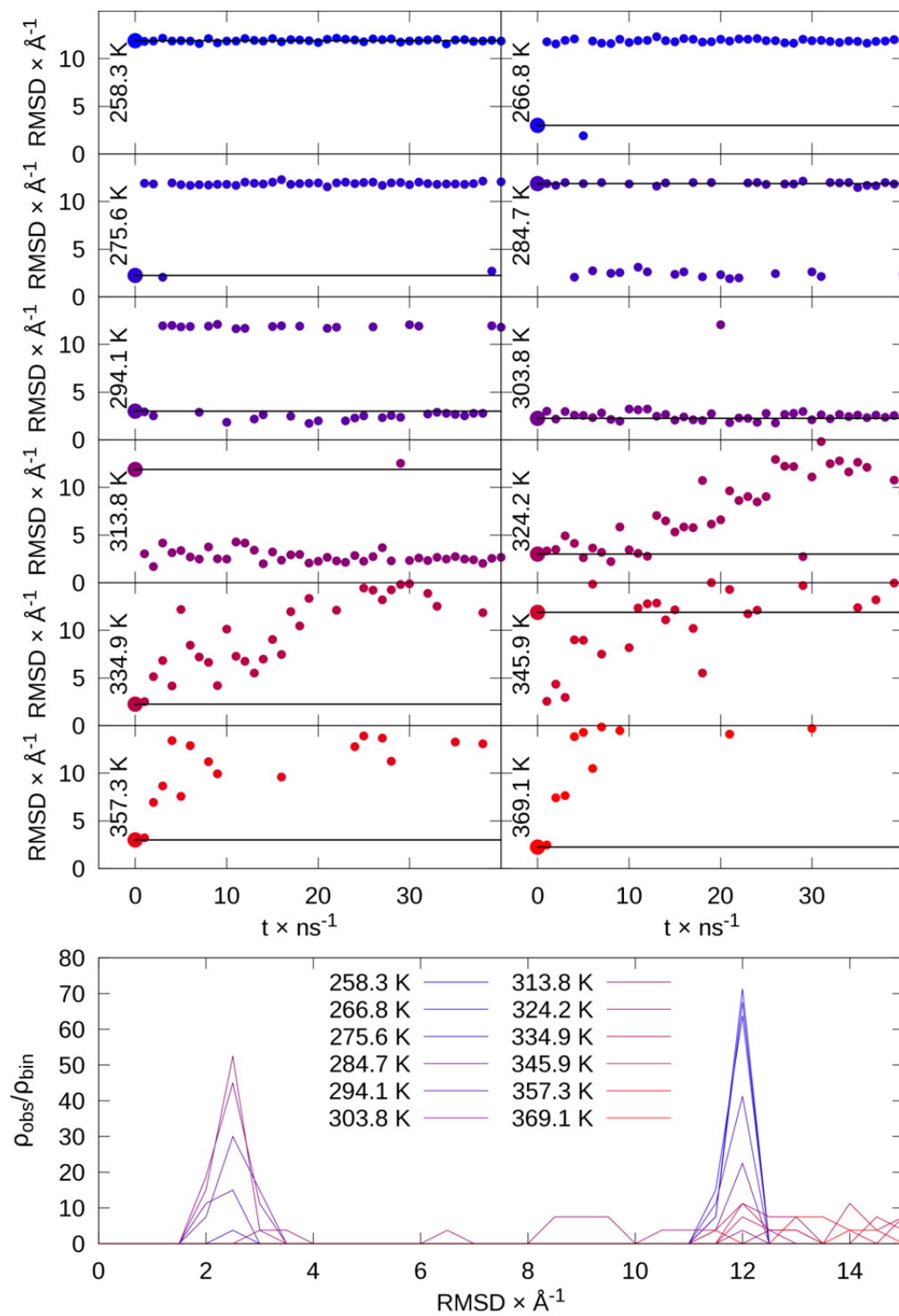


Figure S77. REMD sorting of  $\lambda$ -repressor conformations: unfolded (12.1  $\text{\AA}$ ), partly unfolded (3.0  $\text{\AA}$ ), and native-like (2.3  $\text{\AA}$ ), repeated for 12 replicas. At top, RMSD vs time for each temperature shows sorting of high RMSD conformations to low temperatures. Lines indicate initial RMSD value that each temperature. At bottom, histogram shows preference of high RMSD conformations at low temperatures.

## Top7

The largest system we studied is the designed protein Top7 (92 amino acids). Conformational sampling is observed to be very slow in this system, with the extended conformation and folded structure both stable for the entire  $\sim 5 \mu\text{sec}$  MD runs. Most replicas spend the majority of the simulation trapped in different local minima, suggesting that the data are poorly converged at 20  $\mu\text{sec}$  of REMD. The native topology for Top7 resembles 2 zinc finger domains with the  $\beta$ -hairpins connected through an additional  $\beta$ -strand, forming a 5-stranded sheet in the protein. The misfolded structure with highest population is only sampled by 1 replica. It shows correct placement of strands 3, 4 and 5, as well as the helix between strands 3 and 4, with an RMSD of 3.3  $\text{\AA}$  for the region 42-92.  $\beta$ -strand 1 is also folded, but the subsequent strand 2 and helix are not yet well formed. Seeded REMD combining the misfolded and correctly folded structures showed a strong preference for the correct fold, moving all misfolded structures to higher temperatures. Interestingly, two of the misfolded replicas refolded to the correct structure during this run.

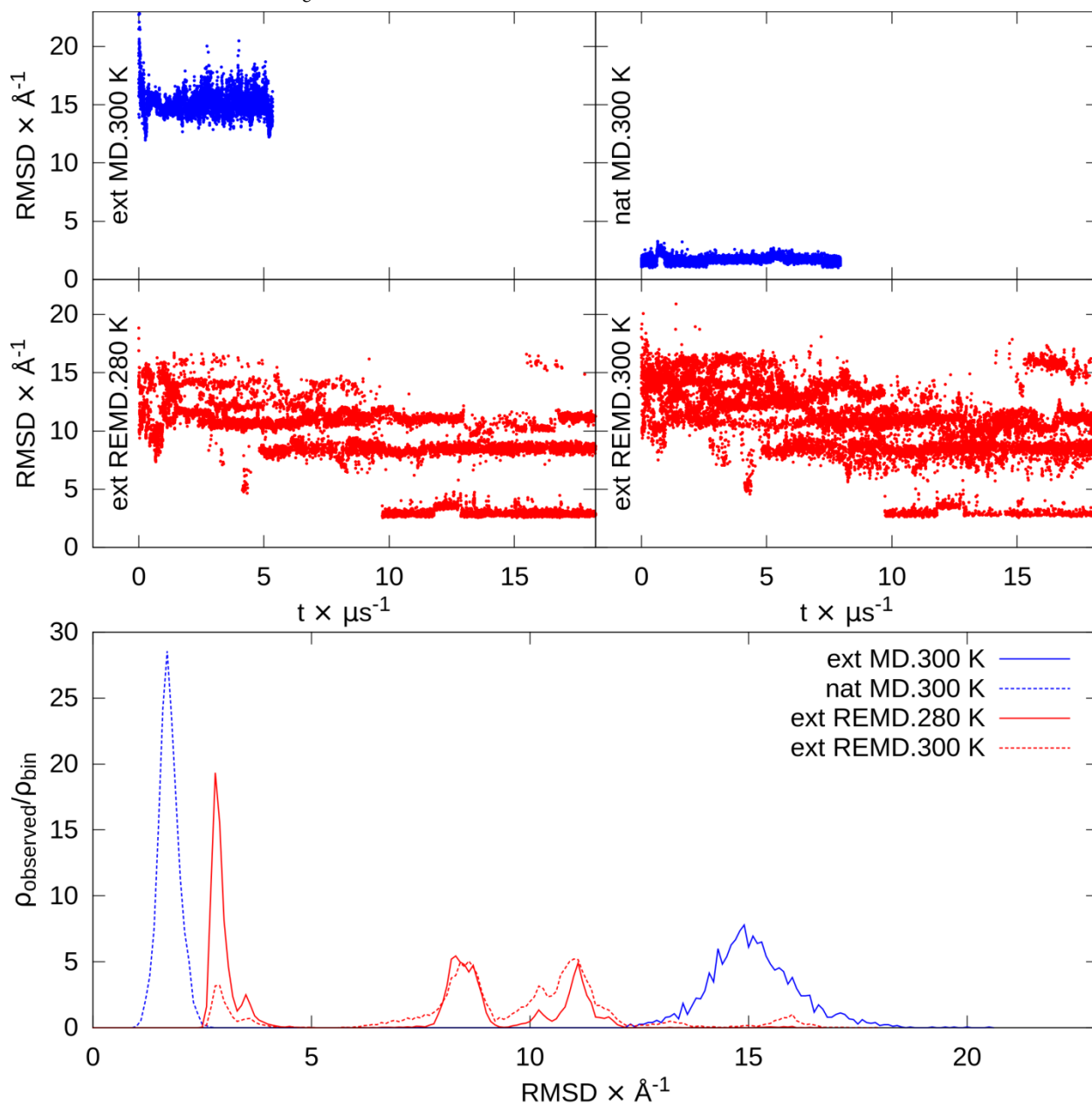


Figure S78. Top7 RMSDs. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

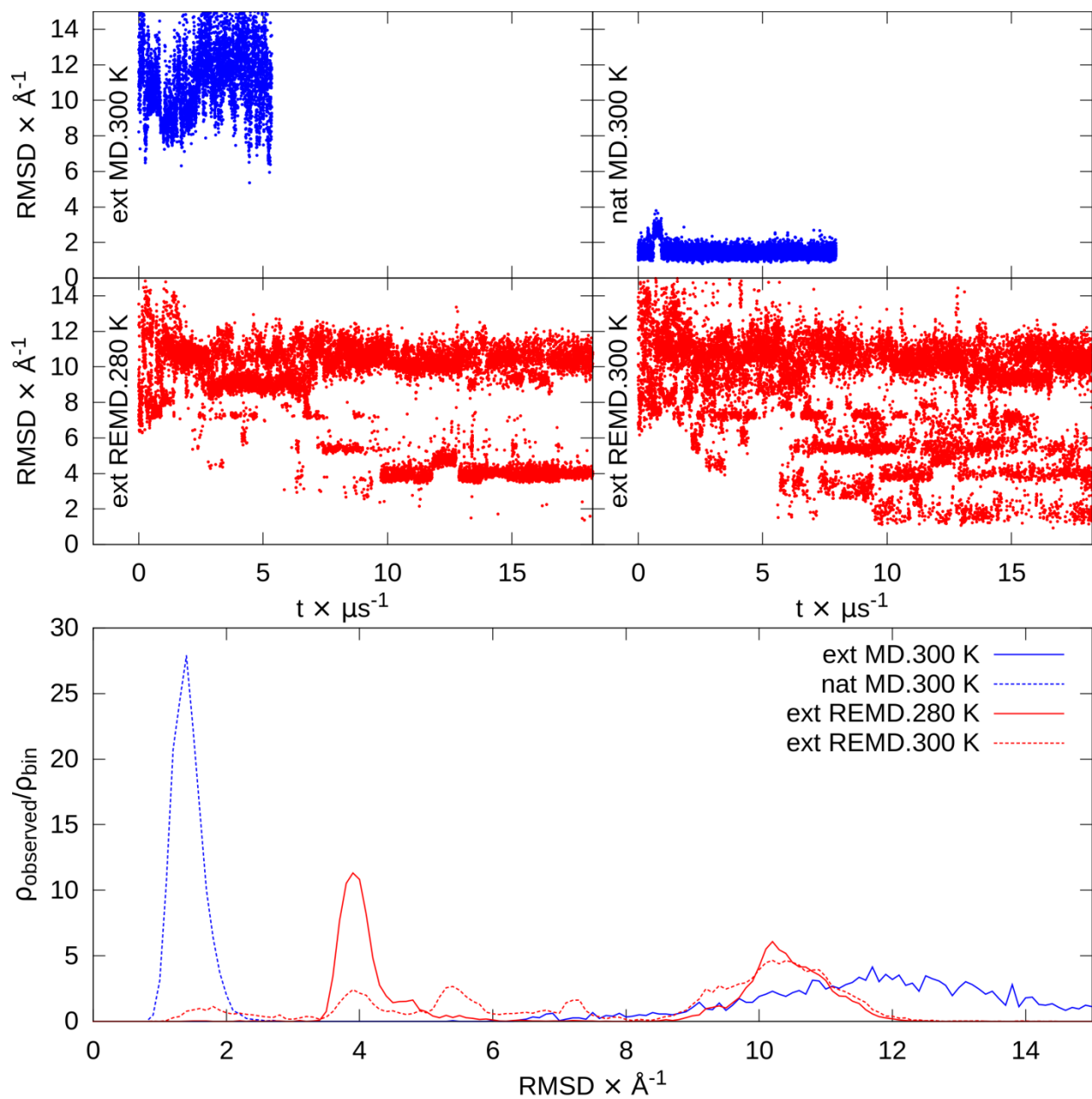


Figure S79. Top7 RMSDs, residues 1 to 40. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the binsize ( $\rho_{\text{bin}}$ ).

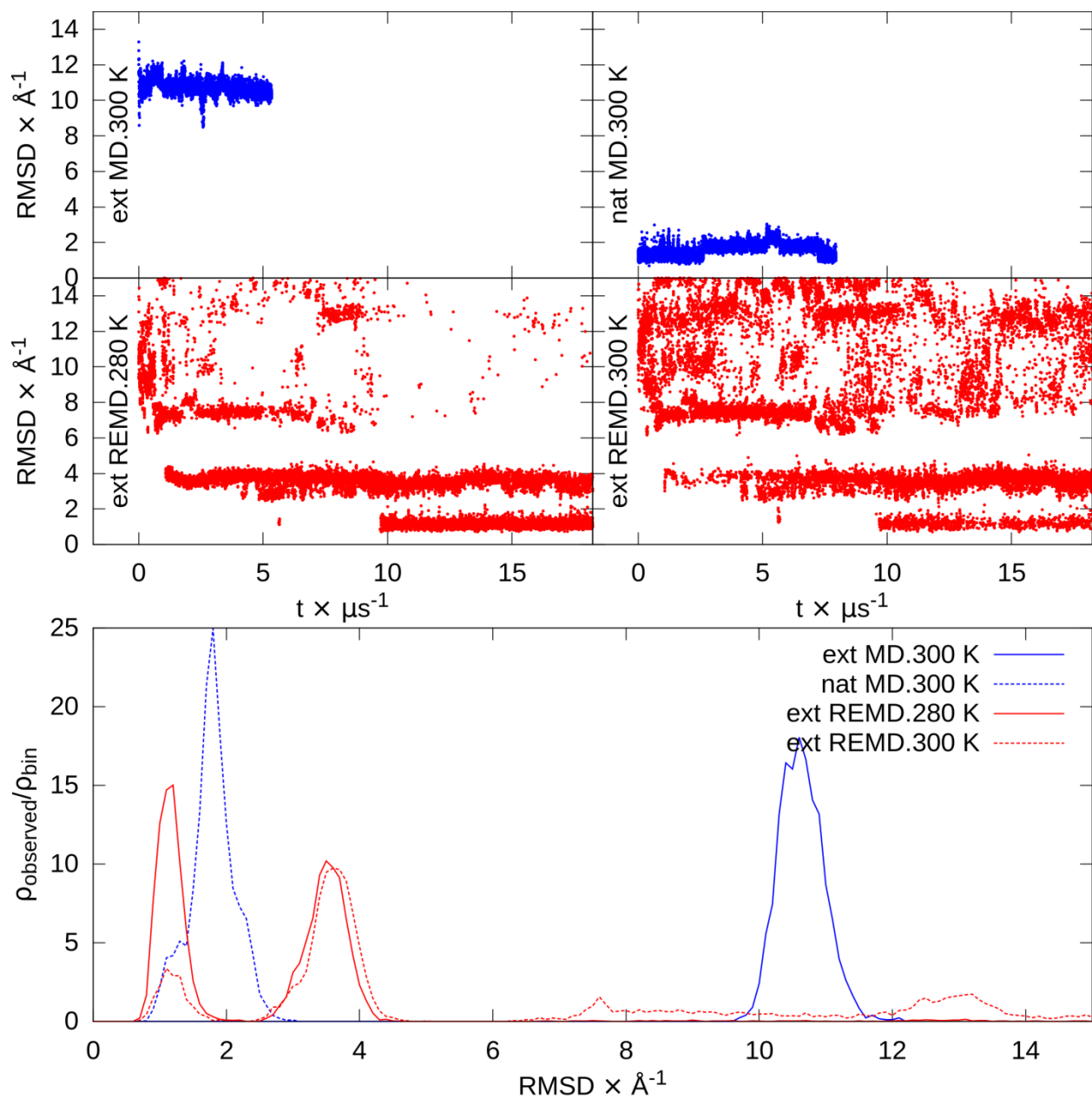


Figure S80. Top7 RMSDs, residues 42 to 92. At top are RMSD versus time for extended and native MD and the lowest temperatures from extended REMD. At bottom are RMSD histograms of the second half of each simulation.  $\rho_{\text{observed}}/\rho_{\text{bin}}$  represents the fraction observed ( $\rho_{\text{observed}}$ ) relative to the fraction of the bin size ( $\rho_{\text{bin}}$ ).

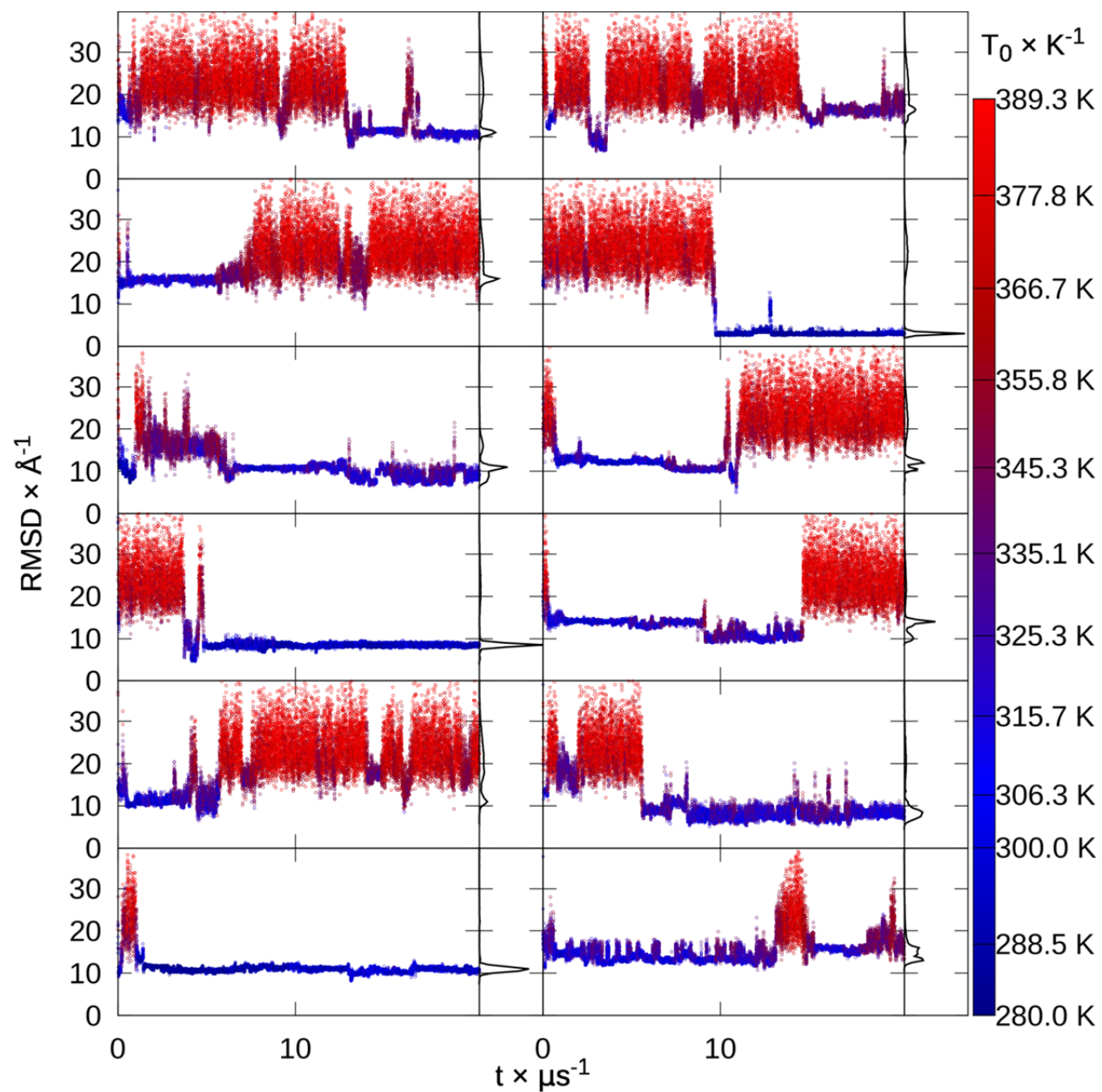


Figure S81. Top7 replica RMSDs. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

Cluster population (%)	35.9	24.1	19.0	3.2	2.2
Centroid $C\alpha$ RMSD (Å)	11.2	2.7	8.3	13.9	8.3

Table S22. Top7top 5 extended REMD cluster populations and centroid  $C\alpha$  RMSDs.

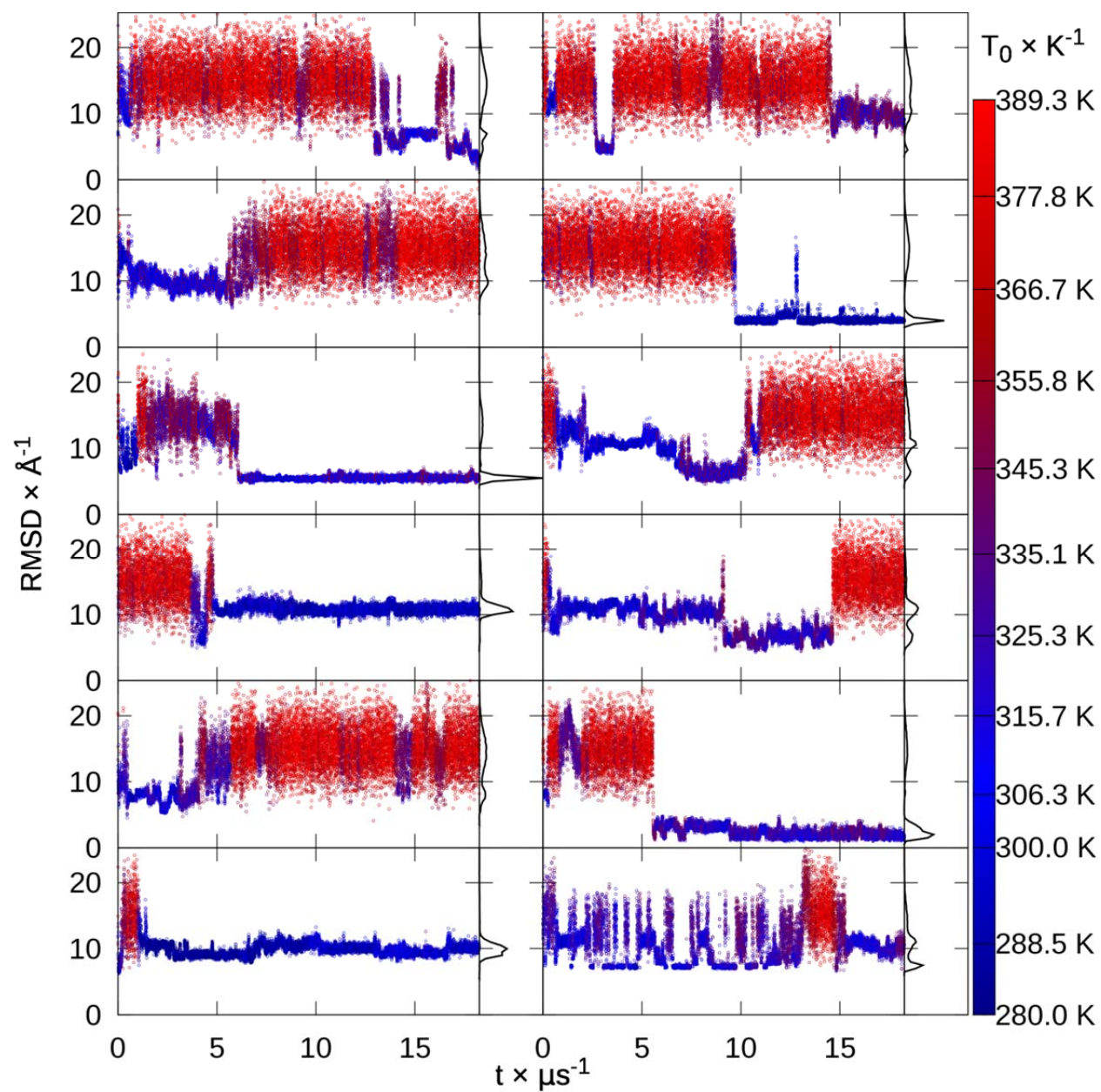


Figure S82. Top7 replica RMSDs, residues 1 to 40. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.

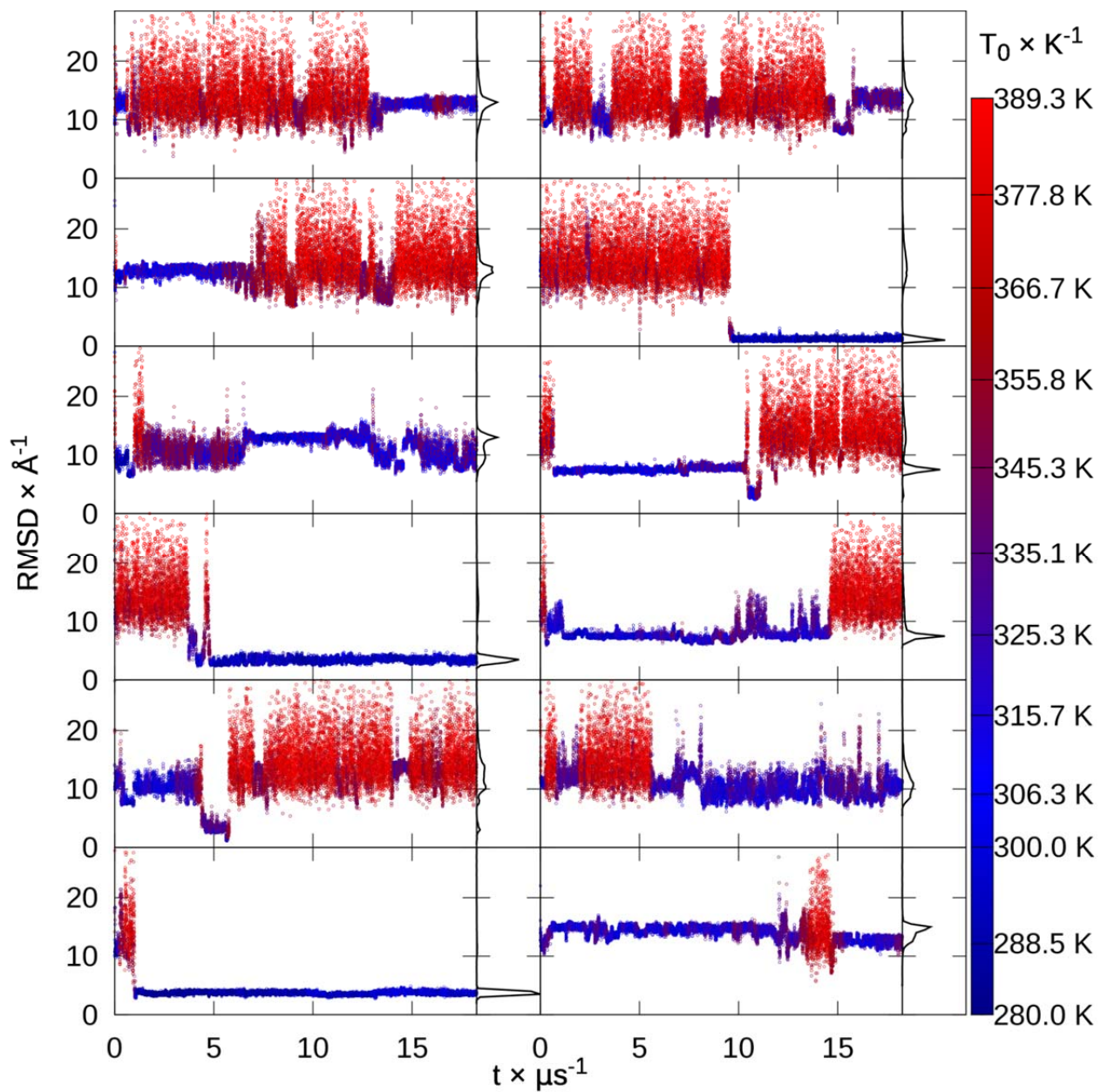


Figure S83. Top7 replica RMSDs, residues 42 to 92. RMSD to native of each replica from extended replica exchange versus time, colored by snapshot temperature from blue to red, with histograms.



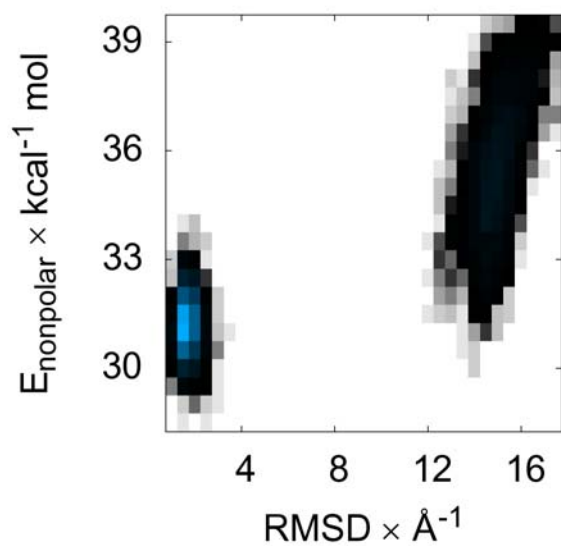


Figure S84. Top7 surface area energy versus RMSD. Color indicates the histogrammed population in each  $0.5 \text{ \AA}$  by  $0.5 \text{ kcal mol}^{-1}$  bin, going from white (no population) to black (1% of maximum bin population) and then to blue (maximum bin population). The correction for the solvent-accessible surface area, determined by recursively optimizing spheres around each atom starting from icosahedra, is more favorable at low ( $1\text{--}3 \text{ \AA}$ ) RMSD.

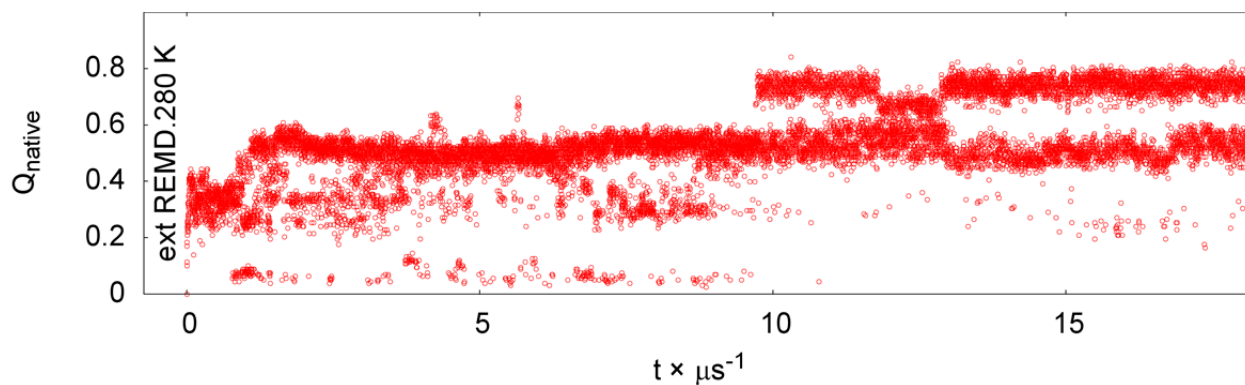


Figure S85. Top7 native contacts versus time.

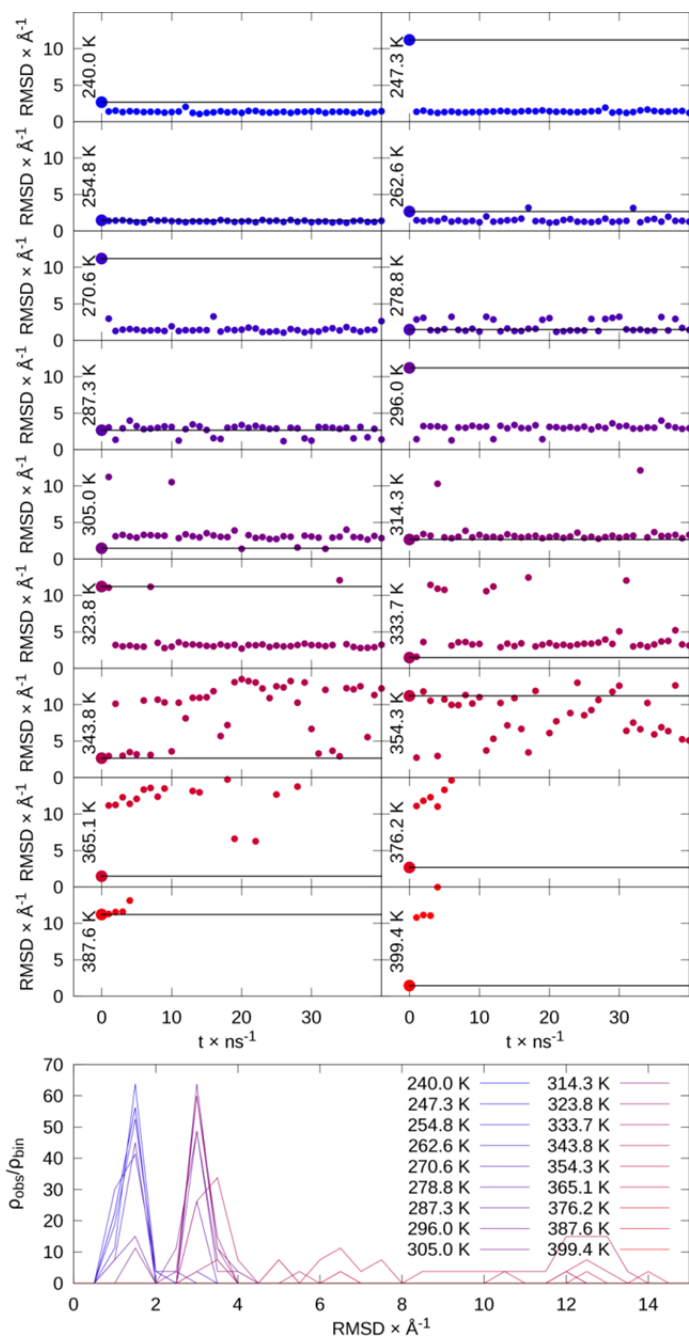
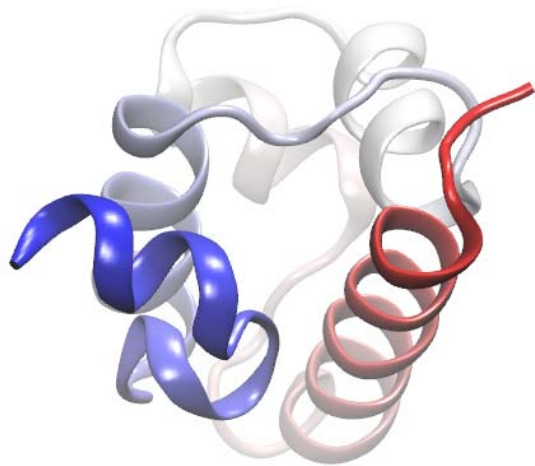
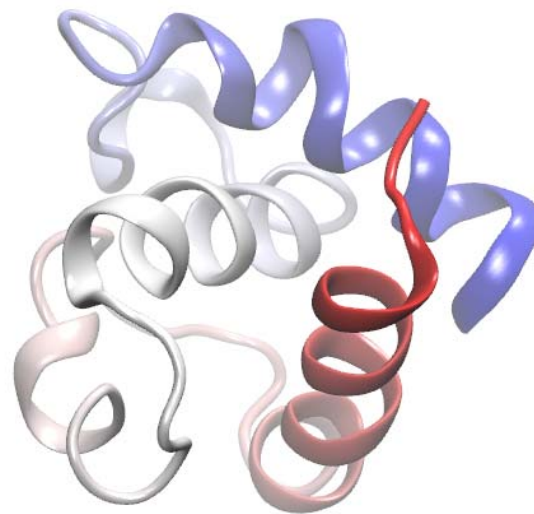


Figure S86. REMD sorting of Top7 conformations: partly folded (2.7 Å), unfolded (11.2 Å), and native-like (1.5 Å), repeated for 18 replicas. At top, RMSD vs time for each temperature shows sorting of native-like conformations to low temperatures, partly folded to intermediate temperatures, and unfolded to high temperatures. Lines indicate initial RMSD value at each temperature. At bottom, histogram shows preference of native-like conformations at low temperatures and partly folded conformations at intermediate temperatures.



**Experiment**



**Simulation preferred**

Figure S87.  $\lambda$ -repressor as in the crystal structure (left) and the largest cluster in REMD simulation (right). Structures were aligned to the first helix. In both structures, the helices are arranged in a circular fashion from N-terminal (red) to C-terminal (blue), but the simulation prefers a structure with the topology progressing in the opposite direction.

## References

- (1) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712.
- (2) Møller, C.; Plesset, M. S. *Physical Review* **1934**, *46*, 618.
- (3) Ditchfield, R.; Hehre, W. J.; Pople, J. A. *The Journal of Chemical Physics* **1971**, *54*, 724.
- (4) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *Journal of the American Chemical Society* **1995**, *117*, 5179.
- (5) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *Journal of Computational Chemistry* **1993**, *14*, 1347.
- (6) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.; Gaussian, Inc.: Pittsburgh, PA, USA, 1998.
- (7) Case, D. A., Cheatham, T. E., Darden, H., Gohlke, R., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. *Journal of Computational Chemistry* **2005**, *26*, 1668.
- (8) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; University of California, San Francisco: 2012.
- (9) Holland, J. H. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*; University of Michigan Press: Ann Arbor, 1975.
- (10) Wall, M.; Massachusetts Institute of Technology.
- (11) Gotz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542.
- (12) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvary, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. *University of California, San Francisco* **2014**.
- (13) Nguyen, H.; Roe, D. R.; Simmerling, C. *Journal of Chemical Theory and Computation* **2013**, *9*, 2020.
- (14) Chen, J.; Brooks, C. L. *Phys. Chem. Chem. Phys.* **2008**, *10*, 471.
- (15) Levy, R. M.; Zhang, L. Y.; Gallicchio, E.; Felts, A. K. *Journal of the American Chemical Society* **2003**, *125*, 9523.
- (16) Wagoner, J. A.; Baker, N. A. *Proceedings of the National Academy of Sciences* **2006**, *103*, 8331.
- (17) D.A. Case, T. A. D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A.W. Goetz, I. Kolossvary, K.F. Wong, F. Paesani, J. Vanicek, R.M. Wolf, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman *University of California, San Francisco* **(2014)**.
- (18) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *Journal of Computational Chemistry* **1999**, *20*, 786.
- (19) Hopkins, C.; Roitberg, A. **unpublished data**.
- (20) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (21) Roe, D. R.; Cheatham, T. E. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084.
- (22) Roe, D. R.; Cheatham, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3084.
- (23) Gohlke, H.; Kiel, C.; Case, D. A. *Journal of Molecular Biology* **2003**, *330*, 891.
- (24) Lipari, G.; Szabo, A. *Biochemistry* **1981**, *20*, 6250.
- (25) Prompers, J. J.; Brüschweiler, R. *Journal of the American Chemical Society* **2002**, *124*, 4522.
- (26) Li, D.-W.; Brüschweiler, R. *Journal of Chemical Theory and Computation* **2011**, *7*, 1773.
- (27) Feng, W.; Tejero, R.; Zimmerman, D. E.; Inouye, M.; Montelione, G. T. *Biochemistry* **1998**, *37*, 10881.
- (28) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Annual Review of Biophysics and Biomolecular Structure* **2005**, *34*, 43.
- (29) Yun-yu, S.; Lu, W.; Van Gunsteren, W. F. *Molecular Simulation* **1988**, *1*, 369.

- (30) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (31) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089.
- (32) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *Journal of Chemical Physics* **1984**, *81*, 3684.
- (33) Ciccotti, G.; Ryckaert, J. P. *Computer Physics Reports* **1986**, *4*, 346.
- (34) Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K. *J. Am. Chem. Soc.* **2008**, *130*, 15327.
- (35) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Mol. Biol.* **2002**, *9*, 425.
- (36) Sarisky, C. A.; Mayo, S. L. *Journal of Molecular Biology* **2001**, *307*, 1411.
- (37) Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008**, *94*, L75.
- (38) Liu, F.; Du, D.; Fuller, A. A.; Davoren, J. E.; Wipf, P.; Kelly, J. W.; Gruebele, M. *Proceedings of the National Academy of Sciences* **2008**, *105*, 2369.
- (39) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. *Journal of Molecular Biology* **2011**, *405*, 43.
- (40) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. *Science* **2011**, *334*, 517.
- (41) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Mol. Biol.* **1997**, *4*, 180.
- (42) Gronwald, W.; Hohm, T.; Hoffmann, D. *BMC Bioinformatics* **2008**, *9*, 109.
- (43) Hornig, J.-C.; Moroz, V.; Raleigh, D. P. *Journal of Molecular Biology* **2003**, *326*, 1261.
- (44) Neuweiler, H.; Sharpe, T. D.; Rutherford, T. J.; Johnson, C. M.; Allen, M. D.; Ferguson, N.; Fersht, A. R. *Journal of Molecular Biology* **2009**, *390*, 1060.
- (45) Shah, P. S.; Hom, G. K.; Ross, S. A.; Lassila, J. K.; Crowhurst, K. A.; Mayo, S. L. *J.Mol.Biol.* **2007**, *372*, 1.
- (46) Nauli, S.; Kuhlman, B.; Le Trong, I.; Stenkamp, R. E.; Teller, D.; Baker, D. *Protein Science* **2002**, *11*, 2924.
- (47) Schindelin, H.; Jiang, W.; Inouye, M.; Heinemann, U. *Proc.Natl.Acad.Sci.USA* **1994**, *91*, 5119.
- (48) Kanagawa, M.; Yokoyama, S.; Kuramitsu, S.
- (49) Walsh, S. T. R.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 5486.
- (50) Beamer, L. J.; Pabo, C. O. *J.Mol.Biol.* **1992**, *227*, 177.
- (51) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364.
- (52) Buck, M.; Boyd, J.; Redfield, C.; MacKenzie, D. A.; Jeenes, D. J.; Archer, D. B.; Dobson, C. M. *Biochemistry* **1995**, *34*, 4041.
- (53) Davis, C. M.; Xiao, S.; Raleigh, D. P.; Dyer, R. B. *Journal of the American Chemical Society* **2012**, *134*, 14476.
- (54) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *Journal of the American Chemical Society* **2002**, *124*, 12952.
- (55) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10648.
- (56) Wang, M.; Tang, Y.; Sato, S.; Vugmeyster, L.; McKnight, C. J.; Raleigh, D. P. *Journal of the American Chemical Society* **2003**, *125*, 6032.
- (57) Hornig, J.-C.; Moroz, V.; Raleigh, D. P. *J. Mol. Biol.* **2003**, *326*, 1261.
- (58) Neuweiler, H.; Sharpe, T. D.; Rutherford, T. J.; Johnson, C. M.; Allen, M. D.; Ferguson, N.; Fersht, A. R. *J.Mol.Biol.* **2009**, *390*, 1060.
- (59) Johansson, M. U.; de Château, M.; Wikström, M.; Forsén, S.; Drakenberg, T.; Björck, L. *J.Mol.Biol.* **1997**, *266*, 859.
- (60) Wang, T.; Zhu, Y.; Gai, F. *The Journal of Physical Chemistry B* **2004**, *108*, 3694.
- (61) Gillespie, B.; Vu, D. M.; Shah, P. S.; Marshall, S. A.; Dyer, R. B.; Mayo, S. L.; Plaxco, K. W. *Journal of Molecular Biology* **2003**, *330*, 813.
- (62) Nauli, S.; Kuhlman, B.; Baker, D. *Nature Structural & Molecular Biology* **2001**, *8*, 602.
- (63) Reid, K. L.; Rodriguez, H. M.; Hillier, B. J.; Gregoret, L. M. *Protein Science* **1998**, *7*, 470.
- (64) Zhu, Y.; Alonso, D. O. V.; Maki, K.; Huang, C.-Y.; Lahr, S. J.; Daggett, V.; Roder, H.; DeGrado, W. F.; Gai, F. *Proceedings of the National Academy of Sciences* **2003**, *100*, 15486.
- (65) Yang, W. Y.; Gruebele, M. *Nature* **2003**, *423*, 193.
- (66) Scalley-Kim, M.; Baker, D. *Journal of Molecular Biology* **2004**, *338*, 573.
- (67) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Research* **2000**, *28*, 235.