Thermodynamic Insights of Base Flipping in TNA Duplex: Force Fields, Salt Concentrations, and Free Energy Simulation Methods

Zhaoxi Sun[1*], and John Z.H. Zhang[1,2,3*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

[3]*Department of Chemistry, New York University, NY, NY 10003, USA*

*To whom correspondence should be addressed: proszx@163.com, john.zhang@nyu.edu

Abstract

Threofuranosyl nucleic acid (TNA) is an analogue of DNA. Its inter-nucleotide linkages are shifted from the wild-type 5'-to-3' one to the 3'-to-2' one. As a result, the number of covalent bonds between consecutive phosphates is reduced from 6 to 5. This leads to higher chemical stability, less reactive groups, and lower conformational flexibility. Experimental observations indicate that the interaction network is perturbed at the minimal level and the thermodynamic stability of the duplex is unaltered upon the TNA mutation. Whether computational modelling could reproduce this result will be studied in the base flipping of the middle T (DNA) residue or its T-to-TFT mutation (TNA). We applied the equilibrium free energy simulation and the nonequilibrium stratification method proposed previously in the base flipping case, proving the applicability of alternative free energy simulation protocols. As the force field is the main accuracy-limiting factor when converged phase space sampling is obtained, we benchmarked three popular AMBER force fields for nucleotides. The last-generation force fields include bsc1 and OL15, both of which perform similarly in reproducing the structures near the crystal conformation in previous benchmark studies. Our results indicate that all these three force fields provide similar descriptions of the base-paired state. However, with free energy simulation constructing the free energy profiles along the conformational change pathway, high-energy regions are explored and these three force fields behave differently. The bsc1 force field is found to perform best in reproducing the similarity of stabilities of DNA and TNA duplexes. The free energy barrier of base flipping under the OL15 force field is lowered modestly in TNA, and thus this force field is also usable. However, the bsc0 force field provides wrong results. The TNA duplex is significantly less stable than the

DNA duplex. Therefore, the bsc0 force field is not recommended in any application in modern nucleotide simulations. The salt concentration in nucleotide simulations is another factor influencing the thermodynamics of the system. Previous reports conclude that the net-neutral and excess-salt simulations provide similar results. However, the simulation method limits the phase space region explored in previous computational modelling. Our free energy simulation explores high-energy regions, where the excess salt does affect the thermodynamic stability. The free energy barrier along the base flipping pathway is generally elevated upon the addition of excess salts, but the relative height of the free energy barriers in DNA and TNA duplexes is not significantly changed. This phenomenon emphasizes the importance of adding sufficient salts to reproduce the experimental condition.

Introduction

The capacity of the Watson-Crick (WC) base pair is observed in various alternative sugars (e.g. hexopyranoses, pentopyranoses and tetrofuranoses).[1-8] One of the simplest analogues of DNA is threofuranosyl nucleic acid (TNA), the inter-nucleotide linkages of which are shifted from the wild-type 5'-to-3' one to the 3'-to-2' one.[9-10] TNA is able to form stable duplexes with DNA, RNA and TNA,[4, 7, 11-16] which enables it to transfer information to DNA or RNA. The enzymatic and non-enzymatic polymerization of the TNA nucleotide is also made possible due to this cross-pairing.[17-23] TNA is often assumed to be a good predecessor of RNA.[24-25] The chemical stability of TNA is higher and the number of reactive groups is smaller, which leads to fewer side reactions and thus more faithful copying. The regioselectivity is no longer a problem due to the presence of only two hydroxyl groups.[24] A direct consequence of the change of the linkage in TNA is the alternation of the number of covalent bonds between consecutive phosphates, i.e. from 6 in DNA to 5 in TNA. The conformational flexibility is reduced, making TNA probably more suitable for information storage than DNA. However, the thermodynamic stability of the duplex and the interaction network are often unaltered upon the TNA mutation.[2] For instance, the stacking interactions are virtually unchanged in TNA-modified Dickerson-Drew dodecamer (DDD). The O4' atoms in the tetrose sugars also share very similar interaction network with those of the deoxyribose O4' atoms in the DDD duplex.[26]

In recent years, the importance of DNA systems are becoming widely recognized and more and more scientists start to work on nucleotide systems.[27-31] The unique functional role of nucleotides makes them one of the key targets in understanding biological processes. The genetic code deposited in the molecule hides inside the duplex and triplex structures and is inaccessible to other biomolecules. When the duplex is activated due to the change in the surrounding conditions or the intrinsic fluctuation of the duplex and the base flips outward to be exposed to solvent,[32-40] the bases can interact with enzymes and triggers some biological processes, such as modifications of the nucleotide sequence.[28, 41] The computer simulation is a powerful tool to study the dynamics of DNA-related molecules at atomic details and numerous investigations of the base flipping process in various systems are reported.[37, 42-48] For example, the computer simulation shows that the conformational ensemble in DNA systems depends on the sequence of the molecule, and the BI<->BII transition is found to depends on the ion included in the simulation.[49] The excitation of DNA molecules is also studied computationally.[50-51] Technically, molecular dynamics (MD) simulations are normally unable to model base flipping due to the huge difference between the timescale of the base flipping event and the time step for integrating the equations of motion.[52-62] Specifically, the (free) energy penalty of flipping a base is about 10 kcal/mol.[63-68] As a result, the timescale of base flipping is about ms.[69] By contrast, to ensure numerical stability

of integrating the equations of motion, the time step in MD simulations cannot exceed several fs. Such a huge gap between the timescale accessible in MD simulations and that of base flipping makes it hard to get converged statistics in computer simulations. To get converged and statistically meaningful data, the equilibrium enhanced sampling techniques such as umbrella sampling[70-72] and replica exchange methods[73-78] are often employed. For instance, the umbrella sampling is applied to study the protonation-dependent behavior of base flipping in RNA systems, where the protonation-dependent base flipping is found to be coupled with a syn-to-anti transformation of a guanine (G) group in the G-adenine (A) mismatch.[65] The polarization-induced change of the height of the free energy profile along the base flipping pathway is also studied with umbrella sampling.[63] Although the nonequilibrium technique of steered MD (SMD)[79-83] is less frequently employed to investigate the thermodynamics in base flipping, they do have significant potentials and show similar performance in the construction of the free energy landscape of base flipping, compared with equilibrium enhanced sampling techniques.[64, 84] Successful applications reported in recent years are, for example, the variation of base flipping free energy landscapes upon the sulfur substitution at G groups in G-cytosine (C) base pairs and G-thymine (T) mismatches.[84] However, the existing publications often focus on the mutation-dependent behaviors of the nucleotide systems, e.g. synthetic nucleobase pairs and naturally occurring mutations.[65, 84-93] The behavior of the duplex could also be influenced by the alteration in the sugar moiety, which is less frequently studied. Therefore, we select a TNA case here to investigate the effect of the alteration in the sugar moiety on the thermodynamics of the duplex. The thermodynamic profiles along the base flipping pathway in wild-type DNA and the mutant TNA duplexes are constructed from equilibrium and nonequilibrium free energy simulations. Comparisons of the performance of the two different free energy simulation methods are provided.

The major interactions stabilizing the base-paired conformation include two parts. The first one is the inner hydrogen bond interactions between in each base pair.[94-100] The number of hydrogen bonds can be 2 in AT pair or various types of mismatches or 3 in GC pair.[101] The second contribution is the stacking interactions between neighboring bases in the same chain.[102-108] The solvent reorganization, the relaxation of the DNA duplex upon base flipping and the DNA-solvent interactions also contribute to the free energy difference.[100, 109-112] Based on the two major interaction parts assumption, various simplified theoretical models are proposed and applied to a number of DNA systems.[113-114] The other minor contributions are included implicitly in the parameterization of the hydrogen bond interactions and the stacking interactions.[100, 108, 110-111, 115] This leads to the two-state helix-coil transition model[103, 116] and the Peyrad-Bishop-Dauxois (PBD) mesoscopic model.[105-106, 117] These models are successfully applied to DNA melting,[39-40, 98-99, 102, 114, 118] bubbling,[119] breathing[37, 100,

[120] and bending.[121] When more details are added to the models, they approach the atomic force field description and simulations with extended Ising models become all-atom MD simulations.

All-atom descriptions of DNA systems are more popular in computational community in recent years due to the increase in computer power. The mostly used derivatives of AMBER force fields are bsc1[122] and OL15,[123-125] both of which are based on the earlier AMBER force field named parm99[126] with its modification bsc0.[127] The dihedral parameter is the most difficult and developing part of nucleic acid force fields and recently developed force fields focus mostly on the refinement of this term. The physical meaning of the dihedral term is not defined clearly. It is closely related to the contributions from electronic structures and is used as the final part tuning the force field. Recent benchmark studies on force fields show that these two last-generation force fields perform similarly in describing the structural ensemble of canonical structures of DNA in long brute-force simulations. The conformational ensemble described by the last-generation AMBER force fields is found to be comparable with the NMR-derived ones.[66] However, in our recent work on base flipping in AT tracts, we observe that the free energy barriers along the base flipping pathway are significantly different under different force fields.[64] The base flipping in mutated DNA duplex also has this behavior,[84] but both last-generation AMBER force fields are able to provide quantitatively correct results for the substitution-induced variation of the free energy barrier along the base flipping pathway. Therefore, in the current work, we perform simulations with all of the three AMBER force fields including the oldest bsc0 modification and the newer OL15 and bsc1 modifications to assess their abilities in describing the thermodynamics in nucleic acid systems (i.e. DNA and TNA duplexes).

The ion concentration in the solvated system may also influence the thermodynamic and kinetic behavior of the system in nucleic acid simulations. As the nucleic acid systems are polyanions with negative charges at the outer phosphate groups, neutralization is often performed by adding Na^+ cations. By adding excess salt to achieve the physiological ion concentration of 0.1 M or 0.15 M, we obtain a system for excess-salt simulations. It has been summarized that the structural dynamics of nucleic acid systems would not be altered significantly with the addition of excess salts.[128-131] However, these observations are obtained from unbiased simulations, where the phase space regions explored are close to the canonical structure. In our enhanced sampling simulations, some high-free-energy regions are visited, which may make the conclusion inapplicable. Therefore, we simulate the systems under different salt concentrations to investigate the salt-concentration effect.

Method and Computational Details

System preparation. The original 2'-deoxy-T (T) in DNA could be mutated to (L)-alpha-threofuranosyl-T (TFT) in TNA. The structures of the DDD and its T-to-TFT mutant are obtained from the crystal structure with the pdbid 1N1O.[26] Based on this structure, we build the DNA and TNA systems simulated in this work. The crystal structure and illustrations of the flipping processes are depicted in Fig. 1a. In the middle of Fig. 1a, the structures of the T residue in DNA and the TFT residue in TNA is compared. The canonical 5'-3' DNA backbone is mutated to the 3'-2' TNA backbone. There are two TFT residues in the system. The 7th residue is at the center of the duplex and thus its flipping should be influenced at the minimal level by the terminal effect. Therefore, we study the flipping of the base of this residue. AM1-BCC[132] charges are derived for the mutated T residue, which is named as TFT in the current work. The other parameters for the TFT residue are obtained from AMBER14SB[133] and GAFF.[134] The flipping 7th residue of the canonical T is mutated to TFT to investigate the variation of thermodynamics in the DNA<->TNA mutation. Three AMBER force fields are tested in the current simulation. Namely, the whole simulation procedure is repeated under force fields including the OL15 combination of modifications of the AMBER force field,[123-125] another last-generation AMBER force field of bsc1,[122] and the oldest bsc0 force field.[127] As the two last-generation AMBER force fields are built on the oldest bsc0, the current work can be seen as an assessment of the improvement introduced in the modifications of the last-generation force fields, from a thermodynamic perspective. The whole system is solvated with TIP3P[135-136] water molecules. The truncated octahedron cell is replicated in whole space by periodic boundary conditions. Non-polarizable spherical counter ions of Na^+ parameterized for the TIP3P water by Joung and Cheatham[137-138] are added for neutralization. Also, we performed simulations to study the variations of the thermodynamic profiles when adding excess salts (Na-Cl ion pairs). The salt concentration of the excess ions is 90 mM, which is consistent with the condition of the experimental study.[26]

Free energy simulation. The equilibrium enhanced sampling method of umbrella sampling and our nonequilibrium stratification method proposed previously are used to bias the sampling.[64] As a traditional and representative equilibrium free energy simulation method, umbrella sampling is well developed and applied to a number of cases. Time-independent (harmonic) biasing potentials are added along the collective variable (CV) to enhance the sampling efficiency. The reaction coordinate or CV is the slow degree of freedom important for describing the process of interest. In this case, the flipping dihedral defined by 4 centers of masses (COM) is used as the CV. Only heavy atoms are included. The definition is illustrated in Fig. 1b. Similar CVs have been widely applied to a number of DNA and RNA systems.[63, 65, 84] After accumulating sufficient statistics in the biased ensemble, we reweight the data to recover the expectations of observables in the original unbiased ensemble. Popular reweighting techniques are maximum likelihood estimators such as

the weighted histogram analysis method (WHAM)[139-140] and the variational free energy profile (vFEP).[141] Note that vFEP uses the cubic spline interpolation and thus the potential of mean force (PMF) estimates obtained from it are often smoother than the WHAM results. Therefore, in the current work, we use the vFEP method to construct the free energy profiles.

The window spacing regime also follows the widely applied one.[63, 65, 142] Specifically, the umbrella windows are equally spaced from 0° to 360° with 5° increments. The force constant of 100 kcal/(mol·rad^2) is used to ensure sufficient phase space overlap for reliable reweighting and enhance the sampling efficiency. In each umbrella window, 2000-cycle energy minimization and 200 ps NPT equilibration are performed. Then we initiate the production run with the sampling interval of 2 ps. The autocorrelation time of the COM pseudo-dihedral is about 1-10 ps.[64] Therefore, such a sampling interval extracts approximately independent samples and avoids excessive output. The simulation proceeds until the convergence is reached. The convergence check of the umbrella sampling simulations is performed with the block averaging method, which is a widely used method to check the convergence of the simulation. In this method, the whole production run is divided into several time blocks to monitor the convergence behavior of the free energy landscape. When similar or identical PMF estimates are obtained from different time blocks, the convergence is reached. The non-equilibrated time blocks can thus be identified and omitted, and the well-equilibrated time blocks are used for later comparisons.

Another free energy simulation method used here to enhance the sampling efficiency is the nonequilibrium stratification method proposed by us previously.[64, 84] It could construct the free energy profile along the base flipping CV with a similar efficiency and accuracy compared with the equilibrium umbrella sampling scheme.[64, 84] The method applies the time-dependent biasing potential to drive the system from one conformational state to another. The whole pulling process from 0° to 360 ° is divided into a series of smaller 2° segments, the dissipation in the pulling process between which is reduced, resulting in improvements in the convergence behavior. In nonequilibrium pulling and initial configuration sampling, a large force constant of 2000 kcal/mol·rad^2 is used to achieve the stiff spring limit, which has been employed in various cases in our previous work.[64, 79, 81] In the nonequilibrium free energy simulation, short configurational sampling is needed to extract equilibrated structures. In each conformational state, we perform 5000-cycle energy-minimization, 200 ps heating from 0 K to 300 K in an NVT ensemble and 150 ps NPT equilibration. Then, short configurational sampling is performed with a sampling interval of 2 ps to extract a set of configurations from equilibrium ensembles. The 2 ps interval is observed to be close to the autocorrelation time of the flipping dihedral in our previous work.[64, 84] To provide a statistically meaningful dataset for the uncertainty estimation

in data analysis, we calculate the autocorrelation time of the pseudo-dihedral in each conformational state τ_i and subsample the whole dataset by the statistical inefficiency $\phi_{eq,i} = 1 + 2\tau_i$ in order to obtain uncorrelated configurations.[143] Bidirectional pulling is then initiated from these equilibrated structures to accumulate the microscopic nonequilibrium works. The pulling speed of 0.5 ps per segment is used due to our previous experience in base flipping simulations.[143] To extract unbiased estimates of the mechanical observables from the nonequilibrium simulations in the statistically optimal way, we employ the Crooks' Equation (CE)[144] or Bennett Acceptance Ratio (BAR)[145] to reweight the statistics. A new convergence criterion for bidirectional reweighting proposed in our previous work is tested here. The criterion is achieved when the standard deviation (SD) is obviously smaller than the overlap scalar,[145-146] which is used to estimate the phase space overlap or the overlap between distributions of nonequilibrium works.[81]

In all simulations, the SHAKE[147] algorithm is applied to perform bond-length constraints for bonds involving hydrogen atoms in all molecules to minimize the fluctuation of chemical bonds.[148] Langevin dynamics[149] with the collision frequency of 4 ps^{-1} are employed for temperature regulation at 300 K. Isotropic position scaling along with the Berendsen barostat is employed to regulate the pressure. A time step of 1 fs is used to integrate the equations of motion. A cutoff of 10 Å for non-bonded interactions in the real space is applied and the long-range electrostatics are treated with the PME method,[150] where the size of the charge grid in each dimension is 64. The direct sum tolerance of PME remains the default value of 10^{-5}. MD simulations are performed with the AMBER[151] 16 suite and all other analyses are performed with homemade codes.

Result and discussion

Equilibrium Simulations. The first thing to check in free energy simulations is the convergence. Block averaging is used to check the convergence. In our previous base flipping simulations, 1 ns, 2 ns or 4 ns time blocks are used to estimate the free energy profiles, as they are estimated to be close to the minimum sampling time for each umbrella window to reach converged sampling results.[64, 84] To make the convergence as reliable as possible, in the current case, we use 4 ns time blocks, the results of which are shown in Fig. 2 (OL15), Fig. S1 (bsc1) and Fig. S2 (bsc0). Note that in our previous study, the statistical error obtained from the bootstrap analysis is much smaller than the systematic error.[64] Therefore, convergence is mainly hindered by bias elimination rather than variance minimization. Thus, in the current study, the statistical error is not shown in the free energy profile. In the first several time blocks, the free energy profiles become lower due to the gradual equilibration of the system in each umbrella window. In the last several time blocks, the free energy profiles

are similar and fluctuations are observed. In this case, we use the overall results obtained from these last time blocks to calculate the PMFs.

As different systems require different lengths of equilibration, the total simulation times for different systems described with different force fields under different salt concentrations differ. The resulting statistics are summarized in Table 1. An empirical conclusion from the statistics is that in base flipping simulations, to obtain converged estimates of the free energy profile from umbrella sampling simulations, 2 μs simulation time is often required to be performed in the whole system. For simpler systems with smaller fluctuations the simulation time could be shorter (e.g. 1 μs).[64, 84] Any result obtained from sub-μs simulations should be treated with care.

Nonequilibrium techniques. The nonequilibrium stratification method coupled with the statistically optimal bidirectional estimator is shown to be very robust in constructing the thermodynamic profiles in base flipping.[64, 84] We then check its applicability in the current base flipping cases. The OL15 force field is used in the illustrative discussion and the behaviors of the other force fields are similar.

The first thing to check is the autocorrelation of the mechanical observable we bias, namely the autocorrelation of the CV or the flipping dihedral. The statistical inefficiency in the equilibrium ensemble calculated during the initial configurational sampling procedure is shown in Fig. 3a. The statistical inefficiency has the lower bound of 2 ps, as the sampling interval is 2 ps. Fluctuations in the value of the statistical inefficiency are observed, but we can still see that the typical value is 2 ps. Therefore, this value is used in our later calculation of the sampling time of the nonequilibrium free energy simulation.

Previously, we have observed the correlation between the convergence of the free energy simulation and the relative size of the dimensionless SD and the overlap scalar. If SD is smaller than the overlap scalar, the convergence is reached. The convergence determined by this criterion agrees with the time-invariant behavior of the free energy profile and the monotonically decreasing behavior of the SD profile during further sampling. Therefore, before checking the free energy profile, we firstly get a glance of the sample size required for converged determined by the SD-smaller-than-overlap criterion. Fig. 3b provides the comparison between the dimensionless SD profile and the overlap profile. We notice that 5-sample SD profile intersects with the overlap profile, while with larger sample sizes such as 50 or 100 samples the criterion seems to be satisfied and the convergence is reached. Therefore, the convergence requires about 50 samples.

To make a more reliable determination of the sample size required for convergence, we plot the iteration-dependence of the free energy profile in Fig. 3c. In each iteration, 5 new samples are added to the dataset. When the simulation converges, the PMF should remain unchanged or fluctuate at a minimal level with the

addition of new samples. In the first 8 iterations, the free energy barrier becomes lower and some fluctuations are observed. Since the 10th iteration, the PMF does not change significantly in further sampling and thus the convergence is reached with 50 samples.

The time-evolution of the SD profile also provides hints on the convergence behavior of the simulation. If the simulation converges, the SDs in all configurational states should decrease monotonically with further sampling. Another worth noting phenomenon is that the variance or SD is often more biased than the free energy itself.[146] Therefore, checking the time-evolution of the SD profile could provide many useful information that cannot be obtained from other statistics (e.g. PMF). Thus, we then check the time-dependence of the SD profile in Fig. 3d. Obviously, in the first several iterations, there are fluctuations in the values of SDs. The 10th iteration seems to be a nice starting point that all SDs decrease monotonically with further sampling. Therefore, the time-dependence of the SD profile also tells us that 50 samples are required for convergence in nonequilibrium stratification.

Therefore, all of the three convergence-check statistics mentioned above indicate that 50 samples should be the minimum sample size for convergence. Therefore, it is used in the later calculation of the efficiency of the nonequilibrium technique. As for the results under the bsc0 force field, we summarized them in Fig. S3 and Fig. S4. The convergence behavior of the simulation under the bsc0 force field is a little better than that under OL15. However, to avoid exaggeration of the performance of the nonequilibrium method, we still estimate the minimum sample size as 50.

Equilibrium vs nonequilibrium. After checking the convergence behavior of the equilibrium and nonequilibrium free energy simulations, we then turn to check the consistency of the free energy estimates obtained from these two methods. In Fig. 4 and Fig. S5, we presented direct comparisons of the free energy profiles constructed from these two free energy methods for the wild-type DNA duplex and the TNA mutant described with the OL15 and bsc0 force fields. We can see that the agreement between the free energy profiles is very good, which indicates the validity and applicability of the nonequilibrium method in constructing the thermodynamic information in the base flipping process of DNA-related systems.

The consistency check provides no information about the efficiency of the method. Therefore, we calculate the simulation time required for convergence in each method and summarize the statistics in Table 2. The equilibrium umbrella sampling requires about 4 ns per umbrella window and 72 windows to converge, which gives a total simulation time of 288 ns. As for the nonequilibrium method, previously we have concluded that 50 samples (i.e. 10 iterations) are needed for convergence and 180 stratification segments are needed, leading to a total simulation time of 27 ns. This indicates that the nonequilibrium method could be

much faster than the equilibrium method in constructing the thermodynamic profiles in base flipping. In our previous work, we also observed that sometimes the nonequilibrium work method could be faster than the equilibrium method.[64, 84] However, these observation does not necessarily indicate that the nonequilibrium method is preferred than the equilibrium one, as there are some non-rigorous parts in the timing data processing. For instance, neither the spacing of the umbrella window nor the stratification spacing is optimal, which is a potential influencing factor of the efficiency comparison. Further, neither the equilibration time in each umbrella window nor that in each configurational state in nonequilibrium simulations is included in the calculation, which could also influence the efficiency of each method. Thus, a modest conclusion should be that the nonequilibrium techniques could be used to obtain the thermodynamic profiles with a similar accuracy and efficiency of the equilibrium method.

Sugar-moiety-mutation-induced variation of the thermodynamic stability of the base pair. The main difference between DNA and TNA duplexes lies in the sugar moiety. The inter-nucleotide linkages in the wild-type DNA system are from the 5' side to the 3' one, while in TNA there is a shift in the position, leading to a quasi *trans*-diaxial 3'-'2' phosphodiester linkages. The number of covalent bonds connecting the consecutive phosphates is thus decreased from 6 in DNA to 5 in TNA.[4, 7, 11-12] Experimental observations indicate that DNA and TNA duplexes are of similar thermodynamic stability.[2, 26] Whether this behavior could be reproduced in computational modelling is then investigated.

In Fig. 5, we provide comparisons between the free energy profiles along the base flipping pathway in DNA and TNA duplexes described with three AMBER force fields in the net-neutral and excess-salt simulations. From the free energy profiles in Fig. 5, we know that the positions of the global minima in both DNA and TNA in different force fields under different salt concentrations are almost the same, which indicates that different force fields describe the canonical structure in a similar fashion. Therefore, in brute-force simulation, the dynamics in the base-paired state are similar in different force fields. It is worth noting that when simulating the system with only the brute-force method, the phase space region explored is very close to the canonical structure. Although different developed force fields have good performance in many benchmarks with long unbiased simulations,[66] they may have different behaviors in regions with higher free energy, which could be of great significances in the simulation of biological processes. Therefore, we then focus on the height of the free energy barrier and the shape of the free energy profile.

In Fig. 5a, we can see that under both salt concentrations, the free energy barriers in DNA and TNA base flipping are of different heights, which indicates that under OL15, the two types of base pair are of different stabilities. Note that the difference between the heights of free energy barriers is not significant. Thus, we

believe that the OL15 force field is still usable. Under bsc1 in Fig. 5b, by contrast, the free energy profiles in DNA and TNA become more similar, and the salt concentration only has a little effect on the relative stability of DNA and TNA base pairs. The situation is significantly different in the bsc0 case shown in Fig. 5c, where the free energy profile of TNA is much lower than that of DNA, regardless of the salt concentration. Therefore, the bsc1 force field could reproduce the experimental findings in the most accurate way among the three force fields, the OL15 force field is usable, and the bsc0 one is unable to describe the DNA-to-TNA mutation properly and is not recommend in any modern application, especially when dealing with the TNA case.

Salt-concentration dependence. After discussing the system-dependence of the thermodynamic profile, we then focus on the influence of the salt concentration. In net-neutral simulations, only Na^+ counter ions are added to neutralize the system, enabling simulations with the periodic boundary condition. More ion pairs are added to achieve the physiological ion concentration to provide an excess-salt system. In the current case, we add 90 mM NaCl ion pairs to reproduce the experimental condition. Many simulation reports observe that the results obtained from net-neutral and excess-salt simulations are similar. However, due to the time-scale limitation in the previous studies, the phase space regions explored are close to the canonical structure. In our enhanced sampling simulations, the high-free-energy regions are also visited in order to construct the free energy profile along the base flipping pathway. In these high-energy regions, the conclusion from previous simulations may not be applicable. Therefore, we then investigate whether the base-flipping free energy profile is invariant upon the addition of the excess salt.

The comparison could still be done in Fig. 5. In Fig. 5a, when the force field is fixed to OL15, for both DNA and TNA duplexes, the free energy barriers are elevated upon the addition of the excess salt. The increases are similar for different systems. Therefore, if the focus of the comparison is the relative stability of DNA and TNA duplexes, there will not be obvious alteration in the conclusion. This phenomenon is also observed in our previous study focusing on the sulfur-substitution effect in DNA duplex.[84] In Fig. 5b, under the bsc1 force field, the free energy profiles for DNA and TNA systems also change due to the increase of the salt concentration. However, the increases in the free energy barrier are smaller than the OL15 case. The bsc0 results in Fig 5c behave similarly, compared with bsc1 ones. It is worth noting that the position of the free energy minimum describing the base-paired state remains untouched in the addition of the ion pairs. Therefore, it is expected that the unbiased simulations are unable to observe the above differences in the free energy profiles. Therefore, we conclude that the salt concentration does have an effect on the thermodynamics of the system, especially when the high-energy regions are explored in free energy simulations. This emphasizes the importance of the addition of sufficient salts to reproduce the experimental condition in a more accurate way.

Force-field-dependence. With converged phase space sampling, the force field is the main factor determining the accuracy of the modelling. We then discuss the behaviors of different force fields when the system and the salt concentration are invariable. Clearer comparisons between force fields when fixing the salt concentration are provided in Fig. 6.

When the salt concentration is set to the minimal net-neutralizing condition, as shown in Fig. 6a, for DNA duplex, the shapes of the free energy profiles under bsc1 and bsc0 are similar, and the height of the free energy barrier along the flipping pathway in bsc0 is very similar to that in bsc1. The OL15 one is significantly lower than the other. Upon the DNA-to-TNA mutation, the system described with OL15 experiences a modest destabilization, while the bsc0 one is significantly destabilized. The change in the free energy profile under bsc1 is relatively small. The resulting rank of the height of the free energy barrier in TNA duplex is bsc1 > OL15 > bsc0. Therefore, even for the same system, different force fields provide different thermodynamic profiles along the base flipping pathway, which emphasizes the significance of the selection of the force fields in base flipping simulations. As the rank of the heights of the free energy barriers in different force fields changes with the system (i.e. DNA or TNA), the system-dependence is also observed. Thus, testing the behaviors of different force fields would be valuable in practical applications. The excess-salt simulations seem to differentiate different force fields better. In Fig. 6b, for the DNA duplex, the height of the free energy barrier has the rank bsc1 > bsc0 > OL15. Namely, in excess-salt simulations, the thermodynamic behaviors of bsc1 and bsc0 are distinguished more clearly. When it comes to the TNA duplex, again the stability of the base-paired state in bsc0 decreases significantly and the stability rank is bsc1 > OL15 > bsc0.

Therefore, considering the discussion about the salt-concentration and system dependence, either under the net-neutral condition or in the excess-salt simulation, the bsc1 force field is preferred for a stable base-paired state, regardless of the system under investigation. The bsc1 force field also performs very well in describing the similarity of the stabilities of DNA and TNA duplexes. The OL15 force field is also usable as its thermodynamic profile shows relatively small dependence on the system or the salt concentration, which indicates that OL15 could describe the similar stabilities of the two systems in an appropriate way. The bsc0 PMF shows significant variations when the DNA-to-TNA mutation happens, which is inconsistent with the experimental observations. Also, as it is the oldest version of AMBER modifications, we do not recommend its practical application in modern computational study.

Conclusion

Compared with the natural nucleic acids of DNA and RNA, TNA has higher chemical stability, less

reactive groups, a smaller number of hydroxyl groups, and less enantiomeric pairs of diastereoisomers. It was simpler than RNA and thus was assumed to be a good predecessor of RNA in the pre-RNA world. It could also be more suitable for information storage due to its reduced conformational flexibility compared with RNA. The canonical conformation was observed to be unaltered in the DNA-to-TNA mutation. The interaction networks in DNA and TNA were similar. The thermodynamic stability was also observed to be unperturbed. In our work, we performed equilibrium and nonequilibrium free energy simulations to construct the free energy profiles along the base flipping pathway, aiming at investigating whether the thermodynamic profile during conformational change in TNA was also similar to that in DNA. Three most popular AMBER force fields were benchmarked and the effect of ion concentrations was studied.

The equilibrium and nonequilibrium free energy simulation methods were found to be of similar efficiency and accuracy in our previous work. In the current work, we still observed that these two simulation methods could provide the base-flipping free energy profiles within similar amounts of simulation time. The agreement was observed regardless of systems and force fields. Therefore, the nonequilibrium stratification method could be a nice alternative to the equilibrium free energy simulation methods in base flipping simulations.

The investigation of the variation of the sugar moiety was then performed. We compared the free energy profiles of base flipping in DNA and TNA duplex and found that all of the three force fields provided similar descriptions of the base-paired state. However, the flipping-out states under different force fields differed significantly. Under the bsc1 force field, the free energy profiles in DNA and TNA were similar, while under OL15 some differences were observed. The bsc0 force field provided significantly different thermodynamic profiles in base flipping in DNA and TNA duplex, which deviates from the experimental observations. Thus, the bsc0 force field would not be applied to any case in modern nucleotide simulations. As bsc1 was shown to have a good description of the base-paired state in several benchmarks, the bsc1 force field was recommended in base flipping simulations in DNA and TNA duplexes.

It has been observed in many MD simulations that there was no marked difference in the nucleic acid structural dynamics under net-neutral and excess-salt conditions. However, previous investigations relied on unbiased simulations, where the phase space regions explored were close to the canonical structure. In our free energy simulation constructing the free energy profiles along the base flipping pathway, the high-energy regions were also explored. The thermodynamic behavior of these regions could be influenced by the salt concentration. Therefore, we checked the validity of the previous conclusion in the base flipping cases. Notably, in our results, the ion concentration had a significant effect on the thermodynamics in the base

flipping process in DNA and TNA duplexes. The free energy barrier along the base-flipping pathway without excess salts showed obvious difference, compared with that under the experimental excess-salt concentration of 0.09 M. Such a phenomenon emphasized the needs of adding excess salts to reproduce the experimental ensemble precisely. Note that the PMFs under the bsc1 and bsc0 force fields were not significantly perturbed by the excess salts. Therefore, considering this salt-concentration stability, these two force fields were preferred in base flipping simulations. As bsc1 was superior to bsc0 in a number of benchmark studies and was an updated version of bsc0, we recommended using bsc1 in practice.

In the last part of our study, we focused on the force-field comparison, identifying the difference between free energy profiles of the same system described with different force fields when the salt concentration was fixed. The base-flipping free energy profiles were different in different force fields. The base-paired state was of the highest stability under the bsc1 force field. As the base-paired state was found to be described better with bsc1 and OL15 than bsc0 in many benchmark studies, the bsc1 force field would be preferred.

Therefore, considering our investigations of the force-field dependence and the salt-concentration dependence of the simulation results, we recommended using the bsc1 force field and adding sufficient excess salts to reproduce the experimental condition in practical application of the base flipping simulations.

Acknowledgement

This work was supported China Scholarship Council and National Key Research and Development Program of China [No. 2016YFA0501700]. Computer access to the CLAIX cluster of RWTH Aachen University and clusters of Forschungszentrum Juelich is gratefully acknowledged. We thank the anonymous reviewers for valuable comments and critical reading.

Conflicts of interest

There are no conflicts of interest to declare.

Supporting Information

The convergence behavior of the free energy profiles obtained from equilibrium umbrella sampling simulations of systems described with bsc1 and bsc0 force fields in DNA and TNA systems, the time-evolution of the free energy profiles and SD profiles and the comparison between dimensionless SD profiles and the dimensionless overlap profile in nonequilibrium stratification, and the comparison between the free energy profiles constructed from equilibrium and nonequilibrium methods are provided in the supporting information.

References

1. Eschenmoser, A., Chemical etiology of nucleic acid structure. *Science* **1999,** *284* (5423), 2118-2124.
2. Schöning, K.-U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A., Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3'→ 2') oligonucleotide system. *Science* **2000,** *290* (5495), 1347-1351.
3. Albert, E., Etiology of potentially primordial biomolecular structures: from vitamin?B12 to the nucleic acids and an inquiry into the chemistry of life's origin: a retrospective. *Angew. Chem. Int. Ed.* **2012,** *43* (13), no-no.
4. Chiba, J.; Inouye, M., Exotic DNAs made of nonnatural bases and natural phosphodiester bonds. *Chem. Biodivers.* **2010,** *7* (2), 259-282.
5. D'Alonzo, D.; Guaragna, A.; Palumbo, G., Exploring the role of chirality in nucleic acid recognition. *Chem. Biodivers.* **2011,** *8* (3), 373-413.
6. Saladino, R.; Crestini, C.; Ciciriello, F.; Costanzo, G.; Di Mauro, E., Formamide chemistry and the origin of informational polymers. *Chem. Biodivers.* **2007,** *4* (4), 694-720.
7. Ichida, J. K.; Allen, H.; Keyong, Z.; Mclaughlin, L. W.; Szostak, J. W., High fidelity TNA synthesis by Therminator polymerase. *Nucleic Acids Research* **2005,** *33* (16), 5219-25.
8. Kempeneers, V.; Froeyen, M.; Vastmans, K.; Herdewijn, P., Influence of threose nucleoside units on the catalytic activity of a hammerhead ribozyme. *Chem. Biodivers.* **2010,** *1* (1), 112-123.
9. Aviles-Moreno, J.-R.; Huet, T. R., Sugars in the gas phase: The conformational properties of erythrose, threose, and erythrulose characterized by quantum chemistry calculations. *Journal of Molecular Structure: THEOCHEM* **2008,** *858* (1-3), 113-119.
10. Bednarko, J.; Stachurski, O.; Wielińska, J.; Kozakiewicz, K.; Liberek, B.; Nowacki, A., Threocytidines: Insight into the Conformational Preferences of Artificial Threose Nucleic Acid (TNA) Building Blocks in B3LYP Studies. *Journal of Molecular Graphics & Modelling* **2018,** *80*, 157.
11. Wu, X.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A., 2, 6-Diaminopurine in TNA: Effect on Duplex Stabilities and on the Efficiency of Template-Controlled Ligations1. *Org. Lett.* **2002,** *4* (8), 1283-1286.
12. Wu, X.; Guntha, S.; Ferencic, M.; Krishnamurthy, R.; Eschenmoser, A., Base-Pairing Systems Related to TNA: α-Threofuranosyl Oligonucleotides Containing Phosphoramidate Linkages1. *Org. Lett.* **2002,** *4* (8), 1279-1282.
13. Mohitosh, M.; Vanessa, S.; Mikhail, A.; Eveline, L.; Helmut, R.; Mathy, F.; Amutha, R.; Arnout, C.; Andreas, M.; Piet, H., Solution structure and conformational dynamics of deoxyxylonucleic acids (dXNA): an orthogonal nucleic acid candidate. *Chemistry* **2012,** *18* (3), 869-879.
14. Anosova, I.; Kowal, E. A.; Sisco, N. J.; Sau, S.; Liao, J. Y.; Bala, S.; Rozners, E.; Egli, M.; Chaput, J. C.; Van Horn, W. D., Structural Insights into Conformational Differences between DNA/TNA and RNA/TNA Chimeric Duplexes. *Chembiochem A European Journal of Chemical Biology* **2016,** *17* (18), 1705-1708.
15. Marc-Olivier, E.; Christian, M.; Ramanarayanan, K.; Albert, E.; Bernhard, J., The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from alpha-(L)-threofuranosyl-(3'-2')-CGAATTCG. *J. Am. Chem. Soc.* **2008,** *130* (45), 15105-15.
16. Eschenmoser, A., The TNA-Family of Nucleic Acid Systems: Properties and Prospects. *Origins of Life & Evolution of the Biosphere the Journal of the International Society for the Study of the Origin of Life* **2004,** *34* (3), 277.
17. Chaput, J. C.; Ichida, J. K.; Szostak, J. W., DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.* **2002,** *125* (4), 856-857.
18. Schmidt, J. G.; Nielsen, P. E.; Orgel, L. E., Enantiomeric cross-inhibition in the synthesis of oligonucleotides on a nonchiral template. *J. Am. Chem. Soc.* **1997,** *119* (6), 1494-1495.
19. Toti, K.; Renders, M.; Groaz, E.; Herdewijn, P.; Van Calenbergh, S., Nucleosides with transposed base or 4′-hydroxymethyl moieties and their corresponding oligonucleotides. *Chem. Rev.* **2015,** *115* (24), 13484-13525.
20. Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P., Recognition of threosyl nucleotides by DNA and RNA polymerases. *Nucleic acids research* **2003,** *31* (21), 6221-6226.
21. Chaput, J. C.; Szostak, J. W., TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* **2003,** *125* (31), 9274-9275.

22. J Craig, B.; Alonso, R.; Szostak, J. W., Synthesis and nonenzymatic template-directed polymerization of 2'-amino-2'-deoxythreose nucleotides. *J. Am. Chem. Soc.* **2014,** *136* (5), 2033-9.
23. Zhang, S.; Chaput, J. C., Synthesis and enzymatic incorporation of α-l-threofuranosyl adenine triphosphate (tATP). *Bioorg. Med. Chem. Lett.* **2013,** *23* (5), 1447-1449.
24. Toti, K.; Renders, M.; Groaz, E.; Herdewijn, P.; Van, C. S., Nucleosides with Transposed Base or 4'-Hydroxymethyl Moieties and Their Corresponding Oligonucleotides. *Chem. Rev.* **2015,** *115* (24), 13484.
25. Loakes, D.; Holliger, P., Polymerase engineering: towards the encoded synthesis of unnatural biopolymers. *Chem. Commun.* **2009,** (31), 4619-4631.
26. Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M., Crystal structure of a B-form DNA duplex containing (L)-α-threofuranosyl (3 '→ 2 ') nucleosides: A four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.* **2002,** *124* (46), 13716-13721.
27. Freudenthal, B. D.; Beard, W. A.; Perera, L.; Shock, D. D.; Kim, T.; Schlick, T.; Wilson, S. H., Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* **2015,** *517* (7536), 635.
28. Banavali, N. K.; Jr, M. K., Free Energy and Structural Pathways of Base Flipping in a DNA GCGC Containing Sequence. *Journal of Molecular Biology* **2002,** *319* (1), 141-160.
29. And, E. G.; Lavery, R., Nucleic Acid Base Pair Dynamics: The Impact of Sequence and Structure Using Free-Energy Calculations. *J. Am. Chem. Soc.* **2003,** *125* (17), 4998-4999.
30. Irobalieva, R. N.; Fogg, J. M.; Catanese Jr, D. J.; Sutthibutpong, T.; Chen, M.; Barker, A. K.; Ludtke, S. J.; Harris, S. A.; Schmid, M. F.; Chiu, W., Structural diversity of supercoiled DNA. *Nature communications* **2015,** *6*, 8440.
31. Vologodskii, A.; Lukashin, A.; Anshelevich, V.; Frank-Kamenetskii, M., Fluctuations in superhelical DNA. *Nucleic acids research* **1979,** *6* (3), 967-982.
32. Keepers, J.; Kollman, P. A.; James, T. L., Molecular mechanical studies of base-pair opening in d (CGCGC): d (GCGCG), dG5· dC5, d (TATAT): d (ATATA), and dA5· dT5 in the B and Z forms of DNA. *Biopolymers: Original Research on Biomolecules* **1984,** *23* (11), 2499-2511.
33. Keepers, J. W.; Kollman, P. A.; Weiner, P. K.; James, T. L., Molecular Mechanical Studies of DNA Flexibility: Coupled Backbone Torsion Angles and Base-pair Openings. *Proceedings of the National Academy of Sciences* **1982,** *79* (18), 5537-5541.
34. And, U. D. P.; Jr, M. K., Computational Approaches for Investigating Base Flipping in Oligonucleotides†. *Cheminform* **2006,** *106* (2), 489-505.
35. Yang, W., An overview of Y-family DNA polymerases and a case study of human DNA polymerase η. *Biochemistry* **2014,** *53* (17), 2793-2803.
36. Kusumoto, R.; Masutani, C.; Iwai, S.; Hanaoka, F., Translesion synthesis by human DNA polymerase η across thymine glycol lesions. *Biochemistry* **2002,** *41* (19), 6090-6099.
37. Duduiala, C. I.; Wattis, J. A.; Dryden, I. L.; Laughton, C. A., Nonlinear Breathing Modes at A Defect Site in DNA. *Phys Rev E* **2009,** *80* (6), 061906.
38. Pincet, F.; Perez, E.; Bryant, G.; Lebeau, L.; Mioskowski, C., Long-range attraction between nucleosides with short-range specificity: Direct measurements. *Phys. Rev. Lett.* **1994,** *73* (20), 2780.
39. Lazurkin, Y. S.; Frank-Kamenetskii, M.; Trifonov, E., Perspectives Report: Melting of DNA: Its Study and Application as A Research Method. *Biopolymers* **1970,** *9* (11), 1253-1306.
40. Gotoh, O., Prediction of Melting Profiles and Local Helix Stability for Sequenced DNA. *Advances in biophysics* **1983,** *16*, 1-52.
41. HART, K.; NYSTRÖM, B.; ÖHMAN, M.; NILSSON, L., Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *RNA* **2005,** *11* (5), 609-618.
42. Qi, W.; Song, B.; Lei, X.; Wang, C.; Fang, H., DNA Base Pair Hybridization and Water-mediated Metastable Structures studied by Molecular Dynamics Simulations. *Biochemistry* **2011,** *50* (44), 9628-9632.
43. Galindo-Murillo, R.; Roe, D. R.; Cheatham, T. E., 3rd, On the absence of intrahelical DNA dynamics on the mus to ms

timescale. *Nat Commun* **2014,** *5*, 5152.
44. Dixit, S. B.; Beveridge, D. L.; Case, D. A.; Cheatham 3rd, T. E.; Giudice, E.; Lankas, F.; Lavery, R.; Maddocks, J. H.; Osman, R.; Sklenar, H., Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. II: sequence context effects on the dynamical structures of the 10 unique dinucleotide steps. *Biophys. J.* **2005,** *89* (6), 3721-3740.
45. Pérez, A.; Luque, F. J.; Orozco, M., Dynamics of B-DNA on the microsecond time scale. *J. Am. Chem. Soc.* **2007,** *129* (47), 14739-14745.
46. Xi, K.; Wang, F.-H.; Xiong, G.; Zhang, Z.-L.; Tan, Z.-J., Competitive Binding of Mg 2+ and Na+ Ions to Nucleic Acids: From Helices to Tertiary Structures. *Biophys. J.* **2018,** *114* (8), 1776-1790.
47. Zhang, Z.-L.; Wu, Y.-Y.; Xi, K.; Sang, J.-P.; Tan, Z.-J., Divalent Ion-mediated DNA-DNA Interactions: A Comparative Study of Triplex and Duplex. *Biophys. J.* **2017,** *113* (3), 517-528.
48. Yongping, P.; Mackerell, A. D., Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. *Nucleic Acids Research* **2003,** *31* (24), 7131-40.
49. Dans, P. D.; Faustino, I.; Battistini, F.; Zakrzewska, K.; Lavery, R.; Orozco, M., Unraveling The Sequence-dependent Polymorphic Behavior of d(CpG) Steps in B-DNA. *Nucleic acids research* **2014,** *42* (18), 11304-11320.
50. Banyasz, A.; Gustavsson, T.; Onidas, D.; Changenet-Barret, P.; Markovitsi, D.; Improta, R., Multi-Pathway Excited State Relaxation of Adenine Oligomers in Aqueous Solution: A Joint Theoretical and Experimental Study. *Chemistry–A European Journal* **2013,** *19* (11), 3762-3774.
51. Conti, I.; Altoè, P.; Stenta, M.; Garavelli, M.; Orlandi, G., Adenine deactivation in DNA resolved at the CASPT2//CASSCF/AMBER level. *Phys. Chem. Chem. Phys.* **2010,** *12* (19), 5016-5023.
52. Parenti, M. D.; Rastelli, G., Advances and applications of binding affinity prediction methods in drug discovery. *Biotechnol Adv* **2012,** *30* (1), 244-50.
53. Gohlke, H.; Case, D. A., Converging free energy estimates: MM-PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004,** *25* (2), 238-250.
54. Olsson, M. A.; Söderhjelm, P.; Ryde, U., Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J. Comput. Chem.* **2016,** *37* (17), 1589-1600.
55. Monroe, J. I.; Shirts, M. R., Converging free energies of binding in cucurbit[7]uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. *J. Comput. Aided Mol. Des.* **2014,** *28* (4), 401-15.
56. Sun, Z.; Wang, X.; Song, J., Extensive Assessment of Various Computational Methods for Aspartate's pKa Shift. *J. Chem. Inf. Model.* **2017,** *57* (7), 1621-1639.
57. Villamaina, D.; Trizac, E., Thinking outside the box: fluctuations and finite size effects. *European Journal of Physics* **2014,** *35* (3), 035011.
58. Heidari, M.; Cortes-Huerto, R.; Kremer, K.; Potestio, R., Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations. *The European Physical Journal E* **2018,** *41* (5), 64.
59. Román, F.; White, J.; Velasco, S., Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.* **1997,** *107* (12), 4635-4641.
60. Lebowitz, J.; Percus, J., Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.* **1961,** *122* (6), 1675.
61. Salacuse, J.; Denton, A.; Egelstaff, P., Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. I. Theoretical method. *Physical Review E* **1996,** *53* (3), 2382.
62. Gallavotti, G., Ergodicity: a historical perspective. Equilibrium and Nonequilibrium. *The European Physical Journal H* **2016,** *41* (3), 181-259.
63. Lemkul, J. A.; Savelyev, A.; MacKerell Jr, A. D., Induced Polarization Influences The Fundamental Forces in DNA Base Flipping. *The journal of physical chemistry letters* **2014,** *5* (12), 2077-2083.
64. Wang, X.; Sun, Z., Determination of Base Flipping Free Energy Landscapes from Nonequilibrium Stratification. *J. Chem. Inf. Model.* **2019,** *59* (6), 2980-2994.

65. Sun, Z.; Wang, X.; Zhang, J. Z. H., Protonation-dependent Base Flipping in The Catalytic Triad of A Small RNA. *Chem. Phys. Lett.* **2017,** *684*, 239-244.
66. Dans, P. D.; Ivani, I.; Hospital, A.; Portella, G.; González, C.; Orozco, M., How Accurate are Accurate Force-Fields for B-DNA? *Nucleic Acids Research* **2017,** *45* (7), 4217-4230.
67. Galindo-Murillo, R.; Robertson, J. C.; Zgarbová, M.; Sponer, J.; Otyepka, M.; Jurec˘ka, P.; Cheatham III, T. E., Assessing The Current State of AMBER Force Field Modifications for DNA. *J. Chem. Theory Comput.* **2016,** *12* (8), 4114-4127.
68. Dans, P. D.; Danilāne, L.; Ivani, I.; Dršata, T.; Lankaš, F.; Hospital, A.; Walther, J.; Pujagut, R. I.; Battistini, F.; Gelpí, J. L., Long-timescale Dynamics of The Drew–Dickerson Dodecamer. *Nucleic acids research* **2016,** *44* (9), 4052-4066.
69. Priyakumar, U. D.; Mackerell, A. D., Base Flipping in a GCGC Containing DNA Dodecamer: A Comparative Study of the Performance of the Nucleic Acid Force Fields, CHARMM, AMBER, and BMS. *J. Chem. Theory Comput.* **2006,** *2* (1), 187-200.
70. Hooft, R. W.; van Eijck, B. P.; Kroon, J., An Adaptive Umbrella Sampling Procedure in Conformational Analysis using Molecular Dynamics and Its Application to Glycol. *J. Chem. Phys.* **1992,** *97* (9), 6690-6694.
71. Mezei, M., Adaptive Umbrella Sampling: Self-consistent Determination of the Non-Boltzmann Bias. *J. Comput. Phys.* **1987,** *68* (1), 237-248.
72. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.
73. Fukunishi, H.; Watanabe, O.; Takada, S., On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002,** *116* (116), 9058-9067.
74. Itoh, S. G.; Damjanovic, A.; Brooks, B. R., pH replica-exchange method based on discrete protonation states. *Proteins* **2011,** *79* (12), 3420-36.
75. Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C., Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* **2006,** *2* (2), 420.
76. T, G.; CM, S., Mechanism of Amyloid-β Fibril Elongation. *Biochemistry* **2014,** *53* (44), 6981-91.
77. Sugita, Y.; Okamoto, Y., Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999,** *314* (1-2), 141-151.
78. Sugita, Y.; Kitao, A.; Okamoto, Y., Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000,** *113* (15), 6042-6051.
79. Wang, X.; Xingzhao, T.; Boming, D.; John Z. H., Z.; Sun, Z., BAR-based Optimum Adaptive Steered MD for Configurational Sampling. *J. Comput. Chem.* **2019,** *40* (12), 1270-1289.
80. Wang, X.; Tu, X.; Zhang, J. Z. H.; Sun, Z., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation: The Nonequilibrium Stratification. *Phys. Chem. Chem. Phys.* **2018,** *20* (3), 2009-2021.
81. Wang, X.; He, Q.; Sun, Z., BAR-Based Multi-Dimensional Nonequilibrium Pulling for Indirect Construction of a QM/MM Free Energy Landscape. *Phys. Chem. Chem. Phys.* **2019,** *21* (12), 6672-6688
82. Sun, Z. X.; Wang, X. H.; Zhang, J. Z. H., BAR-based Optimum Adaptive Sampling Regime for Variance Minimization in Alchemical Transformation. *Phys. Chem. Chem. Phys.* **2017,** *19* (23), 15005-15020.
83. Sun, Z., BAR-based multi-dimensional nonequilibrium pulling for indirect construction of QM/MM free energy landscapes: from semi-empirical to ab initio. *Phys. Chem. Chem. Phys.* **2019,** *21* (39), 21942-21959
84. Sun, Z.; Wang, X.; Zhang, J. Z. H.; He, Q., Sulfur-substitution-induced base flipping in the DNA duplex. *Phys. Chem. Chem. Phys.* **2019,** *21*, 14923-14940.
85. Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Fajardo, D.; Moussatche, P., Alternative Watson–Crick synthetic genetic systems. *Cold Spring Harbor perspectives in biology* **2016,** *8* (11), a023770.
86. Ferencic, M.; Reddy, G.; Wu, X.; Guntha, S.; Nandy, J.; Krishnamurthy, R.; Eschenmoser, A., Base-Pairing Systems Related to TNA Containing Phosphoramidate Linkages: Synthesis of Building Blocks and Pairing Properties. *Chem. Biodivers.* **2004,** *1* (7), 939-979.
87. Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C., An efficient and faithful in vitro replication system for threose nucleic acid. *J. Am. Chem. Soc.* **2013,** *135* (9), 3583-3591.

88. Egli, M.; Pallan, P. S., Crystallographic studies of chemically modified nucleic acids: a backward glance. *Chem. Biodivers.* **2010,** *7* (1), 60-89.
89. Boer, D. R.; Canals, A.; Coll, M., DNA-binding drugs caught in action: the latest 3D pictures of drug-DNA complexes. *Dalton Transactions* **2008,** *3* (3), 399-414.
90. Chaput, J. C.; Yu, H.; Zhang, S., The emerging world of synthetic genetics. *Chem. Biol.* **2012,** *19* (11), 1360-1371.
91. Joyce, G. F., Toward an alternative biology. *Science* **2012,** *336* (6079), 307-308.
92. Herdewijn, P.; Marliere, P., Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem. Biodivers.* **2009,** *6* (6), 791-808.
93. Pinheiro, V. B.; Holliger, P., The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **2012,** *16* (3-4), 245-252.
94. Ferrantini, A.; Carlon, E., Anomalous zipping dynamics and forced polymer translocation. *Journal of Statistical Mechanics Theory & Experiment* **2011,** *2011* (2), P02020.
95. Manghi, M.; Palmeri, J.; Destainville, N., Coupling between denaturation and chain conformations in DNA: stretching, bending, torsion and finite size effects. *Physics* **2008,** *21* (3), 034104-034104.
96. Chakrabarti, B.; Levine, A. J., Nonlinear elasticity of an α-helical polypeptide: Monte Carlo studies. *Physical Review E* **2006,** *74* (3), 031903.
97. Benham, C. J., The equilibrium statistical mechanics of the helix–coil transition in torsionally stressed DNA. *J. Chem. Phys.* **1980,** *72* (6), 3633-3639.
98. Wildes, A.; Theodorakopoulos, N.; Valle-Orero, J.; Cuesta-López, S.; Garden, J.-L.; Peyrard, M., Structural Correlations and Melting of B-DNA Fibers. *Physical Review E* **2011,** *83* (6), 061923.
99. Velizhanin, K. A.; Chien, C.-C.; Dubi, Y.; Zwolak, M., Driving Denaturation: Nanoscale Thermal Transport as A Probe of DNA Melting. *Physical Review E* **2011,** *83* (5), 050906.
100. Peyrard, M.; Cuesta-Lopez, S.; James, G., Nonlinear Analysis of The Dynamics of DNA Breathing. *Journal of biological physics* **2009,** *35* (1), 73.
101. Pechlaner, M.; Donghi, D.; Zelenay, V.; Sigel, R. K., Protonation-Dependent Base Flipping at Neutral pH in the Catalytic Triad of a Self-Splicing Bacterial Group II Intron. *Angew. Chem. Int. Ed.* **2015,** *54* (33), 9687-9690.
102. Blake, R.; Bizzaro, J., Jd; Day, G.; Delcourt, S.; Knowles, J.; Marx, K.; Santalucia, J. J., Statistical Mechanical Simulation of Polymeric DNA Melting with MELTSIM. *Bioinformatics* **1999,** *15* (5), 370-375.
103. Takeno, S.; Homma, S., Topological Solitons and Modulated Structure of Bases in DNA Double Helices: A Dynamic Plane Base-Rotator Model. *Prog. Theor. Phys.* **1983,** *70* (1), 308-311.
104. Wartell, R. M.; Montroll, E. W., Equilibrium denaturation of natura and of periodic synthetic DNA molecules. *Adv. Chem. Phys.* **1970,** *22*, 129-203.
105. Dauxois, T.; Peyrard, M.; Bishop, A. R., Entropy-driven DNA Denaturation. *Physical Review E* **1993,** *47* (1), R44-R47.
106. Muto, V.; Lomdahl, P.; Christiansen, P., Two-dimensional Discrete Model for DNA Dynamics: Longitudinal Wave Propagation and Denaturation. *Phys. Rev. A* **1990,** *42* (12), 7452-7458.
107. Alexandrov, B. S.; Gelev, V.; Monisova, Y.; Alexandrov, L. B.; Bishop, A. R.; Rasmussen, K. O.; Usheva, A., A nonlinear dynamic model of DNA with a sequence-dependent stacking term. *Nucleic Acids Res.* **2009,** *37* (7), 2405-10.
108. Joyeux, M.; Buyukdagli, S., Dynamical Model based on Finite Stacking Enthalpies for Homogeneous and Inhomogeneous DNA Thermal Denaturation. *Physical Review E* **2005,** *72* (5), 051902.
109. Frank-Kamenetskii, M. D.; Prakash, S., Fluctuations in the DNA double helix: A critical review. *Physics of Life Reviews* **2014,** *11* (2), 153-170.
110. ZdravkoviĆ, S., Helicoidal Peyrard–Bishop Model of DNA Dynamics. *Journal of Nonlinear Mathematical Physics* **2011,** *18* (sup2), 463-484.
111. Deng, M. L.; Zhu, W. Q., Stochastic Dynamics and Denaturation of Thermalized DNA. *Physical Review E* **2008,** *77* (2), 021918.
112. Šponer, J.; Riley, K. E.; Hobza, P., Nature and magnitude of aromatic stacking of nucleic acid bases. *Phys. Chem. Chem.*

Phys. **2008,** *10* (19), 2595-2610.
113. Vedenov, A.; Dykhne, A. M.; Frank-Kamenetskiĭ, M., The helix-coil transition in DNA. *Soviet Physics Uspekhi* **1972,** *14* (6), 715-736.
114. Zimm, B. H., Theory of ``Melting'' of The Helical Form in Double Chains of The DNA Type. *J. Chem. Phys.* **1960,** *33* (5), 1349-1356.
115. Kalosakas, G.; Rasmussen, K. Ø.; Bishop, A. R., Non-exponential Decay of Base-pair Opening Fluctuations in DNA. *Chem. Phys. Lett.* **2006,** *432* (1-3), 291-295.
116. Yakushevich, L. V., Is DNA A Nonlinear Dynamical System where Solitary Conformational Waves are Possible? *J. Biosci. (Bangalore)* **2001,** *26* (3), 305-313.
117. Baird, N. C., Simulation of Hydrogen Bonding in Biological Systems: Ab initio Calculations for NH3 NH3 and NH3 NH4+. *Int. J. Quantum Chem* **1974,** *8* (S1), 49-54.
118. Wong, K.; Pettitt, B., The Pathway of Oligomeric DNA Melting investigated by Molecular Dynamics Simulations. *Biophys. J.* **2008,** *95* (12), 5618.
119. Travers, A., DNA Dynamics: bubble 'n'flip for DNA Cyclisation? *Curr. Biol.* **2005,** *15* (10), R377-R379.
120. Alexandrov, B. S.; Fukuyo, Y.; Lange, M.; Horikoshi, N.; Gelev, V.; Rasmussen, K. Ø.; Bishop, A. R.; Usheva, A., DNA Breathing Dynamics Distinguish Binding from Nonbinding Consensus Sites for Transcription Factor YY1 in Cells. *Nucleic acids research* **2012,** *40* (20), 10116-10123.
121. Zeida, A.; Machado, M. R.; Dans, P. D.; Pantano, S., Breathing, Bubbling, and Bending: DNA Flexibility from Multimicrosecond Simulations. *Physical Review E* **2012,** *86* (2), 021903.
122. Ivani, I.; Dans, P. D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Goñi, R.; Balaceanu, A., Parmbsc1: A Refined Force Field for DNA Simulations. *Nat. Methods* **2016,** *13* (1), 55.
123. Zgarbová, M.; Šponer, J.; Otyepka, M.; Iii, T. E. C.; Galindomurillo, R.; Jurečka, P., Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **2015,** *11* (12), 5723-5736.
124. Zgarbová, M.; Luque, F. J.; Šponer, J. í.; Cheatham III, T. E.; Otyepka, M.; Jurecka, P., Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J. Chem. Theory Comput.* **2013,** *9* (5), 2339-2354.
125. Krepl, M.; Zgarbová, M.; Stadlbauer, P.; Otyepka, M.; Banáš, P.; Koca, J.; Cheatham III, T. E.; Jurecka, P.; Šponer, J. i., Reference Simulations of Noncanonical Nucleic Acids with Different χ Variants of the AMBER Force Field: Quadruplex DNA, Quadruplex RNA, and Z-DNA. *J. Chem. Theory Comput.* **2012,** *8* (7), 2506-2520.
126. Wang, J.; Cieplak, P.; Kollman, P. A., How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000,** *21* (12), 1049-1074.
127. Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M., Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of α/γ Conformers. *Biophys. J.* **2007,** *92* (11), 3817-3829.
128. Šponer, J.; Cang, X.; Cheatham III, T. E., Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods* **2012,** *57* (1), 25-39.
129. Krasovska, M. V.; Sefcikova, J.; Réblová, K.; Schneider, B.; Walter, N. G.; Šponer, J., Cations and hydration in catalytic RNA: molecular dynamics of the hepatitis delta virus ribozyme. *Biophys. J.* **2006,** *91* (2), 626-638.
130. Réblová, K.; Šponer, J. E.; Šponer, J.; Šponer, J. i., A-minor tertiary interactions in RNA kink-turns. Molecular dynamics and quantum chemical analysis. *The Journal of Physical Chemistry B* **2011,** *115* (47), 13897-13910.
131. Sklenovský, P.; Florová, P.; Banáš, P.; Réblová, K.; Lankaš, F.; Otyepka, M.; Šponer, J. i., Understanding RNA flexibility using explicit solvent simulations: the ribosomal and group I intron reverse kink-turn motifs. *J. Chem. Theory Comput.* **2011,** *7* (9), 2963-2980.
132. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23* (16), 1623-41.

133. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11* (8), 3696-3713.
134. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *J. Comput. Chem.* **2004,** *25*, 1157–1173.
135. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
136. Price, D. J.; Brooks III, C. L., A Modified TIP3P Water Potential for Simulation with Ewald Summation. *J. Chem. Phys.* **2004,** *121* (20), 10096-10103.
137. Joung, I. S.; Cheatham III, T. E., Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The journal of physical chemistry B* **2008,** *112* (30), 9020-9041.
138. Joung, I. S.; Cheatham, T. E., Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *The Journal of Physical Chemistry B* **2009,** *113* (40), 13279-13290.
139. Hub, J. S.; Groot, B. L. D.; Spoel, D. V. D., g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Phys.* **2015,** *6* (9), 3713-3720.
140. Ono, S.; Nakajima, N.; Higo, J.; Nakamura, H., The Multicanonical Weighted Histogram Analysis Method for the Free-energy Landscape Along Structural Transition Paths. *Chem. Phys. Lett.* **1999,** *312* (2-4), 247-254.
141. Lee, T. S.; Radak, B. K.; Pabis, A.; York, D. M., A New Maximum Likelihood Approach for Free Energy Profile Construction from Molecular Simulations. *J Chem Theory Comput* **2013,** *9* (1), 153-164.
142. Zheng, H.; Cai, Y.; Ding, S.; Tang, Y.; Kropachev, K.; Zhou, Y.; Wang, L.; Wang, S.; Geacintov, N. E.; Zhang, Y., Base Flipping Free Energy Profiles for Damaged and Undamaged DNA. *Chem. Res. Toxicol.* **2010,** *23* (12), 1868-1870.
143. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L., Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2015,** *29* (5), 397-411.
144. Mallick, K.; Moshe, M.; Orland, H., Supersymmetry and Nonequilibrium Work Relations. *arXiv preprint arXiv:0711.2059* **2008**.
145. Bennett, C. H., Efficient Estimation of Free Energy Differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.
146. Wang, X.; Sun, Z., A Theoretical Interpretation of Variance-based Convergence Citeria in Perturbation-based Theories. *arXiv preprint arXiv:1803.03123* **2018**.
147. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of The Cartesian Equations of Motion of A System with Constraints: Molecular Dynamics of n -alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327-341.
148. Miyamoto, S.; Kollman, P. A., Settle: An Analytical Version of The SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992,** *13* (8), 952-962.
149. Pastor, R. W.; Brooks, B. R.; Szabo, A., An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **1988,** *65* (6), 1409-1419.
150. York, D. M.; Darden, T. A.; Pedersen, L. G., The Effect of Long-range Electrostatic Interactions in Simulations of Macromolecular Crystals: A Comparison of The Ewald and Truncated List Methods. *J. Chem. Phys.* **1993,** *99* (10), 8345-8348.
151. Case, D. A.; Cheatham, T. E.; Tom, D.; Holger, G.; Luo, R.; Merz, K. M.; Alexey, O.; Carlos, S.; Bing, W.; Woods, R. J., The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005,** *26* (16), 1668-1688.

Table 1. The summarization of the simulation times in umbrella sampling simulations in the current work. The wild type denotes the T group in DNA, while the mutation TFT represents the mutated T group with 3'-2' linkages in TNA. The net-neutral salt concentration means the counter-ions are added to neutralize the system, while the excess-salt salt concentration indicates that aside from the salts to achieve the net-neutral condition, excess salts (sodium chloride ion pairs) are added to achieve an excess-salt concentration of 0.09 M, corresponding to the experimental condition. Here the equilibrium sampling time only includes the production run after the initial equilibration in each umbrella window.

salt concentration	Terms / System	equilibrium umbrella sampling		
		Sampling time in each umbrella window (ns)	Number of umbrella windows	Total simulation time (ns)
net-neutral	OL15 wild type	32	72	2304
	OL15 mutation TFT	32		2304
	bsc1 wild type	28		2016
	bsc1 mutation TFT	36		2592
	bsc0 wild type	28		2016
	bsc0 mutation TFT	32		2304
excess-salt	OL15 wild type	36		2592
	OL15 mutation TFT	40		2880
	bsc1 wild type	32		2304
	bsc1 mutation TFT	32		2304
	bsc0 wild type	40		2880
	bsc0 mutation TFT	32		2304

Table 2. For the net-neutral systems, the comparison between the minimum simulation times required for converged PMF estimates from equilibrium umbrella sampling and nonequilibrium stratification. In equilibrium umbrella sampling, there are 72 umbrella windows and in each window 4 ns sampling time is required. Thus, the sampling time is about 288 ns. In nonequilibrium stratification, the total sampling time in nonequilibrium stratification is given by $N_{\text{segments}} * N_{\text{traj}} * (\phi_{\text{NEW}} + \phi_{\text{eq}})$, where N_{segments} denotes the number of segments, N_{traj} represents the number of realizations per 2° segment, ϕ_{eq} is the statistical inefficiency in the equilibrium ensemble, and ϕ_{NEW} is the pulling time during the nonequilibrium realization. In each pulling segment, 50 samples are required for converged PMF estimation. Thus, the minimum sampling time in nonequilibrium stratification is about 27 ns.

Terms / Systems	equilibrium umbrella sampling			nonequilibrium stratification				
	Sampling time in each umbrella window (ns)	Number of umbrella windows	Total simulation time (ns)	ϕ_{eq} for each initial configuration (ps)	ϕ_{NEW} in each segment (ps)	Number of segments	Number of realizations per segment	Total simulation time (ns)
OL15 wild type OL15 mutation TFT bsc0 wild type bsc0 mutation TFT	4	72	288.00	2.00	0.5x2=1	180.00	50.00	27.00

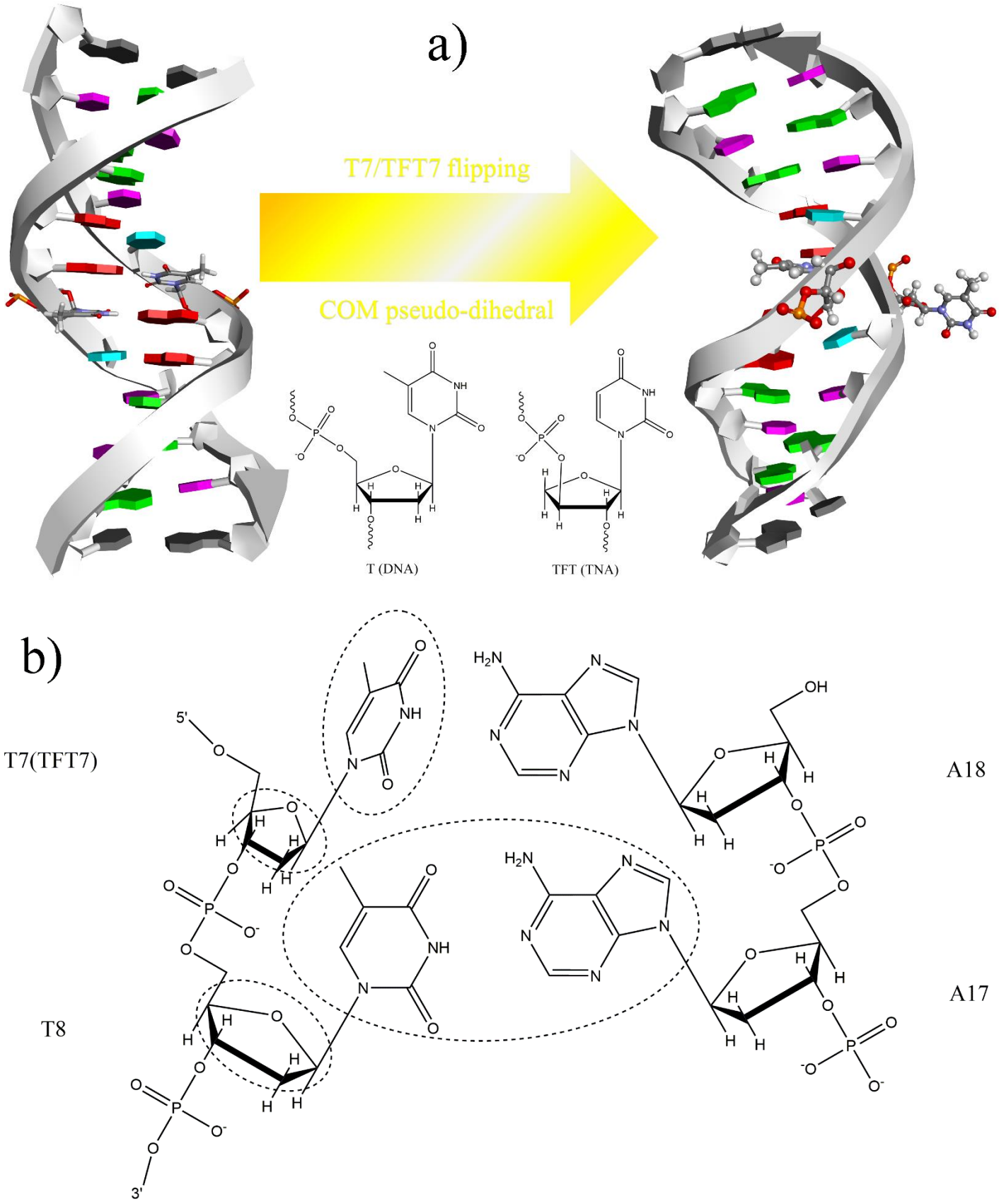


Fig. 1. a) An illustration of the flipping of the 7th residue of T/TFT at the middle of the TNA/DNA duplex and the comparison between the T residue in DNA and the TFT residue in TNA. b) The definition of the CV used to describe the flipping event. The COMs of the 4 groups (heavy atoms only) in circle are used to define the

dihedral.

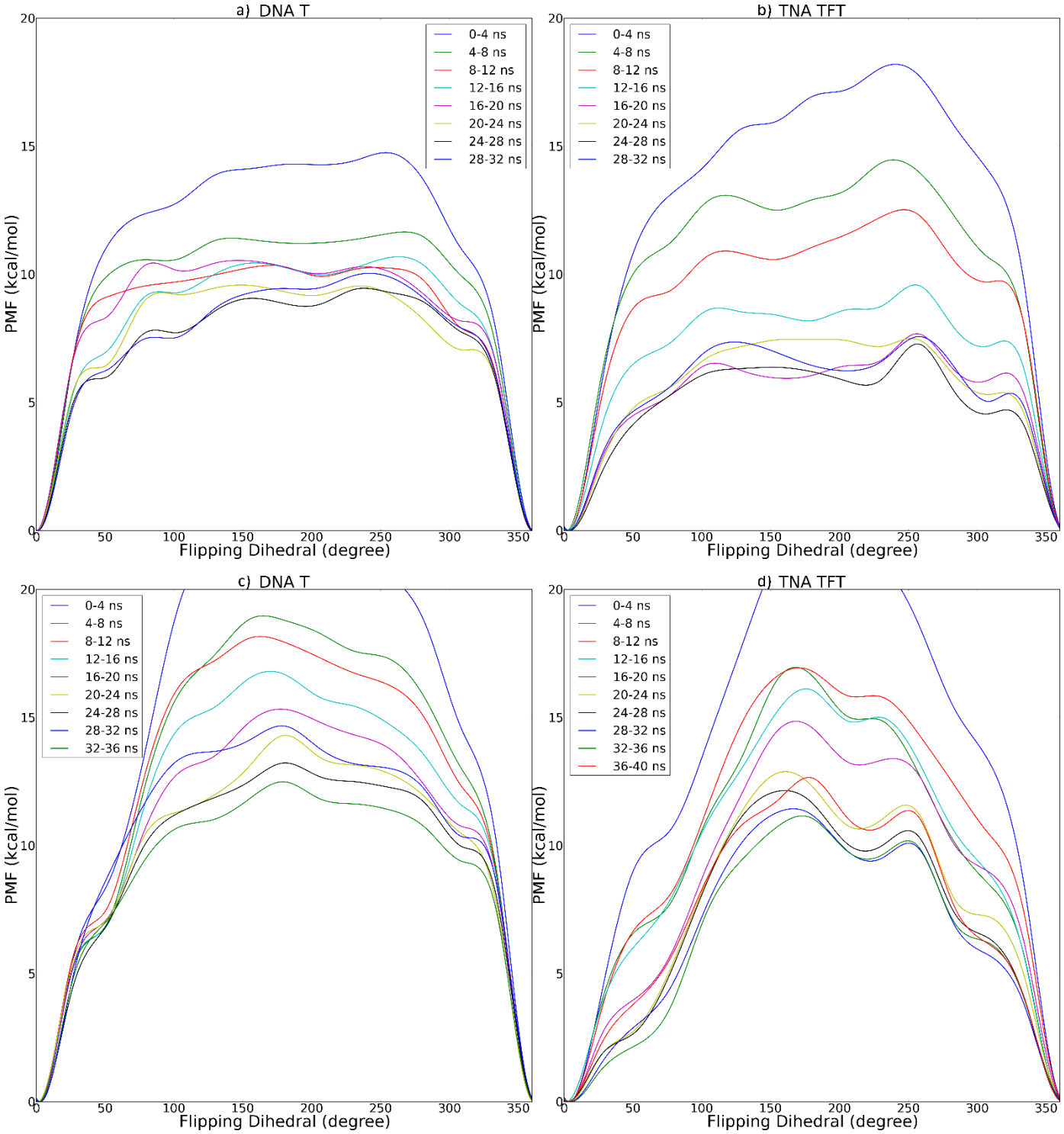
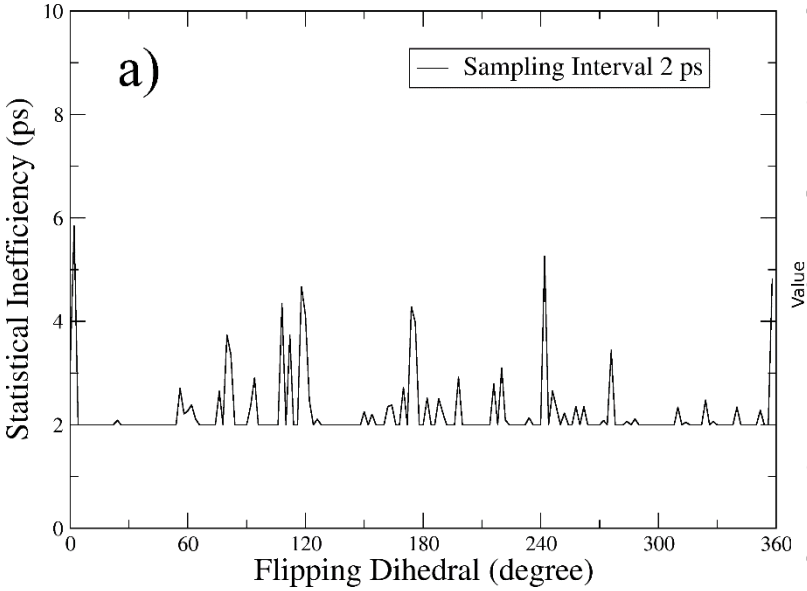
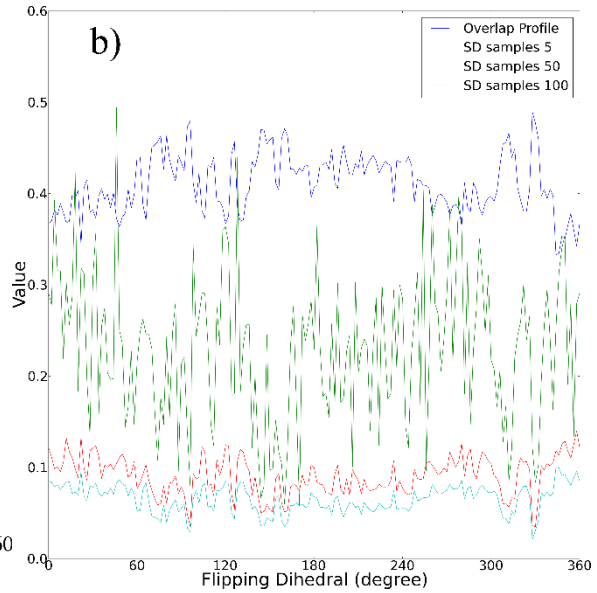
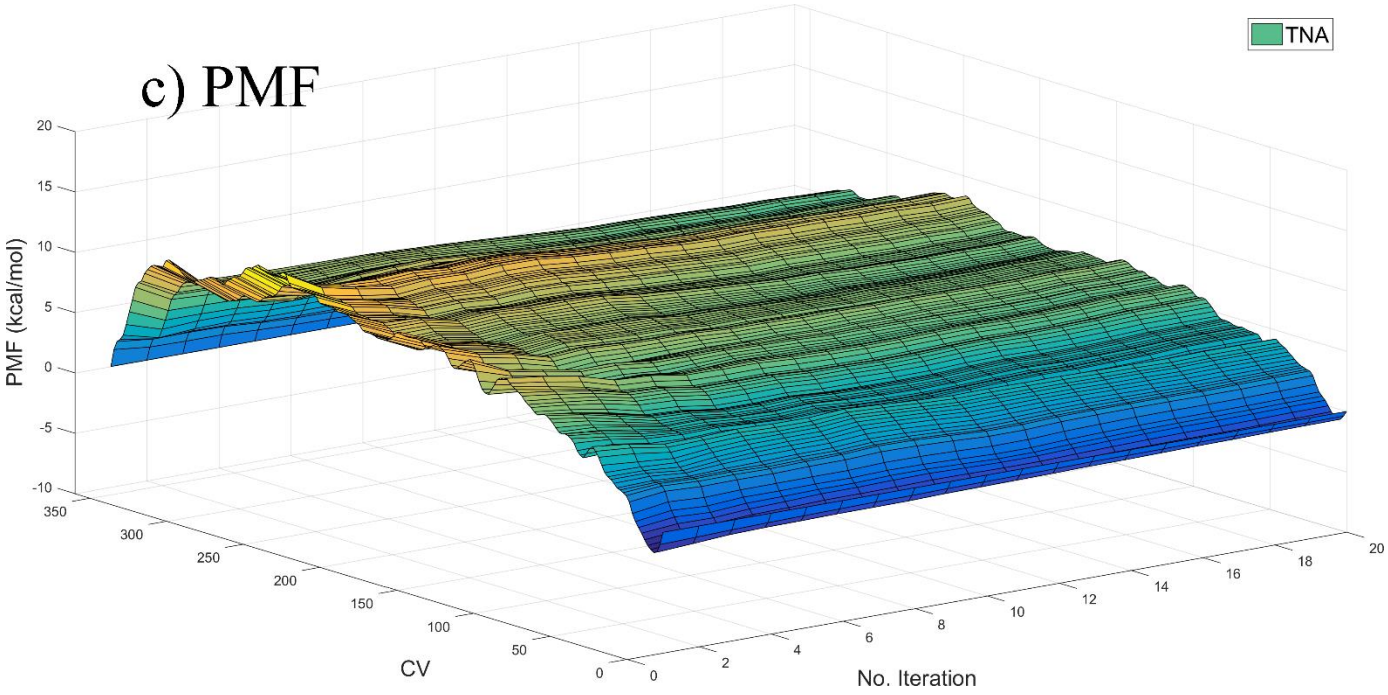


Fig. 2. Under the force field of OL15, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a-b) net-neutral and c-d) excess-salt simulations.


a)
Sampling Interval 2 ps
Statistical Inefficiency (ps)
Flipping Dihedral (degree)


b)
Overlap Profile
SD samples 5
SD samples 50
SD samples 100
Value
Flipping Dihedral (degree)


c) PMF
TNA
PMF (kcal/mol)
CV
No. Iteration

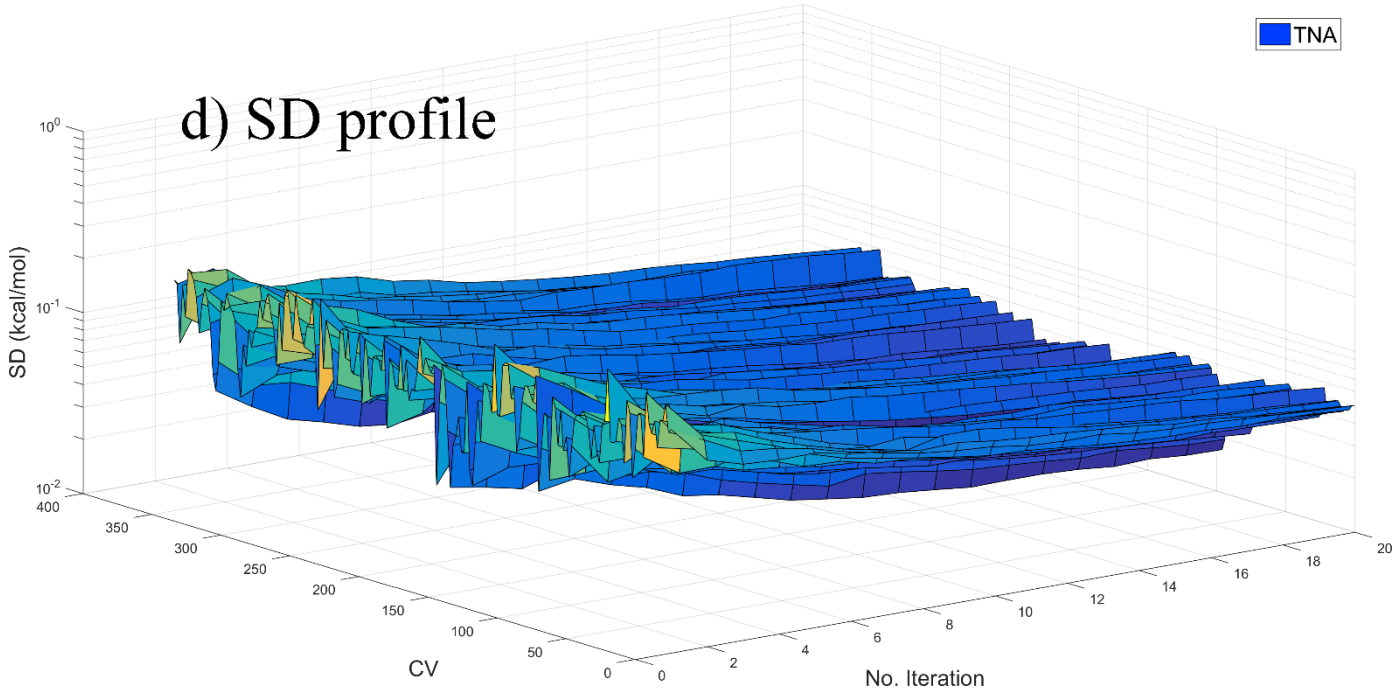


Fig. 3. In the TNA mutant described with the OL15 force field in net-neutral simulations, a) the statistical inefficiency in each configurational state, b) the comparison between dimensionless SD profiles and the overlap profile in nonequilibrium stratification, c) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, and d) the time-evolution of state-specified SD. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$, where n_i represents the number of independent samples in state i, f refers to the Fermi function, $f_{ji} = \left\langle f(W_{ji} - C_{ij}) \right\rangle_j$, $f_{ij} = \left\langle f(W_{ij} + C_{ij}) \right\rangle_i$, j equals i+1, and W_{ij} is the reduced (dimensionless) work for nonequilibrium pulling initiated from state i and ended in state j. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are obvious differences between the 5-sample PMF and the later ones. Since the 10th iteration (i.e. 50 samples), the fluctuation of the PMF is very small, the SDs in all states decrease monotonically with further sampling, and the dimensionless SD profile is much smaller than the overlap profile. Thus, we define this sample size (50 samples) as the minimum sample size required for convergence.

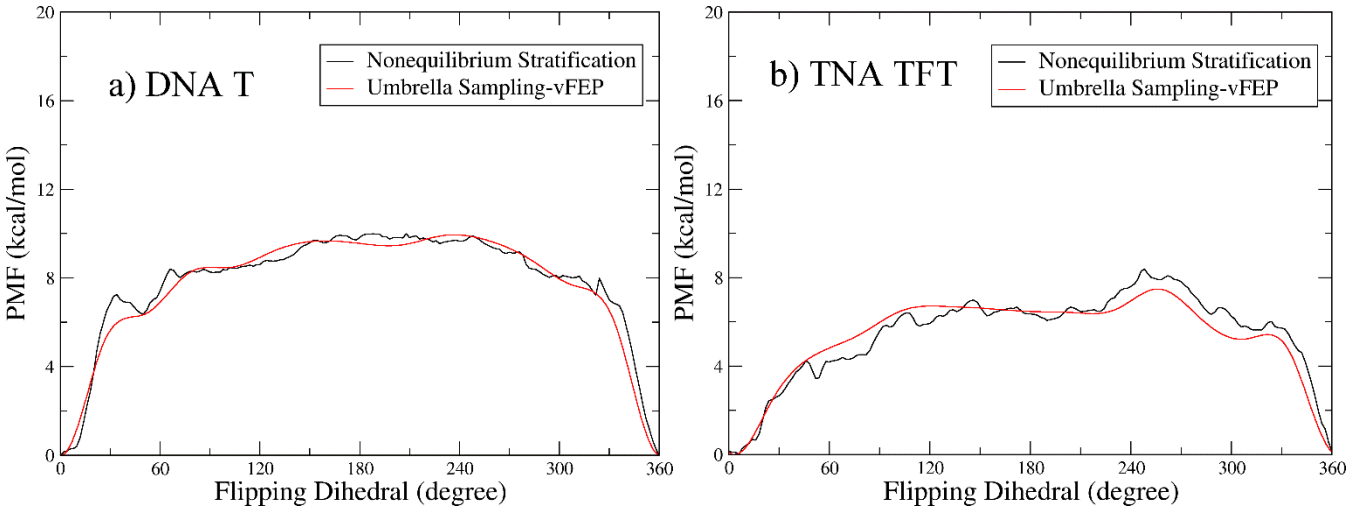


Fig. 4. Under the OL15 force field, the comparisons between the free energy profiles of base flipping in a) the wild-type DNA duplex and b) the TNA mutant constructed from nonequilibrium stratification and equilibrium umbrella sampling with vFEP reweighting.

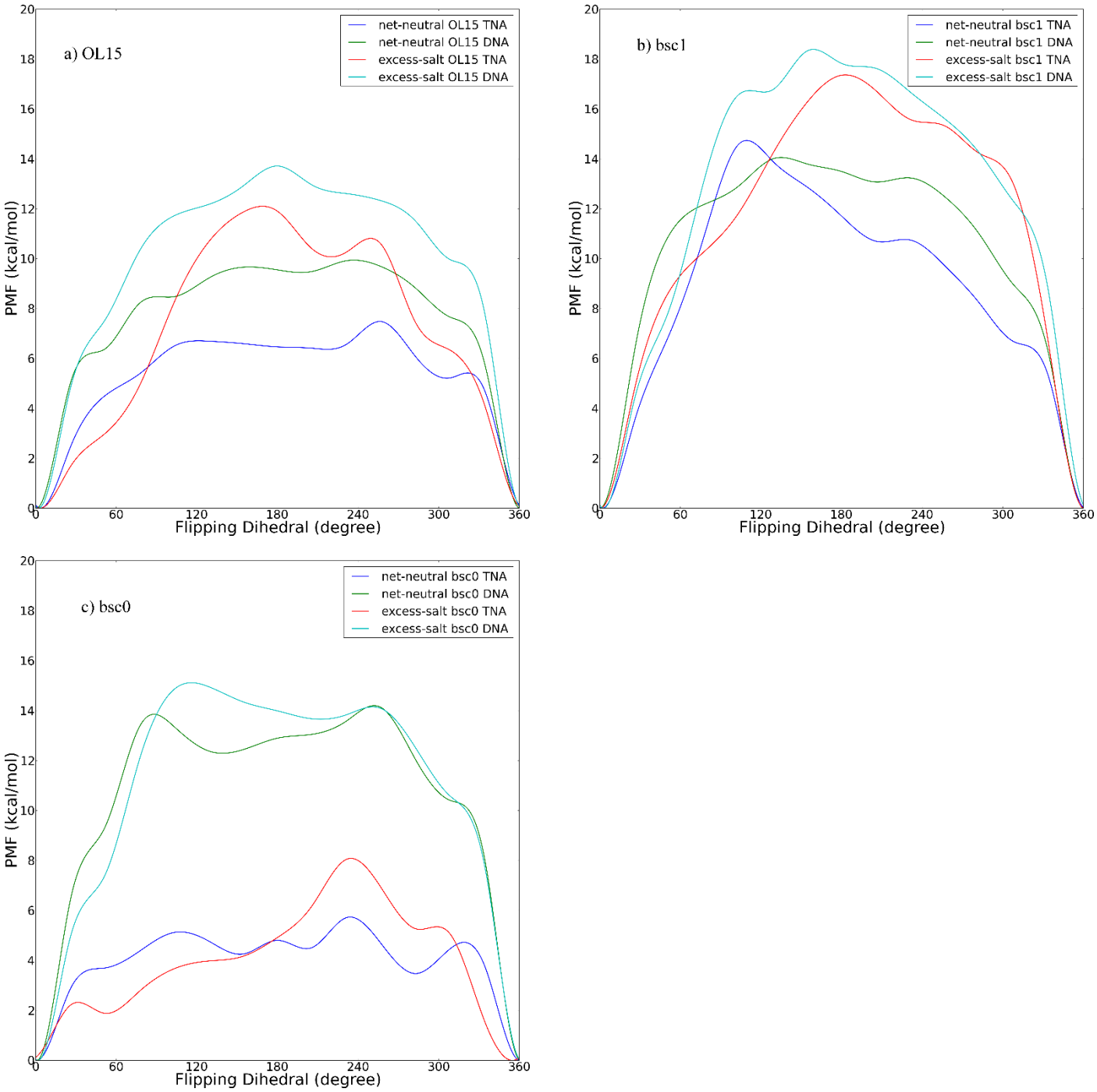
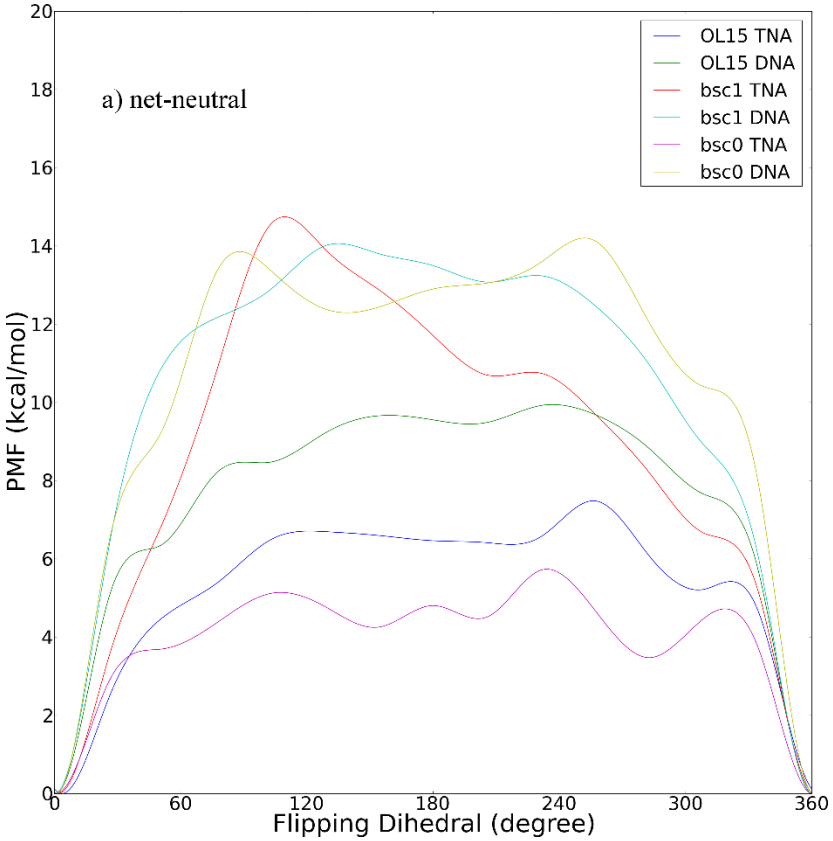


Fig. 5. Comparison between the free energy profiles of the wild-type DNA and the TNA mutant described with the same force field under different ion concentrations (i.e. net-neutral and excess-salt simulations). a) OL15, b) bsc1, and c) bsc0.



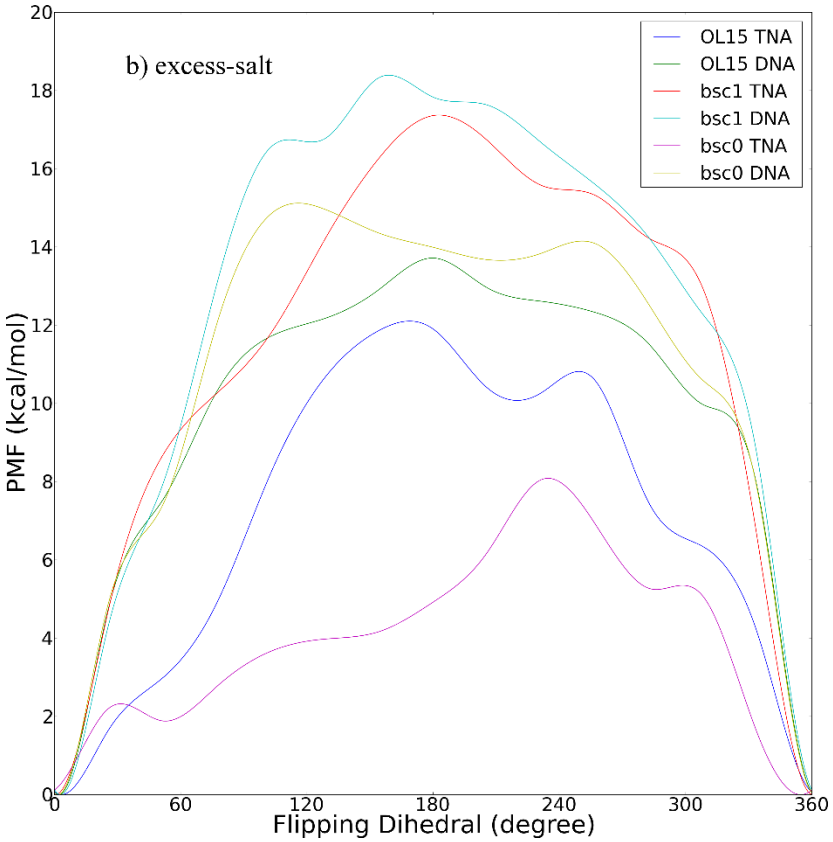


Fig. 6. Comparison between the free energy profiles of the wild-type DNA and the TNA mutant described with the different force fields under the same ion concentration. a) net-neutral and b) excess-salt simulations.

Supporting Information: Thermodynamic Insights of Base Flipping in TNA Duplex: Force Fields, Salt Concentrations, and Free Energy Simulation Methods

Zhaoxi Sun[1*], and John Z.H. Zhang[1,2,3*]

[1]*State Key Laboratory of Precision Spectroscopy, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China*

[2]*NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China*

[3]*Department of Chemistry, New York University, NY, NY 10003, USA*

*To whom correspondence should be addressed: proszx@163.com, john.zhang@nyu.edu

Fig. S1. Under the force field of bsc1, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a-b) net-neutral and c-d) excess-salt simulations.

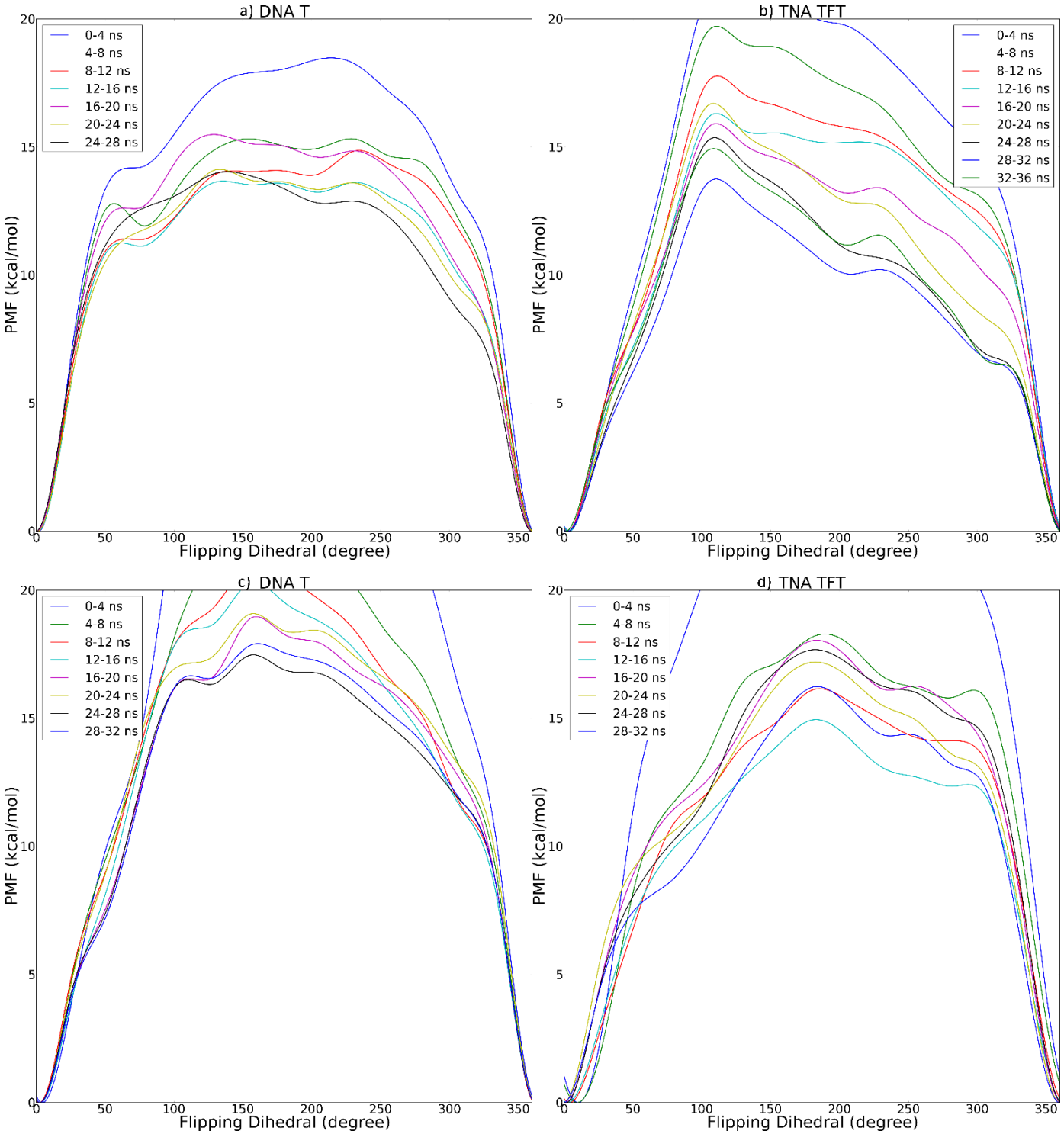


Fig. S2. Under the force field of bsc0, the convergence behavior of free energy profiles constructed from equilibrium umbrella sampling simulations with vFEP reweighting in a-b) net-neutral and c-d) excess-salt simulations.

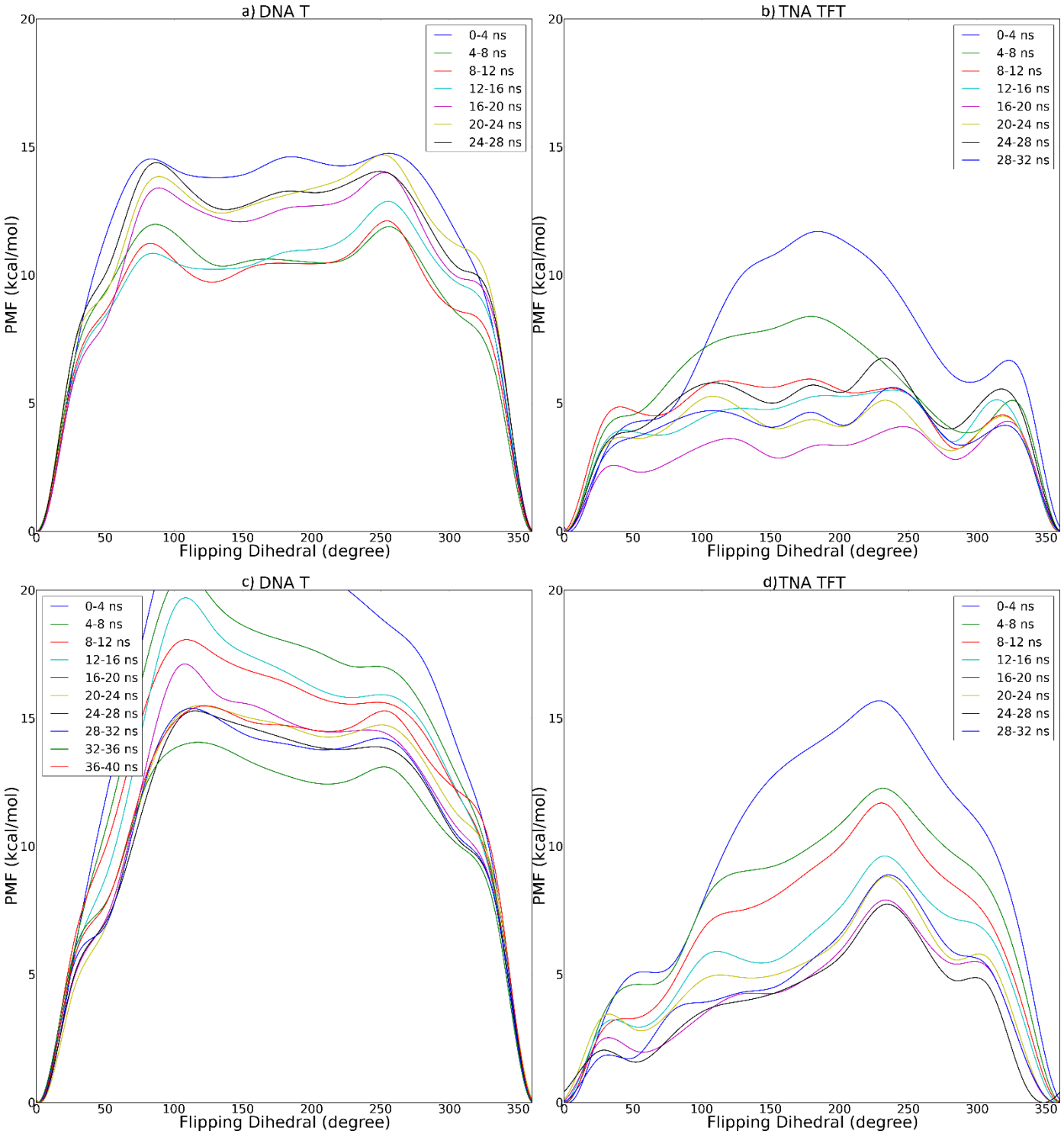
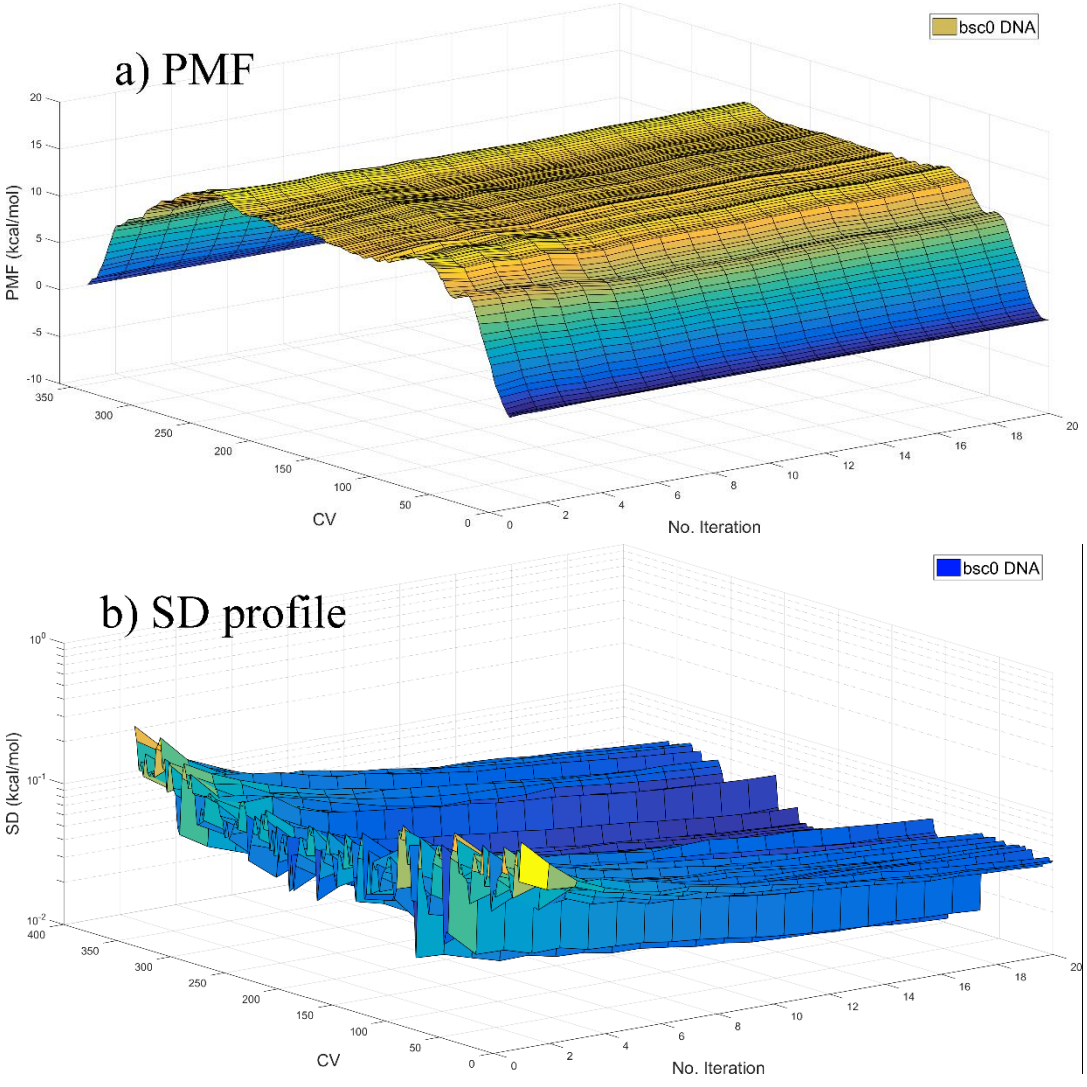


Fig. S3. In the wild-type DNA system described with the bsc0 force field in net-neutral simulations, a) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, b) the time-evolution of state-specified SD, and c) the comparison between dimensionless SD profiles and overlap profile in nonequilibrium stratification. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$, where n_i represents the number of independent samples in state i, f refers to the Fermi function, $f_{ji} = \left\langle f(W_{ji} - C_{ij}) \right\rangle_j$, $f_{ij} = \left\langle f(W_{ij} + C_{ij}) \right\rangle_i$, j equals i+1, and W_{ij} is the reduced (dimensionless) work for nonequilibrium pulling initiated from state i and ended in state j. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are obvious differences between the 5-sample PMF and the later ones. Since the 10th iteration (i.e. 50 samples), the fluctuation of the PMF is very small. To avoid exaggeration of the performance of the nonequilibrium technique, we define this sample size (50 samples) as the minimum sample size required for convergence.



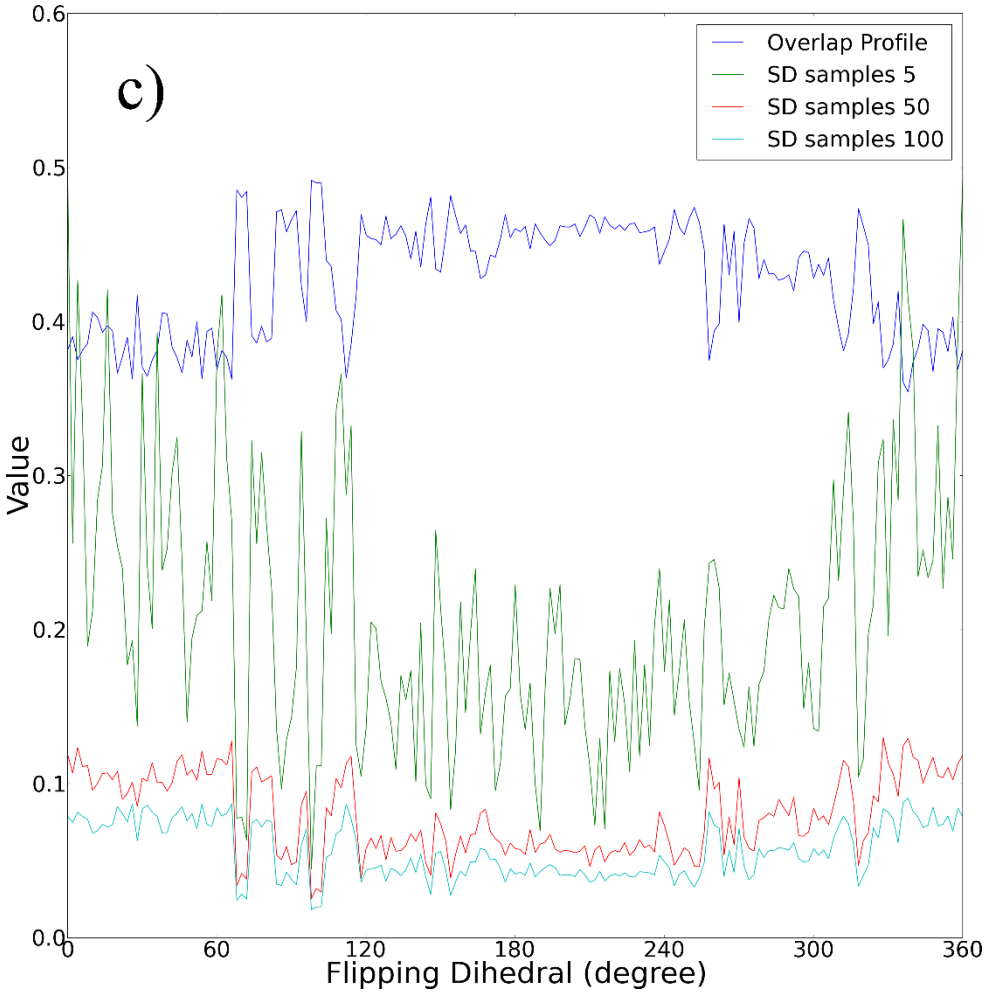
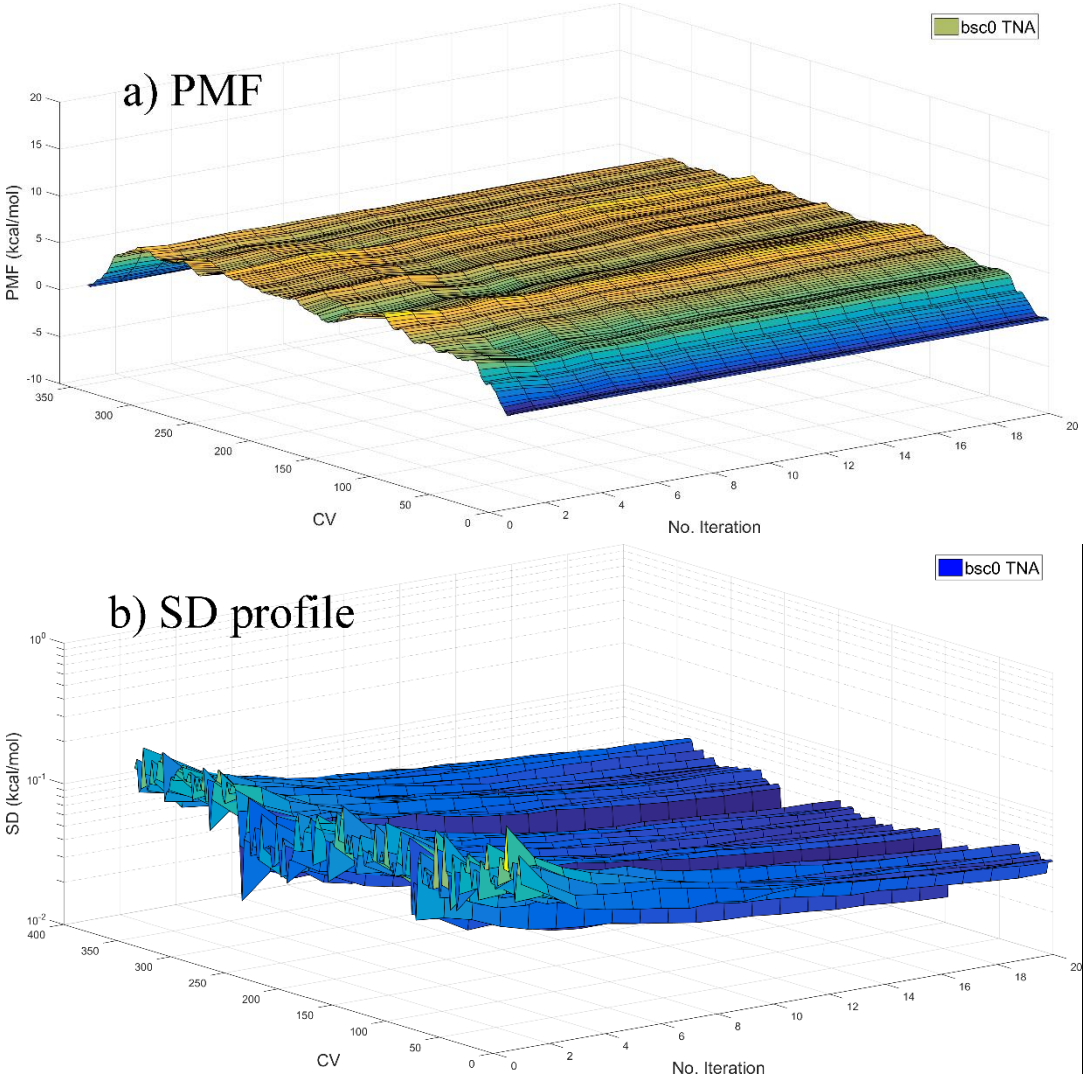

c)
Overlap Profile
SD samples 5
SD samples 50
SD samples 100
Value
Flipping Dihedral (degree)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
60
120
180
240
300
360

Fig. S4. In the TNA mutant described with the bsc0 force field in net-neutral simulations, a) the convergence behavior of free energy profiles from nonequilibrium stratification on the sample size, b) the time-evolution of state-specified SD, and c) the comparison between dimensionless SD profiles and overlap profile in nonequilibrium stratification. The standard deviation in the *i*th state is the sum of the components contributed by samples initiated from that state, namely $\sigma_i = \sqrt{\frac{\mathrm{Var}\left(f_{i,i+1}\right)}{n_i^2 f_{i,i+1}^2} + \frac{\mathrm{Var}\left(f_{i-1,i}\right)}{n_i^2 f_{i-1,i}^2}}$, where n_i represents the number of independent samples in state i, f refers to the Fermi function, $f_{ji} = \left\langle f(W_{ji} - C_{ij}) \right\rangle_j$, $f_{ij} = \left\langle f(W_{ij} + C_{ij}) \right\rangle_i$, j equals i+1, and W_{ij} is the reduced (dimensionless) work for nonequilibrium pulling initiated from state i and ended in state j. The initial sample size is 5 and in each iteration further 5 samples are added to the dataset. There are obvious differences between the 5-sample PMF and the later ones. Since the 10th iteration (i.e. 50 samples), the fluctuation of the PMF is very small. To avoid exaggeration of the performance of the nonequilibrium stratification method, we define this sample size (50 samples) as the minimum sample size required for convergence.



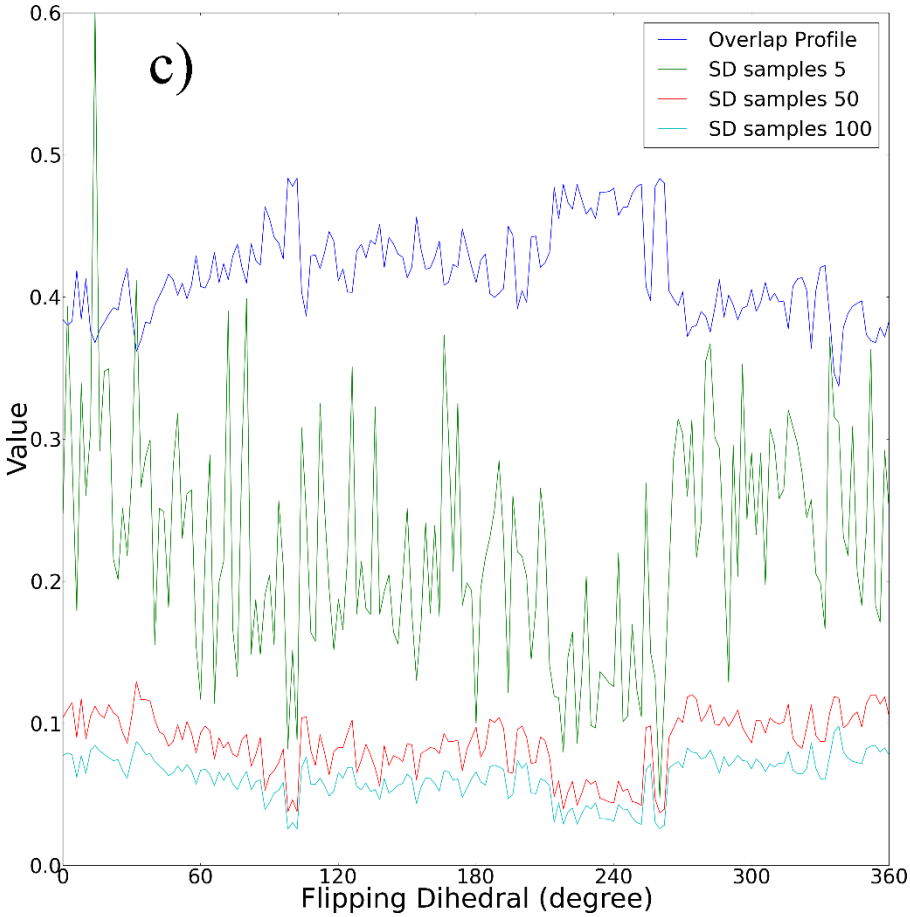

c)
Overlap Profile
SD samples 5
SD samples 50
SD samples 100
Value
Flipping Dihedral (degree)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
60
120
180
240
300
360

Fig. S5. Under the bsc0 force field, the comparisons between the free energy profiles of base flipping in a) the wild-type DNA duplex and b) the TNA mutant constructed from nonequilibrium stratification and equilibrium umbrella sampling with vFEP reweighting.