Investigation of Dynamical flexibility of D5SIC-DNAM inside DNA duplex in Aqueous Solution: A Systematic Classical MD Approach

Tanay Debnath[1], G. Andres Cisneros[1,2,*]

[1]Department of Physics, University of Texas at Dallas, Richardson, TX, USA

[2]Department of Chemistry and Biochemistry, University of Texas at Dallas, Richardson, TX, USA

ABSTRACT: Incorporation of artificial 3rd base pairs (Unnatural base pairs, UBPs) has emerged as a fundamental technique in pursuit of expanding the genetic alphabet. 2,6-dimethyl-2H-isoquiniline-1-thione: D5SIC (DS) and 2-methoxy-3-methylnaphthalene: DNAM (DN), a potential unnatural base pair (UBP) developed by Romesberg and colleagues, has been shown to have remarkable capability for replication within DNA. Crystal structures of a *Taq* polymerase/double-stranded DNA (ds-DNA) complex containing a DS-DN pair in the 3' terminus showed a parallelly stacked geometry for the pre-insertion, and an intercalated geometry for the post-insertion structure. Unconventional orientations of DS-DN inside a DNA duplex have inspired scientists to investigate the conformational orientations and structural properties of UBP-incorporated DNA. In recent years, computational simulations have been used to investigate the geometry of DS-DN within the DNA duplex; nevertheless, unresolved questions persist owing to inconclusive findings. In this work, we investigate the structural and dynamical properties of DS and DN inside a ds-DNA strand in aqueous solution considering both short and long DNA templates using polarizable, and non-polarizable classical MD simulations. Flexible conformational change of UBP with major populations of Watson-Crick-Franklin (WCF) and three distinct non-Watson-Crick-Franklin (nWCFP1, nWCFP2, nWCFO) conformations through intra and inter-strand flipping have been observed. Our results suggest that a dynamical conformational change leads to the production of random distribution of several intermediates. Simulations with a short ds-DNA duplex suggest nWCF (P1 and O) as the predominant structures, whereas long ds-DNA duplex simulations indicate almost equal populations of WCF, nWCFP1, nWCFO. DS-DN in the terminal position is found to be more flexible with occasional mispairing and fraying. Overall, these results suggest flexibility and random conformational distribution of the UBP as well as indicate varied conformational distribution with the increase of the length of the DNA strand.

Introduction

In every DNA-based organism, genetic information is represented by a four-letter genetic alphabet composed of deoxyadenosine (dA), deoxyguanosine (dG), deoxycytidine (dC), and deoxythymidine (dT)[1]. The storage and retrieval of this information depend on the formation of two base pairs, (d)A-dT/U and (d)G-(d)C. Synthetic biology[2], which emerged over a century ago, aims to create new biological forms with potential applicability towards biomedical and bio engineering fields[3]. One promising approach to achieving this goal is to expand the amount of information that can be stored and retrieved in a cell[4, 5]. As a result, scientists have dedicated considerable effort over the last decade to discovering a fifth and sixth nucleotides that can form a third, unnatural base pair (UBP) with increased functionality that can be orthogonally replicated in DNA[4, 6-37]. This would also expand the usefulness of nucleic acids for biological and biotechnological applications.

While several unnatural base pair candidates have been identified, only few have been shown to be able to be efficient for central dogma process[24--37], among them, d5SICS (DS)- dNaM (DN) reported by Romesberg and co-workers, have been PCR amplified without sequence bias and efficiently transcribed in both directions.[32] What makes the DS-DN UBP particularly interesting is that it forms an intercalated structure in duplex DNA and does not rely on complementary hydrogen bond formation for inter-strand pairing. The underlying cause of the unconventional orientations of DS-DN remains uncertain. It is yet to be verified whether the polymerase stabilizes the structure of the UBP-incorporated DNA, or if the UBP itself can be stabilized inside the DNA duplex.

Betz et al. reported several crystal structures of the large fragment of *T. Aquaticus* (*Taq)* DNA polymerase with bound UBP-incorporated DNA duplex. Among these, for the structure in the pre-insertion phase, the UBP are observed to be arranged in a parallel stack (WCF orientation), whereas in the post-insertion structure the DS-DN pair adopts an intercalated structure (nWCF) inside the DNA duplex[32,38]. Several research groups have used computational tools to investigate structural properties of the UBP-incorporated DNA[39-45]. Datta *et al*. investigated the structure of DS-DN incorporated DNA through both QM and MD simulations and showed that the DS-DN distance is found to be consistent with a WCF pairing pattern during MD simulations[39]. Wetmore et al. considered UBP incorporated three nucleotides–long double strands and observed that for DS and DN, QM calculations suggest DS and DN adopt an intercalated nWCF structure, whereas a planar WCF–like configuration has been predicted through MD simulations[43]. Barroso-Flores et al. have reported several conformers of DS and DN inside DNA duplex through their extensive MD simulations[44, 45]. They concluded that an equilibrated structure of DS-DN incorporated DNA duplex may not have been achieved due to sampling time and/or forcefield incompatibility. In summary, unlike natural base pairs, UBPs appear to adopt different conformations in aqueous solution and during different steps of the replication processes as observed from X-Ray crystal structures and computational simulations.

Under these circumstances we have divided our work into two parts. In the first part we investigate the UBP-incorporated DNA duplex in aqueous solution, aiming to investigate the inherent structural attributes of DS-DN within DNA duplex strands through classical MD simulations. Recognizing the significance of non-covalent interactions in UBP stabilization, we have carried out simulations with the multipolar/polarizable AMOEBA forcefield, alongside the fixed-charge non-polarizable AMBER force field. The organization of the paper is as follows: In the next section, we describe the development of forcefield parameters for UBP and simulations details. Next, we discuss the results for the investigation of dynamical structural properties of UBP-incorporated DNA by placing DS-DN in the middle of a dsDNA strand with different orientations with both short and long fragments of DNA. Subsequently, we describe the investigation of the same with UBP placed in terminal position within the DNA, followed by concluding remarks.

Computational Methods

DFT calculations

All gas phase geometry optimizations for the UBs (DS and DN) and UBP (DS-DN) have been performed using Gaussian 16 A.03[46] at the ωB97X-D[47]/6-311++G(d,p)[48, 49] level. Symmetry Adapted Perturbation Theory-DFT, SAPT(DFT)[50] analysis has been done using the PSI4 1.2 software package[51]. NCIPLOT[52] has been employed to investigate the topology of non-covalent interactions between DS and DN. To predict the stability of the UBP in gas phase, we have calculated interaction energy (IE_{UBP}) of the complexes by employing the following equation.

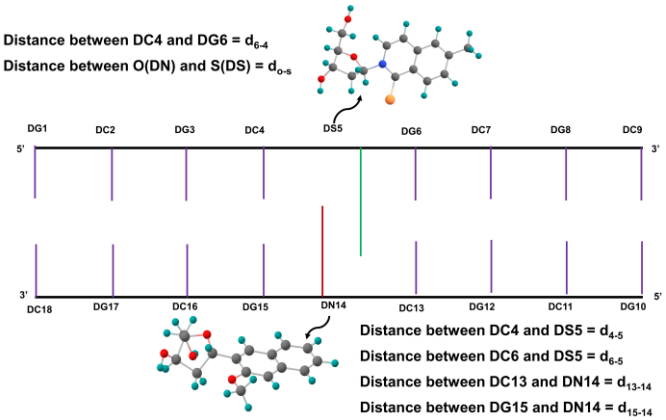
$$IE_{UBP} = E_{UBP}^{opt} - E_{DS}^{Frag} - E_{DN}^{Frag}$$

Molecular Dynamics (MD) Simulations

System Setup

The DNA templates considered here are represented in Scheme 1. The UBP is placed inside the DNA duplex in two possible ways. In one case DS-DN is incorporated into the middle of a 9-mer DNA duplex designated as **MUD (**5'-GCGC**DS**GCGC-3', Scheme 1). In the MUD structure a DS-DN pair has been placed with different orientations. The Non-Watson-Crick DNA models have been created through placement of the UBP as intercalated forms, denoted MUD_{SYN} and MUD_{ANTI}. The parallel model, corresponding to a canonical Watson-Crick DNA duplex, is denoted MUD_{PAR}.

Selected geometrical parameters associated with the UBP have been monitored including the UBP distance (d_{DS-DN}), which has been calculated by measuring the length between C1' of DS and DN. We have also calculated (UB-NB) (NB=Natural Base), for DS the calculated distances are DC4-DS5 (d_{4-5}) and DG6-DS5 (d_{6-5}) whereas for DN, DC13-DN14 (d_{13-14}) and DG15-DN14 (d_{15-14}) are the calculated distances (Scheme 1). The distance between DC4-DG6(d_{4-6}) is also calculated. The distance between sulphur of DS and oxygen (-OMe) of DN is designated as d_{O-S}. Parameters related to angles have been calculated (<NB-UB-UB) to predict the conformational change of the UBP inside the DNA duplex. The measured angles include <DC4DS5DN14(a_{4-5-14}), <DG6DS5DN14(a_{6-5-14}) for DS and <DC13DN14DS5($a_{13-14-5}$) and <DC15DN14DS5($a_{15-14-5}$) for DN.



Scheme 1. Schematic representation of DS-DN incorporated DNA duplex of A) MUD, B) MUDL and C) UUD structures with symbolic representation of distances and angles.

Further, we have considered long DNA (MUDL) with 21 base pairs (scheme 1) having UBP positioned in the middle of the DNA to investigate the size effect on the stability of the UBP incorporated DNA. We have also investigated the UBP incorporated DNA by placing the DS-DN in the 3' terminal position, designated as **UUD**.

General MD Setup

The DS and DN parameters have been calculated with the PYRED program[53] to generate AMBER parameters[53-52]. For the DS and DN AMOEBA parameters[57-61] we have used the parametrization tools available in the TINKER software[62] in tandem with GDMA 2.3 for the atomic multipoles[63] for multipole generation (all parameters are provided in Supplementary Information).

The LEaP module[64] in AMBER20[65] was used to set up the simulation box with UBP-incorporated DNA duplex in water. Neutralization of the system with the required number of counterions (Na^+), and solvation of the system in a cubic box filled with TIP3P water[56], extending at least 12 Å from the DNA duplex was done with the LEAP module in AMBER. All MD simulations were performed with the AMBER20 pmemd.cuda program using the Ol15 AMBER force field.[54] Seven minimization steps were done with decreasing restraint (10.0–0.0 kcal mol−1 $Å^{-2}$) on the solute's heavy atoms. In each stage, the system was minimized with 5000 cycles of minimization of steepest descent, followed by 5000 cycles of conjugate gradient minimization. Subsequently, each system was heated to 300 K using Langevin dynamics[66-67] with a collision frequency of 2 ps^{-1} followed by 7 ns of NVT equilibration with decreasing restraints (10.0–0.0 kcal mol^{-1} $Å^{-2}$) on the system's heavy atoms every ns. Production calculations for each system were performed for 1 μs in the NPT ensemble without

restraints in triplicate–a total of 3 μs for each system. Total simulated time is 9 μs for all MUD structures, 3 μs for the MUDL structure, and 3 μs for the UUD structure. Long-range Coulomb interactions[68] were handled with the smooth particle mesh Ewald method[69-70] using a 10 Å cutoff for real-space non-bonded interactions. The CPPTRAJ module[71] in AMBER18 was used to analyze production dynamics, i.e., RMSD, RMSF and geometrical parameters. In addition, Python libraries NumPy[72], Matplotlib[73], Pandas[74], were also employed for further data processing and graphing. Energy decomposition analysis (EDA) has been employed to investigate the intermolecular interactions between the UBP and residues of the rest of the systems. An in-house Fortan90-based EDA code was employed to calculate the nonbonded intermolecular interaction energies[75].

All simulations with the polarizable AMOEBA[59] (Atomic Multipole Optimized Energetics for Biomolecular) force field were performed with the TINKER HP software.[61] The systems were built using the packmole[76] software. Initially, UBP-incorporated DNA duplex complex was minimized using the BFGS nonlinear optimization algorithm with a convergence criterion (RMS gradient) of 0.1 Å. Subsequently, relaxation via MD in vacuum followed by implicit water with the GBSA model for 2 ns to obtain the starting system was performed. After that the structure was solvated in explicit water in the center of a box with volume 50X50X50 $Å^3$ containing 24000 water molecules, and neutralized using packmole. Subsequently, the system was heated to 300 K in 4 simulation steps (2 ns each) with an NVT ensemble removing all positional restraints (100.0-0.0 kcal/ $Å^{-1}$). After the equilibration step, MD simulations were carried out for 125 ns in an NPT ensemble (1 atm and 298 K) for 3 replicates each (total simulation time 375 ns). The Monte Carlo barostat and Bussi thermostat were used to maintain the pressure and temperature fixed respectively. The duration of the time step was 2 fs using RESPA integrator. The smooth particle mesh Ewald (PME) method[69] was used in the calculation of charge, atomic multipole, and polarization interactions. A cutoff of 10 Å was used for van der Waals potential energy interactions and the real-space distance cutoff in the Ewald summation.[70]

Results and Discussion

DFT analysis

QM calculations have been carried out to investigate the possible geometries of a UBP comprised by DS and DN. DFT calculations indicate two possible intercalating conformers designated as SYN and ANTI through which DS and DN can interact with one another (Figure 1). In the SYN conformer, the sulfur of DS and the methoxy group of DN are on the same side.

The interaction energy calculations indicate that the SYN conformer (IE_{UBP} = -10.8 kcal/mol) is slightly more stable than the ANTI conformer (IE_{UBP} = -9.1 kcal/mol) (Table S1).

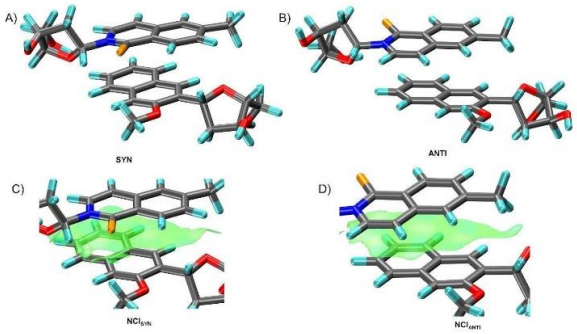


Figure 1. Optimized structures of A) SYN and B) ANTI conformers of DS-DN at ωB97x-D/6-311++g(d,p) level. NCIPLOT of C) SYN and D) ANTI represented the non-covalent interactions between DS and DN in optimized structures.

Our DFT results are consistent with the UBP orientation observed in the post-insertion structure in *Taq*,[32, 38] which show DS and DN form an intercalated structure inside the DNA duplex with the sulfur and methoxy group on the same side. SAPT analysis suggests that the dispersion component is the major contributing factor in total energy to stabilize the UBP as an intercalated structure (Table S1). The NCI index analysis also shows significant non-covalent interaction between DS and DN in the intercalated structures (Figure 1).

AMOEBA Simulations

MD simulations with AMOEBA were performed for both the SYN and ANTI orientations of DS and DN for three replicates, spanning 125 ns each. It has been observed that UBP predominantly forms nWCF structures with occasional flanking and distortion. This distortion leads to the generation of WCF structures on a few occasions during the simulations. It has been noticed that nWCF orientations of the UBP are not static in nature, rather the system explores different geometries. Smaller DS-DN distance generally indicates nWCF geometries of UBP whereas distorted and WCF structures characterized through higher DS-DN distances. Conformational changes of UBP are reflected in $d_{4\text{-}6}$ distances (figure S1); for WCF structures, $d_{4\text{-}6}$ shows reduced distances whereas nWCF and distorted structures display a range of $d_{4\text{-}6}$ (see Scheme 1 for distance definitions) values, which indicates that the flexibility of UBP has an impact on the adjacent BPs.

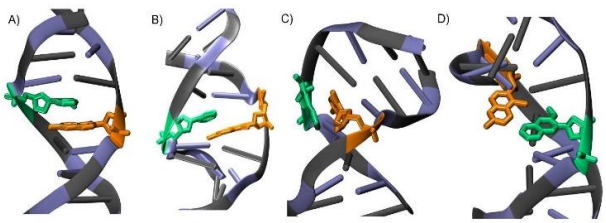


Figure 2. Snapshot of different conformers i.e. A) nWCFP1, B) nWCFP2, C) nWCFO, D) WCF of UBP inside DNA duplex. DS is in orange and DN is in green.

For nWCF orientations, a larger $d_{4\text{-}6}$ represents in-phase placement of UBP whereas a decreased $d_{4\text{-}6}$ denotes outer-phase orientations of UBP inside the DNA duplex. Two distinct structural transformational processes have been observed during the simulations that lead to the generation of several conformers (Figure 2) i) intra-strand flipping which transforms the geometry of the UBP from SYN to ANTI and vice versa, ii) inter-strand flipping leading to reorient the DS and DN upside down (Figure 3).

It should be mentioned that apart from SYN and ANTI, due to dynamical movement of the DS and DN several other intermediates are also produced during the simulations. Inter-strand flipping can be recognized by the pattern shift of UB-NB distance, transitioning between high and low values (Figure 4). Flexibility of the UBP is also synchronized with the RMSD values; distorted UBP-incorporated DNA shows higher RMSD whereas sudden change of RMSD values implies structural transformation. Three sets of replicates have yielded varying conformer distributions, suggesting a stochastic arrangement of the conformers. Overall, AMOEBA force field simulations suggest a dynamical nature of the UBP intercalated ds-DNA, with multiple conformational orientations inside the DNA duplex, where nWCF-DNA structures exhibit the largest occurrence.

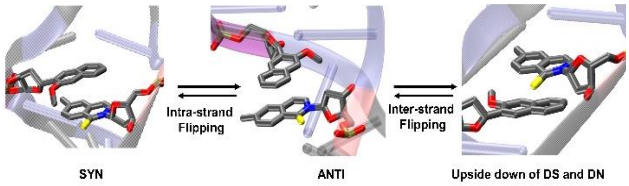


Figure 3. Conformational change through Intra- and Inter-strand flipping observed during the MD simulations.

AMBER simulations

AMBER simulations have been carried out to further investigate the dynamical properties of MUD structure. Here we have considered MUD_{SYN}, MUD_{ANTI} and MUD_{PAR} conformations as the initial intercalated structures. For both cases simulations have been done for 3 replicates with simulation time of 1 μs each.

Structural Analysis of MUD

MUD_{SYN}: Staring from the nWCF conformation with SYN orientation, the simulated outcomes indicate a predominant population of nWCF conformers, with occasional occurrence of WCF structures (see below). Delving into the nWCF structures, it becomes evident that their orientations are not static; instead, the system explores various conformers. Interestingly, dynamical conformational characteristics of UBP predicted by the AMBER forcefield align closely with the results derived from the AMOBA based simulations discussed above.

With these frameworks, the majority of nWCF structures falls into three distinct categories: in-phase-intercalation1

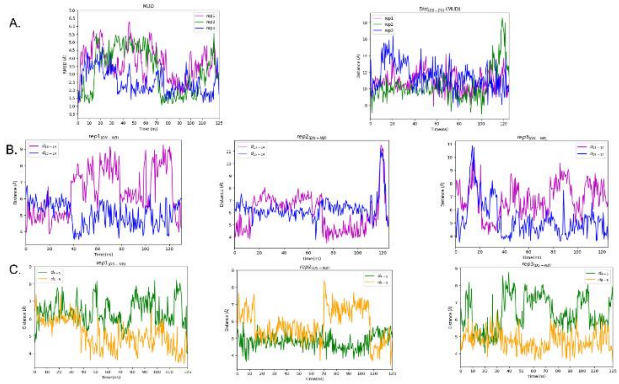


Figure 4. A) RMSD values, B) DS-DN distance values of three replicates, B) DN-NB values obtained from 125 ns simulation employing AMOEBA forcefield.

(nWCFP1), in-phase-intercalation2 (nWCFP2), and outer-phase-intercalation (nWCFO) (Figure 2). Calculated $d_{4\text{-}6}$ distances in nWCFP1 structures is found to be notably high (Figure S2), facilitating the accommodation of the intercalated UBP inside the DNA duplex. For nWCFP2, reduced $d_{4\text{-}6}$ has been observed and it becomes lowest for nWCFO insisting the UBP to settle at the outer phase of the DNA. Interestingly in these two cases occasional distortion has been noticed leading to form mis-paired and flanked structures. During the simulation, SYN-to-ANTI transformation or vice versa through intra-strand flipping is witnessed whereas DS-DN are found to be upside down their position through inter-strand flipping (Figure 4). SYN conformers can be recognized by shorter $d_{O\text{-}S}$ whereas higher $d_{O\text{-}S}$ represent ANTI and distorted orientations of UBP (Figure S2).

Generation of several conformers with different orientations suggests flexibility of the UBP inside the DNA duplex. Higher RMSF for DS and DN further confirms the flexible nature of the UBP (Figure 5). RMSD values of the entire DNA are found to be synchronized with conformational orientations of the UBP; analogous to what is observed with the AMOEBA force field. Here also elevated RMSD values correlate with distorted structures whereas nWCFP1 And WCF structures exhibit comparatively lower RMSD values (Figure 5). It is noticed that conformational change of the UBP is reflected on the associated geometrical parameters related to the UBP depicted in Figure S3 and S4.

The nWCFP1 structures, characterized by intercalation, exhibit the smallest DS-DN distances ($d_{DS\text{-}DN}$) (figure S2). By contrast, the WCF structure, akin to the natural BP orientations, displays an evident increase in distance. Notably, the nWCFP2 and nWCFPO structures reveal fluctuating $d_{DS\text{-}DN}$ values, suggesting the formation of distorted intercalated arrangements. Overall, AMBER based simulations are in agreement with the AMOEBA based simulation which suggest that flexibility and conformational change of the DS-DN are not an artifact of the force fields or sampling time; rather it is a feature of the UBP-incorporated DNA. The analysis of three replicates further emphasizes randomness in conformational change

of the UBP with a biased towards generating nWCF structures as more prevalent conformers.

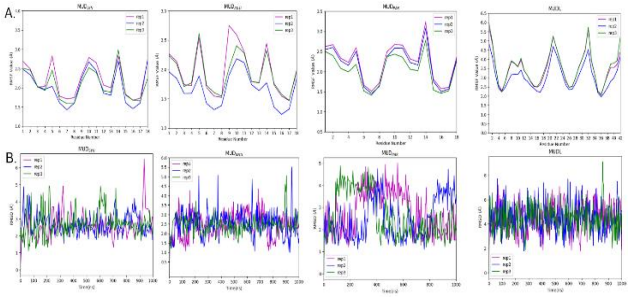


Figure 5. A) RMSF and B) RMSD values obtained from AMBER simulations.

To study the impact of the starting conformation on the geometry of UBP we explored MUD_{ANTI} and MUD_{PAR} structures as starting points for the simulations. Similar to the MUD_{SYN} scenario, commencing with MUD_{ANTI} also revealed a propensity for random conformational changes of the UBP. Here also WCF and nWCF structures are generated during the simulations where nWCF structures are found to be predominant. Noticeably, in this case outer-phase nWCFO structures are not observed. Conformational change between SYN and ANTI further verifies the occurrence of intra-strand flipping, a phenomenon evident through the corresponding O-S distances (Figure S2). Conformational change through Inter-strand flipping is also discernible from the UB-NB distance curves (Figure S5). Here the pattern of the RMSF looks similar to the one obtained for MUD_{SYN}, underscoring the flexibility of the UBP, which remains dynamic in nature and doesn't depend on the initial structure. Stability of WCF and nWCFP1 structures are confirmed by low RMSD values whereas high RMSD value of distorted structures indicates that they are comparatively less stable (Figure 5).

In the context of MUD_{PAR}, the results suggest the random exploration of both WCF and nWCF structures across all the replicates. Notably, the WCF structure exhibits a population exceeding 40%, signifying a higher prevalence than the other scenarios (Figure 6). In this case alongside in-phase nWCF (nWCFP1) structures, outer-phase (nWCFO) are also generated which include occasional distorted structures. Both intra and inter-flipping processes have been noticed during the simulation leading to generate both SYN and ANTI conformers alongside the conformational change through inter-flipping. Flexibility of the UBP is further evident from the RMSF plot (Figure 5), which aligns with the AMOEBA results discussed above. High RMSD values are observed in the region of nWCFO structures indicating the generation of UBP-distortion mainly in the outer-phase region.

Collectively, simulations conducted using both the AMBER and AMOEBA force fields consistently highlight the flexible and random conformational changes of UBP inside DNA duplex which leads to generate both WCF and nWCF structures. It is also evident from the simulation that UBP has an inclined tendency to stay as a nWCF forms throughout the simulations.

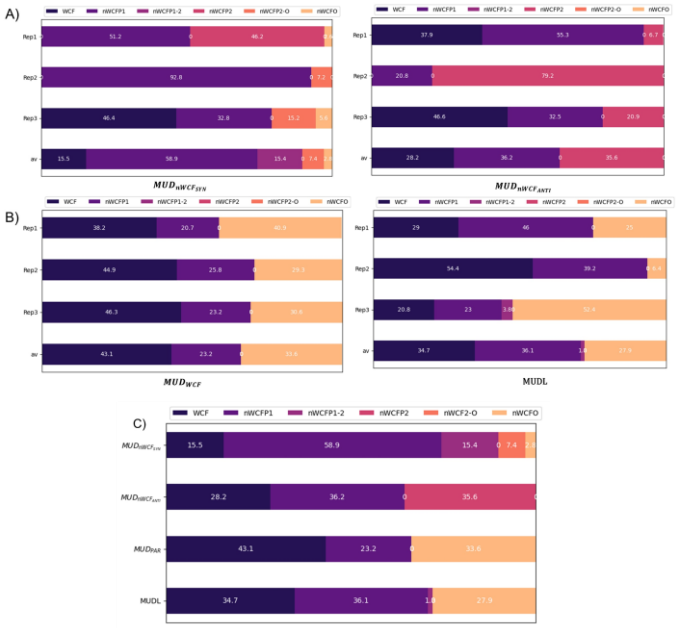


Figure 6. Population of different conformers in different conform.

We have used energy decomposition analysis (EDA) to investigate the interactions between the base pairs using the WCF, nWCFP1, nWCFP2, and nWCFO structures. We have examined the interactions between the DS-DN as well as adjacent complimentary DC-DG base pairs (cBPs) (Table 1). It is observed that for nWCFP1, DS-DN is stabilized through vdW interactions, where the vdW energy is -11.3 kcal/mol. cBPs are stabilized through coulomb interactions with associated E_{Coul}=∼-9.0 kcal/mol, indicating the stability of the base pairs. nWCFP2 structures also show similar interactions between DS-DN as well as cBPs as obtained from EDA analysis. During the calculation of energy decomposition analysis (EDA) for WCF structures, we have identified two distinct interaction regions between cBPs. In one instance (WCF1), Coulombic interactions are approximately around ∼-7.3 kcal/mol, while in the alternative scenario (WCF2), they have escalated to around ∼-8.0 kcal/mol.

The calculated Van der Waals interaction energy, E_{vdW}, for WCF structures is reduced to ∼-1.5 kcal/mol between DS and DN. We have also calculated the interaction energies between DS and DN as well as adjacent DG-DC base pairs for nWCFO structures. Here the Coulomb interactions between cBPs are further decreased to E_{Coul} ∼-7.0 kcal/mol along with a significant reduction of E_{vdW} (-6.4 kcal/mol) for DS-DN. Overall, the population of the different conformers is directly synchronized with the UBP and neighboring cBPs interactions where nWCF which correspond to the highest populations (39.4 %) shows the largest interactions as obtained from the EDA analysis.

MUDL: A system comprising ds-DNA with 21 base pairs with the placement of the UBP at the middle of the DNA duplex has also been considered to investigate the impact of a larger strand on UBP conformational stability (Scheme 1). Our results suggest the random generation of both WCF and

Table 1.Non-covalent interactions between BPs and UBP for short DNA (MUD). E_{col} (Coulamb energy), E_{vdW} (van daar Wal energy) are in kcal/mol.

	E_{Coul}				
cBPs	**nWCFP1**	**nWCFP2**	**nWCFO**	**WCF1**	**WCF2**
DC3-DG16	-9.2	-9.7	-7.0	-7.2	-7.9
DC4-DG15	-9.0	-8.9	-6.3	-7.3	-7.5
DG6-DC13	-9.2	-8.0	-7.2	-7.5	-8.9
DC7-DG12	-9.2	-9.4	-7.2	-7.5	-9.1
	E_{vdW}				
UBP	**nWCFP1**	**nWCFP2**	**nWCFO**	**WCF1**	**WCF2**
DS-DN	-11.3	-10.8	-6.4	-1.5	-1.7

nWCF structures during the simulations in different extent for different replicates. Both in-phase and outer-phase (nWCFP1 and nWCFO) nWCF structures are generated where the distribution and population of the conformers appear arbitrary. Here, conformational change is driven mostly by inter-strand flipping. The $d_{O\text{-}S}$ curve (Figure S2) predicts a majority of the time the system is in the SYN form during the simulation, whereas occasional distorted structure has been generated with larger $d_{O\text{-}S}$. It is found that RMSD values are comparatively higher than that of short DNA, which indicates that flexibility of the UBP transfers to the entire DNA duplex systems, which is further reflected in high RMSF values of all the nucleotides (Figure 5). RMSF value of the DS and DN are also found to be higher for MUDL than other MUD systems which implies that larger DNA duplex is able to give more freedom to the UBP to maintain and amplify its flexible nature. Interestingly like short DNA, here also the RMSF shows similar patterns indicating despite high flexibility, the characteristic of the UBP- incorporated DNA remains similar. High fluctuations of the UBP are also reflected in the high range of DS-DN distance as observed in Figure S2.

EDA analysis has also been employed to predict the interactions between DS and DN as well as complimentary natural base pairs for both WCF and nWCF (nWCFP1 and nWCFO) conformers. vdW energy between DS-DN is found to be the highest for nWCFP1 followed by nWCFO and lowest for WCF. Interestingly, here the interactions between complementary base pairs are similar irrespective of their orientations (Table 2). This suggests that as the length of the DNA increases, flexibility and different conformational orientations do not influence the interactions between the adjacent cBPs. Consequently, unlike MUD, the average population of the conformers for MUDL is almost equal. It further suggests that instead of UBP stability, interaction with adjacent base pairs is a major factor for the conformational distribution and population of the conformations (Figure S6).

Table 2. Non-covalent interactions between BPs and UBP for long DNA (MUDL). E_{col} (Coulamb energy), E_{vdW} (van daar Wal energy) are in kcal/mol.

	E_{Coul}		
cBPs	**nWCFP1**	**nWCFO**	**WCF**
DG9-DG34	-9.1	-9.2	-9.3
DC10-DG33	-8.5	-8.2	-9.1
DG12-DC31	-8.5	-9.1	-8.9
DC13-DG30	-9.5	-9.4	-9
	E_{vdW}		
UBP	**nWCFP1**	**nWCFO**	**WCF**
DS-DN	-9.0	-7.7	-1.7

UUD: Here DS and DN are positioned in the 3' terminus of the DNA to study the dynamical properties of the UBP-incorporated DNA in solution employing AMBER force fields for 1 μs for three replicates. It has been observed that the UBP in a terminal position is more flexible in nature as observed from the RMSF values (Figure S11). This flexibility leads to form frequent mis-paired and frayed structures of the UBP (Figure 7). RMSD values also suggest fluctuating characteristics of UUD structures. Notably, in this case nWCF structures are found to be predominant when the UBP is in the terminal position. Here also conformational change has been observed through both intra and inter-strand flipping. From the EDA analysis, it has been noticed that the Coulomb interaction between adjacent cBPs are smaller compared with the internal UBP systems, indicating flexibility of the UBP also perturb the neighboring cBPs. It has been observed that DS and DN are stabilized through vdW interactions with associated E_{vdW}=-6.5 kcal/mol

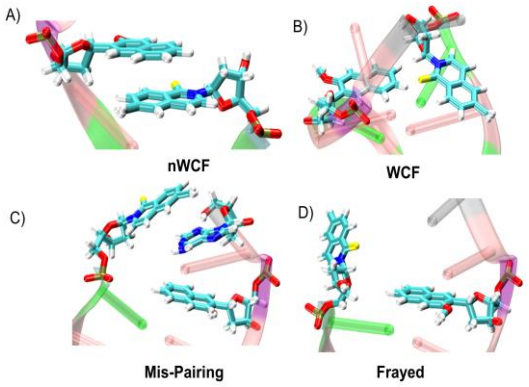


Figure 7. Snapshots of different geometries of UUD form of DNA duplex during simulations through flipping and altering.

indicating the interaction is significantly reduced than what is observed in nWCFP1 structures.

Conclusions

We investigated structural aspects, conformational changes, and stability of DS-DN incorporated DNA duplex by considering both short and long forms of DNA duplex simulated with both polarizable AMOEBA and AMBER force fields. It was found from both force fields that unlike natural base pairs, UBP can persist as both WCF and nWCF conformers inside DNA duplex with flexible orientations and random conformational change, which agrees with the previous simulated results with AMBER[44] and CHARMM[39] forcefields. This could suggest that the flexibility of UBP inside DNA is not an artifact of the forcefields, but rather it is an intrinsic property of this DS-DN incorporated DNA. It is evident from our analysis that conformational orientation perturbs the stability of the neighboring cBP mostly for shorter fragments of DNA, which reflects on the populations of the conformers. In shorter DNA strands, nWCF conformers are predominant whereas equal distributions are noticed for long-DNA. Our simulated results also suggest the fluctuating nature of DS and DN in the terminal position.

ASSOCIATED CONTENT

The Supporting Information is available free of charge at http://pubs.acs.org."
Additional details of MD, EDA (PDF)
Additional ESI for the initial coordinates and parameters for all of the studied systems (ESI-1.zip) (ZIP)

AUTHOR INFORMATION

Corresponding Author

G. Andrés Cisneros–Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, United States; Department of Chemistry and Biochemistry, The University of Texas at Dallas, Richardson, Texas 75080, United States; orcid.org/0000-0001-6629-3430;
Email: andres@utdallas.edu

Author

Tanay Debnath–Department of Physics, The University of Texas at Dallas, Richardson, Texas 75080, United States; orcid.org/0000-0001-8672-1834

Funding Sources

This work was supported by R01GM108583 and Fulbright-Nehru postdoctoral Scholarship. Computing time from UNT CASCaM funded by NSF Grant Nos. CHE1531468 and OAC-2117247 is gratefully acknowledged.

ACKNOWLEDGMENT

Tanay Debnath is a USIEF postdoctoral fellow (Fulbright-Nehru postdoctoral Scholarship). Computational time for this project was provided by the University of North Texas CASCaM CRUNTCh3 and 4 high-performance cluster and the University of Texas at Dallas 'Cyberinfrastructure and Research Services, Ganymede and Titan HPC clusters and NCSA DELTA at University of Illinois Urbana-Champaign.

REFERENCES

(1) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737-738.
(2) Tang, T.-C.; An, B.; Huang, Y.; Vasikaran, S.; Wang, Y.; Jiang, X.; Lu, T. K.; Zhong, C. Materials design by synthetic biology. *Nature Reviews Materials* **2021**, *6* (4), 332-350.
(3) Yan, X.; Liu, X.; Zhao, C.; Chen, G.-Q. Applications of synthetic biology in medical and pharmaceutical fields. *Signal Transduction and Targeted Therapy* **2023**, *8* (1), 199.
(4) Ambrogelly, A.; Palioura, S.; Söll, D. Natural expansion of the genetic code. *Nature Chemical Biology* **2007**, *3* (1), 29-35.
(5) Fischer, E. C.; Hashimoto, K.; Zhang, Y.; Feldman, A. W.; Dien, V. T.; Karadeema, R. J.; Adhikary, R.; Ledbetter, M. P.; Krishnamurthy, R.; Romesberg, F. E. New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nature Chemical Biology* **2020**, *16* (5), 570-576.
(6) Kimoto, M.; Hirao, I. Genetic alphabet expansion technology by creating unnatural base pairs. *Chemical Society Reviews* **2020**, *49* (21), 7602-7626.
(7) Chelliserrykattil, J.; Lu, H.; Lee, A. H. F.; Kool, E. T. Polymerase Amplification, Cloning, and Gene Expression of Benzo-Homologous "yDNA" Base Pairs. *ChemBioChem* **2008**, *9* (18), 2976-2980,
(8) Kaul, C.; Müller, M.; Wagner, M.; Schneider, S.; Carell, T. Reversible bond formation enables the replication and amplification of a crosslinking salen complex as an orthogonal base pair. *Nature Chemistry* **2011**, *3* (10), 794-800.
(9) Meggers, E.; Holland, P. L.; Tolman, W. B.; Romesberg, F. E.; Schultz, P. G. A Novel Copper-Mediated DNA Base Pair. *Journal of the American Chemical Society* **2000**, *122* (43), 10714-10715.
(10) Tarashima, N.; Komatsu, Y.; Furukawa, K.; Minakawa, N. Faithful PCR Amplification of an Unnatural Base-Pair Analogue with Four Hydrogen Bonds. *Chemistry* **2015**, *21* (30), 10688-10695.
(11) Malyshev, D. A.; Seo, Y. J.; Ordoukhanian, P.; Romesberg, F. E. PCR with an Expanded Genetic Alphabet. *Journal of the American Chemical Society* **2009**, *131* (41), 14620-14621.
(12) Lavergne, T.; Malyshev, D. A.; Romesberg, F. E. Major Groove Substituents and Polymerase Recognition of a Class of Predominantly Hydrophobic Unnatural Base Pairs. *Chemistry – A European Journal* **2012**, *18* (4), 1231-1239,
(13) Yamashige, R.; Kimoto, M.; Takezawa, Y.; Sato, A.; Mitsui, T.; Yokoyama, S.; Hirao, I. Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Research* **2012**, *40* (6), 2793-2806.

(14) Yang, Z.; Chen, F.; Alvarado, J. B.; Benner, S. A. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *Journal of the American Chemical Society* **2011**, *133* (38), 15105-15112.
(15) Hashimoto, K.; Fischer, E. C.; Romesberg, F. E. Efforts toward Further Integration of an Unnatural Base Pair into the Biology of a Semisynthetic Organism. *Journal of the American Chemical Society* **2021**, *143* (23), 8603-8607.
(16) Hoshika, S.; Leal, N. A.; Kim, M.-J.; Kim, M.-S.; Karalkar, N. B.; Kim, H.-J.; Bates, A. M.; Watkins, N. E.; SantaLucia, H. A.; Meyer, A. J.; et al. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **2019**, *363* (6429), 884-887.
(17) Wang, H.; Wang, L.; Ma, N.; Zhu, W.; Huo, B.; Zhu, A.; Li, L. Access to Photostability-Enhanced Unnatural Base Pairs via Local Structural Modifications. *ACS Synthetic Biology* **2022**, *11* (1), 334-342.
(18) Hettinger, T. P. Helix instability and self-pairing prevent unnatural base pairs from expanding the genetic alphabet. *Proceedings of the National Academy of Sciences* **2017**, *114* (32), E6476-E6477.
(19) Jena, N. R.; Das, P. Hybrid nucleobases as new and efficient unnatural genetic letters. *Journal of Biomolecular Structure and Dynamics* **2023**, *41* (1), 366-376.
(20) Hirao, I.; Kimoto, M.; Yamashige, R. Natural versus Artificial Creation of Base Pairs in DNA: Origin of Nucleobases from the Perspectives of Unnatural Base Pair Studies. *Accounts of Chemical Research* **2012**, *45* (12), 2055-2065.
(21) Georgiadis, M. M.; Singh, I.; Kellett, W. F.; Hoshika, S.; Benner, S. A.; Richards, N. G. J. Structural Basis for a Six Nucleotide Genetic Alphabet. *Journal of the American Chemical Society* **2015**, *137* (21), 6947-6955.
(22) Marx, A.; Betz, K. The Structural Basis for Processing of Unnatural Base Pairs by DNA Polymerases. *Chemistry – A European Journal* **2020**, *26* (16), 3446-3463.
(23) Dörrenhaus, R.; Wagner, P. K.; Kath-Schorr, S. Two are not enough: synthetic strategies and applications of unnatural base pairs. **2023**.
(24) Feldman, A. W.; Romesberg, F. E. Expansion of the Genetic Alphabet: A Chemist's Approach to Synthetic Biology. *Accounts of Chemical Research* **2018**, *51* (2), 394-403.
(25) Malyshev, D. A.; Dhami, K.; Quach, H. T.; Lavergne, T.; Ordoukhanian, P.; Torkamani, A.; Romesberg, F. E. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proceedings of the National Academy of Sciences* **2012**, *109* (30), 12005-12010.
(26) Mukba, S. A.; Vlasov, P. K.; Kolosov, P. M.; Shuvalova, E. Y.; Egorova, T. V.; Alkalaeva, E. Z. Expanding the Genetic Code: Unnatural Base Pairs in Biological Systems. *Molecular Biology* **2020**, *54* (4), 475-484.
(27) Lee, K. H.; Hamashima, K.; Kimoto, M.; Hirao, I. Genetic alphabet expansion biotechnology by creating unnatural base pairs. *Current Opinion in Biotechnology* **2018**, *51*, 8-15.
(28) Feldman, A. W.; Dien, V. T.; Karadeema, R. J.; Fischer, E. C.; You, Y.; Anderson, B. A.; Krishnamurthy, R.; Chen, J. S.; Li, L.; Romesberg, F. E. Optimization of Replication, Transcription, and Translation in a Semi-Synthetic Organism. *Journal of the American Chemical Society* **2019**, *141* (27), 10644-10653.
(29) Hamashima, K.; Kimoto, M.; Hirao, I. Creation of unnatural base pairs for genetic alphabet expansion toward synthetic xenobiology. *Current Opinion in Chemical Biology* **2018**, *46*, 108-114.
(30) Morihiro, K.; Moriyama, Y.; Nemoto, Y.; Osumi, H.; Okamoto, A. anti–syn Unnatural Base Pair Enables Alphabet-Expanded DNA Self-Assembly. *Journal of the American Chemical Society* **2021**, *143* (35), 14207-14217.
(31) Malyshev, D. A.; Romesberg, F. E. The Expanded Genetic Alphabet. *Angewandte Chemie International Edition* **2015**, *54* (41), 11930-11944,
(32) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Dwyer, T. J.; Ordoukhanian, P.; Romesberg, F. E.; Marx, A. KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nature Chemical Biology* **2012**, *8* (7), 612-614.
(33) Duffy, K.; Arangundy-Franklin, S.; Holliger, P. Modified nucleic acids: replication, evolution, and next-generation therapeutics. *BMC Biology* **2020**, *18* (1), 112.
(34) Dien, V. T.; Holcomb, M.; Feldman, A. W.; Fischer, E. C.; Dwyer, T. J.; Romesberg, F. E. Progress Toward a Semi-Synthetic Organism with an Unrestricted Expanded Genetic Alphabet. *Journal of the American Chemical Society* **2018**, *140* (47), 16115-16123.
(35) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R.; Romesberg, F. E. A semi-synthetic organism with an expanded genetic alphabet. *Nature* **2014**, *509* (7500), 385-388.
(36) Richards, N. G. J.; Georgiadis, M. M. Toward an Expanded Genome: Structural and Computational Characterization of an Artificially Expanded Genetic Information System. *Accounts of Chemical Research* **2017**, *50* (6), 1375-1382.
(37) Oh, J.; Shin, J.; Unarta, I. C.; Wang, W.; Feldman, A. W.; Karadeema, R. J.; Xu, L.; Xu, J.; Chong, J.; Krishnamurthy, R.; et al. Transcriptional processing of an unnatural base pair by eukaryotic RNA polymerase II. *Nature Chemical Biology* **2021**, *17* (8), 906-914.
(38) Betz, K.; Malyshev, D. A.; Lavergne, T.; Welte, W.; Diederichs, K.; Romesberg, F. E.; Marx, A. Structural Insights into DNA Replication without Hydrogen Bonds. *Journal of the American Chemical Society* **2013**, *135* (49), 18637-18643.
(39) Jahiruddin, S.; Datta, A. What Sustains the Unnatural Base Pairs (UBPs) with No Hydrogen Bonds. *The Journal of Physical Chemistry B* **2015**, *119* (18), 5839-5845.
(40) Eberlein, L.; Beierlein, F. R.; van Eikema Hommes, N. J. R.; Radadiya, A.; Heil, J.; Benner, S. A.; Clark, T.; Kast, S. M.; Richards, N. G. J. Tautomeric Equilibria of Nucleobases in the Hachimoji Expanded Genetic

Alphabet. *Journal of Chemical Theory and Computation* **2020**, *16* (4), 2766-2777.
(41) Wang, Q.; Xie, X.-Y.; Han, J.; Cui, G. QM and QM/MM Studies on Excited-State Relaxation Mechanisms of Unnatural Bases in Vacuo and Base Pairs in DNA. *The Journal of Physical Chemistry B* **2017**, *121* (46), 10467-10478.
(42) Jahiruddin, S.; Mandal, N.; Datta, A. Structure and Electronic Properties of Unnatural Base Pairs: The Role of Dispersion Interactions. *ChemPhysChem* **2018**, *19* (1), 67-74.
(43) Negi, I.; Kathuria, P.; Sharma, P.; Wetmore, S. D. How do hydrophobic nucleobases differ from natural DNA nucleobases? Comparison of structural features and duplex properties from QM calculations and MD simulations. *Physical Chemistry Chemical Physics* **2017**, *19* (25), 16365-16374.
(44) Galindo-Murillo, R.; Barroso-Flores, J. Structural and dynamical instability of DNA caused by high occurrence of d5SICS and dNaM unnatural nucleotides. *Physical Chemistry Chemical Physics* **2017**, *19* (16).
(45) Galindo-Murillo, R.; Barroso-Flores, J. Hydrophobic unnatural base pairs show a Watson-Crick pairing in micro-second molecular dynamics simulations. *Journal of Biomolecular Structure and Dynamics* **2020**, *38* (14), 4098-4106.
(46) *Gaussian 16 Rev. A.03*; Wallingford, CT, 2016.
(47) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008**, *10* (44), 6615-6620.
(48) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics* **2008**, *72* (1), 650-654.
(49) McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18. *The Journal of Chemical Physics* **2008**, *72* (10), 5639-5648.
(50) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. J. T. J. o. c. p. Intermolecular potentials based on symmetry-adapted perturbation theory with dispersion energies from time-dependent density-functional calculations. **2005**, *123 21*, 214103.
(51) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E., III; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *Journal of Chemical Theory and Computation* **2017**, *13* (7), 3185-3197.
(52) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *Journal of Chemical Theory and Computation* **2011**, *7* (3), 625-632.
(53) Tutorial-4: R.E.D. Server Development/PyRED (q4md-forcefieldtools.org.
(54) Tian, C.; Kasavajhala, K.; Belfon, K. A. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16* (1), 528-552.
(55) Zgarbová, M.; Šponer, J.; Otyepka, M.; Cheatham, T. E., III; Galindo-Murillo, R.; Jurečka, P. Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *Journal of Chemical Theory and Computation* **2015**, *11* (12), 5723-5736.
(56) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105* (43), 9954-9960.
(57) Ren, P.; Ponder, J. W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry* **2002**, *23* (16), 1497-1506.
(58) Ren, P.; Ponder, J. W. Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. *The Journal of Physical Chemistry B* **2003**, *107* (24), 5933-5947.
(59) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society* **2003**, *125* (50), 15671-15682.
(60) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *The Journal of Physical Chemistry B* **2004**, *108* (35), 13427-13437.
(61) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *Journal of Chemical Theory and Computation* **2011**, *7* (10), 3143-3161.
(62) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; et al. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science* **2018**, *9* (4), 956-972.
(63) Stone, A. J. Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation* **2005**, *1* (6), 1128-1132.
(64) Schafmeister, C.; Ross, W.; Romanovski, V. LEaP. University of California: San Francisco,1995.
(65) D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz,R Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y.

Xue, D.M. York and P.A. Kollman (2020), AMBER 2020, University of California, San Francisco.
(66) Zwanzig, R. Nonlinear generalized Langevin equations.J. Stat.Phys.1973,9, 215–220.
(67) Gillespie, D. T. The chemical Langevin equation. J. Chem. Phys. 2000, 113, 297–306.
(68) Steinbach, P. J.; Brooks, B. R. New spherical-cutoff methods for long-range forces in macromolecular simulation. **1994**, *15* (7), 667-683.

(69) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577-8593.

(70) Salomon-Ferrer, R.; Gotz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine micro second molecular dynamics simulations with AMBER on GPUs.2. Explicit solvent particle mesh Ewald. J. Chem. Theory Comput. 2013, 9, 3878–3888

(71) Roe, D. R.; Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9* (7), 3084-3095.
(72) Walt, S. v. d.; Colbert, S. C.; Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering* **2011**, *13* (2), 22-30.
(73) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* **2007**, *9* (3), 90-95.
(74) McKinney, W. In Data Structures for Statistical Computing in Python,Proceedings of the 9th Python in Science Conference,Austin,TX, 2010, Vol.445, p. 51–56
(75) Emmett Leddin, Cisneros Research Group, & G. Andres Cisneros. (2021). CisnerosResearch/AMBER-EDA: First Release (v0.1). Zenodo.
(76) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30* (13), 2157-2164.

Graphical Abstract: