Supplementary Material For:

Molecular Dynamics of Water Mediated Interactions of a Linear Benzimidazole-Biphenyl Diamidine with the DNA Minor Groove

Prashanth Athri and W. David Wilson

Department of Chemistry, Georgia State University, Atlanta, GA 30303

Section 1: Parametrization of DB921 for use with the Amber Force Field

The additive form of modern force fields (Equation 1) has a typical four-component form accounting for the intra- and intermolecular forces shown in Equation 1.

$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\in R_{ij}}\right]$$

Equation 1

The potential energy, $\boldsymbol{E_{total}}$, of the molecule or biomolecular system under consideration is a function of the atomic co-ordinates of all the atoms present in the molecule (Born-Oppenheimer approximation)[1]. The first term (Hooke's law equation) models the interaction between bonded atoms, and the second term models the interaction between three bonded atoms that form an angle. Both of these interactions use harmonic potentials that increase with the deviation from an equilibrium value, $\boldsymbol{r_{eq}}$ and $\boldsymbol{\theta_{eq}}$, respectively. The bond energy term enumerates the potential energy of a bond and $\boldsymbol{K_r}$ is the force constant that determines how tightly the two atoms are bound, in other words, the width of the curve. As the quadratic form of the equation suggests, the shape of the curve is a harmonic where $\boldsymbol{r_{eq}}$ is the reference bond length, and $\boldsymbol{r}$ the distance between the two atoms that share the bond. The potential energy curve due to variation in angle is similar to the bonded atoms' curves in that both use the functional form of Hooke's law. The energy that is required to change the angle between two bonds is considerably less than that required to stretch a bond between two atoms.

The third term is a summation of various Cosine terms that account for the changes in energy due to dihedral variations. Because of the relatively low energy required to make changes in dihedral angles, this term is the most important bonded contribution for control of the compound conformation either free or bound to DNA. The fourth term represents non-bonded interactions between atoms in the molecule that are separated by at least three bonds. This term models both the electrostatic interaction through the Coulombic potential, as well as the van der Waals interaction using the Lennard-Jones potential. For any two given atoms (***i***, ***j***, notations with respect to Equation 1), which are within a predetermined distance, the electrostatic affinity or repulsion is calculated using their respective partial charges. Partial charges are derived using the RESP[2, 3] fitting method preceded by the calculation of

electrostatic potential through a HF6-31G* basis set (*ab initio* calculation). These charges are represented as q_i and q_j (Equation 1), and R_{ij} represents the distance between these two atoms. Hence the $q_i * q_j / R_{ij}$ term gives the electrostatic contribution with respect to any two atoms, ***i*** and ***j***. The standard Lennard-Jones 6-12 potentials are used to calculate the van der Waals interactions. The functional form of the Lennard-Jones 6-12 potential is a part of the last term of Equation 1. σ and ε, represent the collision diameter and the well depth of the Lennard-Jones 6-12 potential curve, respectively, and R_{ij} is the distance between the atoms being considered.

Force field parameters for DB921

When a new molecular entity like DB921 (Figure S1) is introduced into the system, parameters need to be provided to calculate its conformations and energies. Many of the parameters can be derived by analogy to similar chemical environments that have been parameterized earlier. A very important note with respect to force field development is that a complete array of all parameters, for a given chemical entity developed for a particular force field, is non-transferable[1]. For a given molecular entity, a set of parameters is derived to emulate a chemical phenomenon to an agreeable degree of accuracy. This set works well as long as it is used within the framework of the given force field with that specific set of parameters.

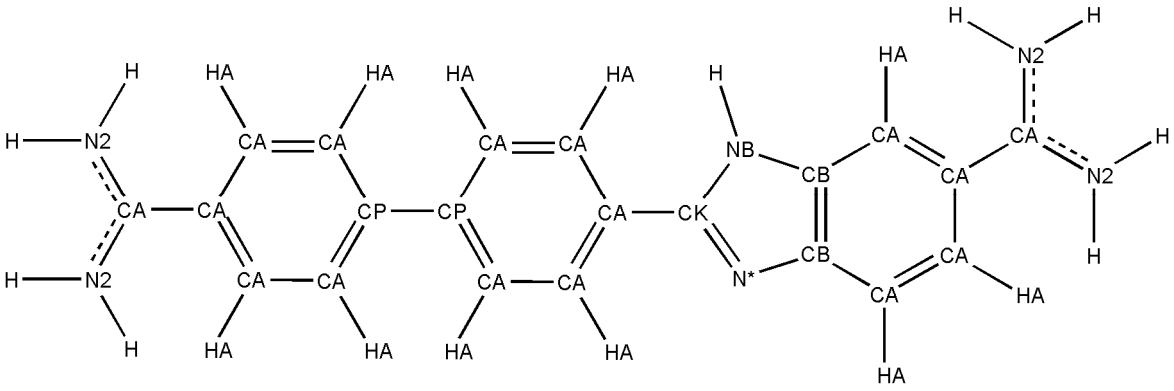


Figure S1: DB921 and atom types used for parameterization. The letters denote the type of atom in a certain type of environment. For example, **CA** denotes a carbon atom in an aromatic-type environment. Note that some critical bonds, such as the biphenyl and phenyl-benzimidazole connection, have been

assigned specific atom type that lets us define new force field parameters for those bonds (see the rows in bold in Figure S2).

The **frcmod** file (a file used with the **AMBER** suite of programs) that contains the parameters for the DB921 molecule is presented in Figure S2 and a guide to interpret each column under each of the individual sections (see below) is presented in Table S1. The figure is in the format accepted by **AMBER.** References to section names below are with respect to this figure. Our intention has been to introduce as few new parameters as possible. Most of the parameters have been adapted from the **ff99**[5] parameter set provided with the **AMBER** distribution. As shown in the figure, the **frcmod** file can be divided into six parts.

Table S1: Column content in any **AMBER** frcmod.dat. This table should be used to read Figure 3.

Section Name	***Information Content In The Respective Columns***					
	Column 1	***Column 2***	***Column 3***	***Column 4***	***Column 5***	***Column 6***
MASS	**Atom Type**	**Atomic Mass**	**Atomic Radius**	**Comments**	**NA**	
BOND	**Bond Definition**	**Force Constant**	**Equilibrium Dist.**	**Comments**	**NA**	
ANGLE	**Angle Definition**	**Force Constant**	**Equilibrium Angle**	**Comments**	**NA**	
DIHE	**Dihedral Definition**	**Multiplicity**	**Force Constant**	**Phase**	**Periodicity**	**Comments**
IMPROPER	**Improper Definition**	**Force Constant**	**Phase**	**Periodicity**	**Comments**	**NA**
NONBON	**Atom Type**	**vdW Parameter**	**vdW Parameter**	**Comments**	**NA**	**NA**

Title: DB921.frcmod – for use with AMBER force field

MASS

NB	14.01	0.530	parm99
CB	12.01	0.360	parm99
CA	12.01	0.360	parm99
HA	1.008	0.167	parm99
N*	14.01	0.530	parm99
CK	12.01	0.360	parm99
N2	14.01	0.530	parm99
H	1.008	0.161	parm99
CP	12.01	0.360	gaff (cp)

BOND

CA-CB	469.0	1.404	parm99
CB-CB	520.0	1.370	parm99
CA-CA	469.0	1.400	parm99
CA-HA	367.0	1.080	parm99
H- N*	434.0	1.010	parm99
CK-CA	469.0	1.475	force const: parm99, CA-CB; length- Gaussian
CB-N*	436.0	1.374	parm99
CB-NB	414.0	1.391	parm99
CK-N*	440.0	1.371	parm99
CK-NB	529.0	1.304	parm99
CA-N2	481.0	1.340	parm99
H-N2	434.0	1.010	parm99
CP-CP	346.5	1.499	force const.-gaff (cp-cp); length-gaussian
CP-CA	466.1	1.395	gaff (cp-ca)

ANGLE

CA-CB-NB	70.0	132.40	parm99
CB-CB-NB	70.0	110.40	parm99
N*-CK-NB	70.0	113.90	parm99
NB-CK-CA	70.0	124.72	force const: parm99, N*-CK-NB; angle-gaussian
CB-NB-CK	70.0	103.80	parm99
CB-CA-HA	50.0	120.00	parm99
CA-CA-CB	63.0	120.00	parm99
CA-CB-CB	63.0	117.30	parm99
CB-CB-N*	70.0	106.20	parm99
CA-CA-HA	50.0	120.00	parm99
CA-CA-CA	63.0	120.00	parm99
CA-CB-N*	70.0	132.89	force const: parm99, CB-CB-N*; angle-gaussian
CB-N*-H	50.0	125.80	parm99
CB-N*-CK	70.0	105.40	parm99
N*-CK-CA	70.0	122.43	force const: parm99, N*-CK-NB; angle-gaussian
CK-N*-H	50.0	128.80	parm99
CA-CA-CK	63.0	121.80	force const.- parm99 (CA-CA-CB); angle-gaussian
CA-CA-N2	70.0	119.80	force const.-parm99 (CM-CA-N2); angle-gaussian
CA-N2-H	50.0	120.00	parm99
N2-CA-N2	70.0	120.00	parm99
H-N2-H	35.0	120.00	parm99
CA-CP-CP	62.6	127.01	gaff (ca-cp-cp)
CP-CA-HA	48.0	121.08	gaff (cp-ca-ha)
CA-CA-CP	67.2	119.07	gaff (ca-ca-cp)
CA-CP-CA	67.1	118.75	gaff (ca-cp-ca)

DIHE

NB-CA-CB-HA	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
NB-CA-CB-CA	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
NB-CB-CB-CA	4	21.80	180.0	2.0	parm99 (X-CB-CB-X)
NB-CB-CB-N*	4	21.80	180.0	2.0	parm99 (X-CB-CB-X)

NB-CK-N*-CB	4	6.80	180.0	2.0	parm99 (X-CK-N*-X)
NB-CK-N*-H	4	6.800	180.0	2.0	parm99 (X-CK-N*-X)
NB-CK-CA-CA	**4**	**-0.6**	**180.0**	**4.0**	**New Parameters**
NB-CK-CA-CA	**4**	**3.1**	**180.0**	**2.0**	**New Parameters**
NB-CK-CA-CA	**4**	**-0.7**	**360.0**	**1.0**	**New Parameters**
CB-CK-NB-N*	2	20.00	180.0	2.0	parm99 (X-CK-NB-X)
CB- CK-NB -CA	2	20.00	180.0	2.0	parm99 (X-CK-NB-X)
CB-CA-CA-HA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CB-CA-CA-CA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CB- CA-CB -CA	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
CB- CA-CB -HA	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
CB-CB-N*-H	4	6.600	180.0	2.0	parm99 (X-CB-N*-X)
CB-CB-N*-CK	4	6.600	180.0	2.0	parm99 (X-CB-N*-X)
CA-CB-NB-CK	2	5.100	180.0	2.0	parm99 (X-CB-NB-X)
CA-CB-CB-CA	4	21.80	180.0	2.0	parm99 (X-CB-CB-X)
CA-CB-CB-N*	4	21.80	180.0	2.0	parm99 (X-CB-CB-X
CA-CA-CA-HA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CA-CA-CA-CA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
HA-CA-CA-HA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CA-CA-CB-N*	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
CA-CB-N*-H	4	6.600	180.0	2.0	parm99 (X-CB-N*-X)
CA-CB-N*-CK	4	6.600	180.0	2.0	parm99 (X-CB-N*-X)
HA-CA-CB-N*	4	14.00	180.0	2.0	parm99 (X-CA-CB-X)
CB-CB-NB-CK	2	5.100	180.0	2.0	parm99 (X-CB-NB-X)
CB- CK-N*-CA	4	6.800	180.0	2.0	parm99 (X-CK-N*-X)
N*-CK-CA-CA	**4**	**-0.6**	**180.0**	**4.0**	**New Parameters**
N*-CK-CA-CA	**4**	**3.1**	**180.0**	**2.0**	**New Parameters**
N*-CK-CA-CA	**4**	**-0.7**	**360.0**	**1.0**	**New Parameters**
H- CK-N*--CA	4	6.800	180.0	2.0	parm99 (X-CK-N*-X)
CK-CA-CA-HA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CK-CA-CA-CA	4	14.50	180.0	2.0	parm99 (X-CA-CA-X)
CA-CA-CA-N2	**4**	**0.789**	**327.0**	**-4.0**	**New Parameters**
CA-CA-CA-N2	**4**	**-3.118**	**0.0**	**-2.0**	**New Parameters**
CA-CA-CA-N2	**4**	**0.609**	**90.0**	**1.0**	**New Parameters**
X -CA-N2- X	4	9.6	180.0	2.0	DAPI paper, Cheatam and coworkers (X-CA-N2-X)
CP-CP-CA-HA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)
CP-CP-CA-CA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)
CP-CA-CA-HA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)
CP-CA-CA-CA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)
CA-CP-CP-CA	**4**	**-0.597**	**180.0**	**4.0**	**New Parameters**
CA-CP-CP-CA	**4**	**1.154**	**180.0**	**2.0**	**New Parameters**
CA-CP-CA-CA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)
CA-CP-CA-HA	4	14.5	180.0	2.0	gaff (x-pc-CA-x or x-CA-CA-x)

IMPROPER

CB-NB-CA-CB	1.1	180.0	2.0	parm99
CA-CB-HA-CA	1.1	180.0	2.0	parm99
CA-CA-HA-CA	1.1	180.0	2.0	parm99
CA-CA-CA-CB	1.1	180.0	2.0	parm99
CB-CB-CA-N*	1.1	180.0	2.0	parm99
N*-CB-H-CK	1.1	180.0	2.0	parm99
CK-NB-N*-CA	1.1	180.0	2.0	parm99
CA-CK-CA-CA	1.1	180.0	2.0	parm99
CA-CA-CA-HA	1.1	180.0	2.0	parm99

NONBON

NB	1.8240	0.1700	parm99 (N)
CB	1.9080	0.0860	parm99 (C*)
CA	1.9080	0.0860	parm99 (C*)
N*	1.8240	0.1700	parm99 (N)
CK	1.9080	0.0860	parm99 (C*)

HA	1.4590	0.0150	parm99 (HA)
N2	1.8240	0.1700	parm99 (N)
H	0.6000	0.0157	parm99 (H)
CP	1.9080	0.0860	parm99 (N)

Figure S2: DB921.frcmod file showing the complete parameterization used for the DB921 molecule. (The column headings are in Table S1 and are not included in the Figure to maintain the original format used in **AMBER**).

The first section, titled **MASS** (Table S1), has the atom type information along with the atomic mass of the atoms and these were taken directly from the **ff99** atom definitions. Force fields use various atom types that help differentiate from one atom to another in different chemical environments, even though they might be the same element, to account for various chemical, structural and electronic variations. The corresponding atom types, adapted from **ff99**, are in the comment column (the last column on the right). No modifications were made in this section, except to change the actual name of the atom type itself, for the sake of clarity and to avoid confusion between **ff99** definitions. It is worthwhile to note that the basic *atom type* entries hardly change among various parameterizations.

Figure S2 shows the **frcmod** file that we have used for DB921. Note that the figure uses atom types adapted from the **ff99** definitions. As seen in the Figure, most parameters were adapted from **ff99** and a few from **GAFF**. In some cases when appropriate **BOND**/ **ANGLE** parameters were not available in either **ff99** or **GAFF**, we have used force constants from comparable atom types and bond length (or bond angle) from a*b initio* calculations. In such cases, the Comments column (last column) has information about the source. All *ab initio* optimizations were carried out using the HF-631G* basis set implemented in **Gaussian**[6].

The detailed analysis of the next section in the parameter **frcmod** file, namely **DIHE**, is presented in the following section. The **IMPROPER** section definitions help force planarity of specific sections of the system, such as aromatic rings and substituents. This section defines atom quartets that are not dihedrals but are four connected atoms as shown in Figure S3. For a two-fold torsional potential (fourth column),

the improper dihedral has its lowest energy when the four listed atoms are planar. Note that the phase (third column) is always 180 degrees. This section was also adopted from the **ff99** database. Finally, the NONBON (non-bonded interactions) section is also from the **ff99** definitions of comparable atom types. This is shown in the final section of the frcmod file.

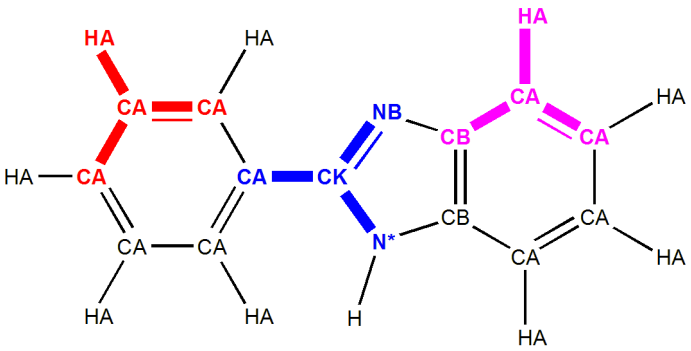


Figure S3: Examples of "Impropers", marked in color, of the phenyl-benzimidazole fragment of DB921.

Dihedral Angle Parameterization

The dihedrals bonded and van der Waals and electrostatic nonbonded parameters are the predominant factors that impact the conformation and structure of the DB921 molecule in solution and in its DNA complex[1]. Adaptation of the nonbonded parameters is straightforward and is briefly covered in the previous section. Dihedrals, on the other hand, have to be carefully evaluated and parameterized because of their affect on 3-D conformations and intermolecular interactions.

There are four undefined dihedrals in DB921 that are critical in determining its conformation and DNA interactions. With respect to Figure S1, they can be identified as the following:

1. NB-CK-CA-CA
2. N*-CK-CA-CA
3. CA-CP-CP-CA
4. CA-CA-CA-N2

The CA-CA-CA-N2 dihedral was previously parameterized by Cheatham and coworkers[7], for the diamidine analog DAPI while the other dihedrals were parameterized as part of this research using the

procedure outlined below. Parameters for CA-CA-CA-N2 were also re-calculated and the results are in good agreement with those of Cheatham and coworkers. To parameterize the new dihedral angles *ab inito* calculations are performed by splitting the molecule into manageable molecular fragments that completely define the dihedrals in question. It is assumed that the derived parameters are valid when the constituent fragments are connected to form the compound.

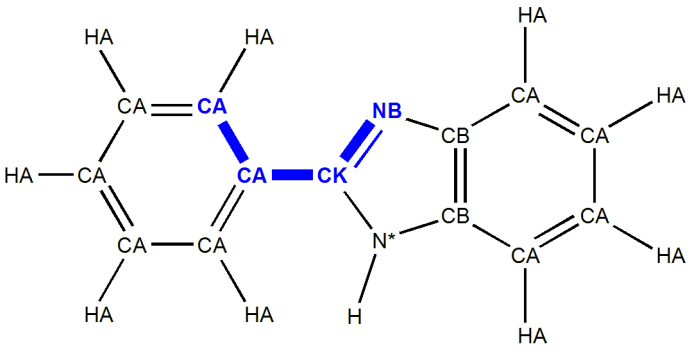


Figure S4: Fragment used to parameterize the dihedral NB-CK-CA-CA

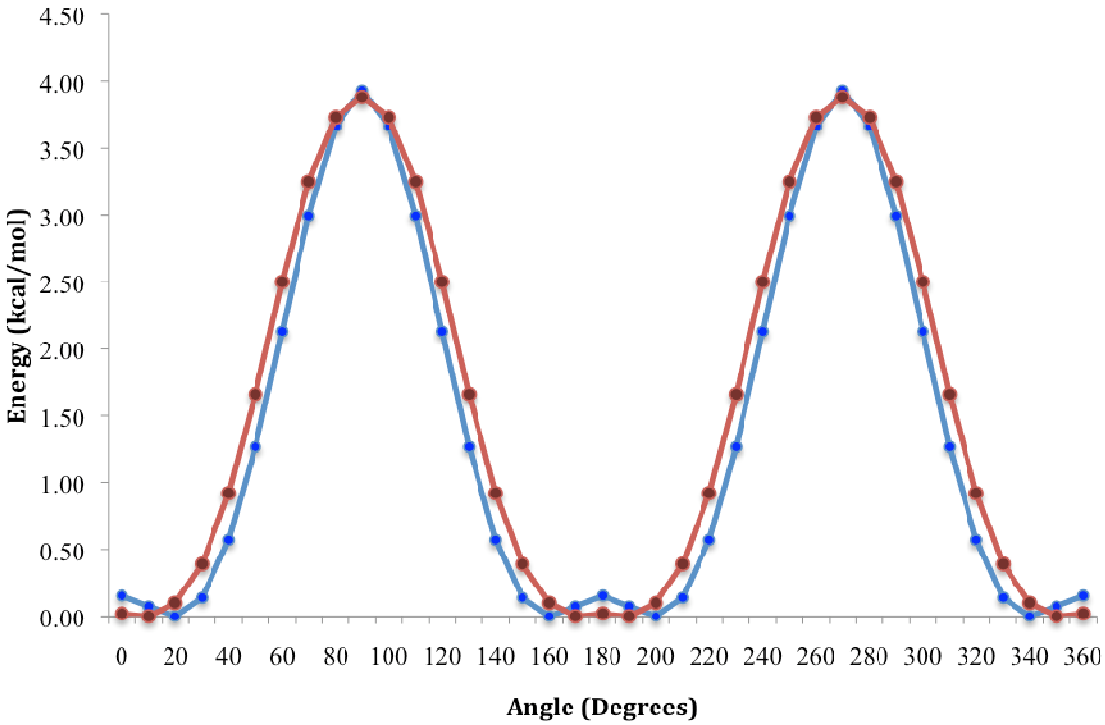
The dihedral defined by NB-CK-CA-CA (Figure S4) is used as an example to outline the procedure used. Figure S4 shows the molecular fragment that was used in the parameterization of the NB-CK-CA-CA dihedral. The aim of dihedral parameterization is to reproduce the plot of the torsional energy curve that is obtained through *ab initio* calculations. Gaussian was used to calculate *ab initio* based potential energy curves for dihedral angles ranging from 0° through 360° for the fragment in question. A constrained equilibrium calculation was performed in 10° intervals between 0° to 360° at the HF/6-31G* level. An energy profile was then constructed using data points at 10° intervals (see Figure S5(a), Table S2). New parameters were then derived such that the dihedral profile with the new parameters fits this curve.

To obtain the new parameters, the partial atomic charges for this fragment were calculated at the HF-631G* level of theory and **RESP** methodology was used to get the final atom centered charges using the inherent two step method outlined in[2, 3]. Standard procedures were used to obtain the dihedral parameters[8, 9]. The fragment was then stored as a Tripos type mol2 file[10]. The mol2 file contains the minimized, charged fragment that needs to be parameterized. The mol2 file was input to the

antechamber program (part of the **AMBER** distribution). This program suggests appropriate atom types, if they are already present in the general **AMBER** force field (**GAFF**[11]). One way of storing parameter information such that it can be used in a MD simulation is to use an **frcmod** file. The **parmchk** utility program (also a part of the **AMBER** distribution) can be used to generate an **frcmod** file that contains all the parameters necessary. Again, this is subject to the precondition that all the parameters are present in **GAFF**. The **frcmod** file can then be used in conjunction with the force field that provides the parameterization for the macromolecule to perform an MD simulation. The **frcmod** file from the **parmchk** program can be used directly or, if **GAFF** based parameters give erroneous results, this file can be constructed manually. It is convenient to use a skeleton of the **frcmod** generated from **parmchk** and populate it with the new parameters.

GAFF is a useful utility program that can be used to obtain parameters for small organic molecules. **GAFF** has parameters for many organic molecules and can be used with new molecules if similar atom types have been defined. In many cases, the parameters present in **GAFF** satisfactorily reproduce the behavior of the molecular fragment and the next step would be to integrate the individual fragments to form the compound, dock the new molecule as a ligand in the DNA or protein target to proceed with the MD simulation. In the case of DB921, **GAFF** generated parameters resulted in erroneous dihedral curves that did not match the *ab initio* generated dihedral plots (see Figure 5(b), Table S2). **GAFF** parameters derived for other small molecules that have conjugated junctions have also been incorrect in predicting torsional potentials[12]. Due to the mismatch in the dihedral plots, we have developed our own parameters for DB921. However, **GAFF** is a force field that has great potential as a complement to **AMBER** but is still in an early stage of development.

(a)



(b)

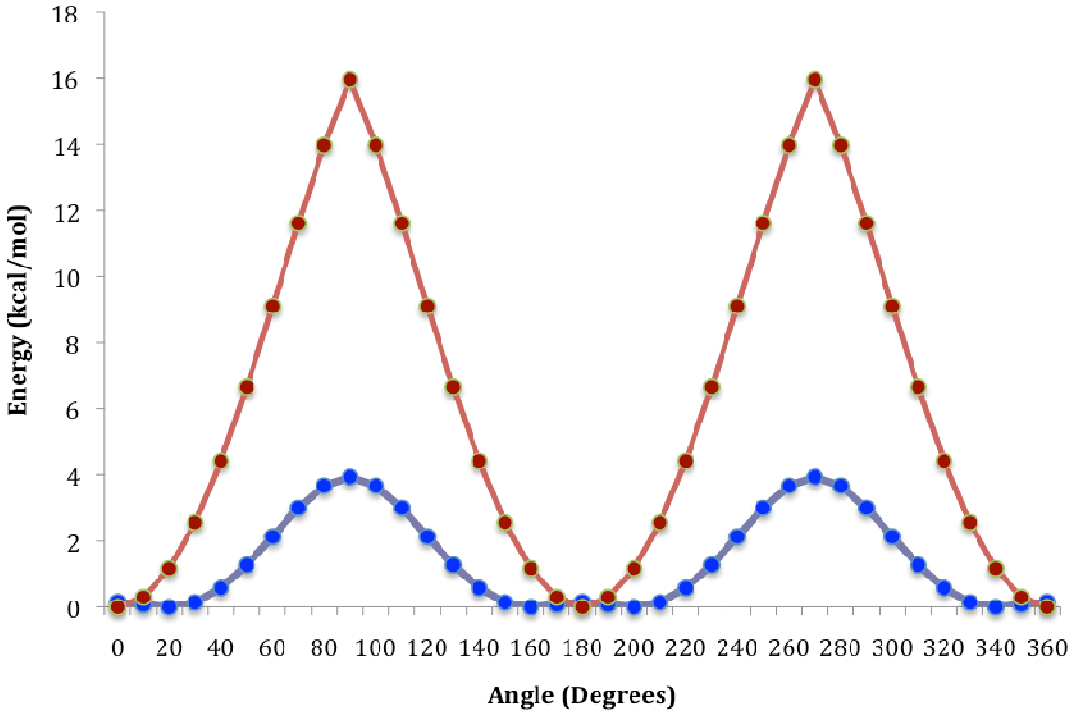


Figure S5: Potential energy profile of NB-CK-CA-CA. Blue (both (a) and (b)): HF-631G*. Red: (a) new parameters: Cosine series: $\Sigma V_4(1+\cos(4\varphi-180)) + V_2(1+\cos(2\varphi-180)) + V_1(1+\cos(\varphi-360))$ (b) **GAFF**.

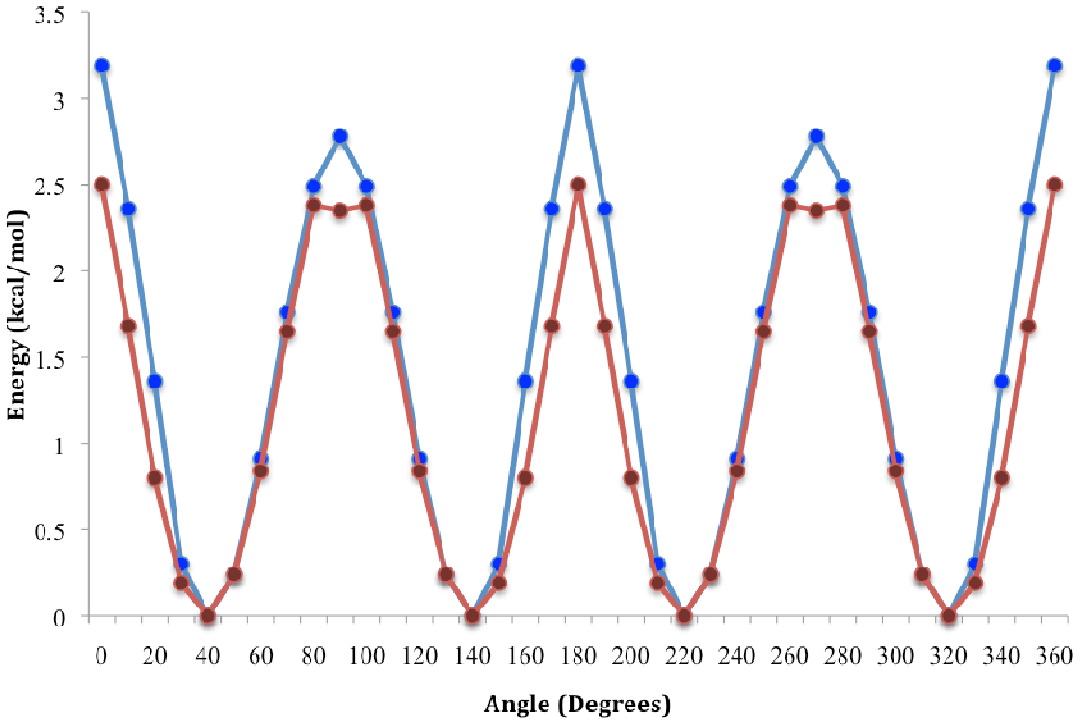
The curves for the NB-CK-CA-CA obtained through *ab initio* calculations and the fitted curve using the newly generated dihedral parameters are shown in Figure S5. The two curves were fitted using the least squares fitting routine[8] in **KaleidaGraph**[13]. The cosine function used to reproduce dihedral energy curves in the **AMBER** force filed is:

$$\sum_{dihedrals} \frac{V_n}{2}[1+\cos(n\phi-\gamma)] \qquad \textit{Equation 2}$$

where n = periodicity = 0, 1, 2, 3, 4 or 6 and γ = phase.

In the case of NB-CK-CA-CA, it was seen that there were significant contributions from the cos(4γ), cos(2γ) and cos(γ) series. Hence all three were included and this gave a good fit with the **Gaussian** generated HF-631G* curve (Figure S5(a)). Figure S5(b) shows the curve obtained by the original **GAFF** parameters. As can be seen, the **GAFF** parameters overestimate the potential energy barriers as well as dynamic characteristics that are of vital importance to MD simulations. Similar plots for the other dihedrals are shown in Figure S6 and S7 with appropriate legends. As can be seen, the above observations, regarding GAFF-generated plots, hold true in these cases as well. Plots from newly generated parameters closely fit the *ab initio* generated plots and all new parameters are collected in Figure S2. The derived cosine functions used to model the dihedrals are shown in the legend to each figure.

(a)



(b)

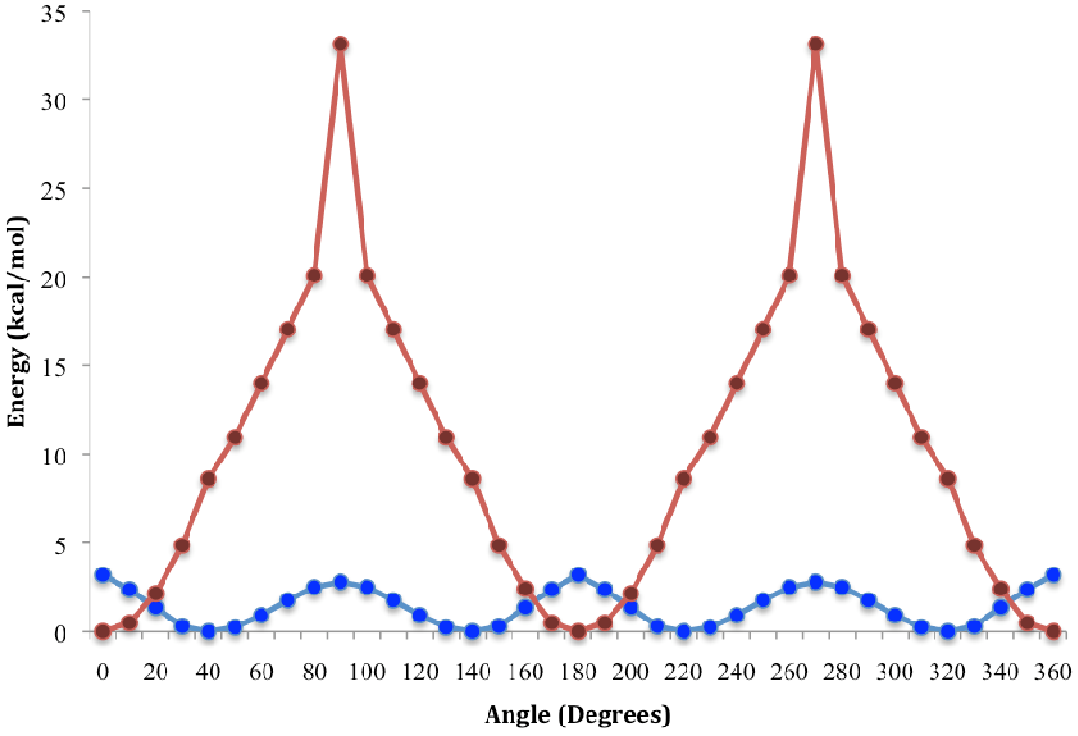
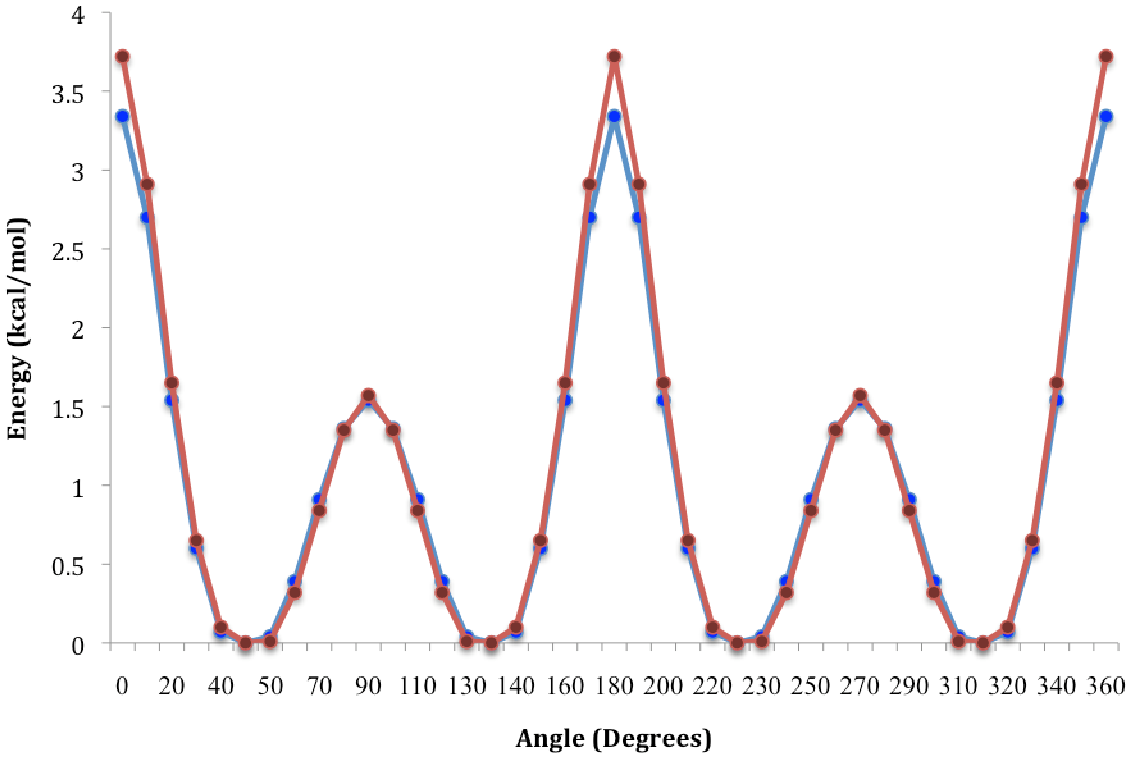


Figure S6: Potential energy profile of CA-CA-CA-N2. Blue (both (a) and (b)): HF-631G*. Red: (a) new parameters: Cosine series: $\Sigma V_4(1+\cos(4\varphi-327)) + V_2(1+\cos(2\varphi)) + V_1(1+\cos(\varphi-90))$ (b) GAFF.

(a)



(b)

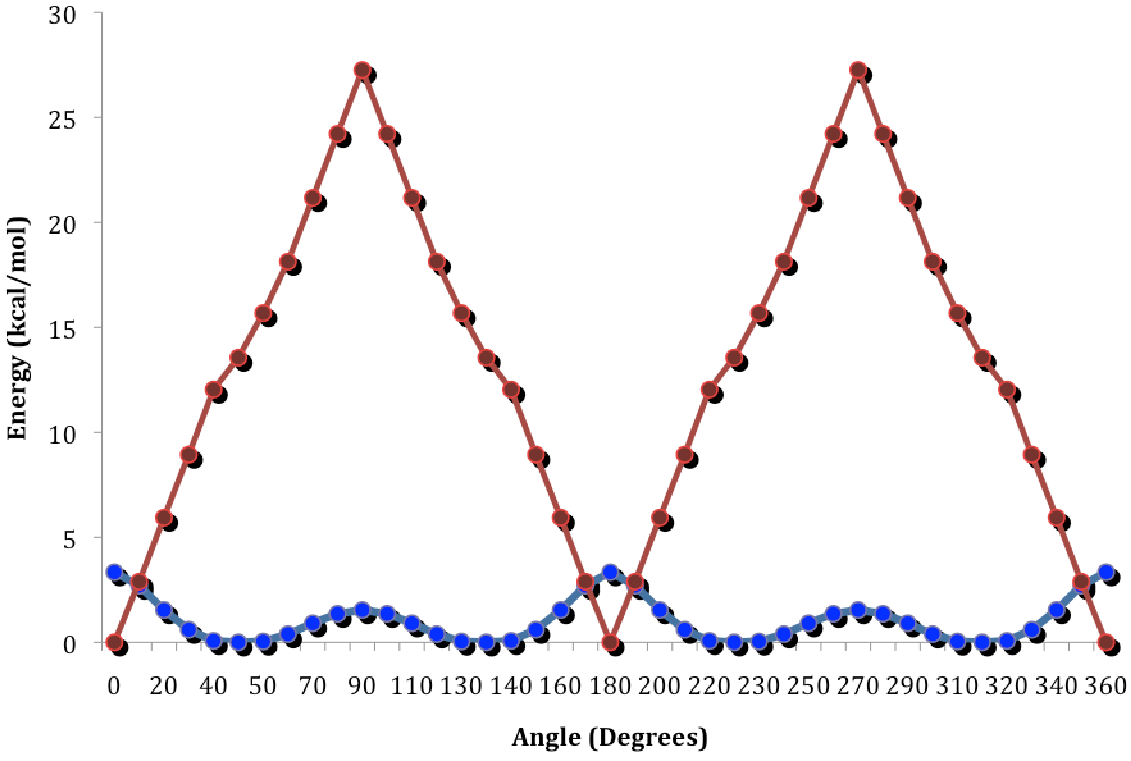


Figure S7: Potential energy profile of CA-CP-CP-CA. Blue ((both (a) and (b)): HF-631G*. Red: (a) new parameters: Cosine series: $\Sigma V_4(1+\cos(4\varphi-180)) + V_2(1+\cos(2\varphi-180))$ (b) GAFF.

In conclusion, we were able to generate force field parameters for DB921, with standardized parameterization for use with the Amber force field, that reproduce an *ab initio* model for the compound. The **frcmod** file can be used as is to reproduce dihedral profiles derived for DB921. The relevant parts can be used to model similar dihedrals in other systems.

Section 2: Construction of the reference plane used to classify the different modes of binding and tracking water molecules in these binding modes

The three atoms used to define the reference plane were three DNA phosphorous atoms (Figures 3 and 4). The reference plane thus defined is approximately parallel to the floor of the DNA minor groove and its orientation with respect to the groove remains fairly consistent across the simulation due to the relatively rigid structure of the DNA backbone atoms that define it. This was confirmed by the low RMSD values of the three atoms across the length of the simulation. The angle formed by the line connecting the two nitrogens of the amidine group (Figure 3 and Figure 4, Label B) to this plane was used as an index to classify each frame into different binding modes. All angles that lie within +/- 20 degrees of 0 were classified as parallel (||) frames and all angle +/- 20 degrees of 90 degrees were classified as perpendicular (⊥) frames. Note that the aim of defining the plane and the line, and calculating the angle formed between them, was to facilitate the classification of each frame into either || or ⊥ modes of binding, and not to accurately measure the angle formed by the terminal amidine to the floor of the groove or any other feature of reference. The precise value of the angle is not critical to the analysis and the values calculated can be expected to deviate from the true value of the angle formed between a perfectly parallel plane to the floor of the minor groove. Nevertheless, the plane serves as an efficient tool for classification, obviating manual parsing of each frame to identify the respective binding mode.

Each frame in the 100ns trajectory file was classified as either ||, ⊥ (**⊥-Major** or **⊥-Xray**) or "transitory". Transitory conformations were those that did not have interactions consistent with either || or ⊥ conformations, nor did they exhibit any consistent interactions. For each frame that belonged to the || class, distances between the closest of the four possible hydrogen atoms of the amidine group of DB921 and the oxygen atoms of the closest sugars were calculated and stored. For frames that had the amidine positioned in a perpendicular alignment to the floor of the groove, a further classification between **⊥-Major** and **⊥-Xray** was needed. The **⊥-Xray** mode of interaction had a hydrogen bonding

network that involved A5-N3, a hydrogen atom of the amidine and a linking water molecule. The following method was applied to each frame to track the number of waters that were involved in the interaction. All waters within hydrogen bonding distance of A5-N3 and waters within hydrogen bonding distance of A5-N3 the closest hydrogen of the phenylamidine group of DB921 were stored in two separate lists. The lists were compared to check for a common entry that would result in the frame being classified as **⊥-Xray**. If there was no common entry, it was classified as **⊥-Major**.

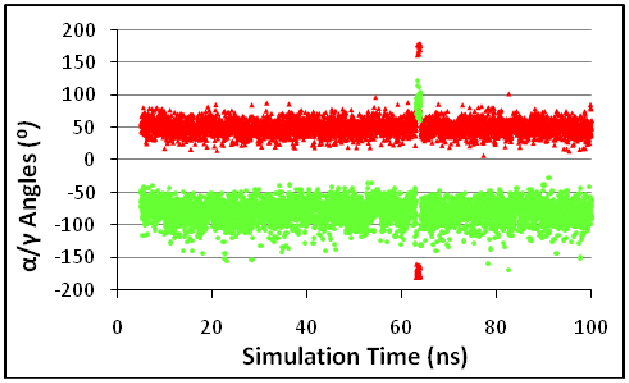
Frames were thus classified into one of the three conformational states. All the frames that belonged to a particular state were saved into different trajectories for further analysis to calculate the distances between partners participating in non-bonded interactions that stabilize that particular state. In the parallel conformation the procedure was used to store the distance between the nitrogens of the amidine and the oxygens of the DNA backbone that were involved in electrostatic interactions.

In the **⊥-Major** mode of interaction (Figure 4), the tracking of water molecules was based on inter-atomic distance based criteria. The specific criteria for capturing the exact water molecules in each frame were: "water molecule that was within hydrogen bonding distance to the DNA atom AND (a) within hydrogen bonding distance of at least one water molecule that was itself within hydrogen bonding distance to the closest amidine hydrogen of DB921 OR (b) within hydrogen bonding distance to amidine of DB921". Condition (a) captures all interactions with two water molecules whereas condition (b) captures one-water interactions. Using this criterion the specific water molecules that were involved in mediating the C21-O2 to DB921 contacts were stored in a database. This database also had the specific distances between each of the molecules along with the respective frame numbers.

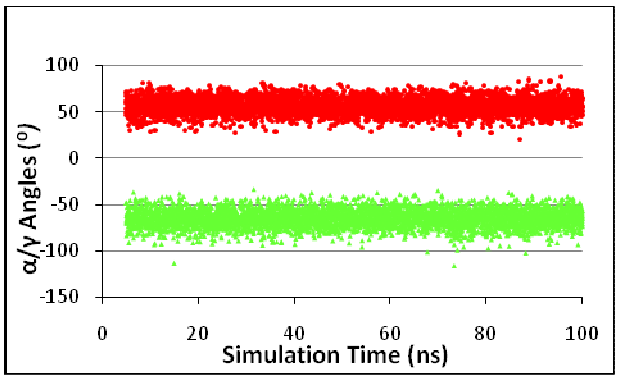
Section 3: Analysis of α/γ angles across the central base steps of the DNA-DB921 complex

The PARMBSC0 force field[14] used in this simulation incorporates changes in the α/γ torsional terms to compensate for the insufficient sampling of these torsions. The study shows that the previous generations of the Amber force fields generate irreversible α/γ concerted transitions that do not effect shorter simulations but result in an aggregative error in longer length simulations. To investigate the effect of the use of this new force field in our relatively long time scale DNA-ligand simulations, we have plotted the α/γ angles for various base steps across the central region of the DNA that includes the binding site of DB921 (Figure S8).

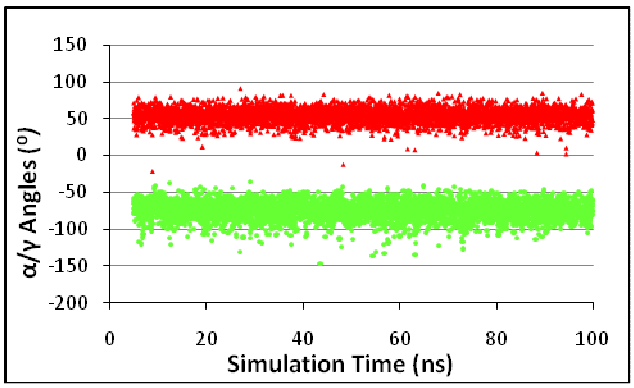
Base Sequence Steps 4 and 5



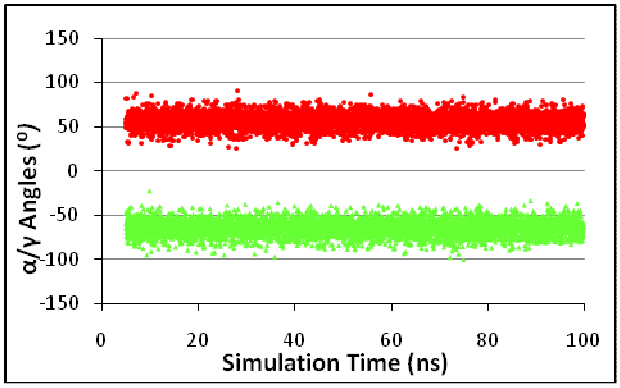
Base Sequence Steps 20 and 21



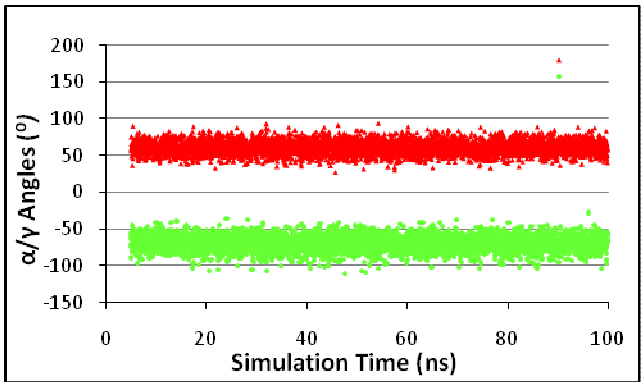
Base Sequence Steps 5 and 6



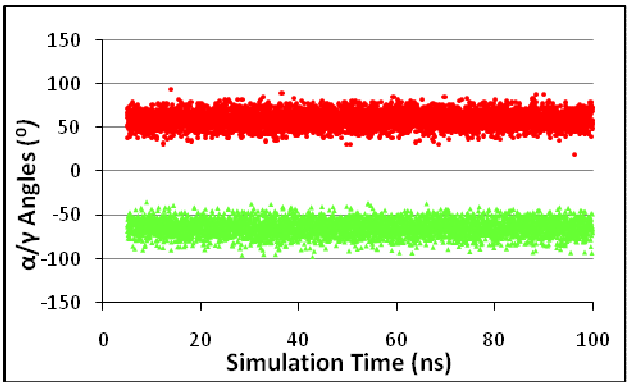
Base Sequence Steps 19 and 20



Base Sequence Steps 6 and 7



Base Sequence Steps 18 and 19



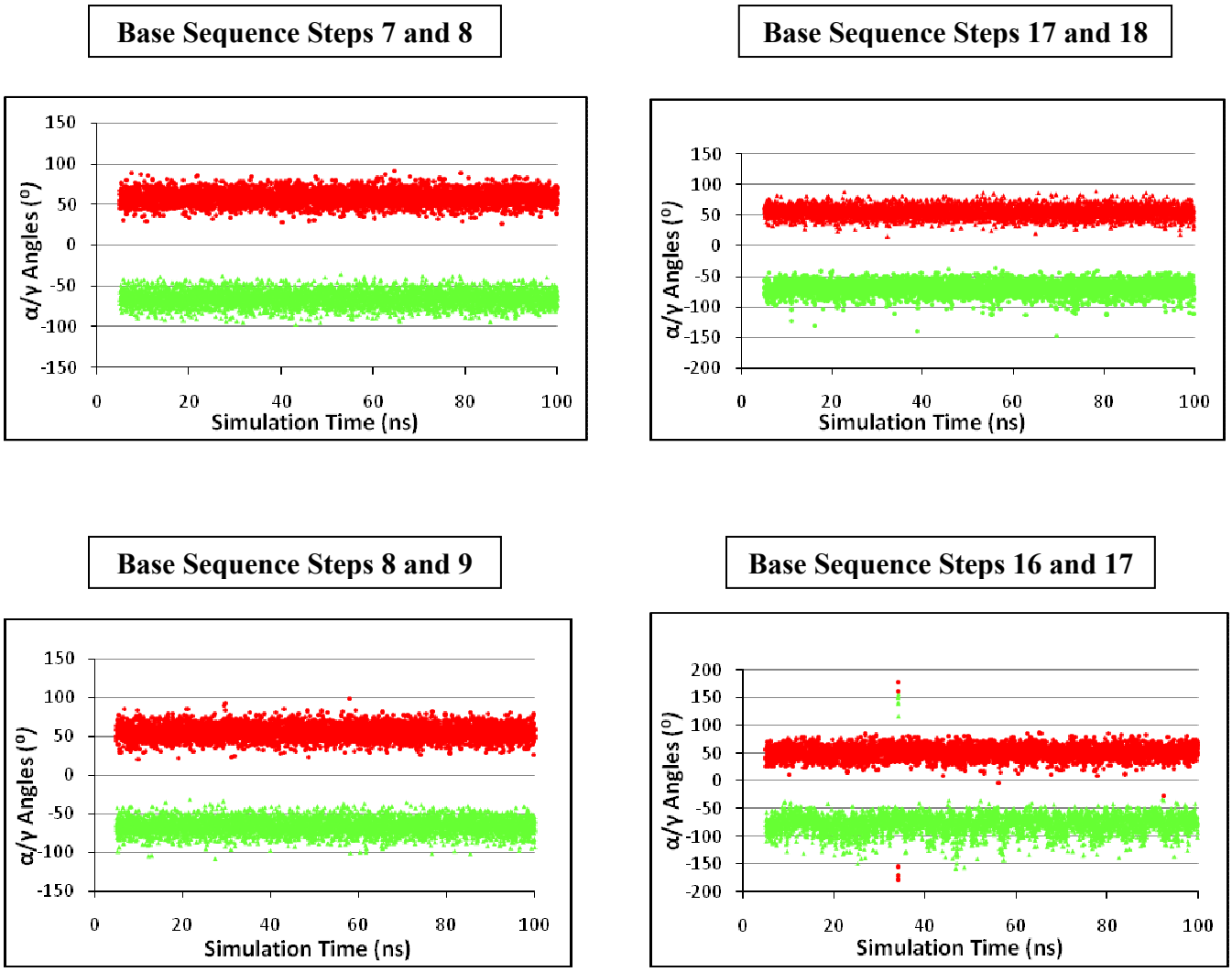


Figure S8: α/γ torsional angles for the five central basepairs across the 5 ns to 100 ns simulation. The left and right columns represent two separate strands with the respective base pairs. α is represented in green and γ in red in all the plots.

As seen in Figure S8, α/γ reversible flips occur in the peripheral regions of the binding site. This can be observed in the cases of α/γ plots for base sequence steps 4-5 and 16-17 above. In the remaining plots, α/γ angles sample canonical values and exhibit a conformational rigidity as compared to the original study[14]. This can be largely attributed to the substantial change in flexibility due to the presence of DB921 in the minor groove. The non-covalent interaction between the DNA and DB921 induces a conformational restriction in the central binding site and thus does not allow α/γ flips to occur across these bases.

References

1. Leach, A. R. *Molecular Modelin: Principles and Applications,* Second Edition; Prentice-Hall: Harlow, England, 2001.

2. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

3. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620-9631.

4. Case, D. A.; et al. *AMBER 9.0*; University of California: San Fransisco, 2006.

5. Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.

6. Frisch, M. J.; et al. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford CT, 2004.

7. Spackova, N.; Cheatham, T. E., 3rd ; Ryjacek, F.; Lankas, F.; van Meervelt, L.; Hobza, P.; Sponer, J. *J. Am. Chem. Soc.* **2003**, *125*, 1759-1769.

8. Hopfinger, A. J.; Pearlstein, R. A. *J. Comput. Chem.* **1984**, *5*, 486-499.

9. Howard, A.; Ross, B. Scripps Institute. http://amber.scripps.edu/doc6/html/AMBER-sh-19.3.html (accessed September 27, 2006).

10. *SYBYL 6.9.2*, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.

11. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

12. Pophristic, V.; Vemparala, S.; Ivanov, I.; Liu, Z.; Klein, M. L.; DeGrado, W. F. *J. Phys. Chem. B* **2006**, *110*, 3517-3726.

13. *Kaleidograph*, Synergy Software: Reading, 2006.

14**.** Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., 3rd; Laughton, C. A.; Orozco, M. *Biophys J.* **2007**, *92*, 3817-3829.