

Molecular modeling and structural studies of 12-mer immobile four-way DNA junction in solution

Ramesh Kumar Yadav¹ & Umesh Yadava^{2*}

¹Department of Physics, B.R.D. P.G. College, Deoria 274001, India; ²Department of Physics, DDU Gorakhpur University, Gorakhpur 273009, India; Umesh Yadava - E mail: u_yadava@yahoo.com; Phone: +91-551-2332398; *Corresponding author

Received May 22, 2014; Accepted May 26, 2014; Published July 22, 2014

Abstract:

Molecular modelling and structural studies of 12-mer immobile four-way DNA junction model is reported here. The DNA junction which was built and investigated, consisted of the following sequences 5'd(GGAAGGGGCTGG), 5'd(CCAGCCTGAGCC), 5'd(GGCTCAACTCGG) and 5'd(CCGAGTCCTTCC). The model was made in such a way that the junction may lack two-fold sequence symmetry at the crossover point. A new version of the AMBER force field has been used, in addition to the Particle Mesh Ewald (PME) method which deals with the refinement treatment of the long range interaction potentials, the well known limitation in MD protocol. After molecular dynamics simulation the backbone parameters and helical parameters of the DNA junction model is calculated and its dynamical pathway is discussed. A close observation near the junction point reveals the shifting in the orientation of some of the P-O bonds from the usual π_3 turn for A- and B- DNA to either π_1 or π_2 type of turn in order to achieve conformational stability. With this study it seems possible to derivatize synthetic DNA molecules with special functional groups both on the bases and at the backbones as in the case of some natural processes by which drugs, particular proteins *etc.* recognizes and binds to the specific sites of DNA.

Keywords: DNA Junction, Molecular Modeling, AMBER, Dynamical pathway.

Background:

Genetic recombination is one of the fundamental processes of biology, leading to genetic variation and diversity. Most of the mechanisms proposed for recombination proceed via a four arm branched DNA intermediate, termed as Holliday junction [1-3]. Survey of literature indicates that the structure of the Holliday junction has a central role in determining the outcome of the recombination event. Thus, in order to understand genetic recombination properly, a detailed knowledge of the structure of Holliday junction is required. Since Holliday junction has a property to undergo branch migration [4], the junction point is translated along stretches of duplex DNA. To avoid this problem, Seeman and Kallenbach [5] proposed the study of synthetic analogs of Holliday junctions in which the sequence is designed in such a way that the branch point is immobile. These synthetic immobile Holliday junctions consist of four stranded

branched DNA structure, whose double helical arms are stacked in two domains: two of the strands are roughly helical and the other two cross over in between.

The significance of the formation and existence of DNA junctions has been confirmed by many researchers [3]. One of the critical stages of recombination is the resolution process, where unconnected duplexes are regenerated. This is brought about by resolvase enzymes, which recognize the structure of the junction and cleave it. It has been shown that the binding of metal ions alter the conformation of such junctions and increase their thermal stability [6, 7]. The first clue as to the three dimensional shape of the junction came from the work of Cooper and Hagermann [8] who studied the relative electrophoretic mobilities of a small synthetic junction to which reporter arms had been ligated in various combinations. They

concluded that the four oligonucleotide strands comprising the junction were structurally non equivalent, their configuration in space being dictated by the sequences that flank the junctional branch point. Duckett *et al.* [9] proposed a general structure for four-way helical junction in DNA, the so called X-structure [10]. They employed a Gel Electrophoretic method similar to that used by Cooper & Hagerman [8] to estimate the relative angles subtended between different arms of a junction, and fluorescence energy transfer to estimate relative end-to-end distances for the arms. The most significant feature of the stacked X-structure was coaxial pair wise stacking between adjacent arms, generating two quasi-continuous helices. There were two ways in which this pairing may occur, and they showed that the local sequence at the midpoint of the junction determines which of the two possible isomers is more stable. The X-structure proposed by them was consistent with the results of hydroxyl-radical probing of small synthetic junctions [11]. The two quasi-continuous helices subtended an angle estimated to be 60°, the X-structure generated was right-handed and the sequence aligned in an anti parallel manner. The cleavage of the junction by the resolvase T4 endonuclease VII was sensitive to the isomeric structure adopted, indicating that the outcome of a recombination event might be influenced by the sequence and structure of the junction formed [9]. Churchill *et al.* probed the equilibrium structure of a junction by means of hydroxyl radicals generated by the reaction of iron (II) EDTA with hydrogen peroxide [11]. The hydroxyl radical-cleavage pattern showed twofold symmetry throughout the molecule, by which they concluded that the junction was a twofold symmetric complex whose four arms form the two stacking domains. In order to understand the structural and biological functions of these biomolecules, molecular modeling has become an important tool in the design and construction of new nucleotide sequences.

A number of MD simulations on B-form DNA oligonucleotides using various force fields have been reported [12-14]. The results to date have provided considerable information on the ability of MD to simulate DNA accurately, as well as details of the dynamical structure of various DNA models. But very little work on the DNA junction model has been reported as yet. Advances in computer power now permits MD to be performed on increasingly realistic models of DNA system, explicitly including solvent water and counter-ions, and to be extended from the pico second well into the nanosecond time frame. In order to find out the possibility of stacking of an arm over the other arm of the junction, we started with a structure which forms same angles with the other two arms. This DNA Junction model study may be used to predict hitherto, unknown properties of genetic recombination and replication process in *in-vivo* systems as well as some futuristic applications of synthetic DNA junction models.

Methodology:

Generation of coordinates

The coordinates of DNA junction were generated according to the ideal geometry with the help of AMBER EDIT module, using following sequences 5'd(GGAAGGGGCTGG), 5'd(CCAGCCTGAGCC), 5'd(GGCTCAACTCGG), and 5'd(CCGAGTCCTICC). The sequence topology for the DNA junction is given in **Figure 1**. The residue numbers are in 5' to 3'

direction with the first strand number 1-12, the second strand 13-24, the third strand 25-36 and the fourth strand 36-48. The DNA junction model has four strands, each having 12 residues. The six base pairs in the first arm are, for example 5' d(G1-G2-A3-A4-G5-G6) and its complement 5' d(C43-C44-T45-T46-C47-C48). Similar trend is followed for the other three arms as well. The first and the third strands are the outer strands of the junction model and they are roughly helical. The second and the fourth strands form the crossover arms and are at the inner portion of the junction model. The second and fourth arm of the DNA-junction model are subtending same angle from their midpoint, and they form the junction crossover in the middle of the model.

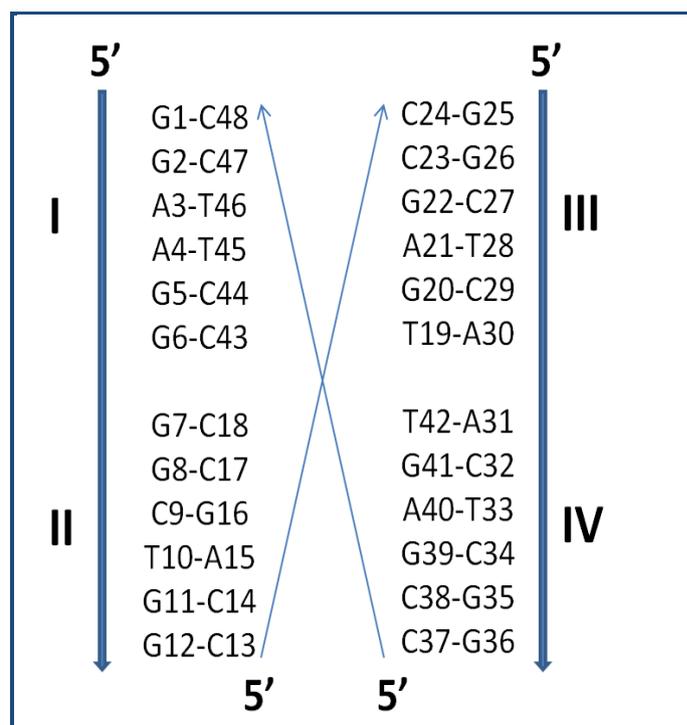


Figure 1: Sequence topology of DNA-Junction model

Placement of counter-ions and water

Forty four Sodium counter ions with cut off distance of 2.0Å were placed at a distance of 4.5Å from the phosphorous atom along the phosphate bisectors. The DNA and the counter ions were "immersed" in a Monte-Carlo-equilibrated, periodic transferable intermolecular potential (TIP3P) water bath. The monopole charges used were derived using 6-31G* basis set. The system thus constructed had 1515 solute atoms, 44 sodium ions and 21705 solvent atoms. The water bath contained 23264 atoms in a cubic box having dimensions of 66.3Å x 61.0Å x 55.0Å.

The AMBER force field [15] parameters were applied in the simulation together with the Particle Mesh Ewald (PME) method [16, 17], for evaluating long-range electrostatic forces. The Verletleap frog algorithm [18] was used in the numerical integration with a 1.0fs time step length for minimization and 2.0fs for dynamics. Hydrogen-atom bonds were constrained to "ideal" lengths. Temperature and pressure were maintained via the Berendsen algorithms [19]. All simulations were run with

the SANDER module of AMBER with SHAKE (tolerance =0.00001Å).

Equilibration

The entire system was subjected to 10000 steps of steepest-descent minimization followed by conjugate-gradient minimization with a satisfactory convergence of rms gradient of 0.1 kcal/mole-Å, keeping bonds involving hydrogen atoms constrained. The velocities were assigned at 100K to all atoms according to Boltzmann's distribution function. The system was heated from 100K to room temperature 298K over 2.0ps by

scaling velocities according to Berendsen algorithm followed by a constant-pressure-dynamics run over 25ps. After the initial equilibration, all subsequent simulations were run by using the Particle Mesh Ewald (PME) method, with the help of AMBER5.0 program with a cubic B-spline interpolation order. The restrained molecular dynamics was run over 600ps, during this dynamics the harmonic restraints were gradually reduced. Molecular dynamics production run were initiated at 298K, without any constraints on the system and was performed for 1250 ps.

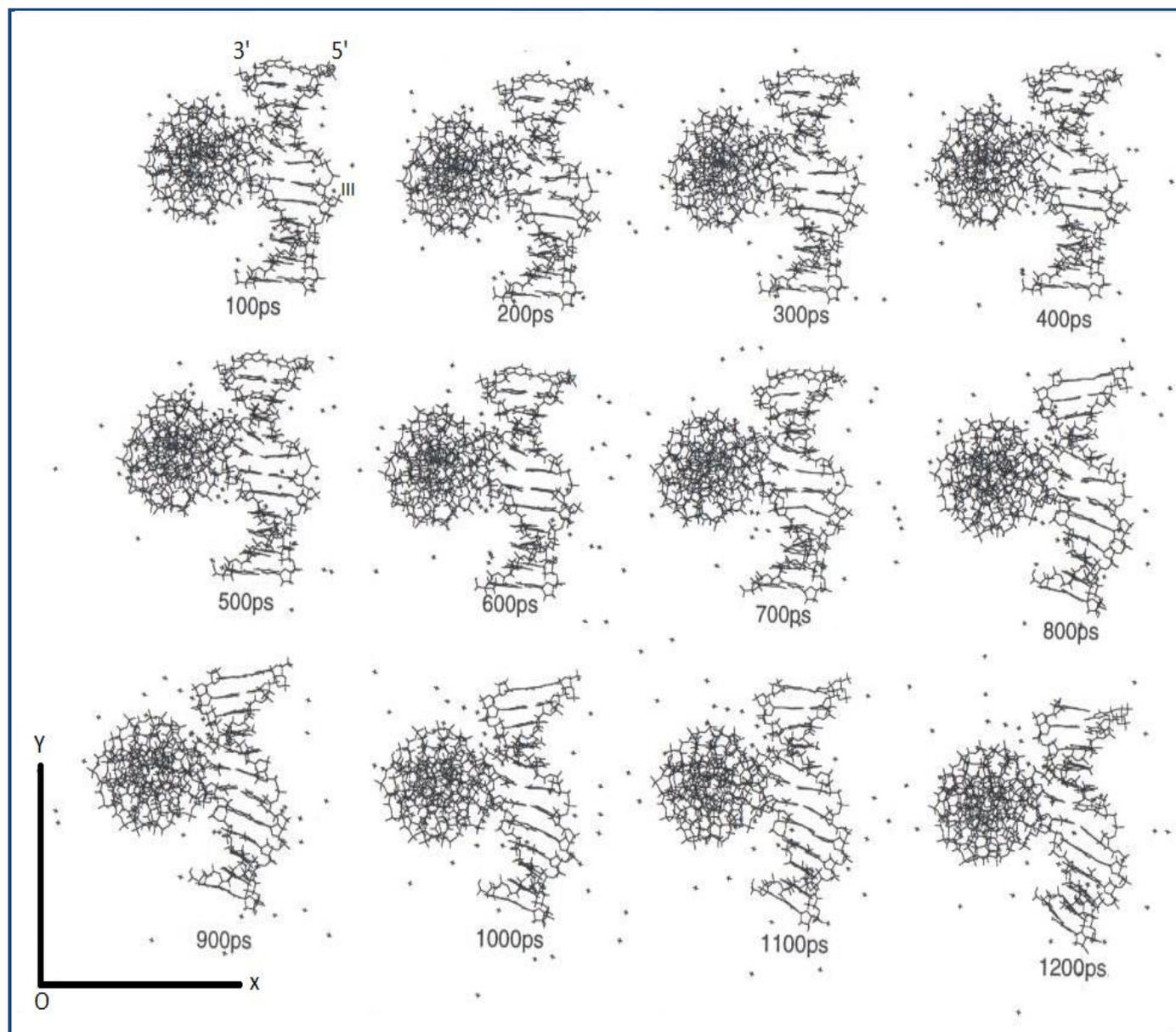


Figure 2: A series of snap shots taken after every 100 ps of DNA Junction during MD simulation

Results & Discussion:

Dynamic structure of the DNA junction model

A series of twelve snapshots taken from the MD trajectory at periodic intervals of 100 ps is shown in (Figure 2). Figure 3 shows the snap shots when the model is rotated by 90° along the x-axis. The variation in DNA junction especially at the

central region of junction model can be seen over the entire course of the simulation. The 100 ps snapshot shows that the base-pairs at the middle portion of the junction are planar. The snapshot taken at 200ps indicates a change in the orientation of the base-pair plane with bases coming closer to each other. The next snapshot at 300ps seems to indicate that the junction is

regaining its initial structure. Only after the 800 ps snapshot onwards, a clear indication of folding of the DNA junction model is observed, with clear changes in the model becoming obvious at the subsequent snapshots taken at 900, 1000, 1100 and 1200 ps. A comparison of the snapshots at 1200 ps with the 100 ps snapshot clearly shows a major change in the shape of the junction model not only at the middle portion but also at the end base-pair. It appears that the third strand has got stretched

from the 5' towards the 3' end. No such stretching in the first strand is apparent from the snapshots in **Figure 3**. In the third strand, the middle part of the junction consists of A: T base pair while in the first strand the middle part consists of G:C base-pairs. The presence of three hydrogen bonds in G: C base-pairs is probably keeping the middle portion of the first strand intact along the helix-axis thereby maintaining the planarity vis-a-vis the A:T base-pair of the third strand.

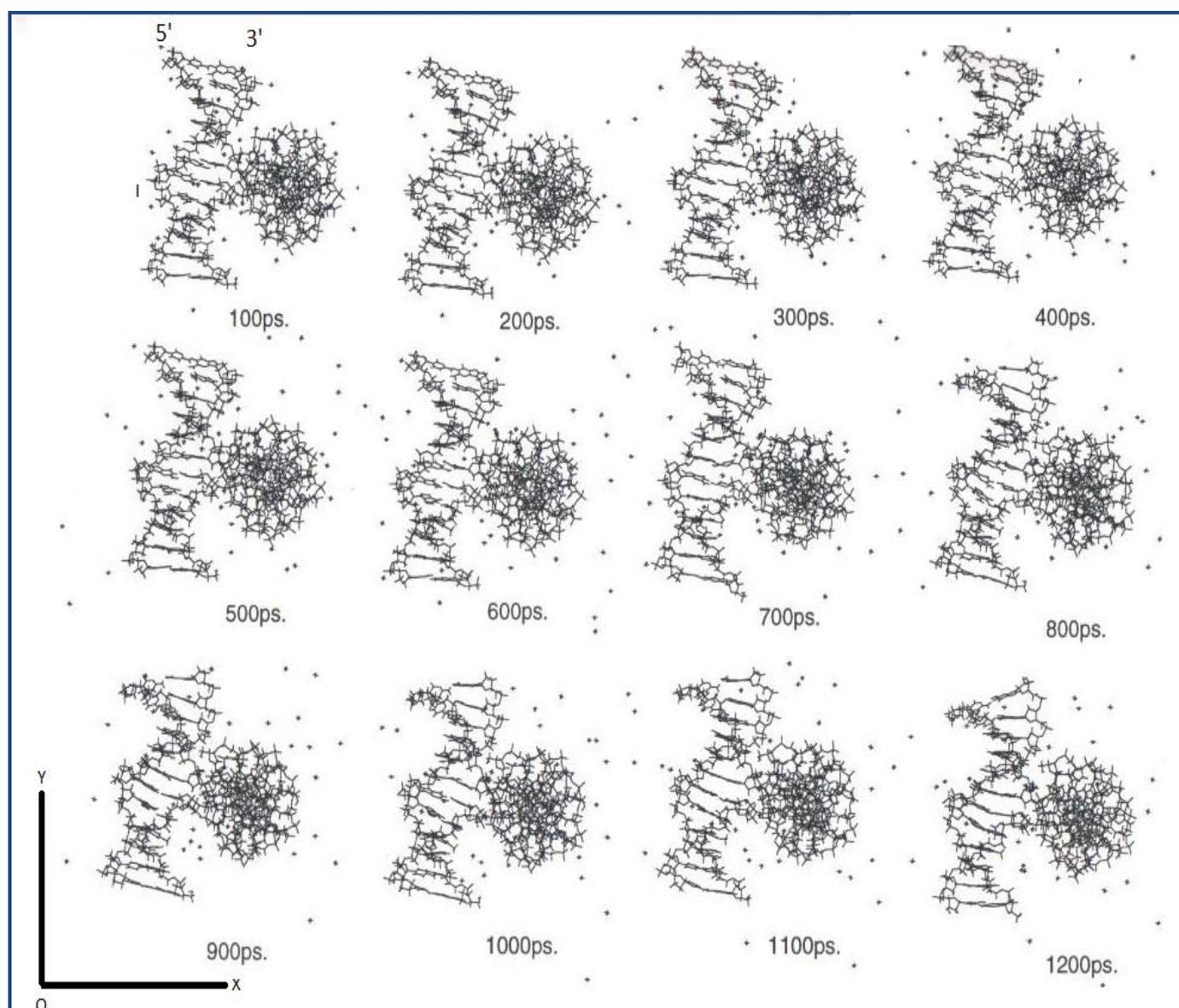


Figure 3: A series of snapshots taken after every 100 ps of DNA Junction (inverted) during MD simulation.

RMS deviation of MD structures

To show whether the structures were fully converged, the root-mean-square (RMS) deviations of MD structure as a function of time (from 600 to 1257 ps) with respect to MD average structure are shown in (**Figure 4**) for all atoms, backbone-atoms and base-atoms. During the first 20 ps range the RMSDs are, larger because at 600 ps point the system was heated from 100K to 298K. The system seems to have stabilized at 298K in about 20 ps (600 to 620 ps). Between these points constrained (5 kcal/mole-Å) MD was done. Beyond 1000 ps all the constraints

were removed allowing greater flexibility to the atoms which obviously resulted in relatively larger RMSDs. Beyond 1000 ps till the the end of simulation (1250 ps) complete convergence has been achieved. **Figure 4** further indicates that the rms deviation for the atoms, when the system is constrained, varies in a small range ($\sim \pm 0.060$) Once the constraints are removed the RMS deviations increase ($\sim 0.150\text{\AA}$) for the next period of 100 ps before attaining a stable configuration in the final simulation of ~ 150 ps when the RMS deviation is ~ 0.140 ps. The RMS deviation for backbone and base atoms are in close proximity of

all atoms RMS deviation. In totality, the small all-atom RMS deviation indicates that the DNA junction model structure is considerably converged

Torsion angles at the junction point

The torsion angles α , β , γ , δ , ϵ and ζ for the initial DNA junction model and for averaged simulated structure are given in **Table 1** (see supplementary material). For the G6 residue the only significant change is in the value of β , – from the initial -ac to +ap. In G7, major changes are observed for α , β , and γ , while for δ , ϵ , and ζ there are very slight changes. The torsion angles α , β , and γ assume the -ac, +ap and +sc conformation for this nucleotide from the initial +sp, -ac, -ac respectively. Residue C18 and T19 are the nucleotides belonging to the crossover strand. For C18 a major change in torsion angle is observed for β which, after, a rotation of 50° has gone into the +ap region from the initial -ac. All the other torsion angles α , β , γ and ζ oscillate in a narrow range. At T19 major changes are seen for α , β , γ and ϵ while the values of δ and ζ remain almost unchanged. These nucleotides assume the -ac, +ap, and -ap conformation from the initial +sc, -ac, +ap respectively.

Residues A30 and A31 belong to the roughly helical third strand of the DNA junction model. For A30 major changes are seen for β , δ and ζ values which have acquired +ap, -sp and +sc conformations respectively from the initial -ac, +ap and -ac conformation. For A31 almost all the angles are changing with the exception of ζ . From the initial +sc, -ac, +sc, +ap and -sc these have gone to -ac, g+, -ap, +ac, and -sc respectively. The Residue T42 and C43 belong to the junction point of the crossover strand, which is part of the fourth strand. In T42 nucleotide major change in torsion angles are observed for β , γ , δ and ϵ , which assumes -sc, -ac, +sp and +ap conformation from the initial -ac, +sc, +ap and -ap. For residue C43, major changes are seen for α , β and ϵ which from the initial +ac, -ac and +ap have shifted to -sc, +ap, and -ap respectively.

The comparison indicates that for the first and second strands, the change in the values of β is correlated in the sense that both the residues G6 and C18 belong to the 5' end of their respective strands. And along the 3' side also G7 and T19 belong to the first and second strand respectively. Both the strands show changes in identical angles, that is α , β and γ with the exception of ϵ for T19. In the case of third and fourth strands there is correlation along the 5' end with the exception of γ , and ϵ and along the 3' end (A31 and C43) there are relative changes in α , β and ϵ with the exception of γ and δ at A31. It is noteworthy that at the junction point with in a strand, from 5' to 3' direction α assumed the values (g^- , -ac) except for the fourth strand which assumes (g^-, g^-) conformation. The β torsion angles assume extended t conformation except for A31 and T42 which assumes g^+ and g^- conformation respectively. The conformation parameter g assumes +sc conformation for most of the nucleotides at the junction point except for T19, A31 and T41 which assumed -ap, -ap and -ac conformations. The exocyclic torsion angle d assumes +ac conformation for most of the nucleotides at the junction except for A30 and T42 which show cis conformation in board terms and -sp and +sp respectively

when defined by Klyne-Prelog notations. The torsion angles ϵ and ζ show correlation at the junction points also where ϵ assumes values in \pm (ap) region and ζ assumes values in \pm (ac) or its vicinity in the \pm (sc) region except for C18 nucleotides.

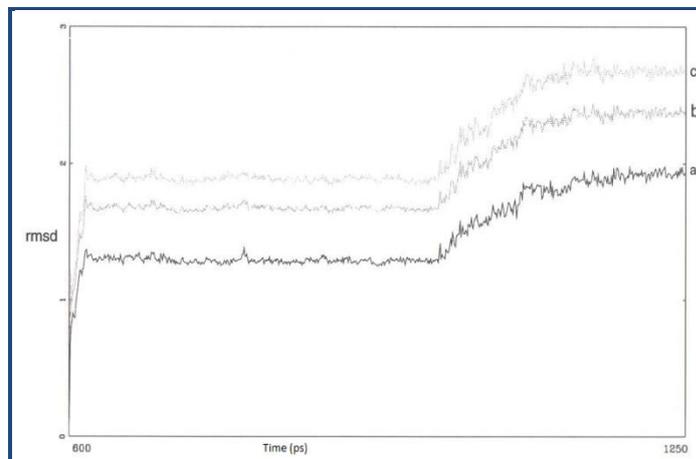


Figure 4: RMS deviation of the average MD simulated Structure as a function of time. a- rmsd for the base atoms, b-rmsd for all atoms and c-rmsd for backbone atoms

Conclusion:

This computational study point out that particle mesh Ewald (PME) method effectively eliminates the usual cut-off and allows user to deal with the refinement treatment of the long range interaction potentials, which was considered as one of the serious limitation in MD protocol. In DNA junction model, the four arms of the junction show B-DNA conformation which is in accord with the experimental observations done so far. The strands containing A-T base pairs at the junction seem to be more flexible, and stretching appears to be from 5' to 3' direction. The strand containing G-C base pairs remain intact at the junction points. At the 5' end, the atoms of the strands show more fluctuations. Near the junction point there is shifting in the orientation of some of the P-O bonds from the usual π_3 turn for A- and B- DNA to either π_1 or π_2 type of turn in order to achieve conformational stability. Finally, we can say that three-dimensional multiply-connected DNA structures, in the time to come, are expected to be of immense use in the assembly of molecular electronic devices, the formation of macromolecular-scale zeolites to host biological complexes for diffraction analysis, development of new catalysts and it seems possible to derivatize synthetic DNA molecules with special functional groups both on the bases and at the backbones as in the case of some natural processes by which drugs, particular proteins *etc.* recognizes and binds to the specific sites of DNA.

Acknowledgement:

R.K. Yadav is much indebted to his research supervisor Professor A.K. Pant for providing valuable guidance and computational facilities. U. Yadava is thankful to Department of Science and Technology(DST), New Delhi for providing financial support through SERC Fast Track Young Scientist scheme (SR/FT/CS-78/2010).

References:

- [1] Holliday R, *Genet Res.* 1964 5: 282

- [2] Lilley DM, *Q Rev Biophys.* 2000 **33**: 109 [PMID: 11131562]
[3] Mandal PK *et al.* *Acta Crystallogr Sect F Struct Biol Cryst Commun.* 2011 **F67**: 1506 [PMID: 22139153]
[4] Biswas I *et al.* *J Mol Biol.* 1998 **279**: 795 [PMID: 9642061]
[5] Seeman NC & Kallenbach NR, *Biophys J.* 1983 **44**: 201 [PMID: 6197102]
[6] Diekmann S & Lilley DMJ, *Nucleic Acids Res.* 1987 **15**: 5765 [PMID: 3039465]
[7] Douglas SM *et al.* *Nature* 2009 **459**: 414 [PMID: 19458720]
[8] Cooper JP & Hagerman PJ, *J Mol Biol.* 1987 **198**: 711 [PMID: 2828644]
[9] Duckett DR *et al.* *Cell* 1988 **55**: 79 [PMID: 3167979]
[10] Khuu PA *et al.* *Mol Recognit.* 2006 **19**: 234 [PMID: 16575941]
[11] Churchill MEA *et al.* *Proc. Nat Acad Sci USA.* 1998 **85**: 4653 [PMID: 3387432]
[12] Yang MS & Millar DP, *Biochemistry* 1997 **35**: 7959 [PMID: 8672499]
[13] Hays FA *et al.* *J Biol Chem.* 2003 **278**: 49663 [PMID: 14563836]
[14] Yadav RK & Yadava U, *Med Chem Res.* 2014 **23**: 280
[15] Cornell WD *et al.* *J Am Chem Soc.* 1995 **117**: 5179
[16] Essmann U *et al.* *J Chem Phys.* 1995 **98**: 10089
[17] Verlet L, *Phys Rev.* 1967 **159**: 98
[18] Miyamoto S & Kollman PA, *J Comput Chem.* 1992 **13**: 952
[19] Ryckaert JP *et al.* *J Comput Phys.* 1977 **23**: 327

Edited by P Kanguane

Citation: Yadav RK & Yadav U, *Bioinformation* 10(7): 394-400 (2014)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: Backbone torsion angles of the junction-point for the initial structure and, the final averaged simulated structure.

Torsion angles	α	β	γ	δ	ϵ	ζ
A-DNA	-50	172	41	79	-146	-78.00
B-DNA	-46	-147	36	157	155	-96.00
G6 initial	-46.81	-146.02	36.36	156.40	-155	-76.40
G6 final	-79.98	155.06	45.40	144.70	-176.08	-95.76
G7 initial	12.89	-146.02	-104.35	156.40	-155.00	-95.19
G7 final	-95.76	178.89	51.54	128.68	-167.91	-86.64
C18 initial	-46.81	-146.02	36.36	156.40	-70.02	-124.29
C18 final	-62.54	173.79	62.37	142.82	-78.58	-111.74
T19 initial	48.83	-146.02	36.36	156.40	155	-95.19
T19 final	-95.84	178.77	-178.70	144.87	-167.54	-95.71
A30 initial	-46.81	-146.02	36.36	156.40	168.71	-96.50
A30 final	-68.58	173.41	64.01	-22.97	167.44	86.26
A31 initial	42.71	-146.02	34.63	156.40	-74.87	-95.19
A31 final	-112.57	83.06	173.11	94.59	-164.50	-82.39
T42 initial	-46.81	-146.02	36.36	156.40	-168.38	169.96
T42 final	-44.17	-42.17	-143.94	28.17	158.49	134.92
C43 initial	144.59	-146.02	36.36	156.40	155.00	-95.19
C43 final	-52.29	176.02	63.57	138.49	-153.34	-84.14