

RESEARCH ARTICLE

Selective Preference of Parallel DNA Triplexes Is Due to the Disruption of Hoogsteen Hydrogen Bonds Caused by the Severe Nonisostericity between the G*GC and T*AT Triplets

Gunaseelan Goldsmith^{1,2}, Thenmalarchelvi Rathinavelan^{1*}, Narayanarao Yathindra^{1*}

1 Institute of Bioinformatics and Applied Biotechnology, Biotech Park, Electronics City Phase I, Bangalore, India, **2** Manipal University, Manipal, India

✉ Current address: Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana State, India

* yathindra@ibab.ac.in



OPEN ACCESS

Citation: Goldsmith G, Rathinavelan T, Yathindra N (2016) Selective Preference of Parallel DNA Triplexes Is Due to the Disruption of Hoogsteen Hydrogen Bonds Caused by the Severe Nonisostericity between the G*GC and T*AT Triplets. PLoS ONE 11 (3): e0152102. doi:10.1371/journal.pone.0152102

Editor: Freddie Salisbury, Jr, Wake Forest University, UNITED STATES

Received: January 11, 2016

Accepted: March 8, 2016

Published: March 24, 2016

Copyright: © 2016 Goldsmith et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by an institutional grant to IBAB from the Department of Electronics and Information Technology, Government of India, under 'Centre of Excellence for Research and Teaching in Bioinformatics' [DIT-R&D-BIO/15(7)/2007] and Department of IT, BT and S&T, Government of Karnataka, India. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Implications of DNA, RNA and RNA-DNA hybrid triplexes in diverse biological functions, diseases and therapeutic applications call for a thorough understanding of their structure-function relationships. Despite exhaustive studies mechanistic rationale for the discriminatory preference of parallel DNA triplexes with G*GC & T*AT triplets still remains elusive. Here, we show that the highest nonisostericity between the G*GC & T*AT triplets imposes extensive stereochemical rearrangements contributing to context dependent triplex destabilisation through selective disruption of Hoogsteen scheme of hydrogen bonds. MD simulations of nineteen DNA triplexes with an assortment of sequence milieu reveal for the first time fresh insights into the nature and extent of destabilization from a single (non-overlapping), double (overlapping) and multiple pairs of nonisosteric base triplets (NIBTs). It is found that a solitary pair of NIBTs, feasible either at a G*GC/T*AT or T*AT/G*GC triplex junction, does not impinge significantly on triplex stability. But two overlapping pairs of NIBTs resulting from either a T*AT or a G*GC interruption disrupt Hoogsteen pair to a noncanonical mismatch destabilizing the triplex by ~10 to 14 kcal/mol, implying that their frequent incidence in multiples, especially, in short sequences could even hinder triplex formation. The results provide (i) an unambiguous and generalised mechanistic rationale for the discriminatory trait of parallel triplexes, including those studied experimentally (ii) clarity for the prevalence of antiparallel triplexes and (iii) comprehensive perspectives on the sequence dependent influence of nonisosteric base triplets useful in the rational design of TFO's against potential triplex target sites.

Competing Interests: The authors have declared that no competing interests exist.

Abbreviations: G_W , A_W , T_W and C_W , Guanine, cytosine, adenine and thymine bases of the WC duplex; G_H and T_H , Guanine and thymine bases of the parallel Hoogsteen strand (third strand); G_{RH} and T_{RH} , Guanine and thymine bases of the antiparallel reverse Hoogsteen strand (third strand); TFOs, Triplex forming oligonucleotides.

Introduction

It is well documented that Watson and Crick paired DNA duplex interacts with sequence specific oligonucleotides to form triple helices stabilised by either Hoogsteen (parallel triplex) or reverse Hoogsteen (antiparallel triplex) pair of hydrogen bonds. Evidences for involvement of triple helices in biological processes come from their participation in regulation of gene expression [1, 2], DNA damage and repair [3, 4], RNA processing and folding [5, 6, 7] and chromatin organization [8]. Triplexes are also known to impair DNA polymerization and influence DNA recombination process [9]. Base triple interactions crucial for function are shown to be present in pseudoknots in telomerase RNA [10] and in transcripts during programmed ribosomal frame shifting in viruses like SARS coronavirus [11]. Most recently, intramolecularly folded RNA triplexes are shown to be formed in the highly abundant and up-regulated long noncoding transcripts such as metastasis associated lung adenocarcinoma transcript 1 (MALAT1) and others to evade exonuclease action [12, 13]. Formation of RNA.DNA hybrid triplexes involving noncoding RNA found to repress genes like DHFR, TGF- β and MAT2A has further enhanced the biological importance of nucleic acid triplexes [14, 15, 16]. Existence of triplexes *in vivo* is supported by the discovery of triplex unwinding helicases [17, 18], triplex-specific antibodies [19, 20] and other endogenous triplex specific proteins [21–24]. Association of triple helices with colorectal cancer [25], neurodegenerative disorder Friedreich's ataxia [26] and a number of inherited as well as acquired human diseases [27] further underscore their profound significance.

DNA triplexes have evoked potential technological applications which include creating high quality DNA vectors for human gene therapy [28], nanomachines for monitoring intracellular pH gradient [29], molecular switches [30,31,32], and for developing drug delivery systems [33,34]. They have also been used as biosensors for the recognition and analysis of toxic metal ions [35], SNPs [36], DNA methylation [37] and cancer cells [38]. Quite recently, DNA triplex formation process has been exploited in developing PCR based biosensor for detection of pathogens [39], high-throughput assays for measuring activity of DNA topoisomerases and other enzymes involved in DNA topology modification [40], controlled assembly of liposomes [41] and as modular probes for DNA detection [42]. Recognition of abundant putative triplex target sites in both prokaryotic and eukaryotic genomes [43, 44, 45], together with new experimental approaches for identification of triple helices [46–50] and development of algorithm/databases for analysis of triplex target sites [51, 52, 53] highlight their increasing relevance in the post genomic era. Several reviews detailing the biological significance [2, 54–60] and potential therapeutic applications [61] of triplexes have been appearing to underscore their importance. Notwithstanding, structural information on triple helices, specifically those addressing the issue of base triplet nonisostericity and their effects, crucial in the design of Triplex forming Oligonucleotides (TFOs) for targeting duplex is rather inadequate.

It is well recognised that TFOs rich in (i) T & C⁺, (ii) G & A, (iii) G & T can interact with purine rich strand of DNA forming a pair of hydrogen bonds to result in T·AT, C⁺·GC, G·GC and A·AT base triplets. Predictably, T·AT & C⁺·GC triplets are isosteric (structurally alike) like the Watson and Crick pairs, while the other triplet pair combinations such as T·AT & G·GC, T·AT & G·GC and A·AT & G·GC are nonisosteric (structurally dissimilar) [56, 62]. While T & C rich TFOs favour parallel triplex [63, 64], those with G & A favour antiparallel triplex [65, 66]. Although TFOs with G & T form both parallel and antiparallel triplexes, they predominantly favour the latter [67–70]. Interestingly, both intramolecular [67, 71] and intermolecular [62, 72–74] parallel triplexes containing a limited number of G·GC & T·AT triplet pairs are observed. They are shown to be effective in interfering with biological functions in several instances. For instance, GT rich 38-mer and 15-mer TFOs inhibit transcription of HIV-1 in

infected human cells [75] and replication in SV40 containing plasmid in COS-1 cells [76] respectively, through formation of parallel triplexes. On the other hand, the 22-mer TFO with 10 T·AT & G·GC juxtapositions, and the 36-mer TFO with 21 T·AT & G·GC juxtapositions, targeted against the promoter region of human *Ki-ras* [77] and the human epidermal growth factor receptor gene [78] respectively, are unable to form parallel triplex, but forms only anti-parallel triplex. Seemingly then, selective ability to form parallel triplex appears to be determined by the number of G·GC & T·AT juxtapositions, their larger incidences, thwarting parallel triplex formation. While this may be notionally attributed to nonisosteric traits between them, their quantitative estimates and precise mechanistic effects on the ability and stability for triplex formation are obscure.

A cursory glance at the superposition of G·GC and T·AT triplet pairs (Fig 1A and 1B), readily point to the existence of an intrinsically large nonisostericity (defined by the residual twist angle, Δt°) in parallel compared to antiparallel orientation. We argue that this might be responsible in imparting differential mechanistic influence in parallel vis-à-vis antiparallel triplexes causing selective preference of the former. To probe this, extensive MD simulations of a variety of parallel and anti-parallel DNA triplexes formed by these triplets, in different sequence contexts (Sequences 1–19 in Table 1) have been carried out. Results have yielded clear insights into the role of nonisostericity on triplex stability, stereochemical reasoning for the prevalence of antiparallel triplex and selective preference for parallel triplex. These should aid in comprehensive understanding of sequence dependent structure of nucleic acid triplexes and their implication in biological processes.

Superposition of G·GC (green) and T·AT (brown) triplets in parallel (A) and anti-parallel (B) orientation to bring out the nature and source of nonisostericity between triplets. Triplets are depicted as circles formed with the three C1' atoms of the triplet bases lying on the circumference. Appearance of a twist between the adjacent Hoogsteen bases (G_H & T_H) or reverse Hoogsteen bases (G_{RH} & T_{RH}), even prior to the application of formal helical twist (t) constitutes the residual twist Δt° . This together with the radial difference Δr Å between the two superposed triplets, suffice to quantify nonisostericity between them. Note the higher value of Δt° in parallel (A) than in antiparallel (B) triplets. Illustration of the effect of over winding ($30^\circ + \Delta t^\circ$) at the $G_H T_H$ step (C) and under winding ($30^\circ - \Delta t^\circ$) at the $T_H G_H$ step (D) of the Hoogsteen strand consequent to residual twist. Two adjacent G·GC triplets with the ideal twist angle of (t) 30° (E) is shown to underscore the effect of residual twist.

Results

Manifestation of the highest nonisostericity between the parallel G*GC & T*AT triplets

Superposition of the parallel G·GC & T·AT base triplets reveals (Fig 1A) misalignment of the C1' atoms of the Hoogsteen strand (see also S1A Fig). This results in the largest value for the residual twist angle $\Delta t^\circ (\pm 21.6^\circ)$ between the G . . G and A . . T Hoogsteen pairs. It is obvious that the effect of residual twist would be to impart over winding ($+\Delta t^\circ$) at the $G_H T_H$ step (G·GC preceding T·AT) (Fig 1C) and under winding ($-\Delta t^\circ$) at the $T_H G_H$ step (T·AT preceding G·GC) in the Hoogsteen strand of the triplex (Fig 1D). Consequently effective triple helical twist would be $30^\circ (t^\circ) + 21.6^\circ (+\Delta t^\circ) = 51.6^\circ$ and $30^\circ (t^\circ) - 21.6^\circ (-\Delta t^\circ) = 8.4^\circ$ at the GT and TG steps of the Hoogsteen strand respectively. 51.6° is close to nearly twice the value of triple helical twist of $30^\circ (t^\circ)$, while 8.4° is an exceptionally low value for triple helical twist. Outcome of these creates disconnects of nearly one nucleotide length (5.5 Å to 6.3 Å) in the sugar-phosphate backbone at the successive steps of the Hoogsteen strand (S1B Fig). They also generate severe steric overlap between the adjacent sugars, especially, at the under wound TG step.

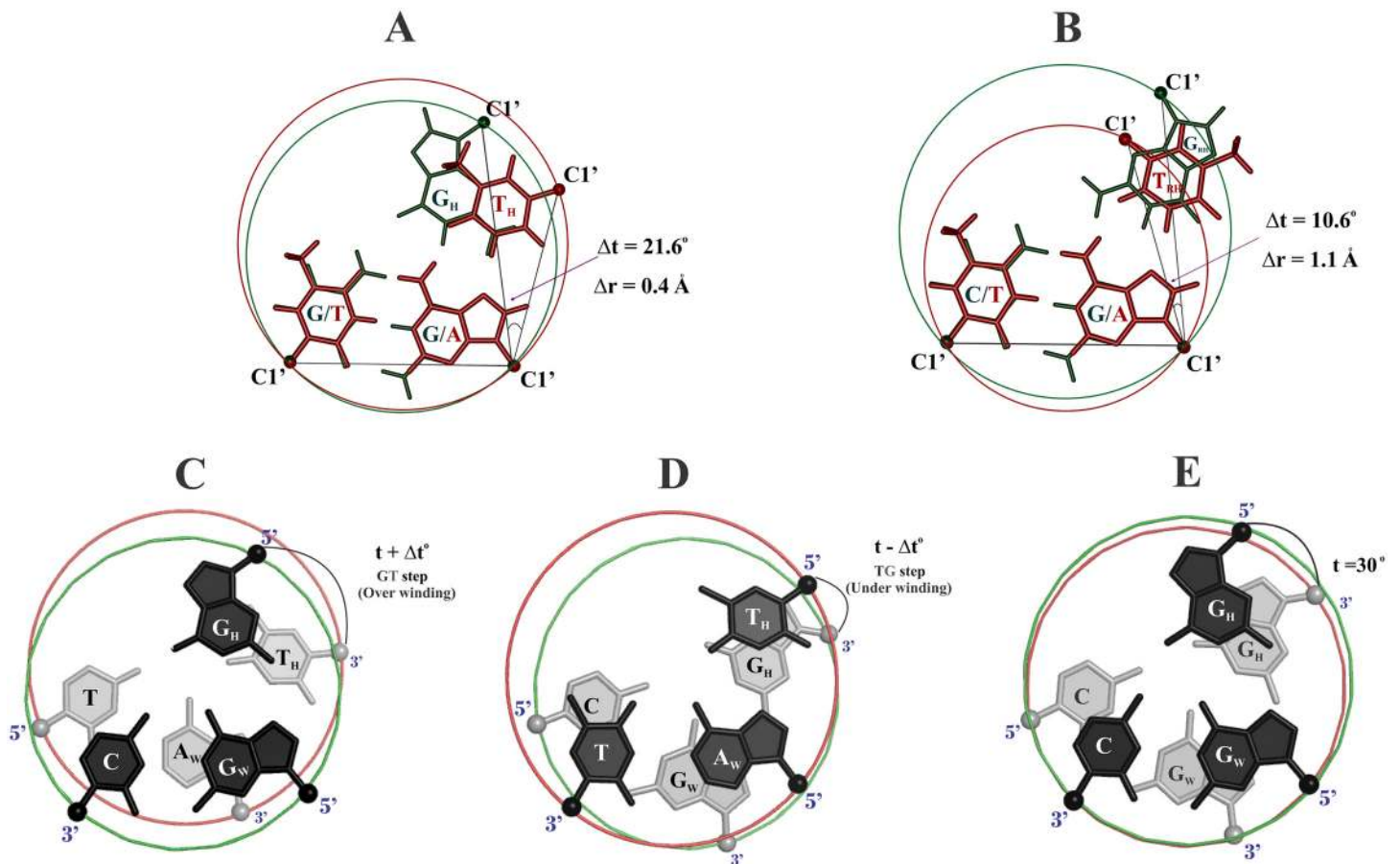


Fig 1. Nature, degree of base triplet nonisostericity and their consequence.

doi:10.1371/journal.pone.0152102.g001

Bridging these disconnects and ensuring a steric free conformation, while retaining Hoogsteen hydrogen bonds in the G-GC & T-AT triplets, demand considerable conformational rearrangements (see below). Incidentally, the other nonisostericity parameter viz., radial difference (Δr Å) is relatively small ($\cong 0.4$ Å) for the parallel (Fig 1A) compared to the antiparallel base triplets ($\cong 1.1$ Å) (Fig 1B). Given this, it is expected that Δt° would impact more than Δr Å in a parallel DNA triplex.

Successive incidence of G*GC & T*AT triplets destabilizes parallel DNA triplex by disrupting Hoogsteen hydrogen bonds

It might be anticipated that the conformational changes necessitated by the unusually high and low twist angles at the alternating $G_H T_H$ and $T_H G_H$ steps might be absorbed by the inherent flexibility of the triplex. In order to examine this, MD simulations (250 ns) have been carried out on a 15-mer parallel triplex comprising alternating G-GC & T-AT triplets (Sequence 1 in Table 1). The results at once indicate a proclivity for the loss of Hoogsteen hydrogen bonds in both the triplets (Fig 2A): N6 (A_W) . . . O4 (T_H) and N7 (A_W) . . . N3 (T_H) hydrogen bonds in the T-AT triplets are absent for over 52% and 49% of the simulation period respectively, while N7 (G_W) . . . N2 (G_H) and O6 (G_W) . . . N1 (G_H) hydrogen bonds are absent in the G-GC triplets for over 64% and 47.7% respectively. This is caused by the movement of T_H and G_H to overcome the effects of high Δt° . In the process, Hoogsteen hydrogen bonds are disrupted. In

Table 1. DNA triplexes considered for the study.

Sequence No. (Oligomer size)	DNA Triplex Sequence	No. of GGC/TAT triplets	No. of GT/TTG Steps	No. and type of NIBT pairs	Simulation Time (ns)	Binding Free energy (kcal/mol)
1. (15-mer)	5' C ₁ T ₂ C ₃ T ₄ C ₅ T ₆ C ₇ T ₈ C ₉ T ₁₀ C ₁₁ T ₁₂ C ₁₃ T ₁₄ C ₁₅	8/7	7GT 7TTG	14 Overlapping	250	-9.4
	3' G ₃₀ A ₂₉ G ₂₈ A ₂₇ G ₂₆ A ₂₅ G ₂₄ A ₂₃ G ₂₂ A ₂₁ G ₂₀ A ₁₉ G ₁₈ A ₁₇ G ₁₆					
	3' G ₄₅ T ₄₄ G ₄₃ T ₄₂ G ₄₁ T ₄₀ G ₃₉ T ₃₈ G ₃₇ T ₃₆ G ₃₅ T ₃₄ G ₃₃ T ₃₂ G ₃₁ TFO					
2. (15-mer) Antiparallel	5' C ₁ T ₂ C ₃ T ₄ C ₅ T ₆ C ₇ T ₈ C ₉ T ₁₀ C ₁₁ T ₁₂ C ₁₃ T ₁₄ C ₁₅	8/7	7GT 7TTG	14 Overlapping	100	-70.8
	3' G ₃₀ A ₂₉ G ₂₈ A ₂₇ G ₂₆ A ₂₅ G ₂₄ A ₂₃ G ₂₂ A ₂₁ G ₂₀ A ₁₉ G ₁₈ A ₁₇ G ₁₆					
	5' G ₃₁ T ₃₂ G ₃₃ T ₃₄ G ₃₅ T ₃₆ G ₃₇ T ₃₈ G ₃₉ T ₄₀ G ₄₁ T ₄₂ G ₄₃ T ₄₄ G ₄₅ TFO					
3. (11-mer)	5' T ₁ T ₂ T ₃ T ₄ T ₅ T ₆ T ₇ T ₈ T ₉ T ₁₀ T ₁₁	11	0	homopolymer	100	-63.6
	3' A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆ A ₁₅ A ₁₄ A ₁₃ A ₁₂					
	3' T ₃₃ T ₃₂ T ₃₁ T ₃₀ T ₂₉ T ₂₈ T ₂₇ T ₂₆ T ₂₅ T ₂₄ T ₂₃ TFO					
4. (11-mer)	5' T ₁ T ₂ T ₃ T ₄ T ₅ C ₆ T ₇ T ₈ T ₉ T ₁₀ T ₁₁	1/10	1TG 1GT	2 Overlapping	100	-53.5
	3' A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆ A ₁₅ A ₁₄ A ₁₃ A ₁₂					
	3' T ₃₃ T ₃₂ T ₃₁ T ₃₀ T ₂₉ G ₂₈ T ₂₇ T ₂₆ T ₂₅ T ₂₄ T ₂₃ TFO					
5. (11-mer)	5' T ₁ T ₂ C ₃ T ₄ T ₅ C ₆ T ₇ C ₈ T ₉ T ₁₀ T ₁₁	2/9	2TG 2GT	4 Overlapping	100	-40.2
	3' A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆ A ₁₅ A ₁₄ A ₁₃ A ₁₂					
	3' T ₃₃ T ₃₂ T ₃₁ G ₃₀ T ₂₉ T ₂₈ T ₂₇ G ₂₆ T ₂₅ T ₂₄ T ₂₃ TFO					
6. (11-mer)	5' T ₁ T ₂ C ₃ T ₄ T ₅ C ₆ T ₇ T ₈ C ₉ T ₁₀ T ₁₁	3/9	3TG 3GT	6 Overlapping	100	-25.9
	3' A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆ A ₁₅ A ₁₄ A ₁₃ A ₁₂					
	3' T ₃₃ T ₃₂ G ₃₁ T ₃₀ T ₂₉ G ₂₈ T ₂₇ T ₂₆ G ₂₅ T ₂₄ T ₂₃ TFO					
7. (11-mer)	5' C ₁ C ₂ C ₃ C ₄ C ₅ C ₆ C ₇ C ₈ C ₉ C ₁₀ C ₁₁	11	0	homopolymer	100	-62.0
	3' G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ G ₁₇ G ₁₆ G ₁₅ G ₁₄ G ₁₃ G ₁₂					
	3' G ₃₃ G ₃₂ G ₃₁ G ₃₀ G ₂₉ G ₂₈ G ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ TFO					
8. (11-mer)	5' C ₁ C ₂ C ₃ C ₄ T ₅ C ₆ C ₇ C ₈ C ₉ C ₁₀ C ₁₁	10/1	1GT 1TTG	2 Overlapping	100	-50.6
	3' G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ A ₁₇ G ₁₆ G ₁₅ G ₁₄ G ₁₃ G ₁₂					
	3' G ₃₃ G ₃₂ G ₃₁ G ₃₀ G ₂₉ T ₂₈ G ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ TFO					
9. (11-mer)	5' C ₁ C ₂ C ₃ T ₄ C ₅ C ₆ C ₇ T ₈ C ₉ C ₁₀ C ₁₁	10/1	2GT 2TTG	4 Overlapping	100	-39.3
	3' G ₂₂ G ₂₁ G ₂₀ A ₁₉ G ₁₈ G ₁₇ G ₁₆ A ₁₅ G ₁₄ G ₁₃ G ₁₂					
	3' G ₃₃ G ₃₂ G ₃₁ T ₃₀ G ₂₉ G ₂₈ T ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ TFO					
10. (15-mer)	5' T ₁ T ₂ T ₃ T ₄ T ₅ T ₆ T ₇ C ₈ C ₉ C ₁₀ C ₁₁ C ₁₂ C ₁₃ C ₁₄ C ₁₅	8/7	1GT	1 Non-Overlapping	100	-71.6
	3' A ₃₀ A ₂₉ A ₂₈ A ₂₇ A ₂₆ A ₂₅ A ₂₄ G ₂₃ G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ G ₁₇ G ₁₆					
	3' T ₄₅ T ₄₄ T ₄₃ T ₄₂ T ₄₁ T ₄₀ T ₃₉ G ₃₈ G ₃₇ G ₃₆ G ₃₅ G ₃₄ G ₃₃ G ₃₂ G ₃₁ TFO					
11. (15-mer)	5' C ₁ C ₂ C ₃ C ₄ C ₅ C ₆ C ₇ C ₈ T ₉ T ₁₀ T ₁₁ T ₁₂ T ₁₃ T ₁₄ T ₁₅	8/7	1TG	1 Non-Overlapping	100	-78.4
	3' G ₃₀ G ₂₉ G ₂₈ G ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆					
	3' G ₄₅ G ₄₄ G ₄₃ G ₄₂ G ₄₁ G ₄₀ G ₃₉ G ₃₈ T ₃₇ T ₃₆ T ₃₅ T ₃₄ T ₃₃ T ₃₂ T ₃₁ TFO					
12. (15-mer)	5' T ₁ T ₂ T ₃ C ₄ C ₅ C ₆ C ₇ C ₈ C ₉ C ₁₀ C ₁₁ T ₁₂ T ₁₃ T ₁₄ T ₁₅	8/7	1TG 1GT	2 Non-Overlapping	100	-57.2
	3' A ₃₀ A ₂₉ A ₂₈ G ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ G ₂₂ G ₂₁ G ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆					
	3' T ₄₅ T ₄₄ T ₄₃ G ₄₂ G ₄₁ G ₄₀ G ₃₉ G ₃₈ G ₃₇ G ₃₆ G ₃₅ T ₃₄ T ₃₃ T ₃₂ T ₃₁ TFO					

(Continued)

Table 1. (Continued)

Sequence No. (Oligomer size)	DNA Triplex Sequence	No. of GGC/TAT triplets	No. of GT/ TG Steps	No. and type of NIBT pairs	Simulation Time (ns)	Binding Free energy (kcal/mol)
13. (15-mer)	5' C ₁ C ₂ C ₃ C ₄ T ₅ T ₆ T ₇ T ₈ T ₉ T ₁₀ T ₁₁ C ₁₂ C ₁₃ C ₁₄ C ₁₅	8/7	1GT 1TG	2 Non-Overlapping	100	-55.8
	3' G ₃₀ G ₂₉ G ₂₈ G ₂₇ A ₂₆ A ₂₅ A ₂₄ A ₂₃ A ₂₂ A ₂₁ A ₂₀ G ₁₉ G ₁₈ G ₁₇ G ₁₆					
	3' G ₄ G ₄ G ₄ G ₄ T ₄₁ T ₄₀ T ₃₉ T ₃₈ T ₃₇ T ₃₆ T ₃₅ G ₃₄ G ₃₃ G ₃₂ G ₃₁ TFO					
14. (15-mer)	5' T ₁ T ₂ C ₃ C ₄ C ₅ C ₆ T ₇ T ₈ T ₉ C ₁₀ C ₁₁ C ₁₂ C ₁₃ T ₁₄ T ₁₅	8/7	2TG 2GT	4 Non-Overlapping	100	-44.3
	3' A ₃₀ A ₂₉ G ₂₈ G ₂₇ G ₂₆ G ₂₅ A ₂₄ A ₂₃ A ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ A ₁₇ A ₁₆					
	3' T ₄ T ₄ T ₄ G ₄ G ₄ G ₄ G ₄ T ₃₉ T ₃₈ T ₃₇ G ₃₆ G ₃₅ G ₃₄ G ₃₃ T ₃₂ T ₃₁ TFO					
15. (25-mer)	5' C ₁ C ₂ C ₃ C ₄ T ₅ T ₆ T ₇ T ₈ T ₉ C ₁₀ T ₁₁ T ₁₂ T ₁₃ T ₁₄ T ₁₅ C ₁₆ C ₁₇ C ₁₈ C ₁₉ C ₂₀ T ₂₁ C ₂₂ C ₂₃ C ₂₄ C ₂₅	14/11	3GT 3TG	4 Overlapping & 2 Non-Overlapping	100	-81.2
	3' G ₅₀ G ₄₉ G ₄₈ G ₄₇ A ₄₆ A ₄₅ A ₄₄ A ₄₃ A ₄₂ G ₄₁ A ₄₀ A ₃₉ A ₃₈ A ₃₇ A ₃₆ G ₃₅ G ₃₄ G ₃₃ G ₃₂ G ₃₁ A ₃₀ G ₂₉ G ₂₈ G ₂₇ G ₂₆					
	3' G ₇₅ G ₇₄ G ₇₃ G ₇₂ T ₇₁ T ₇₀ T ₆₉ T ₆₈ T ₆₇ G ₆₆ T ₆₅ T ₆₄ T ₆₃ T ₆₂ T ₆₁ G ₆₀ G ₅₉ G ₅₈ G ₅₇ G ₅₆ T ₅₅ G ₅₄ G ₅₃ G ₅₂ G ₅₁ TFO					
16. (11-mer) Antiparallel	5' C ₁ C ₂ C ₃ C ₄ C ₅ C ₆ C ₇ C ₈ C ₉ C ₁₀ C ₁₁	11	0	homopolymer	100	-64.2
	3' G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ G ₁₇ G ₁₆ G ₁₅ G ₁₄ G ₁₃ G ₁₂					
	5' G ₂₃ G ₂₄ G ₂₅ G ₂₆ G ₂₇ G ₂₈ G ₂₉ G ₃₀ G ₃₁ G ₃₂ G ₃₃ TFO					
17. (11-mer) Antiparallel	5' C ₁ C ₂ C ₃ C ₄ C ₅ T ₆ C ₇ C ₈ C ₉ C ₁₀ C ₁₁	10/1	1GT 1TG	2 Overlapping	100	-61.4
	3' G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ A ₁₇ G ₁₆ G ₁₅ G ₁₄ G ₁₃ G ₁₂					
	5' G ₂₃ G ₂₄ G ₂₅ G ₂₆ G ₂₇ T ₂₈ G ₂₉ G ₃₀ G ₃₁ G ₃₂ G ₃₃ TFO					
18. (15-mer) Antiparallel	5' C ₁ C ₂ C ₃ C ₄ C ₅ C ₆ C ₇ C ₈ T ₉ T ₁₀ T ₁₁ T ₁₂ T ₁₃ T ₁₄ T ₁₅	8/7	1GT	1 Non-Overlapping	100	-76.4
	3' G ₃₀ G ₂₉ G ₂₈ G ₂₇ G ₂₆ G ₂₅ G ₂₄ G ₂₃ A ₂₂ A ₂₁ A ₂₀ A ₁₉ A ₁₈ A ₁₇ A ₁₆					
	5' G ₃₁ G ₃₂ G ₃₃ G ₃₄ G ₃₅ G ₃₆ G ₃₇ G ₃₈ T ₃₉ T ₄₀ T ₄₁ T ₄₂ T ₄₃ T ₄₄ T ₄₅ TFO					
19. (15-mer) Antiparallel	5' T ₁ T ₂ T ₃ T ₄ T ₅ T ₆ T ₇ C ₈ C ₉ C ₁₀ C ₁₁ C ₁₂ C ₁₃ C ₁₄ C ₁₅	8/7	1TG	1 Non-Overlapping	100	-82.3
	3' A ₃₀ A ₂₉ A ₂₈ A ₂₇ A ₂₆ A ₂₅ A ₂₄ G ₂₃ G ₂₂ G ₂₁ G ₂₀ G ₁₉ G ₁₈ G ₁₇ G ₁₆					
	5' T ₃₁ T ₃₂ T ₃₃ T ₃₄ T ₃₅ T ₃₆ T ₃₇ G ₃₈ G ₃₉ G ₄₀ G ₄₁ G ₄₂ G ₄₃ G ₄₄ G ₄₅ TFO					

List of parallel (Sequences 1, 3–15) and antiparallel (Sequences 2, 16–19) DNA triplexes studied in diverse sequential context. The 15-mer triplexes (Sequences 10–14 & 18–19) comprising of 8 G*GC & 7 T*AT triplets are rearranged to obtain a variety of triplex junctions under different sequence contexts.

doi:10.1371/journal.pone.0152102.t001

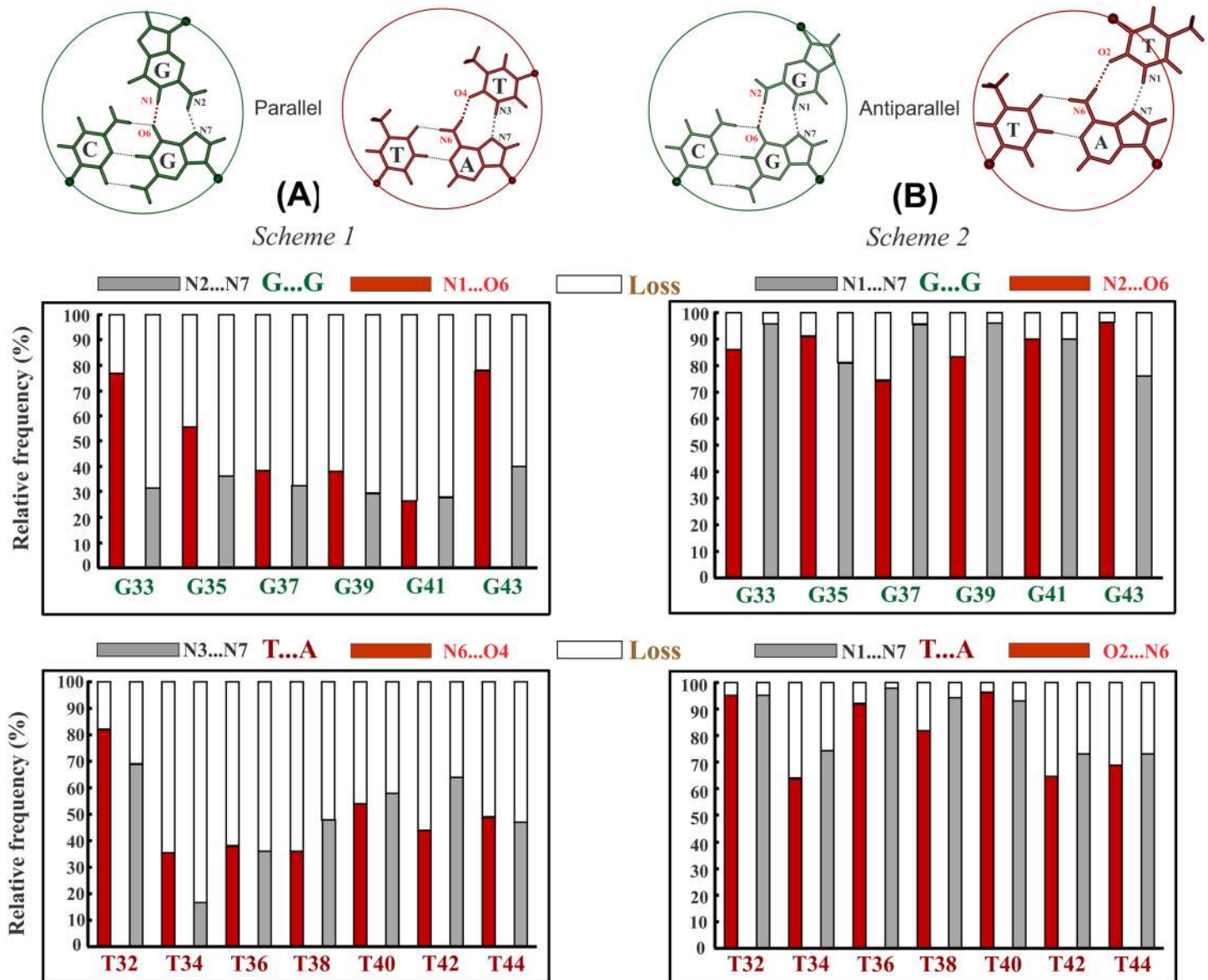


Fig 2. Demonstration of unstable and stable nature of parallel and antiparallel triplex formed by alternating G-GC & T-AT triplets respectively. **A** Frequency of incidence (red & gray filled part) and loss (void part) of canonical (A) Hoogsteen (Sequence 1) and (B) reverse Hoogsteen hydrogen bonds (Sequence 2) in the central 6 G-GC and 7 T-AT triplets of the 15-mer parallel and antiparallel triplex (terminal triplets not considered) over a simulation time of 250 and 100 ns respectively. Wide spread loss of Hoogsteen hydrogen bonds in the (A) parallel triplex (Sequence 1), and retention of reverse Hoogsteen hydrogen bonds in the (B) antiparallel triplex (Sequence 2) are apparent. Canonical Hoogsteen and reverse Hoogsteen hydrogen bonding schemes in the G-GC & T-AT base triplets in parallel & antiparallel orientations are shown on top for reference.

doi:10.1371/journal.pone.0152102.g002

striking contrast, the reverse Hoogsteen bonds of the G-GC & T-AT triplets in the antiparallel DNA triplex (Sequence 2) are largely preserved (Fig 2B). This can be attributed to the much lower value of Δt° (10°) in antiparallel, as opposed to 21.6° in the parallel triplex. Evidently much smaller conformational variations required by the lower residual twists are readily absorbed by the inherent triplex flexibility without the need for the loss of reverse Hoogsteen hydrogen bonds.

Emergence of noncanonical Hoogsteen schemes. Consequent to the disruption of Hoogsteen hydrogen bonds, alternative schemes of hydrogen bonds emerge for both the T·AT (NC1 to NC4; Fig 3D–3G) and G·GC (NC5 to NC8; Fig 3H–3K) triplets. These are referred to as noncanonical Hoogsteen schemes. For T·AT, NC1 scheme (Fig 3D) is predominant and is formed by way of water mediated interaction involving N3 of T_H and N6 & N7 of A_W. In NC2 (Fig 3E), both ion and water interact with T_H (O4 & N3) and N6 & N7 of A_W and O4 of T_H. Interaction via two water molecules describes NC3 (Fig 3F). In all of these none of the canonical Hoogsteen hydrogen bonds are retained. But in NC4 (Fig 3G), canonical N6...O4 Hoogsteen hydrogen bond scheme is preserved while a water molecule mediates the interaction between N3 of T_H and N7 of A_W.

In the case of G·GC triplet, NC5 scheme (Fig 3H) is predominant. This is characterised by bifurcated hydrogen bonds involving N2 of G_H and O6 & N7 of G_W. In the NC6 scheme (Fig 3I), a water molecule mediates interaction between G_H and WC paired bases. In both these situations, Hoogsteen G_H swivels causing the amino group to glide towards the WH groove. But, similar action in the direction of the CH groove results in the NC7 scheme (Fig 3J). Here, an ion comes in between O6's of G_H and G_W, while N7 of G_W forms bifurcated hydrogen bonds with N2 & N1 of G_H. NC8 scheme is alike NC5, but in addition, is mediated by a water molecule between N7 & N3 atoms of G_W and G_H (Fig 3K).

A common feature observed in all of the above is the propensity for reducing the initial large Δt° (21.6°) to lower values (as low as 1.2° in NC3; Fig 3G) to circumvent the large mechanistic effects (steric overlap and disconnect in the sugar-phosphate backbone) discussed above. This can happen only at the expense of losing canonical Hoogsteen hydrogen bonding. Such a loss in a large number of triplets would critically affect triplex stability. This is clearly reflected in the very low binding free energy of only—9.4 kcal/mol for the parallel triplex (Sequence 1) in sharp contrast to the comparatively high value of—70.80 kcal/mol for the antiparallel triplex (Sequence 2). The latter by and large retains all the reverse Hoogsteen hydrogen bonds in all the triplets. Thus, the parallel DNA triplex with consecutive G·GC and T·AT triplets is not expected to be stable due to the mechanistic effects of extreme nonisostericity at every base step of the Hoogsteen strand. Hence the discussion about this unstable structure is not pursued further.

Limited number of nonisosteric base triplets (NIBTs) does not impinge on triplex stability

In the 15-mer parallel DNA triplex (Sequence 1, Table 1) discussed above, each triplet say T·AT, is flanked on either side by the nonisosteric base triplet G·GC, giving rise to two pairs of consecutive and overlapping nonisosteric base triplets (NIBTs), G·GC & T·AT and T·AT & G·GC. The intervening T·AT triplet, being common to both, overlaps with both the flanking G·GC triplets, and it therefore in effect experiences nonisostericity effects due to the preceding and the following G·GC base triplet. Consequently, two Δt° 's, viz., $+\Delta t^\circ$ at the GT step (between the G·GC & T·AT triplets) and $-\Delta t^\circ$ at the TG step (between the T·AT & G·GC triplets) (Fig 1C) operate on the intervening T·AT triplet. We refer to this scheme as overlapping pair of triplets. In fact, the 15-mer triplex (Sequence 1; Table 1) may be deemed to be comprised of such trimer triplets constituting 14 pairs of overlapping NIBTs. The deleterious effect of which has been discussed above. But in reality a limited number of NIBTs may occur in the target duplex. Moreover, these may be spaced out rather than occurring alternately along the sequences. Such circumstances are in fact investigated [79,80]. These prompted us to study the effect of isolated and lone occurrence of such NIBTs. This situation can be visualized when a G·GC or T·AT triplet interrupts an otherwise homopolymeric triplex (Sequences 4 and 8; Table 1). It will be

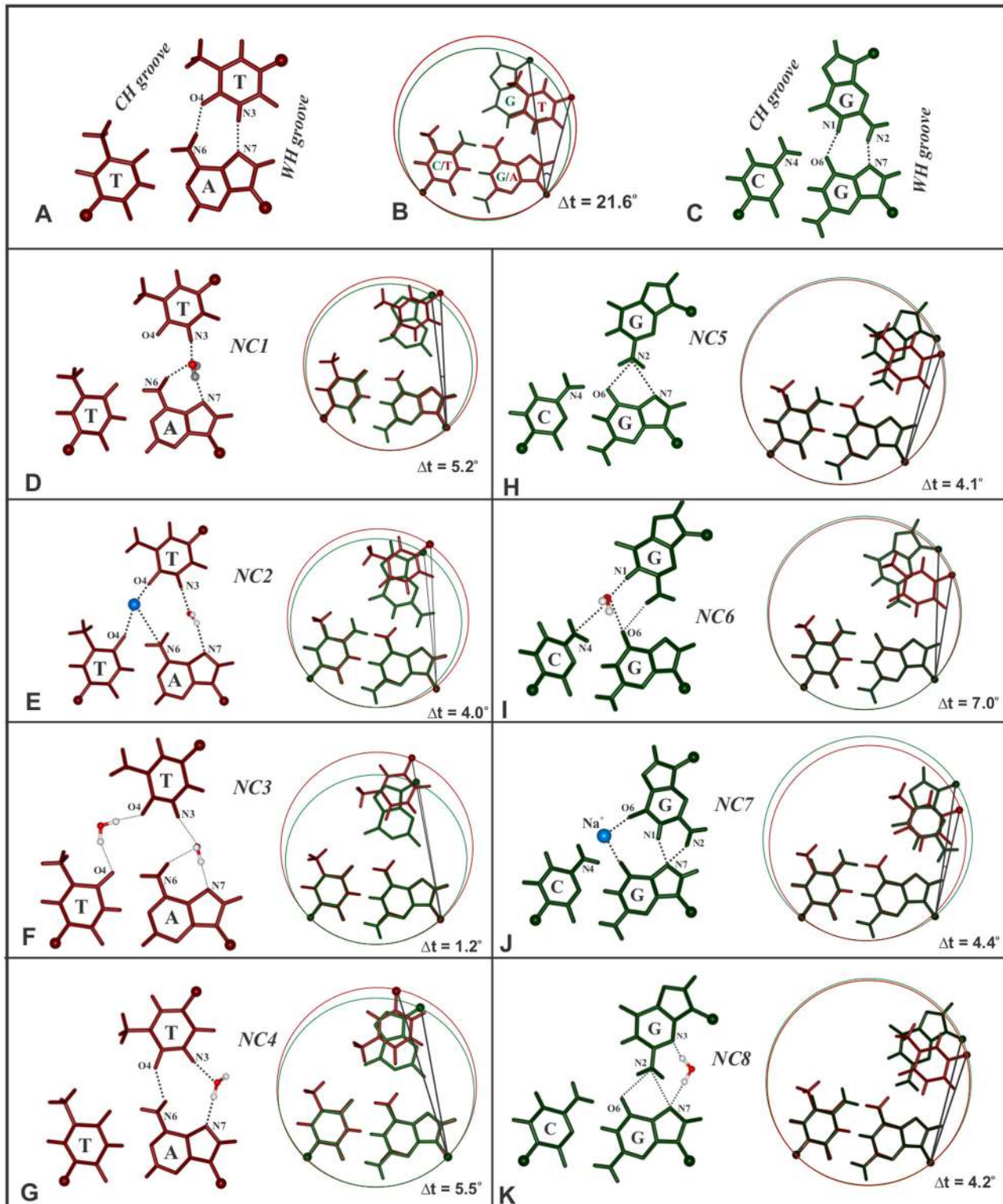


Fig 3. Transition from canonical to noncanonical Hoogsteen hydrogen bond schemes in the T-AT (NC1 to NC4) and G-GC (NC5 to NC8) triplets with reduced residual twists due to the effects of nonisostericity. Typical canonical Hoogsteen hydrogen bonding schemes in T-AT (brown), G*GC (green) triplets & their superposition are shown in A, C and B respectively. Noncanonical Hoogsteen hydrogen bonding schemes of T-AT (D,E,F,G) & G-GC (H,I,J,K) base triplets are shown along with their superposition with the canonical schemes to highlight reduction in the residual twist Δt° (right panel in D to K). Na^+ ion (E & J) and oxygen atom of water molecule (D-G, I & K) are coloured in cyan and red respectively.

doi:10.1371/journal.pone.0152102.g003

intriguing to examine if the Hoogsteen hydrogen bonding pair is retained here and if not, evaluate its influence on the triplex structure and stability.

Structure of a T*AT triplex with a G*GC interruption mediated by overlapping pairs of NIBTs ($-\Delta t^\circ$ followed by $+\Delta t^\circ$). It is readily seen that a G*GC triplet interruption in the parallel homopolymeric T*AT DNA triplex (Sequence 4; [Table 1](#)) generates two sets of NIBTs which are overlapping and consecutive. The first pair of NIBT at the TG ($T_{27}G_{28}$) step results in a residual twist of $-\Delta t^\circ = -21.6^\circ$ and effects under winding. This is followed by a residual twist of $+\Delta t^\circ = +21.6^\circ$ at the GT ($G_{28}T_{29}$) step to effect over winding. MD simulation of this triplex reveals that the interrupting $G_{28} \cdot G_{17} \cdot C_6$ triplet loses the N1...O6 hydrogen bond of the canonical Hoogsteen scheme ([Fig 4A](#) and [S2B Fig](#)) in favour of the noncanonical NC5 Hoogsteen scheme ([Fig 3H](#)). This clearly demonstrates that even an isolated nonisosteric triplet is unable to retain the canonical Hoogsteen pair and instantly transits to a noncanonical scheme (NC5) which reduces the Δt° values to $\sim -4.1^\circ$ and $+4.6^\circ$ at the $T_{27}G_{28}$ and $G_{28}T_{29}$ steps respectively from the initial value of 21.6° . As a result, slightly lower (20.2°) and higher (36.6°) helical twist angles are seen at these steps occurring on either sides of the G*GC interruption ([Fig 5A](#)). Stacking interactions between the Hoogsteen bases (G_{28} and T_{29}) are minimal at the GT step compared to that between T_{27} & G_{28} at the under wound TG step ([Fig 5B](#)). Better stacking between A_{16} & G_{17} and G_{17} & A_{18} and relatively poor stacking between T_5 & C_6 ([Fig 5B](#)) of WC paired duplex prevail. Twist angles at the TT Hoogsteen steps and the TC & CT steps of WC duplex lie around 29.1° .

Average value for the X-displacement of base pair at the interruption site (-3.6 \AA) is slightly higher compared to other places (-2.9 \AA). On the other hand, propeller twist of WC base pairs at the interruption site is slightly lower ($\sim 6.86^\circ$) compared to -9.6° found at other places. Sugar ring of G_{28} of the $G_{28} \cdot G_{17} \cdot C_6$ interruption assumes the C1' *exo* pucker, while thymidines of the Hoogsteen strand favour O4' *endo* pucker. All the torsion angles lie in their preferred ranges. Average value of minor groove width is lowered to 12.2 \AA from the initial value of 14 \AA concomitant with an increase in the major groove width by 1.5 \AA to a value of 20.1 \AA . Increase in the widths of WH and CH groove by 2 and 4 \AA to 9.3 \AA and 16.1 \AA respectively is also seen ([S3A Fig](#)). The binding free energy ($\sim -53.5 \text{ kcal/mol}$) of this triplex (Sequence 4) is reduced by $\sim 10 \text{ kcal/mol}$ compared to $\sim -63.6 \text{ kcal/mol}$ obtained for the uninterrupted homopolymeric T*AT triplex (Sequence 3; [S2A Fig](#)) pointing clearly to the destabilizing effect of base triplet interruption which is tolerated as a triplet mismatch. Incidentally, a T*AT triplex with a G*GC interruption forms a stable parallel triple helix [62].

Structure of a G*GC triplex with a T*AT interruption mediated by overlapping pairs of NIBTs ($+\Delta t^\circ$ followed by $-\Delta t^\circ$). Here, the roles of G*GC and T*AT triplets are swapped. A homopolymeric G*GC triplex is interrupted by the $T_{28} \cdot A_{17} \cdot T_6$ triplet, creating two overlapping and consecutive pairs of NIBTs (Sequence 8; [Table 1](#)). Contrary to the previous case, here the over winding GT step ($+\Delta t^\circ$) precedes the under winding TG step ($-\Delta t^\circ$) in the Hoogsteen strand. MD simulations show that the interrupting $T_{28} \cdot A_{17} \cdot T_6$ loses the canonical Hoogsteen N3 (T_H)...N7 (A_w) hydrogen bond for over 95% of total simulation time ([Fig 4B](#) and [S4B Fig](#)), while the O4 (T_{28})...N6 (A_{17}) hydrogen bond is lost for over 13% of total simulation time. Eventually, a water mediated noncanonical NC1 Hoogsteen scheme ([Fig 3D](#)) prevails, as a triplet mismatch, entailing reduced Δt° values of $+5.8^\circ$ and -6.1° at the Hoogsteen GT and TG steps respectively. Akin to the previous case (Sequence 4; [Table 1](#)), twist angles ([Fig 5C](#)), base stacking pattern ([Fig 5D](#)), groove width variations ([S3B Fig](#)) and base pair parameters exhibit variations at the interruption site similar to above. Average value of minor groove width is $\sim 13.26 \text{ \AA}$ with a reduced value of $\sim 12.1 \text{ \AA}$ at the interruption site ([S3B Fig](#)). Concomitantly, major groove widens by $\sim 2.5 \text{ \AA}$ to a value of $\sim 21.9 \text{ \AA}$. Likewise, CH groove widens by $\sim 1.5 \text{ \AA}$ to $\sim 12.3 \text{ \AA}$ with the WH groove remaining nearly unchanged at $\sim 9.9 \text{ \AA}$ ([S3B Fig](#)). The binding

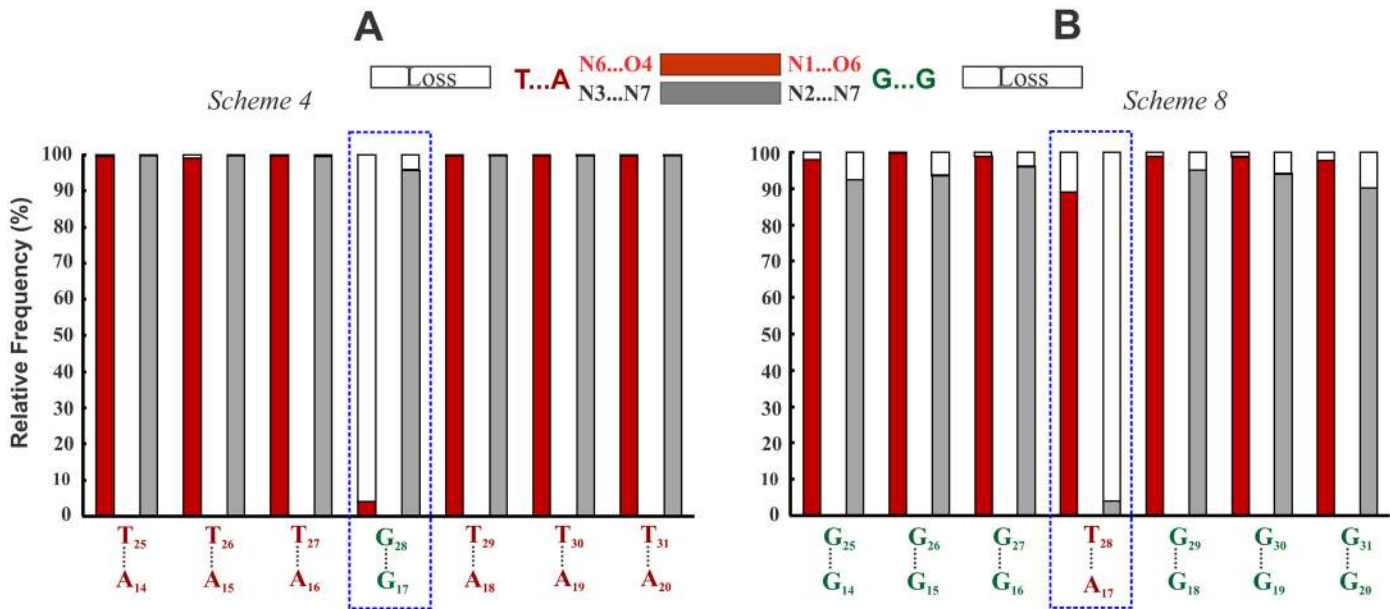


Fig 4. Destabilization of canonical Hoogsteen hydrogen bond in the G*GC and T*AT interruptions. Frequency of incidence (red & gray filled part) and loss (void part) of Hoogsteen hydrogen bonds in the G*GC interruption of the T*AT triplex (**A**) (Sequence 4) and in the T*AT interruption of the G*GC triplex (**B**) (Sequence 8). Loss of the canonical Hoogsteen hydrogen bond (void part) in the interrupting G*GC (N1...O6) and T*AT (N3...N7) triplets are conspicuous (blue box).

doi:10.1371/journal.pone.0152102.g004

free energy of this triplex with an interruption is reduced by ~ 11.4 kcal/mol to ~ -50.6 kcal/mol compared to ~-62.0 kcal/mol obtained for the uninterrupted homopolymeric G*GC triplex (Sequence 7; [S4A Fig](#)).

On the other hand, MD simulation of an 11-mer antiparallel G*GC triplex with (Sequence 17) and without a T*AT triplet interruption (Sequence 16) entails only a marginal loss of ~ 3 kcal/mol in the binding free energy due to the retention of canonical reverse Hoogsteen hydrogen bonds even in the T*AT interrupt ([S5 Fig](#)).

Multiple interruptions progressively destabilize the triplex. To envisage further the effect of multiple interruptions (NIBTs), a parallel T*AT DNA triplex with 2 and 3 G*GC interruptions, separated by 3 and 2 T*AT triplets (Sequence 5 and Sequence 6), are investigated by MD simulation. Expectedly, the canonical Hoogsteen hydrogen bonds are lost in all the G*GC interruptions, G₁₅...G₂₆ & G₁₉...G₃₀ in sequence 5 and in G₁₄...G₂₅, G₁₇...G₂₈ & G₂₀...G₃₁ in sequence 6 ([S2C and S2D Fig](#)) and figure as triplet mismatches. Other conformational changes are similar to those seen with only one interruption (Sequence 4). The binding free energy of the triplex with 2 and 3 G*GC interruptions (Sequence 5; -40.2 kcal/mol and Sequence 6; -25.9 kcal/mol; [Table 1](#)) is reduced further by ~13 and 27 kcal/mol respectively compared to the triplex with a single interruption (Sequence 4) indicating progressive destabilization of triplex with increasing number of interruptions. Likewise, 2 T*AT interruptions in the parallel G*GC DNA triplex (Sequence 9) lose the canonical Hoogsteen hydrogen bonds involving A₁₅...T₂₆ & A₁₉...T₃₀ pair ([S4C Fig](#)) concomitant with binding energy loss of ~-11 kcal/mol compared to the triplex with a single T*AT interruption (Sequence 8 vs. Sequence 9). These indicate progressive destabilisation of triplex with increasing number of interruptions with a near linear correlation ([S6 Fig](#)). This could impact triplex stability or even its formation. While longer sequences might be able to endure interruptions, shorter sequences with frequent interruptions may fail to form triplex.

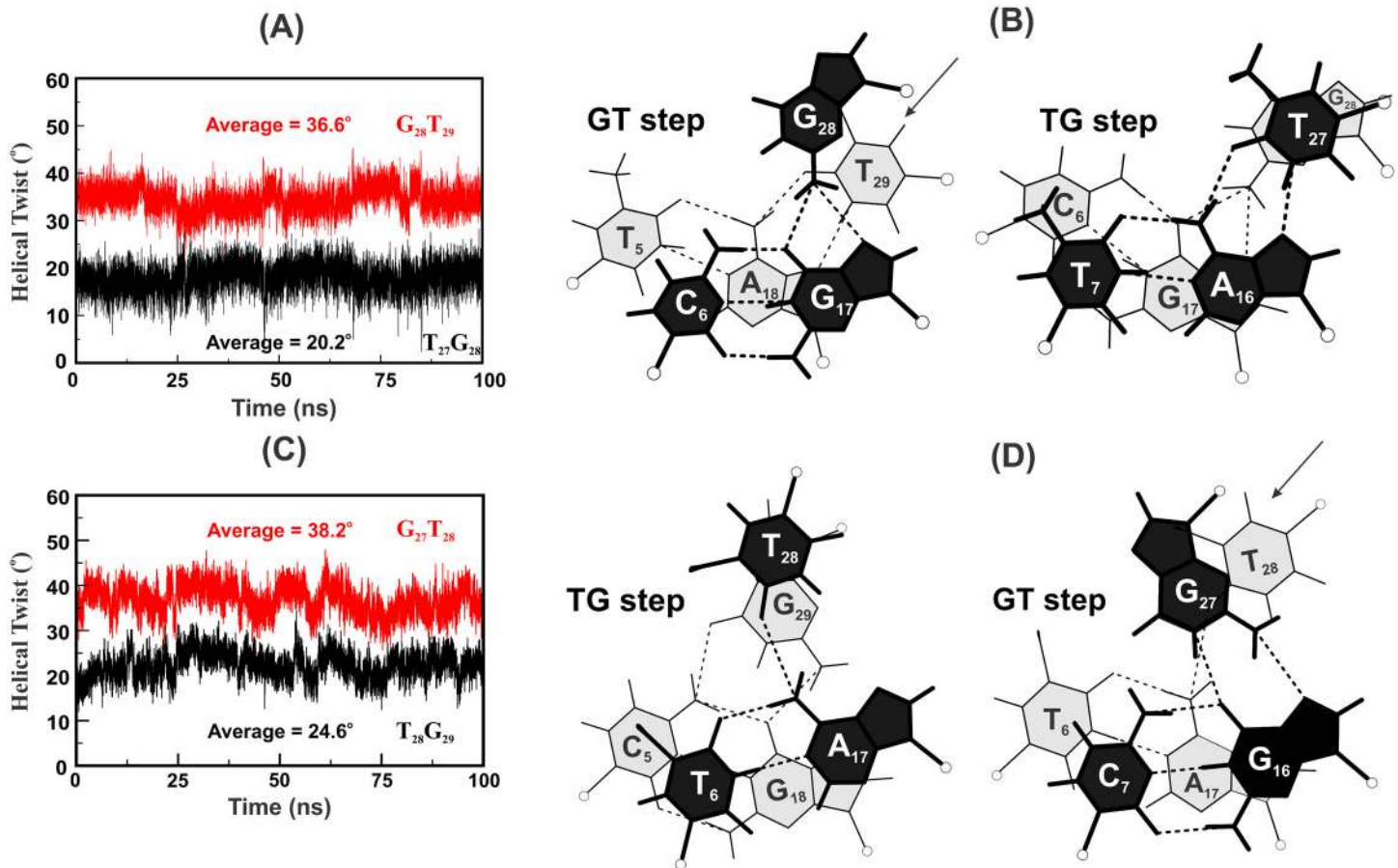


Fig 5. High and low twist at the GT & TG steps concomitant with disrupted stacking at the interruption sites. Helical twist angle variation at the overlapping GT and TG steps at the G*GC interruption site (A) (Sequence 4) in the T*AT triplex and in the G*GC triplex (Sequence 8) with the T*AT interruption (C). Note the loss of significant base stacking at the GT step of the interruption site (B) and (D) and partial stacking at the TC & CT steps of WC duplex.

doi:10.1371/journal.pone.0152102.g005

Influence of a solitary pair of NIBTs on triplex stability

In view of the above findings it is extremely important to ascertain the influence of a single pair vis-à-vis 2 overlapping pairs of NIBTs in retaining or disruption of the canonical Hoogsteen hydrogen bonds and their impact on triplex stability. A single pair of (non-overlapping) NIBTs is harbored uniquely at a triplex junction interface. This can be differentiated by an over winding GT (+ Δt°) or an under winding TG (- Δt°) step along the Hoogsteen strand. Results of MD simulations of 5 DNA triplexes differing in the number of triplex junctions and GT/TG step junction interfaces (Sequences 10–14; Table 1), mimicking a few experimentally studied instances [72, 76, 80, 81] are discussed below.

Structure of the DNA triplex junction with a GT step (+ Δt°) interface. Sequence 10 in Table 1 represents a mini 8-mer G*GC triplex followed by a mini 7-mer T*AT triplex constituting a triplex junction harboring a lone pair of NIBTs mediated by the over winding G₃₈T₃₉ step (+ Δt°). MD simulation of this structure surprisingly shows the retention of canonical Hoogsteen hydrogen bonds (Fig 6A) in both the NIBTs. Effect of overwinding at the GT step (+ Δt°) is reflected by way of large X-displacement of WC pairs in the junction neighborhood (S7A Fig) viz., C₈...G₂₃ exhibiting larger value (-9.23 Å) followed by C₉...G₂₂ (-4.7 Å) and C₁₀...G₂₁ (-3.7 Å). X-displacement is around -3.4 Å and -3.1 Å for the other C...G pairs (in

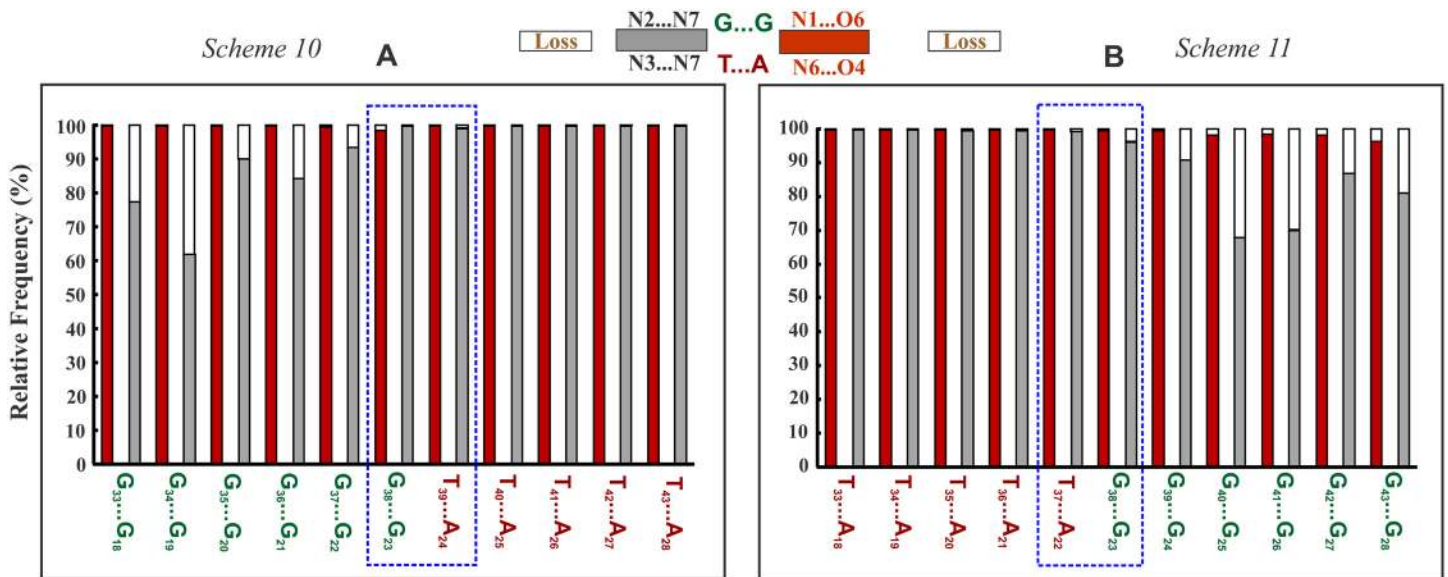


Fig 6. Demonstration of the retention of Canonical Hoogsteen hydrogen bonds at the G*GC/T*AT triplex junction interface. Frequency of incidence (red & gray colour part) and loss (void part) of Hoogsteen hydrogen bonds in the NIBTs at the GT step—Sequence 10 (A) and at the TG step—Sequence 11 (B) triplex junction interfaces. Hydrogen bonds retention is highlighted by the enclosed blue box.

doi:10.1371/journal.pone.0152102.g006

G*GC triplex) and T...A pairs (in T*AT triplex) respectively. As expected, higher twist angle of $\sim 37.5^\circ$ at the G₃₈T₃₉ step of the junction interface concomitant with a low twist of $\sim 20^\circ$ prevails at the T₇C₈ step of WC duplex (S8A Fig). These cause a conspicuous bend in the triplex (Fig 7A). Twist angles at the TT and CC steps of the WC duplex fluctuate around 29.1° , while propeller twists of C...G and T...A pairs remain around -1.3° and -6.33° respectively. Stacking interaction is affected in the neighborhood of junction interface with minimal stacking between the G₃₇ & G₃₈ and T₃₉ & T₄₀ of the Hoogsteen strand, while good stacking prevails between the adjacent triplets in the rest of triplex (S8C Fig). Both major and minor groove expand by ~ 2.5 Å and ~ 1 Å to settle around ~ 23.71 Å and 13.8 Å respectively. WH groove of the mini T*AT triplex widens from 7.3 Å to 9.2 Å with negligible change in mini G*GC triplex (~ 9.0 Å). CH groove of the T*AT mini triplex widens to ~ 18 Å, while that of G*GC mini triplex narrows to ~ 12.0 Å starting from ~ 14 Å respectively (S3C Fig). Backbone torsion angles lie in the preferred range (S9A Fig). The binding free energy of this triplex is ~ -71.6 kcal/mol.

Structure of the triplex junction harboring the TG step (-Δt°) interface. Sequence 11 in Table 1 corresponds to a triplex junction with a single pair of NIBTs intervened by the T₃₇G₃₈ step. Here again the results of MD simulation reveal preservation of canonical Hoogsteen hydrogen bonds in the nonisosteric triplet pairs (Fig 6B). The under winding effect caused by the T₃₇G₃₈ step is reflected in the low X-displacement for the WC pairs, C₈...G₂₃ (~ -1.31 Å) and T₉...A₂₂ (~ -0.65 Å) respectively, proximal to the triplex junction. This is in contrast to large X-displacement seen at the GT junction interface (see above). X-displacement for the other C...G and T...A pairs correspond to ~ -3.23 Å and -2.45 Å respectively. Although a small warp is observed at the triplex junction (Fig 7B) it is not as prominent as in GT step junction interface (Sequence 10). Interestingly BII phosphodiester conformation is found at the under wound T₃₇G₃₈ (S9B Fig) while the rest favour BI. As expected, a low twist angle of $\sim 15^\circ$ is found at the Hoogsteen T₃₇G₃₈ step concomitant with a higher twist at the abutting C₈T₉ ($\sim 38.3^\circ$) and C₇C₈ ($\sim 35.1^\circ$) steps of WC duplex (S8B Fig). Twist angles at the other CC & TT steps of the WC duplex remain at $\sim 30^\circ$. As in the previous case, propeller twist for the C...G

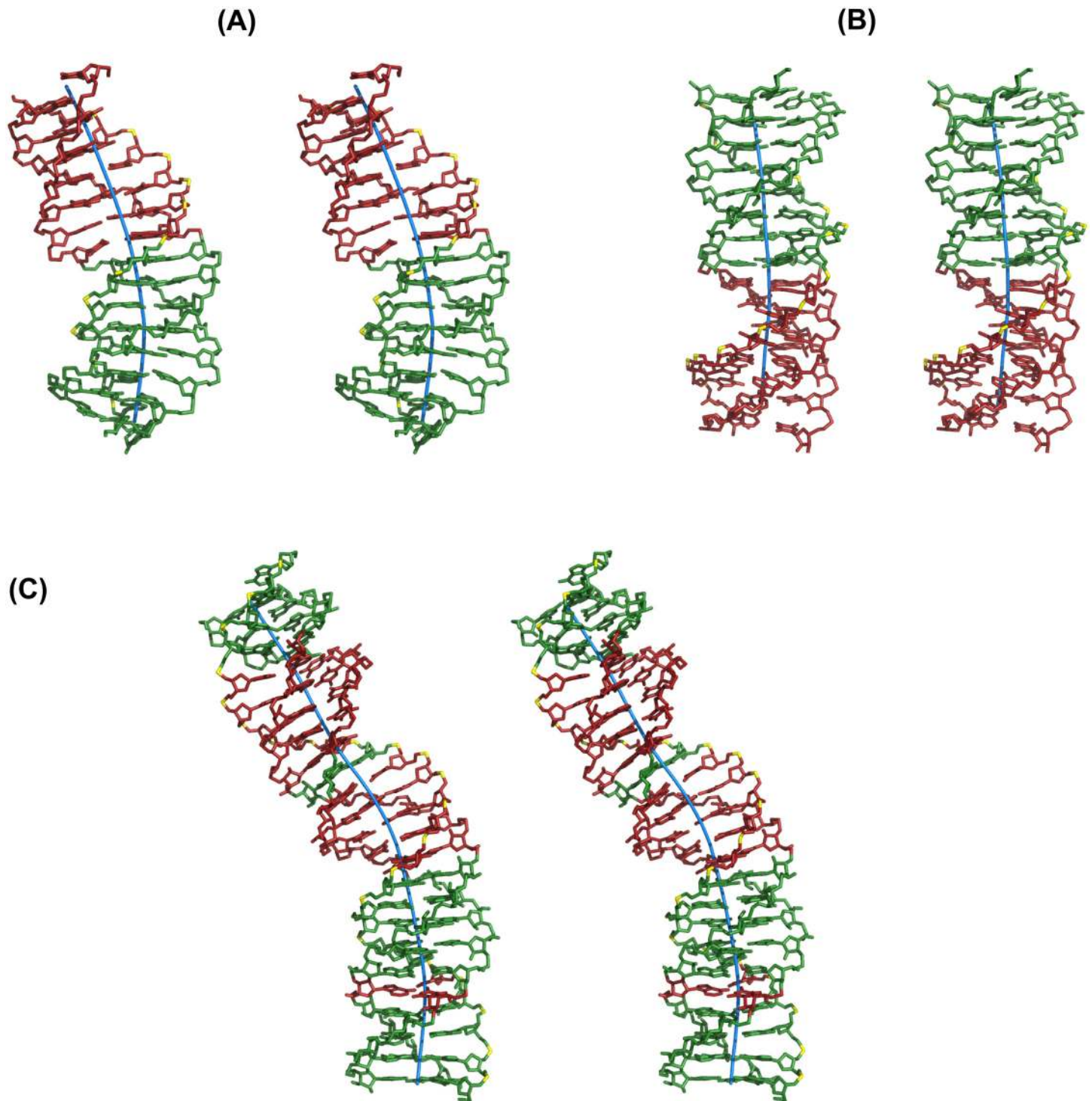


Fig 7. TFOs leading to an overwinding GT step triplex junction interface act as bending agent. Stereo view of parallel triplex with (A) a GT step junction interface (Sequence 10), (B) a TG step junction interface (Sequence 11) and (C) a triplex containing a GT and a TG step junction interface as well as interruptions (1 T*AT & 1 G*GC) (Sequence 15). Note a conspicuous bend in the triplex structure with GT step junction interface in the sequences 10 & 15 (A & C), while near uniform triplex structure prevails in the sequence 11 with a TG step junction interface (B). G*GC and T*AT triplets are coloured green and red respectively. Helical axis (blue stick) w.r.t WC duplex is shown. Phosphate atom of the third strand (TFO) is coloured yellow for reference.

doi:10.1371/journal.pone.0152102.g007

pair (-4.02°) is lower than T...A pair (-13.7°). Average minor groove width is ~13.12 Å in the mini G*GC triplex and it is ~11.7 Å in the mini T*AT triplex. Average major groove width in both the triplexes is ~21.1 Å. Likewise, WH and CH groove width lie in the range of ~9 Å and 13 Å respectively (S3D Fig). Stacking interaction is minimal at the neighborhood of triplex junction (S8D Fig) involving Hoogsteen bases, G₃₈ & G₃₉ and WC duplex pyrimidines, C₇ and C₈. However partial stacking prevails between the purines, A₂₂ & G₂₃ of WC duplex. Binding free energy of this triplex corresponds to ~ -78.4 kcal/mol. These data clearly bring about the distinctions in the influence of overlapping and non-overlapping NIBTs.

Incidence of multiple junctions contributes to instability factor. MD simulations of 15-mer triplexes with the incidence of two (Sequences 12–13; Table 1) and four (Sequence 14; Table 1) triplex junctions under different sequence contexts are performed to discern the limit on the number of occurrences of non-overlapping pair of NIBTs that might sustain, hinder or completely prevent triplex formation. In sequence 12, the 8-mer G*GC triplex is flanked by T*AT triplexes creating a T₃₄G₃₅ step junction interface on the 5'-side and G₄₂T₄₃ step junction interface on the 3'-side of TFO. On the other hand, in sequence 13, the 7-mer T*AT triplex is flanked by G*GC triplexes creating a G₃₄T₃₅ junction interface on the 5'-side and a T₄₁G₄₂ step interface on the 3'-side of TFO. Both these triplexes possess two sets of non-overlapping pair of NIBTs spaced apart by at least 7 or more triplets. Sequence 14 is designed to contain 4 sets of non-overlapping NIBTs to examine the effect of increased incidence of triplex junctions. Results reveal canonical Hoogsteen hydrogen bonds are retained at the junction NIBTs akin to sequences 10 & 11 with only junction. Features characteristic to the GT or TG step junction interface including conspicuous bend at the overwound Hoogsteen GT (+Δt°) steps (G₄₂T₄₃ in Sequence 12, G₃₄T₃₅ in Sequence 13 and G₃₆T₃₇ in Sequence 14; S10A–S10C Fig) are manifested here as well. The 15-mer triplexes with 2 junctions or 2 sets of non-overlapping pairs of NIBTs have binding free energies of ~ -57.2 kcal/mol (Sequence 12) and ~ -55.8 kcal/mol (Sequence 13). The 15-mer triplex with 4 junctions or 4 sets of non-overlapping pairs of NIBTs shows a further decrease in binding free energy of ~ -44.3 kcal/mol (Sequence 14) reflecting the progressively declining stability of triplexes with increase in number of non-overlapping pairs of NIBTs as well (S11 Fig). This clearly indicates that although not as detrimental as the overlapping NIBTs incidence of large non-overlapping NIBTs in a given sequence could affect triplex stability.

Structure of a triplex with simultaneous incidence of junctions and interruptions

Here, we have investigated the impact of 2 interruptions (4 pairs of overlapping NIBTs) and 2 junctions (2 pairs of non-overlapping NIBTs), which together make up six pairs of NIBTs (~25%) in the 25-mer triplex (Sequence 15; Table 1). Results are on the expected lines with disengagement of canonical Hoogsteen hydrogen bonds to noncanonical Hoogsteen sequences in the T₅₅*A₃₀T₂₁ and G₆₆*G₄₁C₁₀ interruptions as in the sequences 4 to 9, while they are retained in at the triplex junction interfaces G₆₀*G₃₅C₁₆ & T₆₁*A₃₆T₁₅ and T₇₁*A₄₆T₅ & G₇₂*G₄₇C₄ triplets as in the sequences 10–14. Likewise, trend of high (~37°) and low twist (~12.7°) angles at the Hoogsteen G₆₀T₆₁ and T₇₁G₇₂ steps respectively concomitant with the reverse trend at the corresponding steps of WC duplex namely; low (~20.6°) and high (~37.2°) twist angles at the T₁₅C₁₆ and C₄T₅ steps respectively, are noticed. Also, larger X-displacement (-7.6 Å) is seen for the C₁₇...G₃₄ and C₁₈...G₃₃ WC base pairs causing a bend in the triplex (Fig 7C). This triplex with 6 pairs of NIBTs is associated with a binding free energy of ~ -81.2 kcal/mol.

Discussion

Ever since the finding of the ability of nucleic acid duplex to accommodate a TFO along its major groove, a variety of biological roles have been demonstrated for DNA, RNA and DNA. RNA hybrid triplexes. Distinctive disposition of donors and acceptors in guanine and thymine of the GT rich TFOs enables them to interact with the purine rich strands of WC duplexes in both parallel and antiparallel orientations forming G·GC and T·AT triplets, with unique geometries rendering them to be nonisosteric. Knowledge of the nature and magnitude of nonisostericity between triplet pairs is crucial in the design of TFOs as they determine the degree of influence on triplex conformation as well as stability. Nonisostericity can be effectively and elegantly defined by residual twist and radial difference [82] which can relate directly to structural traits. More assuring is that they have proven to be valuable in assessing the impact of nonisostericity on DNA triplexes [82, 83].

Surprisingly G·GC & T·AT parallel triplets exhibit strikingly large nonisostericity with the highest value of $\sim 21.6^\circ$ for the residual twist (Δt°) making it a potential culprit for the observed selective preference of parallel triplexes [73, 74, 75, 77] as this is expected to strongly perturb triplex. Qualitative arguments based on nonisostericity are unable to offer clarity to explicate the above, leave alone a mechanistic rationale. MD simulations of 14 parallel and 5 antiparallel DNA triplexes (Table 1), comprising these triplets in an assortment of sequence contexts, are carried out to critically evaluate the decisive influence of single, double and multiple pairs of overlapping and non-overlapping nonisosteric base triplets (NIBTs) on triplex stability. Results have provided new insights for a comprehensive understanding of nonisostericity effects which should aid in the efficient design of TFO to target nucleic acids duplexes to form triplexes.

Relative influence of residual twist vis-à-vis radial difference

One of the key revelations from the present study is the disruption of canonical Hoogsteen hydrogen bonds in both the G·GC and T·AT parallel triplets when they occur alternately (Sequence 1; Fig 2A) and as interruptions (Sequences 4–6, 8, 9; Fig 4A and 4B & S2 and S4 Figs) in a parallel triplex in sharp contrast to antiparallel triplex where reverse Hoogsteen hydrogen bonds remain intact (Fig 2B and S5 Fig). This is directly correlatable to the vastly contrasting nonisostericity between the triplet pairs in parallel and antiparallel orientations, as reflected in the disparate values of Δt° and Δr Å. While Δt° is twice (21.6°) the value in parallel compared to 10.6° in antiparallel triplexes, Δr Å is slightly larger in antiparallel (1.1 Å) than in the parallel (0.4 Å) triplexes. Though both Δt° and Δr Å independently impact in creating backbone disconnect at the successive phosphodiester links, influence of Δr Å seems to be more in the antiparallel triplex although the effects are absorbed by the intrinsic triplex flexibility. On the other hand, mechanistic influence of the large residual twist of 21.6° in parallel triplexes is reflected in lowering the stability sharply through disruption of Hoogsteen hydrogen bonds and associated effects through significant conformational modifications (Figs 2A, 3, 4 & 7). In contrast, smaller Δt° ($\sim 10.6^\circ$) merely causes a zig-zag sugar-phosphate-sugar backbone conformation concomitant with minor twist angle variations in the antiparallel triplex and these by no means affect its formation [83]. Therefore, it can be reckoned that the role of Δt° is more dominant than Δr Å in influencing the formation as well as stability of triplexes.

Overlapping NIBTs disrupt canonical Hoogsteen hydrogen bond and stacking

Isosteric base triplets (with inherent $\Delta t^\circ = 0$ and Δr Å = 0) endow a 'uniform' DNA triplex. However, presence of residual twist (Δt°), as large as 21.6° between the parallel G·GC & T·AT

triplets undermines the canonical Hoogsteen hydrogen bonds in the DNA triplex causing them to assume noncanonical Hoogsteen schemes (NC1 to NC4 for T...A and NC5 to NC8 for G...G; [Fig 3](#)). By doing so, nonisosteric effects are offset via reduction in Δt° ([Fig 3D–3I](#)) concomitant with loss of adjacent base stacking ([Fig 8](#)). Thus, loss of canonical Hoogsteen G...G and T...A hydrogen bonds together with interrupted stacking decrease triplex stability. Absence of these in the antiparallel DNA triplex [[82](#)] results in highly favorable binding free energy of -70.8 kcal/mol compared to -9.4 kcal/mol for the parallel triplex. It is evident then that parallel triplexes with frequent juxtaposition of G*GC and T*AT triplets would be less favored.

Comparative assessment of overlapping vis-à-vis non-overlapping NIBTs on structure, energetics and stability of triplexes

As mentioned earlier, a single base triplet interruption in an otherwise homopolymeric triplex (typified by isosteric base triplets) results in two consecutive and overlapping pairs of NIBTs, with the interrupting NIBT simultaneously bearing the brunt of nonisostericity effects of $+\Delta t$ ($+21.6^\circ$) and $-\Delta t$ (-21.6°) (Sequences 4,8; [Table 1](#)). This causes the canonical Hoogsteen hydrogen bond to switch to noncanonical scheme ([Fig 3D and 3H](#)), resulting in disrupted base stacking ([Fig 5B and 5D](#)). It is obvious that loss of even one of the canonical hydrogen bonds in the triplet interruption enables the Hoogsteen base to sway in a direction to elude the deleterious effect of large Δt° . The base triplet interruption under these circumstances is reckoned as a triplet mismatch and a limited number of them, spaced apart, could be tolerated in a sufficiently long sequence (Sequence 15). Systematic investigations made here suggests that an interrupting G*GC or T*AT triplet decreases the binding energy by nearly 10–11 kcal/mol with reference to the homopolymeric triplex (Sequence 3 vs. Sequence 4 and Sequence 7 vs. Sequence 8). This reduction nearly doubles to ~ 23 kcal/mol with 2 interruptions (Sequence 3 vs. Sequence 5 and Sequence 7 vs. Sequence 9, [Table 1](#)). Increase in the incidence of interruptions from 1 to 3 (Sequence 4 vs. Sequence 6) further reduces the binding energy by ~ 37 kcal/mol indicating a linear correlation between the base triplet interruptions and the binding energy penalty. Hence, it is obvious that multiple incidence of interruptions ie. overlapping pair of NIBTs leads to less stable nature of parallel triplex leading even to inhibition of triplex formation. On the other hand, in the antiparallel triplex (Sequence 16 vs. Sequence 17), the estimated loss in the binding energy (~3 kcal/mol) caused by an interruption due to retention of canonical reverse Hoogsteen hydrogen bond ([S5 Fig](#)) and marginal loss in stacking is not severe or detrimental to antiparallel triplex formation.

On the other hand incidence of non-overlapping pairs of NIBTs harboring at triplex junctions (Sequences 10–14) experiences nonisostericity effects of either $+\Delta t^\circ$ or $-\Delta t^\circ$ alone. The over winding influence of $+\Delta t^\circ$ at the GT step causes the triplex to bend ([Fig 7A and 7C](#) and [S10 Fig](#)). Unwinding influence of $-\Delta t^\circ$ at the TG step merely brings about change from BI to BII phosphodiester conformation ([S9 Fig](#); Sequences 11–14). Comparison of the binding energies of the five 15-mer triplexes (Sequences 10 to 14) designed out of 8 G*GCs and 7 T*ATs suggest a reduction in the binding energy with the increase from 2 (Sequences 12 and 13) to 4 (Sequence 14) in the number of non-overlapping NIBTs (triplex junctions). Sequences 12 and 13 have 1 TG and 1 GT junction while in Sequence 14 there are 2 TG and 2 GT junctions. This causes a reduction in binding energy of ~ 11.5 to 13 kcal/mol (Sequence 13 & Sequence 14 and Sequence 12 & Sequence 14) when non-overlapping NIBTs are increased from 2 to 4. Using this, a crude estimate of the binding energy penalty for a non-overlapping pair of NIBTs turns out to be ~ 5.8 to 6.5 kcal/mol. This is nearly one half compared to the binding energy penalty (~ 11 to 14 kcal/mol) for an interruption entailing 2 overlapping pairs of NIBTs.

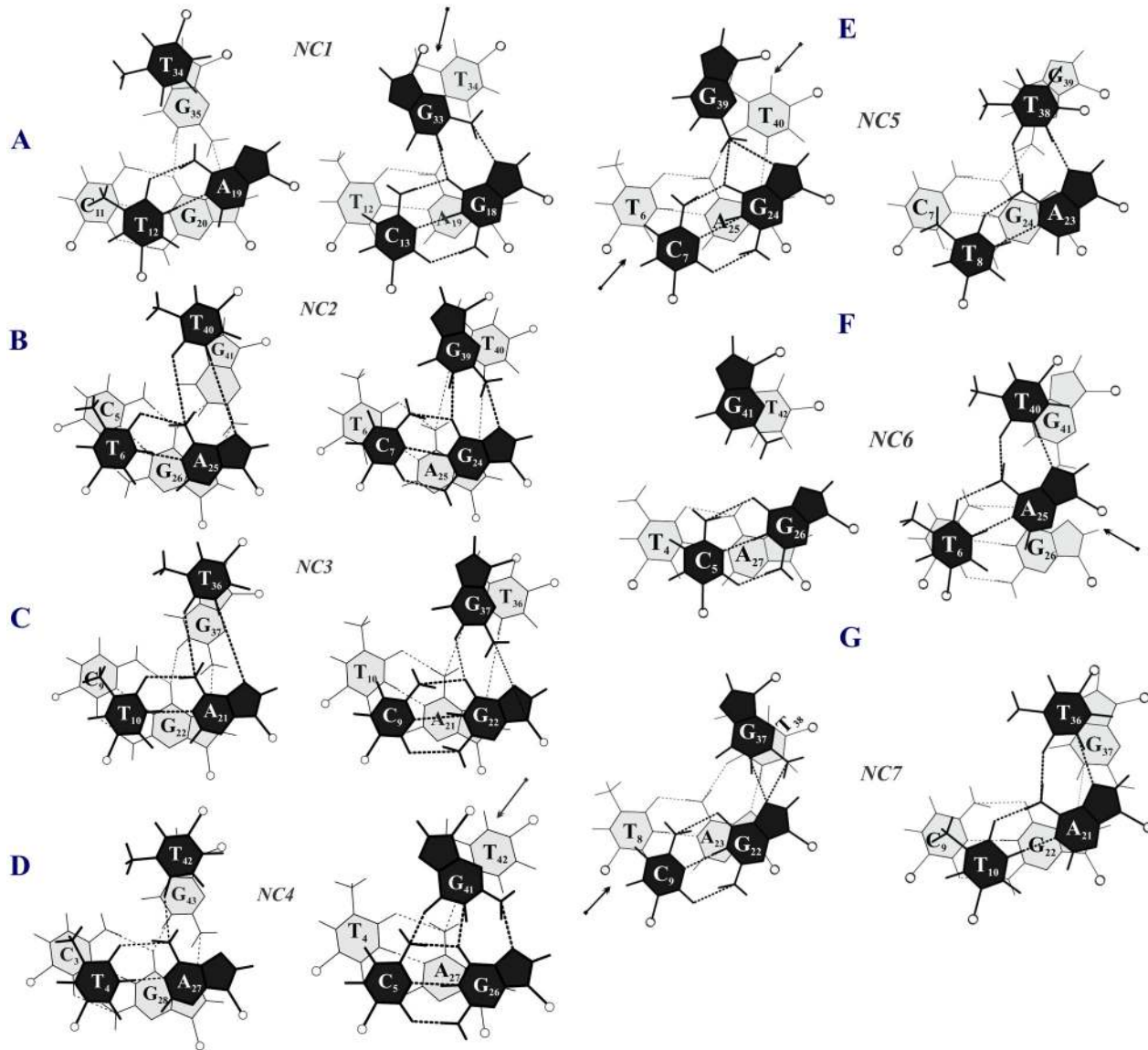


Fig 8. Incidence of alternating G-GC and T-AT triplets in a parallel triplex disrupts base stacking. Stacking interaction at the GT and TG steps of the Hoogsteen strand in the various noncanonical Hoogsteen schemes seen for the *sequence 1* during simulation: NC1 to NC4 for T*AT and NC5 to NC7 for G*GC triplets. Loss of stacking is indicated by dark arrows. C1' atom of the sugar is shown as open circle.

doi:10.1371/journal.pone.0152102.g008

Antiparallel triplexes with GT or TG junction interfaces (non-overlapping pair of NIBTs) (Sequence 18 & Sequence 19) retain canonical reverse Hoogsteen hydrogen bonds (data not shown) at the junction interface like in parallel triplexes. Their binding energies of -76.4 kcal/mol (Sequence 18) and -82.3 kcal/mol (Sequence 19) are slightly more favorable by ~ 5 kcal/mol than their parallel counterparts (-71.6 kcal/mol and -76.4 kcal/mol) due to better stacking at the interface.

Bent triplex as a possible recognition feature for triplex binding proteins

It is significant that characteristic bending of ~ 25° (Fig 7A and 7C and S10 Fig) is seen in the GT step triplex junction interfaces (Sequences 10, 12–14). Also, it has been shown earlier [83]

that the GA step triplex junction interface also exhibits bending of $\sim 17^\circ$. Besides, it is known that a tethered GT rich TFO induces a bend in the triplex formed [84, 85] which regulates the transcription of Luciferase gene [86]. Therefore it might be speculated that bent triplexes could serve as recognition sites for triplex binding proteins. Interestingly, an intramolecular triplex formed by GT rich TFO is demonstrated to bind a few hnRNP family proteins [87]. Likewise an intermolecular triplex formed by a CT rich TFO is shown to bind the GAGA transcription factor from *Drosophila melanogaster* [88].

Parallel between the number and variety of NIBTs and experimental data

Present results convincingly provide a mechanistic argument for the experimental observations concerning the ability or lack of it for formation of parallel DNA triplex with G*GC & T*AT triplets and preference of antiparallel over triplex. For instance the 22-mer TFO, **GGGT***GTGGGT***GTGGGGTGG** targeted to form parallel triplex with the human *Ki-ras* promoter [77], entails 5 interruptions (marked bold) resulting in 8 sets of overlapping pairs of NIBTs, of which 3 occur consecutively twice in the sequence. Apart from this, 2 sets of non-overlapping pairs of NIBTs (mini triplex junctions marked in *italics* and underlined) occur in the sequence. As a consequence, it is expected that a parallel triplex would be less stable owing to the disruption of canonical Hoogsteen hydrogen bonds (5 out of 22) in the 5 triplet interruptions (3 T*AT & 2 G*GC). Indeed, this TFO forms only antiparallel triplex. Similarly, the 31-mer TFO, 5' GTTTTTGG**GT****TGT**GGGTGTGTGGTTTG, designed to inhibit transcription of HIV-1 in infected human cells [75] could bind target duplex only in anti-parallel and not in parallel orientation due to presence of 12 sets of overlapping pairs of NIBTs due to 8 interruptions, spaced in proximity, effecting disruption of Hoogsteen hydrogen bonding in 8 out of 31 triplets. On the other hand, a G*GC triplet interruption in 5' TTT**TGT**TTT [62] and 5' TTC TTCT**TGT**TTCT [89] does not inhibit parallel triplex formation due to loss of only one Hoogsteen hydrogen bond out of 9. Likewise the 14-mer TFO, 5' **GTGGTGGTGTGTG** can form only antiparallel triplex [73] since in parallel triplex it will result in 5 interruptions (10 overlapping pairs of NIBTs) causing loss of Hoogsteen pairs in 5 out of 14 triplets. Similarly the 36-mer TFO, 5' **GGTGGTGG**GT**GGTGGTGGTGGTGGTGGTGGT** 3' (10 interruptions with 19 overlapping pairs of NIBTs), and the 38-mer TFO, 5' GGGGG**GTGGGGTGT** TTGG**GTGGTGGTGGGGTGG** 3' (8 interruptions with 15 overlapping pairs of NIBTs) targeted against the promoter region of human epidermal growth factor receptor gene and the transcription start site of mouse insulin receptor gene respectively, form only antiparallel triplexes [78] to effect transcription inhibition. Their inability to form parallel triplex is obvious in view of expected loss of Hoogsteen pairs at all the interruption sites (9 out of 36 in 36-mer TFO and 8 out of 38 in 38-mer TFO) making the triplex less viable due to lower stability.

On the other hand, non-overlapping NIBTs in 5' TTTTC_mTTTTGGGGGG (one TG step junction) enables it to form a stable triplex with the oligopurine sequence flanking the SV40 origin of replication containing plasmid in COS-1 cells to stall replication [76]. Likewise, a stable parallel triplex [81] is formed with the TFO 5' GGGGGTTTCTTTT 3', in spite of GT step triplex junction (a pair of non-overlapping NIBTs) which facilitates retention of Hoogsteen pair. This finding that triplex junctions (non-overlapping NIBTs) do not significantly influence triplex formation or its stability gains further support from the observation that the 5' TTTTGGTTTTGG 3' forms a triplex in both parallel and antiparallel orientation with a T_m of $\sim 46^\circ\text{C}$ [80]. Likewise, parallel and antiparallel triplex formed by 5' GGGGTTTTGGGG 3' with 2 triplex junctions (2 sets of non-overlapping pairs of NIBTs) display a similar T_m of 57°C & 58°C [72]. But, T_m is lowered by $\sim 10^\circ\text{C}$ for the parallel triplex formed by the TFO 5'

GGTT**TGTTT**TGTTT 3' compared to its antiparallel counterpart triplex formed by the TFO 5' TTT**TGTTT**TGTTTGG 3'. This can be readily attributed to the loss of 2 (out of 14) canonical Hoogsteen hydrogen bonds at the two G-GC interruptions (4 overlapping NIBTs) in the parallel triplex [80]. In the antiparallel triplex, reverse Hoogsteen hydrogen bonds are preserved since minor effects of low residual twist of $\Delta t = 10.6^\circ$ are absorbed by the inherent triplex flexibility. Thus, there exists a direct bearing between the experimental observations and the number and type of NIBTs with MD simulations clearly bringing out the criticality of the mechanistic attributes of base triplet nonisostericity.

Growing evidence for occurrence of nucleic acid triple helices *in vivo* and their implication in diverse critical biological processes, disease and therapy, makes their structural study imperative. Unlike DNA and RNA duplexes where isosteric WC base pairs dominate, nucleic acid triplexes encounter mixture of isosteric and nonisosteric base triplets. Nature and magnitude of base triplet nonisostericity is solely determined by neighbouring triplets. This together with the number of nonisosteric base triplets (NIBTs) determines triplex forming ability, parallel or non-parallel, with or without sequence dependent structural variations as well as their stability. Although base triplet nonisostericity between the parallel G-GC & T-AT triplets is seemingly obvious [74] quantitative estimates of its nature and magnitude and its mechanistic influence are by no means readily comprehensible. The work presented here fills this critical gap by providing quantitative description of base triplet nonisostericity as well as their mechanistic effects in influencing stability and formation of DNA triplexes. It is found that the extreme base triplet nonisostericity (residual twist) prevalent between the parallel G-GC & T-AT triplets adds an element of instability, owing to the inevitability of significant stereochemical rearrangements, to the extent of even precluding triplex formation when their incidences are recurrent. Thus, the concept of residual twist has provided a means not only to assess the source and degree of nonisostericity but also has enabled in providing a mechanistic basis for the experimental observations in relation to the feasibility or the lack of same to form parallel DNA triplexes. Although these results are derived from analyses of DNA triplexes they should hold valid in assessing the role of nonisostericity in RNA and RNA-DNA hybrid triplexes as well thus enabling a comprehensive understanding of sequence dependent structural variation in nucleic acid triplexes in general. In fact, this concept has found an application in developing an algorithm to identify triplex forming sequences [90]. The results presented here should aid in designing more specific and efficient TFOs as gene targeting agents (antigene therapeutics). Interestingly, the concept of residual twist proved immensely useful in predicting mechanistic effects of base pair nonisostericity due to juxtaposition of Watson-Crick and different non-Watson-Crick pairs in RNA duplexes [91].

Materials and Methods

Triplex model generation

Parallel and anti-parallel DNA triplexes (Sequences 1–19 in Table 1), were built conforming to a 12-fold helix [92] with stereochemistry of the third strand regularized by constrained-restrained molecular geometry optimization and van der Waals energy minimization using X-PLOR [93]. Generated models were then subjected to steepest descent energy minimization using the Sander module of AMBER 12.0 [94] and were used as starting model for MD simulation.

MD simulation

MD simulation is a technique employed to obtain the dynamical aspects of a protein or nucleic acid structure obtained from X-ray, NMR or modelling investigations. AMBER [94] and CHARMM [95] are the widely used force-fields. Modified version of the Cornell et al. force field [96] as in parmbsc0 and its minor variations in the AMBER [97–100] has provided not

only a decisive stabilization of nucleic acid simulations, but also have been able to reproduce with reasonable accuracy several structural and dynamical features of BDNA and non BDNA structures which includes detailed insights into BDNA \leftrightarrow ADNA [101], B \rightarrow ZDNA transition [102,103], base flipping [104], DNA bending [105] and deformation of DNA/RNA in the presence of proteins [106] etc. AMBER has also been successfully utilised to study properties like stability, folding pathway, role of ions, interaction with drugs and hydration property of various non-BDNA structures like RNA [91], DNA.RNA hybrids [97,107], t-RNA [108,109], i-motif [110], Holliday junction [111], PNA.DNA [112], G-quadruplexes [113,114] etc.

Likewise, MD simulations [115] using AMBER force field bring out the feature intermediate to that of A and B DNA of the duplex in a triplex in accordance with geometrical stipulation [116] and NMR data [117,118]. Also, it has been competently used to explore conformational flexibility [119], stability [120], hydration [121], protonation and folding property [122] etc. of DNA triplexes. In fact, our own earlier investigations on antiparallel triplexes brought out base triplet nonisosteric effects [82, 83]. Given its established utility, AMBER force field is used in the present study of parallel triple helices to enable comparison with antiparallel triplexes.

All MD simulations were performed using AMBER 12.0 program [94]. Triplexes were solvated in a periodic box of TIP3P waters and net-neutralized with Na⁺ counter ions using Leap module of AMBER Tools 13.0 [123]. Following this, initial equilibration minimization was carried out using steepest descent (2000 cycles) and conjugate-gradient (500 cycles) algorithms, with a positional restraint of 500 kcal/mol \AA^2 on the solute atoms. Further minimization was carried out with positional restraints on the solute reduced in steps of 100 kcal/mol \AA^2 . Subsequently, minimization was effected for 1000 cycles without any positional restraints. MD simulation was then carried out in an NVT ensemble, during which the entire system was heated from 0 to 300K over 100 ps. Further equilibration MD was carried out for 200 ps under constant pressure (NPT). Following this, production run MD simulations were initiated for 250 ns for Sequence 1 and 100 ns each for triplex sequences in Sequences 2–19 (Table 1). Equilibration and production run simulations were carried out using the Sander and PMEMD module (optimized for CUDA) of AMBER 12.0 (ff99+parmbsc0) respectively. Periodic boundary conditions and PME under isothermal isobaric conditions (T = 298 K; P = 1 atm) with an integration time step of 2 fs was employed in the simulation. All bonds involving hydrogen were constrained using SHAKE algorithm (tolerance = 0.0005 \AA) [124]. Trajectories were analyzed using Ptraj module of Amber Tools 13.0. Base pair, base step parameters and backbone torsion angles were extracted from the output of 3DNA Ver 2.1 [125] using in-house programs. Helical twist angles were calculated with respect to C1' . . . C1' vector due to the presence of nonisosteric base pairs [82, 83]. Hydrogen bond distances and angles were computed using the criteria of donor . . . acceptor distance of < 3.6 \AA and hydrogen bond angle >120°. Figures were prepared either using PyMOL (ww.pymol.org) or graphical interface of insight II [126].

Description of residual twist (Δt°) and radial difference ($\Delta r \text{\AA}$)

Residual twist (Δt°) is calculated by measuring the angle between the line joining C1' . . . C1' atoms of the Hoogsteen/Reverse Hoogsteen pair of the superimposed base triples prior to the application of helical twist angle (t) requisite to generate a triple helical structure. Radial difference ($\Delta r \text{\AA}$) corresponds to half the difference between diameters of the G-GC (14.5 \AA) & T-AT (15.3 \AA) base triples.

Binding free energy calculations

Binding free energy (ΔG_{Bind}) of the triplex structures is estimated by using MM-PBSA [127, 128] module of AMBER 12.0. It is calculated as the binding free energy difference between the

bound (triplex) and the unbound (WC duplex and TFO) states.

$$\Delta G_{(\text{bind})} = G_{(\text{Triplex})} - (G_{(\text{WC duplex})} + G_{(\text{TFO})})$$

G_{Triplex} , $G_{\text{WC duplex}}$ & G_{TFO} are individually calculated by post-processing of the simulation trajectories taken at 20 ps interval after removing counter ions and water molecules. Free energy (G) for each component is given by; $G = \Delta E_{(\text{gas})} - \Delta G_{(\text{sol})} - T\Delta S_{(\text{gas})}$. The gas-phase energy ΔE^{gas} is calculated as summation over the bond length, bond angle, dihedral angle, van der Waals and electrostatics energy contributions. ΔG^{sol} is calculated by summing the polar (G_{PB}) and non-polar ($G_{\text{nonpol, sol}}$) contributions. G_{PB} is calculated by the Poisson-Boltzmann (PB) equation with a dielectric constant value of 1.0 and 80.0 set for solute and solvent respectively. Non-polar contribution $G_{\text{nonpol, sol}} = \gamma^* \text{SASA}$, wherein SASA stands for solvent accessible surface area. Value of γ corresponds to 0.0072 kcal/mol. Using normal mode analysis, entropy term ($T\Delta S$) is calculated as the sum of translational, rotational and vibrational components. For the normal mode calculation distance-dependent dielectric constant energy minimization is used with $\epsilon = 4r$ and convergence value of 0.5. Terminal triplet at the 5' and 3' ends of the triplexes is not considered due to end-effects.

Supporting Information

S1 Fig. Pronounced intrinsic nonisostericity between G*GC and T*AT triplets leads to non-uniform triplex. (A) Illustration of how the residual twist between the nonisosteric G*GC (green) and T*AT (brown) triplets lead to skewed displacement of the C1' atoms of the Hoogsteen bases implying the need for large conformational changes in triplex. Note that the C1' atoms of WC duplex are aligned vertically. Base triplets are related to each other by $t = 0^\circ$ and $h = 3.26 \text{ \AA}$. (B) Stereo view of a G*GC triplet sandwiched between T*AT triplets (S1A Fig) demonstrates that the skewed arrangement of C1' atoms disconnects it to the extent of nearly one nucleotide length at both the GT & TG steps, besides causing steric overlap at TG step due to close proximity of adjacent sugar residues. WC duplex and TFO are coloured blue and red respectively. Hydrogen bond is denoted by dashed line (green). (TIF)

S2 Fig. Disruption of canonical Hoogsteen hydrogen bond in G interrupts in a T*AT triplex. Average canonical Hoogsteen hydrogen bond distance corresponding to T...A pairs in a (A) poly T*AT triplex (Sequence 3), (B) with a single G interruption (Sequence 4), (C) with 2 G interruptions (Sequence 5) and (D) with 3 G interruptions (Sequence 6). Standard deviation w.r.to mean distance is indicated above the bar. G interruptions are marked by dashed blue rectangle. Note the large fluctuation in flanking T...A triplets with increase in G interruption in C & D (denoted by orange circle). (TIF)

S3 Fig. Variation in groove widths. Changes in minor groove (m), CH groove (CH) and WH groove (WH) widths in different triplexes: T*AT triplex (Sequence 4) with a G*GC interruption (A); G*GC triplex (Sequence 8 with a T*AT interruption (B)); a triplex (Sequence 10) with a GT step junction interface (C); a triplex (Sequence 11) with a TG step junction interface (D). Groove widths corresponding to starting model (thick black line) and average structure (dashed line) calculated for the last 5 ns are shown. (TIF)

S4 Fig. Canonical Hoogsteen hydrogen bond variation in G*GC triplex with T interrupts. Average canonical Hoogsteen hydrogen bond distance of G...G pairs constituting a (A) poly G*GC triplex (Sequence 7), (B) with a single T...A interruption (Sequence 8) and (C) with

2 T . . A interruptions (Sequence 9). Standard deviation w.r.to mean distance is indicated above the bar. Intervening T . . A pair are marked by dashed blue rectangle. Note the less fluctuation in flanking T . . A triplets with increase in G interruption in C & D (denoted by orange circle) as compared to [S3 Fig.](#)
(TIF)

S5 Fig. Stable nature of antiparallel G-GC triplex with T-AT interruption. Frequency of incidence (red & gray colour filled part) and loss (void part) of reverse Hoogsteen hydrogen bonds in the T-AT interruption of the G-GC triplex (Sequence 17). Conservation (filled part) of canonical hydrogen bonds O2 . . N6 in and N1 . . N7 the interrupting T-AT triplet is conspicuous (blue box).
(TIF)

S6 Fig. Progressive destabilisation of triplex with increasing number of overlapping NIBTs. Variation of binding free energy w.r.t triplexes comprising different number of G-GC (green) and T-AT (red) interruptions (overlapping pair of NIBTs). Sequences 3 & 7 (zero interruption); Sequences 4 & 8 (1 interruption); Sequences 5 & 9 (2 interruptions); Sequence 6 (3 interruptions). Proportional decrease in energy with increase in overlapping NIBTs denotes a linear correlation (indicated by least square line fit).
(TIF)

S7 Fig. Large variation of X-displacement at the non-overlapping GT step junction. Illustration of large X-displacement (dashed line) of base pairs of WC duplex near GT junction interfaces in different triplexes viz., (A) Sequence 10; (B) Sequence 12; (C) Sequence 13; (D) Sequence 14. X-displacement corresponding to the starting model is depicted as thick black line.
(TIF)

S8 Fig. Helical twist and stacking in and around the GT & TG junction. Twist angle variation at the WC T₇C₈ (black) and WH G₃₇T₃₈ (red) steps in the junction triplex with GT interface—Sequence 10 (A); at the WC C₈T₉ (black) T₇C₈ and WH T₃₇G₃₈ (red) steps in the junction triplex with TG interface- Sequence 11 (B). Nature of base stacking in and around the neighbourhood of junction interface in Sequence 10 (C); and in Sequence 11 (D). Minimal stacking is indicated by arrows. C1' atom of the sugar is shown as open circle.
(TIF)

S9 Fig. BI to BII transition at the non-overlapping TG step junction. Variation of backbone torsion angles around the C3'-O3' (ϵ ; black) and P-O3' bonds (ζ ; red) in different triplexes (Sequences 10–14). Note the switch from BI to BII conformation at the TG step in Sequences 11–14 (indicated by arrow). The GT step assumes the preferred BI conformation.
(TIF)

S10 Fig. Kinked triplexes. Snapshot of parallel triplexes showing a curvature near the neighbourhood of GT step: Sequence 12 (A); Sequence 13 (B); and Sequence 14 (C). G-GC and T-AT mini triplexes are coloured green and red respectively. Bent curvature is indicated by a black circle. 5'-terminus of TFO is indicated. Helical axis of WC duplex of the triplex is shown (blue sphere).
(TIF)

Acknowledgments

We thank CDAC, Pune, India for computational resources and R.Srivatsan (IBAB) for his help in creating one of the plots.

Author Contributions

Conceived and designed the experiments: NY GG TR. Performed the experiments: GG TR. Analyzed the data: GG NY. Contributed reagents/materials/analysis tools: NY. Wrote the paper: GG TR NY.

References

1. Hégarat N, Novopashina D, Fokina AA, Boutorine AS, Venyaminova AG, Praseuth D, et al. Monitoring DNA triplex formation using multicolor fluorescence and application to insulin-like growth factor I promoter downregulation. *FEBS J*. 2014; 281: 1417–1431. doi: [10.1111/febs.12714](https://doi.org/10.1111/febs.12714) PMID: [24423253](https://pubmed.ncbi.nlm.nih.gov/24423253/)
2. Jain A, Wang G, Vasquez KM. DNA Triple Helices: biological consequences and therapeutic potential. *Biochimie*. 2008; 90: 1117–1130. doi: [10.1016/j.biochi.2008.02.011](https://doi.org/10.1016/j.biochi.2008.02.011) PMID: [18331847](https://pubmed.ncbi.nlm.nih.gov/18331847/)
3. Chin JY, Glazer PM. Repair of DNA lesions associated with triplex-forming oligonucleotides. *Mol Carcinog*. 2009; 48: 389–399. doi: [10.1002/mc.20501](https://doi.org/10.1002/mc.20501) PMID: [19072762](https://pubmed.ncbi.nlm.nih.gov/19072762/)
4. Rogers FA, Tiwari MK. Triplex-induced DNA damage response. *Yale J Biol Med*. 2013; 86: 471–478. PMID: [24348211](https://pubmed.ncbi.nlm.nih.gov/24348211/)
5. Holland JA, Hoffman DW. Structural features and stability of an RNA triple helix in solution. *Nucleic Acids Res*. 1996; 24: 2841–2848. PMID: [8759020](https://pubmed.ncbi.nlm.nih.gov/8759020/)
6. Mitton Fry RM, De Gregorio SJ, Wang J, Steitz TA, Steitz JA. PolyA tail recognition by a viral RNA element through assembly of a triple helix. *Science*. 2010; 330: 1244–1247. doi: [10.1126/science.1195858](https://doi.org/10.1126/science.1195858) PMID: [21109672](https://pubmed.ncbi.nlm.nih.gov/21109672/)
7. Vuyisich M, Beal PA. Regulation of the RNA-dependent protein kinase by triple helix formation. *Nucleic Acids Res*. 2000; 28: 2369–2374. PMID: [10871369](https://pubmed.ncbi.nlm.nih.gov/10871369/)
8. Hampel KJ, Burkholder GD, Lee JS. Plasmid dimerization mediated by triplex formation between polypyrimidine-polypurine repeats. *Biochemistry*. 1993; 32: 1072–1077. PMID: [8424937](https://pubmed.ncbi.nlm.nih.gov/8424937/)
9. Chin JY, Schleifman EB, Glazer PM. Repair and recombination induced by triple helix DNA. *Front Biosci*. 2007; 12: 4288–4297. PMID: [17485375](https://pubmed.ncbi.nlm.nih.gov/17485375/)
10. Shefer K, Brown Y, Gorkovoy V, Nussbaum T, Ulyanov NB, Tzfati Y. A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol Cell Biol*. 2007; 27: 2130–2143. PMID: [17210648](https://pubmed.ncbi.nlm.nih.gov/17210648/)
11. Chou MY, Chang KY. An intermolecular RNA triplex provides insight into structural determinants for the pseudoknot stimulator of -1 ribosomal frame shifting. *Nucleic Acids Res*. 2010; 38: 1676–1685. doi: [10.1093/nar/gkp1107](https://doi.org/10.1093/nar/gkp1107) PMID: [20007152](https://pubmed.ncbi.nlm.nih.gov/20007152/)
12. Brown JA, Valenstein ML, Yario TA, Tycowski KT, Steitz JA. Formation of triple-helical structures by the 3'-end sequences of MALAT1 and MEN β noncoding RNAs. *Proc Natl Acad Sci U S A*. 2012; 109: 19202–19207. doi: [10.1073/pnas.1217338109](https://doi.org/10.1073/pnas.1217338109) PMID: [23129630](https://pubmed.ncbi.nlm.nih.gov/23129630/)
13. Wilusz JE, JnBaptiste CK, Lu LY, Kuhn CD, Joshua-Tor L, Sharp PA. A triple helix stabilizes the 3' ends of long noncoding RNAs that lack poly (A) tails. *Genes Dev*. 2012; 26: 2392–2407. doi: [10.1101/gad.204438.112](https://doi.org/10.1101/gad.204438.112) PMID: [23073843](https://pubmed.ncbi.nlm.nih.gov/23073843/)
14. Martjanov I, Ramadass A, Serra Barros A, Chow N, Akoulitchev A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*. 2007; 445: 666–670. PMID: [17237763](https://pubmed.ncbi.nlm.nih.gov/17237763/)
15. Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, et al. MEG3 long noncoding RNA regulates the TGF- β pathway genes through formation of RNA-DNA triplex structures. *Nat Commun*. 2015; 6: 7743. doi: [10.1038/ncomms8743](https://doi.org/10.1038/ncomms8743) PMID: [26205790](https://pubmed.ncbi.nlm.nih.gov/26205790/)
16. O'Leary VB, Ovsepian SV, Carrascosa LG, Buske FA, Radulovic V, Niyazi M, et al. PARTICLE, a Triplex-Forming Long ncRNA, Regulates Locus-Specific Methylation in Response to Low-Dose Irradiation. *Cell Rep*. 2015; 11: 474–485. PMID: [25900080](https://pubmed.ncbi.nlm.nih.gov/25900080/)
17. Jain A, Bacolla A, Chakraborty P, Grosse F, Vasquez KM. Human DHX9 helicase unwinds triple-helical DNA structures. *Biochemistry*. 2010; 49: 6992–6999. doi: [10.1021/bi100795m](https://doi.org/10.1021/bi100795m) PMID: [20669935](https://pubmed.ncbi.nlm.nih.gov/20669935/)
18. Guo M, Hundseth K, Ding H, Vidhyasagar V, Inoue A, Nguyen CH, et al. A distinct triplex DNA unwinding activity of ChIR1 helicase. *J Biol Chem*. 2015; 290: 5174–5189. doi: [10.1074/jbc.M114.634923](https://doi.org/10.1074/jbc.M114.634923) PMID: [25561740](https://pubmed.ncbi.nlm.nih.gov/25561740/)
19. Ohno M, Fukagawa T, Lee JS, Ikemura T. Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma*. 2002; 111: 201–213. PMID: [12355210](https://pubmed.ncbi.nlm.nih.gov/12355210/)

20. Agazie YM, Burkholder GD, Lee JS. Triplex DNA in the nucleus: direct binding of triplex-specific antibodies and their effect on transcription, replication and cell growth. *Biochem J.* 1996; 316: 461–466. PMID: [8687388](#)
21. Musso M, Nelson LD, Van Dyke MW. Characterization of purine-motif triplex DNA-binding proteins in HeLa extracts. *Biochemistry* 1998; 37: 3086–3095. PMID: [9485462](#)
22. Rao JE, Craig NL. Selective recognition of pyrimidine motif triplexes by a protein encoded by the bacterial transposon Tn7. *J Mol Biol.* 2001; 307: 1161–1170. PMID: [11292332](#)
23. Van Dyke M, Nelson L. Triple helix-interacting proteins and cancer. *OA Mol Oncol.* 2013; 1: 5. PMID: [24910779](#)
24. Kusic J, Tomic B, Divac A, Kojic S. Human initiation protein Orc4 prefers triple stranded DNA. *Mol Biol Rep.* 2010; 37: 2317–2322. doi: [10.1007/s11033-009-9735-8](#) PMID: [19690980](#)
25. Nelson LD, Bender C, Mannsperger H, Buergy D, Kambakamba P, Mudduru G, et al. Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Mol Cancer.* 2012; 11: 38. doi: [10.1186/1476-4598-11-38](#) PMID: [22682314](#)
26. Ohshima K, Montermini L, Wells RD, Pandolfo M. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication *in vivo*. *J. Biol. Chem.* 1998; 273: 14588–14595. PMID: [9603975](#)
27. Bissler JJ. Triplex DNA and human disease. *Front Biosci.* 2007; 12: 4536–4546. PMID: [17485395](#)
28. Hou XH, Guo XY, Chen Y, He CY, Chen ZY. Increasing the minicircle DNA purity using an enhanced triplex DNA technology to eliminate DNA contaminants. *Mol Ther Methods Clin Dev.* 2015; 1: 14062. doi: [10.1038/mtm.2014.62](#) PMID: [26052527](#)
29. Yang M, Zhang X, Liu H, Kang H, Zhu Z, Yang W, Tan W. Stable DNA Nanomachine Based on Duplex-Triplex Transition for Ratiometric Imaging Instantaneous pH Changes in Living Cells. *Anal Chem.* 2015; 87: 5854–5859. doi: [10.1021/acs.analchem.5b01233](#) PMID: [26016566](#)
30. Xiao Z, Zhu H, Xin A, Li Y, Ling L. Triplex DNA logic gate based upon switching on/off their structure by Ag(+)/cysteine. *Analyst.* 2015; 140: 7322–7326. doi: [10.1039/c5an01371b](#) PMID: [26359516](#)
31. Idili A, Vallée-Bélisle A, Ricci F. Programmable pH-triggered DNA nanoswitches. *J Am Chem Soc.* 2014; 136: 5836–5839. doi: [10.1021/ja500619w](#) PMID: [24716858](#)
32. Zheng J, Nie Y, Hu Y, Li J, Li Y, Jiang Y, Yang R. Time-resolved fluorescent detection of Hg²⁺ in a complex environment by conjugating magnetic nanoparticles with a triple-helix molecular switch. *Chem Commun (Camb).* 2013; 49: 6915–6917.
33. Vollmer S, Richert C. DNA Triplexes That Bind Several Cofactor Molecules. *Chemistry.* 2015 Nov 12;
34. Kröner C, Göckel A, Liu W, Richert C. Binding cofactors with triplex-based DNA motifs. *Chemistry.* 2013; 19: 15879–15887. doi: [10.1002/chem.201303098](#) PMID: [24194407](#)
35. Zheng J, Jiao A, Yang R, Li H, Li J, Shi M, et al. Fabricating a reversible and regenerable Raman-active substrate with a biomolecule-controlled DNA nanomachines. *J Am Chem Soc.* 2012; 134: 19957–19960. doi: [10.1021/ja308875r](#) PMID: [23190376](#)
36. Zhu X, Liu Y, Yang J, Liang Z, Li G. Gold nanoparticle-based colorimetric assay of single-nucleotide polymorphism of triplex DNA. *Biosens. Bioelectron.* 2010; 25: 2135–2139. doi: [10.1016/j.bios.2010.02.017](#) PMID: [20233656](#)
37. Johannsen MW, Gerrard SR, Melvin T, Brown T. Triplex-mediated analysis of cytosine methylation at CpA sites in DNA. *Chem Commun (Camb).* 2014; 50: 551–553.
38. Xi D, Wang X, Ai S, Zhang S. Detection of cancer cells using triplex DNA molecular beacons based on expression of enhanced green fluorescent protein (eGFP). *Chem Commun (Camb).* 2014; 50: 9547–9549.
39. Aghamollaei H, Moghaddam MM, Kooshki H, Heiat M, Mirnejad R, Barzi NS. Detection of *Pseudomonas aeruginosa* by a triplex polymerase chain reaction assay based on *lasI/R* and *gyrB* genes. *J Infect Public Health.* 2015; 8: 314–322. doi: [10.1016/j.jiph.2015.03.003](#) PMID: [25863575](#)
40. Burrell MR, Burton NP, Maxwell A. A high-throughput assay for DNA topoisomerases and other enzymes, based on DNA triplex formation. *Methods Mol Biol.* 2010; 613: 257–66. doi: [10.1007/978-1-60327-418-0_16](#) PMID: [19997889](#)
41. Jakobsen U, Vogel S. Assembly of Liposomes Controlled by Triple Helix Formation. *Bioconjug Chem.* 2013; 24: 1485–1495. doi: [10.1021/bc300690m](#) PMID: [23885785](#)
42. Grossmann TN, Röglin L, Seitz O. Triplex molecular beacons as modular probes for DNA detection. *Angew Chem Int Ed Engl.* 2007; 46: 5223–5225. PMID: [17535003](#)
43. Behe MJ. An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Res.* 1995; 23: 689–695. PMID: [7899090](#)

44. Goñi JR, de la Cruz X, Orozco M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res.* 2004; 32: 354–360. PMID: [14726484](#)
45. Holder IT, Wagner S, Xiong P, Sinn M, Frickey T, Meyer A, Hartig JS. Intrastrand triplex DNA repeats in bacteria: a source of genomic instability. *Nucleic Acids Res.* 2015 Oct 7. pii: gkv1017. [Epub ahead of print].
46. Riechert-Krause F, Autenrieth K, Eick A, Weisz K. Spectroscopic and calorimetric studies on the binding of an indoloquinoline drug to parallel and antiparallel DNA triplexes. *Biochemistry.* 2013; 52: 41–52. doi: [10.1021/bi301381h](#) PMID: [23234257](#)
47. Eick A, Riechert-Krause F, Weisz K. Spectroscopic and calorimetric studies on the triplex formation with oligonucleotide-ligand conjugates. *Bioconjugate Chem.* 2010; 21: 1105–1114.
48. Rusling DA, Nandhakumar IS, Brown T, Fox KR. Triplex-directed recognition of a DNA nanostructure assembled by crossover strand exchange. *ACS Nano.* 2012; 6: 3604–3613. doi: [10.1021/nn300718z](#) PMID: [22443318](#)
49. Xu N, Yang H, Cui M, Wan C, Liu S. High-performance liquid chromatography-electrospray ionization-mass spectrometry ligand fishing assay: a method for screening triplex DNA binders from natural plant extracts. *Anal Chem.* 2012; 84: 2562–2568. doi: [10.1021/ac202796v](#) PMID: [22220694](#)
50. Arya DP. New approaches toward recognition of nucleic acid triple helices. *Acc Chem Res.* 2011; 44: 134–146. doi: [10.1021/ar100113q](#) PMID: [21073199](#)
51. Buske FA, Bauer DC, Mattick JS, Bailey TL. Triplex-Inspector: an analysis tool for triplex-mediated targeting of genomic loci. *Bioinformatics.* 2013; 29: 1895–1897. doi: [10.1093/bioinformatics/btt315](#) PMID: [23740745](#)
52. Jenjaroenpun P, Kuznetsov VA. TTS mapping: integrative WEB tool for analysis of triplex formation target DNA sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome. *BMC Genomics.* 2009; 10 Suppl 3:S9. doi: [10.1186/1471-2164-10-S3-S9](#) PMID: [19958507](#)
53. Jenjaroenpun P, Chew CS, Yong TP, Choowongkamon K, Thammasorn W, Kuznetsov VA. The TTSMI database: a catalog of triplex target DNA sites associated with genes and regulatory elements in the human genome. *Nucleic Acids Res.* 2015; 43: D110–116. doi: [10.1093/nar/gku970](#) PMID: [25324314](#)
54. Wang E, Feigon J. Structures of Nucleic Acid Triplexes. In: Neidle S, editor. *Oxford Handbook of Nucleic Acid Structure.* New York: Oxford University Press Inc; 1999. pp. 355.
55. Soyfer VN, Potaman VN. *Triple-Helical Nucleic Acids.* New York: Springer Verlag; 1996.
56. Frank-Kamenetskii MD, Mirkin SM. Triplex DNA structures. *Annu Rev Biochem.* 1995; 64: 65–95. PMID: [7574496](#)
57. Sun JS, Garestier T, Helene C. Oligonucleotide directed triple helix formation. *Curr Opin Struct Biol.* 1996; 6: 327–333. PMID: [8804836](#)
58. Duca M, Vekhoff P, Oussedik K, Halby L, Arimondo PB. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* 2008; 36: 5123–5138. doi: [10.1093/nar/gkn493](#) PMID: [18676453](#)
59. Buske FA, Mattick JS, Bailey TL. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol.* 2011; 8: 427–439. PMID: [21525785](#)
60. Devi G, Zhou Y, Zhong Z, Toh DF, Chen G. RNA triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscip Rev RNA.* 2015; 6: 111–128. doi: [10.1002/wrna.1261](#) PMID: [25146348](#)
61. Chubb JM, Hogan ME. Human therapeutics based on triple helix technology. *Trends Biotechnol.* 1992; 10: 132–136. PMID: [1374262](#)
62. Giovannangeli C, Rougee M, Garestier T, Thuong NT, Helene C. Triple-helix formation by oligonucleotides containing the three bases thymine, cytosine, and guanine. *Proc Natl Acad Sci U S A.* 1992; 89: 8631–8635. PMID: [1528873](#)
63. Keppler MD, Fox KR. Relative stability of triplexes containing different numbers of T.AT and C+.GC triplets. *Nucleic Acids Res.* 1997; 25: 4644–4649. PMID: [9358177](#)
64. Asensio JL, Lane AN, Dhese J, Bergqvist S, Brown T. The contribution of cytosine protonation to the stability of parallel DNA triple helices. *J Mol Biol.* 1998; 275: 811–822. PMID: [9480771](#)
65. Svinarchuk F, Monnot M, Merle A, Malvy C, Fermandjian S. The high stability of the triple helices formed between short purine oligonucleotides and SIV/HIV-2 vpx genes is determined by the targeted DNA structure. *Nucl. Acids Res.* 1995; 23: 3831–3836. PMID: [7479024](#)
66. Noonberg SB, François JC, Garestier T, Hélène C. Effect of competing self-structure on triplex formation with purine-rich oligodeoxynucleotides containing GA repeats. *Nucleic Acids Res.* 1995; 23: 1956–1963. PMID: [7596824](#)

67. Radhakrishnan I, de los Santos C, Patel DJ. Nuclear magnetic resonance structural studies of intramolecular purine.purine.pyrimidine DNA triplexes in solution. Base triple pairing alignments and strand direction. *J.Mol. Biol.* 1991; 221: 1403–1418. PMID: [1942059](#)
68. Beal PA, Dervan PB. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science.* 1991; 251: 1360–1363. PMID: [2003222](#)
69. Radhakrishnan I, Patel DJ. Solution structure of a purine.purine.pyrimidine DNA triplex containing G. GC and T.AT triples. *Structure.* 1993; 1: 135–152. PMID: [8069626](#)
70. Chandler SP, Fox KR. Specificity of antiparallel DNA triple helix formation *Biochemistry.* 1996; 35: 15038–15048. PMID: [8942670](#)
71. Gondeau C, Maurizot JC, Durand M. Spectroscopic studies on ethidium bromide binding to intramolecular parallel and antiparallel triple helices containing T*A:T and G*G:C triplets. *J Biomol Struct Dyn.* 2000; 17: 879–886. PMID: [10798532](#)
72. Scaria PV, Shafer RH. Calorimetric analysis of triple helices targeted to the dG3A4G3.dC3T4C3 duplex. *Biochemistry.* 1996; 35: 10985–10994. PMID: [8718892](#)
73. Clarenc JP, Lebleu B, Léonetti JP. Base changes and triple-helix hybridization properties of GT containing third strands: A systematic study. *Nucl. Nucl.* 1994; 13: 799–809.
74. Sun JS, De Bizemont T, Duval-Valentin G, Montenay-Garestier T, Helene C. Extension of the range of recognition sequences for triple helix formation by oligonucleotides containing guanines and thymine. *C R Acad Sci III.* 1991; 313: 585–590. PMID: [1782564](#)
75. McShan WM, Rossen RD, Laughter AH, Trial J, Kessler DJ, Zendegeui JG, Hogan ME, Orson FM. Inhibition of transcription of HIV-1 in infected human cells by oligodeoxynucleotides designed to form DNA triple helices. *J. Biol. Chem.* 1992; 267: 5712–5721. PMID: [1544943](#)
76. Diviacco S, Rapozzi V, Xodo L, Helene C, Quadrifoglio F, Giovannangeli C. Site-directed inhibition of DNA replication by triple helix formation. *FASEB J.* 2001; 15: 2660–2668. PMID: [11726542](#)
77. Mayfield C, Squibb M, Miller D. Inhibition of nuclear protein binding to the human Ki-ras promoter by triplex-forming oligonucleotides. *Biochemistry.* 1994; 33: 3358–3363. PMID: [8136373](#)
78. Durland RH, Kessler DJ, Gunnell S, Duvic M, Pettitt BM, Hogan ME. Binding of triple helix forming oligonucleotides to sites in gene promoters. *Biochemistry.* 1991; 30: 9246–9255. PMID: [1892832](#)
79. Macaya RF, Gilbert DE, Malek S, Sinsheimer JS, Feigon J. Structure and stability of X.G.C mismatches in the third strand of intramolecular triplexes. *Science.* 1991; 254: 270–274. PMID: [1925581](#)
80. de Bizemont T, Duval-Valentin G, Sun JS, Bisagni E, Garestier T, Helene C. Alternate strand recognition of double-helical DNA by T,G-containing oligonucleotides in the presence of a triple helix-specific ligand. *Nucleic Acids Res.* 1996; 24: 1136–1143. PMID: [8604349](#)
81. Brodin P, Sun JS, Mouscadet JF, Auclair C. Optimization of alternate-strand triple helix formation at the 5"-TpA-3" and 5"-ApT-3" junctions. *Nucleic Acids Res.* 1999; 27: 3029–3034. PMID: [10454596](#)
82. Thenmalarchelvi R, Yathindra N. New insights into DNA triplexes: residual twist and radial difference as measures of base triplet nonisomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acids Res.* 2005; 33: 43–55. PMID: [15657986](#)
83. Rathinavelan T, Yathindra N. Base triplet nonisomorphism strongly influences DNA triplex conformation: Effect of nonisomorphic G*GC and A*AT triplets and bending of DNA triplexes. *Biopolymers.* 2006; 82: 443–461. PMID: [16493655](#)
84. Liberles DA, Dervan PB. Design of artificial sequence-specific DNA bending ligands. *Proc Natl Acad Sci USA.* 1996; 93: 9510–9514. PMID: [8790361](#)
85. Akiyama T, Hogan ME. The design of an agent to bend DNA. *Proc Natl Acad Sci USA.* 1996; 93: 12122–12127. PMID: [8901543](#)
86. Bednarski D, Firestone SM. Regulation of Transcription by Synthetic DNA-Bending Agents. *Chem. Bio Chem.* 2006; 7: 1715–1721. PMID: [17004274](#)
87. Guillonneau F, Guieysse AL, Caer JPL, Rossier J, Praseuth D. Selection and identification of proteins bound to DNA triple-helical structures by combination of 2D-electrophoresis and MALDI-TOF mass spectrometry. *Nucleic Acids Res.* 2001; 29: 2427–2436. PMID: [11376162](#)
88. Jiménez-García E, Vaquero A, Espinás ML, Soliva R, Orozco M, Bernués J, Azorín F. The GAGA factor of *Drosophila* binds triple-stranded DNA. *J Biol Chem.* 1998; 273: 24640–24648. PMID: [9733760](#)
89. Kukreti S, Sun JS, Garestier T, Helene C. Extension of the range of DNA sequences available for triple helix formation: stabilization of mismatched triplexes by acridine-containing oligonucleotides. *Nucleic Acids Res.* 1997; 25: 4264–4270. PMID: [9336456](#)
90. Lexa M, Martínek T, Burgetová I, Kopeček D, Brázdová M. A dynamic programming algorithm for identification of triplex-forming sequences. *Bioinformatics.* 2011; 27: 2510–2517. doi: [10.1093/bioinformatics/btr439](#) PMID: [21791534](#)

91. Ananth P, Goldsmith G, Yathindra N. An innate twist between Crick's wobble and Watson-Crick base pairs. *RNA*. 2013; 19: 1038–1053. doi: [10.1261/rna.036905.112](https://doi.org/10.1261/rna.036905.112) PMID: [23861536](https://pubmed.ncbi.nlm.nih.gov/23861536/)
92. Arnott S, Selsing E. Structures for the polynucleotide complexes poly(dA) with poly (dT) and poly(dT) with poly(dA) with poly (dT). *J Mol Biol*. 1974; 88: 509–521. PMID: [4453005](https://pubmed.ncbi.nlm.nih.gov/4453005/)
93. Brunger AT. X-PLOR Ver 3.851: Yale University, NewYork. 1996
94. Case DA, Darden TA, Cheatham TE 3rd, Simmering CL, Wang J, Duke RE, et al. 2012, AMBER 12, University of California, San Francisco.
95. Hart K, Foloppe N, Baker CM, Denning EJ, Nilsson L, Mackerell AD Jr. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput*. 2012; 8:348–362. PMID: [22368531](https://pubmed.ncbi.nlm.nih.gov/22368531/)
96. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc*. 1995; 117: 5179–5197.
97. Perez A, Marchan I, Svozil D, Sponer J, Cheatham TE III, Laughton CA, Orozco M. Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* 2007; 92:3817–3829. PMID: [17351000](https://pubmed.ncbi.nlm.nih.gov/17351000/)
98. Krepl M, Zgarbová M, Stadlbauer P, Otyepka M, Banáš P, Koča J, et al Reference simulations of non-canonical nucleic acids with different χ variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA. *J Chem Theory Comput*. 2012; 8: 2506–2520. PMID: [23197943](https://pubmed.ncbi.nlm.nih.gov/23197943/)
99. Zgarbová M, Luque FJ, Sponer J, Cheatham TE 3rd, Otyepka M, Jurečka P. Toward Improved Description of DNA Backbone: Revisiting Epsilon and Zeta Torsion Force Field Parameters. *J Chem Theory Comput*. 2013; 9: 2339–2354. PMID: [24058302](https://pubmed.ncbi.nlm.nih.gov/24058302/)
100. Zgarbová M, Otyepka M, Sponer J, Mládek A, Banáš P, Cheatham TE 3rd, Jurečka P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput*. 2011; 7: 2886–2902. PMID: [21921995](https://pubmed.ncbi.nlm.nih.gov/21921995/)
101. Noy A, Pérez A, Laughton CA, Orozco M. Theoretical study of large conformational transitions in DNA: the B \leftrightarrow A conformational change in water and ethanol/water. *Nucleic Acids Res*. 2007; 35: 3330–3338. PMID: [17459891](https://pubmed.ncbi.nlm.nih.gov/17459891/)
102. Moradi M, Babin V, Roland C, Sagui C. Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study. *Nucleic Acids Res*. 2013; 41: 33–43. doi: [10.1093/nar/gks1003](https://doi.org/10.1093/nar/gks1003) PMID: [23104380](https://pubmed.ncbi.nlm.nih.gov/23104380/)
103. Khan N, Kolimi N, Rathinavelan T. Twisting right to left: A . . . A mismatch in a CAG trinucleotide repeat overexpansion provokes left-handed Z-DNA conformation. *PLoS Comput Biol*. 2015; 11(4): e1004162. doi: [10.1371/journal.pcbi.1004162](https://doi.org/10.1371/journal.pcbi.1004162) PMID: [25876062](https://pubmed.ncbi.nlm.nih.gov/25876062/)
104. Mura C, McCammon JA. Molecular dynamics of a kappaB DNA element: base flipping via cross-strand intercalative stacking in a microsecond-scale simulation. *Nucleic Acids Res*. 2008; 36: 4941–4955. doi: [10.1093/nar/gkn473](https://doi.org/10.1093/nar/gkn473) PMID: [18653524](https://pubmed.ncbi.nlm.nih.gov/18653524/)
105. Dršata T, Zgarbová M, Špačková N, Jurečka P, Šponer J, Lankaš F. Mechanical Model of DNA Allostery. *J Phys Chem Lett*. 2014; 5: 3831–3835. doi: [10.1021/jz501826q](https://doi.org/10.1021/jz501826q) PMID: [26278756](https://pubmed.ncbi.nlm.nih.gov/26278756/)
106. Chakravorty DK, Merz KM Jr. Studying allosteric regulation in metal sensor proteins using computational methods. *Adv Protein Chem Struct Biol*. 2014; 96: 181–218. doi: [10.1016/bs.apcsb.2014.06.009](https://doi.org/10.1016/bs.apcsb.2014.06.009) PMID: [25443958](https://pubmed.ncbi.nlm.nih.gov/25443958/)
107. Noy A, Perez A, Marquez M, Luque FJ, Orozco M. Structure, recognition properties, and flexibility of the DNA–RNA hybrid. *J Am Chem Soc*. 2005; 127: 4910–4920. PMID: [15796556](https://pubmed.ncbi.nlm.nih.gov/15796556/)
108. McCrate NE, Varner ME, Kim KI, Nagan MC. Molecular dynamics simulations of human tRNA Lys₃ UUU: the role of modified bases in mRNA recognition. *Nucleic Acids Res*. 2006; 34: 5361–5368. PMID: [17012271](https://pubmed.ncbi.nlm.nih.gov/17012271/)
109. Ishida H, Matsumoto A Free-energy landscape of reverse tRNA translocation through the ribosome analyzed by electron microscopy density maps and molecular dynamics simulations. *PLoS One*. 2014; 9(7):e101951. doi: [10.1371/journal.pone.0101951](https://doi.org/10.1371/journal.pone.0101951) PMID: [24999999](https://pubmed.ncbi.nlm.nih.gov/24999999/)
110. Singh RP, Blossey R, Cleri F. Structure and mechanical characterization of DNA i-motif nanowires by molecular dynamics simulation. *Biophys J*. 2013; 105: 2820–2831. doi: [10.1016/j.bpj.2013.10.021](https://doi.org/10.1016/j.bpj.2013.10.021) PMID: [24359754](https://pubmed.ncbi.nlm.nih.gov/24359754/)
111. Wheatley EG, Pieniazek SN, Mukerji I, Beveridge DL. Molecular dynamics of a DNA Holliday junction: the inverted repeat sequence d(CCGGTACCGG)₄. *Biophys J*. 2012; 102: 552–560. doi: [10.1016/j.bpj.2011.11.4023](https://doi.org/10.1016/j.bpj.2011.11.4023) PMID: [22325278](https://pubmed.ncbi.nlm.nih.gov/22325278/)
112. Rathinavelan T, Yathindra N. Molecular dynamics structures of peptide nucleic acid x DNA hybrid in the wild-type and mutated alleles of Ki-ras proto-oncogene—stereochemical rationale for the low affinity of PNA in the presence of an AC mismatch. *FEBS J*. 2005; 272: 4055–4070. PMID: [16098189](https://pubmed.ncbi.nlm.nih.gov/16098189/)

113. Islam B, Stadlbauer P, Krepl M, Koca J, Neidle S, Haider S, et al. Extended molecular dynamics of a c-kit promoter quadruplex. *Nucleic Acids Res.* 2015; 43: 8673–8693. doi: [10.1093/nar/gkv785](https://doi.org/10.1093/nar/gkv785) PMID: [26245347](https://pubmed.ncbi.nlm.nih.gov/26245347/)
114. Spomer J, Cang X, Cheatham TE 3rd. Molecular dynamics simulations of G-DNA and perspectives on the simulation of nucleic acid structures. *Methods.* 2012; 57: 25–39. doi: [10.1016/j.ymeth.2012.04.005](https://doi.org/10.1016/j.ymeth.2012.04.005) PMID: [22525788](https://pubmed.ncbi.nlm.nih.gov/22525788/)
115. Sekharudu CY, Yathindra N, Sundaralingam M. Molecular dynamics investigations of DNA triple helical models: unique features of the Watson-Crick duplex. *J Biomol Struct Dyn.* 1993; 11: 225–44. PMID: [8286053](https://pubmed.ncbi.nlm.nih.gov/8286053/)
116. Raghunathan G, Miles HT, Sasisekharan V. Symmetry and molecular structure of a DNA triple helix: d(T)n.d(A)n.d(T)n. *Biochemistry.* 1993; 32: 455–462. PMID: [8422354](https://pubmed.ncbi.nlm.nih.gov/8422354/)
117. Macaya R, Schultze P, Feigon J. Sugar conformations in intramolecular DNA triplexes determined by coupling constants obtained by automated simulation of P.COSY cross peaks. *J. Am. Chem. Soc.* 1992; 114: 781–783.
118. Macaya R, Wang E, Schultze P, Sklenár V, Feigon J. Proton nuclear magnetic resonance assignments and structural characterization of an intramolecular DNA triplex. *J Mol Biol.* 1992; 225: 755–773. PMID: [1318385](https://pubmed.ncbi.nlm.nih.gov/1318385/)
119. Ojha RP, Tiwari RK. Molecular dynamics simulation study of DNA triplex formed by mixed sequences in solution. *J Biomol Struct Dyn.* 2002; 20: 107–126. PMID: [12144358](https://pubmed.ncbi.nlm.nih.gov/12144358/)
120. Tateishi-Karimata H, Nakano M, Sugimoto N. Comparable stability of Hoogsteen and Watson-Crick base pairs in ionic liquid choline dihydrogen phosphate. *Nature Sci. Rep.* 2014; 4: 3593.
121. Ojha RP, Tiwari RK. Triplex hydration: nanosecond molecular dynamics simulation of the solvated triplex formed by mixed sequences. *Nucleic Acids Res.* 2003; 31: 6373–6380. PMID: [14576325](https://pubmed.ncbi.nlm.nih.gov/14576325/)
122. Petrov AS, Lamm G, Pack GR. The Triplex-Hairpin Transition in Cytosine-Rich DNA. *Biophys J.* 2004; 87: 3954–3973. PMID: [15377537](https://pubmed.ncbi.nlm.nih.gov/15377537/)
123. Case DA, Darden TA, Cheatham TE 3rd, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, et al. 2012, AMBER 13, University of California, San Francisco.
124. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics.* 1977; 23: 327–341.
125. Lu XJ, Olson WK. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 2003; 31: 5108–5121. PMID: [12930962](https://pubmed.ncbi.nlm.nih.gov/12930962/)
126. Insight II 2005L. 2005 Accelrys Inc, San Diego, CA, USA.
127. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Accounts of Chemical Research.* 2000; 33: 889–897. PMID: [11123888](https://pubmed.ncbi.nlm.nih.gov/11123888/)
128. Gohlke H, Case DA. Converging free energy estimates: MM-PB GB SA studies on the protein–protein complex Ras–Raf. *Journal of Computational Chemistry.* 2004; 25: 238–250. PMID: [14648622](https://pubmed.ncbi.nlm.nih.gov/14648622/)