JBC Papers in Press. Published on September 18, 2017 as Manuscript M117.796235
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M117.796235

A B–Z junction induced by an A...A mismatch in GAC repeats in the gene for cartilage oligomeric matrix protein promotes binding with the hZα$_{ADAR1}$ protein

Narendar Kolimi†, Yogeeshwar Ajjugal† and Thenmalarchelvi Rathinavelan*

Department of Biotechnology, Indian Institute of Technology Hyderabad,
Kandi, Telangana State 502285, India

Running title: A model for GAC duplex…hZα$_{ADAR1}$ complex

†These authors contributed equally
*For Correspondence: tr@iith.ac.in

Keywords: GAC repeat expansion, pseudoachondroplasia, multiple epiphyseal dysplasia, B- Z junction, human ADAR1 protein, COMP, RNA editing, passive binding mechanism, backbone flipping, A…A mismatch, nonisostericity, nuclear magnetic resonance, circular dichroism, molecular dynamics simulation, microscale thermophoresis, dissociation constant

ABSTRACT

GAC repeat expansion from five to seven in the exonic region of the gene for cartilage oligomeric matrix protein (COMP) leads to pseudoachondroplasia, a skeletal abnormality. However, the molecular mechanism by which GAC expansions in the *COMP* gene lead to skeletal dysplasias is poorly understood. Here, we used MD simulations which indicate that an A...A mismatch in a d(GAC)$_6$.d(GAC)$_6$ duplex induces negative supercoiling, leading to a local B-to-Z DNA transition. This transition facilitates the binding of d(GAC)$_7$.d(GAC)$_7$ with the Zα-binding domain of human adenosine deaminase acting on RNA 1 (ADAR1, hZα$_{ADAR1}$), as confirmed by CD, NMR and microscale thermophoresis studies. The CD results indicated that hZα$_{ADAR1}$ recognizes the zigzag backbone of d(GAC)$_7$.d(GAC)$_7$ at the B–Z junction and subsequently converts it into Z-DNA via the so-called passive mechanism. MD simulations carried out for the modeled hZα$_{ADAR1}$–d(GAC)$_6$.d(GAC)$_6$ complex confirmed the retention of previously reported important interactions between the two molecules. These findings suggest that hZα$_{ADAR1}$ binding with the GAC hairpin stem in *COMP* can lead to a non-genetic, RNA editing–mediated substitution in the COMP that may then play a crucial role in the development of pseudoachondroplasia.

INTRODUCTION

Expansion of GAC repeat sequences can be observed in exonic region of the genome which can lead to poly-Asp track in cartilage oligomeric matrix protein (COMP) [1]. COMP is a noncollagenous pentameric extracellular matrix protein that is localized in the chromosome 19p13.1 [2]. It is expressed predominantly in cartilage as well as transiently in tendons, ligaments, smooth muscles *etc* [3, 4] and is important for growth plate organization & its function [5]. COMP gene contains five tandem GAC repeats and expansion of even one or two repeats cause multiple epiphyseal dysplasia or pseudoachondroplasia respectively. Such expansion in pseudoachondroplasia results in short stature, early onset-osteoarthritis and limb dwarfism [3]. Mutations in COMP also cause disruption of calcium/ligand binding, intramolecular interactions and disulfide bond formation [5].

Although an *in vitro* study has shown that GAC repeat exhibit orientation dependent instability that subsequently leads to repeat deletion and expansion during replication and transcription respectively [6], the exact mechanism about how GAC expansion in COMP gene leads to skeletal dysplasias is

Copyright 2017 by The American Society for Biochemistry and Molecular Biology, Inc.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

poorly understood. Interestingly, a very recent molecular dynamics (MD) study indicates that nonisostericity of A...A mismatch with respect to the flanking canonical base pairs provokes left-handed Z-DNA conformation in CAG repeat expansion [7]. As d(GAC) repeats can also have similar periodic A…A mismatches, we investigate here its effect on DNA conformation using MD simulation and circular dichroism (CD) studies. Indeed, earlier investigations report that d(GAC)15 [8] and r(CGA)17 [9] form stable hairpin structures that may have periodic A…A mismatches. Previous CD studies also indicate that d(GAC) repeats exhibit multiple conformations and even form parallel duplex under acidic environment [10, 11]. Nonetheless, stereochemical rationale behind such non-B-DNA secondary structural preference by GAC repeats and the consequent biological significances are unknown. MD simulations performed here to explore this clearly underpin that nonisosteric character of A...A mismatch leads to B-Z junction by inducing negative supercoiling akin to CAG repeats [7] as also confirmed by CD. Further, CD and nuclear magnetic resonance (NMR) titration experiments reveal for the first time that GAC interacts with Zα binding domain of human adenosine deaminase acting on RNA (ADAR1) (herein onwards, $hZ\alpha_{ADAR1}$ that stands for human Zα binding domain of ADAR1) through specific recognition of B-Z junctions. Microscale thermophoresis (MST) experiment further reveals that $d(GAC)_7.d(GAC)_7$ duplex interacts with $hZ\alpha_{ADAR1}$ in nanomolar binding affinity.

ADAR family of proteins mainly catalyze adenosine to inosine editing process in pre-mRNA substrates [12]. In fact, the role of hyper-/altered A-to-I editing mediated by ADAR in several neurological disorders is well established [13-16]. The N-terminus of ADAR1 contains Z-DNA binding winged helix turn helix (wHTH) domain. Using MD derived structure of $d(GAC)_6.d(GAC)_6$, published crystal structure of $hZ\alpha_{ADAR1}$ (PDB ID: 2ACJ) and NMR chemical shift mapping of $hZ\alpha_{ADAR1}$ & $d(GAC)_7.d(GAC)_7$ titration, their complex model is proposed here. Subsequent MD simulation of the complex model confirms the importance of certain amino acids in $hZ\alpha_{ADAR1}$ recognizing B-Z/Z-DNA conformation. Based on $hZ\alpha_{ADAR1}$ & d(GAC).d(GAC) binding studies, a model about how $hZ\alpha_{ADAR1}$ can anchor onto Z-philic GAC repeat and facilitate A-to-I editing of the corresponding mRNA transcript of COMP and lead to pseudoachondroplasia is also proposed.

RESULTS

A…A mismatch amidst G...C & C...G base pairs imposes B-Z junction formation

For MD simulation, $d(GAC)_6.d(GAC)_6$ repeat sequence has been considered, so as to complete one helical turn of a DNA duplex (i.e. 10 bases per turn in a normal B-DNA). Flanking sequences (one GAC repeat) on either sides are added to avoid the end fraying effect during MD simulation. Effect of A...A mismatch in $d(GAC)_6.d(GAC)_6$ duplex (Fig. 1A) has been investigated at the atomistic level using molecular dynamics (MD) simulations by considering two different starting *glycosyl* conformations for the mismatch following the earlier studies [7, 17]. In the first model, both A's are chosen to have *anti* conformation (*anti...anti*). On the other hand, one of the two A's in the second model is chosen to be in *anti glycosyl* conformation, and the other is chosen to be in *+syn glycosyl* conformation (*+syn...anti*).

Starting model with anti...anti glycosyl conformation

Analysis of 500ns simulation of $d(GAC)_6.d(GAC)_6$ duplex that comprises of 6 A...A mismatches with *anti...anti glycosyl* conformation shows that nonisostericity of A...A mismatch with respect to canonical G...C and C...G base pairs induces distortions in the helix. Distortions are seen within ~2ns of the simulation and become prominent ~63ns through the unwinding of the helix and stays in the same conformation till the end of the simulation (Fig. 1B, supplemental Movie

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

M1). This eventually reflects in the root mean square deviation (RMSD) whose average value stays ~4.5Å between 0.5 to 10.5ns and ~6.5Å between 10.5-60.3ns, before finally reaching the highest value of ~9 Å (Fig. 1C).

Conformational features that are associated with such helix unwinding are: Firstly, A's taking up *high-anti glycosyl* conformation (61%), as well as exhibiting preference for -*syn* (39%) transiently during the simulation (Fig. 1D) and G's that are engaged in canonical hydrogen bonding with C's profoundly favor ±*syn glycosyl* conformation (92%) after 50ns (Fig. 1E).

Secondly, backbone torsion angles (ε,ζ,α,γ) that are calculated for the last 400ns (after reaching the equilibration state) show the possibility of BIII(g-,g-,g-,g+) [18] & Z(ZI=(g-,g+,g+,t)/ZII=(t,g-,t,g+)) [19] conformations at base steps such as GA (44%) and CG (23%) leading to zigzag backbone (Fig. 2A). Interestingly other unusual conformations mainly (t,t,g-,g+), (g-,g-,t,t), (t,g-,g+,t), (t,g-,t,t) and (g-,t,g+,t) are also seen predominantly (58%) (Fig. 2B). Such unusual conformations can be attributed to the interaction of cations with the duplex. However, nonisomorphic nature of A...A mismatch dominates over the interaction of counter ions in contribution for such unusual conformations [20] (supplemental Fig. S1A). AC steps have shown 36% preference for BI(t,g-,g-,g+)/BII(g-,t,g-,g+) conformations, which is higher than GA (11%) and CG (5%) steps (Fig. 2A). Such a mixed occurrence of a variety of conformations at different steps leads to B-Z junction in the duplex. As the result, the final conformation of the duplex deviates significantly from the starting B-form conformation (Fig. 1B), as also reflects in RMSD.

Concomitantly, helical twist angles at GA, AC and CG steps also exhibit variations. Among the three steps, CG steps exhibit a higher population of low helical twists (64% of helical twists lower than 10°), compared to AC (22%) and GA (20%) steps (Fig. 2E (left)). Yet another property that can support the formation of Z-DNA in the midst of B-DNA in $d(GAC)_6.d(GAC)_6$ duplex is the angle formed by three adjacent phosphates. For the average structure calculated over the last 10ns, the angle at the central phosphate in the following steps are below 110° that further supports the formation of local Z-DNA (~110°) [21]: G4pA5 (119°), C6pG7 (96°), G7pA8 (90°), C9pG10 (81°), G10pA11 (99°), C12pG13 (92°), G13pA14 (108°), A14pC15 (111°), C15pG16 (85°), A20pC21 (104°), C21pG22 (88°), C24pG25 (93°), G25pA26 (102°), A26pC27 (117°), C27pG28 (87°), G28pA29 (99°), C30pG31 (90°) & G31pA32 (104°). Remaining steps are confined to B-DNA conformation with angle at central phosphate close to ~150° (supplemental Fig. S2A).

Starting model with +syn...anti glycosyl conformation

To further explore the global conformational preference for A…A mismatch, another starting conformation, namely, *+syn...anti glycosyl* conformation is considered for the mismatch. This conformation is specifically chosen based on the *glycosyl* angle preference for A…A mismatch in the CAG containing RNA duplex [17, 22].

As seen above, nonisosteric character of A...A mismatch with respect to the flanking canonical G…C/C…G base pairs triggers unwinding of the helix ~90ns of the simulation (Fig. 3A, supplemental Movie M2). Time vs RMSD profile calculated with respect to the initial model (Fig. 3B) also indicative of significant deviation from initial model (~5Å). Associated conformational changes are: G's predominantly taking ±*syn glycosyl* conformation (76%) (Fig. 3C) along with A's in *anti/+syn* conformation (Fig. 3D). Exceptionally, some of the G's and A's briefly take *high-anti* and *–syn* conformation respectively. Backbone torsion angles such as (ε,ζ,α,γ) in GA steps exhibit the characteristics of Z-DNA (35%), whereas, AC (64%) and CG (45%) steps favor to form BI and BII conformations (Fig. 2C). CG (37%), GA (52%) & AC (28%) steps are also

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

populated by other conformations such as (t,t,g-,g+), (g-,g-,t,t), (t,g-,g+,t), (t,g-,t,t) and (g-,t,g+,t) (Fig. 2D). As discussed earlier, these unusual conformations may be due to the interaction of counter ions with the duplex (supplemental Fig. S1B). As before, CG steps (46% of helical twists less than 10°) exhibit a lower twist angle compared to GA (2%) & AC (2%) steps (Fig. 2E (right)). The angle at central phosphate in the steps C6pG7 (90°), G7pA8 (90°), G22pA23 (89°), C24pG25 (93°), G25pA26 (102°), A26pC27 (117°), C27pG28 (87°), A29pC30 (105°) and G31pA32 (85°) exhibit a very low value during the last 10ns, yet another supportive evidence for the presence of Z-DNA (supplemental Fig. S2B).

Together, these properties confirm the presence for B-Z junction in the GAC repeat containing duplex with *+syn...anti* starting *glycosyl* conformation for A…A mismatch. Nonetheless, preference for Z-conformation is less prominent compared to *anti...anti* starting *glycosyl* conformation.

Canonical base pairs containing $d(GAC)_6.d(GTC)_6$ and T...T mismatch containing $d(GTC)_6.d(GTC)_6$ duplexes retain B-form geometry

Control simulations carried out for $d(GAC)_6.d(GTC)_6$ duplex with G…C and A…T canonical base pairs (Fig. 4A) to pinpoint that B-Z junction formation observed in $d(GAC)_6.d(GAC)_6$ is purely due to nonisomorphism of A…A mismatch indicate the dominance of B-form geometry (Fig. 4B). Although, Z-DNA characteristics are observed during the simulation by 34% CG steps having helical twists lower than 10° (Fig. 4C) along with 42% of G's prefer ±*syn glycosyl* conformation (Fig. 4D), this is comparatively lower than in the mismatch situations (Figs. 1E and 3C). Its noteworthy, A's (96%) prefer predominantly *anti/high-anti glycosyl* conformations (Fig. 4E). Few GA/GT steps also show Z-DNA backbone conformation (Figs. 4F and G). In fact, such a minor population can be related to the presence of cations in the minor groove as pointed out in an earlier study (supplemental Fig. S1C) [20].

Similarly, MD simulation of $d(GTC)_6.d(GTC)_6$ duplex that contains 6 T…T mismatches in the place of 6 A…A mismatches in $d(GAC)_6.d(GAC)_6$ duplex (Fig. 5A), clearly indicate the preponderance for B-form geometry (Fig. 5B). Overall percentage of helical twists below 10° are 5%, 9% and 11% at CG, TC & GT steps respectively (Fig. 5C). Unlike A...A mismatch, G's (90%) favor *anti/high-anti glycosyl* conformations (Fig. 5D). Thus, Z-DNA backbone conformation is less observed here (Figs. 5E and F).

In summary, only a minor population of Z-DNA is observed in $d(GAC)_6.d(GTC)_6$ and $d(GTC)_6.d(GTC)_6$ duplexes compared to $d(GAC)_6.d(GAC)_6$ duplexes. To further validate that B-Z junction formation is mainly induced by A…A mismatch, circular dichroism (CD) studies have been carried out (see below).

CD confirms B-Z junction formation in $d(GAC)_7.d(GAC)_7$ duplex

At 50mM NaCl salt concentration, CD spectrum of $d(GAC)_7.d(GAC)_7$ shows a positive peak between 270-280nm and a negative peak around 260nm, a typical characteristic of B-form DNA (Fig. 6A). However, with the increase in NaCl concentration in the range of 0.05M to 4.2M, the negative ellipticity around 260nm moves towards positive ellipticity. Additionally, the spectra start developing two negative peaks (~290nm and ~205nm) with respect to increase in salt concentration that are all Z-DNA signature peaks (Fig. 6A). Nevertheless, canonical base pairs containing $d(GAC)_7.d(GTC)_7$ duplex doesn't exhibit any B-to-Z transition with respect to increase in NaCl concentration and stays in B-form with positive and negative peaks around 285nm & 260nm respectively (Fig. 6B). $D(GTC)_7.d(GTC)_7$ duplex that has 7 T...T mismatches also exhibit same tendency as $d(GAC)_7.d(GTC)_7$ duplex (supplemental Fig. S3). Thus, salt-dependent CD spectra clearly

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

indicate that A…A mismatch dictates B-Z junction formation, that subsequently converts the duplex to complete Z-form at a higher salt concentration.

Zα domain of human ADAR1 binds with d(GAC)$_7$.d(GAC)$_7$ duplex

CD spectra of d(GAC)$_7$.d(GAC)$_7$ duplex (N) and hZα$_{ADAR1}$ protein (P) titration clearly show that increasing the concentration of protein (P) (viz., increasing P/N ratio by keeping N as a constant) completely changes the duplex to left-handed Z-form. As the concentration of hZα$_{ADAR1}$ (P) increases, the negative peak ~255nm gradually diminishes accompanied with the appearance of a new negative peak ~295nm and a shift in the positive peak from 280nm to 275nm, characteristic features of Z-DNA conformation (Fig. 6C). In contrast, d(GAC)$_7$.d(GTC)$_7$ duplex that contains only canonical base pairs doesn't exhibit such tendency for B to Z transition (Fig. 6D). Such scenario is seen irrespective of the number of repeats in the duplex. For instance, d(GAC)$_6$.d(GAC)$_6$ duplex that has 6 A…A mismatches also takes up Z-form upon increasing P/N ratio (supplemental Fig. S4A), while the corresponding canonical duplex does not exhibit such characteristics (supplemental Fig. S4B). In fact, d((GAC)$_3$T$_4$(GAC)$_3$) (wherein, one GAC in the d(GAC)$_7$ is replaced by T$_4$ to facilitate the hairpin formation) that is expected to form a hairpin with 3 A…A mismatches also exhibit B to Z transition upon titrating with hZα$_{ADAR1}$ (supplemental Fig. S5). This situation mimics hairpin formation in d(GAC)$_7$ by having one GAC repeat in the hairpin loop and 6 GAC repeats in the stem with 3 A…A mismatches. Although there is a possibility that d(GAC)$_7$ can take up either intramolecular hairpin conformation (with 3 A…A mismatches) or intermolecular duplex conformation (with 7 A…A mismatches) in solution (supplemental Fig. S6), it is difficult to identify the preferred conformation from CD data. Indeed, both the conformations may equally be populated *in vitro*, unlike *in vivo*, wherein, it can take up only hairpin conformation. Thus, d((GAC)$_3$T$_4$(GAC)$_3$) titration with hZα$_{ADAR1}$ confirms that d(GAC)$_7$ can adopt stable hairpin conformation with 3 A…A mismatches that subsequently facilitate the binding with the protein.

hZα$_{ADAR1}$ binds d(GAC)$_7$.d(GAC)$_7$ with nanomolar affinity

In accordance with CD results, 1D proton NMR spectra of hZα$_{ADAR1}$ and d(GAC)$_7$.d(GAC)$_7$ duplex titration also confirms the interaction between these two. Overall, the spectra show gradual reduction in peak intensity as the concentration of hZα$_{ADAR1}$ increases. Though it may be difficult to identify the amino acids/nucleotides that are associated with the proton chemical shifts simply from the 1D spectra, the signature chemical shifts around 0 to -1ppm, 9.8ppm and 9.6ppm can be assigned to protons corresponding to Thr191 [23], Hε1 proton of Trp195 and amide proton of Ala158 [24] respectively that are located in & around the binding site of hZα$_{ADAR1}$ (PDB ID: 2ACJ). Likewise, the chemical shifts between 5 to 6ppm belong to backbone protons of DNA duplex [25]. Notably, protein and DNA chemical shifts in these regions do not overlap with each other (Fig. 7A).

Microscale thermophoresis (MST) exhibits nanomolar binding affinity between hZα$_{ADAR1}$ and d(GAC)$_7$.d(GAC)$_7$ duplex with a dissociation constant (K$_D$) of 41 nanomolar (nM) (Fig. 7B).

d(GAC)$_7$.d(GAC)$_7$...hZα$_{ADAR1}$ complex model

Aforementioned information about hZα$_{ADAR1}$ amino acids (Thr191, Trp195 & Ala158) that may be involved in interaction with d(GAC)$_7$.d(GAC)$_7$ duplex (Fig. 7A) along with the readily available complex structure of hZα$_{ADAR1}$ and a B-Z junction (PDB ID: 2ACJ) have been used to model hZα$_{ADAR1}$...d(GAC)$_6$.d(GAC)$_6$ complex. Fig. 7C shows the modeled complex derived from X-ray (former) and MD (latter) structures. As more than one hZα$_{ADAR1}$ can bind to a single duplex depending on the availability of Z-philic centers (PDB ID: 2ACJ) [21],

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

$d(GAC)_6.d(GAC)_6$ can also accommodate more than one $hZ\alpha_{ADAR1}$ molecule (Fig. 7D).

MD simulation retains the conserved interactions between $hZ\alpha_{ADAR1}$ and DNA duplex

Modeled $hZ\alpha_{ADAR1}...d(GAC)_6.d(GAC)_6$ complex has been subjected to 300ns MD simulations to optimize the interaction between the two. Its noteworthy that the complex has been modeled such that two monomers of $hZ\alpha_{ADAR1}$ interact with 2 different strands of the duplex (Fig. 7C) as reported earlier (PDB ID: 2ACJ). Analysis of the MD trajectories reveals that $hZ\alpha_{ADAR1}$ interacts with the duplex through its minor groove (Fig. 8A). K169, N173 & R174 residues of $hZ\alpha_{ADAR1}$ monomers participate in hydrogen bonding interaction with the duplex backbone atoms (like O5', O1P & O2P) either transiently or persistently (Figs. 8B-D). This is consistent with previous mutagenesis and NMR studies [26, 27] that show the importance of the above mentioned residues in facilitating the interaction between the two. However, minor difference in the nature of interaction is also seen. For instance, Y177 that is involved in stacking interaction in the crystal structure (PDB ID: 2ACJ) is engaged in transient hydrogen bonding interaction with the sugar-phosphate backbone atoms (Fig. 8E). Similarly, W195 does not participate in any direct hydrogen bonding interaction with the duplex though it lies in the proximity of the duplex (Fig. 8E). Thus, the unwinding of $d(GAC)_6.d(GAC)_6$ duplex due to the presence of A…A mismatch (Figs. 1B and 3A) facilitate the interaction of $hZ\alpha_{ADAR1}$ protein at the minor groove.

DISCUSSION

Left-handed Z-DNA is a higher energy conformation compared to canonical B-DNA conformation [28] and *in vitro,* $d(GC)_n$ sequences are shown to choose Z-form at extreme conditions like high salt concentrations [29]. There are increasing evidences about the participation of Z-DNA in gene regulation, formation of which is believed to relieve the stress on DNA structure through negative supercoiling [30]. Proteins that specifically recognize and bind to Z-DNA are also identified: $hZ\alpha_{ADAR1}$ [27], E3L [31], DLM1 [32] and PKZ [33]. Interconversion between B & Z-DNA is believed to take place either through 'stretch collapse mechanism' or through 'zipper mechanism' [34, 35], facilitated by base extrusion and base and/or backbone flipping. Intriguingly, A…A mismatch in the hairpin stem of CAG repeat readily exhibits preponderance for Z-DNA conformation through 'zipper mechanism' [7]. As GAC repeats that are responsible for pseudoachondroplasia also contain periodic A…A mismatches, we investigate here the ability of the same to adopt Z-form structure by employing MD simulation, CD, MST and NMR techniques. Subsequently, its ability to bind with $hZ\alpha_{ADAR1}$ protein is also explored.

A…A mismatch induces local B-to-Z transition through backbone flipping

MD simulations carried out by considering two different models to explore all the possible conformational preference for $d(GAC)_6.d(GAC)_6$ duplex reveal that A…A mismatch leads to local Z-DNA formation irrespective of *anti…anti* & *+syn…anti* starting *glycosyl* conformations. Such a conformational change is facilitated by backbone flipping through the base steps taking local Z-DNA and other non-B-DNA backbone conformations. This is further concomitant by G's and a few A's favoring *±syn glycosyl* conformation (Figs. 1D and 3C). Not surprisingly, pyrimidines favor *anti glycosyl* conformation. Indeed, such a mixed occurrence of *syn* and *anti glycosyl* conformations together with aforementioned non-B-DNA backbone conformations lead to Z-DNA features along side with B-DNA conformation. Normalized frequency of occurrence of G's falling in *±syn* conformation in $d(GAC)_6.d(GAC)_6$ (*+syn…anti*), $d(GAC)_6.d(GTC)_6$ and $d(GTC)_6.d(GTC)_6$ duplexes with respect to $d(GAC)_6.d(GAC)_6$ (*anti…anti*) duplex are 0.8, 0.4 and 0 respectively. This clearly indicates the influence of A…A mismatch in inducing

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

B-to-Z transition in $d(GAC)_6.d(GAC)_6$ in contrast to $d(GAC)_6.d(GTC)_6$ and $d(GTC)_6.d(GTC)_6$ duplexes. A minor population of B-Z junction is observed in $d(GAC)_6.d(GTC)_6$ may also be attributed to the interaction with counter ions (supplemental Fig. S1C). This eventually reflects in CG step taking low twist in the midst of high twists at AC & GA steps in $d(GAC)_6.d(GAC)_6$ duplexes (64% and 46% in *anti...anti* and *+syn...anti glycosyl* conformations respectively) (Fig. 2E) leading to B-Z junction in the vicinity of the mismatch. Such an occurrence of high and low twists in the duplex leads to unwinding of the helix, a typical characteristic of B-Z junction as observed in the crystal structures (PDB ID: 1FV7 and 2ACJ) (supplemental Fig. S7). It is noteworthy that B-Z junction does not show alternating *glycosyl* (*±syn* and *anti*) and backbone conformations as in the Z-form, wherein, alternating *glycosyl* conformation lead to zig-zag backbone [36, 37]. Instead, B-Z junction possesses the characteristics of both B-&Z-forms.

In sharp contrast to $d(GAC)_6.d(GAC)_6$ duplex, $d(GAC)_6.d(GTC)_6$ duplex with canonical base pairs and $d(GTC)_6.d(GTC)_6$ duplex with T…T mismatch have preference for B-form geometry. This finding is further confirmed by CD spectroscopy by titrating NaCl with $d(GAC)_7.d(GAC)_7$, $d(GTC)_7.d(GTC)_7$ and $d(GAC)_7.d(GTC)_7$ duplexes: while $d(GAC)_7.d(GAC)_7$ clearly displays B-Z to Z transition with respect to the increase in salt concentration, the other two do not exhibit such transition (Figs. 6A, 6B and supplemental Fig. S3). Such an inclination of A…A mismatch towards Z-form is due to its nonisostericity exemplified by a high residual twist and radial difference with the flanking C…G/G…C base pairs [18], offering discomfort to accommodate itself in a B-DNA. Thus, it unwinds the helix to relieve the mechanistic effect arising from the nonisostericity of A…A mismatch with respect to the flanking canonical base pairs as well as to retain the backbone connectivity [18, 38-40]. As seen in $d(CAG)_6.d(CAG)_6$ duplex [7], B-to-Z transition takes place through 'zipper mechanism' rather than 'stretch-collapse mechanism'. One can envisage similar situation in the case of $(GA)_n$ homoduplex, wherein, the nonisostericity between G...G and A...A may provoke parallel duplex formation [41].

Inclination of A...A mismatch towards Z-DNA leads to passive binding with $hZ\alpha_{ADAR1}$

The mechanism of recognition and binding of $hZ\alpha_{ADAR1}$ protein with B-Z junction/Z-DNA is still a matter of debate. As per the active mechanism, $hZ\alpha_{ADAR1}$ binds to B-DNA and subsequently converts it into Z-DNA [24]. Nevertheless, the passive mechanism suggests that $hZ\alpha_{ADAR1}$ traps the transient B-Z junction/Z-DNA and subsequently converts it into Z-DNA [42]. MD simulation (Figs. 1 and 3), CD (Fig. 6C), NMR (Fig. 7A) and microscale thermophoresis data presented here (Fig. 7B) conjointly identify that $hZ\alpha_{ADAR1}$ binds to $d(GAC)_7.d(GAC)_7$ in a 'passive mechanism' due to the formation of B-Z junction induced by A…A mismatch.

As discussed above, MD simulation clearly shows the preference for B-Z junction formation in $d(GAC)_6.d(GAC)_6$ duplex (Figs. 1B and 3A) in accordance with CD spectra of NaCl titration with $d(GAC)_7.d(GAC)_7$ duplex (Figs. 6A and 6B). While the mismatch containing duplex (former) has the proclivity towards Z-DNA transition (Fig. 6A), the canonical base pair containing duplex (later) doesn't possess such property (Fig. 6B). In line with this, titration of $hZ\alpha_{ADAR1}$ with $d(GAC)_7.d(GAC)_7$ converts the duplex completely to Z-form irrespective of the duplex length (Figs. 6C, supplemental Figs. S4 and S5). Although some studies have shown that GAC sequence is prone to form Z-DNA [10, 11], the rationale behind such conformational preference is unknown. For the first time it has been shown here that nonisosteric A...A mismatch provokes B-Z transition in GAC repeats that subsequently facilitates the binding with $hZ\alpha_{ADAR1}$ protein through 'passive mechanism'. This is further confirmed by 1D proton NMR spectroscopy that indicates tighter

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

affinity between the two (Fig. 7A). Additionally, K_D measured by MST also indicates that $d(GAC)_7.d(GAC)_7$ duplex bind with $hZ\alpha_{ADAR1}$ in nanomolar affinity (Fig. 7B). MD simulation carried out on the modeled $hZ\alpha_{ADAR1}$ dimer...$d(GAC)_6.d(GAC)_6$ complex (Figs. 7C, 7D and 8A) subsequently confirms that the protein residues interact with the duplex through the minor groove in accordance with the earlier studies (PDB ID: 2ACJ & 3IRQ).

Model for pathogenicity in $d(GAC)_n$ expansion disorders through RNA editing mediated by h_{ADAR1}

Current study clearly shows that $d(GAC)_7.d(GAC)_7$ duplex is not only prone to form Z-DNA, but also, binds to the Z-DNA binding domain of human ADAR1 protein. Intriguingly, expansion in d(GAC) trinucleotide repeat is shown to cause skeletal dysplasias such as multiple epiphyseal dysplasia and pseudoachondroplasia [3, 6]. Hence, based on the results from the current study we propose a model that explains how d(GAC) trinucleotide expansion in COMP gene may lead to skeletal dysplasia such as pseudoachondroplasia. As per our model (Fig. 9), $d(GAC)_7$ that can form a hairpin structure with the stem possessing B-Z junction (induced by A…A mismatches) facilitates the anchorage of Z-DNA binding domain of h_{ADAR1} onto the DNA during transcription. Succeeding to this event, the double stranded RNA specific deaminase domain of h_{ADAR1} performs A-to-I editing in GAC either in the corresponding nascent RNA duplex (Fig. 9A (top)) or in the downstream (Fig. 9A (bottom)). This eventually codes for 'Gly' instead of 'Asp' in the COMP. In fact, samples isolated from pseudoachondroplasia patients show that genomic point mutations in d(GAC) track of COMP gene that can code for Gly instead of Asp473/Asp482 are among the 70 possible mutations in the COMP gene [2, 43]. As per the current model, A-to-I editing can lead to such Asp to Gly mutation in the protein level during transcription and thus, can reflect the effect of genomic point mutations as mentioned above. Such A-to-I editing in the downstream of GAC repeat expansion (Asp482) can also take place (Fig. 9A (bottom)) resulting in Asp to Gly in COMP, which is already shown to be deleterious [44, 45]. On the other hand, when d(GAC) repeat is in normal number, hairpin formation may not take place. Thus, h_{ADAR1} may not be able to bind to the DNA duplex and A-to-I editing may not occur (Fig. 9B). Its noteworthy that although direct evidence for the role of h_{ADAR1} in pseudoachondroplasia is not well established, its hyper-/altered-editing in several neurodegenerative disorders has been well documented [13, 14, 16]. In line with this, the hypothesis presented here offers a new insight about the role of non-genetic A-to-I mutation in pseudoachondroplasia.

We have shown here that nonisomorphic nature of A...A mismatch with respect to the flanking base pairs is the underlying factor for Z-philic nature observed in $d(GAC)_7.d(GAC)_7$ repeat expansion that is found in pseudoachondroplasia. We have shown here for the first time that such a structural trait of A...A mismatch facilitates the binding of $hZ\alpha_{ADAR1}$ to $d(GAC)_{n=6,7}.d(GAC)_{n=6,7}$ duplex irrespective of the repeat length. A model for the complex of $hZ\alpha_{ADAR1}$...$d(GAC)_7.d(GAC)_7$ duplex and the consequent A-to-I editing during the transcription by double stranded RNA specific deaminase domain of h_{ADAR1} under the disease condition are also presented.

Experimental procedures

Molecular dynamics simulations

Starting models of $d(GAC)_6.d(GAC)_6$ DNA duplexes were manually modeled using Pymol suite (www.pymol.org, Schrödinger, LLC). Subsequently, the models were refined using constrained-restrained molecular geometry optimization using XPLOR-NIH [46]. MD simulations of the modeled duplexes (Fig. 1A) were carried out in explicit solvent environment following the protocol described earlier [7] using AMBER 12 suite [47]. FF99SB forcefield was used during the simulation. The systems were initially equilibrated for 50ps, following which, production runs were extended to 0.5μs for

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

each system using isobaric & isothermal conditions (NPT), 2fs integration time and 9Å cut-off distance for Lennard-Jones interaction. Following the above procedure, MD simulations of $d(GTC)_6.d(GTC)_6$ and $d(GAC)_6.d(GTC)_6$ duplexes have been carried out for 0.5μs timescale each. Note that 3D-NuS web server was used to built these models [48].

Analysis of the trajectories

Ptraj module of Amber 12 was used to post-process the trajectories corresponding to $d(GAC)_6.d(GAC)_6$, $d(GTC)_6.d(GTC)_6$ and $d(GAC)_6.d(GTC)_6$ simulations. Root mean square deviation (RMSD) was calculated to acquire quantitative information about the deviation or proximity of the trajectories from the initial structure. Backbone conformational angles and helical parameters were extracted from 3DNA [49] output using in house programs. Pymol (www.pymol.org, Schrödinger, LLC) & VMD [50] were used for visualization and MATLAB software (The MathWorks Inc., Natick, Massachusetts, United States) was used for plotting the graphs. Note that for the analysis, the central 14mer alone was considered.

Docking of $d(GAC)_6.d(GAC)_6$ DNA duplex with $hZ\alpha_{ADAR1}$ protein

Complex structure of $d(GAC)_6.d(GAC)_6$ DNA duplex and $hZ\alpha_{ADAR1}$ protein was manually modeled by replacing the duplex present in the crystal structure (PDB ID: 2ACJ) with our MD derived $d(GAC)_6.d(GAC)_6$ duplex. Subsequently, the complex model was subjected to 0.3μs MD simulations using *pmemd.cuda* module of AMBER 16 suite. Analysis was carried out by using cpptraj module of AMBER 16.

Duplex preparation

HPLC grade $d(GAC)_{n=6,7}$ and $d(GTC)_{n=6,7}$ oligonucleotides were purchased from Sigma-Aldrich. The oligonucleotides were dissolved in 50mM Tris-HCl and 50mM NaCl (pH 7.4). DNA duplex with canonical base pairs was formed by annealing $(GAC)_7$ and the complementary $(GTC)_7$ oligonucleotides at 95°C, and cooled them down to room temperature for 3 hours. On the other hand, only the former was considered for the formation of the duplex with A...A mismatch and the later was used for the formation of T...T mismatch. Subsequently, duplex formation was verified by acquiring CD spectrum (see below). Likewise, hairpin formation of $d((GAC)_3T_4(GAC)_3)$ was carried out. Its noteworthy that, to investigate salt-dependent behavior of the duplex, the above process was repeated in the presence of appropriate salt concentrations (0.05M, 0.5M, 1M, 2M, 3M, 4M, 4.2M). Base line correction was done using 50mM Tris-HCl along with the corresponding salt concentration (pH 7.4).

Sub-cloning of $hZ\alpha_{ADAR1}$ gene into pET21b expression vector

$hZ\alpha_{ADAR1}$ gene cloned in pMAT cloning vector was acquired from Invitrogen with Ndel and Sal1 restriction sites at 5'- and 3'-ends respectively. Subsequently, PCR amplified double digested $hZ\alpha_{ADAR1}$ gene was sub-cloned into ampicillin resistant pDZ1 expression vector, a modified form of pET-21a vector with T7 promoter [51-53]. The construct was organized in the following order: a N-terminal His_6-tag, GB1 solubility tag and tobacco etch virus (TEV) protease cleavage site that are followed by $hZ\alpha_{ADAR1}$ gene (225 bp).

Protein expression and purification

pDZ1 expression vector was transformed into *E.coli* BL21 (DE3) (Bioline) cells for the overexpression of hZαADAR1 protein. Pre-inoculum cells grown overnight were transferred into LB medium containing 100μg/ml ampicillin and incubated at 37°C until the optical density (OD) reaches to 0.6 at A_{600}. Protein expression was induced by 1mM isopropyl-thio-D-galactopyranoside (IPTG) followed by overnight incubation at 15°C to attain the OD at A_{600} in the range of 1.4 to 1.6. Cells were then harvested and sonicated in the binding buffer containing 20mM Tris-HCl, 500mM NaCl, 5mM imidazole (pH 8.0) and 0.1mM phenyl methane sulfonyl fluoride (PMSF). $HZ\alpha_{ADAR1}$

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

protein was eluted in a buffer containing 20mM Tris-HCl, 500mM NaCl & 200mM imidazole (pH 8.0) using Ni^{2+}-nitrilo tri acetate affinity column chromatography that was treated with 50mM $NiSO_4$ solution.

Purification involves two steps: At first, $hZ\alpha_{ADAR1}$ protein tagged with GB1 protein was purified as described above (supplemental Fig. S8A) followed by the removal of GB1-tag through overnight digestion with TEV protease. During the second round of purification, $hZ\alpha_{ADAR1}$ protein was isolated from the cleaved GB1-tag and the fractions were collected (supplemental Fig. S8B) in the binding buffer. Finally, the protein was dialyzed in NMR buffer (10mM phosphate buffer and 10mM NaCl, pH 7.4). Protein concentration was measured by UV absorption at 280nm using the extinction coefficient value of 8480 M^{-1} cm^{-1}.

$d((GAC)_n.(GAC)_n)_{n=6,7}$ duplex...$hZ\alpha_{ADAR1}$ complex formation

$d((GAC)_n.(GAC)_n)_{n=6,7}$…$hZ\alpha_{ADAR1}$ complex was prepared by changing the concentration of hZαADAR1 while retaining the concentration of DNA. For NMR experiments, the following protein(P)/nucleic acid (N) ratios were used by keeping the DNA concentration at 120μM: 0.25, 0.5, 0.75, 1. For CD experiments, P/N ratios of 0.25, 0.5, 0.75, 1, 1.5 & 2 were used by keeping the DNA concentration as 40μM. The samples were prepared in buffer containing 10mM sodium phosphate and 10mM NaCl (pH 7.4) and 10% of D_2O was added to the NMR sample. The complex was prepared by adding the protein to DNA sample in fractions of 10μl in 2 minutes intervals and incubated for 1 hour at 25°C.

CD spectroscopy

All the CD spectra reported here were acquired in JASCO-1500 and processed by spectra manager software. The data was collected in triplicate in the wavelength region of 320nm to 200nm and the baseline correction was done with respect to the appropriate buffer. Average of triplicate spectra was reported here.

NMR spectroscopy

1D proton NMR experiments were performed on a Bruker 700 MHz instrument equipped with a room temperature probe. Zggpw5 pulse sequence [54] is used to acquire the data at 25°C. All the acquisition parameters were kept identical for all the experiments: 768 scans and 32768 ^1H complex points. Bruker top spin was used for data processing and analysis.

Dissociation constant measurement using microscale thermophoresis

Dissociation constants of $hZ\alpha_{ADAR1}$ binding with $d(GAC)_7.d(GAC)_7$ duplex was estimated using microscale thermophoresis assay [55, 56]. The assay was carried out using $(His)_6$-GB1-$hZ\alpha_{ADAR1}$ tag protein. MST assay requires one fluorescent binding partner (protein) and one non-fluorescent binding partner (DNA). Prior to the titration, NT-647 fluorescent dye was non-covalently attached to the histidine residues of $hZ\alpha_{ADAR1}$. DNA was titrated to $hZ\alpha_{ADAR1}$ in serial dilutions with concentrations ranging from 0.313μM to 0.000153μM. Subsequently, the assay was carried out in 10mM phosphate buffer by keeping the concentration of labeled $hZ\alpha_{ADAR1}$ protein as a constant (15nM). These samples were loaded into monolith NT.115 MST premium coated capillaries and the MST analysis was performed using 100% LED power and 60% MST power in NanoTemper monolith NT.115 at 24°C. Using NanoTemper software, K_D was calculated using the mass action equation from triplicate experiments.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

Acknowledgments

The authors thank Prof. Roberto De Guzman, The University of Kansas for generously gifting pDZ1 construct. High Performance Computing facility of Indian Institute of Technology Hyderabad, Center for

Development of Advance Computing (Government of India) and Inter University Accelerator Center (Government of India) is acknowledged. The authors thank TIFR Mumbai and TCIS-Hyderabad for the help in preliminary NMR data collection and Drs. Saji Menon and Sivaramaiah Nallapeta for K_D measurement. The authors also thank Mr. Venkata Subbaiah, Ms. Komal Ghule and Ms. Kripi Tomar for some technical support.

Funding

The work was supported by Department of Biotechnology, Government of India (To TR): IYBA-2012 (D.O.No.BT/06/IYBA/2012), BIO-CaRE (SAN.No.102/IFD/SAN/1811/2013-2014), and R&D (SAN.No.102/IFD/SAN/3426/2013-2014) and Indian Institute of Technology Hyderabad (IITH) (To TR). Ministry of Human Resources Development, Government of India provided fellowships to NK & YA.

Conflict of interests

None

Supporting information

Attached

Author contributions

NK carried out molecular dynamics simulations & analyzed the data and also performed salt-dependent CD experiments. YA carried out sub-cloning, expression and purification of $hZ\alpha_{ADAR1}$ along with CD, NMR titration and K_D experiments of $hZ\alpha_{ADAR1}$…DNA complex followed by docking & MD simulations of $hZ\alpha_{ADAR1}$…DNA complex. TR designed and supervised the entire project. NK, YA & TR wrote the manuscript. †These authors contributed equally.

REFERENCES

1. Pearson, C.E., K. Nichol Edamura, and J.D. Cleary, *Repeat instability: mechanisms of dynamic mutations.* Nat Rev Genet, 2005. **6**(10): p. 729-42.
2. Briggs, M.D., et al., *Pseudoachondroplasia and multiple epiphyseal dysplasia due to mutations in the cartilage oligomeric matrix protein gene.* Nat Genet, 1995. **10**(3): p. 330-6.
3. Delot, E., et al., *Trinucleotide expansion mutations in the cartilage oligomeric matrix protein (COMP) gene.* Hum Mol Genet, 1999. **8**(1): p. 123-8.
4. Tufan, A.C., et al., *Serum or plasma cartilage oligomeric matrix protein concentration as a diagnostic marker in pseudoachondroplasia: differential diagnosis of a family.* Eur J Hum Genet, 2007. **15**(10): p. 1023-8.
5. Tan, K., et al., *The crystal structure of the signature domain of cartilage oligomeric matrix protein: implications for collagen, glycosaminoglycan and integrin binding.* FASEB J, 2009. **23**(8): p. 2490-501.
6. Mochmann, L.H. and R.D. Wells, *Transcription influences the types of deletion and expansion products in an orientation-dependent manner from GAC*GTC repeats.* Nucleic Acids Res, 2004. **32**(15): p. 4469-79.
7. Khan, N., N. Kolimi, and T. Rathinavelan, *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): p. e1004162.
8. Mitas, M., *Trinucleotide repeats associated with human disease.* Nucleic Acids Res, 1997. **25**(12): p. 2245-54.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

9. Sobczak, K., et al., *Structural diversity of triplet repeat RNAs.* J Biol Chem, 2010. **285**(17): p. 12755-64.
10. Kejnovska, I., M. Tumova, and M. Vorlickova, *(CGA)(4): parallel, anti-parallel, right-handed and left-handed homoduplexes of a trinucleotide repeat DNA.* Biochim Biophys Acta, 2001. **1527**(1-2): p. 73-80.
11. Vorlickova, M., et al., *Conformational properties of DNA fragments containing GAC trinucleotide repeats associated with skeletal displasias.* Eur Biophys J, 2001. **30**(3): p. 179-85.
12. Barraud, P. and F.H. Allain, *ADAR proteins: double-stranded RNA and Z-DNA binding domains.* Curr Top Microbiol Immunol, 2012. **353**: p. 35-60.
13. Meier, J.C., et al., *RNA Editing-Systemic Relevance and Clue to Disease Mechanisms?* Front Mol Neurosci, 2016. **9**: p. 124.
14. Singh, M., *Dysregulated A to I RNA editing and non-coding RNAs in neurodegeneration.* Front Genet, 2012. **3**: p. 326.
15. Nishikura, K., *A-to-I editing of coding and non-coding RNAs by ADARs.* Nat Rev Mol Cell Biol, 2016. **17**(2): p. 83-96.
16. Farajollahi, S. and S. Maas, *Molecular diversity through RNA editing: a balancing act.* Trends Genet, 2010. **26**(5): p. 221-30.
17. Yildirim, I., et al., *A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations.* J Am Chem Soc, 2013. **135**(9): p. 3528-38.
18. Thenmalarchelvi, R. and N. Yathindra, *New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex.* Nucleic Acids Res, 2005. **33**(1): p. 43-55.
19. Krepl, M., et al., *Reference simulations of noncanonical nucleic acids with different chi variants of the AMBER force field: quadruplex DNA, quadruplex RNA and Z-DNA.* J Chem Theory Comput, 2012. **8**(7): p. 2506-2520.
20. Dans, P.D., et al., *Unraveling the sequence-dependent polymorphic behavior of d(CpG) steps in B-DNA.* Nucleic Acids Res, 2014. **42**(18): p. 11304-20.
21. Ha, S.C., et al., *Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases.* Nature, 2005. **437**(7062): p. 1183-6.
22. Kiliszek, A., et al., *Atomic resolution structure of CAG RNA repeats: structural insights and implications for the trinucleotide repeat expansion diseases.* Nucleic Acids Res, 2010. **38**(22): p. 8370-6.
23. Ulrich, E.L., et al., *BioMagResBank.* Nucleic Acids Res, 2008. **36**(Database issue): p. D402-8.
24. Kang, Y.M., et al., *NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z alpha domain of human ADAR1.* J Am Chem Soc, 2009. **131**(32): p. 11485-91.
25. Wijmenga, S.S. and B.N.M. van Buuren, *The use of NMR methods for conformational studies of nucleic acids.* Progress in Nuclear Magnetic Resonance Spectroscopy, 1998. **32**(4): p. 287-387.
26. Jeong, M., et al., *NMR study of the Z-DNA binding mode and B-Z transition activity of the Zalpha domain of human ADAR1 when perturbed by mutation on the alpha3 helix and beta-hairpin.* Arch Biochem Biophys, 2014. **558**: p. 95-103.
27. Schade, M., et al., *The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA.* Proc Natl Acad Sci U S A, 1999. **96**(22): p. 12465-70.
28. Wang, A.H., et al., *Molecular structure of a left-handed double helical DNA fragment at atomic resolution.* Nature, 1979. **282**(5740): p. 680-6.
29. Pohl, F.M. and T.M. Jovin, *Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC).* J Mol Biol, 1972. **67**(3): p. 375-96.
30. Nordheim, A., et al., *Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding.* Cell, 1982. **31**(2 Pt 1): p. 309-18.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

31. Ha, S.C., et al., *A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA.* Proc Natl Acad Sci U S A, 2004. **101**(40): p. 14367-72.
32. Schwartz, T., et al., *Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins.* Nat Struct Biol, 2001. **8**(9): p. 761-5.
33. Kim, D., et al., *Distinct Z-DNA binding mode of a PKR-like protein kinase containing a Z-DNA binding domain (PKZ).* Nucleic Acids Res, 2014. **42**(9): p. 5937-48.
34. Moradi, M., et al., *Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study.* Nucleic Acids Res, 2013. **41**(1): p. 33-43.
35. Kim, D., et al., *Base extrusion is found at helical junctions between right- and left-handed forms of DNA and RNA.* Nucleic Acids Res, 2009. **37**(13): p. 4353-9.
36. Berger, I., et al., *Spectroscopic characterization of a DNA-binding domain, Z alpha, from the editing enzyme, dsRNA adenosine deaminase: evidence for left-handed Z-DNA in the Z alpha-DNA complex.* Biochemistry, 1998. **37**(38): p. 13313-21.
37. Kimura, T., et al., *One-electron attachment reaction of B- and Z-DNA modified by 8-bromo-2'-deoxyguanosine.* J Org Chem, 2004. **69**(4): p. 1169-73.
38. Rathinavelan, T. and N. Yathindra, *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**(5): p. 443-61.
39. Goldsmith, G., T. Rathinavelan, and N. Yathindra, *Correction: 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, 2016. **11**(5): p. e0155090.
40. Ananth, P., G. Goldsmith, and N. Yathindra, *An innate twist between Crick's wobble and Watson-Crick base pairs.* RNA, 2013. **19**(8): p. 1038-53.
41. Vorlickova, M., et al., *Dimerization of the guanine-adenine repeat strands of DNA.* Nucleic Acids Res, 1999. **27**(2): p. 581-6.
42. Bae, S., et al., *Intrinsic Z-DNA is stabilized by the conformational selection mechanism of Z-DNA-binding proteins.* J Am Chem Soc, 2011. **133**(4): p. 668-71.
43. Hecht, J.T., et al., *Mutations in exon 17B of cartilage oligomeric matrix protein (COMP) cause pseudoachondroplasia.* Nat Genet, 1995. **10**(3): p. 325-9.
44. Jackson, G.C., et al., *Pseudoachondroplasia and multiple epiphyseal dysplasia: a 7-year comprehensive analysis of the known disease genes identify novel and recurrent mutations and provides an accurate assessment of their relative contribution.* Hum Mutat, 2012. **33**(1): p. 144-57.
45. Susic, S., J. Ahier, and W.G. Cole, *Pseudoachondroplasia due to the substitution of the highly conserved Asp482 by Gly in the seventh calmodulin-like repeat of cartilage oligomeric matrix protein.* Hum Mutat, 1998. **Suppl 1**: p. S125-7.
46. Schwieters, C.D., et al., *The Xplor-NIH NMR molecular structure determination package.* J Magn Reson, 2003. **160**(1): p. 65-73.
47. Case, D.A., et al., *AMBER 12*. 2012, University of California, San Francisco.
48. Patro, L.P.P., et al., *3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-Dimensional Nucleic Acid Structures.* J Mol Biol, 2017. **429**(16): p. 2438-2448.
49. Lu, X.J. and W.K. Olson, *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures.* Nucleic Acids Res, 2003. **31**(17): p. 5108-21.
50. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
51. Wang, Y., et al., *Differences in the electrostatic surfaces of the type III secretion needle proteins PrgI, BsaL, and MxiH.* J Mol Biol, 2007. **371**(5): p. 1304-14.
52. Rathinavelan, T., C. Tang, and R.N. De Guzman, *Characterization of the interaction between the Salmonella type III secretion system tip protein SipD and the needle protein PrgI by paramagnetic relaxation enhancement.* J Biol Chem, 2011. **286**(6): p. 4922-30.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

53. Rathinavelan, T., et al., *NMR model of PrgI-SipD interaction and its implications in the needle-tip assembly of the Salmonella type III secretion system.* J Mol Biol, 2014. **426**(16): p. 2958-69.
54. Liu, M., et al., *Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy.* Journal of Magnetic Resonance, 1998. **132**(1): p. 125-129.
55. Parker, J.L. and S. Newstead, *Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1.* Nature, 2014. **507**(7490): p. 68-72.
56. Jerabek-Willemsen, M., et al., *Molecular interaction studies using microscale thermophoresis.* Assay Drug Dev Technol, 2011. **9**(4): p. 342-53.

FIGURES AND FIGURE LEGENDS

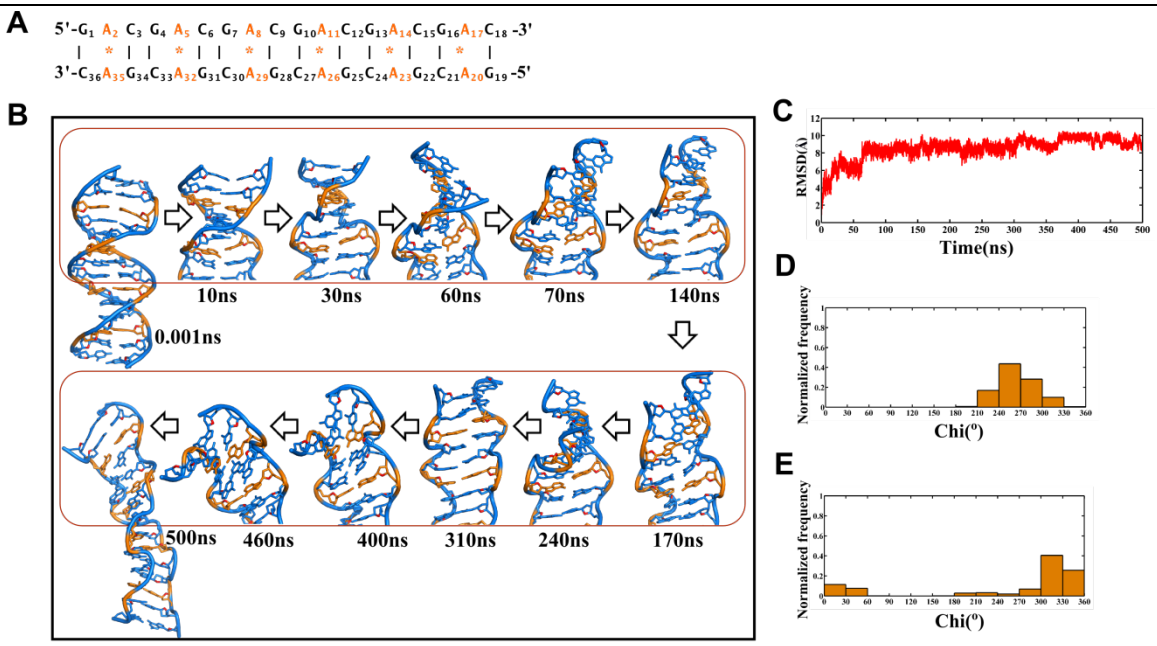

Figure 1. B-DNA to B-Z junction transition in d(GAC)$_6$.d(GAC)$_6$ duplex with *anti…anti* starting *glycosyl* conformation for A…A mismatch. (A) Sequence of 18mer DNA duplex that is subjected to MD simulation in the current investigation. (B) Cartoon representation of the duplex at various time intervals during the 500ns simulation which indicates the formation of left-handed conformation (shown in box). (C) Time vs RMSD profile showing significant conformational changes in the duplex as indicated by a high RMSD value with respect to the starting model. (D&E) *Glycosyl* torsion (*chi*) showing *high-anti* and *±syn* conformational preference for A's and G's respectively. Note that the noncanonical A...A mismatches are indicated in orange in (B).

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

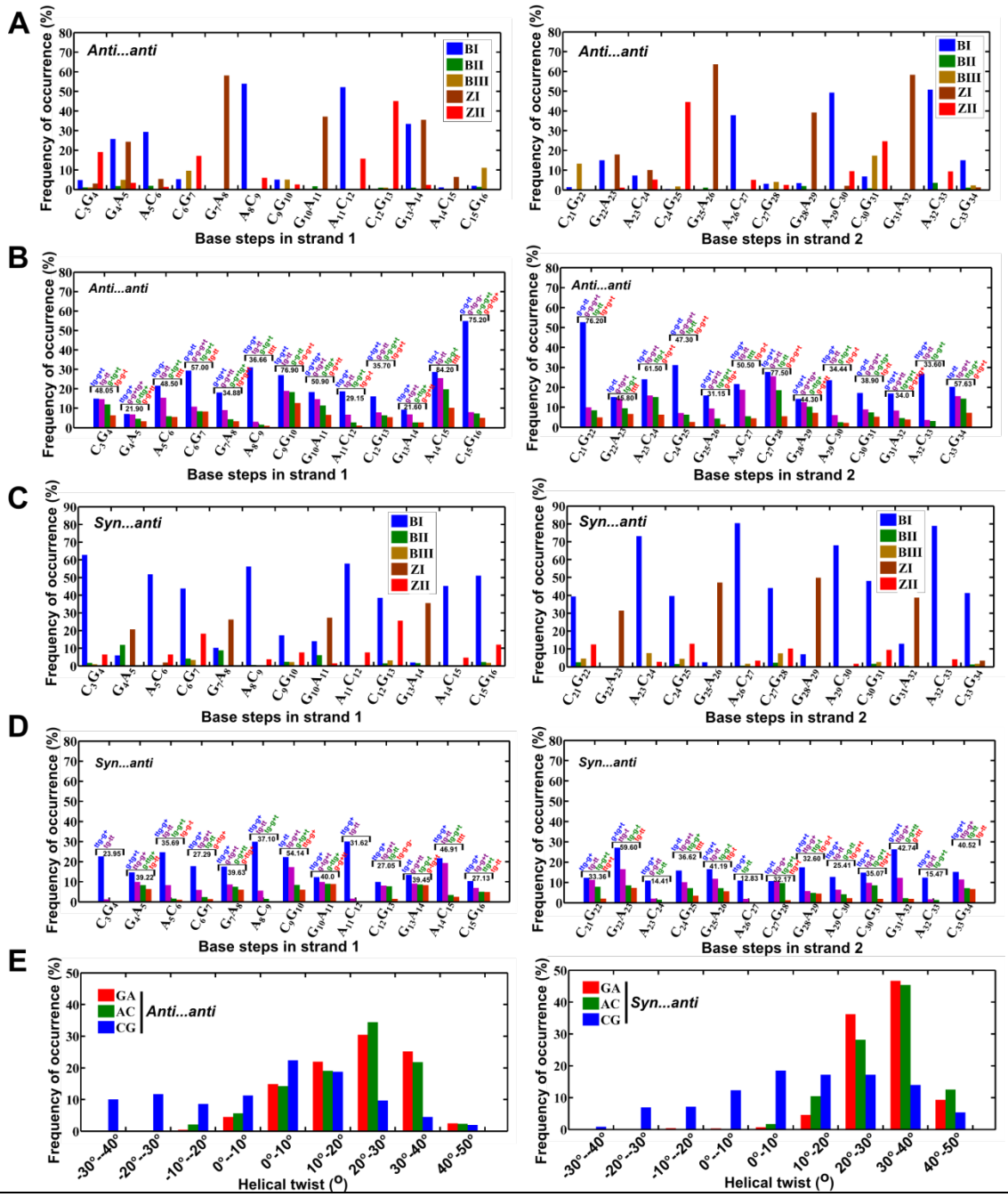

Figure 2. Parameters associated with B-Z junction formation during 101-500ns simulation time in the duplex containing A…A mismatches with *anti…anti* and *syn…anti* starting *glycosyl* conformations. (A&C) Frequency of occurrence of BI, BII, BIII, ZI & ZII conformations defined in terms of (ε,ζ,α,γ) at different steps of A…A mismatch containing duplex. Preponderance for BIII/Z conformations compared to BI/BII at GA/CG steps can be seen irrespective of the starting *glycosyl* conformation. (B&D) Frequency of occurrence of (ε,ζ,α,γ) conformations other than BI, BII, BIII and Z during 2 different MD simulations of $d(GAC)_6.d(GAC)_6$. Populations that occupy top 4 places are alone shown. Note that (ε,ζ,α,γ) conformational preference and the corresponding frequency of occurrence of top 4 populations (%) are given for each base step. (E) Histogram showing the distribution of helical twists at different steps of DNA duplex. Strikingly, CG steps exhibit the tendency

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

for lower twists (between -40° to 10°) compared to GA and AC steps. Percentage of such low twists is significantly higher in A…A mismatches compared to canonical duplex (Fig. 4C).

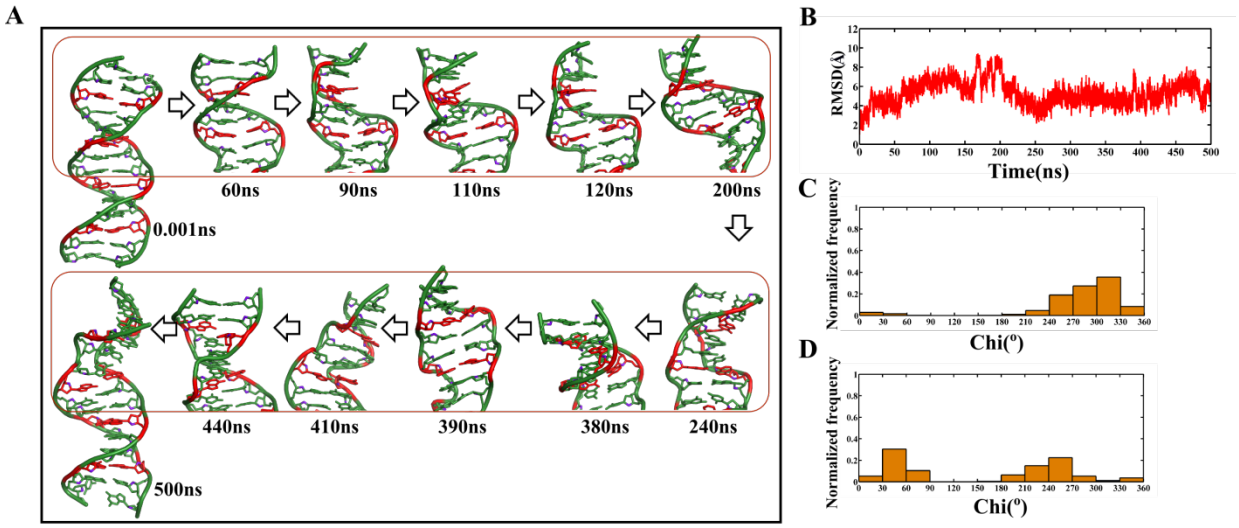

Figure 3. Evolution of B-Z junction in $d(GAC)_6.d(GAC)_6$ duplex with *+syn...anti glycosyl* conformation for the mismatch. (A) Snapshots showing the unwinding of helix due to Z-DNA evolution, resulting in negative supercoiling can be seen. A...A mismatch is colored red. (B) Time vs RMSD profile illustrating (RMSD > 4Å) significant conformational changes taking place with respect to the starting model. *Glycosyl* torsions showing (C) G's favoring ±*syn* and (D) A's favoring *high-anti/+syn* conformations respectively.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

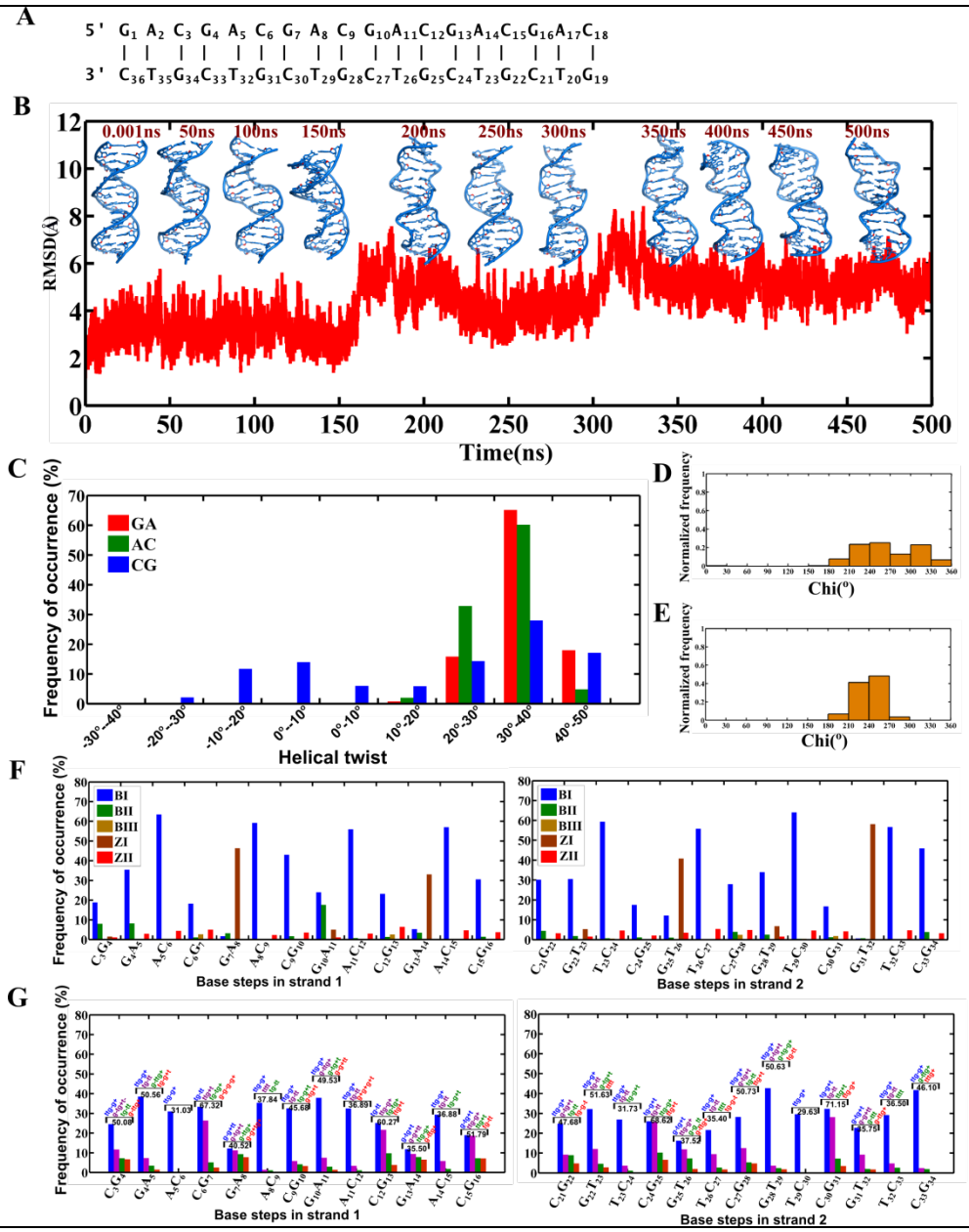

Figure 4. d(GAC)$_6$.d(GTC)$_6$ duplex containing canonical base pairs retains B-form geometry. (A) Sequence of 18mer DNA duplex containing canonical base pairs used for MD simulation. (B) Time vs RMSD profile along with cartoon representation of d(GAC)$_6$.d(GTC)$_6$ duplex at various time intervals are shown. (C) Histograms corresponding to helical twist values of GA, AC and CG steps over last 400ns. (D) *Glycosyl chi* angle values for guanine (favor ±*syn* conformation) and (E) adenine (favor *anti* conformation) residues. (F&G) Backbone conformational angles for canonical base pairs over last 400ns simulations. Note that (ε,ζ,α,γ) conformational preference and the corresponding frequency of occurrence of top 4 populations (%) other than BI, BII, BIII, ZI and ZII are given for each base step (G).

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

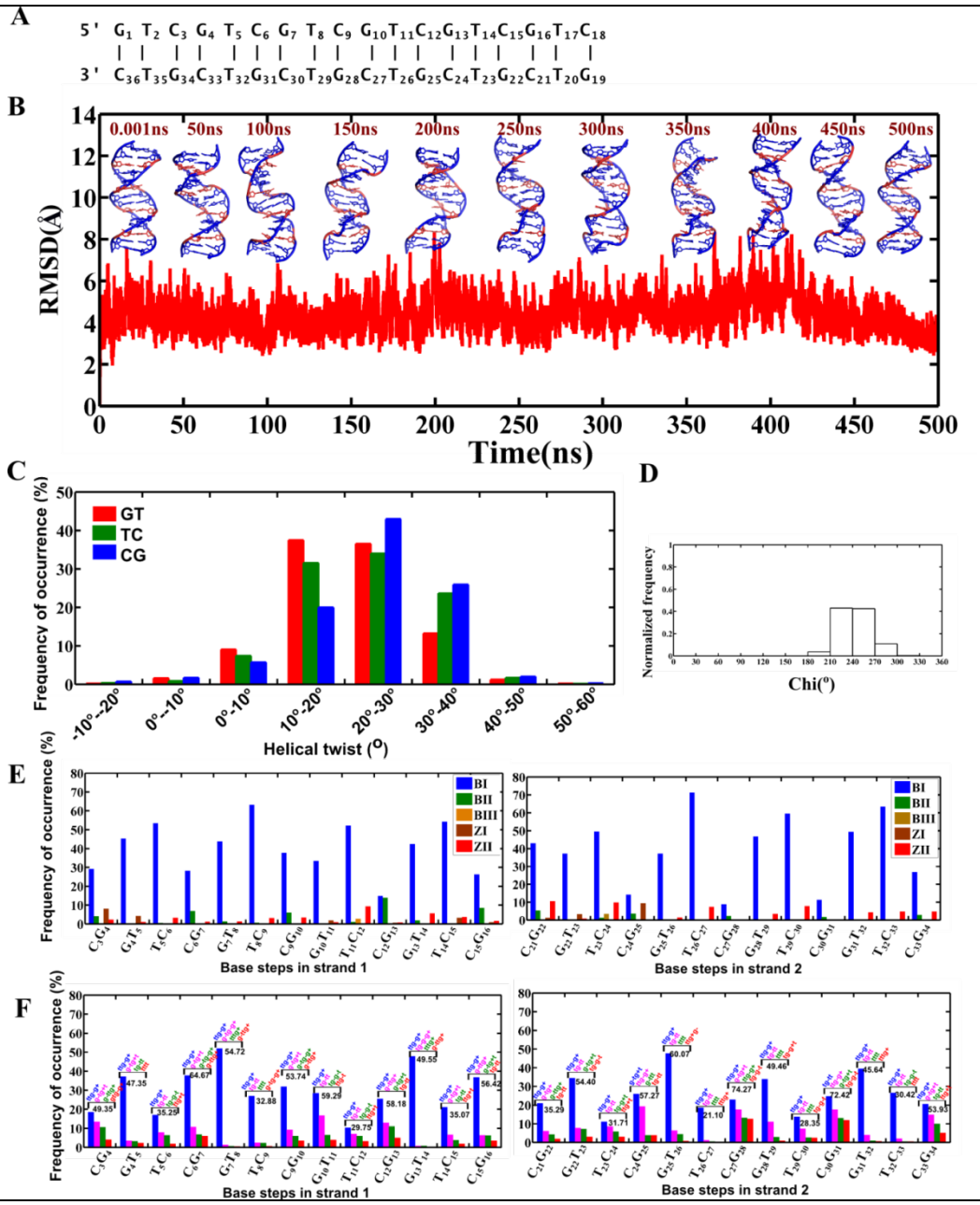

Figure 5. d(GTC)$_6$.d(GTC)$_6$ duplex containing T...T mismatch retains B-form geometry. (A) Sequence of 18mer DNA duplex containing T...T mismatch used for MD simulation. (B) Time vs RMSD profile along with cartoon representation of d(GTC)$_6$.d(GTC)$_6$ duplex at various time intervals are shown. (C) Histograms corresponding to helical twists at GT, TC and CG steps over last 400ns. (D) *Glycosyl chi* angle values for guanine residues (favor *anti/high-anti* conformation). (E) Backbone conformational (ε,ζ,α,γ) angles for canonical base pairs over last 400ns simulations. (F) Frequency of occurrence of top 4 populations other than BI, BII, BIII, ZI and ZII are given for each base step. Note: (GTC)$_6$.(GTC)$_6$ sequences were carried out using *pmemd.cuda* of AMBER 16 suite.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

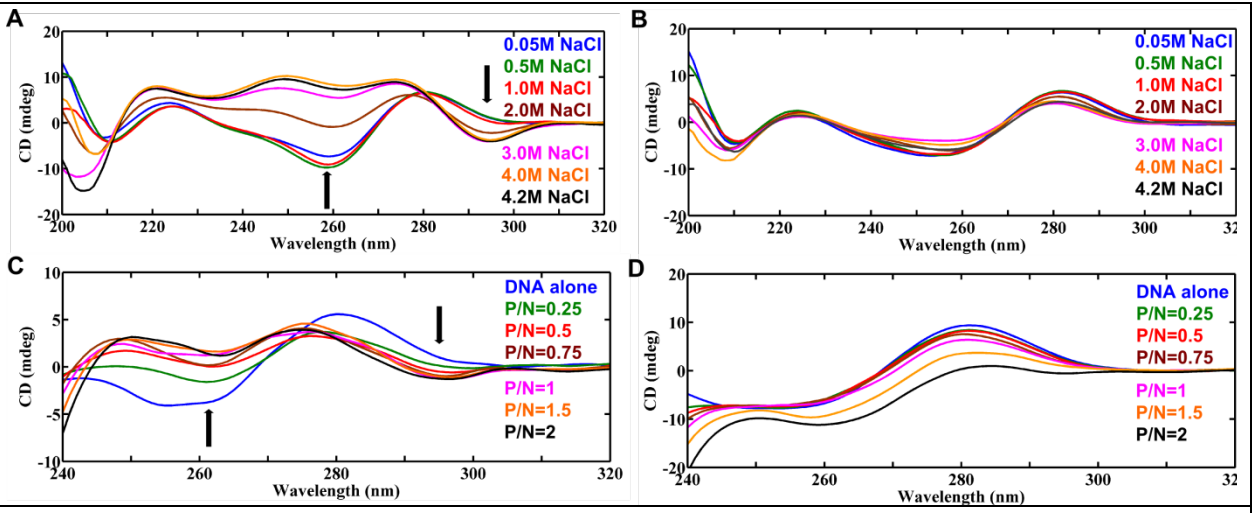

Figure 6. CD spectra showing the role of A…A mismatch in promoting Z-phillicity in $d(GAC)_7.d(GAC)_7$ duplex. Salt dependent (A) B-to-Z transition in $d(GAC)_7.d(GAC)_7$ (contains 7 A…A mismatches) and (B) absence of the same in $d(GAC)_7.d(GTC)_7$ (contains only canonical base pairs) can be seen. Titration of (C) $d(GAC)_7.d(GAC)_7$ and (D) $d(GAC)_7.d(GTC)_7$ with $hZ\alpha_{ADAR1}$ indicating complete B-Z to Z transition in the former and the absence of the same in the later.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

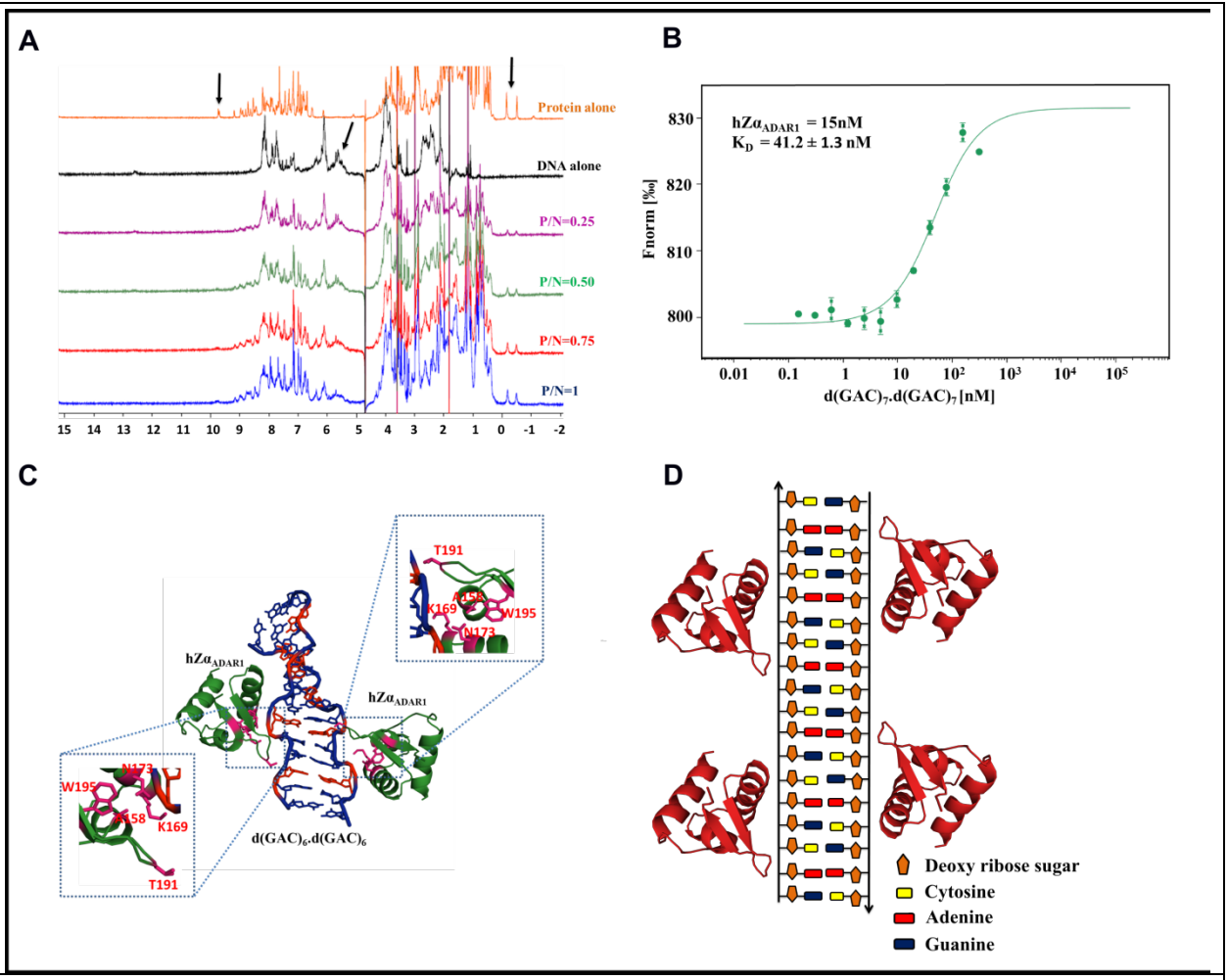

Figure 7. D(GAC)$_7$.d(GAC)$_7$…hZα$_{ADAR1}$ complex model. (A) ^{1}H NMR spectra corresponding to d(GAC)$_7$.d(GAC)$_7$ duplex titration with hZα$_{ADAR1}$. Arrows indicate the reduction in peak intensities (peak broadening) as the concentration of protein increases, suggestive of an intermediate chemical exchange between the two. (B) DNA concentration dependent (protein concentration is kept constant, while the DNA concentration is varied as described in experimental procedures) binding isotherms obtained from microscale thermophoresis assay indicate that d(GAC)$_7$.d(GAC)$_7$ and hZα$_{ADAR1}$ exhibit nanomolar binding affinity with a dissociation constant (K$_D$) of 41nM. (C) ^{1}H NMR based docked model of hZα$_{ADAR1}$ (PDB ID: 2ACJ)…d(GAC)$_6$.d(GAC)$_6$ (MD derived) complex (red color represents A...A mismatch). Note that the important interactions are zoomed and boxed. (D) Schematic representation of hZα$_{ADAR1}$ binding at multiple mismatch sites of d(GAC)$_6$.d(GAC)$_6$ duplex.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

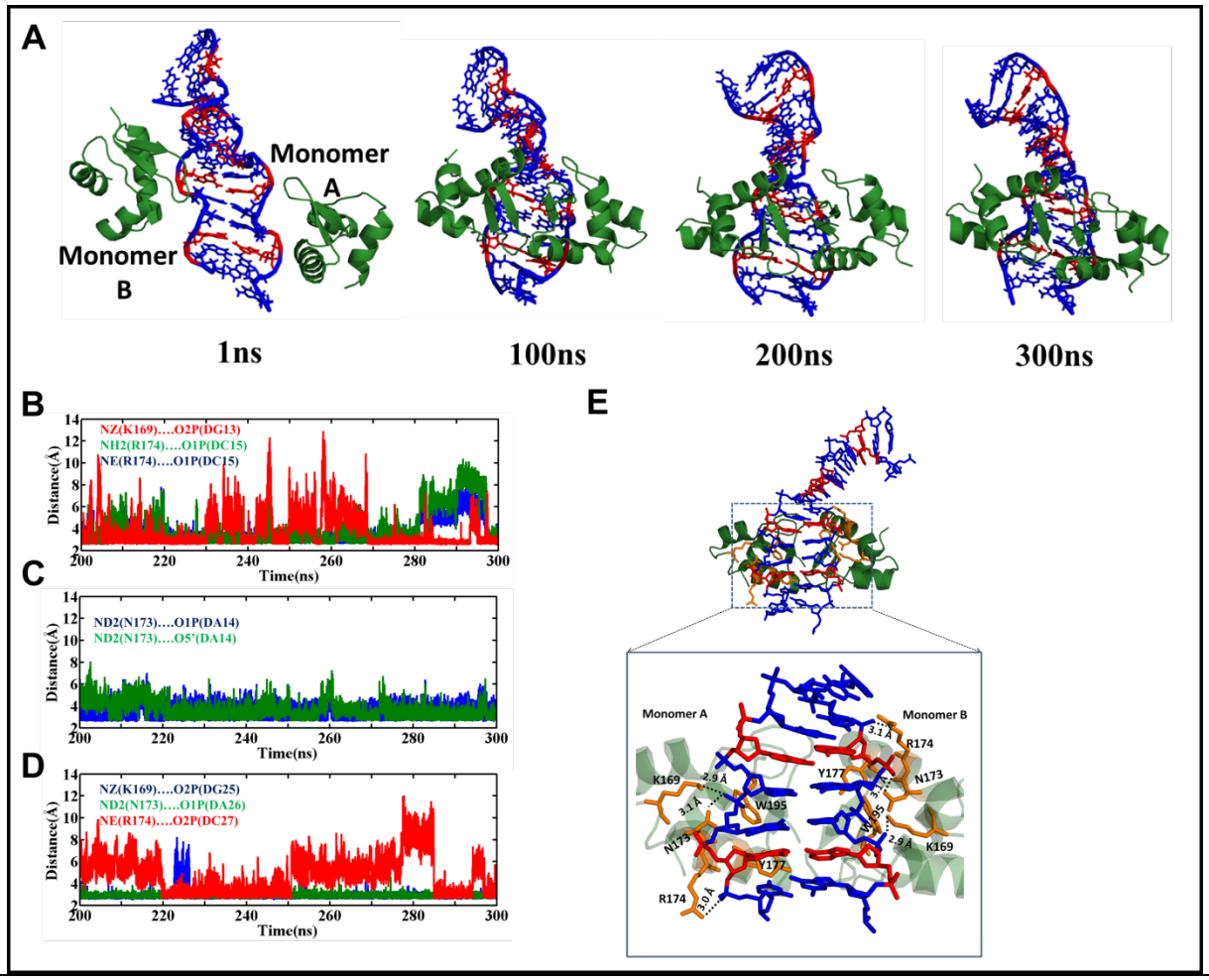

Figure 8. B-Z junction formation in $d(GAC)_6.d(GAC)_6$ duplex facilitates the accommodation of $hZ\alpha_{ADAR1}$ in the minor groove. A) Snapshots of $d(GAC)_6.d(GAC)_6$...$hZ\alpha_{ADAR1}$ complex MD simulation reveals that the β-hairpins of $hZ\alpha_{ADAR1}$ dimer interact with the duplex through its minor groove (A...A mismatches are colored red). (B-D) Time vs hydrogen bond distance profile corresponding to $hZ\alpha_{ADAR1}$ monomers A (B&C) and B (D). See text for details. E) Snapshot illustrating all the hydrogen bonding interactions between $d(GAC)_6.d(GAC)_6$ and $hZ\alpha_{ADAR1}$ during the simulation.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

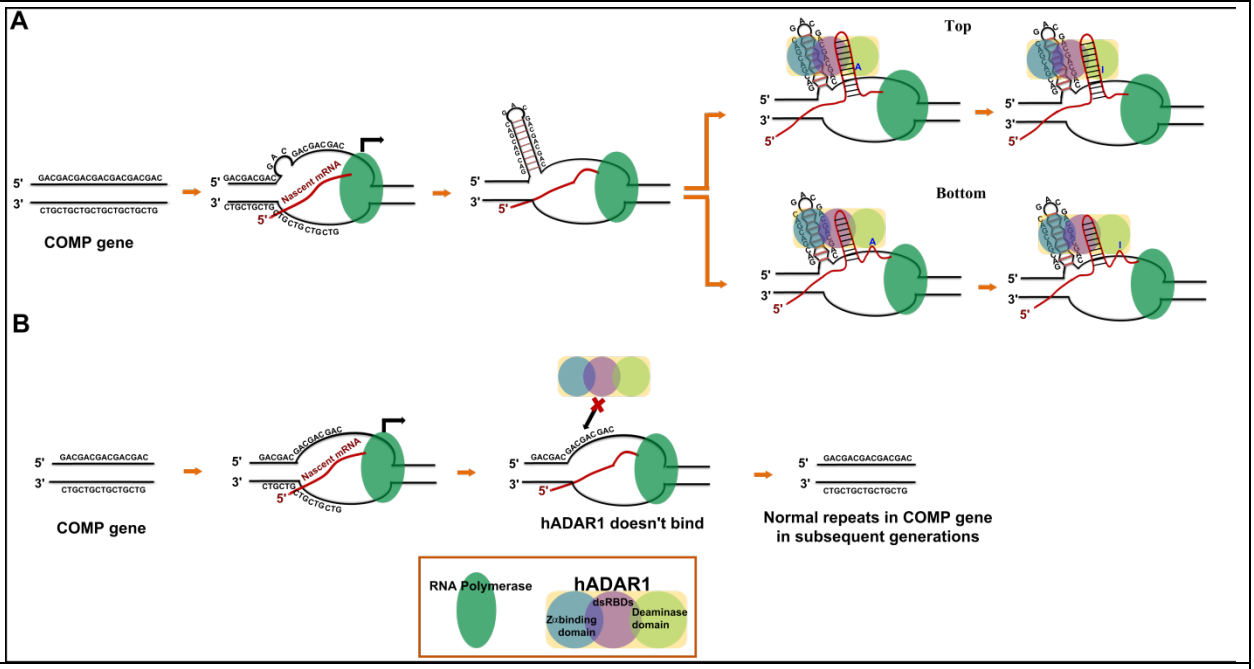

Figure 9. Proposed model for RNA editing by neuronal h_{ADAR1} in COMP in the perspective of pseudoachondroplasia disease. During transcription, (A) formation of d(GAC) hairpin that containing Z-conformation facilitates $hZ\alpha_{ADAR1}$ to anchor onto the hairpin stem and aids in A-to-I editing either (Top) in the corresponding nascent RNA or (Bottom) in the downstream. (B) Under normal condition, $hZ\alpha_{ADAR1}$ doesn't bind to the duplex due to the absence of B-Z/Z conformations, resulting in wild-type protein expression.

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017

A B−−Z junction induced by an A...A mismatch in GAC repeats in the gene for cartilage oligomeric matrix protein promotes binding with the hZ αADAR1 protein

Narendar Kolimi, Yogeeshwar Ajjugal and Thenmalarchelvi Rathinavelan

J. Biol. Chem. published online September 18, 2017

Access the most updated version of this article at doi: 10.1074/jbc.M117.796235

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2017/09/18/M117.796235.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2017/09/18/jbc.M117.796235.full.html#ref-list-1

Downloaded from http://www.jbc.org/ at INDIAN INSTITUTE OF TECHNOLOGY HYDERABAD on September 24, 2017