


RNAHelix: computational modeling of nucleic acid structures with Watson–Crick and non-canonical base pairs

Dhananjay Bhattacharyya¹  · Sukanya Halder¹ · Sankar Basu^{2,4} · Debasish Mukherjee¹ · Prasun Kumar³ · Manju Bansal³

Received: 19 September 2016 / Accepted: 29 December 2016
© Springer International Publishing Switzerland 2017

Abstract Comprehensive analyses of structural features of non-canonical base pairs within a nucleic acid double helix are limited by the availability of a small number of three dimensional structures. Therefore, a procedure for model building of double helices containing any given nucleotide sequence and base pairing information, either canonical or non-canonical, is seriously needed. Here we describe a program *RNAHelix*, which is an updated version of our widely used software, NUCGEN. The program can regenerate duplexes using the dinucleotide step and base pair orientation parameters for a given double helical DNA or RNA sequence with defined Watson–Crick or non-Watson–Crick base pairs. The original structure and the corresponding regenerated structure of double helices were found to be very close, as indicated by the small RMSD values between positions of the corresponding atoms. Structures of several usual and unusual double helices have been regenerated and compared with their original

structures in terms of base pair RMSD, torsion angles and electrostatic potentials and very high agreements have been noted. *RNAHelix* can also be used to generate a structure with a sequence completely different from an experimentally determined one or to introduce single to multiple mutation, but with the same set of parameters and hence can also be an important tool in homology modeling and study of mutation induced structural changes.

Keywords Molecular modeling · RNA · Non Watson–Crick base pairs · Base pair parameters · Dinucleotide step parameters · Electrostatic potential

Introduction

RNA, structures are comprised of several double helical stretches which are generally well ordered and flanked by unpaired regions [1]. Quite often, these unpaired nucleotides form either non-canonical base pairs (bps) or make contacts with more than one base [2, 3]. They even flip out of the helix for providing proper interactions in a complex or crystal lattice [4, 5]. Moreover, atoms of the bps within double helices undergo least amount of thermal fluctuations compared to the atoms of sugar or phosphate group as (1) these are located towards the center of the helix, (2) these are formed by two rigid planar moieties and (3) are stabilized by multiple hydrogen bonds. The canonical Watson–Crick (WC) interactions between two bases that form a bp are most abundant in majority of non-coding RNAs. However, significant number of non-Watson–Crick (NWC) bps have also been classified [6] and were found to be almost as stable as the Watson–Crick bps [7]. The overall double helical nucleic acid structure can be described by relative orientations of the two bases with respect to each other in

Electronic supplementary material The online version of this article (doi:10.1007/s10822-016-0007-0) contains supplementary material, which is available to authorized users.

✉ Dhananjay Bhattacharyya
dhananjay.bhattacharyya@saha.ac.in

✉ Manju Bansal
mb@mbu.iisc.ernet.in

¹ Computational Science Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India

² Computational Biophysics, IFM, Bioinformatics Division, University of Linköping, 581 83 Linköping, Sweden

³ Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

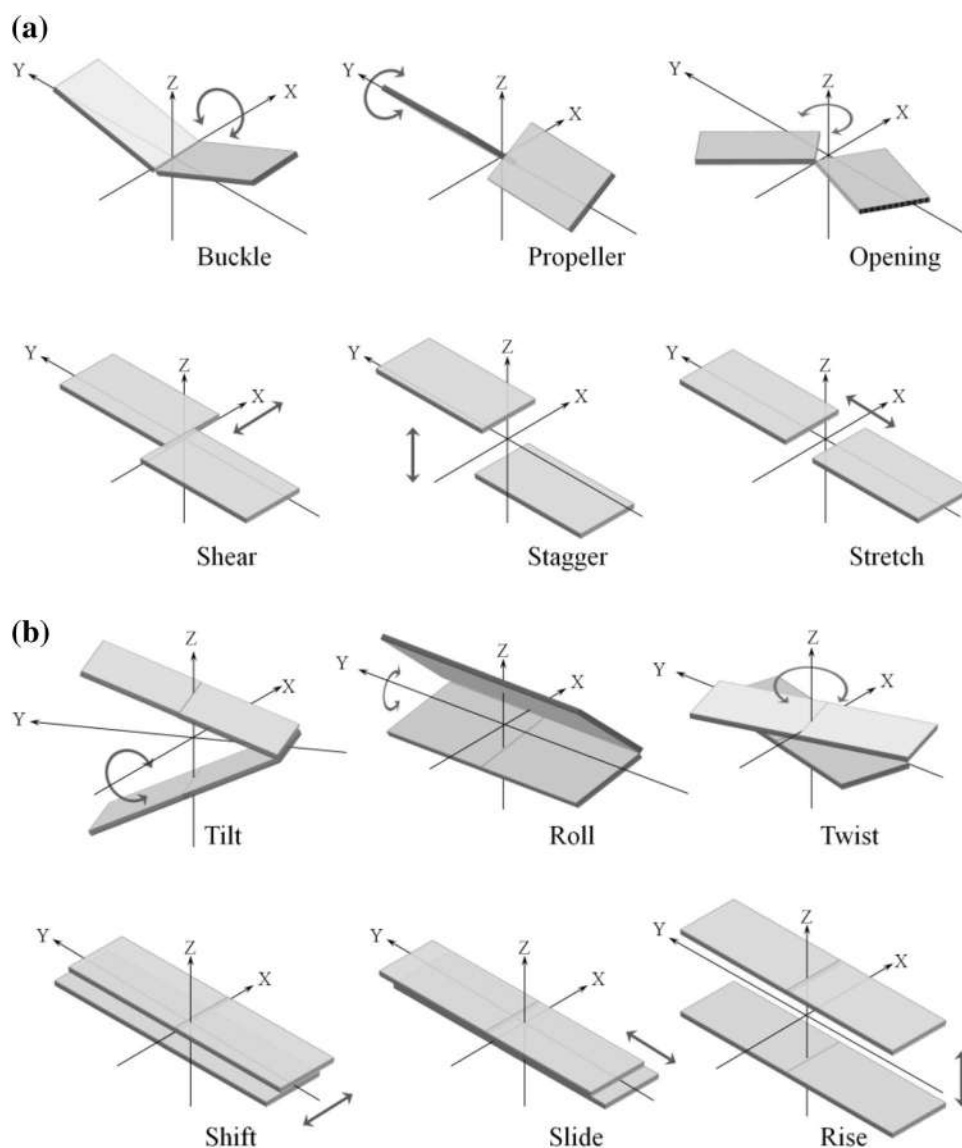
⁴ Present Address: Department of Biochemistry, University of Calcutta, Kolkata 700019, India

two strands along with those of the two stacked bps with respect to each other, which are defined by the IUPAC-IUB recommendations, in terms of three translational and three rotational parameters [8, 9]. It was suggested by IUPAC-IUB that six intra basepair parameters, suitable for visualization of geometry of a bp (Propeller, Buckle, Open, Stagger, Shear and Stretch) and six inter basepair parameters (Tilt, Roll, Twist, Shift, Slide and Rise) to describe geometry of a base paired dinucleotide step can be used (Fig. 1), Using these parameters, a nucleic acid double helical structure can be classified into various forms like A, B or Z. These parameters also reflect dinucleotide sequence dependent structural features [10–13], which are also sometimes influenced by the neighboring sequences [14–16]. Though several protein-bound nucleosome structures of ~140 nucleotides long DNA molecules have been solved [17–19], bent or curved free-DNA structures, with

varying base sequences are difficult to determine by experimental methods like X-ray crystallography or NMR. Similar sequence-directed bending has also been observed for RNA double helices, though the effect is less pronounced because of their shorter length. Furthermore, major variations in A-form RNA duplexes come from the inclusion of non-canonical bps at their terminal or central regions. In our previous studies, we have analyzed the structural features of RNA double helical stretches containing non-canonical bps at the central regions [20, 21], which indicated that these bps also possess some structural features similar to canonical ones. The availability of only small number of 3D coordinates for such molecules in PDB [22] constrained our studies to this limited dataset of double helices containing non-canonical bps.

The nucleic acid model building approach is an important aspect of molecular biology and can be applied to (1)

Fig. 1 Schematic representations of different types of orientational variations between a two bases of a bp, describing Shear (relative motion about X-axis), *Stagger* (relative motion about Z-axis) and Stretch (relative motion about Y-axis), *Buckle* (relative rotation through X-axis), Propeller (relative rotation through Y-axis), *Open* (relative rotation through Z-axis) and **b** two two bps of a dinucleotide step, describing *Tilt* (relative rotation through X-axis), Roll (relative rotation through Y-axis), *Twist* (relative rotation through Z-axis), *Shift* (relative translation about X-axis), *Slide* (relative translation about Y-axis) and *Rise* (relative translation about Z-axis). The bases (and basepairs) are shown as solid slab, translational motions are shown by straight bidirectional arrow and rotational motions are shown curved arrows. Both the axis systems shown here correspond to the axis system used for calculation of inter-bp parameters



identification and study of various structural properties of different sequences; (2) recognition of the sequences that can possibly interact with protein in protein-nucleic acid complexes and (3) computational analysis of stacking interactions in a multi-dimensional parameter space [23–25]. Such stacking interaction energy would be quite useful for predicting secondary structures of different small non-coding RNA, such as miRNA. It can also help in more extensive studies of sequence-dependent bending of nucleic acid structures and the influence of bp geometries on the overall three-dimensional structures of DNA/ RNA double helices.

Considering the above importance, several methods including 3DNA [26–28], NUCGEN [29] and Nucleic Acid Builder (NAB) [30] were developed to regenerate and analyze the nucleic acid structures. Among these, NAB uses a molecular manipulation language for representing nucleic acid structures in a hierarchical fashion, with principal focus on constructing models for both helical and non-helical nucleic acids. Fiber models of A-form and B-form DNA double helices can also be generated by the model building utility of AMBER suite and the Nucleic Acid Canonical Coordinates (NACC) web server (<http://nucleix.mbu.iisc.ernet.in/nacc/index.html>). There are few other servers for RNA structure prediction, such as MC-Sym [31] and RNAComposer [32], which predicts three-dimensional structure from sequence of the fragment based on homology modeling and free-energy minimizations. On the other hand, NUCGEN and 3D-DART [28] are capable of generating non-uniform curved DNA structures using the IUPAC-IUB recommended parameters. The program 3DNA uses the formulation implemented in SCHNARP [33] and NUCGEN uses a self-consistent formulation for structure generation of nucleic acid double helices containing canonical Watson–Crick bps [34–36]. It generates coordinates of only the base atoms along with the C1'-atoms, using the intra-bp and inter-bp (wedge) parameters as calculated by NUPARM [29]. The NUPARM software, also available now as web-server (<http://nucleix.mbu.iisc.ernet.in/nuparmplus>), calculates the IUPAC-IUB recommended parameters. However, none of the above algorithms are capable of (re)generating helical structures containing non-Watson–Crick (NWC) base pairs, viz. the bps using their Hoogsteen or sugar edges for hydrogen bonding or the bps in *trans* geometry. These NWC base pairs require special attention as the bp parameters, such as Propeller, Buckle, Shear, etc., calculated in the usual way, appear to be unrealistic. Furthermore, it is difficult to visualize the distortions of the bps from these parameter values. Only NUPARM-2 [36] gives physically meaningful values of these parameters (Buckle, Open, Propeller, Stagger, Shear and Stretch) in a self-consistent formalism. In this algorithm, the base fixed axis system is obtained from the atoms which are involved in hydrogen bonding to form the base pair. Hence

the bp parameter values of even a good *trans* base pair, for example, can be near zero. Such favorable base pairs are supposed to have two strong hydrogen bonds and the bases nearly coplanar.

We have developed an upgradation to the existing NUCGEN software and named it *RNAHelix*, for the generation of double helical structures with Watson–Crick as well as non-Watson–Crick bps. The base pair parameters calculated using base pairing edge specific axis system are used for regeneration, along with the information about the base pairing edges and orientation of the base pair (*trans* or *cis*). The original and regenerated model structures are compared in terms of RMS deviations between the base atoms. This, however, does not give the complete three-dimensional model of nucleic acid and hence the backbone atoms were generated followed by constrained minimization using CHARMM [37]. The electrostatic potential surfaces of the original and regenerated structures were also compared. We have observed excellent agreement between the original structure and the corresponding regenerated structures of double helices. With the help of few examples, we have shown the accuracy of our program and believe that *RNAHelix* will be very useful for structural biologists.

Methods

Composition of dataset

The performance of the *RNAHelix* was validated by considering both DNA and RNA consisting of Watson–Crick (WWC) and non-WWC bps. Systematic searches for the non-WWC bps were performed on Protein Data Bank (PDB) for both DNA and RNA and the structures were downloaded. We have used BPFIND [38] for basepair detection and in house perl script to select at least five basepair long helices with NWC bp within the helix. It may be noted that the BPFIND algorithm detects bps only when two hydrogen bonds (including C–H...O/N type) are possible between two bases, hence there is a chance that we do not find some well-documented bp. This search gave fifteen representative RNA fragments and eight representative DNA fragments with various characteristic features (Table 1). We have intentionally avoided those NWC bps which are formed by hydrogen bonds involving 2'-OH group of RNA sugar, as these bps mostly appear to be non-planar and distort significantly during geometry optimization using DFT [39]. Among the DNA structures, we have chosen B-DNA, A-DNA, two representative protein-DNA complexes with two different types of bending and a nucleosomal DNA structure. We have also analyzed a few DNA structures with non Watson–Crick bps. However, we have avoided G-quadruplex structure containing

Table 1 Descriptions of the regenerated double helices considered for *RNAHelix* validation

S. No	PDB	Reason for selection and unusual basepair type ^a	Residue numbers
RNA			
1	1J5A	RNA with U:U W:WC at center	2066–2072 (A) 2215–2209 (A)
2	1FJG	RNA with A:G W:WC at center	1409–1416 (A) 1491–1484 (A)
3	1N33	RNA with G:U W:WC at center	1421–1431 (A) 1479–1469 (A)
4	4V4Q	RNA with G:A S:HT and A:G H:ST bps at center	533–543 (B) 560–550 (B)
5	1N32-a	RNA with A:A s:hT, A:U H:WT and A:G H:ST bps at center	778–786 (A) 804–796 (A)
6	354D	RNA with G:A S:HT, A:U H:WT, U:A W:HT and A:G H:ST bps	71–80 (A) 105–96 (B)
7	1XMQ	RNA with A:U H:WT, A:G H:ST and G:A S:HT bps at terminii	439–448 (A) 495–486 (A)
8	1N32-b	RNA with A:A W:HT and C:A WH:T at one terminal	1241–1249 (A) 1296–1288 (A)
9	5J7L-bundle3	RNA with G:A S:HT, A:U H:WT and A:A H:WT bps at one terminal	150–161 (d) 176–165 (d)
10	4V9R-bundle4	RNA with G:A S:HT and A:A s:hT bps at one terminal	1198–1204 (A) 1247–1241 (A)
11	3R1C	RNA with G:G W:HC bps at center	1–8 (Q) 8–1 (R)
12	5DM6	RNA with C:A S:WC bp at terminal	3–13 (Y) 122–112 (Y)
13	4V88-bundle2	RNA with G:A S:HT, A:G H:ST and A:A H:HT bps	1645–1656 (A) 1810–1799 (A)
14	2L3C	RNA with G:G W:WT bp at center	1–15 (B) 34–20 (B)
15	2AZX	RNA with A:A W:SC bp at one terminal	510–514 (C) 525–521 (C)
DNA			
1	1BNA	Dickerson's dodecamer with Watson–Crick (W:W C) bps	1–12 (A) 24–13 (B)
2	1K61	Matalpha2 homeodomain bound DNA	2–21 (E) 42–23 (F)
3	1RSB	DNA with only A:T H:WC bps	1–6 (A) 12–7 (B)
4	1KX5	Nucleosomal DNA	–73–73 (I) –73–73 (J)
5	1QNC	TATA-box DNA complexed with TBP, with G:C H:+C bp at center	201–214 (E) 228–215 (F)
6	111D	DNA with A(Anti).G(Syn) bps A:G +:HC G:A H:+C	1–12 (A) 24–13 (B)
7	1DNM	DNA with A:G H:WC bps at center	1–12 (A) 24–13 (B)

Table 1 (continued)

S. No	PDB	Reason for selection and unusual basepair type ^a	Residue numbers
8	399D	A-DNA with Watson–Crick (W:W C) bps	1–12 (A) 24–13 (B)

Residue numbering is as given in the corresponding PDB

^aType of basepairing is indicated by three-letter code. The first letters specify edges of the two bases involved in hydrogen bonding. As for example, W:W, H:W, H:S represent basepair formation involving Watson–Crick edges of both the paired bases, Hoogsteen edge of the first base paired to Watson–Crick edge of the second base, Hoogsteen edge of the first base paired to Sugar edge of the second base, respectively. Smaller-case letters, “w”, “s” or “h” representing involvement of non-polar C–H...O/N type hydrogen bond in base pairing using Watson–Crick, Hoogsteen or Sugar edge, respectively. Similarly “+” represents requirement of protonation of N1 (of Adenine) or N3 (of Cytosine) imino nitrogen atom for formation of two hydrogen bonds. The last character (C or T) represents Cis or Trans (regular or reverse) orientation of the two bases with respect to each other considering hydrogen bond as the virtual middle bond of a torsion angle

Hoogsteen base pairs as it involves four strands of DNA and the segments are very small. The representative structures, downloaded from Protein Data Bank (PDB), are mentioned in Table 1. The base pairing information, as given in Table ST2, indicates capability of the program to regenerate wide variety of possible structures.

The regenerated structures were compared with other crystal structures having identical nucleotide sequence. A set of 5 such structures with identical sequences to each of the helices were selected from the classification database HD-RNAS [40] and BGSU [41] (Supplementary Table ST1).

Adding backbone to the regenerated structures

We have used the internal coordinates to Cartesian coordinate conversion (IC PARAM followed by IC BUILd) module of charmm [37] for adding sugar-phosphate backbone atoms to the regenerated structures. The IC values used for torsion angles α , β , γ , δ , ϵ , ζ , χ and sugar puckers are chosen to mimic the values observed in canonical A or B-form fiber-model helices. These topology files, along with the software, can be downloaded from <http://www.saha.ac.in/biop/bioinformatics.html>. The package (RNAHelix.tar.gz) contains the fortran-77 program, a linux executable, an input file containing ideal coordinates of the base atoms (DataSet.dat) [42] a perl-script (prm2parm.pl) to convert NUPARM-2 output to input (parameter.loc) for *RNAHelix*, the charm topology and parameter files and a charmm input script for constrained energy minimization along with README and sample output files. We minimized the regenerated structure using restrained energy minimization for 100 steps of steepest descent followed by 1000 steps of conjugate gradient and 20,000 steps of adopted basis Newton–Raphson methods using charmm. The nitrogen, carbon and oxygen atoms belonging to each base, including C1' atoms of the sugars were restrained to their *RNAHelix* generated coordinates by strong harmonic potential (CONSHARM FORCE 2000) at all stages of minimizations. The

energy minimized structures were superposed on the corresponding original ones for calculating the respective RMSD values.

Comparison of electrostatic potentials

In order to compute electrostatic potentials on the molecular surfaces of the poly-nucleotides, partial charges and atomic radii for all the atoms were assigned from the AMBER94 all-atom molecular-mechanics force field [43]. Detailed care was taken to discriminate between 5' or 3' terminal and internal residues in assignment of their partial charges so that every internal nucleotide along the negatively charged phosphodiester backbone eventually has a net charge of -1 whereas the 5'- and 3'-terminal residues have net charges of -0.31 and -0.69 , respectively, following the AMBER force-field design.

Molecular surfaces (MS) were generated by EDTsurf [44] which uses fast Euclidean Distance Transform (EDT) technique to generate triangulated mesh surfaces. The scale factor was set to 4.0 to generate around 24 triangles/Å² (nearly 10 dots/Å²). The calculation of surface electrostatic potential was done following our previous report [45] using finite-difference Poisson–Boltzmann method as implemented in DelPhi-v.7 [46, 47]. We used $\epsilon=2$ for interior, $\epsilon=80$ for surrounding solvent with 0.15 M ionic strength and 0.2 Å as ionic radii. All the calculations were performed at 298 K. The dielectric boundary and the partial charges were mapped onto a cubic grid of 201 × 201 × 201 in size. The percentage grid fill was set to 80% with a scale of 1.2 grid points/Å. Boundary potentials were approximated by the Debye–Huckel potential of the dipole equivalent to the molecular charge distribution. A probe radius of 1.4 Å was used to delineate the dielectric boundary. The Non-linear Poisson–Boltzmann equation (PBE) was then solved iteratively until convergence to 0.0001 kT/e.

The electrostatic focusing files, containing the surface potential values in units of kT/e (setting PHICON = f), were generated. The potentials were written in .cube format to

visualize the maps on Chimera [48]. In order to compare surface electrostatic potentials of the native and the regenerated structures, the nearest neighboring dot point to every dot surface point on the original (native) structure was detected on the surface of its regenerated partner within a distance cutoff of 2.0 Å subsequent to the superposition of the two structures. Surface potential values of these pairs of nearest neighboring dot surface points (from the two structures) were then stored as ‘target - nearest neighbor’ ordered pairs. The RMS deviation and the Pearson's Correlation Coefficients were then calculated (with an associated p-value for the second measure) between these pairs of potential values.

NUPARM and RNAHelix self-consistent algorithm

The geometry of two bases of a bp or two bps in a dinucleotide step is defined in terms of three translational and three rotational parameters. The three rotational parameters are similar to the three Eulerian angles along the moments of inertia, but the use of Eulerian rotation matrices for structure analysis and generation has some disadvantages—(1) in case of a Gimbal lock (when the second rotation is zero) one degree of freedom may be lost and (2) the order in which the rotations and translations should be applied remains ambiguous. Another common method of applying transformation to the rigid bodies is the use of quaternions. However, quaternions are, till date, applied only for rotations and the formulation is not well defined for translation. These parameters, moreover, fail to describe the orientation in a physically meaningful way. It may be noted also that the parameters calculated by the other similar software, namely Curves and 3DNA [26, 49], also are not physically meaningful for the non-canonical bps. Following recommendation of IUPAC-IUB, the intra-bp and inter-bp step parameters as defined in NUPARM are mathematically as well as physically meaningful.

The relative positioning of individual bps in a dinucleotide step are calculated by NUPARM as three rotational parameters (tilt, roll, and twist) and three translational parameters (shift, slide, and rise) along mutually perpendicular axis system (Fig. 1). According to IUPAC-IUB convention, these parameters are defined based on a local axis system situated on the imaginary mean plane between the bps. In an individual bp, the Y-axis is generally obtained as a vector connecting C8(R)/C6(Y) atom of the base in the second (descending) strand (bp2) to C6(Y)/C8(R) atom of the base of the first (ascending) strand (bp1), while the bp centre is defined as the mid-point of these two atoms. This axis, at least in case of WC base pairing, passes approximately through their centre of gravity and is along the longest dimension. There is another option in NUPARM to calculate Y-axis as a vector passing through

the C1' atoms of the two paired bases. This axis is nearly parallel to the default Y-axis passing through C8 and C6 atoms of the bases for Watson–Crick bps. However, direction of this vector is quite different from the other for non Watson–Crick bps. The Z-axis is determined as an average vector perpendicular to the bp plane along the direction of the first strand and the X-axis is perpendicular to the others ($\bar{X} = \bar{Y} \times \bar{Z}$). It may be noted that the above procedure ensures X-axis direction towards major groove, defined by the nearby WC bps, even for consecutive non-canonical trans bps (Fig. 2). The local axis system of a dinucleotide step is defined as the mean of the two sets of axes (Eq. 1), obtained from the two bps constituting the step and the origin is located midway between them. This local axis system of each bp step is chosen in such a manner that the same numerical values are obtained when going from bp1 to bp2, as from bp2 to bp1, with only a possible change of sign for some of the parameters. The values of the inter-bp or local wedge parameters are given by Eq. 1:

$$\bar{X}_m = \frac{\bar{x}_1 + \bar{x}_2}{|\bar{x}_1 + \bar{x}_2|}, \quad \bar{Y}_m = \frac{\bar{y}_1 + \bar{y}_2}{|\bar{y}_1 + \bar{y}_2|}, \quad \bar{Z}_m = \bar{X}_m \times \bar{Y}_m \quad (1a)$$

$$Tilt(\tau) = -\sin^{-1}(\bar{Z}_m \cdot \bar{y}_1) \quad (1b)$$

$$Roll(\rho) = \sin^{-1}(\bar{Z}_m \cdot x_1) \quad (1c)$$

$$Twist(\omega) = \cos^{-1}[(x_1 \times \bar{Z}_m) \cdot (x_2 \times \bar{Z}_m)] \quad (1d)$$

$$Shift(Dx) = \bar{M} \cdot \bar{X}_m \quad (1e)$$

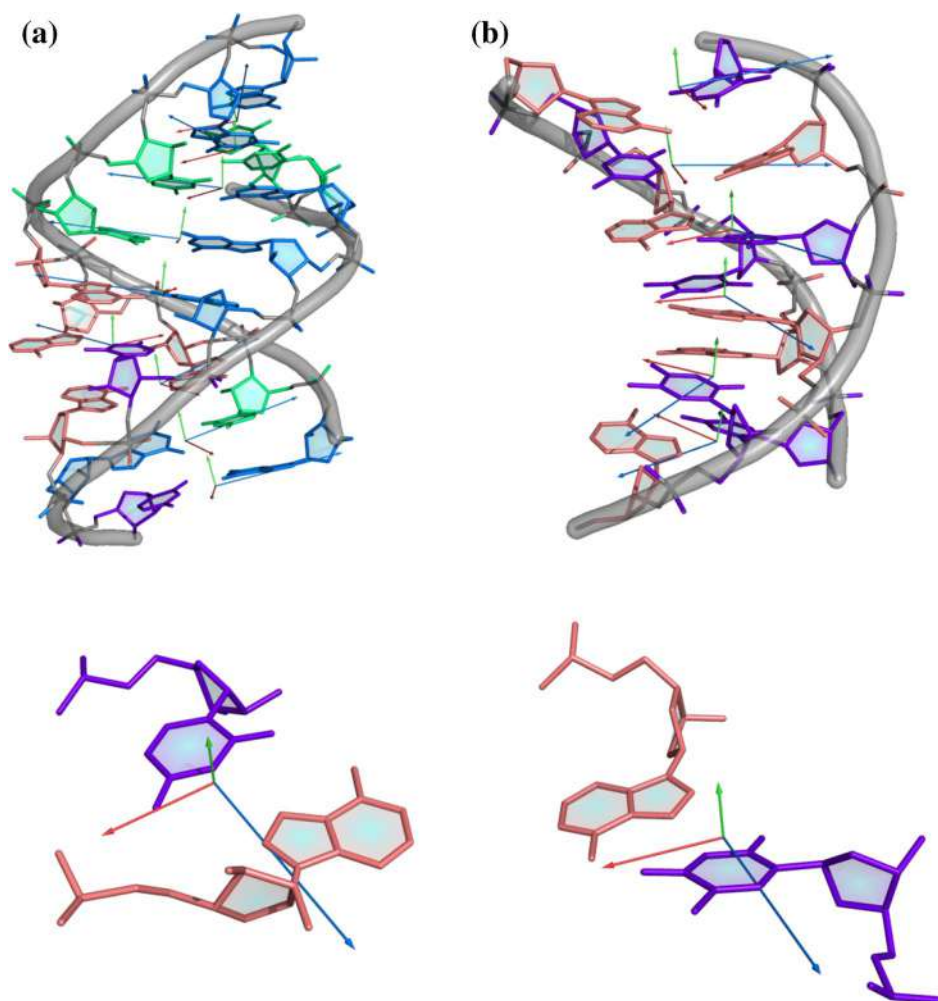
$$Slide(Dy) = \bar{M} \cdot \bar{Y}_m \quad (1f)$$

$$Rise(Dz) = \bar{M} \cdot \bar{Z}_m \quad (1g)$$

and (\bar{M}) is the vector joining the centers of two consecutive bps.

Similarly for a bp, the relative orientation of the two bases are represented by three rotational, namely Buckle (κ), Open (σ) and Propeller (π) and three translational, namely Stagger (Sx), Shear (Sy) and Stretch (Sz) (Fig. 1) parameters (Eq. 2). Calculation of these parameters again requires definition of axis system of the two paired bases, which is quite hassle-free for the Watson–Crick bps, due to their iso-stericity. In NUPARM-2 and RNAHelix we tried to follow similar standard for non-canonical bps also. Considering hydrogen bonds in A:U or G:C Watson–Crick bps, the Y-axis of each base is defined using N1 → N6/O6 atoms of purines and N3 → O4/N4 atoms of pyrimidines and Watson–Crick basepairs are represented as W:WC. The reverse

Fig. 2 Representative double helices **a** 1N32-a with non-canonical bps in trans orientation and **b** 1RSB with only A:T H:W C bps are shown along with the axis system used for calculation of inter-bp parameters. A trans Hoogsteen bp (A:U H:W T) from 1N32-a and one of the constituent A:T H:WC bps are also shown to highlight the weird natures. The colors used for the bases, *red*, *blue*, *green* and *purple* are representing Adenine, Guanine, Cytosine and Uracil respectively



Watson–Crick basepairs (W:WT) are stabilized by similar hydrogen bonds, hence the axis system for the bases are also similar. This axis is directed towards the major groove for Watson–Crick base paired helices but is ill-defined for non-canonical or trans basepairs. In case of bps involving Hoogsteen edge (H:X C or T, X being any edge), we define the Y-axis considering N7(R)/C6(Y) \rightarrow N6/O6(R)/O4/N4(Y) atoms and for sugar edge (S:X C or T, X being any edge) we define the Y-axis through C1' \rightarrow N3(R)/O2(Y) atoms. The X-axes of the bases are the base normals obtained by least-squares fit. An ambiguity remains regarding the results of such least-squares fit—whether the normals are along or opposite to the strand direction. This and the ambiguity of Y-axis are both fixed simultaneously after calculating the Z-axis, which must be directed from C1' of the second strand to C1' of the first strand. Thus, even the structure of DNA double helix containing all A:T W:HC basepairing or RNA with trans basepairs give meaningful parameters and axes directions (Fig. 2).

Thus to describe or to generate a dinucleotide step, a set of 18 parameters (6 each of the two bps and 6 for the bp

step) are required. Stretch or separation of the bases from each other was described by NUCGEN [29] as C8–C6 distance and opening angle was described in terms of two glycosidic angles. Such choice was sufficient to generate double helices with only Watson–Crick base pairs, although it does not follow the IUPAC-IUB standard. In the *RNAHelix*, we have employed all the 18 parameters to overcome the constraint of regenerating non-canonical geometries.

$$\bar{X}_m = \frac{\bar{x}_1 + \bar{x}_2}{|\bar{x}_1 + \bar{x}_2|}, \quad \bar{Y}_m = \frac{\bar{y}_1 + \bar{y}_2}{|\bar{y}_1 + \bar{y}_2|}, \quad \bar{Z}_m = \bar{X}_m \times \bar{Y}_m \quad (2a)$$

$$Buckle(\kappa) = -2\sin^{-1}(\bar{Z}_m \cdot \bar{x}_1) \quad (2b)$$

$$Open(\sigma) = -2\sin^{-1}(\bar{Z}_m \cdot \bar{y}_1) \quad (2c)$$

$$Propeller\ twist(\pi) = \cos^{-1}\left[(x_1 \times \bar{Z}_m) \cdot (x_2 \times \bar{Z}_m)\right] \quad (2d)$$

$$\text{Stagger } (S_x) = \bar{M} \cdot \bar{X}_m \tag{2e}$$

$$\text{Shear } (S_y) = \bar{M} \cdot \bar{Y}_m \tag{2f}$$

$$\text{Stretch } (S_z) = \bar{M} \cdot \bar{Z}_m \tag{2g}$$

and the vector \bar{M} is obtained by joining the two base atoms, one from each base of the pair, chosen according to the hydrogen bonding edge of the particular base.

Helical transformation

The relative orientation of two bps in three-dimensional space could also be described with respect to an external cylindrical coordinate system, where the laboratory-frame Z-axis is considered to be the helix axis [29, 34]. A set of parameters (tip, inclination, helical twist, X-displacement, Y-displacement and helical rise) was defined with respect to this helix axis [50], which are identical for two bps constituting the local helix. Finally, its relative orientation with the adjacent bp is described by the helical twist and rise about the helical Z-axis. It was also observed that the inter-bp or local wedge parameters (Tilt, Roll, etc) and the local helical parameters are analytically related to each other (Eqs. 3–5) and one can obtain the helical parameters from the local doublet parameters and vice versa [35]. The local doublet parameters are descriptive, so that one can understand kink etc, from these values. These parameters have been found important to understand base sequence effect on DNA or RNA double helical structures. One can also predict DNA curvature from vector summation of the inter-bp or wedge parameters [51],

of parameters was derived [29, 34], and was implemented in NUCGEN [29] for generation of curved DNA structures.

In *RNAHelix*, we have extended the application of the self-consistent formulation to generate canonical and non-canonical bps. The intra-bp parameters, such as Propeller, Buckle, Open, etc., calculated using a mean axis system by NUPARM-2, are converted into helical sense and applied accordingly. In analogy with dinucleotide steps, the bases are rotated about and translated along a set of helical bp axis system, where the body frame Z-axis lies along the base pair frame Y-axis (or long axis) of external coordinate system. The relation between intra-bp parameters (Propeller, Buckle, etc.) in the mean axis system and those in helical sense ($\pi_h, \kappa_h, \sigma_h, S_{hx}, S_{hy}, S_{hz}$ representing Propeller, Buckle, Open, Stagger, Shear, and Stretch equivalents in helical sense, respectively) are given by (Eq. 3). These are similar to tip, inclination, helical twist, X-displacement, Y-displacement and helical rise for a bp step. It is important to mention here that these equations are extremely accurate for numerical evaluation of small angles. However, numerical evaluation of σ_h and κ_h using inverse sin function (ASIN in FORTRAN) around 90° or -90° can be error-prone, even in double precision code, due to the inherent nature of inverse sin function. The error in calculating helical parameters would increase with increase in the absolute values of the local parameters and hence the generated structures may deviate more from the initial starting structure.

$$\pi_h = \sin^{-1} \left[\sin^2 \left(\frac{\pi}{2} \right) \cos^2 \left\{ \sin^{-1} (R + T)^{1/2} + R + T \right\} \right]^{1/2} \tag{3a}$$

$$\kappa_h = \sin^{-1} \left[\left(\frac{1}{2T} \right) \cot^2 \left(\frac{\pi_h}{2} \right) \left\{ (1 - R - T) - \sqrt{(1 - R - T)^2 - 4RT} \right\} \right]^{1/2} \tag{3b}$$

$$\sigma_h = \sin^{-1} \left[\left(\frac{1}{2R} \right) \cot^2 \left(\frac{\pi_h}{2} \right) \left\{ (1 - R - T) - \sqrt{(1 - R - T)^2 - 4RT} \right\} \right]^{1/2} \tag{3c}$$

while the local helical parameters are better for model building. Thus, to generate a model structure, the local step parameters can be converted to the helical frame (Eqs. 3–5) and then these helical parameters can be applied to the bp steps, such that the final structure of the dinucleotide steps have the applied values of local step or wedge parameters. In this formulation, the values of tip, inclination, X-displacement and Y-displacement are calculated from the local doublet parameters and applied to both the bps in each step. Finally, helical Twist and helical Rise are applied, in a negative sense, to all the previously generated bps, so that the last generated bp is always situated on the X–Y plane, with the bp Y-axis (or long axis) lying along the positive Y-axis of the external coordinate system. The analytical relation between these two sets

$$S_{hx} = 2B_1 \cos(\kappa_h) \sin(\pi_h) + 2B_3 \sin(\kappa_h) \tag{3d}$$

$$S_{hy} = -2B_2 \cos(-\sigma_h) \sin(\pi_h) + 2B_3 \sin(-\sigma_h) \tag{3e}$$

$$S_{hz} = -2B_1 \cos(-\sigma_h) \sin(\kappa_h) \sin(\pi_h) + 2B_2 \cos(\kappa_h) \sin(-\sigma_h) \sin(\pi_h) + B_3 [1 + \cos(\pi_h)] \cos(\kappa_h) \cos(-\sigma_h) \tag{3f}$$

$$R = \sin^2 \left(\frac{\sigma}{2} \right); \quad T = \sin^2 \left(\frac{\kappa}{2} \right) \tag{4}$$

$$B_1 = S_y \sqrt{2[1 + \cos(\pi_h) + \{1 - \cos(\pi_h)\sin^2(-\sigma_h)\}]} \quad (5a)$$

$$B_2 = S_x \sqrt{2[1 + \cos(\pi_h) + \{1 - \cos(\pi_h)\sin^2(\kappa_h)\}]} \quad (5b)$$

$$B_3 = S_z [1 + \cos(\pi_h) + \{1 - \cos(\pi_h)\sin^2(-\sigma_h)\}] [1 + \cos(\pi_h) + \{1 - \cos(\pi_h)\sin^2(\kappa_h)\}] \quad (5c)$$

In accordance with NUPARM formalism of intra-bp parameters, we have employed the edge-specific axis system for the regeneration of nucleotide structure. Thus, in a bp, the Y-axis is directed towards the major groove sides of the bases, unambiguously defined by edge-specific hydrogen bonding atom positions, whereas that for a bp step is directed from strand II to strand I and the X-axis is directed towards major groove which leads to small values of the bp parameters for even a good Hoogsteen or a trans bp.

Results

Comparison of structures and backbone generation

The *RNAHelix* software can regenerate any right handed double helical structure of DNA or RNA containing canonical Watson–Crick or non-canonical basepairs using the given set of intra-bp and inter-bp parameters. The intra-bp parameters, according to design, are supposed to have small values for a good planar bp with two strong hydrogen bonds between the bases. Our previous analysis indicated definite trends of the values of these parameters for all the frequently occurring basepairs [39]. The values of Propeller, Buckle and Stagger are generally small for all such bp (Supplementary Table ST3) although the Propeller values are generally negative (around -10°). The Stretch values are mostly around 3Å similar to hydrogen bonding distance between non-hydrogen atoms. The Shear and Open values, however, depend on base pairing, particularly orientations of hydrogen bonding atoms. For example, Shear values are around 2 Å for U:G W:WC, A:G H:ST, A:A H:HT, etc. bps. The Open angles are also around 13° for A:G H:ST, G:A S:WT and few other bps. Open angles of some of the basepairs, such as A:G w:sC, A:U w:sT, etc. are, however, too large and not indicative of planarity. These base pairs are stabilized by hydrogen bonds involving 2'-OH group of one of the sugars and adopted quite different hydrogen bonding geometry after geometry optimization by DFT or other quantum chemical methods [39]. Considering these discrepancies we are not confident about the quality of regeneration of structures of double helices containing such a base pair, which is stabilized by hydrogen bond involving

a sugar-OH group. One can generate a viable structure of a double helix, containing any of the listed non-canonical basepair using their mean parameters. The values of expected inter-bp parameters for all the Watson–Crick bp containing steps are also available in literature [52]. However, values of twist, for example, for non-canonical bp containing steps can be quite different from those found for steps with canonical bps. As for example, twist and slide values for the 3rd and 4th steps of 111D system are quite different from typical values in B-DNA (36° and ~ 0 Å). The deviations are even larger in 1RSB, 1st and 2nd steps of 4V88, 5th and 6th steps of 4V9R, etc. with twist values as small as 2° or as large as 98° (Supplementary Table ST2). Hence an approximate prior knowledge about stacking geometry is expected to get a feasible model structure. Furthermore, generation of a double helix alone is not sufficient to test our algorithm and a comparison with an original structure is mandatory. We have identified all five base pair or longer double helices from DNA and RNA databases [22] containing at least one non-canonical basepair. There are many examples of double helical region of RNA where the terminal bp is non-canonical and some examples where some middle bps are non-canonical. We have considered representatives of those, as shown in Table 1.

The *RNAHelix* software generates coordinates of atoms of each base, along with their C1' atoms, using the input intra-bp and inter-bp dinucleotide step parameters. We have used the bp and bp-step parameters calculated by NUPARM-2 (using C8(R)/C6(Y)–C6(Y)/C8(R) as bp long axis and C8...C6 midpoint as bp center for inter-bp parameters) for the structures listed in Table 1 to regenerate double helical structures of several DNA and RNA and have compared the bp and bp-step parameters of the original crystal structures with those of the regenerated structures (Supplementary Table ST2). Near identical values of intra-bp and inter-bp parameters of the initial and regenerated double helices, suggest identity of overall curvature of the helices and the relative disposition of the bases in space. The values of intra-bp and inter-bp parameters of some of the bp-steps are quite unusual. As for example, twist values of the second and third steps of 1N32 are 11.4° and 92.8° , respectively. Such unusual values appear due to involvement of A:A s:h T and A:U H:W T basepairs in those steps. The regenerated structures also have similar unusual values (Supplementary Table ST2). Such unusual base pairs are known to possess large Open (near 90°), large Shear and Stretch values, if calculated by 3DNA or Curves [11, 26]. The intra-bp parameters of these bps, calculated by NUPARM-2 and used for regeneration by *RNAHelix* are, however, quite small. Such small values of Propeller, Buckle and Stagger are indicative of coplanar orientation of the two bases, paired involving their sugar or Hoogsteen edges in trans orientation. Similarly small values of Open,

Shear and Stretch are indicative of good hydrogen bonds between the two bases. Some of the intra-bp parameters of 354D (original or regenerated) are quite high—Open values ca 50° for the 4th, 5th and 6th steps, Shear values around 4 Å for the 4th and 6th steps. It may be noted that these bps were not detected by our base pair detection program BPFIND [38], as two hydrogen bonds are not possible between these bases. Nevertheless *RNAHelix* could regenerate such unusual structure with high accuracy as the parameters of the two structures (original and regenerated) have similar parameter values.

We have also compared the two sets of models (original and regenerated) by superposing one on another using CHARMM and have calculated their root mean square deviations (RMSD). All the RMSD values between the base atoms (4th column of Table 2) are generally less than 0.2 Å. Few large molecules, such as the nucleosomal DNA (1KX5) show larger RMSD. These values also indicate that

the method is not only applicable to the regeneration of the structures with only WC bps, but can also be used to regenerate the helices with various NWC bps with equally higher accuracy. Original and regenerated structures after superposing the base atoms for representative cases are shown in Fig. 3. In order to confirm that the RMSD values are really small, we have used Non-Redundant lists of HD-RNAS [40] and BGSU [41, 53] databases to obtain structures similar to each of the selected models. We found there are at least five structures in PDB, such as 1FJG, 1N33, 1IBM, 1HNW and 1HNZ, all similar to 1N32 (Supplementary Table ST1). We obtained the same fragment as given in Table 1 from all these structures and calculated RMSD between 1N32 and 1FJG, 1N32 and 1N33, 1N32 and 1IBM, etc. Averages of these RMSD values for the base atoms are shown in 5th column of Table 2. It is seen that the RMSD values between original and regenerated structures are always lower than those between different structures of same class (except for 4V88) confirming that the regenerated structures closely reflect the properties of original structure and differentiate between similar ones.

Structures consisting of only bases are not sufficient to understand biochemical role of such double helix. Hence backbone atoms were added and optimized using CHARMM considering base atoms as precursors for Internal Coordinates to Cartesian Coordinates conversion. The backbone generation and restrained optimization using charmm27 force-field by CHARMM produced well connected sugar-phosphate atoms of all the bases of all the regenerated structures. The torsion angles of the original and regenerated structures are listed in Supplementary Table ST2. The torsion angles of individual residues of the original and regenerated structures are generally very similar to each other. Most of the sugars adopt C3'-endo puckering for the regenerated RNA structures, as characterized by δ (delta) torsion angles. Similarly most of the sugars adopt C2'-endo puckering for the DNA structures. Although we started constrained minimization of the TATA-box DNA sequence (1QNC) with all sugars in C2'-endo conformation, as initial geometries, the final optimized sugars adopt C3'-endo like geometry in the central TATA-box region, similar to the original structure. Exceptions are often seen near the non-canonical bps for sugar-pucker or α , γ , or ϵ torsion angles. For example, α (alpha) torsion angle of 9th residue of the regenerated 111D is 112.5° while the same is -43.7° in the original. It should be noted that the strain of accommodating non-canonical bp in the helix was adjusted by adopting unusual γ (gamma) torsion angle of the same residue (4.2°) in the crystal structure. Similar adjustments of backbone by alternate variations in torsion angles of nearby residues were found in many cases, e.g. 8th and 16th residues of 1DNM, 23rd residue of 4V88, 11th and

Table 2 RMSD values between coordinates of original and regenerated structure and the selected and their similar crystal structures

S. No.	PDB ID	RMSD (Å) between original and regenerated structure		Average RMSD (Å) with similar crystal structures
		All-atom	Base-atom	Base-atom
RNA				
1	1J5A	1.87	0.11	0.44
2	1FJG	1.83	0.11	0.30
3	1N33	1.84	0.17	0.39
4	4V4Q	1.87	0.17	0.38
5	1N32-a	1.65	0.10	0.22
6	354D	1.91	0.18	0.22
7	1XMQ	1.74	0.10	0.33
8	1N32-b	1.93	0.13	0.33
9	5J7L-bundle3	1.89	0.08	0.24
10	4V9R-bundle4	2.12	0.19	0.35
11	3R1C	1.85	0.06	— ^a
12	5DM6	1.69	0.07	0.64
13	4V88-bundle2	1.81	0.24	0.22
14	2L3C	2.09	0.33	0.81
15	2AZX	1.49	0.12	— ^a
DNA				
1	1BNA	1.13	0.11	0.32
2	1K61	1.21	0.08	— ^a
3	1RSB	1.38	0.17	0.49
4	1KX5	1.46	0.84	— ^a
5	1QNC	1.43	0.32	— ^a
6	111D	1.18	0.06	— ^a
7	1DNM	1.53	0.06	— ^a
8	399D	1.09	0.07	— ^a

^aWe did not find any structure in PDB or NDB with sequence similar to these

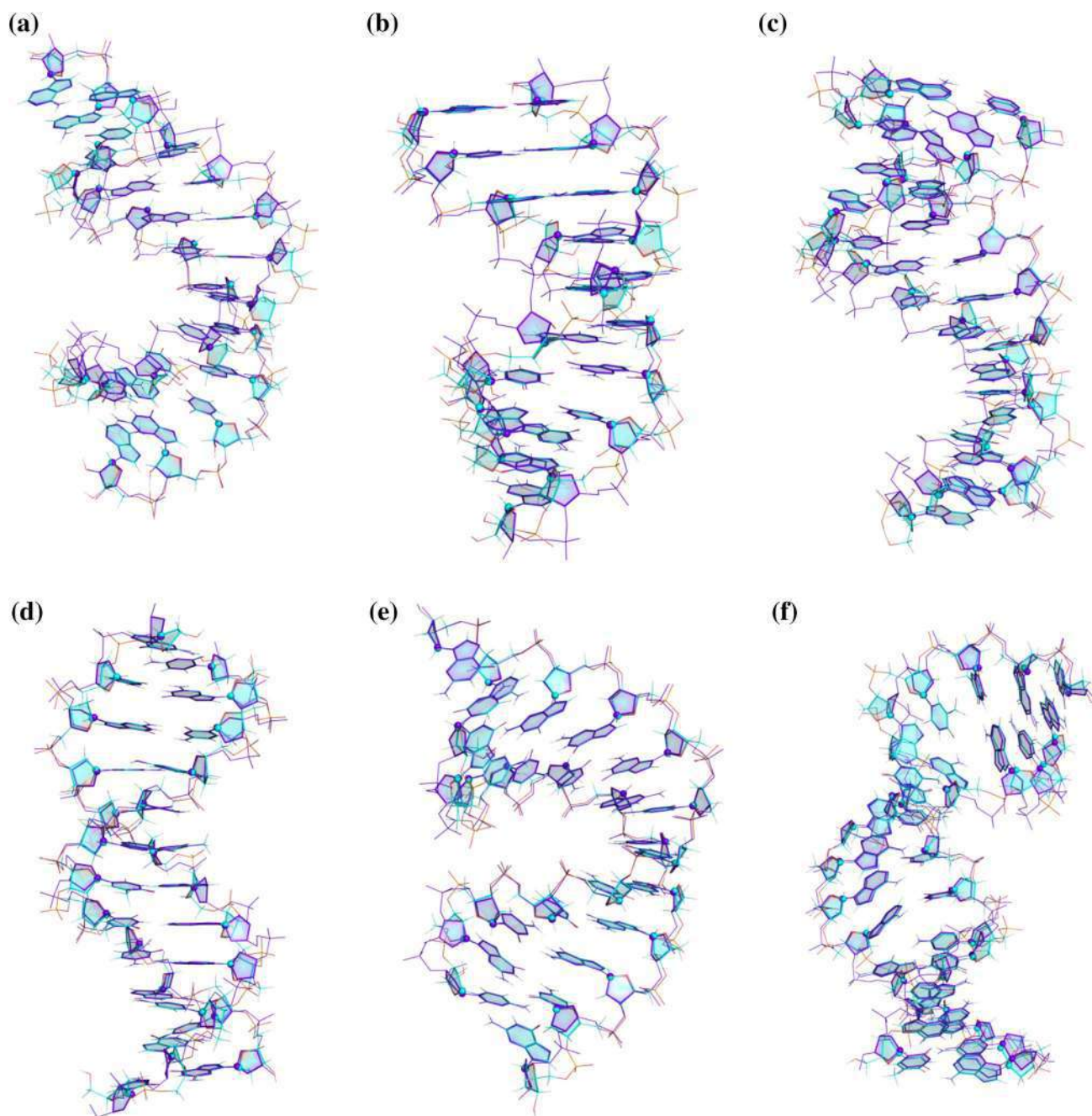


Fig. 3 Superposition of regenerated structures of representative systems, **a** 5J7L, **b** 354D, **c** 4V88, **d** 1BNA, **e** 399D and **f** 1QNC on top of the original ones are shown here. The original structures are drawn by *red color* while the regenerated structures are drawn by *CPK color* scheme

14th residue of 3R1C, etc. Such preference of *trans* values of α torsion angle, may be a feature of the force-field, needs to be addressed by detailed molecular dynamics simulations. The terminal regions of DNA or RNA double helices may have unusual geometry due to fraying effect; hence we feel mismatch between torsion angles at the terminals may not be that significant. Furthermore, we used energy minimization only to obtain reasonable preliminary structures, which can be further improved by

performing restrained molecular dynamics simulations. Nevertheless, frequency distributions of all the torsion angles of the two sets of structures indicate very similar nature, as shown in Fig. 4. All-atom RMSD between the initial crystal and regenerated model structures are shown in (Table 2) though these values are somewhat dependent on the use of appropriate force-field and a point that is still being debated [54].

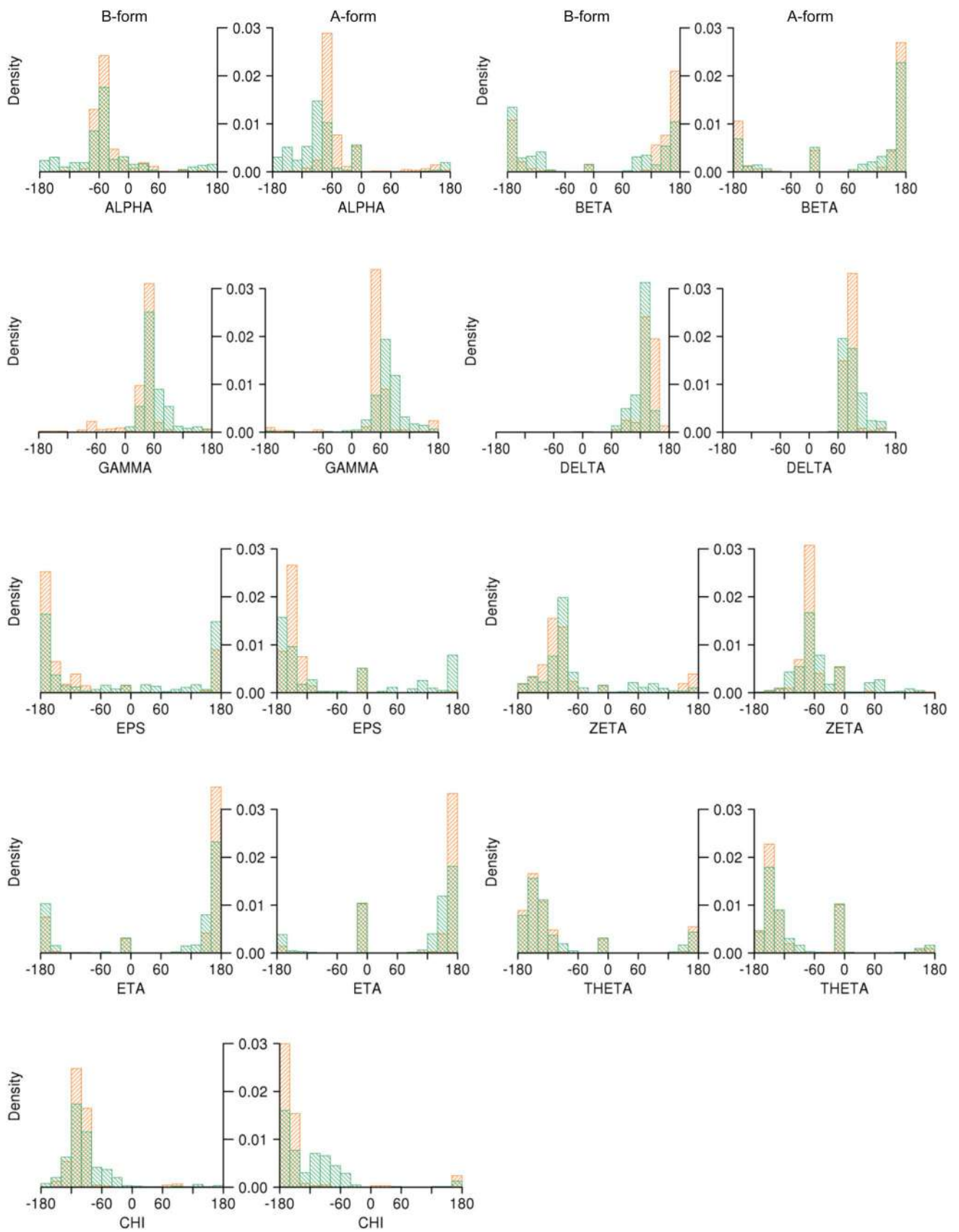


Fig. 4 Frequency distributions of all the backbone torsion angles (α , β , γ , δ , ϵ , ξ , χ) and the two pseudo-torsion angles, η and θ [56], for the original structures and regenerated structures, classifying them into A-form (right panels) and B-form (left panels), are shown. The distributions of the original torsion angles are shown in *red* while the *green colored bars* are created from torsion angles of the regenerated and constrained minimized structures

Comparison of electrostatic potential

Variations in torsion angles of the original and regenerated structures may give rise to changes in groove dimension of

the double helices. This can affect molecular recognition through alteration of electrostatic potentials (ESP) around the molecules. Hence, we have compared the electrostatic surface potential (ESP) values calculated by Poisson–Boltzmann method (Fig. 5). The ESP values are also compared quantitatively, by calculating the pair-wise Pearson correlation coefficients and their RMS deviations. Each surface point on the original crystal structure and its corresponding nearest neighboring surface point on the regenerated model were identified prior to the calculation of the two measures (Table 3). For all the correlation coefficient values, the

Fig. 5 Electrostatic Potential calculated by Delphi, mapped onto the original (*left panel*) and regenerated (*right panel*) structures for **a** 1KX5; **b** 1RSB (DNA), **c** 4V4Q and **d** 4V9R (RNA). This file was used to draw the images on the molecular surfaces (MS) in Chimera

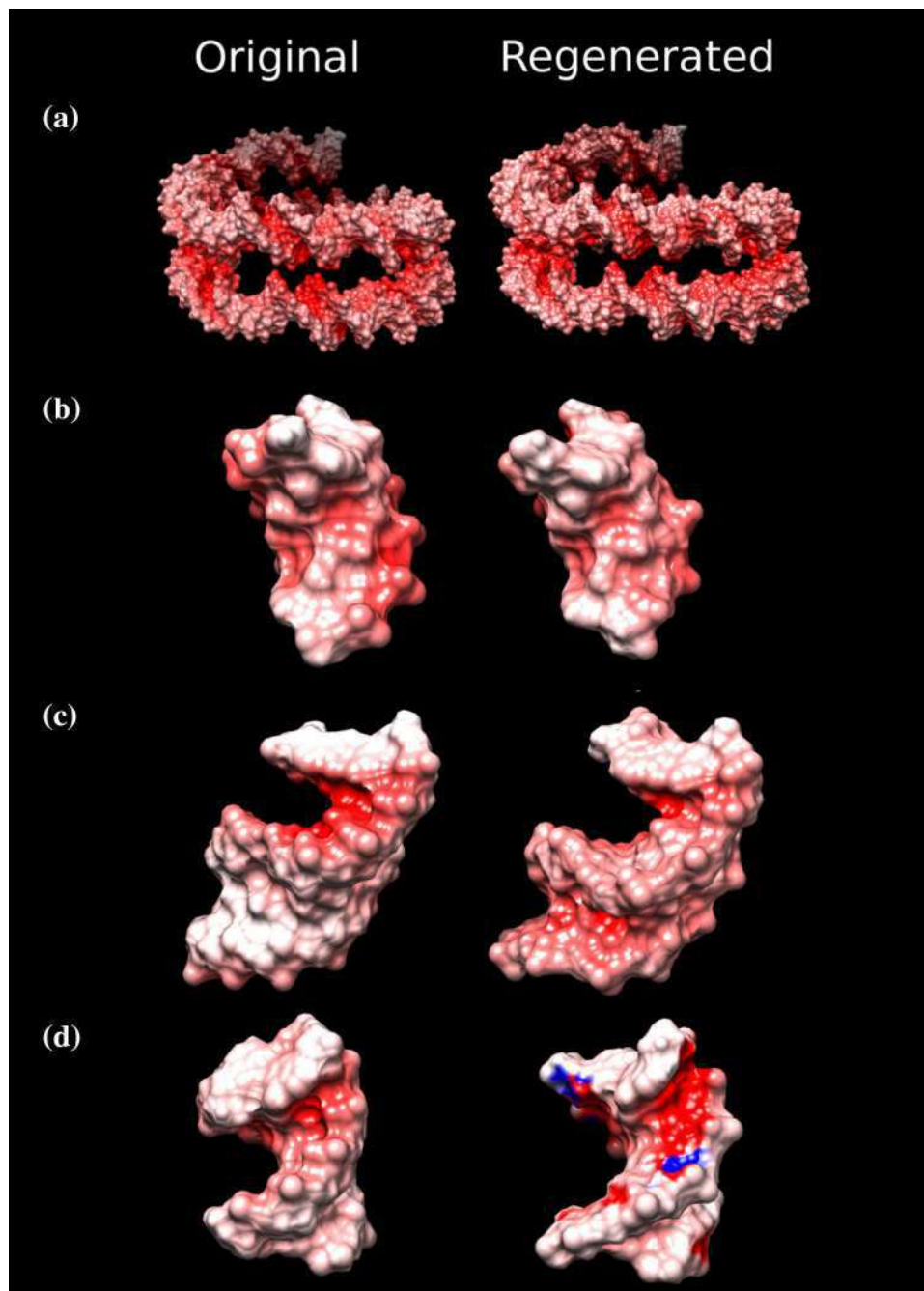


Table 3 Correlation and RMSD values of grid potentials between the original (experimental) structure with regenerated structures

S. No.	Model	RMSD (kT/e)	Correlation coefficient (Pearsons)	Number of nearest neighboring dot points
RNA				
1	1J5A	19.74	0.49	24,238
2	1FJG	19.41	0.48	27,214
3	1N33	20.17	0.46	37,790
4	4V4Q	19.25	0.46	35,957
5	1N32-a	19.19	0.47	28,911
6	354D	18.53	0.45	33,928
7	1XMQ	18.39	0.49	29,687
8	1N32-b	18.44	0.46	29,649
9	5J7L-bundle3	18.41	0.50	38,856
10	4V9R-bundle4	18.95	0.48	22,653
11	3R1C	20.30	0.43	26,545
12	5DM6	19.23	0.51	37,405
13	4V88-bundle2	18.37	0.51	37,940
14	2L3C	20.38	0.41	48,122
15	2AZX	18.80	0.51	18,021
DNA				
1	1BNA	15.18	0.68	41,833
2	1K61	14.95	0.65	65,370
3	1RSB	15.95	0.60	20,618
4	1KX5	18.93	0.56	202,378
5	1QNC	16.07	0.62	49,625
6	111D	16.55	0.65	41,255
7	1DNM	17.54	0.56	41,082
8	399D	12.68	0.76	41,766

Corresponding helix information is given in Table 1. Correlation and RMSD values (in units of kT/e) were computed on electrostatic surface potentials. P-values for all correlation coefficients were found to be $<10^{-5}$ indicating that the correlations are significant at 99.9% confidence and hence could not have occurred by chance

associated p-value was found to be 0.00001 ruling out the fact that they could have been obtained randomly. These agreements strongly suggest that the model structures are near-native and can thus effectively help in understanding the molecular recognition properties of the helices.

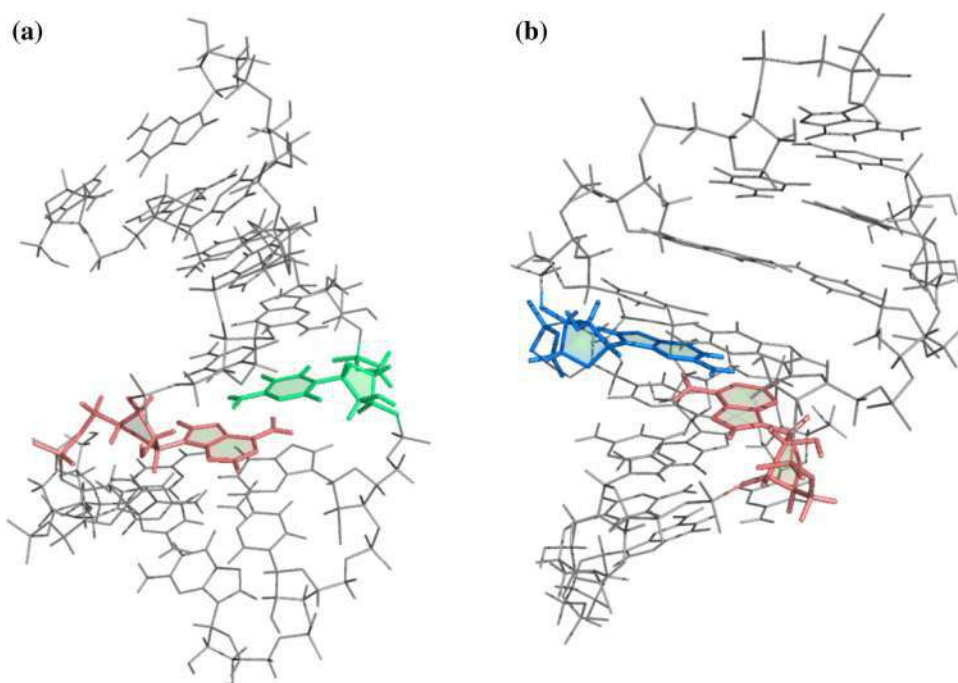
Discussion

Presence of increasing number of NWC bps reported in double helices of DNA/ RNA molecules, has prompted us to update our program NUCGEN to *RNAHelix*, which can generate nucleic acid duplexes, compatible with these NWC bps. It employs a self-consistent formulation for model building of helical stretches with user-defined intra-bp and inter-bp parameters. *RNAHelix* uses an edge-specific axis system, similar to NUPARM-2, for generation of non-canonical base pairing geometries. The required input parameters can be obtained by

NUPARM-2 or similar software utilities, available for nucleic acid structure analysis. The regenerated models showed very low RMSD, when compared to corresponding original crystal structure, indicating high accuracy of the program. We have further shown that generation of sugar-phosphate backbone using restrained energy minimization is a suitable method for obtaining coordinates of the complete model structure. Moreover, using *RNAHelix*, we have generated dinucleotide steps containing non-canonical bps to identify their stable geometries. *RNAHelix* can also be applied for modeling a variety of isolated non-canonical bps, which has not been reported in functional RNA structures till date, for carrying out in silico structural studies. We believe that *RNAHelix* could make an important contribution towards a better understanding of comparatively lesser occurring NWC bps and homology modeling of nucleotide sequences.

As indicated earlier, this method can generate double helical nucleic acid structures containing any type of

Fig. 6 Two hypothetical structures using intra-bp and inter-bp parameters for Watson–Crick basepairs from fiber-diffraction derived structure of A-RNA [55] having **a** A:C H:WT and **b** G:A S:WT non-canonical bps at the center. The intra-bp parameters of these non-canonical basepairs were taken from Supplementary Table ST3. The complete input files are given in Supplementary Table ST4 and ST5, respectively. Only the representative pairs are in color. The colors used for the bases, red, blue, green are representing Adenine, Guanine, Cytosine respectively



non-canonical basepairing. Two such hypothetical models generated using intra-bp parameters for A:C H:WT and G:A S:WT bps from Table ST3 and inter-bp parameters and intra-bp parameters for W:WC bps following A-RNA fiber model [55] generated using NACC server (Supplementary Tables ST4 and ST5) are shown in Fig. 6. One can generate models for such DNA or RNA double helical structure containing any non-Watson–Crick base pairing. We, however, could not compare the accuracy of this generation, as no reference structure is available. These models, however, may not be optimum, as twist or other inter-bp parameters for such stacking between W:WC and S:WT bps, for example, can be quite different from regular values ($\rho \sim 10$, $\omega \sim 33$, etc. for A-RNA helix). We believe the software can be adjusted easily if any difference between original and regenerated structures for an unusual helix is found at a later date.

Funding This work has been supported by the Department of Atomic Energy, Govt. of India and Department of Biotechnology, Govt. of India. MB is recipient of J.C. Bose National Fellowship from DST, India.

References

- Kim SH, Suddath FL, Quigley GJ et al (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185:435–440. doi:10.1126/science.185.4149.435
- Holbrook SR, Cheong C, Tinoco I, Kim SH (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson–Crick base pairs. *Nature* 353:579–581
- Cruse WB, Saludjian P, Biala E et al (1994) Structure of a mispaired RNA double helix at 1.6-Å resolution and implications for the prediction of RNA secondary structure. *Proc Natl Acad Sci USA* 91:4160–4164
- Egli M, Portmann S, Usman N (1996) RNA hydration: a detailed look †, ‡. *BioChemistry* 35:8489–8494. doi:10.1021/bi9607214
- Lenz T, Bonnist EYM, Pljevaljčić G et al (2007) 2-aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M. TaqI: crystal structures and time-resolved fluorescence. *J Am Chem Soc* 129:6240–6248. doi:10.1021/ja069366n
- Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* 7:499–512
- Bhattacharya S, Mittal S, Panigrahi S, et al. (2015) RNABP COGEST: a resource for investigating functional RNAs. Database (Oxford) bav011. doi:10.1093/database/bav011
- Olson WK, Bansal M, Burley SK et al (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* 313:229–237. doi:10.1006/jmbi.2001.4987
- Dickerson RE (1989) Definitions and nomenclature of nucleic acid structure parameters. *J Biomol Struct Dyn* 6:627–634. doi:10.1080/07391102.1989.10507726
- Calladine CR (1982) Mechanics of sequence-dependent stacking of bases in B-DNA. *J Mol Biol* 161:343–352. doi:10.1016/0022-2836(82)90157-7
- Ravishanker G, Swaminathan S, Beveridge DL et al (1989) Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: “curves”, dials and windows. *J Biomol Struct Dyn* 6:669–699. doi:10.1080/07391102.1989.10507729
- Bhattacharyya D, Bansal M (1990) Local variability and base sequence effects in DNA crystal structures. *J Biomol Struct Dyn* 8:539–572. doi:10.1080/07391102.1990.10507828
- Babcock MS, Olson WK (1994) The effect of mathematics and coordinate system on comparability and “dependencies” of nucleic acid structure parameters. *J Mol Biol* 237:98–124. doi:10.1006/jmbi.1994.1212
- Bandyopadhyay D, Bhattacharyya D (2000) Effect of neighboring bases on base-pair stacking orientation: a molecular

- dynamics study. *J Biomol Struct Dyn* 18:29–43. doi:[10.1080/07391102.2000.10506645](https://doi.org/10.1080/07391102.2000.10506645)
15. Beveridge DL, Barreiro G, Suzie Byun K et al (2004) Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. I. Research design and results on d(CpG) steps. *Biophys J* 87:3799–3813. doi:[10.1529/biophysj.104.045252](https://doi.org/10.1529/biophysj.104.045252)
 16. Fujii S, Kono H, Takenaka S et al (2007) Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res* 35:6063–6074. doi:[10.1093/nar/gkm627](https://doi.org/10.1093/nar/gkm627)
 17. Davey CA, Sargent DF, Luger K et al (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 319:1097–1113. doi:[10.1016/S0022-2836\(02\)00386-8](https://doi.org/10.1016/S0022-2836(02)00386-8)
 18. Wu B, Mohideen K, Vasudevan D, Davey CA (2010) Structural insight into the sequence dependence of nucleosome positioning. *Structure* 18:528–536. doi:[10.1016/j.str.2010.01.015](https://doi.org/10.1016/j.str.2010.01.015)
 19. Andrews AJ, Luger K (2011) Nucleosome structure(s) and stability: variations on a theme. *Annu Rev Biophys* 40:99–117. doi:[10.1146/annurev-biophys-042910-155329](https://doi.org/10.1146/annurev-biophys-042910-155329)
 20. Halder S, Bhattacharyya D (2010) Structural stability of tandemly occurring noncanonical basepairs within double helical fragments: molecular dynamics studies of functional RNA. *J Phys Chem B* 114:14028–14040. doi:[10.1021/jp102835t](https://doi.org/10.1021/jp102835t)
 21. Halder S, Bhattacharyya D (2012) Structural variations of single and tandem mismatches in RNA duplexes: a joint MD simulation and crystal structure database analysis. *J Phys Chem B* 116:11845–11856. doi:[10.1021/jp305628v](https://doi.org/10.1021/jp305628v)
 22. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. *Nucleic Acids Res* 28:235–242. doi:[10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235)
 23. Mohanty D, Bansal M (1991) DNA polymorphism and local variation in base-pair orientation: a theoretical rationale. *J Biomol Struct Dyn* 9:127–142. doi:[10.1080/07391102.1991.10507898](https://doi.org/10.1080/07391102.1991.10507898)
 24. Hunter CA (1993) Sequence-dependent DNA structure. The role of base stacking interactions. *J Mol Biol* 230:1025–1054. doi:[10.1006/jmbi.1993.1217](https://doi.org/10.1006/jmbi.1993.1217)
 25. Mondal M, Halder S, Chakrabarti J, Bhattacharyya D (2016) Hybrid simulation approach incorporating microscopic interaction along with rigid body degrees of freedom for stacking between base pairs. *Biopolymers* 105:212–226. doi:[10.1002/bip.22787](https://doi.org/10.1002/bip.22787)
 26. Lu X-J, Olson WK (2003) 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res* 31:5108–5121. doi:[10.1093/nar/gkg680](https://doi.org/10.1093/nar/gkg680)
 27. Lu X-J, Olson WK (2008) 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* 3:1213–1227. doi:[10.1038/nprot.2008.104](https://doi.org/10.1038/nprot.2008.104)
 28. van Dijk M, Bonvin AMJJ (2009) 3D-DART: a DNA structure modelling server. *Nucleic Acids Res* 37:W235–W239. doi:[10.1093/nar/gkp287](https://doi.org/10.1093/nar/gkp287)
 29. Bansal M, Bhattacharyya D, Ravi B (1995) NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. *Bioinformatics* 11:281–287. doi:[10.1093/bioinformatics/11.3.281](https://doi.org/10.1093/bioinformatics/11.3.281)
 30. Macke TJ, Case DA (1997) Modeling unusual nucleic acid structures. In: *ACS Symp. Ser.* pp 379–393
 31. Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452:51–55. doi:[10.1038/nature06684](https://doi.org/10.1038/nature06684)
 32. Popenda M, Szachniuk M, Antczak M et al (2012) Automated 3D structure composition for large RNAs. *Nucleic Acids Res* 40:e112. doi:[10.1093/nar/gks339](https://doi.org/10.1093/nar/gks339)
 33. Lu X-J, El Hassan MA, Hunter CA (1997) Structure and conformation of helical nucleic acids: rebuilding program (SCHNArP). *J Mol Biol* 273:681–691. doi:[10.1006/jmbi.1997.1345](https://doi.org/10.1006/jmbi.1997.1345)
 34. Bhattacharya D, Bansal M (1988) A general procedure for generation of curved DNA molecules. *J Biomol Struct Dyn* 6:93–104. doi:[10.1080/07391102.1988.10506484](https://doi.org/10.1080/07391102.1988.10506484)
 35. Bhattacharyya D, Bansal M (1989) A self-consistent formulation for analysis and generation of non-uniform DNA structures. *J Biomol Struct Dyn* 6:635–653. doi:[10.1080/07391102.1989.10507727](https://doi.org/10.1080/07391102.1989.10507727)
 36. Mukherjee S, Bansal M, Bhattacharyya D (2006) Conformational specificity of non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Comput Aided Mol Des* 20:629–645. doi:[10.1007/s10822-006-9083-x](https://doi.org/10.1007/s10822-006-9083-x)
 37. Brooks BR, Brooks CL, Mackerell AD et al (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614. doi:[10.1002/jcc.21287](https://doi.org/10.1002/jcc.21287)
 38. Das J, Mukherjee S, Mitra A, Bhattacharyya D (2006) Non-canonical base pairs and higher order structures in nucleic acids: crystal structure database analysis. *J Biomol Struct Dyn* 24:149–161. doi:[10.1080/07391102.2006.10507108](https://doi.org/10.1080/07391102.2006.10507108)
 39. Panigrahi S, Pal R, Bhattacharyya D (2011) Structure and energy of non-canonical basepairs: comparison of various computational chemistry methods with crystallographic ensembles. *J Biomol Struct Dyn* 29:541–556. doi:[10.1080/07391102.2011.10507404](https://doi.org/10.1080/07391102.2011.10507404)
 40. Ray SS, Halder S, Kaypee S, Bhattacharyya D (2012) HD-RNAS: an automated hierarchical database of RNA structures. *Front Genet* 3:59. doi:[10.3389/fgene.2012.00059](https://doi.org/10.3389/fgene.2012.00059)
 41. Petrov AI, Zirbel CL, Leontis NB (2013) Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. *RNA* 19:1327–1340. doi:[10.1261/rna.039438.113](https://doi.org/10.1261/rna.039438.113)
 42. Clowney L, Jain SC, Srinivasan AR et al (1996) Geometric parameters in nucleic acids: nitrogenous bases. *J Am Chem Soc* 118:509–518. doi:[10.1021/ja952883d](https://doi.org/10.1021/ja952883d)
 43. Cornell WD, Cieplak P, Bayly CI et al (1995) A Second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197. doi:[10.1021/ja00124a002](https://doi.org/10.1021/ja00124a002)
 44. Xu D, Zhang Y (2009) Generating triangulated macromolecular surfaces by Euclidean distance transform. *PLoS One* 4:e8140. doi:[10.1371/journal.pone.0008140](https://doi.org/10.1371/journal.pone.0008140)
 45. Basu S, Bhattacharyya D, Banerjee R (2012) Self-complementarity within proteins: bridging the gap between binding and folding. *Biophys J* 102:2605–2614. doi:[10.1016/j.bpj.2012.04.029](https://doi.org/10.1016/j.bpj.2012.04.029)
 46. Rocchia W, Sridharan S, Nicholls A et al (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J Comput Chem* 23:128–137. doi:[10.1002/jcc.1161](https://doi.org/10.1002/jcc.1161)
 47. Li L, Li C, Sarkar S, et al. (2012) DelPhi: a comprehensive suite for DelPhi software and associated resources. *BMC Biophys* 5:9. doi:[10.1186/2046-1682-5-9](https://doi.org/10.1186/2046-1682-5-9)
 48. Pettersen EF, Goddard TD, Huang CC et al (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. doi:[10.1002/jcc.20084](https://doi.org/10.1002/jcc.20084)
 49. Blanchet C, Pasi M, Zakrzewska K, Lavery R (2011) CURVES + web server for analyzing and visualizing the helical, backbone and groove parameters of nucleic acid structures. *Nucleic Acids Res* 39:W68–W73. doi:[10.1093/nar/gkr316](https://doi.org/10.1093/nar/gkr316)
 50. Saenger W (1984) Principles of nucleic acid. Structure. doi:[10.1007/978-1-4612-5190-3](https://doi.org/10.1007/978-1-4612-5190-3)
 51. Ulanovsky LE, Trifonov EN (1987) Estimation of wedge components in curved DNA. *Nature* 326:720–722. doi:[10.1038/326720a0](https://doi.org/10.1038/326720a0)

52. Kailasam S, Bhattacharyya D, Bansal M et al (2014) Sequence dependent variations in RNA duplex are related to non-canonical hydrogen bond interactions in dinucleotide steps. *BMC Res Notes* 7:83. doi:[10.1186/1756-0500-7-83](https://doi.org/10.1186/1756-0500-7-83)
53. Leontis NB, Zirbel CL (2012) In: Leontis N, Westhof E (eds) *Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking*. Springer Berlin Heidelberg, Berlin, pp 281–298t;/bib>
54. Cheatham TE, Case DA (2013) Twenty-five years of nucleic acid simulations. *Biopolymers* 99:969–977. doi:[10.1002/bip.22331](https://doi.org/10.1002/bip.22331)
55. Arnott S, Hukins DW, Dover SD (1972) Optimised parameters for RNA double-helices. *Biochem Biophys Res Commun* 48:1392–1399
56. Duarte CM, Pyle AM (1998) Stepping through an RNA structure: A novel approach to conformational analysis. *J Mol Biol* 284:1465–1478. doi:[10.1006/jmbi.1998.2233](https://doi.org/10.1006/jmbi.1998.2233)