J Comput Aided Mol Des (2014) 28:851–867
DOI 10.1007/s10822-014-9767-6

Analysis of stacking overlap in nucleic acid structures: algorithm and application

Pavan Kumar Pingali · Sukanya Halder · Debasish Mukherjee · Sankar Basu · Rahul Banerjee · Devapriya Choudhury · Dhananjay Bhattacharyya

Received: 1 April 2014 / Accepted: 23 June 2014 / Published online: 3 July 2014

Abstract RNA contains different secondary structural motifs like pseudo-helices, hairpin loops, internal loops, etc. in addition to anti-parallel double helices and random coils. The secondary structures are mainly stabilized by base-pairing and stacking interactions between the planar aromatic bases. The hydrogen bonding strength and geometries of base pairs are characterized by six intra-base pair parameters. Similarly, stacking can be represented by six local doublet parameters. These dinucleotide step parameters can describe the quality of stacking between Watson–Crick base pairs very effectively. However, it is quite difficult to understand the stacking pattern for dinucleotides consisting of non canonical base pairs from these parameters. Stacking interaction is a manifestation of the interaction between two aromatic bases or base pairs and thus can be estimated best by the overlap area between the planar aromatic moieties. We have calculated base pair overlap between two consecutive base pairs as the buried van der Waals surface between them. In general, overlap values show normal distribution for the Watson–Crick base pairs in most double helices within a range from 45 to 50 $Å^2$ irrespective of base sequence. The dinucleotide steps with non-canonical base pairs also are seen to have high overlap value, although their twist and few other parameters are rather unusual. We have analyzed hairpin loops of different length, bulges within double helical structures and pseudo-continuous helices using our algorithm. The overlap area analyses indicate good stacking between few looped out bases especially in GNRA tetraloop, which was difficult to quantitatively characterise from analysis of the base pair or dinucleotide step parameters. This parameter is also seen to be capable to distinguish pseudo-continuous helices from kinked helix junctions.

Keywords Base stacking · DNA · RNA · Double helices · Tetraloop conformation · Bulges in RNA helix · Non-Watson–Crick base pair

Introduction

Ribonucleic acids (RNA) are large biomolecules and their functional forms usually consist of different structural subunits or motifs. These are collections of short helices interspersed by unpaired regions, packed together into compact structures. Similar to proteins, the structure of RNA can be divided into four levels of organization—primary, secondary, tertiary and quaternary. Primary structure is described by the nucleotide sequence of the RNA. Secondary structural motifs are mainly comprised of helices and intervening unpaired loop regions. RNA can form a number of secondary structures by folding upon itself and forming base pairs between complementary regions, e.g., hairpins, internal loops, junctions etc. The tertiary and quaternary structures connect distant regions of

Electronic supplementary material The online version of this article (doi:10.1007/s10822-014-9767-6) contains supplementary material, which is available to authorized users.

P. K. Pingali · S. Halder · D. Mukherjee · D. Bhattacharyya (✉)
Computational Science Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India
e-mail: dhananjay.bhattacharyya@saha.ac.in

S. Basu · R. Banerjee
Crystallography and Molecular Biology Division, Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Kolkata 700064, India

D. Choudhury
School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India

the RNA strands through extensive hydrogen bonding and backbone interactions. Tertiary structural motifs are comprised of triple helices, kissing loops, co-axial helices, pseudoknots, etc. Quaternary structures, as a form of multiple RNA chains acting together, are seen in a few cases, such as ribosomes.

Hydrogen bonding and stacking interactions between the nucleobases are fundamental to the organization of nucleic acids. The stacking of bases contributes significantly to the stability of RNA folds as it minimizes the exposure of hydrophobic base surfaces to the polar solvent. Double helical regions of nucleic acids usually possess G:C and A:U Watson–Crick base pairs along with G:U wobble base pairs. In case of DNA, base pairs observed are mostly of Watson–Crick type, the only exception being the G:G base pairs in telomeric DNA forming Guanine quadruplexes [1]. In RNA, however, several types of non-canonical base pairs can be found along with standard Watson–Crick base pairs. These base pairs are formed through planar edge-to-edge hydrogen bonding between RNA bases involving one of three distinct edges—Watson–Crick edge, Sugar edge and Hoogsteen edge [2]. The base pairs can either be in *cis* or *trans* orientation. Reports suggest that only 60 % of the base pairs in RNA are Watson–Crick type, the rest being non-canonical in nature [2]. Studies have also shown that many of the non-canonical base pairs are as stable as canonical base pairs [3, 4]. Non canonical base pairs often appear within the RNA double helical regions and it was shown that tandem occurrence of such non-canonical base pairs within double helical structures flanked by Watson–Crick base pairs are quite stable [5, 6]. They also appear consistently in other structural motifs like hairpin loops, loop–loop recognition, coaxial helices etc. They play an integral part in RNA structural organization by spatially connecting structural motifs formed by distant parts of the RNA chains.

A large number of secondary structural motifs of RNA have been identified by FR3D software suite [7] and are categorically listed in the RNA Structure Atlas and RNA 3D Motif Atlas databases [8]. Among the secondary structural motifs found in RNA, bulge loops or bulges have a universal distribution in all types of structured and functional RNAs. A bulge loop forms when a double helix is interrupted by unpaired stretches of nucleotides on only one strand [9]. Bulge sizes can range from a single unpaired residue to several unpaired nucleotides. If the bulge size is several residues long so that it appears to be formed by two different parts of the RNA chain, it is described as a pseudo-continuous helix or co-axial stack. The strand of continuously base paired residues, which is opposite to the bulged strand, facilitates the stacking of the flanking stems. Bulges can create recognition sites in RNA three-dimensional structures both directly, by acting as molecular handles within helical regions which are otherwise uniform, and indirectly, by distorting the RNA backbone and allowing access to base pairs in a widened deep groove. They may also act as contact points in the tertiary folding of RNA [9].

The conformation of bulge loops is governed by the competing interactions of both the unpaired residues and the surrounding base pairs. Unpaired residues may participate in continuous stacking of the flanking regions, or they can be extruded from the duplex with the bases pointing out into the solvent. Unpaired residues in the bulge can favour kinking of the helix axis at the bulge site [10], e.g., Hook-turn, K-turn and Reverse K-turn motifs. On the other hand, when the bulge nucleotides are looped out, the composite duplex geometries can remain close to the regular A-form [11]. The dimensions of the major and minor groove in RNA duplexes can be altered by bulges which introduce distortions in the nucleic acid backbone. Widening of the major groove by bulges, which exposes the hydrogen bonding edges of base pairs, is frequently found at interaction sites where RNA domains dock into duplexes, or ligands bind to RNA [12, 13].

Hairpin loops are the most frequently observed secondary structural elements in RNA. These are formed by an unpaired region joining the two strands of a double helical stretch. Hairpin loops play an important role in providing sites for protein folding and also act as structural scaffolds and recognition sites for both proteins and nucleic acids [14]. Among different hairpin loops that occur in RNA, tetraloops are the most common ones. Tetraloops usually form compact and stable structures which may involve base–base or base-sugar interactions. Sometimes the first and fourth nucleotides in the loop are found paired, leaving two constrained bases in the loop. The second and third nucleotides form a turn in the RNA strand. Theoretically, 256 types of tetraloop sequences are possible, most of which are very rare and uncommon. The most common tetraloop motifs found in functional RNAs are GNRA [15–19], UNCG [20, 21], CUYG [15, 22], ANYA [23–25] and (A/U)GNN [26–28] [N → any residue, R → purine, Y → pyrimidine]. These five families account for 54 types of tetraloop sequences leaving out 202 possibilities. Among these families, GNRA, CUYG and UNCG are especially abundant in naturally occurring RNAs [15, 29] with GNRA and UNCG families accounting for up to 70 % of the loop sequences [30]. Other less frequently observed tetraloop families have also been reported such as YNMG [31], GANC [32], UNAC [33], GYYA [34] and GNAR [35] tetraloops (M → Ade or Cyt). Because of the conformational similarity, Sakamoto et al. [36] suggested to combine GNRA and GNAR families to GNRR family. The conformations of nucleotides in the loop regions differ completely from one family to another [17]. Tetraloops

have an A-form stem while the bases in the loop region can either assume an *anti* conformation or *syn* conformation [37]. In addition to tetraloops, there are other distinct types of loops in RNA structures depending on the number of unpaired residues, such as triloop, pentaloop, hexaloop, etc. Among them, triloops are expected to have some well defined local structural features [38, 39], while the longer loops are generally more flexible.

Non-canonical basepairs are found to be present within various types of secondary structural motifs of RNA, e.g., within double helical stretches, at the helix-loop junctions, or surrounding the bulges, as they are supposed to introduce flexibility in the backbone of RNA through their unusual conformations [40]. It has previously been reported that non-canonical basepairs are often associated with anomalous torsion angles [5, 6]. However, characterization of even double helical structures from torsion angle analysis is quite difficult [41] as the seven freely variable torsion angles can accommodate many strains. Hence, several groups have used backbone dihedrals in the form of pseudo-torsion to envisage the structural correlations between base pairs and backbone conformations [42–45]. However, inherent flexibility of nucleic acid backbones limits the accuracy of using torsion angles to describe structural variations of RNA structures.

The base pairing and stacking stabilities can be described by two sets of geometrical parameters. The parameters which describe orientation of one of the base with respect to its paired one are generally referred to as intra base pair parameters. Among the intra base pair parameters, propeller and buckle, which are rotational parameters, and stagger, a translational parameter, describes the overall non-planarity of a base pair. On the other hand, two translational parameters, shear and stretch along with open angle describe the hydrogen bonding strength between the bases. Similarly, there are three rotational and three translational parameters which describe the relative orientation between two successive base pairs. Tilt and roll indicate the wedge formation between the base pairs. Shift and slide describe the lateral movements between the stacked base pairs while twist and rise are usual helical parameters. There are reports in literature [46–49] which describe preferences of these parameters depending on the base pairing, sequence, etc., in crystallographic structures of DNA and double helical RNA. These analyses indicate several features for the stacking between canonical base pairs.

1. Tilt and shift values are generally zero,
2. Roll values strongly depend on the dinucleotide sequence both in DNA and RNA. These values are generally large positive ($\sim 10°$) in RNA,
3. Twist values are mostly around 30°–40° with smaller twist values being more preferred in A-form RNA helices,
4. Slide values are generally close to zero in B-DNA while in A-form RNA, the values are around -1.5 Å,
5. Rise values are around 3.4 Å in DNA while it reduces slightly in RNA.

In case of DNA, the most common form of structural organization is the double helix, where much of the flexibility is lost because of the base pairing and stacking interactions. Thus, the above mentioned set of intra- and inter-basepair parameters also fall within the normal range of values defined by IUPAC-IUB convention [46]. Any anomaly in these values indicate deviation from regular double helical structure. RNA, however, is not restricted by the limitations of a double helix and retains its flexibility in many of its structural motifs. Analyses of these parameters in non-uniform RNA structures as well as non-canonical basepairs, thus, are expected to result in irregular values, even though the structure shows stability. Recent studies [5] have shown that the base pair step parameters for some dinucleotides consisting of one or two non canonical base pairs are not always similar to those observed for canonical ones and appear to be anomalous. For example, twist values of non-canonical base pair containing dinucleotides are sometimes close to zero or 90°, yet they are found to be as stable as the canonical ones [5, 6]. It is also difficult to define the stability and structural features of nucleic acids in a multi-dimensional space consisting of all the intra- and inter-basepair parameters. Hence, there is a need for a parameter which describes the overall stacking interaction between successive base pairs in a double helical region, manifesting their stability. Calculation of explicit stacking energy to characterize the stability is non-trivial and computationally demanding. Therefore, we have calculated the overlap area between successive base pairs to analyze the stacking interactions between them. As the inter-base pair parameters describe the relative geometric orientation of the two successive base pairs, overlap area between successive base pairs can be considered to manifest the composite effect of all the six inter-base pair parameters. Overlap values of dinucleotide steps are also calculated by 3DNA software [50, 51]. However, the values calculated by 3DNA appear to be too small as compared to the total area covered by the base pairs. We have, therefore, defined a new method for calculation of this parameter depending on the buried surface area between two consecutive base pairs. We have calculated overlap values for the available B-DNA and A-DNA crystal structures, which can be used as standards. The base pair overlap values for different dinucleotide sequences consisting of canonical as well as non Watson–Crick base pairs have been found to be quite

high. We have analyzed the RNA and B-DNA structures solved by NMR spectroscopy and found similar trends. We have further used these values for the other structural motifs of RNA, e.g., tetraloops, pseudo-continuous helices and bulge loops. We find that the overlap values estimated from buried surface area represent the extent of stacking interactions reliably and are consistent with the stability features of these structural motifs.

Methods

We have prepared three datasets from crystal structures of A-DNA, B-DNA and RNA. A non-redundant dataset of 123 functional RNA crystal structures with resolution better than 3.5 Å containing chains longer than 30 nucleotides was obtained from HD-RNAS database [52]. These structures were downloaded from Protein Data Bank [53] and the PDB-IDs of the selected structures are given in the supplementary information (Table S1). All the biological assemblies of A-DNA and B-DNA crystal structures, solved at a resolution better than 3.0 Å, were downloaded from NDB [54] and the structures with modified or mismatched base pair were rejected. PDB-IDs of the 188 A-DNA crystal structures and 162 B-DNA crystal structures selected are given in supplementary information (Table S1).

We have also analyzed the nucleic acid structures solved from NMR spectroscopic data. We found only 5 A-DNA structures characterized by NDB [54] and hence ignored them. There are 109 B-DNA structures available in NDB (List S1) without any ligand or chemical modification; we have analyzed all of them to understand their structural features. Similarly we found 68 RNA structures solved by NMR without ligand or chemical modification (List S2). Although some of these structures possess non-canonical base pairs, their total number is quite insignificant. Hence, we have analyzed the properties of the base pair stacks formed by Watson–Crick base pairs only. Quite often there are several model conformations of same molecule, given in a pdb file solved by NMR spectroscopy. In these cases we have considered only the first models.

The base pairs in DNA and RNA were detected by BPFIND [55] with default hydrogen bond length cut off of 3.8 Å between the acceptor and donor atoms and angle cut off of 120°. The usual intra and inter base pair parameters were calculated by NUPARM [56, 57] along with the overlap values. Similar calculations were also done using 3DNA for A-DNA and B-DNA crystal structures for comparison between the two sets of values. Throughout this paper, we have used the nomenclature system described by Das et al. [55] for representing base pairing geometries.

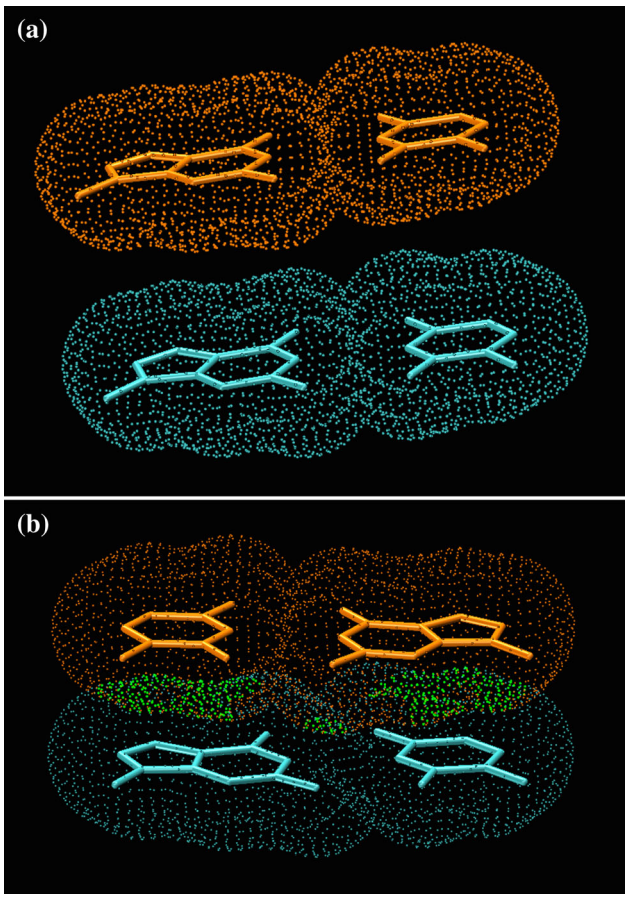

Fig. 1 Van der Waals surface of **a** two G:C W:WC base pairs when they do not stack (VDW surface is shown as points in *orange* and *cyan*) and **b** when they stack [(G:C W:WC)::(G:C W:WC) dinucleotide step], the overlap area has been shown by the points buried between the two base pairs (shown in *green*) [':' represents base pairing and '::' represents stacking]

Method of calculation of stacking overlap by NUPARM

The overlap area is calculated as a measure of buried van der Waals (VDW) surface area (in $Å^2$) between the two base pairs. If S_A and S_B are the VDW surface area of the two stacked base pairs, and S_{AB} is the VDW surface area of the stacked dinucleotide step, the overlap is measured as:

$$Overlap = (S_A + S_B) - S_{AB}$$

The VDW surfaces of the base pairs have been calculated using the method described by Banerjee et al. [58]. According to this method, the van der Waals surface is a collection of uniformly sampled points on the van der Waals sphere of each atom. The overall van der Waals surface of a nucleotide or base pair moiety is generated by excluding the overlapping points between covalently bonded atoms with surface point density of 30 dots/$Å^2$. The radii of each atom has been considered as (R_{min} + 0.2),

where R_{min} is the distance corresponding to minimum van der Waals energy [59]. Similarly, van der Waals surface area is also generated for the composite base pair step. The overlap area is then calculated as the number of surface points that are buried between the two stacked bases or base pairs when a composite VDW surface is generated for the dinucleotide step (Fig. 1).

$$Overlap = \frac{N_1 + N_2 - N_{12}}{2 \times 30}$$

N_1 and N_2 are number of van der Waals surface points for base pair-1 and base pair-2 and N_{12} is number of van der Waals surface points for the composite dinucleotide step.

This method of calculation of stacking overlap has been implemented in NUPARM software and the code is available for download from http://www.saha.ac.in/biop/bioinformatics.html. It should be further noted that one can calculate and analyse stacking overlap between two unpaired bases or between a base pair and an unpaired base also by this method.

Results and discussion

Validation of overlap calculation method

We have estimated the values of surface area covered by the four nucleobases occurring in DNA or RNA as well as canonical Watson–Crick base pairs as two-dimensional projections. The surface area values for the canonical base pairs are found to be around 85 Å^2 (Supplementary Information Table S2; Fig. SF1–3). We can thus expect stacking overlap value of about 85 Å^2 for ideal stacks, where there is no relative rotation and translation between the base pairs. We have generated coordinates of ideal planar base pair stacks with 0° twist and 3.34 Å rise [Roll = 0°; Tilt = 0°; Shift = 0 Å; Slide = 0 Å] using NUCGEN [56] and have calculated their overlap values by 3DNA as well as by overlap calculation utility of our method (NUPARM). We observe that the overlap areas for ideal stacks in most of the dinucleotide steps are around 57 Å^2. By visual inspection, it was confirmed that this reduction in overlap area is due to some of the surface points, which remain exposed as holes in between two base pairs (Fig. 2). However, overlap values calculated by 3DNA are even smaller and found to be within 15 Å^2 (Tables 1, 2; Supplementary Information Table S3). To evaluate the stacking overlap for regular double helices, we have calculated the overlap areas for all the unique dinucleotide steps in canonical B-DNA and A-DNA structures obtained from fiber-diffraction studies [60, 61]. It is observed that purine–pyrimidine steps have higher overlap areas than purine–purine and pyrimidine–purine steps for both A- and B-form helices (Supplementary Information Table S4). The overlap areas between base pairs in A-DNA and B-DNA, depending on dinucleotide sequences, are quite similar to each other. The larger negative slide values in A-DNA do not seem to affect the overlap values. The overlap in A-RNA fiber models [62], however, is significantly smaller as compared to the above. This could be due to the difference in propeller twist values between the structures. As expected, the overlap decreases with increasing slide as the stacking interaction between the base pairs reduces with increase in separation between them. This indicates that the overlap area calculated by NUPARM is a representation of stacking interaction between the base pairs, rather than being geometric overlap of the base pair planes. However, the effect of other base pair parameters like roll, twist and slide on overlap area are also observed, though the effects are less pronounced.

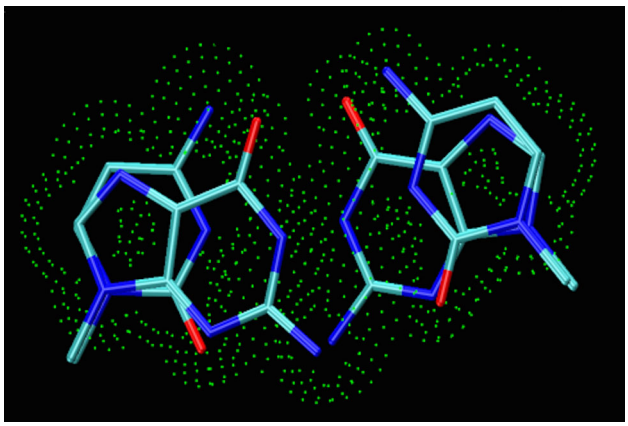

Fig. 2 The reduction in surface overlap is due to the surface points that remain exposed as holes

Stacking overlap values in A-DNA and B-DNA structures

We have calculated stacking overlap in all the selected A-DNA and B-DNA crystal structures. In general, stacking overlap values calculated using NUPARM show normal distributions, within a range of 35–65 Å^2 with mean values around 45–50 Å^2 (Tables 1, 2). Although helical twist, slide, roll, etc., movements between the base pairs in a dinucleotide step are expected to reduce stacking overlap between two base pair planes, we do not observe significant reduction of overlap values for the steps in crystal structures from those observed in idealized orientation. This is possibly due to positioning of atoms of one base pair on top of ring centres of the other base pairs, similar to A-B type of stacking between graphene planes giving rise to graphite [63]. It can be expected from the larger sliding motion between successive base pairs in A-form helices that

Table 1 Mean and SD (in parentheses) values of base step parameters and stacking overlap values in B-DNA

Base pair step	No. of data points	Tilt	Roll	Twist	Shift	Slide	Rise	Overlap	
								NUPARM	3DNA
AA STEP	193	−0.09 (1.75)	−0.53 (3.44)	36.03 (2.67)	−0.00 (0.23)	−0.28 (0.25)	3.23 (0.08)	53.31 (4.03)	8.80 (1.27)
AC STEP	72	−0.32 (1.69)	1.35 (3.56)	31.90 (3.66)	−0.01 (0.49)	−0.19 (0.35)	3.26 (0.11)	53.41 (4.07)	10.07 (1.30)
AG STEP	62	−0.17 (1.86)	4.71 (3.56)	31.44 (5.80)	0.08 (0.52)	0.36 (0.49)	3.33 (0.10)	46.93 (5.44)	10.12 (2.71)
AT STEP	96	0.24 (0.92)	−1.13 (3.07)	33.07 (2.30)	0.02 (0.21)	−0.49 (0.17)	3.14 (0.08)	56.49 (3.30)	11.39 (1.24)
CA STEP B–I	45	−0.15 (2.01)	4.60 (3.56)	35.43 (5.93)	0.07 (0.38)	0.80 (0.60)	3.46 (0.11)	43.27 (4.75)	5.11 (2.16)
CA STEP B-II	26	0.11 (1.80)	−6.72 (3.55)	50.03 (2.14)	0.05 (0.10)	2.43 (0.28)	3.44 (0.08)	40.72 (3.24)	4.17 (0.30)
CC STEP	120	0.33 (3.34)	4.86 (4.15)	33.11 (4.93)	−0.00 (0.59)	0.46 (0.55)	3.35 (0.13)	45.93 (5.72)	9.00 (2.06)
CG STEP	290	0.17 (2.18)	5.09 (4.36)	34.09 (4.76)	0.05 (0.52)	0.52 (0.40)	3.43 (0.11)	44.42 (4.98)	4.74 (1.69)
GA STEP	129	0.07 (1.98)	0.49 (2.64)	38.14 (2.81)	0.02 (0.36)	−0.02 (0.36)	3.28 (0.09)	50.42 (4.04)	7.91 (1.05)
GC STEP	201	−0.25 (2.12)	−3.65 (5.13)	37.26 (4.14)	−0.08 (0.72)	0.40 (0.42)	3.31 (0.11)	50.07 (4.40)	6.41 (1.97)
TA STEP	49	0.24 (2.17)	1.18 (4.39)	39.55 (5.48)	0.00 (0.30)	0.33 (0.85)	3.41 (0.12)	43.29 (5.87)	3.38 (2.48)

Table 2 Mean and SD (in parentheses) values of base step parameters and stacking overlap values in A-DNA

Base pair step	No. of data points	Tilt	Roll	Twist	Shift	Slide	Rise	Overlap	
								NUPARM	3DNA
AA STEP	4	−2.82 (3.04)	16.50 (11.97)	32.27 (3.48)	−0.59 (0.19)	−1.69 (1.10)	3.45 (0.15)	43.12 (4.81)	5.12 (2.56)
AC STEP	161	−0.24 (1.72)	4.47 (2.75)	33.16 (2.39)	−0.12 (0.69)	−1.27 (0.29)	3.19 (0.11)	50.07 (4.15)	11.8 (0.90)
AG STEP	19	0.15 (1.69)	2.98 (5.23)	33.40 (4.64)	−0.04 (0.55)	−1.34 (0.46)	3.34 (0.16)	48.67 (6.29)	5.57 (1.52)
AT STEP	13	−0.14 (0.92)	4.88 (2.09)	32.36 (2.35)	0.17 (0.15)	−1.00 (0.33)	3.29 (0.13)	50.84 (5.39)	12.48 (0.59)
CA STEP	24	0.27 (2.24)	11.29 (3.44)	30.15 (2.39)	−0.01 (0.38)	−1.34 (0.19)	3.38 (0.11)	45.72 (3.55)	2.99 (0.55)
CC STEP	291	−0.04 (2.59)	6.68 (4.20)	30.95 (3.27)	0.00 (0.56)	−1.69 (0.32)	3.34 (0.12)	46.56 (4.30)	4.12 (0.68)
CG STEP	186	−0.23 (1.48)	11.03 (4.47)	28.65 (3.06)	−0.10 (0.40)	−1.91 (0.30)	3.37 (0.09)	45.45 (3.79)	4.74 (0.53)
GA STEP	20	−1.20 (2.48)	7.45 (7.83)	29.25 (3.87)	0.02 (0.58)	−1.28 (0.50)	3.29 (0.14)	46.69 (5.72)	5.61 (2.16)
GC STEP	179	0.50 (1.52)	3.58 (3.93)	33.58 (2.87)	0.10 (0.36)	−1.29 (0.29)	3.26 (0.11)	46.58 (3.65)	11.51 (0.74)
TA STEP	39	−0.01 (1.13)	10.73 (4.92)	28.14 (2.56)	−0.07 (0.26)	−1.37 (0.16)	3.36 (0.13)	45.80 (5.15)	2.29 (0.34)

dinucleotide steps in B-DNA structures would have higher overlap values than those observed in A-DNA structures. However, in crystal structures, the dinucleotide steps tend to attain stable structures by maximizing the stacking overlap between the successive base pairs through an optimum orientation. This is shown by similar overlap values between the dinucleotide steps of A-DNA and B-DNA, even though A-DNA structures conform to larger roll and slide values and the base pairs are shifted along the base pair long axis compared to B-DNA.

The purine-pyrimidine base pair steps, which have small roll values, show maximum overlap as the base pairs in such cases are nearly parallel to each other. Maximum average stacking overlap is observed for (A:T W:WC)::(T:A W:WC) steps [50.8 $Å^2$ for A-DNA (Table 2) and 56.5 $Å^2$ for B-DNA (Table 1)]. On the other hand (G:C W:WC)::(C:G W:WC), a purine-pyrimidine step, has significant negative roll values in B-DNA. Hence, these are non parallel and have smaller overlap values [46.6 $Å^2$ for A-DNA and 50.1 $Å^2$ for B-DNA] with respect to (A:T W:WC)::(T:A W:WC). In B-DNA, two types of (C:G W:WC)::(A:T W:WC) steps are known—B-I and B-II (Table 1). Even though the relative orientations of the base pairs are extremely different in these cases, stacking overlap values are in the same range, indicating both types of stacking are favourable.

We have calculated the correlation coefficients between the local wedge parameters and the overlap values of dinucleotide steps from A-DNA and B-DNA crystal structure data (Table 3). In both the datasets, there is significant negative correlation between rise and overlap values in all forms of double helices, which is an expected feature. However, as the base pairs in B-DNA are located on the helix axis, larger slide values between two consecutive base pairs decrease the overlap considerably, giving rise to a negative correlation between them (−0.6). On the other hand, the base pairs in A-form nucleic acid double

Table 3 Correlation between stacking overlap and base step parameters in canonical base pair steps

	BDNA	ADNA	RNA
No. of data points	1,283	935	6,507
Tilt-overlap	−0.03528	−0.03977	−0.01821
Roll-overlap	−0.16347	−0.14085	0.040375
Twist-overlap	−0.11579	0.332913	0.12336
Shift-overlap	−0.0179	−0.09952	−0.04423
Slide-overlap	−0.59998	0.227952	0.049036
Rise-overlap	−0.77154	−0.68635	−0.58791

helices are shifted from the helix axis and have slide values about −1.5 Å in general. Thus, the stacking interaction appears to be implicitly dependent on twist and slide motions. We do not find any notable correlation between overlap and the other translational and rotational motions, e.g., roll, tilt and shift.

The general trend of the sequence dependence of base pair step parameters, as found from analysis of B-DNA crystal structures, is also seen in the structures solved from NMR data. The mean values of twist for most of the dinucleotide sequences appear to be slightly smaller in the NMR derived structures (Table S5) as compared to their mean values in the B-DNA crystal structure database [47]. Such small values of twist are generally associated with systematically negative values of slide. The base pair stacking overlap values, however, show a trend very similar to that observed in the B-DNA crystal structures, i.e., the purine–pyrimidine sequences have larger base pair overlap than those of the pyrimidine–purine sequences. This possibly indicates the overlap values are not artefacts of crystal packing effect.

Stacking overlap values in RNA crystal structures

The stacking overlap values in RNA double helical regions containing canonical Watson–Crick base pairs calculated using NUPARM show normal distributions within a range of 35 to 65 $Å^2$ (Table 4). The trend of sequence dependent variation of stacking overlap, as found in DNA crystal structures, is also reflected in RNA double helices. The only exception being the overlap in (A:U W:WC)::(U:A W:WC) vs. (U:A W:WC)::(A:U W:WC) steps—the (A:U W:WC)::(U:A W:WC) step has slightly smaller mean overlap as compared to the other. It may be noted that mean roll values of (A:U W:WC)::(U:A W:WC) dinucleotide sequence is unusually high positive and these are probably stabilized by some other interactions 49, 64]. (C:G W:WC)::(C:G W:WC) dinucleotide steps shows good stacking overlap values of 49.8 $Å^2$ (Table 4). The RNA structures, solved by NMR spectroscopy, also show features similar to those observed from analysis of X-ray crystal structures (Table S6) [49].

Canonical base pair steps represent ten unique base pair sequences in case of DNA, but one cannot ignore the G:U base pairs found frequently within double helical regions of RNA. A consideration of G:U W:WC base pairs gives 21 types of unique base pair steps in RNA. Dinucleotide steps containing wobble G:U W:WC base pairs with an adjacent canonical G:C or A:U W:WC base pair also generally follow Calladine's rule [65], i.e., the pyrimidine–purine sequences have larger positive roll and larger negative slide values. Stacking overlap values also follow a trend similar to the steps containing canonical base pairs, i.e., purine–pyrimidine steps have larger base pair overlap values. It may be noted that G:U W:WC base pairs usually have large shear value (around 2 Å), which moves one of the bases within a base pair along the base pair short axis. This gives rise to artificially large or small twist in a dinucleotide step due to rotation of base pair long axis obtained by connecting C8 (purine)–C6 (pyrimidine) atoms of the bases. Thus, twist angles for the (G:C W:WC)::(G:U W:WC), (A:U W:WC)::(G:U W:WC), (U:A W:WC)::(G:U W:WC) and (U:G W:WC)::(G:U W:WC) steps are large as compared to A-form structures (Table 4). On the other hand (G:C W:WC)::(U:G W:WC), (C:G W:WC)::(U:G W:WC), (A:U W:WC)::(U:G W:WC), (U:A W:WC)::(U:G W:WC) steps have twist angles less than 28° due to reverse effect. Stacking overlap values, however, remain similar among these two sets of dinucleotide sequences.

We have also analysed stacking overlap values for dinucleotides containing noncanonical base pairs in RNA (Table 4), which have high frequencies of occurrences in RNA crystal structures. It may be noted here that these sets of dinucleotide structures have similar stacking arrangements, as reflected from small standard deviations (SDs) of all the parameters. Thus, twist value of 6.94° for the (C:G W:WC)::(G:A S:HT), for example, did not arise from averaging of several large and small twist values. It has been observed previously that non-canonical base pairs have anomalous stacking geometry and the inter-base pair parameters also lie outside the usual range. For example, the twist angle values of non-canonical base pairs have been seen to vary within a wide range of ∼5° to ∼90° [5, 6]. Consequently, it can be assumed that noncanonical base pair steps would have a lower stacking overlap value compared to those between canonical base pairs. However, it is found that, in most of the cases, the surface overlap values for dinucleotide steps containing non-canonical base pairs are similar to those of the canonical base pair containing steps, indicating high stacking interaction between them. Among the various types of noncanonical base pair steps, high stacking overlap values are observed for A:G W:WC containing base pair steps, possibly due to larger

Table 4 Mean and SD (in parentheses) values of base step parameters and stacking overlap values in RNA

Base pair step	No. of data points	Tilt	Roll	Twist	Shift	Slide	Rise	Overlap
AA STEP	241	−0.21 (2.66)	6.76 (5.15)	30.93 (3.12)	−0.03 (0.46)	−1.47 (0.53)	3.28 (0.11)	46.10 (3.75)
AC STEP	552	−0.16 (2.66)	5.33 (5.13)	32.50 (3.62)	0.01 (0.54)	−1.37 (0.51)	3.18 (0.14)	48.07 (4.64)
AG STEP	643	0.06 (2.89)	8.46 (4.63)	31.23 (3.36)	0.02 (0.50)	−1.60 (0.49)	3.29 (0.12)	48.14 (4.41)
AU STEP	104	0.12 (2.76)	8.53 (6.13)	31.90 (2.86)	0.02 (0.44)	−1.25 (0.43)	3.21 (0.13)	45.40 (4.82)
CA STEP	630	0.06 (2.51)	10.93 (5.39)	30.00 (2.92)	−0.02 (0.49)	−1.67 (0.31)	3.29 (0.15)	47.25 (4.84)
CC STEP	2,190	0.30 (3.01)	7.60 (4.44)	31.13 (3.35)	0.05 (0.52)	−1.88 (0.37)	3.29 (0.13)	49.80 (4.47)
CG STEP	628	0.06 (2.59)	11.56 (5.84)	29.71 (3.48)	0.00 (0.53)	−1.88 (0.38)	3.33 (0.16)	47.21 (5.68)
GA STEP	640	−0.05 (2.77)	5.53 (5.68)	30.85 (3.36)	0.04 (0.49)	−1.66 (0.43)	3.26 (0.13)	48.20 (4.62)
GC STEP	786	−0.08 (2.67)	3.03 (4.84)	31.90 (3.33)	−0.03 (0.55)	−1.57 (0.50)	3.20 (0.13)	49.18 (4.52)
UA STEP	93	0.18 (2.40)	12.78 (4.83)	29.77 (2.58)	0.06 (0.37)	−1.52 (0.26)	3.30 (0.12)	46.41 (3.89)
G:C W:WC::U:G W:WC	456	0.96 (2.00)	5.71 (4.22)	27.80 (2.87)	0.21 (0.29)	−1.01 (0.33)	3.17 (0.11)	50.47 (4.26)
C:G W:WC::U:G W:WC	333	1.62 (2.43)	5.62 (4.10)	25.70 (3.04)	0.39 (0.44)	−1.63 (0.38)	3.30 (0.11)	49.20 (4.03)
G:C W:WC::G:U W:WC	230	−2.56 (2.86)	8.48 (4.36)	37.54 (3.08)	−0.17 (0.37)	−1.84 (0.43)	3.10 (0.15)	46.86 (4.37)
U:G W:WC::G:C W:WC	120	0.95 (3.45)	12.43 (6.60)	34.99 (5.10)	−0.21 (0.69)	−2.11 (0.70)	3.24 (0.19)	41.83 (5.70)
A:U W:WC::U:G W:WC	120	0.78 (2.26)	5.56 (4.35)	27.27 (2.86)	0.31 (0.31)	−1.17 (0.26)	3.19 (0.10)	49.77 (3.88)
U:A W:WC::U:G W:WC	82	1.33 (2.04)	5.55 (3.38)	26.40 (3.29)	0.26 (0.38)	−1.42 (0.50)	3.31 (0.12)	49.27 (4.62)
A:U W:WC::G:U W:WC	59	−1.05 (3.00)	10.91 (4.02)	35.81 (3.68)	0.15 (0.57)	−1.95 (0.44)	3.14 (0.14)	44.72 (3.46)
U:A W:WC::G:U W:WC	52	−0.69 (3.30)	13.09 (5.44)	36.42 (2.49)	0.14 (0.61)	−1.80 (0.50)	3.20 (0.18)	41.57 (6.21)
U:G W:WC::G:U W:WC	41	0.49 (4.41)	12.02 (6.76)	40.47 (4.11)	0.18 (0.75)	−2.53 (0.46)	3.15 (0.18)	38.68 (3.95)
U:G W:WC::U:G W:WC	40	3.54 (2.48)	8.50 (4.56)	31.77 (2.80)	0.35 (0.35)	−1.78 (0.32)	3.17 (0.10)	46.91 (4.18)
C:G W:WC::G:A S:HT	230	0.74 (1.96)	9.38 (5.63)	6.94 (6.32)	−0.26 (0.57)	−0.87 (0.47)	3.70 (0.29)	47.90 (5.33)
A:G H:ST::C:G W:WC	124	−1.08 (2.01)	9.72 (5.53)	10.88 (5.25)	0.04 (0.61)	−0.57 (0.34)	3.14 (0.32)	46.70 (4.81)
G:A S:HT::A:U H:WT	88	1.91 (5.32)	4.19 (4.43)	84.13 (4.21)	−1.53 (0.37)	−1.39 (0.39)	3.47 (0.31)	39.22 (5.16)
U:G W:WC::G:A S:HT	80	2.87 (2.10)	10.16 (4.90)	15.90 (4.49)	0.24 (0.46)	−0.49 (0.72)	3.64 (0.25)	45.43 (5.08)
C:G W:WC::U:U W:WC	64	−1.42 (3.01)	11.32 (6.84)	27.65 (6.25)	0.26 (0.45)	−1.42 (0.45)	3.41 (0.21)	42.78 (5.76)
G:A S:HT::A:G H:ST	52	−1.09 (10.63)	−0.68 (4.83)	93.75 (7.55)	−0.21 (0.78)	−1.46 (0.49)	2.87 (0.53)	45.89 (5.66)
G:C W:WC::A:G W:WC	51	4.29 (2.70)	4.76 (5.16)	34.18 (3.55)	0.06 (0.57)	−1.46 (0.63)	3.17 (0.13)	51.53 (4.20)
G:C W:WC::G:A W:WC	43	3.35 (2.79)	4.20 (4.37)	30.36 (3.15)	−0.33 (0.29)	−0.61 (0.46)	3.19 (0.14)	51.53 (5.44)
G:A S:HT::A:A H:HT	40	1.25 (3.51)	−2.11 (4.27)	24.06 (2.92)	1.02 (0.59)	0.65 (0.66)	3.57 (0.22)	45.52 (6.04)
U:A W:WC::G:A S:HT	36	1.24 (2.18)	9.13 (4.81)	8.97 (5.74)	0.06 (0.55)	−0.67 (0.55)	3.60 (0.25)	47.22 (5.38)
A:G H:ST::A:G H:ST	35	−3.24 (1.88)	10.11 (3.34)	26.59 (3.00)	−0.75 (0.37)	0.53 (0.35)	4.58 (0.20)	48.56 (3.72)

Table 4 continued

Base pair step	No. of data points	Tilt	Roll	Twist	Shift	Slide	Rise	Overlap
C:G W:WC::G:A s:wT	34	−8.10 (3.61)	29.44 (6.04)	−23.48 (3.00)	−2.73 (0.56)	−3.27 (0.25)	4.16 (0.38)	40.66 (4.18)
A:U W:WC::U:U W:WC	32	−0.32 (2.84)	5.58 (5.83)	29.72 (4.17)	0.36 (0.42)	−1.33 (0.30)	3.25 (0.13)	44.73 (5.18)
G:C W:WC::U:U W:WC	32	−1.09 (3.52)	6.04 (5.75)	27.82 (5.51)	0.34 (0.33)	−0.97 (0.39)	3.33 (0.11)	41.66 (5.38)

surface area of the A:G W:WC base pairs. On the other hand, small overlap values are mostly seen for U:U W:WC base pair containing steps or the (C:G W:WC)::(G:A s:wT) step having very large mean roll values. Smaller overlap values are also observed for base pair steps like (G:A S:HT)::(A:U H:WT) [~ 39 Å^2] as the base pairs are oriented almost perpendicularly to each other [Twist = 84.2°]. Similar high twist values can be observed for (A:G H:ST)::(G:A S:HT) steps [Twist = 93.7°], though they have large overlap area of 45.9 Å^2 because of cross-strand stacking.

Analysis of bulge loops and pseudo continuous helices

A bulge loop is formed by the unpaired residues situated within either or both strands of a double helical stretch of RNA structures. Bulges with one to four unpaired residues in the discontinuity are referred to as bulge loops or bulges while those with five or more residues in the discontinuity are referred to as pseudocontinuous helices or pseudo helices. Earlier report [14] suggests that two types of bulges are possible—one with the discontinuous bases bulged out and the other with the discontinuous bases intercalated into the helix. However, detection and differentiation of such bulges is difficult, as the rise and other inter-base pair parameters between the base pairs surrounding the bulge can have anomalous values and do not properly represent the extent of stacking.

We have calculated stacking overlap between the base pairs surrounding the bulges in the non-redundant set of RNA crystal structures. Tables 5, 6, 7, 8 and 9 contain the structural parameters and overlap values for 1-residue, 2-residue, 3-residue, 4-residue bulge loops and pseudo-continuous helices, respectively. The overall distribution patterns of overlap values show two peaks—one around 0 Å^2 and the other at 45–50 Å^2. On closer analysis of the structures, it is observed that if the unpaired nucleotides, situated in bulge regions, are intercalated into the helix and stacked with the adjacent base pairs, no stacking interaction can be observed between the base pairs residing above and below the bulge, thus giving rise to zero overlap area. The twist, however, may adopt unusual values if non-Watson–Crick base pairs are present. On the other hand, if the bulge residues are looped out of the helical scaffold, the surrounding base pairs maintain A-form like stacking between themselves with regular dinucleotide step parameters, e.g., small tilt, positive roll, negative slide, etc., and the stacking overlap values are similar to those observed for A-form helices. In such cases, the overlap areas between the base pairs in looped out bulges have values around 50 Å^2. Most of the single residue bulges are seen to have good stacking overlap between the base pairs, indicating the unpaired bases generally project out of the helix. On the other hand, stacking overlap between the base pairs surrounding two to four residue bulges are either zero or very small. These bulges probably introduce kinks between the two double helical segments. Thus, stacking overlap values can provide a good measure for the detection of continuation of double helical nature across the junction between two helices when there is a discontinuation in one of the strands at the junction.

Analysis of hairpin loops

As indicated above, there are hairpin loops with varying number of residues in the looped out regions, starting from diloop to significantly longer ones. Many of the diloops are, however, considered as tetraloops where the two bases next to the double helical stem region also form base pairs. We observed 41 hairpin loops with three looped out residues in the non-redundant RNA structural database. These triloops can adopt different conformational folds depending on sequence variation in the loop regions as well as on the nature of closing basepair [38]. There are no triloop sequence with high occurrence frequency, the loop containing UCG sequence capping a U:A W:H T base pair is found eight times, having maximum frequency and all of them are lonepair triloop [39]. In this loop the unpaired Cytosine residue is seen to stack well with the Guanine (Table S7). It is to be noted that the triloop sequences capping a U:A W:H T base pair (11 times), the second residue is seen to stack quite well on the third. The UCU triloop, which is found in three structures, have Watson–Crick base pair at the end and the looped out residues do

Table 5 Mean and SD (in parentheses) values of single residue bulge loops. The symbol "^" represents discontinuity of the strand

Base pair step	Frequency	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
CG/C^G	15	−1.31 (4.71)	10.24 (7.22)	40.62 (12.23)	0.82 (1.56)	−1.65 (0.71)	3.25 (0.19)	44.59 (8.71)
GG/U^C	15	−5.68 (4.25)	8.46 (5.21)	53.84 (5.11)	1.55 (0.57)	−1.76 (0.49)	3.23 (0.17)	41.47 (3.71)
A:U H:WT:^: A:A H:HT	14	−1.60 (3.58)	−6.07 (3.55)	9.86 (3.70)	−1.35 (0.25)	1.85 (0.21)	3.62 (0.28)	46.37 (4.66)
C:G W:WC:^: A:G H:ST	8	−1.12 (3.13)	3.46 (3.48)	76.85 (5.99)	2.39 (0.93)	−0.24 (0.22)	3.56 (0.13)	41.00 (2.91)
AC/G^U	7	−1.40 (5.96)	6.53 (8.78)	44.51 (7.85)	1.51 (0.65)	−0.94 (0.33)	3.22 (0.25)	45.09 (10.62)
CG/U^G	7	−1.44 (5.00)	7.20 (5.00)	59.92 (8.11)	2.85 (0.65)	−1.26 (0.55)	3.29 (0.19)	37.16 (2.58)
GG/C^C	7	−3.99 (4.63)	6.96 (4.13)	44.94 (8.45)	1.33 (0.77)	−1.71 (0.66)	3.22 (0.11)	46.54 (8.07)
UG/C^A	6	−2.24 (2.48)	8.73 (6.25)	43.20 (10.25)	1.24 (1.45)	−2.11 (0.21)	3.28 (0.13)	36.88 (6.27)
U:G W:WC:^: A:A s:hT	5	−21.65 (1.44)	21.04 (3.64)	25.01 (4.01)	1.70 (0.41)	0.23 (0.30)	5.50 (0.27)	13.51 (3.18)
C:G W:WC:^: U:A W:HT	5	1.56 (3.55)	4.46 (4.45)	36.90 (3.45)	−0.28 (0.25)	0.50 (0.14)	3.23 (0.22)	53.03 (3.45)
CG/A^G	5	−4.21 (1.55)	−0.59 (3.75)	37.11 (1.88)	1.94 (0.16)	−3.28 (0.16)	3.36 (0.08)	39.90 (1.06)
UA/U^A	5	−2.73 (7.36)	12.88 (9.62)	33.49 (5.32)	0.81 (0.55)	−1.47 (0.35)	3.39 (0.42)	38.72 (7.18)
UC/G^U	5	−6.90 (2.08)	−3.10 (2.04)	58.83 (3.99)	3.19 (0.42)	−2.99 (0.16)	3.42 (0.15)	25.92 (4.34)
G:A S:HT:^: G:G H:ST	4	−9.09 (18.68)	33.92 (23.58)	65.14 (82.62)	2.04 (0.94)	2.09 (0.72)	4.95 (0.24)	19.24 (1.03)
CC/G^G	4	2.31 (1.73)	6.22 (4.70)	42.25 (3.47)	0.98 (0.78)	−1.47 (0.79)	3.22 (0.16)	46.74 (7.53)
GU/G^C	4	3.14 (0.89)	−0.40 (2.70)	27.71 (3.39)	0.18 (0.22)	−1.33 (0.09)	3.38 (0.09)	47.33 (3.16)
UU/G^G	4	4.53 (2.20)	11.56 (3.65)	36.34 (3.94)	0.67 (0.08)	−1.86 (0.09)	3.19 (0.21)	49.75 (5.06)
UG/C^A	1	−12.29	4.00	32.00	1.86	−4.13	6.76	0.00

Table 6 Mean and SD (in parentheses) values of two residue bulge loops. The symbol "^" represents discontinuity of the strand

Base pair step	Frequency	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
CC/G^G	7	−29.33 (10.37)	11.81 (9.32)	64.27 (5.54)	3.46 (0.68)	−1.27 (0.44)	5.28 (0.73)	15.67 (4.92)
UG/C^G	5	−35.49 (17.03)	25.51 (9.32)	55.18 (17.52)	1.71 (2.31)	−2.15 (0.66)	6.19 (1.85)	9.82 (18.67)
AG/U^U	4	−15.13 (8.55)	56.55 (25.76)	−0.70 (53.89)	−3.42 (5.45)	−1.95 (1.00)	−0.58 (2.62)	16.79 (9.45)
CU/A^G	4	0.78 (1.47)	10.04 (1.91)	11.36 (2.46)	−2.67 (0.48)	−1.79 (0.18)	3.20 (0.16)	37.41 (4.11)
UG/C^G	2	−38.94 (1.61)	29.58 (1.10)	62.77 (2.62)	2.78 (0.11)	−1.72 (0.52)	7.05 (0.02)	0.00 (0.00)

Table 7 Mean and SD (in parentheses) values of three residue bulge loops. The symbol "^" represents discontinuity of the strand

Base pair step	Frequency	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
A:A s:hT:^: C:G W:WC	5	−18.54 (1.02)	−50.16 (3.52)	−50.51 (1.01)	3.15 (0.12)	8.11 (0.08)	3.10 (0.18)	5.79 (1.96)
G:A S:HT:^: A:G w:sT	4	−24.88 (4.69)	21.59 (3.56)	24.81 (1.91)	−1.93 (0.19)	−5.95 (0.14)	3.47 (0.06)	24.14 (1.18)
GG/U^C	4	−15.37 (6.78)	−3.81 (5.04)	91.03 (3.21)	8.51 (0.31)	0.05 (0.09)	4.82 (0.67)	3.06 (1.10)
UC/G^G	4	−70.02 (2.17)	43.74 (5.07)	80.34 (3.65)	3.10 (0.20)	−0.73 (0.27)	9.09 (0.28)	4.98 (1.63)
C:A S:WC:^: G:C W:WC	1	−52.72	−15.67	80.23	2.10	−3.78	6.93	0.00
UC/G^U	1	−73.43	38.19	77.34	2.18	−0.59	9.24	0.00
UG/C^G	1	−21.26	85.89	−47.68	−9.97	−4.91	4.94	0.00

Table 8 Mean and SD (in parentheses) values of four residue bulge loops

Base pair step	Frequency	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
G:A S:HT:^: A:G w:sT	6	−7.39 (1.08)	33.63 (4.30)	0.09 (0.72)	−3.23 (0.22)	−3.76 (0.05)	4.31 (0.20)	24.78 (2.20)
UG/C^A	3	−14.08 (1.46)	61.30 (0.53)	1.53 (0.38)	−9.11 (0.30)	−2.51 (0.37)	3.79 (0.49)	0.00 (0.00)
CG/C^G	1	−33.67	−84.51	13.79	−0.39	−1.27	13.32	0.00
UG/U^A	1	82.70	28.81	41.26	−3.75	10.41	1.88	0.00

not have any stacking interaction. We also observed few other triloop structures, such as AAU, which show no stacking between single stranded residues. On the other hand the UCA lonepair triloop occurs four times, but their looping residues do not show any consistent stacking interaction. The ACG triloop is seen in two structures, one of them appears after a G:A S:H T base pair and the third residue seems to stack on top of the 3′ base pair. We found two triloop sequences, namely UGA and UUG, both appear after a G:A s:w T base pair stabilized by C–H…O/N hydrogen bond. Similarly the AGC triloop has A:A H:W T base pair as the closing one. The unpaired residues in the UGA triloop do not stack with each other while the first Uracil residue of the UUG triloop stack reasonably well with the s:w T base pair. Three other loops, AGU and UUC, showed no consistency regarding stacking parameter within them. As frequency of observation of these triloops are quite small, a statistically meaningful conclusion is difficult to arrive at but hopefully on availability of more RNA structures one would find important features of these motifs.

Among the tetraloop sequences, GNRA loops are the most commonly observed type from our non-redundant dataset of RNA crystal structures. We have also identified UNCG (YNMG), CUYG, UGNN, GANC and GNAR tetraloop families from base pairing features, although their occurrences are very small (Tables 10, 11, 12; Supplementary Information Table S8). GNRA tetraloops can either be of paired type (Table 10) or unpaired type (Table 11). In the paired type of GNRA tetraloops, the first and fourth nucleotides in the loop region form a G:A S:HT base pair. These loops cap a double helix where the last base pair is generally a G:C W:WC type. We observed high stacking overlap between the G:C base pair and the G:A S:HT base pair in paired type of GNRA loops [mean stacking overlap 48.8(4.7)]. According to IUPAC-IUB suggestion [46] the values of all the parameters, including stacking overlap, remain same when one looks from two chains of the double helix, even containing the non-Watson–Crick base pair (G:A S:HT, for example). In case of unpaired GNRA tetraloops, large stacking overlap is also seen between the terminal G:C base pair and the unpaired Guanine residue (G_n). The A_{n+3}, however, does not stack well with the terminal Watson–Crick base pair. Usually, GNRA tetraloops have a distinctive fold, where one base is stacked with the 5′-strand and the other three are stacked in the 3′-strand. Thus, the G_n and A_{n+3} residues have high overlap values with the stem region, even when these are not paired (Mean stacking overlap 30.4 and 9.0). The N_{n+1}, R_{n+2} and A_{n+3} are stacked consecutively with the 3′-strand, giving rise to considerable amount of overlap area between them. The single base overlap areas between N_{n+1} and R_{n+2} residues, as well as between R_{n+2} and A_{n+3} residues for unpaired types, lie within the range of 25–30 $Å^2$, indicating good stacking. The values of the other parameters, such as tilt, roll, etc., vary considerably at these single stranded regions, as expected and are reflected in their large SD values. The stacking overlap between G_n and N_{n+1} is close to zero, as there is a U-turn in the backbone direction between these residues. However, in one case (paired GAAA sequence; 2ZJR; 623–626), a G:A s:sT base pair is observed (Supplementary Information Table S8-A). Moreover, no stacking overlap is observed between the second, third and fourth nucleotides in the loop in this case, which appears to be a deviation from the usual GNRA loop structure. Thus the sequence of the hairpin loop alone is not sufficient to classify the tetraloop.

It was hypothesized that YNMG tetraloops are an extension of UNCG sequences [31]. Usually, the UNCG tetraloops have a C:G W:WC closing base pair in the stem (Table 12), whereas YNMG loops can have either a C:G W:WC or a U:G W:WC closing base pair. These tetraloops have good stacking between the closing base pair and the U_n and also between G_{n+3} and the closing base pair. However, we noticed an example of UNCG tetraloop in 3AM1 (Supplementary Information Table S8-B) where good stacking is present also between N_{n+1}, C_{n+2} and G_{n+3} residues. On the other hand, the YNMG loops are found to have C:G S:WT base pairing between Y_n and G_{n+3} (Supplementary Information Table S8-C) and their stacking overlap values are also quite different from the above. This perhaps indicates they are different types of loops. Structural data about all the other tetraloops are quite rare and are given in Supplementary Information (Table S8). Due to less frequency of these structures, conclusive statements cannot be made but their structural properties can be

Table 9 Mean and SD (in parentheses) values of pseudo continuous helices. The symbol "^" represents discontinuity of the strand

Base pair step	Frequency	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
CC/G^G	43	−2.05 (7.57)	10.47 (8.23)	36.07 (10.63)	0.40 (1.14)	−1.72 (0.51)	3.40 (0.49)	45.50 (9.88)
CG/C^G	23	24.76 (16.12)	2.85 (37.83)	−5.37 (60.12)	−4.46 (8.27)	7.55 (9.27)	3.03 (2.03)	0.00 (0.00)
GG/C^C	16	−0.68 (4.49)	10.76 (7.47)	34.00 (8.53)	0.68 (1.30)	−2.03 (0.64)	3.22 (0.22)	45.86 (3.32)
CG/A^G	14	−16.25 (3.63)	9.31 (6.22)	60.92 (4.06)	3.08 (0.62)	−2.19 (0.32)	3.81 (0.19)	34.76 (3.28)
GU/A^C	13	−3.81 (10.61)	4.33 (5.85)	38.03 (5.58)	0.48 (0.65)	−1.14 (0.51)	3.40 (0.81)	43.46 (11.62)
AC/G^U	10	1.15 (2.04)	2.58 (3.38)	44.90 (2.53)	1.43 (0.22)	−0.80 (0.32)	3.15 (0.05)	46.73 (3.49)
GC/G^C	9	−0.10 (4.30)	−0.03 (5.81)	42.63 (11.07)	1.10 (1.53)	−0.74 (0.90)	3.14 (0.12)	43.34 (7.37)
UC/G^A	9	−2.50 (13.12)	12.54 (6.28)	42.97 (10.04)	1.61 (0.57)	−2.03 (0.84)	3.42 (0.92)	42.78 (11.84)
GC/G^C	8	43.02 (9.46)	23.30 (10.64)	45.87 (−11.72)	2.96 (−4.88)	11.13 (7.99)	4.28 (0.77)	0.00 (0.00)
CA/U^G	7	−0.42 (2.43)	3.06 (3.52)	8.63 (1.94)	−1.54 (0.22)	−3.34 (0.21)	3.40 (0.30)	41.82 (5.60)
CG/C^G	7	2.51 (4.33)	8.94 (8.70)	27.09 (19.86)	0.09 (1.77)	−2.62 (1.26)	3.22 (0.06)	43.96 (5.43)
UC/G^G	7	1.96 (2.56)	8.51 (7.36)	40.56 (1.77)	0.24 (0.33)	−1.54 (1.56)	2.94 (0.36)	46.53 (2.09)
C:G W:WC:^:A:G H:ST	6	−12.19 (9.14)	30.95 (20.53)	34.88 (41.66)	−4.08 (6.20)	−1.55 (0.90)	−0.74 (3.32)	14.78 (18.72)
CG/U^G	6	−3.65 (5.66)	11.68 (3.64)	61.63 (13.14)	2.87 (1.48)	−1.31 (0.54)	3.46 (0.19)	34.87 (5.63)
UA/U^U	6	−18.68 (4.60)	0.15 (4.41)	64.62 (3.10)	2.52 (0.32)	−3.29 (0.29)	3.88 (0.19)	24.00 (4.50)
UC/G^G	6	−34.74 (36.67)	0.28 (35.33)	−12.45 (48.85)	−6.20 (3.58)	2.10 (10.25)	8.91 (3.05)	0.00 (0.00)
UG/C^G	6	−20.26 (46.59)	45.22 (38.36)	−1.53 (28.37)	−6.52 (2.45)	6.87 (12.60)	1.87 (4.25)	0.00 (0.00)
A:A H:WT:^:U:A W:WC	5	0.83 (0.79)	−20.46 (1.58)	−96.83 (0.23)	−10.16 (0.14)	−2.97 (0.07)	4.20 (0.13)	4.16 (0.66)
C:G W:WC:^:A:C +:WC	5	−0.88 (3.32)	10.54 (3.35)	71.29 (1.73)	4.10 (0.25)	−1.68 (0.41)	3.26 (0.11)	31.08 (0.83)
CU/A^G	5	1.14 (1.68)	4.05 (4.35)	43.64 (8.99)	1.07 (0.92)	−1.39 (0.19)	3.24 (0.11)	42.45 (5.82)
AG/C^G	5	31.79 (18.35)	34.67 (18.47)	66.09 (13.51)	−2.80 (6.48)	13.02 (7.77)	5.62 (0.60)	0.00 (0.00)
A:A s:hT:^:A:G H:ST	4	0.87 (1.27)	12.79 (3.93)	−47.47 (0.79)	0.85 (0.19)	4.97 (0.07)	2.94 (0.41)	24.27 (3.16)
G:A S:HT:^:C:G W:WC	4	−15.74 (4.16)	−71.03 (3.94)	−56.73 (0.99)	3.83 (0.46)	8.13 (0.23)	2.62 (0.40)	6.48 (3.13)
AG/C^U	4	0.89 (1.07)	13.76 (6.94)	37.57 (5.41)	0.96 (0.71)	−1.53 (0.57)	3.35 (0.07)	43.84 (3.67)
CA/G^G	4	−8.83 (17.13)	7.71 (6.03)	50.99 (18.24)	2.01 (1.99)	−1.85 (0.44)	3.64 (0.80)	41.75 (18.53)
GA/G^C	4	9.59 (2.04)	15.40 (1.28)	46.57 (1.87)	2.04 (0.11)	−1.05 (0.17)	3.21 (0.06)	43.51 (2.69)
G:A S:HT:^:A:U W:WC	4	8.41 (7.05)	7.83 (15.17)	−40.36 (74.50)	−11.34 (0.68)	−3.57 (4.54)	0.44 (1.88)	0.00 (0.00)

Table 10 Mean and SD (in parentheses) values of paired GNRA tetraloops (frequency = 20)

Base pair	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
X:X	−0.24 (2.3)	3.64 (5.4)	29.80 (3.6)	−0.15 (0.4)	−1.84 (0.6)	3.33 (0.2)	46.01 (6.5)
X:X	−0.56 (3.5)	6.49 (8.9)	−4.63 (4.1)	−0.89 (0.7)	−1.39 (0.4)	3.56 (0.2)	45.84 (4.7)
G:A S:HT	−15.70 (153.8)	9.51 (21.3)	3.54 (24.7)	−1.47 (1.2)	1.57 (3.7)	−0.33 (3.7)	0.00 (0.0)
N:	6.59 (28.3)	0.36 (24.5)	46.98 (29.8)	1.94 (1.6)	−1.33 (2.3)	3.09 (0.3)	23.70 (5.5)
R:	−39.30 (69.8)	42.32 (44.2)	55.67 (42.5)	2.38 (3.5)	0.61 (0.8)	3.77 (2.2)	27.43 (4.5)
A:G H:ST	0.56 (3.5)	6.49 (8.9)	−4.63 (4.1)	0.93 (0.8)	−1.41 (0.4)	3.65 (0.8)	45.84 (4.7)
X:X	0.24 (2.3)	3.64 (5.4)	29.80 (3.6)	0.15 (0.4)	−1.84 (0.6)	3.33 (0.2)	46.01 (6.5)
X:X	1.07 (3.4)	4.84 (5.1)	29.62 (8.9)	−0.18 (1.1)	−1.63 (0.8)	3.32 (0.1)	43.32 (11.4)

Table 11 Mean and SD (in parentheses) values of unpaired GNRA tetraloops (frequency = 39)

Base pair	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
X:X	−0.33 (2.2)	7.23 (6.1)	28.19 (6.3)	0.03 (0.4)	−1.56 (0.6)	3.33 (0.2)	45.22 (5.4)
X:X	6.96 (5.6)	9.94 (11.8)	43.36 (28.5)	0.80 (1.6)	−1.01 (0.6)	3.47 (0.8)	30.45 (8.2)
G:	98.40 (37.4)	−58.82 (42.3)	21.04 (21.6)	−4.84 (2.4)	6.76 (2.3)	1.79 (2.3)	1.17 (4.3)
N:	3.88 (25.1)	−5.56 (21.8)	46.08 (31.8)	1.09 (2.6)	−0.74 (3.0)	3.03 (0.9)	20.59 (8.4)
R:	8.22 (27.5)	2.21 (18.4)	32.32 (19.6)	0.35 (1.6)	−0.41 (1.5)	2.61 (2.0)	22.53 (7.5)
A:	19.84 (44.5)	−39.40 (98.7)	10.10 (29.6)	2.47 (3.1)	−1.07 (2.4)	−1.60 (6.1)	9.00 (6.1)
X:X	0.33 (2.2)	7.23 (6.1)	28.19 (6.3)	−0.03 (0.4)	−1.56 (0.6)	3.35 (0.2)	45.22 (5.4)
X:X	1.17 (5.0)	8.52 (5.9)	33.71 (12.1)	0.22 (0.6)	−1.47 (1.1)	3.27 (0.2)	44.98 (8.6)

Table 12 Mean and SD (in parentheses) values of unpaired UNCG tetraloops (frequency = 25)

Base pair	Tilt mean (SD)	Roll mean (SD)	Twist mean (SD)	Shift mean (SD)	Slide mean (SD)	Rise mean (SD)	Overlap mean (SD)
X:X	0.23 (2.7)	4.37 (6.2)	30.79 (3.2)	0.15 (0.5)	−1.86 (0.4)	3.31 (0.1)	47.97 (5.7)
X:X	14.48 (23.7)	0.97 (8.5)	31.75 (9.3)	0.84 (0.8)	0.92 (0.8)	2.97 (0.7)	25.60 (3.8)
U:	28.45 (54.7)	−56.86 (72.4)	1.82 (42.4)	−7.28 (2.0)	7.06 (3.9)	4.14 (3.7)	0.00 (0.0)
N:	−36.22 (39.0)	−67.41 (66.0)	−11.40 (35.9)	1.19 (5.5)	7.24 (5.4)	3.82 (3.1)	0.95 (4.7)
C:	−1.38 (39.3)	7.56 (29.5)	34.10 (17.5)	−0.83 (2.4)	3.53 (1.8)	−2.86 (2.4)	2.64 (4.6)
G:	10.65 (64.1)	5.48 (40.0)	17.43 (13.6)	−1.51 (2.4)	−2.71 (3.8)	−1.77 (2.8)	12.99 (5.1)
X:X	−0.23 (2.7)	4.37 (6.2)	30.79 (3.2)	−0.15 (0.5)	−1.86 (0.4)	3.31 (0.1)	47.97 (5.7)
X:X	1.87 (5.6)	7.19 (7.0)	32.87 (8.0)	0.49 (0.9)	−1.60 (0.8)	3.28 (0.3)	46.12 (9.3)

understood from these values. Their detail study using different other methods can be presented elsewhere.

We observed many other families of tetraloops, such as CUYG, UGNN, GNAR, UNAC, GANC and GYYA, from base pairing analysis by BPFIND in our non-redundant database of functional RNA crystal structures. It should be noted that BPFIND detects a base pair only when it finds possibility of at least two hydrogen bonds between the bases with maximum one hydrogen bond involving sugars. Hence, there is a possibility that some loop detected by other methods is not found in our analysis. It is seen that in all the four structures of UUAG tetraloop the U_{n+1} stacks reasonably well with the U_{n+2} residue. The two structures of CUYG tetraloop family, however, do not show any structural similarity in terms of stacking overlap (Table S8-D) between the looped out bases. Two structures each of GNAR and UNAC tetraloop types show similarity in stacking patterns indicating these can be clubbed together. The looped out residues of single representatives of GANC show good stacking between them. On the other hand the looped out residues of single representative of GYYA tetraloop family have no stacking (Table S8-I). We also observed structures of tetraloops having UAGC, AUUC, AUUU, CUUU, GAAG, UAUU, GAAU, GAGU, GUAC,

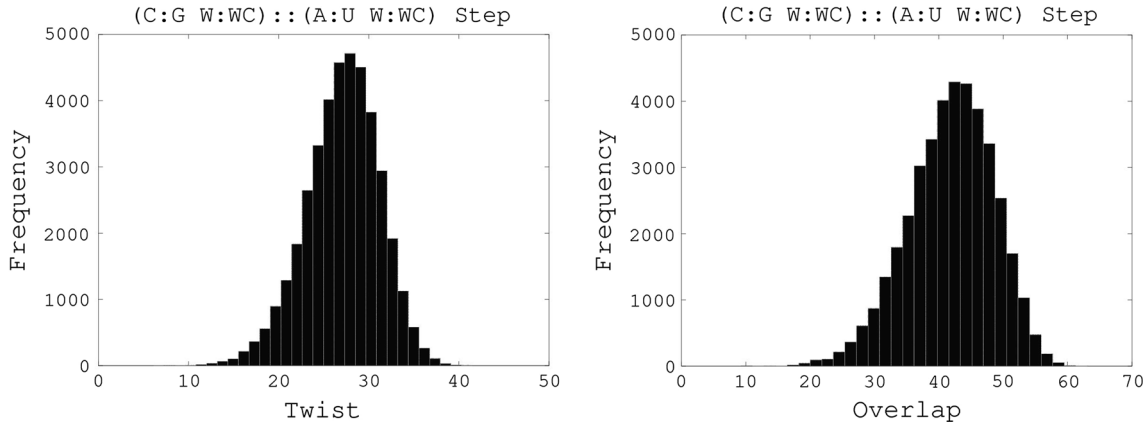

Fig. 3 Frequency distribution for twist and overlap for the canonical step (C:G W:WC)::(A:U W:WC) (eighth step) in 2AW4

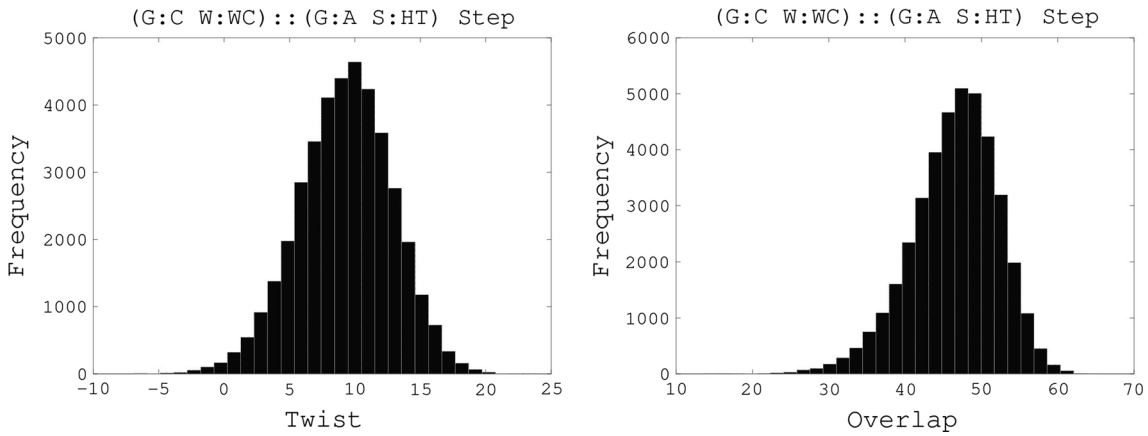

Fig. 4 Frequency distribution for twist and overlap for the non-canonical step (G:C W:WC)::(G:A S:HT) (fourth step) in 2AW4

GUCU, UAAU, GAAC and GCCA sequences but statistically meaningful conclusions can not be generated from their lone examples.

Similarly we observed 19 structures of double helices with loops having five residues, which can be termed as pentaloop (Table S9). Considering the possible number of sequences of five residues and configurational entropy of such structures, we could not carry out systematic structural analysis of these loops from the small number of available RNA structures.

Double helices with non Watson–Crick base pairs

As indicated earlier, there are quite a few double helices in different functional RNA molecules with one to three non-canonical base pairs flanked by regular G:C, A:U or G:U W:WC base pairs [40] and the steps involving non-canonical base pairs often adopt unusual twist. Hence a proper analysis of their stacking preferences becomes difficult. We have analysed overlap values from 50 ns long molecular dynamics simulation trajectories reported in our previous report [6] of these double helices using NUPARM. The distributions of twist as well as stacking overlap values for the canonical base pair containing steps from these MD snapshots for the helix from 2AW4 (Fig. 3) show high similarity with the values given in Table 3. Although twist values of the non-canonical base pair containing steps are around 10° (for the fourth step) (Fig. 4) and around 85 (for the fifth step) (Fig. 5), their stacking overlaps are always very high (Figs. 4,5).

Conclusion

We have developed a novel method for calculation of stacking overlap, which is not merely a geometric overlap, between nucleic acid bases and base pairs appearing in DNA or RNA structures. Analyses of crystal structures in terms of stacking overlap indicate that base pair stacking also follow a sequence directed feature—the purine–pyrimidine sequences prefer to have better stacking in all

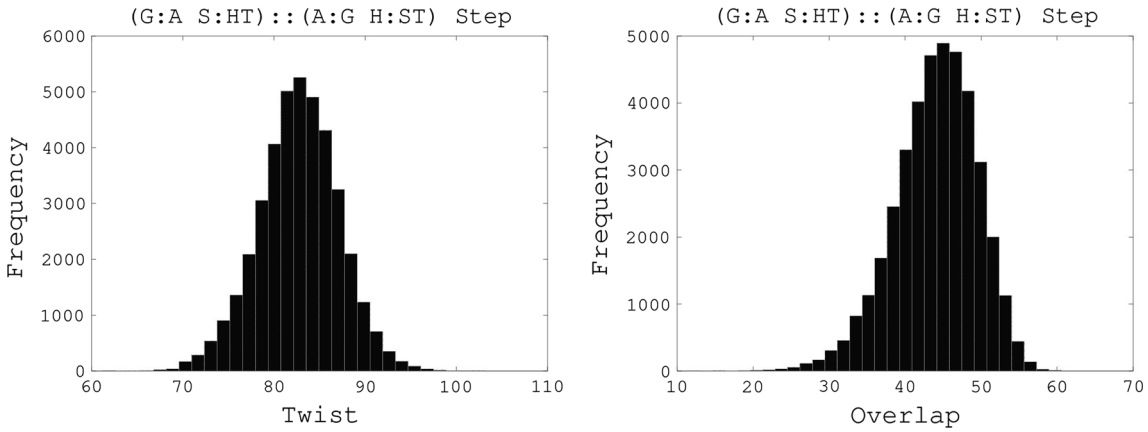

Fig. 5 Frequency distribution for twist and overlap for the non-canonical step (G:A S:HT)::(A:G H:ST) (fifth step) in 2AW4

kinds of double helical structures. The stacking overlap is thus correlated to some of the base pair step orientation parameters, such as roll, slide, twist, etc., although the correlation coefficients are not always statistically significant. Statistically significant correlation between rise and stacking is observed in all types of structures, as expected. These values have the capability to detect kinks in RNA double helices appearing due to bulge loops. Stacking and hence stabilities of hairpin loops can be easily characterised by this parameters, which was only qualitatively detected from molecular graphics visualization. There are now huge number of databases and servers for detection of unusual structural motifs in RNA three-dimensional fold [7, 8]. We believe that our methodology can assist such finding with great success rate.

Non-canonical base pairs appear quite frequently in RNA and when they stack with a canonical or another non-canonical base pair, it gives special features to the double helix in terms of molecular recognition. We found that stacking overlap in dinucleotide steps containing non-canonical base pairs are also quite similar to those found between two canonical base pairs. Hence it should now be possible to quantitatively evaluate stacking free energy differences from MD simulation of RNA for various sequences.

Acknowledgments We are thankful to Prof. Manju Bansal, Indian Institute of Science, Bangalore, India for discussions. This work was partially supported by Department of Biotechnology, Govt. of India. PKP is thankful to National Institute of Pharmaceutical Education and Research, Kolkata for supports.

References

1. Neidle S (2002) Nucleic acid structure and recognition. Oxford University Press, Oxford
2. Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. RNA 7:499–512
3. Šponer JE, Špačková N, Leszczynski J, Šponer J (2005) Principles of RNA base pairing: structures and energies of the trans Watson – Crick/sugar edge base pairs. J Phys Chem B 109:11399–11410
4. Roy A, Panigrahi S, Bhattacharyya M, Bhattacharyya D (2008) Structure, stability, and dynamics of canonical and noncanonical base pairs: quantum chemical studies. J Phys Chem B 112:3786–3796
5. 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
6. 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
7. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. J Math Biol 56:215–252
8. Petrov AI, Zirbel CL, Leontis NB (2013) Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. RNA 19:1327–1340
9. Woese CR, Gutell RR (1989) Evidence for several higher order structural elements in ribosomal RNA. Proc Natl Acad Sci USA 86:3119–3122
10. Varani L, Hasegawa M, Spillantini MG, Smith MJ, Murrell JR, Ghetti B, Klug A, Goedert M, Varani G (1999) Structure of tau exon 10 splicing regulatory element RNA and destabilization by mutations of frontotemporal dementia and parkinsonism linked to chromosome 17. Proc Natl Acad Sci USA 96:8229–8234
11. Portmann S, Grimm S, Workman C, Usman N, Egli M (1996) Crystal structures of an A-form duplex with single-adenosine bulges and a conformational basis for site-specific RNA self-cleavage. Chem Biol 3:173–184
12. Hermann T, Patel DJ (1999) Stitching together RNA tertiary architectures. J Mol Biol 294:829–849
13. Hermann T, Westhof E (1999) Non-Watson–Crick base pairs in RNA–protein recognition. Chem Biol 6:R335–R343
14. Chastain M, Tinoco I (1991) Structural elements in RNA. Prog Nucleic Acid Res Mol Biol 41:131–177
15. Woese CR, Winkers S, Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of tetra-loops. Proc Natl Acad Sci USA 87:8467–8471
16. Heus HA, Pardi A (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. Science 253:191–194

17. Jucker FM, Pardi A (1995) GNRA tetraloops make a U-turn. RNA 1:219–222
18. Leontis NB, Westhof E (2002) The annotation of RNA motifs. Comp Funct Genomics 3:518–524
19. Correll CC, Swinger K (2003) Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4Å resolution. RNA 9:355–363
20. Tuerk C, Gauss P, Thermes C, Groebe DR, Gayle M, Guild N, Stormo G, D'aubenton-Carafa Y, Uhlenbeck OC, Tinoco I Jr, Brody EN, Gold L (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. Proc Natl Acad Sci USA 85:1364–1368
21. Cheong C, Varani G, Tinoco I Jr (1990) Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. Nature 346:680–682
22. Baumruk V, Gouyette C, Huynh-Dinh T, Sun JS, Ghomi M (2001) Comparison between CUUG and UUCG Tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy. Nucleic Acids Res 29:4089–4096
23. Convery MA, Rowsell S, Stonehouse NJ, Ellington AD, Hirao I, Murray JB, Peabody DS, Phillips SE, Stockley PG (1998) Crystal structure of an RNA aptamer–protein complex at 2.8 Å resolution. Nat Struct Biol 5:133–139
24. Rowsell S, Stonehouse NJ, Convery MA, Adams CJ, Ellington AD, Hirao I, Peabody DS, Stockley PG, Phillips SEV (1998) Crystal structures of a series of RNA aptamers complexed to the same protein target. Nat Struct Biol 5:970–975
25. Klosterman PS, Hendrix DK, Tamura M, Holbrook SR, Brenner SE (2004) Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns. Nucleic Acids Res 32:2342–2352
26. Butcher SE, Dieckmann T, Feigon J (1997) Solution structure of the conserved 16S-like ribosomal RNA UGAA tetraloop. J Mol Biol 268:348–358
27. Lebars I, Lamontagne B, Yoshizawa S, Elela SA, Fourmy D (2001) Solution structure of conserved AGNN tetraloops: insights into Rnt1p RNA processing. EMBO J 20:7250–7258
28. Wu H, Yang PK, Butcher SE, Kang S, Chanfreau G, Feigon J (2001) A novel family of RNA tetraloop structures forms the recognition site for *Saccharomyces cerevisiae* RNaseIII. EMBO J 20:7240–7249
29. Woese CR, Gutell R, Gupta R, Noller HF (1983) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol Rev 4:621–669
30. Antao VP, Tinoco I Jr (1992) Thermodynamic parameters for loop formation in RNA and DNA hairpin tetraloops. Nucleic Acids Res 20:819–824
31. Proctor DJ, Schaak JE, Bevilacqua JM, Falzone CJ, Bevilacqua PC (2002) Isolation and characterization of a Family of Stable RNA Tetraloops with the motif YNMG That Participate in Tertiary Interactions. Biochemistry 41:12062–12075
32. Keating KS, Toor N, Pyle AM (2008) The GANC tetraloop: a novel motif in the group IIc intron structure. J Mol Biol 383:475–481
33. Zhao Q, Huang HC, Nagaswamy U, Xia Y, Gao X, Fox GE (2012) UNAC tetraloops: to what extent do they mimic GNRA tetraloops? Biopolymers 97:617–628
34. Melchers WJG, Zoll J, Tessari M, Bakhmutov DV, Gmyl AP, Agol VI, Heus HA (2006) A GCUA tetranucleotide loop found in poliovirus oriL by in vivo SELEX (un)expectedly forms a YNMG-like structure: extending the YNMG family with GYYA. RNA 12:1671–1682
35. Zwieb C (1992) Conformity of RNAs that interact with tetranucleotide loop binding proteins. Nucleic Acids Res 20:4397–4400
36. Sakamoto T, Morita S, Tabata K, Nakamura K, Kawai G (2002) Solution structure of a SRP 19 binding domain in human SRP RNA. J Biochem 132:177–182
37. Varani G, Cheong C, Tinoco I Jr (1991) Structure of an unusually stable RNA hairpin. Biochemistry 30:3280–3289
38. Lisi V, Major F (2007) A comparative analysis of the triloops in all high-resolution RNA structures reveals sequence–structure relationships. RNA 13:1537–1545
39. Lee JC, Cannone JJ, Gutell RR (2003) The lonepair triloop: a new motif in RNA structure. J Mol Biol 325:65–83
40. Halder S, Bhattacharyya D (2013) RNA structure and dynamics: a base pairing perspective. Prog Biophys Mol Biol 113:264–283
41. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, Headd JJ, Richardson DC, Ham D, Hershkovits E, Williams LD, Keating KS, Pyle AM, Micallef D, Westbrook J, Berman HM, RNA Ontology Consortium (2008) RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). RNA 14:465–481
42. Malathi R, Yathindra N (1985) Backbone conformation in nucleic acids: an analysis of local helicity through heminucleotide scheme and a proposal for a unified conformational plot. J Biomol Struct Dyn 3:127–144
43. Duarte CM, Pyle AM (1998) Stepping through an RNA structure: a novel approach to conformational analysis. J Mol Biol 284:1465–1478
44. Duarte CM, Wadley LM, Pyle AM (2003) RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. Nucleic Acids Res 31:4755–4761
45. Wadley LM, Keating KS, Duarte CM, Pyle AM (2007) Evaluating and learning from RNA pseudotorsional space: quantitative validation of a reduced representation for RNA structure. J Mol Biol 372:942–957
46. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, Sklenar H, Suzuki M, Tung CS, Westhof E, Wolfberger C, Berman HM (2001) A standard reference frame for the description of nucleic acid base-pair geometry. J Mol Biol 313:229–237
47. Samanta S, Mukherjee S, Chakrabarti J, Bhattacharyya D (2009) Structural properties of polymeric DNA from molecular dynamics simulations. J Chem Phys 130:115103
48. Marathe A, Bansal M (2011) An ensemble of B-DNA dinucleotide geometries lead to characteristic nucleosomal DNA structure and provide plasticity required for gene expression. BMC Struct Biol 11:1
49. Kailasam SK, Bhattacharyya D, Bansal M (2014) Sequence dependent variations in RNA duplex are related to non-canonical hydrogen bond interactions in dinucleotide steps. BMC Res Notes 7:83
50. Lu XJ, 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
51. Gabb HA, Sanghani SR, Robert CH, Prévost C (1996) Finding and visualizing nucleic acid base stacking. J Mol Graph 14:6–11
52. Ray SS, Halder S, Kaypee S, Bhattacharyya D (2012) HD-RNAS: an automated hierarchical database of RNA structures. Front Genet 3:59
53. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
54. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B (1992) The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 63:751–759

55. 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
56. Bansal M, Bhattacharyya D, Ravi B (1995) NUPARM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures. CABIOS 11:281–287
57. 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
58. Banerjee R, Sen M, Bhattacharyya D, Saha P (2003) The jigsaw puzzle model: search for conformational specificity in protein interiors. J Mol Biol 333:211–226
59. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117:5179–5197
60. Chandrasekaran R, Wang M, He R-G, Puigjaner LC, Byler MA, Millane RP, Arnott S (1989) A re-examination of the crystal structure of A-DNA using fiber diffraction data. J Biomol Struct Dyn 6:1189–1202
61. Chandrasekaran R, Arnott S (1996) The structure of B-DNA in oriented fibers. J Biomol Struct Dyn 13:1015–1027
62. Arnott S, Hukins DWL, Dover SD (1972) Optimised parameters for RNA double-helices. Biochem Biophys Res Commun 48:1392–1399
63. Zhang F, Sahu B, Min H, MacDonald AH (2010) Bond structure of ABC-stacked graphene trilayers. Phys Rev B 82:035409
64. Mukherjee S, Kailasam SK, Bansal M, Bhattacharyya D (2014) Energy hyperspace for stacking interaction in AU/AU dinucleotide step: dispersion-corrected density functional theory study. Biopolymers 101:107–120
65. Calladine CR (1982) Mechanics of sequence-dependent stacking of bases in B-DNA. J Mol Biol 25:343–352