A Quantitative Measure of Chirality Inside Nucleic Acid Databank

ADRIANA PIETROPAOLO1,2,3* AND MICHELE PARRINELLO1
1Computational Science, Department of Chemistry and Applied Biosciences, ETH Zürich, USI Campus, Lugano, Switzerland
2Dipartimento di Scienze Chimiche, Università di Catania, V.le A. Doria 6, Catania, Italy
3Dipartimento di Scienze Farmacobiologiche, Università Magna Graecia di Catanzaro, Campus Universitario, Viale Europa, Catanzaro, Italy

ABSTRACT We show the capability of a chirality index (Pietropaolo et al., Proteins 2008;70:667–677) to investigate nucleic acid structures because of its high sensitivity to helical conformations. By analyzing selected structures of DNA and RNA, we have found that sequences rich in cytosine and guanine have a tendency to left-handed chirality, in contrast to regions rich in adenine or thymine which show strong negative, right-handed, chirality values. We also analyze RNA structures, where specific loops and hairpin motifs are characterized by a well-defined chirality value. We find that in nucleosome the chirality is exalted, whereas in ribosome it is reduced. Our results illustrate the sensitivity of this descriptor for nucleic acid conformations. Chirality 00:000–000, 2011. © 2011 Wiley-Liss, Inc.

KEY WORDS: DNA; RNA; supramolecular order; guanine; left-handed

INTRODUCTION

Besides the canonical B-DNA double helixes, DNA is known to adopt a variety of structures,1–10 which include bent,11 triplex,12 and quadruplex12,13 conformations. Also RNA can take a variety of structural motifs and beside its typical A-form,14 exhibits loops and hairpins,15–18 with many relevant motifs like loop-E,19 loop-C,20 and the kink-turn.21 These motifs share the feature of containing noncanonical base pairs, like A-G or G-U in loop-E19 or G-A base pairs in the second helical stem of the kink-turn.22

Such structural plasticity suggests a packing strategy essential to their supramolecular functionality. Classifying these structures, relating them to the nucleotide sequence and the nucleic acid function, is an important task. Much effort has been devoted to the determination of the secondary structures of nucleic acids,23–30 their evolution,31 and their folding.32–37

Previous works of DNA23,24 were based on the description of the backbone structure through the torsional angle space or through a specific restriction of it, as the virtual bond formalism proposed by Olson.25 Also for RNA, Richardson et al. have recently reported a classification of RNA backbone structures based on torsional angles.36 Also based on torsional angles, Wadley and Pyle have identified recurring RNA motifs37 and simplified the description by projecting it on a Ramachandran-like plot. The use of torsional space to describe nucleic acid structures is accurate but requires at least six backbone angles for each base pair along the chain.38 This has to be compared with proteins where only two backbone dihedrals per residue are necessary.

Recently, a successful classification of protein structures based on a chirality index has been introduced.39 Protein chirality is particularly effective in discriminating helix handedness, which is negative for right-handed helices and positive for left-handed ones and can discriminate between β turn and hairpin motifs.

In this study, we will use the recently proposed local chirality index,39 with the aim of providing a quantitative answer to how nucleic acid conformations can be represented in a low-dimensional space.

We here explore the use of this index to classify nucleic acid structures. Given the intrinsic chiral structure of nucleic acids, a chirality index appears as a natural descriptor. We have analyzed X-ray and NMR structure database of DNA and RNA, focusing on structural motifs and their relation to specific nucleotide sequence. Sequences rich in cytosine and especially guanine adopt small or slightly positive chirality values, which implies that consecutive guanines favor flat or left-handed chirality regions. In contrast, sequences rich in adenine or thymine show strong negative values, consistent with their tendency to form narrow DNA grooves.40 Moreover, RNA hairpins and loops can be identified by a specific negative chirality value. The stronger helix character of DNA relative to RNA is reflected in the more negative chirality, which is exalted in nucleosome and reduced in ribosome. In all cases, investigated the chirality index appears as a very sensitive descriptor of local conformations.

To the best of our knowledge, this is the first time that the specific nucleotide chirality is used to define nucleic acids structures.

METHODS

Chirality Calculation on Nucleic Acids

A simple descriptor of chirality can be expressed as a pseudoscalar combination of three molecular vectors, akin to the calculation of a dihedral angle. Osipov et al.31 proposed a pseudoscalar quantity useful for determining molecular handedness41 and a subsequent scaled version was introduced to compare molecules of different sizes.42 These former studies considered the total molecular chirality, with a single value which identified a specific molecule. A recent description of chirality was

© 2011 Wiley-Liss, Inc.
proposed for analyzing protein structures,\textsuperscript{39} representing chirality as a local property of molecular backbone and expressed as follows:

\[ G_m = \frac{4!}{3N^4} \sum_{\text{all permutations}} \left( \frac{\{r_{ij} \times r_{kl}\} \cdot r_{il} \{r_{ij} \cdot r_{lk}\} \{r_{jk} \cdot r_{kl}\}}{r_{ij} r_{jk} r_{kl} r_{il}} \right) \quad (1) \]

where \(i, j, k,\) and \(l\) are four of the \(N\) atoms belonging to a given molecular fragment. The permutation of \(i, j, k, l\) atoms are required to preserve the translation invariance of chirality, whereas the normalization factor preserves the size invariance, to be only a shape-dependent quantity. Analogously with the previous formalism of virtual bond representation\textsuperscript{25} to relate the backbone of nucleic acids and proteins, we here analyze a sub-space of coordinates identified with the \(O5', C3',\) and \(P\) atoms (see Fig. 1 for clarity). As inclusions of backbone atoms different from \(O5', C3',\) and \(P\) did not change significantly the value of chirality index, we decided to use these sets of coordinates as a reference set for oxygen, carbon, and phosphorous atoms belonging to the nucleic acid backbone and starting from the \(O5'\) to follow the \(5'-3'\) convention. In terms of this coordinate, an index \(G_m\) is calculated. Given an atom in this subset of index \(m,\) we consider a fragment which goes from \(m\) to \(m + N - 1\) and define the relative chirality index expressed in eq. 1. The sum is restricted to the fragment defined above and is performed over all possible \(4!\) permutations of \(i, j, k, l\) atoms, belonging to the \(O5', C3',\) and \(P\) atom fragment. Finally, a nucleotide chirality value is defined as an average over the \(G_m\) of its associated \(O5', C3',\) and \(P\) atoms (Fig. 1).

After some trial and error, we have found \(N = 30\) to be optimal, which corresponds to \(30/3\) nucleotides associated as we have considered only \(O5', C3',\) and \(P.\) In fact for this value of \(N,\) which corresponds to the helix pitch, \(G_m\) reaches the highest sensitivity, understandable from the more negative values. Similarly, sensitive is the choice \(N = 33,\) but as the difference between \(N = 30\) and \(N = 33\) is small we have used \(N = 30\) throughout (Fig. 2).

**Nucleic Acid Datasets**

The structures of nucleic acids considered in this work were downloaded from the nucleic acid database (NDB)\textsuperscript{43} (http://ndbserver.rutgers.edu). The list of the analyzed structures is reported in the Supporting Information and consisting for DNA of 1569 from X-ray and 492 from NMR, whereas for RNA 374 from X-ray and 273 from NMR, excluding the old and the obsolete coordinates, as specified in the NDB server\textsuperscript{43} website and checking their consistence with the SCOR database.\textsuperscript{44} As we are interested the connection between chirality index and nucleotide sequence, we have checked the uniformity of the nucleotide distribution in the datasets, whose occurrences are reported in Table 1. The structures of the ideal 30-mer duplex of poly-d(CG) in the B and Z form were generated using 3DNA.\textsuperscript{45}
Using eq. 1, we first calculated the chirality index for an ideal 30-mer duplex of poly-d(CG) in the B and Z form, to prove the effectiveness of chirality index in discriminating their different handedness (see Fig. 3). Later on, we have calculated the chirality index for all the selected structures and this resulting distribution $P(G)$. In Figure 4, we see that DNA tends to have more negative $G$ values reflecting a stronger tendency to form helical structures. It is also interesting to note a multi-varied nature of $P(G)$ in the NMR derived DNA structures, whose structures are less helicoidal. This result led us to examine in further detail which nucleotide possesses a particular chirality and with this purpose we have calculated the probability surfaces of chirality belonging to two consecutive nucleotides of DNA and RNA structures.

**Sequence-Specific Motifs in DNA**

We show in Figures 1 and 2 of Supporting Information the isochirality surfaces for selected pairs of consecutive nucleotides (TT, AG, GG, and CG steps) for X-ray and NMR DNA structures. As expected, the X-ray shows a marked tendency to negative values. Instead, the NMR structures which include less ordered structures exhibit significant positive values. The positive peak is strongly correlated to GG sequences, in which the intensity of the negative peak is lower than that of Thy-Thy sequence. These are expected in telomers which exhibit TTAGGG repeats. A particular sample can be found in Figure 5a where the local variation in chirality is shown together with a corresponding structure. Sequences rich in GG steps are also to be found in junctions (Fig. 5b). We also note that small negative $G$ values can be found also in DNA structures when it forms complex architectures like the cross in Figure 5c.

**Sequence-Specific Motifs in RNA**

The isochirality surfaces of selected nucleotide pairs of RNA (UU, AG, GG, and CG steps) are plotted in Figure 3 of Supporting Information. We decided to combine X-ray and NMR datasets, as no particular difference in their chirality are observed (see Fig. 4). These isochirality surfaces show the tendency of RNA to acquire a larger variety of conformations, and at variance with DNA, there are not two peaks, but instead a broad isochirality distribution. As also found in DNA, the peak at small values of chirality is strongly correlated to GG sequences. Among the isochirality values, we have identified a recurrent peak at $G$ values of $[-0.03 \pm 0.01, -0.06, \pm 0.01]$ which corresponds to RNA hairpin motifs, an example of which is reported in Figure 6 and the list of chirality values for specific loop sequences is reported in Tables 2 and 3. In particular, it comes out that specific negative chirality values are stabilized by uracil-rich sequences, which are more negative and hence more right-handed, if compared with adenine–guanine-rich nucleotide sequences belonging to the third position of the tetraloop sequence. In the light of the DNA and RNA chiralities, it is interesting to compare uracil isochiralities with those of thymine. Thymine shows more populated negative values of isochiralities with

**TABLE 1. Percentages of DNA and RNA bases inside X-ray and NMR datasets**

<table>
<thead>
<tr>
<th></th>
<th>DNA X-ray</th>
<th>DNA NMR</th>
<th>RNA X-ray</th>
<th>RNA NMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ade</td>
<td>30</td>
<td>27</td>
<td>Ade</td>
<td>27</td>
</tr>
<tr>
<td>Thy</td>
<td>27</td>
<td>30</td>
<td>Ura</td>
<td>20</td>
</tr>
<tr>
<td>Cyt</td>
<td>22</td>
<td>17</td>
<td>Cyt</td>
<td>22</td>
</tr>
<tr>
<td>Gua</td>
<td>21</td>
<td>26</td>
<td>Gua</td>
<td>31</td>
</tr>
</tbody>
</table>

**Fig. 3.** Chirality indexes calculated for ideal structures of a 30-mer duplex of poly-d(CG) in B and Z form. It is worth noting the opposite chirality of the two fragments, which is negative for the right-handed and positive for the left-handed helixes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
respect to uracil and this is expected as uracil stabilizes U-turns or tetraloops, as graphically shown in the poly-U uracil sequence of an RNA pseudoknot in Figure 6. Thymine, instead, tends to narrow the DNA grooves, apart the positive values in telomeric repeat regions, and thus it has the tendency to adopt the negative chirality, typical of helix architectures.

To further study RNA motifs, we show the values of chirality for specific loops, as loop-E, loop-C, and the kink-turn. From Figure 7, the three motifs are distinct in their intrinsic chirality, which here properly deals with their bending capability. In particular, in loop-E, the two facing strands having nonstandard base pairings show in the pairing region one maximum, which becomes a minimum in the facing strand (see Figs. 7a and 7b).

Loop-C instead shows lower chirality values, starting from positive indexes and progressively decreasing (see Figs. 7c and 7d), which are more similar to hairpin values, as indicated by the higher values at the beginning of the sequence. The kink-turn, instead, shows two sharp peaks: the first one

Fig. 5. Two examples of regions of DNA that are not right-handed. (a) A guanine-rich telomere structures (pdb code 143D) with chirality indexes reported from the side. Thymine is highlighted by red dots, adenine by violet ones, and guanine by green points, consistent with the colors of the pdb sketches. (b) A cytosine-guanine rich DNA three-way junction (pdb code 1EZN) and chirality is reported from the side. DNA backbone is shown in blue, whereas cytosine-guanine motif is highlighted in red. (c) A right-handed Holliday junction (pdb code 1FLO) mostly containing adenine and thymine shown, respectively, in the pdb sketch in orange and blue. Both of them stabilize negative values of chirality, which are reported from the side. Chirality values shown in (a) and (b) highlight less negative, right-handed, values for guanine and cytosine rich systems at variance with those of adenine-thymine rich junction reported in (c). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
Fig. 6. (a) RNA pseudoknot with poly-uracil loop sequence shown in red (pdb code 1A60). (b) Chirality indexes of RNA pseudoknot, the values of chirality belonging to poly-uracil loop sequence are shown in red. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

<table>
<thead>
<tr>
<th>Gi</th>
<th>Gi+1</th>
<th>Gi+2</th>
<th>Gi+3</th>
<th>Motifs (PDB code)</th>
</tr>
</thead>
<tbody>
<tr>
<td>−0.037</td>
<td>−0.030</td>
<td>−0.027</td>
<td>−0.043</td>
<td>UUUU-classical pseudoknot (1A60)</td>
</tr>
<tr>
<td>+0.009</td>
<td>−0.007</td>
<td>−0.030</td>
<td>−0.058</td>
<td>UUGC-tetraloop (1AUD)</td>
</tr>
<tr>
<td>+0.010</td>
<td>−0.008</td>
<td>−0.036</td>
<td>−0.083</td>
<td>UUGC-hairpin ribozyme loop B domain (1B36)</td>
</tr>
<tr>
<td>+0.018</td>
<td>−0.003</td>
<td>−0.032</td>
<td>−0.040</td>
<td>UUGC-FMN-RNA aptamer complex (1FMN)</td>
</tr>
<tr>
<td>−0.010</td>
<td>−0.014</td>
<td>−0.034</td>
<td>−0.066</td>
<td>UUGC-fragment of 18S ribosomal RNA (1FYO)</td>
</tr>
<tr>
<td>0.003</td>
<td>−0.014</td>
<td>−0.035</td>
<td>−0.053</td>
<td>UUGC-complex of HIV-1 RRE-IB RNA (1G70)</td>
</tr>
<tr>
<td>−0.006</td>
<td>−0.016</td>
<td>−0.039</td>
<td>−0.079</td>
<td>UUGC-trNA ALA acceptor STEM (1IKD)</td>
</tr>
<tr>
<td>−0.001</td>
<td>−0.016</td>
<td>−0.042</td>
<td>−0.087</td>
<td>UUGC-influenza A virus promoter RNA (1J07)</td>
</tr>
<tr>
<td>+0.022</td>
<td>0.001</td>
<td>−0.031</td>
<td>−0.071</td>
<td>UUUC-RNA hairpin binding site for the histone stem-loop (1JU7)</td>
</tr>
<tr>
<td>−0.012</td>
<td>−0.016</td>
<td>−0.031</td>
<td>−0.058</td>
<td>UUUU-histone mRNA Hairpin required (1KKS)</td>
</tr>
<tr>
<td>−0.012</td>
<td>−0.024</td>
<td>−0.040</td>
<td>−0.067</td>
<td>UUGC-Complementary RNA promoter of influenza A virus (1M82)</td>
</tr>
<tr>
<td>−0.033</td>
<td>−0.031</td>
<td>−0.042</td>
<td>−0.061</td>
<td>UUGA23S ribosomal RNA hairpin (1MT4)</td>
</tr>
<tr>
<td>+0.019</td>
<td>0.009</td>
<td>−0.023</td>
<td>−0.066</td>
<td>UUGG-stem loop from the activation domain of HTR (1OQ0)</td>
</tr>
<tr>
<td>+0.008</td>
<td>−0.012</td>
<td>−0.040</td>
<td>−0.064</td>
<td>UUGC-ATP binding RNA aptamer in complex with AMP (1RAW)</td>
</tr>
<tr>
<td>0.000</td>
<td>−0.012</td>
<td>−0.032</td>
<td>−0.052</td>
<td>UUGC-RNA aptamer complexed with HIV-1 REV peptide (1ULL)</td>
</tr>
<tr>
<td>0.000</td>
<td>−0.018</td>
<td>−0.027</td>
<td>−0.042</td>
<td>UUAU-PREQ1 riboswitch (2KFC)</td>
</tr>
</tbody>
</table>

Hairpin values are shown in bold.

<table>
<thead>
<tr>
<th>Gi</th>
<th>Gi+1</th>
<th>Gi+2</th>
<th>Gi+3</th>
<th>Motifs (PDB code)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>−0.007</td>
<td>−0.020</td>
<td>−0.046</td>
<td>AGAA-hairpin with AGAA tetraloop (1ATV)</td>
</tr>
<tr>
<td>0.000</td>
<td>−0.002</td>
<td>−0.021</td>
<td>−0.052</td>
<td>AGAU-hairpin with AGAU tetra-Loop (1ATW)</td>
</tr>
<tr>
<td>+0.018</td>
<td>+0.003</td>
<td>−0.017</td>
<td>−0.045</td>
<td>AGAA-RNA tetraloop (1K6H)</td>
</tr>
<tr>
<td>+0.016</td>
<td>+0.001</td>
<td>−0.029</td>
<td>−0.067</td>
<td>AGUG-IRE hairpin loop (1NB0)</td>
</tr>
<tr>
<td>+0.018</td>
<td>+0.003</td>
<td>−0.013</td>
<td>−0.054</td>
<td>AGAA-RNA Aptamer (1NEM)</td>
</tr>
<tr>
<td>+0.014</td>
<td>+0.003</td>
<td>−0.016</td>
<td>−0.040</td>
<td>AGUU-first stem loop of the of the SL1 RNA (1SLP)</td>
</tr>
<tr>
<td>0.000</td>
<td>−0.011</td>
<td>−0.032</td>
<td>−0.052</td>
<td>UUGC-RNA Aptamer complexed with HIV-1 REV peptide (1ULL)</td>
</tr>
<tr>
<td>+0.007</td>
<td>+0.008</td>
<td>−0.015</td>
<td>−0.057</td>
<td>AGUA-RNA packaging signal complex (2HIX)</td>
</tr>
<tr>
<td>−0.006</td>
<td>−0.016</td>
<td>−0.039</td>
<td>−0.079</td>
<td>UUUC-TRNA ALA Acceptor Stem (1IKD)</td>
</tr>
<tr>
<td>0.004</td>
<td>−0.010</td>
<td>−0.030</td>
<td>−0.045</td>
<td>AGCU-RNA recognition (1TOB)</td>
</tr>
<tr>
<td>−0.003</td>
<td>−0.002</td>
<td>−0.015</td>
<td>−0.037</td>
<td>UUAC-STEM-LOOP (1TXS)</td>
</tr>
<tr>
<td>+0.009</td>
<td>−0.005</td>
<td>−0.025</td>
<td>−0.054</td>
<td>ACAA-RNA signal for translational frameshifting in HIV-1(1ZC5)</td>
</tr>
<tr>
<td>−0.072</td>
<td>−0.028</td>
<td>−0.017</td>
<td>−0.044</td>
<td>ACGG-HIV-1 RRE RNA (1BF)</td>
</tr>
<tr>
<td>−0.037</td>
<td>−0.024</td>
<td>−0.028</td>
<td>−0.049</td>
<td>CUGU-Cell cycle/RNA (1EKZ)</td>
</tr>
<tr>
<td>−0.052</td>
<td>−0.040</td>
<td>−0.038</td>
<td>−0.052</td>
<td>AGAG-Stem-Bulge-Stem Region of the HIV-dimerization (2D17)</td>
</tr>
<tr>
<td>−0.058</td>
<td>−0.036</td>
<td>−0.034</td>
<td>−0.053</td>
<td>AGAG-HIV-1 dimerization initiation site (2D1A)</td>
</tr>
<tr>
<td>−0.048</td>
<td>−0.036</td>
<td>−0.044</td>
<td>−0.070</td>
<td>UGAA-HIV-1 dimerization initiation site (2D1A)</td>
</tr>
<tr>
<td>−0.056</td>
<td>−0.036</td>
<td>−0.034</td>
<td>−0.053</td>
<td>AGAG-HIV-1 dimerization initiation site (2D1B)</td>
</tr>
<tr>
<td>−0.005</td>
<td>−0.013</td>
<td>−0.029</td>
<td>−0.061</td>
<td>ACG25S Ribosomal RNA (1ZS8)</td>
</tr>
<tr>
<td>−0.007</td>
<td>−0.014</td>
<td>−0.032</td>
<td>−0.066</td>
<td>ACG25S ribosomal subunit (2I2T)</td>
</tr>
<tr>
<td>−0.007</td>
<td>−0.013</td>
<td>−0.033</td>
<td>−0.068</td>
<td>ACG7OS ribosomal subunit (3I20)</td>
</tr>
</tbody>
</table>

Hairpin values are shown in bold.
is a maximum, similar to the hairpin maximum and the second one is a minimum, less intense with respect to loop-C and loop-E.

NUCLEOSOME AND RIBOSOME STRUCTURES

By means of chirality, we can now see some insights into the complexity of DNA and RNA structures and their molecular periodicity. With this in mind, we investigate nucleosome and ribosome structures. Nucleosome, being a storage material, must be an ordered system and has to be able to pack efficiently very long DNA fragments. Ribosome, instead, being a molecular machine that carries out complex cellular processes, must be able to assume complex structures. If we look at the chirality index of nucleosome and ribosome structures along the backbone, these features are clearly brought out. Nucleosome chirality reported in Figure 8a distinctly indicates a high periodicity in this complex systems, with only negative chirality, typical of right-handed helixes. In contrast, ribosome chirality shows a wide range of values (see Fig. 8b). The different function and structure of nucleosome and ribosome is clearly reflected in their chirality. Our method is also effective in determining the intrinsic chirality of RNA junctions, which plays a fundamental role in the function of RNA.
role in the folding of RNA. Figure 9 reports the chirality values along Hammerhead ribozyme highlighting, as experimentally shown, a fine regulation in the handedness of RNA junctions.

CONCLUSIONS

A chirality index has been presented to investigate nucleic acid structural motifs. We have shown its effectiveness in capturing structural details in telomeres, DNA junctions, and RNA hairpins. Small or slightly positive chirality values are observed in guanine-rich sequences, whereas adenine or thymine-rich sequences show more negative values, consistent with their tendency to narrow DNA grooves. We have considered the consecutive repeat TTAGGG of telomeres, which are known to be essential for genome stability. Such telomeric repeats end in a guanine-rich single-stranded overhang, crucial for telomere protection and extension. Telomeres possess G-quadruplex architectures, highlighting the importance of guanine conformational preferences and in
these regards, we have observed it to adopt a left-handed chirality. We have compared nucleosome and ribosome and a high level of periodicity has been detected in nucleosome, whereas ribosome structures show an aperiodic chirality pattern along RNA backbone. Although the nucleotides shared by DNA and RNA differ slightly in their chemistry, they lead to different spatial arrangement. A chirality analysis brings out clearly the different structures.

LITERATURE CITED


Fig. 9. Hammerhead ribozyme (pdb code 2OEU(25)); chain A and chain B are shown in blue and red, respectively. It is worth noting the positive chirality of chain B and the negative chirality of chain A, which becomes more negative moving along the backbone chain. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]


49. Dallas A, Moore PB. The loop E-loop D region of Escherichia coli 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. Structure 1997;5:1639–1653.


