Article No. mb982233 *J. Mol. Biol.* (1998) **284**, 1465–1478

JMB

Stepping Through an RNA Structure: A Novel Approach to Conformational Analysis

Carlos M. Duarte and Anna Marie Pyle*

Department of Biochemistry and Molecular Biophysics Columbia University New York, NY 10032, USA

Drawing from the growing database of complex three-dimensional RNA structures, a systematic method has been developed for classifying and analyzing the variety of conformations adopted by nucleic acids. This method is based on the development of a reduced representation for nucleic acid backbone conformation, simplifying the formidable eight-dimensional problem that has long complicated nucleic acid conformational analysis. Two pseudotorsion angles (η and θ) have been defined, based on the selection of two appropriate pivot points along the RNA backbone, P and C4′. These pseudotorsions, together with a complete library of conventional torsion angles, can be calculated for any RNA structure or all-atom model using a new program called AMIGOS. Having computed η and θ pseudotorsions for each position on an RNA molecule, they can be represented on a two-dimensional plot similar to the ϕ-φ plots that have traditionally been used for protein conformational analysis. Like a Ramachandran plot, clusters of residues appear at discrete regions on an η-θ plot. Nucleotides within these clusters share conformational properties, often belonging to the same type of structural motif such as A-platforms, sheared tandem purine-purine pairs and GNRA tetraloops. An η-θ plot provides a two-dimensional representation of the conformational properties of an entire RNA molecule, facilitating rapid analysis of structural features. In addition to the utility of η-θ plots for intuitive visualization of conformational space, the pseudotorsional convention described here should significantly simplify approaches to macromolecular modeling of RNA structure.

Keywords: RNA tertiary structure; torsion angle; Ramachandran; computational analysis; nucleic acid

**Corresponding author*

Introduction

''Various types of poly**nucleotide** conformations have been proposed in recent years for **RNA**, e.g. the **hairpin tetraloops, sheared tandem purine-purine base-pairs, adenosine platforms**. However, there appears to be no analytical method for writing down these configurations. In connection with the studies on the **group II intron** carried out in this laboratory, the authors have worked out a convenient notation of this type and this is briefly described here.''

Boldface portions of the text above are substitutions that we have made into the original text from Ramachandran *et al*. (1963).

There has been an explosion of information on the three-dimensional structure, dynamics and conformational energetics of RNA. While this type of data is rapidly accumulating, there is no systematic way to describe RNA structural elements, or to classify them for comparison. For example, common names for RNA motifs (such as the UUCG or GAAA tetraloop) often include members that differ markedly in structure (Allain & Varani, 1995; Peterson & Feigon, 1996) or are seen in alternative structural contexts (Heus & Pardi, 1991; Abramovitz & Pyle, 1997; Legault *et al*., 1998), respectively. In general, RNA structural features are examined visually and are described anecdotally in a phenomenological rather than a quantitative manner. This makes it difficult to evaluate RNA structural elements and tertiary architecture in the rigorous fashion that is employed for analysis and modeling of protein structure.

E-mail address of the corresponding author: amp11@columbia.edu

0022–2836/98/501465–14 $30.00/0

The most useful way of visualizing, analyzing, and predicting polymeric structure is to classify the range of conformations that individual monomers can adopt in the context of a polymer. Two-dimensional Ramachandran plots (Ramachandran *et al.*, 1963) are the primary basis for amino acid conformational classification and description, and they are now commonly used to analyze, predict and verify structural models of proteins (Laskowski *et al.*, 1993). This approach is made possible by the fact that only two torsion angles (ϕ and φ) are required to define and describe the backbone configuration of a protein molecule.

By contrast, the conformation of a nucleotide within an RNA molecule is described by six backbone torsional angles (α, β, γ, δ, ε and ζ; Figure 1), the torsional angle of the bond between the ribose ring and the base (χ), and the pucker of the ribose ring (minimally described as C3′-*endo* or C2′-*endo* (Saenger, 1984)). Although some correlations have been observed between α and β, α and γ, β and ε, ε and ζ, and δ and the ribose pucker, these do not significantly reduce the large dimensionality of the problem. In addition, correlations between individual parameters are not absolute. Nucleotide analysis is therefore an eight-dimensional problem that cannot be evaluated intuitively and represents a formidable computational task. This is exemplified by attempts to model even small architectural elements in RNA.

For example, the program MC-SYM (Major *et al.*, 1991) has been successfully used to create structural models of RNA by combining computational approaches and empirical data (Laing & Hall, 1996; Klinick *et al.*, 1997). We used the program to create structural representations of the two-nucleotide asymmetric loop in D5 of group II introns, surrounded on either side by standard duplex. When given a broad set of unbiased input parameters (using the full 718-nucleotide conformational database) and no empirical restraints, the program created over 750 sterically allowed structures (data not shown). We explored several approaches for parsing these into groups of similar structure (including global comparisons of base-plane dihedrals, etc.), all of which failed. Individual structures had to be visually and subjectively inspected for ''reasonableness'' one by one. When this exercise was carried out, it became apparent that most of the solutions are almost identical, differing only in tiny alterations of individual torsions that have little or no consequence for the morphology of the loop. This suggests that individual dihedrals can vary significantly without altering overall structure and may, in fact, provide too much information in certain contexts. This exercise exemplifies the need to derive a simplified means for describing nucleic acid structure.

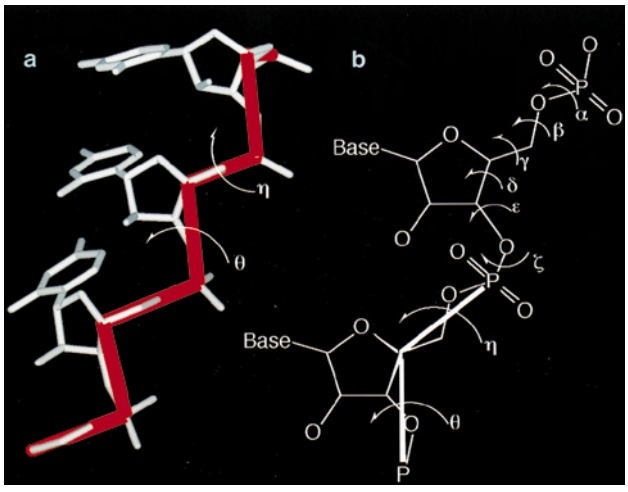

Figure 1. The η and θ pseudotorsion angles are contrasted with conventional torsions α, β, γ, δ, ε and ζ. (a) Canonical type-*A* form RNA strand with the pseudobonds connecting P and C4′ along the backbone. These are the central bonds of the pseudotorsions η and θ, shown. (b) Diagram of an RNA backbone showing both the conventional torsions and the pseudotorsions η and θ.

A common way to depict RNA structure employs a series of circles, rectangles and lines (Saenger, 1984; Varani, 1995). Circles of different sizes are commonly used to represent sugars and phosphate groups, respectively, while rectangles represent bases and lines connect them to each other appropriately. When this schematic is used to depict a normal helix, the backbone can have the appearance of a spiral staircase. We have applied this conceptual shorthand to the quantitative structural analysis of RNA by selecting two backbone atoms, P and C4′, whose positions were most like the corners between the steps and rises of an RNA helical staircase. Conceptually, P and C4′ can be connected to each other sequentially by a series of ''pseudobonds''. In this description, each nucleotide is spanned by two pseudobonds, the first from P to C4′ of a nucleotide, and the second from C4′ to P of the sequentially successive, or 3′, nucleotide (Figure 1). The two pseudobonds of each nucleotide, N, have corresponding pseudotorsional angles. The first of these, η, is defined by the atoms, $C4'_{N-1}$-P_N-$C4'_N$-P_{N+1}, while the second, θ, is defined by P_N-$C4'_N$-P_{N+1}-$C4'_{N+1}$ (Figure 1).

In this work, we have measured η and θ for all nucleotides in an extensive database of three-dimensional RNA structures. We show that η and θ of all nucleotides in an RNA molecule can be plotted on the axes of a two-dimensional plot, in the manner of a Ramachandran plot. This two-dimensional representation provides an intuitively accessible, graphic representation of quantitatively distinct structural features. We show that clusters of nucleotides with similar η and θ torsional angles have similar conformational properties and *vice versa*. In addition, we demonstrate that the pseudotorsions η and θ are at least as sensitive as conventional torsion angles, if not more so, for characterizing specific structural distortions within a RNA polymer. We therefore propose that the η-θ

conformational shorthand is useful for analyzing and modeling RNA structure.

Results

Features of the η-θ plot

The η and θ angles of all nucleotides within a structural database (see Methods) were plotted in two dimensions (Figure 2(a)). The plot reveals one particularly distinct grouping of nucleotides with similar η-θ values that is centered around $\eta = 170^\circ$ and $\theta = 225^\circ$, and covers an area roughly equivalent to the intersection of the two perpendicular gray bars (Figures 2(b) to (e) and 3(a)). This region is designated the helical region. The identity of these nucleotides was evaluated and they were generally found to be residues within helical secondary structure. Because the bulk of RNA crystal structures and NMR structures to date have been carried out on molecules that are simply helices or composed primarily of helices, it is reasonable that nucleotides with this conformation should dominate the plot.

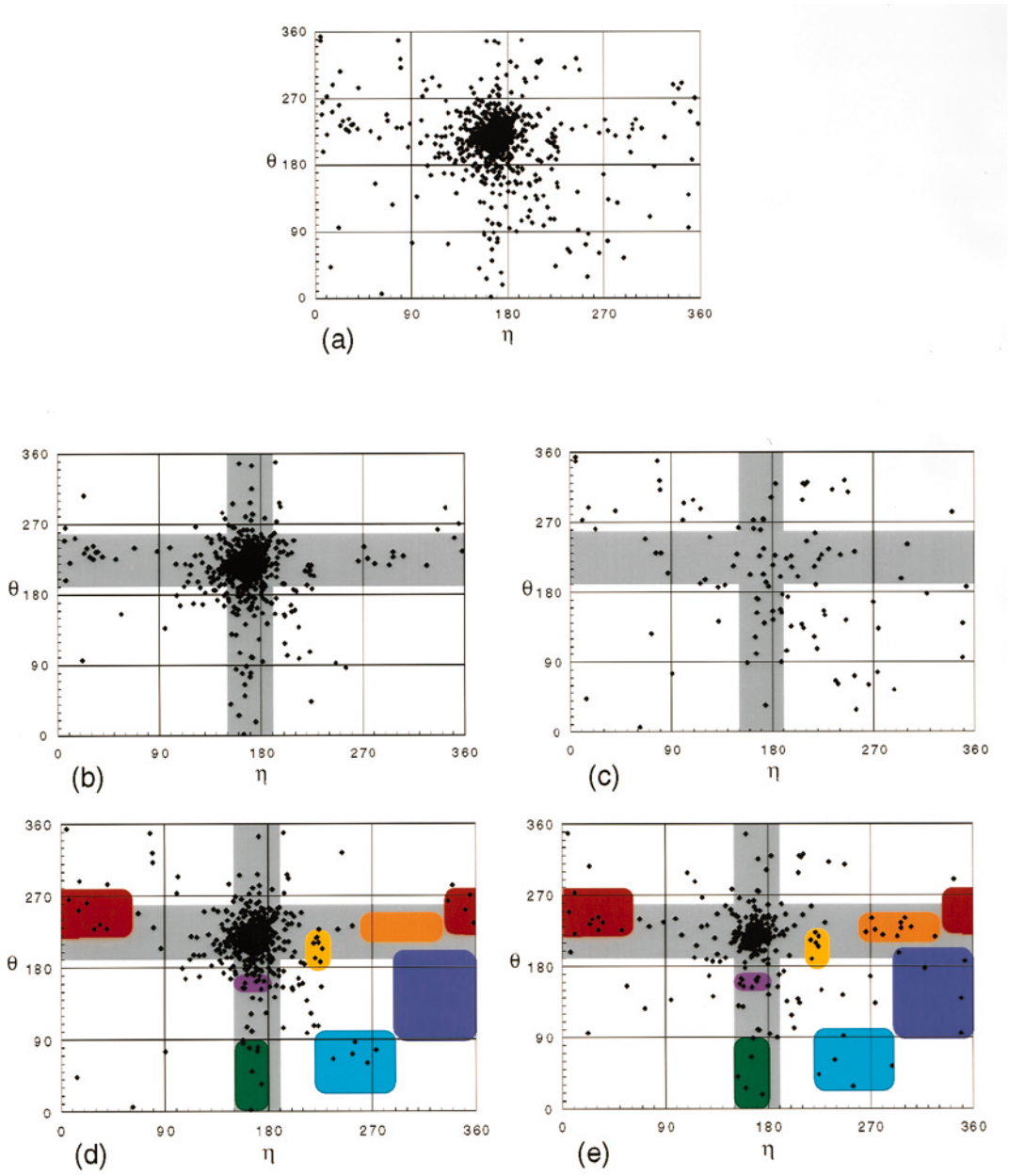

Figure 2. Plots of η and θ for different types of nucleotides and databases. (a) All non-terminal nucleotides from all molecules in the database. (b) All C3′-*endo* nucleotides from the database. Gray bars represent areas of the plot where either η or θ is in the same range as nucleotides in the helical region. The helical region is the intersection of these bars. (c) All C2′-*endo* nucleotides from the database. (d) All nucleotides from molecules with no tertiary interactions. Colored regions correspond to regions discussed in the text (and see Figure 3). (e) All nucleotides from molecules that feature tertiary interactions.

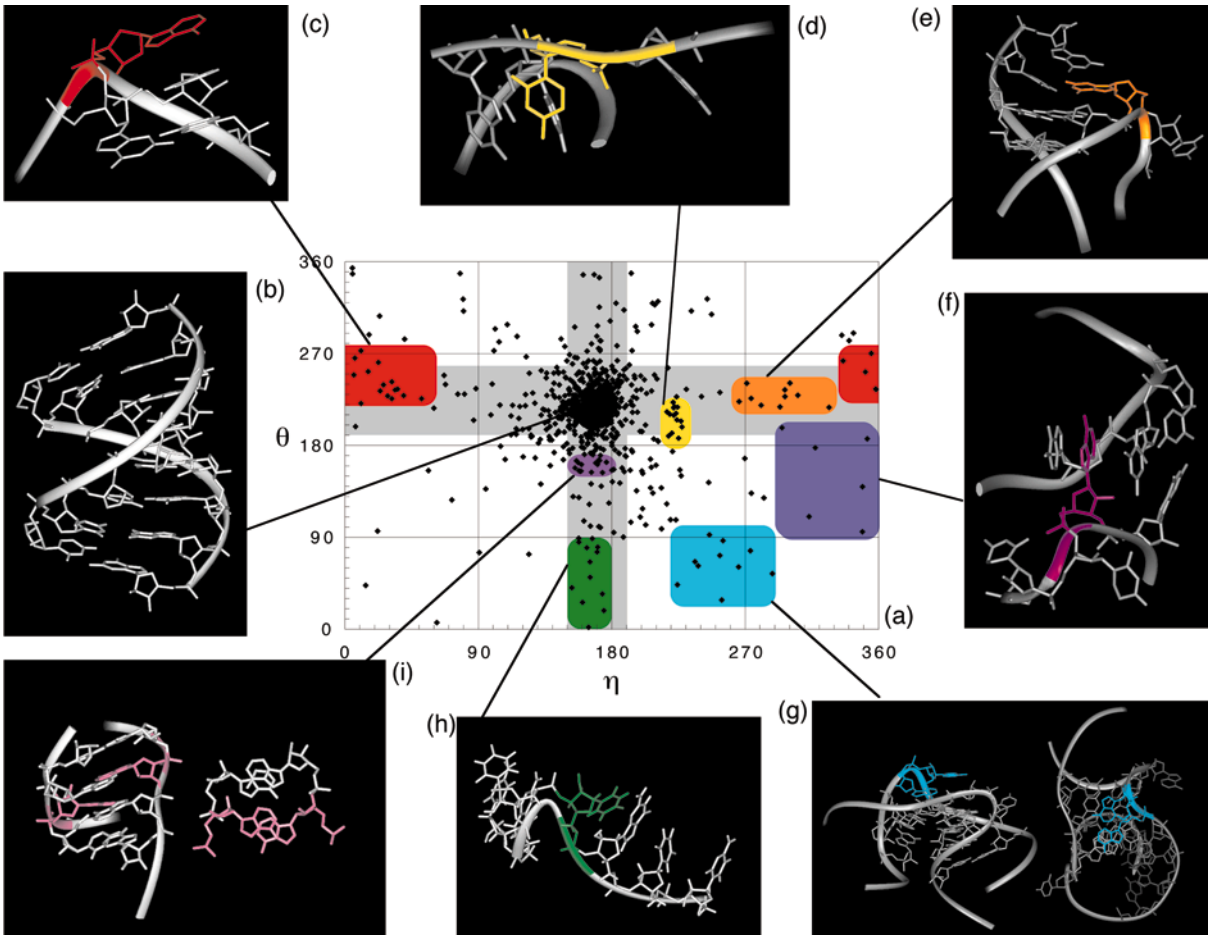

Figure 3. (a) An η-θ plot of all nucleotides from the database of 53 RNA structures. Gray bars are as described for Figure 2(b). Colored areas are regions of the plot that contain nucleotides that share similar structural features (see the text). (b) to (i) Representative nucleotides from the regions of the plot indicated. (b) The helical region; at the intersection of the two gray bars includes nucleotides from the crystal structure of an *A*-form duplex (from PDB file 1rxa). (c) Stacked turn region; exemplified by the second nucleotide of a GNRA loop (PDB file 1zif, Ade 5). (d) The χ-switch region; includes the nucleotide 5′ to the cleavage site of the hammerhead ribozyme (PDB file 300d, Cyt B170). (e) Flip turn region, exemplified by APKA27G pseudoknot nucleotide G9 (PDB file 1kpd, Gua 9). (f) The C2′-bend region, includes $tRNA^{phe}$ tertiary contact nucleotide G18 (PDB file 1tra, Gua 18). (g) The stack switching region, exemplified by P456 domain pivot nucleotides (PDB file 1gid, nucleotides Ade A122, Ade A123). (h) The base twist region; includes the last stem nucleotide of a kissing hairpin (PDB file 1kis, Ura 21). (i) The cross-strand stack region; includes all 5′ nucleotides in sheared tandem R-R pairs (PDB file 1gid, Ade A113,Ade A206).

Ability of η and θ to discern major conformational differences between nucleotides

One of the major differences in nucleotide conformational state stems from whether a ribose pucker is C3′-*endo* (the norm in RNA) or C2′-*endo*. To determine if the plot was sensitive to sugar conformation, all nucleotides in the database were subdivided into C3′-*endo* or C2′-*endo* categories. A comparison of the η-θ plots of these two groups revealed striking differences (Figure 2(b) and (c)). Over 90% of C3′-*endo* nucleotides fall within a sector defined by the two perpendicular gray bars (Figure 2(b)) and therefore have either the same η and/or θ measurements as nucleotides in the previously described helical region. In contrast, C2′-*endo* nucleotides are more variably distributed, and there are several clearly identifiable regions where only C2′-*endo* nucleotides were found. These results strongly suggest that the η-θ shorthand is sensitive to conformational differences in sugar pucker. Sugar pucker, while critical, is not the sole determinant of backbone morphology. Both C3′-*endo* and C2′-*endo* nucleotides adopt a variety of conformational states and this is reflected in the distribution of points for each on their respective η-θ plots.

Nucleotides involved in tertiary interactions often adopt highly unusual conformations relative to more simple structures. To get a preliminary idea of whether the plot would reflect these differences, we divided the database into sets of structures that contained or lacked extensive tertiary interactions, respectively. All nucleotides from

both subsets were plotted and examined (Figure 2(d) and (e)). The first subset consisted of three pseudoknot structures (Shen & Tinoco, 1995; Kang *et al*., 1996; Kang & Tinoco, 1997), a kissing hairpin (Chang & Tinoco, 1997), domain P4-6 of the group I intron (Cate *et al*., 1996), yeast $tRNA^{Phe}$ (Westhof & Sundaralingam, 1986) and the all-ribose form of the hammerhead ribozyme (Scott *et al*., 1995) (Figure 2(e)). In the second group were all other molecules in the database, including many straight duplexes (Dock-Bregeon *et al*., 1989; Conte *et al*., 1997) and others containing internal loops designed to study mismatch pairing arrangements or bending (Baeyens *et al*., 1996; Peterson & Feigon, 1996) (Figure 2(d)). A comparison of the two η-θ plots shows that the vast majority of nucleotides where η is greater than 270° are in molecules containing tertiary interactions. Visual inspection of these nucleotides revealed that nearly all were either directly involved in tertiary interactions or flipped out of a helix.

Ability of the η-θ plot to specify individual folding motifs

Inspection of the η-θ plot reveals that, in addition to the helical region, there are sets of nucleotides with similar η and θ angles that cluster in several different regions of the plot. In order to determine if nucleotides in these clusters share conformational similarities, we chose eight regions of the η-θ plot for characterization (Figures 2(d) and (e), and 3(a)). Clearly, there are more regions of potential interest on the plot, all of which will become more pronounced as the database expands. But in the interest of time, we endeavored to analyze only eight regions that appear to be relatively distinct at this time.

Each nucleotide within the eight regions was visually examined by referring back to the original structures (Figure 3(a)) and conformational features were tabulated in order to statistically evaluate the extent to which a given region contains nucleotides with specific structural features (see Methods). We find that nucleotides within a region share distinctive conformational characteristics, which are classified below. The eight regions we examined do not represent a complete survey or analysis of the map, which contains additional potentially interesting clusters of nucleotides (e.g. a region centered at η 200°, θ 140°). The structural features described below do not necessarily apply to every nucleotide in a region, but do describe the vast majority in each, and, in some regions, describe all members.

For purposes of the following discussion, it should be noted that nucleotides sharing an η value typical for *A*-form helical residues are defined as having ''normal'' structural features leading into their 5′ side, while those with the same θ as *A*-form nucleotides have ''normal'' structural features leading from their 3′ side. Interestingly, by this definition we find that most non-helical nucleotides have one normal pseudotorsion and one that is irregular. Throughout this work, we define as normal any structural feature (such as a conventional or pseudotorsion angle) that falls in a range empirically found to be typical for nucleotides within a standard *A*-form helix (see Table 1 and The helical region, below). Note that a nucleotide can have normal conformational parameters whether or not it is actually part of a duplex structure.

The helical region

The helical region (centered at η 170°, θ 225°; Figure 3(b)) corresponds to the intersection of two perpendicular gray bars and is the most densely populated area of the plot. Due to the sheer number of nucleotides in this area, we did not examine the conformational characteristics of every member nucleotide, but did test a sufficient number to make the following generalizations. Nucleotides within this region are generally involved in Watson-Crick base-pairs, as part of an *A*-form helical structure. This region also contains unpaired nucleotides that still adopt the geometrical features characteristic of nucleotides in a helix (defined as normal conformation) The mode values (although not necessarily individual values) of the standard torsions for nucleotides throughout this region all fall within the normal range using either one of two standard classifications (Table 1). This area comprises over 70% of all nucleotides in the database.

The stacked turn region

The stacked turn region is typified by the conformational state of the second nucleotide (N) in a GNRA hairpin loop. The region extends out the left side of the plot and into the right, centered at η 25°, θ 250° (Figure 3(c)). The region appears to be split only due to the choice of axis ranges, since the map should actually be thought of as a cylinder in which η specifies positions around the circumference. Nucleotides in this region are found at turns in the backbone and are the 5′ base of a new stacking axis. This definition accurately describes 72% of the 18 nucleotides that fall in this region. For the five residues that do not fit this definition, four are in large extended loops and one is part of an unusual bend and tertiary interaction in tRNA. For nucleotides within this region, the conventional torsion angles α and β fall outside the normal range, while mode values for γ, δ, ε and ζ fall within a range commonly specified for normal structure (Table 1). The second base of all GNRA tetraloops, both bound and unbound, fall within the stacked turn region.

The χ-switch region

The χ-switch region includes the nucleotide 5′ to the cleavage site in crystal structures of the ham-

Table 1. Ability of RNA backbone torsion angle ranges to identify *A*-form structure

I Screening criterion (by torsion angle)	II Saenger[a] range	III Crystallographic[b] range	IV η-θ[c] range
α	265-310	155-360	–
β	165-210	149-210	–
γ	45-60	−5-180	–
δ	75-95	65-100	–
ε	170-210	155-225	–
ζ	280-320	250-340	–
η	–	–	150-190
θ	–	–	190-260
Screening results			
All nucleotides[d]	208 of 1311	715 of 1311	944 of 1311
A-form crystal 1[e]	0 of 12	12 of 12	12 of 12
A-form crystal 2[e]	3 of 12	12 of 12	12 of 12
P456 canonical[f]	17 of 76	50 of 76	68 of 76
P456 non-canonical[g]	9 of 80	32 of 80	36 of 80

Shown are two sets of angle ranges reported in the literature that can be used as standards (top, columns II and III), together with their ability to predict conformational features typical of nucleotides in *A*-form helices (bottom, columns II and III). This is compared with the predictive capability of η and θ parameters (bottom, column IV).

[a,b] Ranges for backbone torsion angles taken from published texts (Saenger, 1984) and derived from a set of high-resolution *A*-form crystal structures (Portmann *et al.*, 1995).

[c] The values shown for η and θ (top, column IV) correspond to the helical region of an η-θ plot (see Methods).

[d] The total number of nucleotides in the database, of which 208/1311 are identified as having normal conformation using the Saenger torsion range, in contrast to 715/1311 using *A*-form crystal parameters and 944/1311 using pseudotorsions.

[e] The number of nucleotides in the high-resolution *A*-form crystal structures identified as having normal conformation using the three standards.

[f] The 76 nucleotides from the P456 structure reported to participate in conventional base-pairs (Cate *et al.*, 1996), and the number identified as having normal conformation by each standard (columns II-IV).

[g] The 80 remaining nucleotides from the crystal structure of P456 and the number identified as having normal conformation by the three standards (columns II-IV).

merhead ribozyme (Figure 3(d), centered at η 220°, θ 200°). This region includes a discrete group of nucleotides that are exclusively composed of C3′-*endo* residues (compare with Figure 2(b)). These nucleotides have χ-angles that differ markedly from the nucleotide to either the 5′ and/or 3′ side, thereby placing the base at an unusual angle relative to surrounding structure. This definition applies to 90% of all nucleotides that are found in this region. Despite the large differences in χ (>40°) between these nucleotides and their neighbors, in many cases both residues are still *anti* (rather than alternating *syn* and *anti*). The α, β, ε and γ angles for this region are variable, while mode values for δ and ζ fall within the normal range. Members of this class are found adjacent to bulges, aptamer-binding sites, purine-purine base-pairs and platforms.

The flip-turn region

The flip-turn region includes nucleotide G9 from the APKA27G frameshifting pseudoknot (Figure 3(e), centered at η 285°, θ 230°). The region contains nucleotides involved in tertiary interactions (Figure 2(e)) which are at a sharp turn within a strand, where a turn is defined as a full reversal of strand direction. Nucleotides in this region contain bases that are oriented ∼90° out of the plane from bases that are immediately adjacent to the 5′ side. These definitions apply to 80% of the nucleotides in this region. Of the two regional nucleotides that do not fit this description, one is involved in a bend as a base triple (A107 from the P456 structure) and another is in a large loop. All of the nucleotides in this region are flipped (extrahelical) with respect to their original (5′) helix axis. They have normal δ and ζ values but the other torsion angles are highly variable. Nucleotide C48 from yeast $tRNA^{Phe}$, which sits at the interface between two coaxially stacked helices, is a representative member of this region.

The C2′ bend region

The C2′ bend region is typified by G18 of yeast $tRNA^{Phe}$, which is stacked between nucleotides mA58 and G57 to form a critical tertiary interaction (Figure 3(f); centered at η 325°, θ 140°). The region contains nucleotides involved in bends or turns. All nucleotides within this region are C2′-*endo* and all are extrahelical. These definitions describe all five residues in this region, four of which are involved in tertiary interaction (Figure 2(e)). The ε and ζ values for nucleotides in this region are variable, γ falls within the mode value for normal conformation and the mode values for α, β and δ are well defined, but outside the bounds of normal conformation.

The stack switching region

The stack switching region includes two consecutive nucleotides found in the P456 crystal structure (Figure 3(g); centered at η 255°, θ 60°) that form the ''pivot'' between two large helical domains. Each base caps one of the two stacking axes at the terminus of each helical domain. Another notable example of this type of nucleotide is residue 48 of yeast $tRNA^{Phe}$, which is immediately 5′ of the coaxially stacked T arm. Nucleotides in this conformational region lie between helices that do not share the same stacking axis. Eight of the ten nucleotides in this region fit these definitions. Of the two that do not, one is involved in an extended loop, while another is an extrahelical base involved in a tertiary interaction. Values for α, β, ε and ζ are variable, and well-defined mode values for γ and δ are not within normal ranges.

The base twist region

The base twist region is typified by U21 of the kissing hairpin structure, which is the last stem nucleotide immediately 5′ of the inter-hairpin helix (Figure 3(h); centered at η 165°, θ 90°). As expected for residues with a normal η (vertical gray bar, Figure 3), residues in this region are stacked on the adjacent 5′ base. To the 3′ side, these residues initiate a bend in the strand. Their bases are found at an acute angle relative to the plane of the adjacent 3′ base (between 45 and 90°). These definitions strictly fit 11 of the 13 residues found in this region. The two that do not fit the description are adjacent nucleotides in a loop. By this description, nucleotides in this region have a structure that is opposite those found in the flip turn region (Figure 3(e)). Mode values for α, β, δ and γ are normal while ζ and ε are variable.

The cross-strand stack region

The cross-strand stack region includes all 5′ nucleotides involved in sheared tandem purine-purine base-pairs, including the example shown, from the P456 structure (Figure 3(i); centered at η 168°, θ 160°). Of the 16 nucleotides in this region, 12 are involved in purine-purine base-pairs and eight of these are involved in a cross-strand stack. For nucleotides in this region, there is no case in which the base stacks on its immediate neighbor to the 3′ side, although 12 of the nucleotides in this region stack on the adjacent 5′-residue. Their mode values for α, β, γ and δ all fall within normal ranges, while ε and ζ are variable.

Although the vast majority of nucleotides within a given region belong to a specific conformational group or motif, it is important to consider characteristics of the nucleotides that do not fit the definitions supplied for each region: for all regions, nucleotides that defy specific definition tend to originate from large loops of at least four nucleotides. That these nucleotides adopt a diversity of conformations (they are, in fact, scattered throughout the plot) is probably linked to the fact that the constituents of large loops have a high degree of freedom. Alternatively, given their potential for motion, these loops may represent relatively disordered regions of the structural database and therefore regions of high RMSD.

The accuracy of η and θ relative to standard torsion angles

The biggest concern in using a shorthand notation like η and θ is that one might lose a great deal of important, subtle structural information. We were concerned that the pseudotorsions might not describe nucleotide structure as accurately as conventional nucleotide torsion angles. To address this, we decided to compare the ability of ''standard torsions'' to specify residues belonging to a particular structural class relative to the new pseudotorsions. Before this could be done, it was necessary to define the range of standard torsion angles for RNA. It should be noted that there is great variability in χ, even within *A*-form structures, so we limited this exercise to examining the variability of backbone angles α through ζ.

The most commonly used set of torsion angles for *A*-form RNA nucleotides are derived from a range provided by Saenger 1984; and see Table 1. To see if these angles can be used to describe *A*-form structures, we used AMIGOS to measure each standard torsion for individual nucleotides within two high-resolution crystal structures of *A*-form helices. Interestingly, only three of the 24 residues in these two structures have standard torsion angles whose measurements all fall within the Saenger standards, although the RNA appears to be quite regular (Table 1). Standard torsion angles for *A*-form structure can also be defined by using the ranges found within two high-resolution crystal forms of *A*-form helices (Portmann *et al.*, 1995). These provided a more realistic range; however, they cannot uniquely identify a nucleotide as belonging exclusively to an *A*-form helical family (Table 1). By broadening the ranges of acceptable torsions, one begins to identify nucleotides that clearly do not have ''normal'' conformation (such as the adenosine base at position 2 of a GAAA tetraloop in the hammerhead crystal structure).

To explicitly specify that the nucleotides within these structures belong to a specific configuration, Egli and co-workers were required to employ non-torsional parameters for description (Portmann *et al.*, 1995). Helical parameters such as inclination, roll, groove width, etc. were invoked to classify the structures as normal *A*-form. This fact exemplifies the problem with description of structure through application of standard torsion angles. The standard angles, complex as they are, cannot specify a unique conformation and must therefore be supplemented by additional descriptors that relate to overall three-dimensional form. However, it is problematic to apply these descriptors to RNA struc-

ture that is not helical. RNA involved in loops, tertiary interactions and folds other than secondary structure is difficult to classify by this means. While the program CURVES (Lavery & Sklenar, 1988) can calculate parameters related to inter-base rise, roll and other parameters for single-stranded nucleotides, we are not aware of any correlations between these parameters and specific non-helical motifs or structures.

The preceding analysis indicates that conventional torsion angles are by no means an ideal way to evaluate the conformational state of an RNA motif. To determine whether η and θ might actually be more accurate in this regard, η and θ were compared to conventional torsion angles as a means of defining specific conformational features. As shown in Table 1, the values for η and θ range from 150-190° and 190-260° for regular *A*-form geometry as defined by the central ''helical'' region of an η-θ plot of all nucleotides in the current database (Figures 2 and 3). The P456 crystal structure, which contains a large distribution of nucleotides in both canonical (normal) and non-canonical conformations, was used as a test case for monitoring the ability of different standards to differentiate normal structure from that belonging to other subgroupings. In the P456 crystal structure, 76 nucleotides are part of helices and involved in standard Watson-Crick base-pairs (canonical), while 80 nucleotides are involved in a wide variety of other structures and alternative pairings (non-canonical) (Cate *et al.*, 1996). The standard Saenger torsional ranges defined previously identify only 17 of the 76 canonical nucleotides as having normal (as in *A*-form) conformation. By contrast, the pseudotorsions η and θ identify 68 of 76 as being normal. This is significantly higher even than the *A*-form torsion ranges defined by the Egli crystal structures (Table 1), which identify 50 of the normal 76 nucleotides. Conversely, the Saenger torsion ranges specify 9 of the 80 non-canonical nucleotides in P456 as normal, while the empirical *A*-form ranges and the η and θ ranges specify that 32 and 36 of them are normal, respectively. That all three types of parameters result in values that are greater than zero reflects the fact that nucleotides in non-helical arrangements may still adopt local backbone conformations like that found in helical *A*-form structure. Taken together, these results suggest that the pseudotorsions η and θ are at least as descriptive of backbone morphology as standard torsions and, in terms of specifying the backbone conformation of an individual nucleotide, they may in fact be superior.

These findings indicate that the use of regular torsions to classify structure is problematic, as they can reflect subtle variations in structure that do not change the overall fold of the nucleotide. Any deviating standard torsion can be accompanied by an adjustment in one or more of the other torsions to yield highly regular *A*-form structure. Even with the relaxed criteria described above, many nucleotides that by visual inspection had seemingly regular structure were not identified as such. Additionally, as the acceptable range for each torsion is expanded to fit more regular nucleotides, the resulting parameter set mistakenly identifies as normal, increasing numbers of nucleotides with highly unusual conformation. Given the shortcomings of standard torsions, a shorthand description such as η and θ is useful for accurately describing the ensemble of backbone torsion angles without loss of information content.

Do similar structural motifs have similar η and θ angles?

Thus far, our observations have suggested that nucleotides with given η-θ parameters tend to belong to similar types of motifs. To test this further, it is important to turn the problem around and determine whether nucleotides that all adopt a specific motif share similar η and θ angles. To address this problem, we examined the η-θ signatures for four of the most common structural motifs in the database: sheared tandem purine-purine base-pairs, GNRA tetraloops, UNCG tetraloops, and adenosine platforms (Figure 4).

Sheared tandem purine-purine base-pairs can be characterized by an irregular nucleotide conformation at the 5′ purine of the motif on each strand, and normal conformation of the 3′-purine. The conformational deviation imposed at the 5′ position is sufficient to kink the 3′ nucleotide onto the cross-strand stack. The η-θ parameters for this motif are the same for all examples studied thus far (Figures 3(i) and 4).

There are seven examples of GNRA motifs in our database, all of which have similar η and θ values (Figures 4 and 5(a)). According to η-θ parameters obtained for this motif, GNRA tetraloops have unusual backbone conformation only at the second (GNRA) position (Figure 5(a)). There is

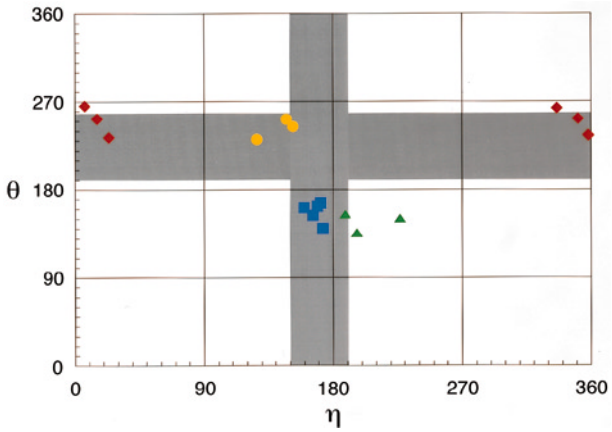

Figure 4. An η-θ plot showing the coordinates of common RNA structural motif nucleotides, which are indicated by color. Red diamond, GNRA second position; blue square, sheared tandem R-R base-paired 5′ base; green triangle, A-platform 5′ A; yellow circle, A-platform 3′ A.

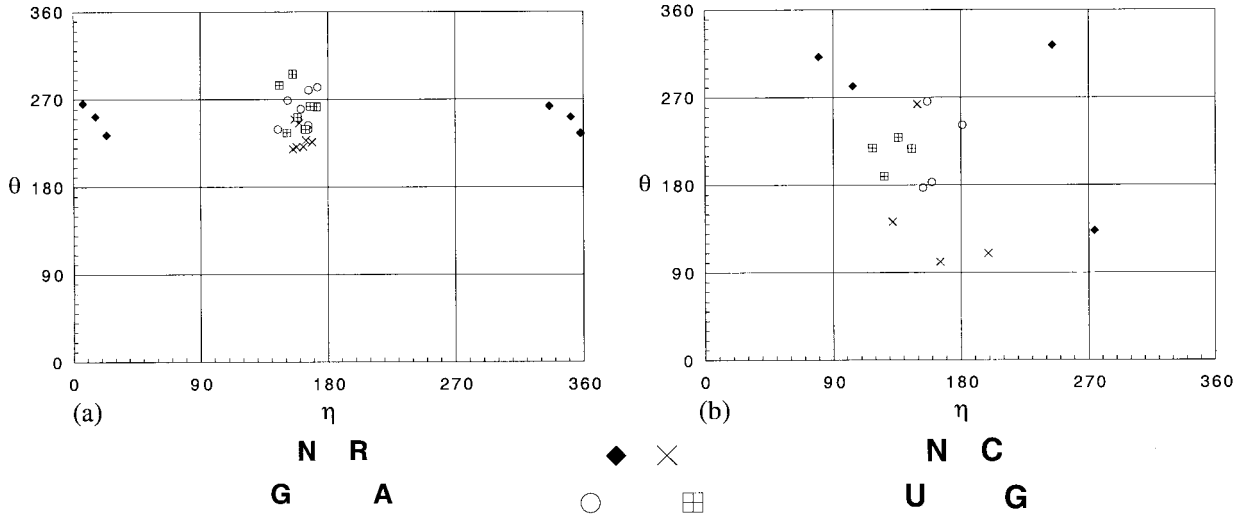

Figure 5. The η-θ plots of (a) all GNRA tetraloops and (b) all UNCG tetraloops in the database. Position 1 (○); position 2 (◆); position 3 (×); position 4 (⊞).

some variation in θ for nucleotides at the 1 and 4 positions of this motif but these do not lie far outside of the helical region. Our initial characterization of trends in GNRA conformation was made at a point where our database contained only 35 structures. Recent additions to the database include two mini GNRA tetraloop molecules, whose sequences are GAGA and GCAA (Jucker *et al.*, 1996). In both cases, our initial observations about the characteristics of these structural motifs has been substantiated. This supports the consistent and predictive properties of the η-θ parameters with respect to GNRA motifs.

Unlike GNRA tetraloops, UNCG tetraloops do not have consistent, characteristic η-θ values for all known examples (Figure 5(b)). This indicates either that a uniform structural definition may not be possible through conformational analysis or that there may be a flaw in our methodology. To explore the latter possibility, we examined backbone structural alignments of all examples of UNCG and GNRA tetraloops from our database. The average RMSD between all GNRA tetraloops was 0.8 Å, compared to 2.0 Å for UNCG. While many studies have shown that the UNCG tetraloop is more thermodynamically stable (Varani, 1995), our comparison of the conformations of these motifs as determined by NMR and X-ray crystallography indicates that there is a much wider range of UNCG tetraloop conformations compared with GNRA tetraloops. This may reflect the quality of the structures determined, multiple stable conformational states, or dynamic motion in the structure of UNCG tetraloops. Thus, the different η-θ angles that are observed for the UNCG tetraloops reflects the fact that these loops exist in a variety of different conformations and that the η-θ analysis is sensitive to this fact.

The crystal structure of P456 was found to contain three adenosine platform motifs (Cate *et al.*, 1996). This motif consists of two contiguous adenosine bases in which adjacent bases lie side-by-side, in the same plane. On an η-θ plot, the 5′ nucleotides of this motif cluster near η 205°, θ 150°, while the 3′ nucleotides are found near η 140°, θ 240° (Figure 4). The 3′ nucleotides all have regular θ values corresponding to their positioning in two instances at the start of a new helix. In all cases, the ensemble of η-θ angles was similar. Taken together, these results all show that if a given motif actually does represent a single conformation or shape of a molecule, then a limited set of η-θ angles consistently and predictively describes them (Figure 4).

Can η and θ discriminate between similar, yet conformationally distinct structures?

The most remarkable example to date of structural alteration through the mutation of a single nucleotide is apparent from the structures of two pseudoknots (32 nt each). The APK pseudoknot is incapable of causing ribosomal frameshifting (Kang *et al.*, 1996). Conversely, APKA27G can carry out ribosomal frameshifting (Kang & Tinoco, 1997). Structural differences due to the A27G mutation in the latter include the formation of a G27-U13 base-pair, the appearance of stacking interactions between G27 and G28 and the introduction of a bend between the two helices of the molecule.

The formation of the G27-U13 base-pair and G27-G28 stacking interactions result in a lengthening of the helix in which they are located (Kang & Tinoco, 1997). These changes are reflected in comparative η-θ plots of the two structures (Figure 6). G28 from APK has shifted into the ''helical region'' of the η-θ plot and G27 has shifted to a position where its θ value is normal (Figure 6). The bend between the two helices introduced by the A27G mutation is evident from the shift of A14 out of the

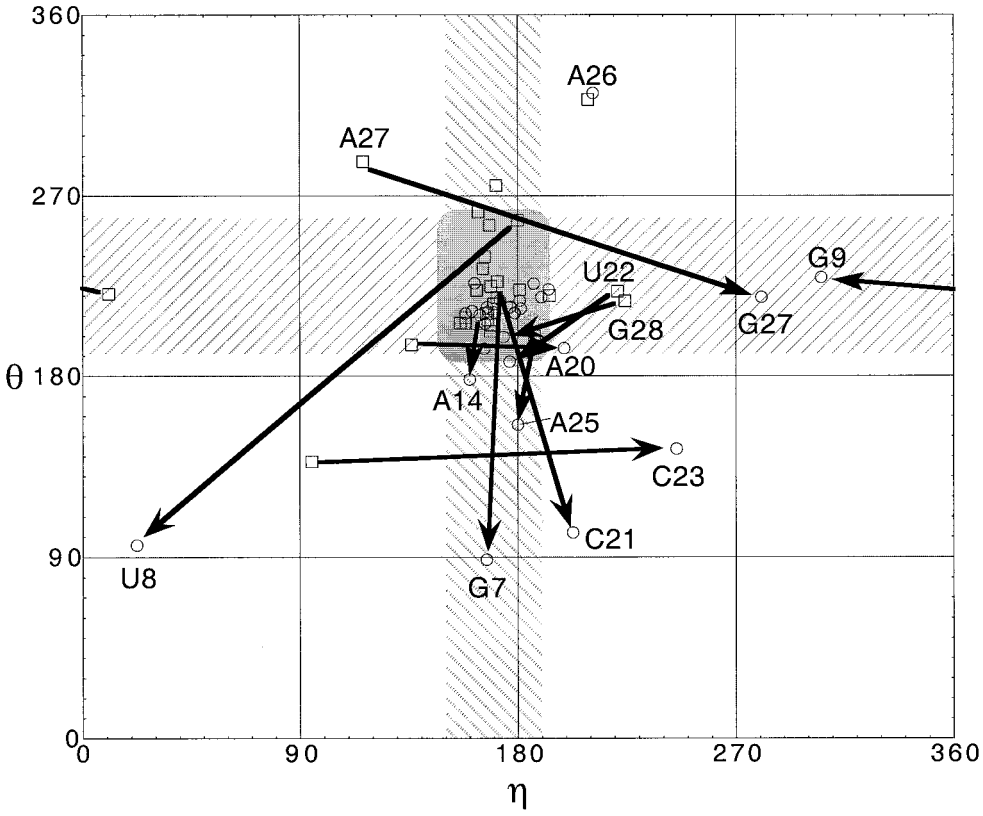

Figure 6. Conformational differences between two different pseudoknots, illustrated on a η-θ plot. Arrows indicate significant shifts in conformational state of corresponding nucleotides from the APK(□) to the APKA27G(○) pseudoknots. Significant shifts include movements into or out of the helical region. The helical region is in gray and the hatched bars are areas with regular η or θ measurements.

helical region. The nature of nucleotide conformation with respect to overall structure can therefore be quantified by η-θ measurements. These include identification of bends, turns, the beginning and end of helices.

In order to determine if morphological differences between these structures could be identified quantitatively, η-θ values for corresponding nucleotides in the two molecules were compared. Movement of corresponding nucleotides from one position to another in an η-θ plot was used was used as a graphical indication of differences in conformation (Figure 6). A strong correlation was observed between shifts in η-θ coordinates and the RMSD of corresponding nucleotides when their surrounding substructures were superimposed. For example, two sets of contiguous nucleotides that undergo considerable shifts in η and θ coordinates are nucleotides 7-9 and 20-23 from the APK and APKA27G pseudoknots (Figure 6). Upon superimposition of their substructures, the overall RMSD of nucleotides in this region were 5.2 and 4.6 Å, respectively. In contrast, when nucleotides 2-6, 10-13 or 15-19 are superimposed, all of which have relatively unchanged η-θ values, their overall RMSD values are 1.7, 0.9 and 1.1 Å, respectively.

Discussion

The η-θ parameters presented here define a useful shorthand reference for describing nucleic acid structure. We have shown above that η-θ parameters can identify groupings of nucleotides that share common conformational characteristics. We have shown that structurally conserved motifs can be characterized by specific sets of predictable η and θ angles. Finally, we have shown that gross morphological changes in a molecule can be characterized by shifts in η-θ parameters. Although we have not examined DNA structure, we expect that it can be analyzed in a similar manner.

There have been other attempts to simplify nucleotide conformational description. These have included modular treatment of structure (Westhof *et al.*, 1996) and reduction of the dimensionality of backbone description. Examples of the latter include virtual bond descriptions (Olson & Flory, 1972; Olson, 1975), and principal component analysis of torsional correlations (Beckers & Buydens, 1998). However, the work presented here is the first study we know of that ascribes specific non-

helical nucleotide conformations to well-defined two-dimensional measurements.

It is worthwhile to consider the general characteristics of the η-θ plot and to compare them with historical analyses of peptide structure. In addition to regions of highly favorable conformational distribution, there are "forbidden regions" in a Ramachandran plot, which describe configurations of the peptide backbone that are sterically or electronically unfavorable. It is tempting to speculate that there are similar "forbidden regions" in the η-θ plot presented here, particularly when one considers the low density of points in the continuous region that encompasses the upper right ($\sim\eta = 270\text{-}360^\circ$, $\theta = 300\text{-}360^\circ$), and lower left ($\sim\eta = 0\text{-}90^\circ$, $\theta = 0\text{-}90^\circ$) areas of the plot. While this is suggestive of a forbidden zone, the limited size of the current database makes it premature to define conformationally undesirable regions of the plot. One regional feature that may be relatively independent of database size and content is the significance of distance between a conformational region and the center of the plot. We observe that η-θ coordinates that are found the farthest from the center (or from the helical region) represent particularly large deviations from normal RNA helical structure (Figure 3(c) and (e) to (h)). Therefore, regions descriptive of the largest backbone distortions are found closer to the edges of the plot.

A particularly surprising finding was that η and θ were able to uniquely specify motifs in which the distinguishing feature is an altered orientation of the nucleotide base plane (consult sections on individual regions of the η-θ plot). After all, the glycosidic angle, χ, is not implicitly included in the calculation. Parameters η and θ are backbone torsion angles and might not be expected to reflect base orientation. That they do must be related to the fact that conventional torsion δ is known to vary with χ, and δ is part of pseudotorsion θ. Nonetheless, it is remarkable that the connection between backbone and base-plane disposition is so faithfully maintained through the pseudotorsional analysis.

It is important to emphasize that the characterized regions were chosen by attempting to sample a wide variety of different positions on the plot and by simple visual inspection of nucleotide clustering. It would have been possible to choose the regions by using a clustering algorithm capable of computationally identifying groupings of points on the η-θ plot. This would have applied powerful statistical analysis to choose regions of a specific shape and to discard some of the data that did not fit specific structural categories. Perhaps this would be, in some ways, more precise. However, we felt that there are several problems with an algorithmic treatment of the current database. (A) The database is too young, undeveloped, and dependent on a relatively small group of structures that represent a non-random sampling of RNA structural form. However, the database has been expanding rapidly (Uhlenbeck *et al.*, 1997) and in a few years it will be appropriate to repeat our entire analysis, potentially with the help of a clustering algorithm to define regions with greater accuracy. (B) We wanted to examine each point in our chosen regions, including those that did not fit our descriptors. A statistical analysis that resulted in elimination of data points might abrogate our attempts to understand the trends in random scatter through the plot and the reasons why certain nucleotides defy specific regional definitions. (C) The limited maturity of the database makes it more appropriate to use simple shapes (rather than carefully calculated probability distributions) to define the regions, as this underscores the low resolution of our knowledge at the present time. We hope that our entire mode of analysis will be subject to large improvements over time: better regions may be chosen or redefined, and perhaps some regions will even be removed as the database becomes more complete and more representative of RNA structure in general.

Because the η-θ analysis allows a two-dimensional representation of conformational state in a macromolecule, there are obvious analogies between this approach and Ramachandran maps. However, there are also several important differences: (1) initial Ramachandran maps were determined, in part, by modeling the allowed conformations of polyalanine (Ramachandran *et al.*, 1963). By contrast, the studies presented here are based on the complete body of known RNA structure. Since we have not determined unique sets of actual backbone dihedrals that correlate directly with η and θ, we are presently unable to relate η and θ to conformational energetics. In fact, an attempt to correlate the single-point AMBER forcefield (Weiner *et al.*, 1986) determined conformational energies of all nucleotides in the database with measured η-θ parameters showed no meaningful relationship (data not shown). (2) Ramachandran studies, as described above, were actively predictive of eventual protein conformation. Our RNA measurements are reflective of what has been seen as conformationally possible thus far, from a limited body of structural knowledge. (3) The energetic conformational characterization of the complete range of nucleotide backbone conformation and its relation to overall nucleotide conformation is, at best, an eight-dimensional problem. The parametric reduction approach of this work does not change the level of this problem. A survey of correlations between previously identified backbone torsion angles and those identified here (data not shown) showed a weak correlation between ζ and θ and no correlation between any others. Pseudotorsional approaches to minimization of polynucleotides have been reported (Harvey & McCammon, 1982), but a systematic determination of nucleotide conformational preference remains a daunting problem.

Despite these distinctions from a Ramachandran map, the η-θ plot is a valuable means for rapidly

representing the conformational state of a molecule and parsing individual nucleotides into specific regions likely to have a certain motif. There are, in fact, many reasons why the simplified η and θ formalism may be useful. First of all, the AMIGOS program provides a way to rapidly analyze a new RNA structure for regions that contain unusual conformation. At the present time, structures are simply examined visually, with a potential for bias and non-systematic classification of observed motifs. By contrast, a rapid AMIGOS analysis of a structure provides a complete set of conventional and pseudotorsion angles from which one can detect deviations from standard *A*-form geometry, or from other structures. As we report here, it is the pseudotorsions η and θ that are particularly useful in this regard, since standard torsions vary too greatly even within *A*-form structure to be used as descriptors of altered conformation. These tools and approaches are a potentially useful form of quality control for the evaluation of new NMR and crystal structures of nucleic acids. Second, the η and θ formalism provides an intuitively accessible representation of overall conformation, since the relevant parameters can be presented on a two-dimensional (rather than eight-dimensional) plot. Thus an individual can immediately see which residues in a large structure or database fall into specific conformational states. This has important implications for the application of η and θ parameters in research and as an educational resource for teaching.

Perhaps the most important implication of reducing dimensionality in nucleotide analysis is that it may now be possible to write RNA and DNA modeling programs that create a realistic set of conformational ensembles without biasing the search through subjective limitation of input parameters. With this approach, it should be possible to reduce the computer time and input complexity that are necessarily required for tackling the multi-dimensional problem of nucleotide conformational analysis.

Methods

Pseudotorsions η and θ were assigned from among different pairs of atoms that can be used to define pseudobonds and pseudotorsions for use in conformational analysis. All possible combinations were tested. In addition to P and C4′, there are two other potential sets of points along the RNA backbone (O5′ and C3′; C5′ and O3′) that can reduce it to two similarly spaced pseudobonds that encompass three atoms. Using each of these atom pairs, we measured the pseudotorsions that they define for all nucleotides in the database (see below) and then plotted these on two-dimensional graphs. No discrete groupings were apparent using C5′ and O3′ as pivot points. Using O5′ and C3′, there were more defined distribution patterns, but the relationships between specific types of conformation and specific pseudotorsional coordinates were considerably less distinct than in the case of P and C4′ (which define η and θ). For example, nucleotides in the C2′-*endo* and C3′-*endo* conformation could not be distinguished and were scattered throughout the plot in contrast to what is seen when using P and C4′ as pivot points (Figure 2). Thus, we focussed on characterizing the plots that result from defining P and C4′ as pivot points and η and θ as pseudotorsions.

To create a database for conformational analysis, we conducted searches of the NDB (Berman *et al.*, 1992) and PDB (Bernstein *et al.*, 1977) for all non-redundant examples of exclusively RNA molecules whose structures had been solved by X-ray or NMR techniques, were larger than four nucleotides and featured secondary structure interactions. The resultant database of 52 structures was used for this analysis. For NMR structures with multiple models, the first model in the file was used for comparisons unless otherwise indicated. For crystal structures where the asymmetric unit contains more than one biologically relevant structure, the first in the file was used, while in those cases where there is more than one example of a particular macromolecule (such as tRNA), the most established version was used (in this case, the highest-resolution structure of $tRNA^{Phe}$; Westhof & Sundaralingam, 1986).

The database was analyzed using a program we have developed: Algorithmic Method of Identifying and Grouping Overall Structure (AMIGOS). Since AMIGOS was written as a perl4 script, file handling is simple and there is no need to compile before running. For every PDB formatted coordinate file in a user-specified directory, AMIGOS will create tables of torsional angles between four designated atoms for each non-terminal nucleotide in a nucleic acid chain. By default it tabulates α, β, γ, δ, ε, ζ, χ, η and θ. It will also identify all nucleotides where a given measurement falls within user-specified ranges. This last feature can identify nucleotides with common parameters from a variety of molecules.

Because there is a limited set of native mathematical functions in perl4, our determination of torsional measurements is slightly non-conventional. For the four atoms that define a torsion, the algorithm calculates three vectors, $\overrightarrow{\mathbf{v}_{2-1}}$, $\overrightarrow{\mathbf{v}_{2-3}}$ and $\overrightarrow{\mathbf{v}_{2-4}}$, where $\overrightarrow{\mathbf{v}_{i-j}}$ is the vector from atom **i** to atom **j**. $\overrightarrow{\mathbf{v}_{2-1}}$ and $\overrightarrow{\mathbf{v}_{2-3}}$ define a plane. The normal unit vector to this plane is calculated by taking the cross product of these two vectors:

$$\overrightarrow{\mathbf{E}_{321}} = \frac{\overrightarrow{\mathbf{v}_{2-3}} \times \overrightarrow{\mathbf{v}_{2-1}}}{\left| \overrightarrow{\mathbf{v}_{2-3}} \times \overrightarrow{\mathbf{v}_{2-1}} \right|}$$

The normal unit vector to the plane defined by the second, third and fourth atomic coordinates, $\overrightarrow{\mathbf{E}_{324}}$, is calculated in an analogous manner. The angle between the normal vectors is equal to the angle between the planes, which is equal to the torsion angle, φ, that we are interested in. The cosine of φ is calculated using the following relation:

$$\cos\phi = \overrightarrow{\mathbf{E}_{321}} \bullet \overrightarrow{\mathbf{E}_{324}}$$

The sine of φ is calculated using the following relation:

$$\sin\phi = sgn\left(\overrightarrow{\mathbf{v}_{2-4}} \bullet \overrightarrow{\mathbf{E}_{321}} \right)[1 - \cos^2\phi]^{1/2}$$

In which $sgn(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|$. These are combined and the angle is calculated using the arctan2 function. Measurements are converted from radians to degrees and from

the range of −180°-180° to 0°-360° to facilitate analysis. The validity of the computed torsion angles was checked by comparing AMIGOS calculations of all standard torsion angles to empirical values reported in the literature and to those directly measured using INSIGHT II (MSI) representations of specific structures. After obtaining η and θ values for a structure, the parameters were plotted in two dimensions using Kaleidagraph (Abelbeck Software).

Visual analysis of RNA structures was performed using INSIGHT II (MSI). Key structural features we sought for describing components of individual motifs and η-θ regions included the nature of base-pairing partners, the presence of 3′ and 5′ base-stacking partners, orientation of a base relative to surrounding stacking axis (the imaginary line that passes through the planes of bases that are stacked upon each other), involvement in secondary or tertiary interactions, angles between the planes of adjacent bases relative to each other, the involvement in or with structural motifs and effect on the overall direction of the backbone (bends or turns). The Supplementary Material includes a Table containing the η and θ coordinates of each nucleotide included in the main η-θ plot (the contents of Figure 3), together with the sequence position of each residue, the structure from which it was obtained and the PDB database entry. AMIGOS is distributed through our website (http://cpmcnet.columbia.edu/dept/gsas/biochem/labs/pyle) or by request to A.M.P. or C.M.D. In this work, all calculations and molecular representations were performed on an Indigo 2 computer (Silicon Graphics) containing an R10,000 chip.

Acknowledgments

The authors thank Steve Harvey, Art Palmer and Dana Abramovitz for helpful discussions and comments on the manuscript. In addition, we thank Francois Major, Eric Westhof, Tao Pan and Olke Uhlenbeck for helpful comments on poster presentations of this work. The computing resources for supporting this work were provided through an NSF National Young Investigator grant to A.M.P., who is an assistant investigator with the Howard Hughes Medical Institute. Additional support was also provided by a Ford Foundation Pre-Doctoral Fellowship to C.M.D.

References

Abramovitz, D. L. & Pyle, A. M. (1997). Remarkable morphological variability of a common RNA folding motif: the GNRA tetraloop-receptor interaction. *J. Mol. Biol.* **266**, 493–506.

Allain, F. H. T. & Varani, G. (1995). Structure of the P1 helix from group-I self-splicing introns. *J. Mol. Biol.* **250**, 333–353.

Baeyens, K. J., DeBondt, H. L., Pardi, A. & Holbrook, S. R. (1996). A curved RNA helix incorporating and internal loop with G-A and A-A non-Watson-Crick base-pairing. *Proc. Natl Acad. Sci. USA,* **93**, 12851–12855.

Beckers, M. L. M. & Buydens, L. M. C. (1998). Multivariate analysis of a data matrix containing A-DNA and B-DNA dinucleoside monophosphte steps: multidimensional Ramachandran plots for nucleic acids. *J. Comput. Chem.* **19**, 695–715.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* **63**, 751–759.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macrmolecular structures. *J. Mol. Biol.* **112**, 535–542.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science,* **273**, 1678–1685.

Chang, K. Y. & Tinoco, I. (1997). The structureof an RNA ''kissing'' hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.* **269**, 52–66.

Conte, M. R., Conn, G. L., Brown, T. & Lane, A. N. (1997). RNA duplex R(CGCAAAUUUGCG)2: comparison with the DNA analogue D(CGCAAATTTGCG)2. *Nucl. Acids Res.* **25**, 2627–2634.

Dock-Bregeon, A. C., Chevrier, B., Podjarny, A., Johnson, J., Debear, J. S., Gough, G. R., Gilham, P. T. & Moras, D. (1989). Crystallographic structure of an RNA helix–(U $(UA)_6$ $A)_2$. *J. Mol. Biol.* **209**, 459–474.

Harvey, S. C. & McCammon, J. A. (1982). Macromolecular conformational energy minimization: an algorithm varying pseudodihedral angles. *Comput. Chem.* **6**, 173–179.

Heus, H. A. & Pardi, A. (1991). Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science,* **253**, 191–194.

Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H. M. & Pardi, A. (1996). A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.* **264**, 968–980.

Kang, H. & Tinoco, I. (1997). A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. *Nucl. Acids Res.* **25**, 1943–1949.

Kang, H., Hines, J. V. & Tinoco, I. (1996). Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J. Mol. Biol.* **259**, 135–147.

Klinick, R., Sprules, T. & Gehring, K. (1997). Structural characterization of three RNA hexanucleotide loops from the internal ribosome entry site of polioviruses. *Nucl. Acids Res.* **25**, 2129–2137.

Laing, L. G. & Hall, K. B. (1996). A model of the iron responsive element RNA hairpin loop structure determined from NMR and thermodynamic data. *Biochemistry,* **35**, 13586–13596.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.

Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic-acids. *J. Biomol. Struct. Dynam.* **6**, 63–91.

Legault, P., Li, J., Mogridge, J., Kay, L. E. & Greenblatt, J. (1998). NMR structure of the bacteriophage lambda N peptide/box B RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell,* **93**, 289–299.

Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E. & Cedergren, R. (1991). The combination of symbolic and numerical computation for 3–dimensional modeling of RNA. *Science,* **253**, 1255–1260.

Olson, W. K. (1975). Configurational statistics of polynucleotide chains: single virtual bond treatment. *Macromolecules,* **8**, 272–275.

Olson, W. K. & Flory, P. J. (1972). Spatial configurations of polynucleotide chains. I. Steric interactions in polyribonucleotides: a virtual bond model. *Biopolymers,* **11**, 1–23.

Peterson, R. D. & Feigon, J. (1996). Structural change in rev responsive element RNA of HIV-1 on binding rev peptide. *J. Mol. Biol.* **264**, 863–877.

Portmann, S., Usman, N. & Egli, M. (1995). The crystal structure of r(CCCCGGGG) in two distinct lattices. *Biochemistry,* **34**, 7569–7575.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99.

Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Scott, W. G., Finch, J. T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell,* **81**, 991–1002.

Shen, L. X. & Tinoco, I. (1995). The structure of an RNA pseudoknot that causes efficient frameshifting in mouse mammary-tumor virus. *J. Mol. Biol.* **247**, 963–978.

Uhlenbeck, O. C., Pardi, A. & Feigon, J. (1997). RNA structure comes of age. *Cell,* **90**, 833–840.

Varani, G. (1995). Exceptionally stable nucleic hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 379–404.

Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986). A new forcefield for molecular mechanical simulation of nucleic-acids and protein (AMBER). *J. Comput. Chem.* **7**, 230–252.

Westhof, E. & Sundaralingam, M. (1986). Restrained refinement of the monoclinic form of yeast phenylalanine transfer RNA. Temperature factors and dynamics, cooordinated waters, and base-pair propellor twist angles. *Biochemistry,* **25**, 4868–4878.

Westhof, E., Masquida, B. & Jaeger, L. (1996). RNA tectonics: towards RNA design. *Folding Des.* **1**, R78–R88.

Edited by D. Draper

(Received 10 July 1998; received in revised form 14 September 1998; accepted 16 September 1998)

http://www.hbuk.co.uk/jmb

Supplementary material comprising two Tables is available from JMB Online