ELSEVIER

MANIP: An interactive tool for modelling RNA

C. Massire and E. Westhof

Equipe de Modélisation et de Simulation des Acides Nucléiques, UPR 9002 Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, France

Large RNA structures can be viewed as assemblies of smaller units or modules that are usually clearly identified (helices, hairpin loops, other recurrent motifs, etc.). We have developed a program, MANIP, which allows the rapid assembly of separate motifs (each with a specified sequence) into a complex three-dimensional architecture. The already determined modules are present in a database from which they can be extracted with the appropriate sequence. Their assembly is performed in real time on the computer screen with buttons and dials that command rotation and translation of any chosen fragment with respect to the chosen pivot, or that generate all possible variations of any torsion angle within a specified segment either in the 5′ or in the 3′ direction. The possible in-built manipulations follow the general stereochemical rules of RNA structure. MANIP automatically recognizes and displays the allowed and nonallowed hydrogen bonds between the residues. The program is interfaced with a rapid and automatic online refinement tool of partial or full assemblies, NUCLIN-NUCLSQ. The refinement protocol incorporates canonical as well as noncanonical base pairing constraints together with restraints imposed by covalent geometry, stereochemistry, and van der Waals contacts. The computer package runs on UNIX Silicon Graphics workstations and is written in C with OpenGL and X11/Motif libraries. © *1999 by Elsevier Science Inc.*

Keywords: RNA modelling, RNA modules, 3D structure

INTRODUCTION

Ribonucleic acids (RNAs) are negatively charged biopolymers assembled from four different types of monomers or residues. Each monomer is made of an invariant phosphorylated sugar to which is attached one of the four standard nucleic acid bases; the pyrimidines uracil (U) and cytosine (C) and the purines guanine (G) and adenine (A). The first level of organization is, thus, the sequence of bases attached to the sugar-phosphate backbone. In salty water, the RNA molecules fold back on themselves via Watson–Crick base pairing between the bases (A with U, G with C or U), leading to double-stranded helices interrupted by single-stranded regions that form bulges, internal loops, or hairpin loops. In helices, the phosphate groups are exposed to the solvent and the base pairs stack upon each other with some of their hydrophilic and polar atoms accessible to solvent molecules in either the shallow or the deep groove, both typical of double-stranded RNA helices. The enumeration of the base-paired regions or helices constitutes a description of the secondary structure, i.e., the second level of organization. The secondary structure of an RNA molecule is experimentally accessible and its content measurable. Under appropriate conditions, RNA molecules undergo a transition to a three-dimensional (3D) architecture in which the helices and the unpaired regions are precisely organized in space. This folding process depends on the presence of divalent ions, such as magnesium ions, and on the temperature.[1] The tertiary structure is the level of organization relevant for biological function.

The basis of the modelling approach[2] for which a computer program is presented in this article assumes a working secondary structure that can be partitioned in recognizable elements, some of which are recurrent and even autonomous in folding. According to the hierarchical view of RNA folding,[2–4] helices and hairpin loops would form first. Helices would then interact locally end-to-end through stacking or by forming pseudoknots or triple helices. And, finally, these autonomously folded subdomains or modules would associate cooperatively by loop–loop base pairings and by numerous contacts involving loops, bulges, and helices, leading to a final, compact structure. Recent breakthroughs in RNA structure determination establish that complex RNA structures result from the hierarchical assembly of recurrent three-dimensional motifs.[5,6] Our modelling approach follows this framework and consists of assembling on a graphics system the overall architecture from preconstructed 3D motifs (see Figure 1). More precisely, we understand a motif

Color Plates for this article are on pages 255–257.

Address reprint requests to: Dr. E. Westhof, Equipe de Modélisation et de Simulation des Acides Nucléiques, UPR 9002 Structure des Macromolécules Biologiques et Mécanismes de Reconnaissance, Institut de Biologie Moléculaire et Cellulaire du CNRS, 67084 Strasbourg, France. Fax: +33 (0) 3-88 417066. E-mail: westhof@ibmc.u-strasbg.fr.

Journal of Molecular Graphics and Modelling 16, 197–205, 1998

655 Avenue of the Americas, New York, NY 10010

1093-3263/99/$–see front matter
PII S1093-3263(99)00010-8

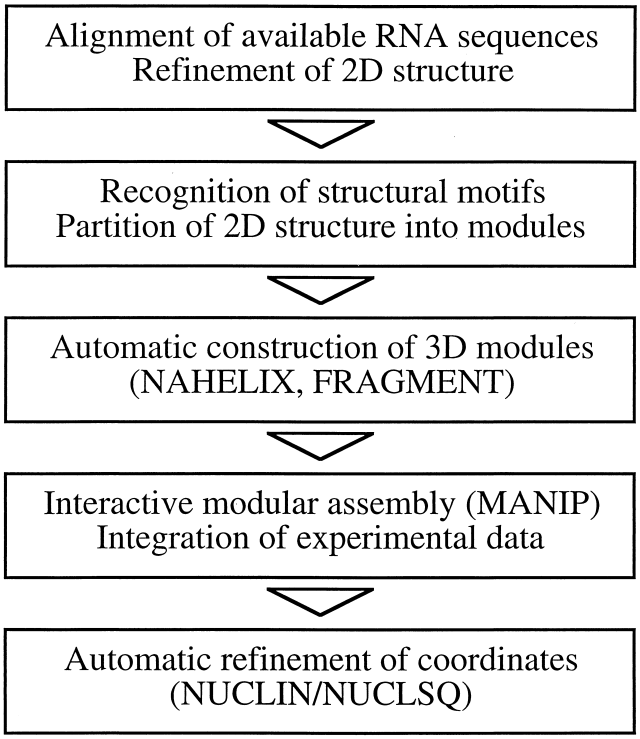


Figure 1. Main steps of the modelling process used with MANIP.

as a set of nucleotides that are in close vicinity and that interact in a specific way independently and autonomously of the rest of the sequence or other motifs (at least at the level of the resolution of the modelling technique). Alone, nucleotides can adopt a huge number of conformations, owing to the six internal torsion angles and sugar puckering. Within a polynucleotide strand, however, the conformational space of each nucleotide is dramatically reduced as steric congestion and internucleotide interactions take place, primarily through base stacking and base pairing via the formation of hydrogen bonds.[7] Therefore, most nucleotides seen in known structured RNA tend to adopt a conformation chosen within a limited set of possible conformations.[8] In the most extreme case of regular helices, the same known conformation is shared by all nucleotides, allowing thus the construction of a model of any regular A-form helix with good accuracy. In a sense, regular helices constitute the most common RNA motif. Comparative sequence analysis has proved to be the most successful way to infer the presence of regular helices from a set of aligned, homologous sequences.[9] Indeed, the formation of helices is revealed by the identification of nucleotides that covary, i.e., which change accordingly to each other in order to maintain a strict Watson–Crick complementarity (A with U or C with G).

Nucleotides that do not form regular helices are involved in terminal loops, internal loops, multibranched loops, or single-stranded junctions, where they likely interact in a more specific way, noticeably through nonstandard base pairings, resulting in motifs that are either sequence specific or for which base covariations exist. Once again, comparative analysis of sequences will often be the key to the identification of nucleotides that are involved in noncanonical pairings and display a specific pattern of variation. GNRA tetraloops probably constitute the best examples of 3D structural motifs. These loops were first pointed out to be abundant in 16S and 23S rRNA sequences[10] before being recognized as well in other structured RNAs, such as group I or group II introns[3] or RNase P RNA.[9] Modelling studies first[11] and then NMR[12] and crystallographic[13] studies of loops of the GNRA family, belonging to different RNAs, show how these peculiar loops are similarly stabilized through the stacking of the three last bases and the formation of a sheared G:A base-pair, thus justifying the conservation of the consensus sequence. Other recurrent motifs, like the GAAA receptor[14] or the internal loop E of 5S rRNA[15] have similarly been identified through their specific patterns of sequence variation (see Color Plate 1 and Table 1[16–27]). It seems reasonable to assume that, as for GNRA tetraloops, identical consensus sequences would lead to identical folding for any given motif, at least to the resolution presently accessible to modelling techniques.

DESCRIPTION OF THE PROGRAM

Overview

MANIP constitutes the central step of a whole modelling process resumed in Figure 1. A MANIP work session typically contains three consecutive steps. It begins with the building of secondary structure motifs using the NAHELIX and FRAGMENT programs. The RNA fragments are then assembled into a three-dimensional architecture according to various biological and stereochemical data (with the help of the MANIP toolbox itself). Finally, the assembled units are refined using NUCLIN-NUCLSQ.[27] MANIP should be viewed as a toolbox, where the user can found a variety of tools that help to design a three-dimensional model, rather than an *ab initio* automatic modelling program.

Main drawing options, including side-by-side or overlapping stereo as well as depth-cueing, are available from a pop-up menu. MANIP uses a dial device extensively for rotations, translations, zoom, and slab. These features are also provided

Table 1. Common structural motifs extracted from NMR or crystallographic studies gathered in the FRAGMENT database

Motif	PDB access number	Reference
UNCG loop		16
CUUG loop	1RNG	17
UGAA loop	1AFX	18
GNRA loops	1ZIF, 1ZIG, 1ZIH	12, 19
GAAA loop and receptor	1GID	5
GA:GA tandem	1HMH	13
UG:UG tandem	315D	20
GU:GU tandem	332D	21
J4/5 internal loop	1GID	5
Bacterial 5S rRNA loop E	354D, 1A4D, 1A51	22, 23
Sarcin/Ricin loop	1SCL, 430D	24–26
$tRNA^{Asp}$ anticodon loop	3TRA	27
$tRNA^{Asp}$ thymine loop	3TRA	27

by the pointer and modifier keys, making the program easy to use on standard workstations. Transient windows allow the partition of RNA strands into multiple fragments, which can be independently colored or displayed (see Color Plate 2). MANIP computes the coordinates of missing hydrogen atoms, a feature that allows an accurate recognition of hydrogen bonds. Structures are displayed either as a wireframe (including hydrogen atoms or not) or as a backbone trace that passes only through phosphorus atoms. Both types of visualization may be combined as requested. The MANIP package incorporates some older programs written in F77, although all new routines are in C. The package will therefore use various coordinate files (format *.hd of NUCLIN-NUCLSQ, especially) although the *.pdb format is read as input and furnished as output. Several small tools allow for conversion between *.hd, *.pdb, or mmCIF format as well as for file and coordinate manipulations (HD2PDB, PDB2HD, etc.).

Building of 3D motifs (NAHELIX and FRAGMENT)

MANIP is interfaced with two programs that generate structural motifs, NAHELIX and FRAGMENT. The NAHELIX program allows the creation of regular or irregular helices of any sequence, according to different conformations (although A-type RNA helices are the most frequent in RNAs). In the case of helices following a strict Watson–Crick complementarity, only the sequence of the first strand need be entered (see the NAHELIX interface in Color Plate 3). However, entering noncomplementary sequences for both strands allows the incorporation of mismatches such as G:U wobble pairs. Moreover, bulges or dangling ends may be obtained by inserting gaps in the opposite strand (denoted as X in the input sequence). NAHELIX computes, thus, any composite helix (made of a variable number of separated strands) where all nucleotides are stacked within the frame of a regular helix. The gap closure is performed at the next stage by using the Torsion tool.

The FRAGMENT program allows the extraction of any interesting structural motif or module from an already existing structure (see Color Plate 2). The user can fit the RNA sequence of interest, typically somewhat similar to the original one, into the desired three-dimensional frame. In this two-step process, the program MAKEFGM is first used for the writing of a database file that contains the coordinates of the sugar-phosphate backbone, if the motif of interest is not yet present in the database. This file contains also the coordinates of the N1 or N9 atom of the base, as well as the computed value of the χ angle (O4′–C1′–N1–C2 for U or C and O4′–C1′–N9–C8 for G or A). In a second step, the program FRAGMENT itself uses this database file to fit a specific sequence into the frame. During this operation, coordinates of the sugar-phosphate backbone are left unchanged. The replacement of the original base by the new one is done while keeping the same orientation of the base with respect to the sugar (χ angle). It is the user's responsibility to check whether the change of sequence results in the formation of bad contacts or overlap of bases, in which case minor adjustments help the refinement step that will eventually lead to a clean structure. Common motifs have been collected in a motif database (see Table 1 and Figure 2). In most cases, these motifs end with a canonical base pair that will facilitate their insertion in future models. Motifs in this database have undergone a standard refinement procedure that leaves minimized and relaxed structures. This database is continuously growing with the crystallisation of interesting RNA motifs with an increasing resolution.

Assembly of motifs (MANIP)

The assembly of structural motifs is made interactively on the computer screen, the user being responsible for the three-dimensional arrangement of the motifs with respect to each other. MANIP provides the user with rulers and tools that allow the independent manipulation of structural motifs in a way that respects RNA stereochemistry.

Most of these tools may be activated through a 32-button box device or from a command window that mimicks the button box. The standard appearance of the command window is shown in Figure 3A. The buttons from the rightmost column of the box allow the user to pick atoms on the screen in order to display distances, angles, or dihedral angles (buttons DIST, ANGL, and DIED, respectively) at any time during the session. There is no button to add atom labels since this is simply done by clicking the desired atom in the viewing area. Atom, distance, angle, and dihedral angle labels may be erased independently of each other by pressing the LABL_D, DIST_D, ANGL_D, or DIED_D button, respectively.

The GROUP option allows the user to gather different strands in order to apply the same transformations simultaneously to these strands, while the UNGRP button performs the reverse operation. The three buttons in the centre of the box are used to quickly display transient windows that allow the partition of the model into strands, fragments, as well as the visualization of the list of base pairs recognized by the program. These windows are visible in Color Plate 2.

On the left side of the box are gathered the commands that modify model coordinates (buttons CUT, BIND, MOVE, TORS_3, and TORS_5). The CUT button allows the user to cut a given strand at the level of the phosphodiester bond to give 5′-P and 3′-OH ends, while the BIND button performs the reverse operation.

Pressing the MOVE button after having picked an atom on the screen will shift the program into the MOVE mode (see Figure 3B). An action on the dials will then move only the fragment (fragment A) that contains the picked atom, which acts as a pivot. A simple press on the SWAP button toggles the use of the dialbox—i.e., it allows one to choose whether rotations and translations apply to the fragment A only or to the complete model. Once an atom (which is not part of the selected fragment) is picked in the MOVE mode, buttons SUPER, ADFR_3, ADFR_5, OK, and CANCEL are made available. The SUPER button allows the superposition of the pivot atom on the last picked atom, while the ADFR_3 and ADFR_5 buttons allow the stacking of the moving fragment under the 3′ end or the 5′ end of fragment B. The ADFR_3 and ADFR_5 functionalities are, thus, a convenient way to assemble different motifs to each other, provided that each motif ends with a canonical base pair to ensure helical continuity. The user exits the MOVE mode by pressing either the OK or CANCEL button, depending of whether or not the modifications must be saved.

Torsion angles either down the 5′ or the 3′ direction of the last selected single-stranded fragment are performed by pressing the TORS_5 or TORS_3 button, respectively. The com-

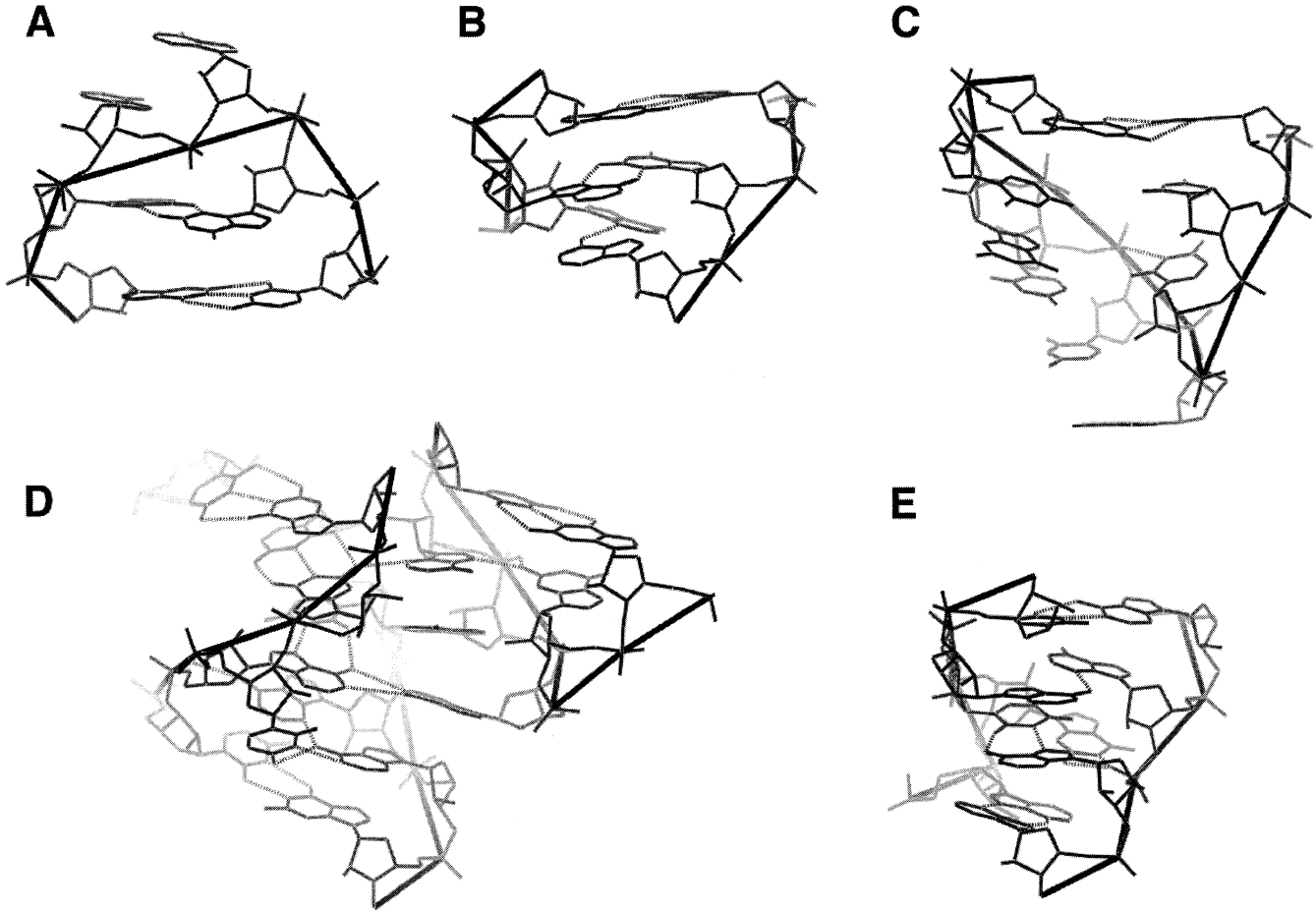


Figure 2. Three-dimensional representation of some common structural motifs: (A) A GAAA tetraloop[12]; (B) a tandem of sheared GA:GA base pairs[13]; (C) the anticodon loop of tRNA^{Asp}[27]; (D) the GAAA loop–GAAA receptor motif from the P4–P6 domain of the Tetrahymena *intron[5]; and (E) an internal loop (taken from loop E of 5S rRNA).[26]*

mand box then displays the additional buttons SWAP, PREV, NEXT, TORS_ED, OK, and CANCEL buttons (see Figure 3C). Through the dial box, the user can then interactively modify torsion angles in the neighbourhood of the selected atom. A dial box typically contains eight dials, allowing the simultaneous mapping of the five torsion angles (ϵ, ζ, α, β, and γ) that separate two consecutive sugar rings. The sixth dial is used to map the torsion angle along the glycosidic bond (χ) of the selected nucleotide. The last torsion angle (δ) is not directly mappable since it is internal to the sugar ring. Instead, the last two remaining dials are used to trace the pucker phase (P) and the amplitude (τ_m) of the current nucleotide, allowing a smooth deformation of the nucleotide along the pseudorotation wheel.[28] There is no limit in the number of nucleotides or in the size of the fragment being carried by variations in any of the torsion angles in the selected fragment. As in the MOVE mode, pressing the SWAP button will change the affectation of the dials, allowing the user to perform torsions along the natural torsion angles of the current nucleotide, or alternatively to apply rotations or translations to the entire model. One can easily jump from one nucleotide to another by pressing the PREV or NEXT button. Finally, modifications must be confirmed by pressing the OK button, or may be ignored by pressing the CANCEL button.

Stereochemical refinement (NUCLIN-NUCLSQ)

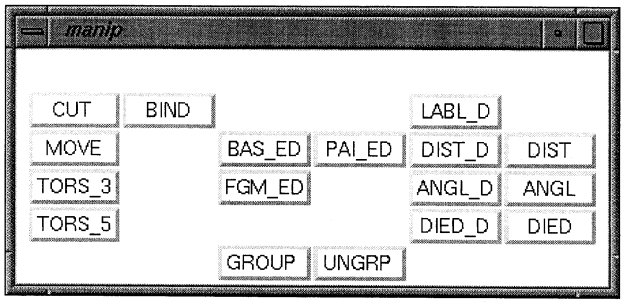
MANIP provides an interface with the stereochemical refinement programs NUCLIN-NUCLSQ[27] that ensure correct stereochemistry at the nucleotide level by removing bad contacts between atoms and the refinement of canonical base pairs. We extended here the refinement process to the automatic treatment of noncanonical base pairs, with the recognition of common base pairings from the network of identified H-bonds. This feature allows one to sort quickly hydrogen bonds between regular interactions and short distance contacts (see Implementation, below). Rather extensive write-ups on NUCLIN-NUCLSQ are present in the package.

IMPLEMENTATION

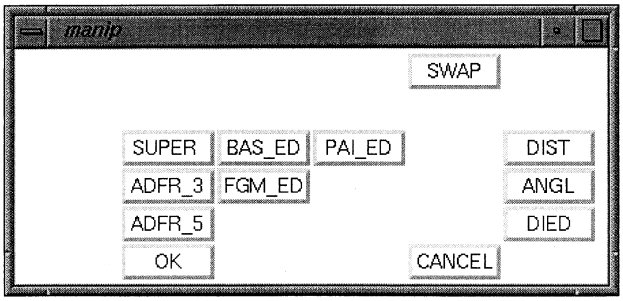
Connection tables

The stereochemical reliability of the model under construction often conditions that of the final structure. For ribonucleic acids, biochemists know how nucleotides are first topologically defined by a set of chemical bonds between given atoms, and naturally expect to recognize this topology in the structure that is displayed on the screen—although it represents most of the

A



B



C

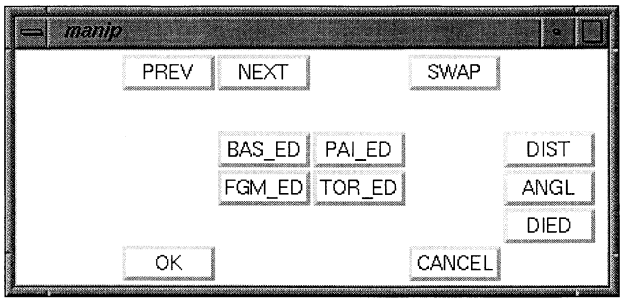


Figure 3. (A) View of the command window in the default command mode. (B) View of the command window in the MOVE mode. (C) View of the command window in the TORS_3 or TORS_5 mode.

time an unfinished model with little stereochemical relevance since it is precisely under construction! It is quite common indeed to have to deal, at least for a while, with "structures" that include overlapping parts, and the ability to distinguish such situations is therefore essential. This is not obvious, however, since most modelling programs handle only coordinate files that will not explicitly contain topological data. The Protein Data Bank (PDB) file format, the most widely used one for dealing with protein or RNA structures, does indeed include two kinds of fields for atom coordinates, depending on their belonging to an oligomer of supposedly known topology (ATOM fields) or to a uncommon substructure, such as a drug or a highly modified nucleotide (HETATM fields). In the latter case, additional CONNECT fields may explicitly define the localization of the chemical bonds to display but, in most cases, the topological information is not present. A common way to bypass this lack of data is to recognize what atoms are actually bound to each other on the basis of considerations of interatomic distances and angles. This approach has, however, two main drawbacks: first, the calculation of the huge number of interatomic distances may be time consuming and, second, it demands that the set of coordinates be refined enough to allow the recognition of all actual bonds and of only those. This approach is therefore well suited for the modelling of a large variety of relatively small chemical or biochemical compounds. In the case of the modelling of RNA, the decomposition of a large, complex 3D structure into small elementary bricks, chosen among a limited set of nucleotides, is already included in the coordinate files through the recurrent use of ATOM fields. We take benefit of the polymeric nature of RNA with the implementation of connection tables that define once and for all the topology of nucleotides.

An example of such a table is reported in Figure 4 for the uridine residue. Connection files are a collection of different data records, each of these records beginning with a four-character label. The REF1, REFR, and NAME records give the one-character abbreviation, the three-character abbreviation, and the full name of the nucleotide, respectively. Intranucleotidic connections are specified by ATOMS records. An ATOM record is assigned to each atom within the nucleotide and begins with an integer (italicized in Figure 4) indicating the entry rank of the atom in the table. The next four characters are the atom name as it appears in GenBank files, and the remaining numbers are the entry numbers of the atoms that are linked to this atom. Entry numbers are also used in other records that define the way nucleotides are bound to each other. The CTRL records precisely what atom is used for the drawing of the trace of the backbone (typically the phosphorus atom for a nucleotide). The CNX5 and CNX3 fields contain the entry number of atoms at 5′ and 3′ ends, respectively, that must be connected to the 5′ and 3′ nucleotides, respectively. Torsion angles are defined by the specification of entry numbers of four joining nucleotides within a DHDR field. Since we found it more convenient to define torsion angles between sugar rings rather than between phosphates, angles ϵ and ζ contain atoms O3′, C3′, and C4′ of the 5′ nucleotide (see Figure 4, right). These atoms are referred to as entry numbers -1, -2, and -3 in the DHDR fields, respectively, and are defined by the CNX3, CNX2, and CNX1 fields in the connection table of the 5′ nucleotide. Interactive modification of pucker parameters are allowed thanks to the specification of the atoms that constitute the ribose ring in the RING field. Finally, the HBDN and HBAC fields reveal precisely what heteroatoms are hydrogen bond donors and acceptors, respectively.

Base-pair tables

Aside from the nucleotide connection tables, MANIP uses a set of base-pair tables that define the common base pairs that may be recognized and refined. The recognition of base pairs is a two-step process.

First, MANIP looks for all putative hydrogen bonds between atoms declared as hydrogen bond donors and acceptors. In practice, MANIP computes the position of missing hydrogen atoms; then, all hydrogen atom–hydrogen bond acceptor distances are computed, and the ones shorter that a parametrable threshold value (typically in the range 1.5–2.1 Å) are considered as hydrogen bonds. The computation of hydrogen atom positions allow a more accurate determination of hydrogen bonds rather than the direct measurement of the distance between hydrogen bond donors and acceptors, which would have

```
REF1 U
REFR URI
NAME Uridine
CNX5   0
CNX1   5
CNX2   6
CNX3  19
CTRL   0

ATOM   0  P      1   2   3
ATOM   1  O1P    0
ATOM   2  O2P    0
ATOM   3  O5'    0   4
ATOM   4  C5'    3   5
ATOM   5  C4'    4   6   7
ATOM   6  C3'    5   9  19
ATOM   7  O4'    5  10
ATOM   8  O2'    9
ATOM   9  C2'    6   8  10
ATOM  10  C1'    7   9  11
ATOM  11  N1    10  12  18
ATOM  12  C6    11  13
ATOM  13  C5    12  15
ATOM  14  O4    15
ATOM  15  C4    13  14  16
ATOM  16  N3    15  18
ATOM  17  O2    18
ATOM  18  C2    11  16  17
ATOM  19  O3'    6

DHDR epsilon   -3  -2  -1   0
DHDR dzeta     -2  -1   0   3
DHDR alpha      1   0   3   4
DHDR beta       0   3   4   5
DHDR gamma      3   4   5   6
DHDR khi        7  10  11  18
RING ribose     7  10   9   6   5

HBDN   8  16
HBAC   1   2  14  17
```

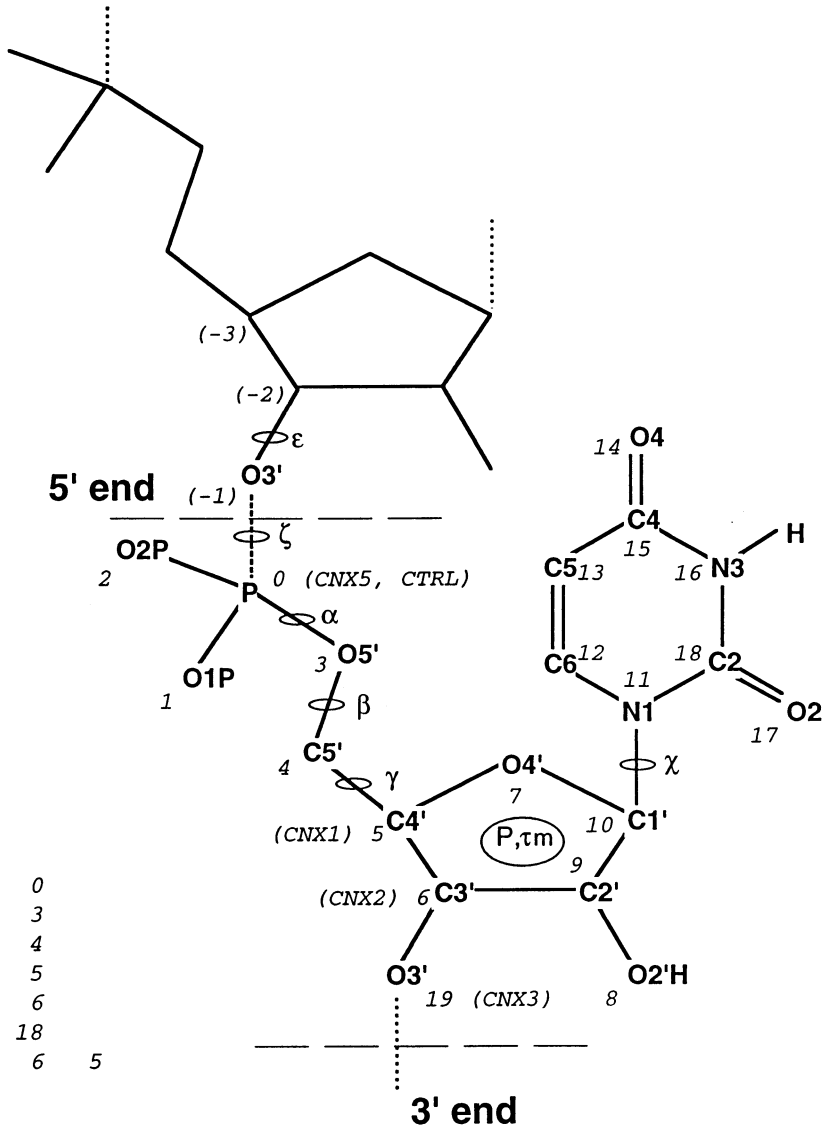


Figure 4. Contents of a simplified connection table for uridine bases (left) and its associated graph (right). For clarity, hydrogen atoms bound to carbon atoms are not represented, although they are normally present in connection tables. Moreover, only the sketch of the 5′ nucleotide is drawn. Standard atom names are reported in boldface characters whereas italicized numbers refer to the internal numbering used by the program (usually the entry number of an atom within the table).

to tolerate a much wider range. Moreover, usual distances between hydrogen atoms and hydrogen bond acceptors are short enough to rely on a single distance measurement to assess confidently the formation of the hydrogen bond without the help of further angle calculations.

Second, hydrogen bonds are sorted by pairing type, according to the specifications included in base-pair tables (see Figure 5). These tables are also made of different records, and the PAIR and NUCL records allow one to distinguish between different types of pairings that will be recognized during the refinement process. The BOND records contains the optimal length of the hydrogen bonds that constitute the pairing. No angle considerations are taken into account during the refinement process. Instead, DIAG records define additional constraints that must be followed to ensure a optimal geometry of the base pair.

A WORKING SESSION: THE MODELLING OF THE L12-P10.1 INTERACTION OF TYPE B RNase P RNA

Color Plates 1 to 6 describe the main steps of a short modelling session. This example is taken from the modelling of RNase P RNA, the ribozyme responsible for the maturation of the 5′ end of tRNAs.[29] The RNase P RNA sequences that follow the so-called type B[30] possess an helical extension, P10.1, that is involved in a loop–helix

interaction with the apical loop of stem P12.[31] A similar loop–helix interaction between a GAAA loop and a so-called GAAA receptor was already identified from comparative sequence analysis of group I and group II introns[14] and later crystallized.[5] Another complex motif is found in the P10.1 stem (see Color Plate 1): an asymmetrical internal loop follows the sequence consensus of the loop E of eukaryotic 5S rRNA, the structure of which has also been determined.[26] Both motifs involve the formation of non-standard base pairs shown in purple, cyan, and blue in Color Plates 1–6. The recognition of these motifs allows the construction of a rather well-detailed model of the L12–P10.1 interaction with good confidence, thus providing a useful frame for the modelling of less constrained parts of RNase P RNA.

The modelling process itself begins with the extraction of the GAAA loop–GAAA receptor motif from the crystallographic structure of the P4–P6 domain of group I intron (see Color Plate 2). Both loop and receptor are extracted at the same time, thus ensuring the conservation of the right geometry of the interaction. The extracted motif is shown in Color Plate 3. In this case, the extraction is not limited to the 11 nt constituting the GAAA receptor and the 4 nt of the GAAA loop; but the closing base pair of the GAAA loop was also extracted in order to allow later the stacking of the regular stem P12. Three nucleotides beyond the G of the G:U pair of the receptor and 5 nt beyond the U of the G:U pair were also extracted in order to close the loop.

The interface to the NAHELIX program, visible in Color Plate 3, was then used to create the 2-bp stem that separates the GAAA receptor from the loop E motif. Regular helices are usually created *de novo*, although in some cases they could be obtained from base pairs that surround interesting motifs in crystallographic structures. The automatic creation of helical segments has, however, the advantage of resulting in more regular helices. At this level of accuracy, we do not attempt to reproduce or include distortions found in crystallographic structures. For example, in the present case, the base pairs that would have been extracted below the GAAA receptor are part of a triple helix. The 2-bp helical fragment is displayed in Color Plate 4, with both strands linked by a solid red line to remind the user that these strands are grouped together. The helical fragment was selected by clicking one of its atoms (for example, G144 O3′) and pressing the MOVE button. A simple click on the C145 P atom, followed by a click on the ADFR_5 button, will make the 2-bp helix stack below the GAAA receptor motif with the optimal RNA helix geometry.

The next step consists of the closing of the loop (see Color Plate 5). The nucleotides that constitute the tetraloop UUUG were already extracted (the first U on the 5′ strand and the remaining bases on the 3′ strand). The closing is partially done by torsion of the dangling 3′ end, in order to bring the 5′ and 3′ ends about 2 Å away from each other. In practice, one selects a nucleotide of the dangling end, then presses the TORS_5 button, thus modifying directly the torsion angles of this nucleotide by action on the dials. The final closure of the loop will be made during the refinement step, using the program NUCLIN-NUCLSQ. Color Plate 6 shows the complete model of the L12-P10.1 interaction, once the extraction and the assembly of the loop E motif, as well as the assembly of the remaining helical fragments, were performed in a similar way.

```
PAIR Watson-Crick
NUCL   A    U
BOND  N6    O4   2.9
BOND  N1    N3   2.9
DIAG  N6    N3   3.7
DIAG  N1    O4   3.7
```

Figure 5. Contents of the file describing the canonical A:U base pair. Diagonal distance restraints are added to prevent horizontal sliding of the bases within a given base pair.

CONCLUSION

MANIP constitutes a quick and easy way to model small- to large-size structured RNAs. The used of multiple connection and pairing tables, although it may seem at first glance tedious, opens indeed further development perspectives and allows, for instance, the precise modelling of RNA–protein interactions.

In constrast to other modelling programs such as MC-SYM,[32,33] MANIP is not an automatic search procedure at the nucleotide level. With our approach, the user, rather than the modelling program, is at every moment responsible for the choice of motifs or nucleotides conformations that must be integrated into the model. This task is, however, made much easier with the interactive access to a motif database. Therefore, in order to construct plausible models, MANIP may demand from the user a more intimate knowledge of RNA structure than do other, more automated, modelling tools. However, in some instances, the conformational search capabilities of MC-SYM could be useful. Therefore, a use of these two packages in combination could constitute a powerful approach.

Models constructed with MANIP will, therefore, reflect first and foremost the knowledge content of the database and the ability of the model builder to extract and use that knowledge. This program has been used to model RNase P RNA,[29] the hairpin ribozyme,[34] and complete group I introns.[35]

AVAILABILITY

MANIP has been developed on a Silicon Graphics Indigo2 workstation, with the help of X11/Motif and OpenGL libraries. The graphic capabilities that the program demands (1280 × 1024 screen with 24 RGB bitplanes and Z-buffer) are nowadays common. More unusual is the use of a dial and button box; this device is, however, hardly dispensable, although its functionalities may be alternatively provided by the pointer and the command window. The program is freely available from the website http://ibmc.u-strasbg.fr/upr9002/westhof/.

ACKNOWLEDGEMENTS

We thank Julien de Murcia for his help in the code writing, as well as Valérie Lehnert and Benoît Masquida for their numerous suggestions. C.M. acknowledges the Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la

Recherche and the Human Frontier Science Program (CRG0291/1997-M to E.W.).

REFERENCES

1 Brion, P., and Westhof, E. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* 1997, **26**, 113–137

2 Westhof, E. Modelling the three-dimensional structure of ribonucleic acids. *J. Mol. Struct.* 1993, **286**, 203–210

3 Michel, F., and Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* 1990, **216**, 585–610

4 Westhof, E., and Michel, F. Some tertiary motifs on RNA foldings. In: *Structural tools for the Analysis of Protein–Nucleic Acids Complexes* (Lilley, D.M.J., Heumann, H., and Sucks, D., eds.). Birkhäuser-Verlag, Basel, 1992, pp. 255–267

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

6 Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., and Woodson, S.A. RNA folding at millisecond intervals by synchrotron hydroxyl radical footprinting. *Science* 1998, **279**, 1940–1943

7 Westhof, E., Masquida, B., and Jaeger, L. RNA tectonics: towards RNA design. *Folding Design* 1996, **1**, R78-R88

8 Sundaralingam, M. The concept of a conformationally "rigid" nucleotide and its significance in polynucleotide conformational analysis. In: *Conformation of Biological Molecules and Polymers* (Bergmann, V.E.D., and Pullmann, B., eds.), Vol. V. The Israel Academy of Sciences and Humanities, Jerusalem, 1973, pp. 417–456

9 James, B.D., Olsen, G.J., and Pace, N.R. Phylogenetic comparative analysis of RNA secondary structure. *Methods Enzymol.* 1989, **180**, 227–239

10 Woese, C.R., Winkler, S., and Gutell, R.R. Architecture of ribosomal RNA: Constraints of the sequences of "tetra-loops." *Proc. Natl. Acad. Sci. U.S.A.* 1990, **87**, 8467–8471

11 Westhof, E., Romby, P., Romaniuk, P.J., Ebel, J.P., Ehresmann, C., and Ehresmann, B. Computer modeling from solution data of spinach chloroplast and of *Xenopus laevis* somatic and oocyte 5S rRNAs. *J. Mol. Biol.* 1989, **207**, 417–431

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

13 Pley, H.W., Flaherty, K.M., and McKay, D.B. Three-dimensional structure of a hammerhead ribozyme. *Nature (London)* 1994, **372**, 68–74

14 Costa, M., and Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* 1995, **14**, 1276–1285

15 Leontis, N.B., and Westhof, E. The 5S RNA loop E: Chemical probing and phylogenetic data vs. crystal structure *RNA* 1998, **4**, 1134–1153

16 Cheong, C., Varani, G., and Tinoco, I., Jr. Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC *Nature (London)* 1990, **346**, 680–682

17 Jucker, F.M., and Pardi, A. Solution structure of the CUUG hairpin loop: A novel RNA tetraloop motif. *Biochemistry* 1995, **34**, 14416–14427

18 Butcher, S.E., Dieckmann, T., and Feigon, J. Solution structure of the conserved 16S-like ribosomal RNA UGAA tetraloop. *J. Mol. Biol.* 1997, **268**, 348–358

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

20 Biswas, R., Wahl, M.C., Ban, C., and Sundaralingam, M. Crystal structure of an alternating octamer r(GUAU-GUA)dC with adjacent G × U wobble pairs. *J. Mol. Biol.* 1997, **267**, 1149–1156

21 Biswas, R., and Sundaralingam, M. Crystal structure of r(GUGUGUA)dC with tandem G × U/U × G wobble pairs with strand slippage. *J. Mol. Biol.* 1997, **270**, 511–519

22 Correll, C.C., Freeborn, B., Moore, P.B., and Steitz, T.A. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain *Cell* 1997, **91**, 705–712

23 Dallas, A., and Moore, P.B. 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

24 Szewczak, A.A., Moore, P.B., Chan, Y.-L., and Wool, I.G. The conformation of the sarcin/ricin loop from 28S ribosomal RNA *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90**, 9581–9585

25 Szewczak, A.A., and Moore, P.B. The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.* 1995, **247**, 81–98

26 Correll, C.C., Munishkin, A., Chan, Y.L., Ren, Z., and Steitz, T.A. Crystal structure of the ribosomal RNA domain essential for binding elongation factors *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95**, 13436–13441

27 Westhof, E., Dumas, P., and Moras, D. Crystallographic refinement of yeast aspartic acid transfert RNA. *J. Mol. Biol.* 1985, **184**, 119–145

28 Westhof, E., and Sundaralingam, M. Interrelationships between the pseudorotation parameters P and τ_m and the geometry of the furanose ring. *J. Am. Chem. Soc.* 1980, **102**, 1493–1500

29 Massire, C., Jaeger, L., and Westhof, E. Derivation of the three-dimensional architecture of bacterial ribonuclease P RNAs from comparative sequence analysis. *J. Mol. Biol.* 1998, **279**, 773–793

30 Haas, E.S., Banta, A.B., Harris, J.K., Pace, N.R., and Brown, J.W. Structure and evolution of ribonuclease P RNA in gram-positive bacteria. *Nucleic Acids Res.* 1996, **24**, 4775–4782

31 Tanner, M.A., and Cech, T.R. An important RNA tertiary interaction of group I and group II introns is implicated in gram-positive RNase P RNAs. *RNA* 1995, **1**, 349–350

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

33 Major, F., Gautheret, D., and Cedergren, R. Reproducing the three-dimensional structure of a tRNA molecule from structural constraints, *Proc. Natl. Acad. Sci. U.S.A.* 1993, **90**, 9408–9412

34 Earnshaw, D.J., Masquida, B., Müller, S., Sigurdsson,

S.T., Eckstein, F., Westhof, E., and Gait, M.J. Interdomain cross-linking and molecular modelling of the hairpin ribozyme *J. Mol. Biol.* 1997, **274**, 197–212
35 Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. New loop–loop tertiary interactions in self-splicing introns of subgroups IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem. Biol.* 1996, **3**, 993–1009