View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Carolina Digital Repository

NIH Public Access

Author Manuscript

J Math Biol. Author manuscript; available in PMC 2008 January 1.

Published in final edited form as:

J Math Biol. 2008 January ; 56(1-2): 253–278.

RNABC: Forward Kinematics to Reduce All-Atom Steric Clashes in RNA Backbone

Xueyi Wang[1], **Gary Kapral**[2], **Laura Murray**[2], **David Richardson**[2], **Jane Richardson**[2], and **Jack Snoeyink**[1]

1 *Department of Computer Science, UNC Chapel Hill, Chapel Hill, NC, 27599-3175, USA Email: xwang@cs.unc.edu*

2 *Department of Biochemistry, Duke University, Durham, NC, 27710-3711, USA*

Abstract

Although accurate details in RNA structure are of great importance for understanding RNA function, the backbone conformation is difficult to determine, and most existing RNA structures show serious steric clashes (≥ 0.4Å overlap) when hydrogen atoms are taken into account. We have developed a program called RNABC (RNA Backbone Correction) that performs local perturbations to search for alternative conformations that avoid those steric clashes or other local geometry problems. Its input is an all-atom coordinate file for an RNA crystal structure (usually from the MolProbity web service), with problem areas specified. RNABC rebuilds a suite (the unit from sugar to sugar) by anchoring the phosphorus and base positions, which are clearest in crystallographic electron density, and reconstructing the other atoms using forward kinematics. Geometric parameters are constrained within user-specified tolerance of canonical or original values, and torsion angles are constrained to ranges defined through empirical database analyses. Several optimizations reduce the time required to search the many possible conformations. The output results are clustered and presented to the user, who can choose whether to accept one of the alternative conformations.

Two test evaluations show the effectiveness of RNABC, first on the S-motifs from 42 RNA structures, and second on the worst problem suites (clusters of bad clashes, or serious sugar pucker outliers) in 25 unrelated RNA structures. Among the 101 S-motifs, 88 had diagnosed problems, and RNABC produced clash-free conformations with acceptable geometry for 71 of those (about 80%). For the 154 worst problem suites, RNABC proposed alternative conformations for 72. All but 8 of those were judged acceptable after examining electron density (where available) and local conformation. Thus, even for these worst cases, nearly half the time RNABC suggested corrections suitable to initiate further crystallographic refinement. The program is available from http://kinemage.biochem.duke.edu.

Keywords

kinematic chain; RNA backbone conformation; RNA backbone adjustment; RNA crystallography; automated rebuilding; steric clash; S-motifs; all-atom contacts; structure validation

1. Introduction

RNA plays many important roles in organisms, with new ones being discovered constantly [Soukup & Soukup 2004, Nielson *et al*. 2005, Salehi-Ashtiani *et al*. 2006]. RNA stores and transmits genetic information [Crick 1970, Sussman & Kim 1976, Lolle *et al.* 2005], provides and regulates molecular-binding interactions [Huang *et al*. 2003, Lukavsky *et al*. 2003, Mattick 2001], maintains chromosome length [Chen & Greider 2004], controls metabolic processes [Winkler *et al*. 2002, Serganov *et al*. 2006], and catalyzes chemical reactions [Nissen *et al*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2000, Lilley 2005, Kelin & Ferre-D'Amare 2006]. RNA plays a central role in all aspects of gene expression and its control [Claverie 2005], such as performing and regulating RNA interference [Tomari & Zamore 2005], co-suppression and silencing [Mattick 2001], and especially splicing and alternative splicing of exons [Nilsen 1994, Murray & Jarrell 1999, Stahley & Strobel 2005].

The biological function of an RNA molecule depends on the local details of its 3D structure. Those details can be determined at varying resolutions by X-ray crystallography [Jovine *et al*. 2000, Yusupov *et al*. 2001, Correll *et al*. 2003], NMR [Kolk *et al*. 1998, Oberstrass *et al*. 2006] and electron microscopy [Frank 2003]. Although RNA crystallography has seen revolutionary progress recently [Ban *et al*. 2000, Schluenzen *et al*. 2000, Wimberley *et al*. 2000, Batey *et al*. 2004, Torres-Larios *et al*. 2005, Martick & Scott 2006], determining RNA structure remains a difficult task.

Large RNA or RNP (ribonucleoprotein) structures are typically determined at resolutions of 2.5Å or worse; at that level of detail the phosphates and bases can be seen clearly and accurately positioned (see Figure 1a), but the remaining backbone atoms and the sugar puckers are underdetermined, with too many variable parameters per observable feature. All-atom-contact analysis [Word *et al.* 1999a, Davis *et al.* 2004] of deposited RNA structures commonly shows steric clashes between backbone and base atoms or among backbone atoms, as illustrated in Figure 1b. Thus, there is a need for new methodology for backbone fitting.

The reason for these problems with determining RNA backbone conformation can be appreciated by comparing the full atomic detail seen in an electron density map at 1.04Å resolution (Figure 2a) with the same piece of structure in a map at 2.4Å resolution (Figure 2b). In the latter, the P (phosphorus) atom of the PO_4 (phosphate) group is still well located by a strong peak (in purple) but the surrounding O atoms cannot be seen individually; the base planes are still clear but sugar pucker cannot be observed directly; and between sugar and phosphate the density necks down evenly with no indication of the zigzag that determines the backbone dihedral angles.

Base pairing and stacking are the dominant features determining RNA structure and energetics. However, the 3D structure of the RNA backbone is at least equally important in functional interactions such as drug binding [Hansen *et al.* 2003], protein/RNA interactions [Klein *et al.* 2004], aptamer binding [Huang *et al*. 2003], and ribozyme catalysis [Doudna & Cech 2002], which often occurs at sites with unusual backbone conformations [Ferre-D'Amare *et al.* 1998, Adams *et al*. 2004, Golden *et al*. 2005] that require careful and accurate analysis. The partner molecules in all these systems interact with the full all-angle, all-atom detail of the RNA, and our structural biology should aim to accurately determine that same level of detail.

The currently-available tools for fitting, refining, rebuilding, and validating crystal structures for proteins are significantly more rich and mature than those for RNA. For proteins, initial model building (“chain tracing”) can be done automatically by ARP/wARP [Perrakis *et al.* 1999] or Resolve [Terwilliger 2002], but for RNA, such tools do not yet exist. Almost all large RNA and RNP structures are refined in CNS [Brunger *et al.* 1998], which has provided parameter sets and other support for nucleic acids. CNS optimizes agreement of model to data by minimization or simulated annealing protocols, using a simple atomic force field weighted relative to an experimental data term. Energy parameters, weightings, and procedural strategies are not yet fully optimized for RNA: for example, sugar puckers are restrained to the default C3’-endo configuration unless explicitly set by the user, and there are not yet good diagnostics to help make that decision. Model rebuilding between rounds of refinement is traditionally performed by visually comparing the model to the electron density map and manually adjusting

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

it, in software such as O [Jones *et al*. 1991], XFit [McRee 1999], or Coot [Emsley & Cowtan 2004]. This process is especially time-consuming and error-prone for RNA.

Some model evaluation measures work equally well for nucleic acids as for proteins, such as the crystallographic residuals R and R_{free} [Brunger 1992], difference density (F_{obs}–F_{calc}), and all-atom steric clashes [Word *et al.* 1999a, Davis *et al*. 2004]. Other tools that are effective on protein do not yet have equivalently versions for RNA rebuilding, including 2-D Ramachandran plots that compactly assess all available protein backbone dihedral angles [Morris *et al.* 1992, Lovell *et al.* 2003]. Protein backbones have the advantage of only 2 major degrees of freedom per residue (ϕ and ψ), while RNA backbones have at least 6 degrees of freedom per nucleotide (depending on how sugar pucker is represented), meaning that the equivalent plot for RNA would be 6-D or 7-D. Simplifications using 2-D projections of pairs of adjacent dihedral angle values [Sasisekharan and Lakshminarayanan 1969; Murthy *et al*. 1999] have not led to practical tools. Simplification by defining virtual dihedral angles at 2 atoms per residue [Duarte *et al*. 2003] is very valuable for locating structural motifs, largely because it is designed to be insensitive to errors. For that same reason, however, it is not useful for building or correcting the all-atom models needed for refining crystallographic or NMR experimental structures. Recent work has identified clusters of preferred RNA backbone conformations [Murray *et al*. 2003, Schneider *et al*. 2004, Leontis *et al*. 2006], but these cannot be represented as a simple 2-D plot and have not yet been incorporated into rebuilding tools. Most steric clashes in refined protein structures are caused by incorrect positions of sidechain atoms, while most steric clashes in refined RNA structures are caused by incorrect positions of backbone atoms. Amino acid sidechains, which have only one end fixed, are easier to adjust than nucleic acid backbone fragments, which have both ends fixed.

For proteins, recent progress has been made in decision algorithms that can largely replace manual rebuilding in an automated refinement pipeline [Adams *et al*. 2002]. Our work responds to the challenge of developing such an automated rebuilding functionality for RNA backbone structures, where the multidimensional fitting problem makes it especially needed. RNABC (RNA Backbone Correction) produces new alternative conformations with equal or better geometry and fewer steric clashes. It applies the robotics technique of forward kinematics [McCarthy 1990] to recalculate backbone conformation across a dinucleotide, subject to anchored positions of the best-known features: phosphates and base planes. (Forward kinematics is the problem of determining the conformation of a robot or molecule given its parameters, which is considerably easier than the inverse kinematics problem of determining the parameters given the conformation. Forward kinematics makes it easier to sample the entire conformation space.) The user can specify most parameters and procedures, or use default values. RNABC finds and clusters all possible conformations within the specified constraints and outputs those with the best geometry and clash scores. Future work (see section 5) will develop further scoring functions to prioritize the output conformations. RNABC executables for multiple platforms, plus source code, can be downloaded from http://kinemage.biochem.duke.edu/.

After describing the details of the RNABC program, we show results from two extensive tests done on sets of existing RNA structures at widely varying resolutions. One tests typical performance, reproducibility, and success at removing clashes in a set of locally similar S-motif structures. The second tests ability to improve the worst local conformations in a set of completely unrelated RNA structures.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2. Method

We aim to remove steric clashes within an individual suite (from sugar to sugar, as illustrated in Figure 3) by considering the possible configurations of the dinucleotide that contains the suite (also in Figure 3).

There are many parameters needed to specify the conformation of a dinucleotide, so we begin our description by making clear which are obtained from the input, which are specified by the user or from standard values, which are constrained, and which are free to be determined by the program. It is important to realize that parameters cannot be set arbitrarily because of constraints that sugars are closed loops, that backbone remains connected, and that certain atom positions (particularly phosphorus and base planes) are usually defined by clear electron density. We conceptually break the bonds of the sugars, so that what remain are three *backbone segments*, the *main segment* inside the suite and two *supplementary segments* outside. Our method samples the configurations of these segments and considers how they can be joined — it emphasizes early filtering to reduce the number of tested conformations.

2.1. Description of the method

We read PDB-format [Berman *et al.* 2000] files using CCP4 utilities [Krissinel 2004] to parse the coordinates of the RNA backbone. The input file is assumed to include hydrogen atoms, which can be added and optimized conveniently using Reduce [Word *et al*. 1999b] via the structure validation service provided by the MolProbity web site [Davis *et al*. 2004]. MolProbity can also help the user decide which backbone suites need attention by flagging serious clashes between atoms [Word *et al*. 1999a] and suspicious sugar puckers. We hold fixed the positions of the bases (defined by the C1'–N1/9 bond) and the phosphorus atoms, since these are the features of RNA structure seen most clearly in X-ray crystallography, and reconstruct the positions of all other backbone atoms in the dinucleotide. With a few exceptions noted below, we also hold the bond lengths and angles fixed to the canonical values used by CNS [Parkinson *et al*. 1996]. Alternatively, the user can specify the target bond lengths and angles directly (e.g., from parameter files of a different refinement program), or from the input values, or from the average of the input and canonical values. The user can specify sugar puckers explicitly, keep them from the original coordinates, or let the software determine them by geometric rules based on the perpendicular distance from 3' phosphorus to base plane or to the C1'–N1/9 vector. The user can even move the position of a phosphorus or base to a specified new location (e.g., to a local peak in the density); this has been used successfully in other work not presented here.

It is common to describe an RNA backbone conformation by the dihedral angles α-ζ illustrated in Figure 3. Because of our decomposition into segments, we make a different choice of dihedral and bond angles which is mathematically equivalent but is easier to filter for disallowed atom positions. Our method samples all dihedrals α, β, and ζ, phosphate orientations, and two of the three bond angles at C2', C3', and C4' atoms of the sugars. It then determines three bond lengths (C4'–C3', O4'–C1', C2'–C1') and five bond angles (C5'–C4'–C3', C4'–C3'–O3', C4'–O4'–C1', O4'–C1'–C2', and C1'–C2'–C3') so as to satisfy geometry and closure constraints, thus positioning all of the atoms. Note that every atom type (e.g., C4') and every bond length, angle, and dihedral, occurs at least twice within a target dinucleotide. Conditions defined below presume that distances or angles are between nearest atoms of the given type (i.e., within a residue, or within a segment) and hold for all instances, unless otherwise specified.

We use three types of criteria for evaluating the positions of RNA backbone atoms.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

1. NOCLASH: selected atoms should not have steric clashes with the atoms in the suite or the atoms out of the dinucleotide. NOCLASH has two categories:

NOCLASH_M: Atoms O5', C5', C4', C3', O3', O1P, O2P, 1H5', and 2H5' in the main segment should have no steric clashes with the atoms in the suite or out of the dinucleotide.

NOCLASH_S: Atoms O4', C2', O2', H1', 1H2', H3', and H4' in the two sugars should have no steric clashes with the atoms in the suite or out of the dinucleotide.

Atoms within the dinucleotide but out of the suite being adjusted are allowed to clash because local flexibility is not enough to avoid clashes between these and atoms in the suite; clashes related to these atoms may be corrected by running RNABC on adjacent suites.

2. PUCKERTYPE: The two sugar puckers satisfy designated sugar pucker types. For C3'-endo sugar pucker, the perpendicular distance from C3' to plane C4'–O4'–C1' should be longer than the perpendicular distance from C2' to plane C4'–O4'–C1' by a threshold value (default = 0.2Å), and the perpendicular distance from C2' to plane C4'–O4'–C1'should be shorter than a threshold value (default = 0.4Å). Among all the satisfied sugar puckers generated by one sextuple {C5', C4', C3', O3', C1', N1/9}, the best sugar pucker, e.g. C3'-endo, is chosen so that C3' is the farthest from the plane C4'–O4'–C1'. The δ dihedral is also kept within a range compatible with C3'-endo pucker, but quite permissive (51 to 110°). The C2'-endo sugar pucker has similar criteria.

3. INRANGE: distances of atom pairs, angles of certain atom triples and dihedrals of certain atom quadruples that are not pre-specified should be in certain ranges. INRANGE has three categories:

INRANGE_BB: Backbone atoms O5', C5', C4', C3' and O3' in the main and supplementary segments satisfy: the 2-bond to 4-bond distances of O5'–C1', C4'–C1', C5'–C1', C4'–N1/9, C3'–C1', O3'–C1' and C3'–N1/9 and the multi-bond virtual angles of C5'–C4'–C1', C4'–C1'–N1/9, O3'–C3'–C1' and C3'–C1'–N1/9 should be within certain ranges (e.g. within 3 or 4 standard deviations (σ) of the range implied by combining specified values of the intervening parameters; see section 2.2.3), and multi-bond virtual dihedrals C5'–C4'–C1'–N1/9 and O3'–C3'–C1'–N1/9 should be within certain ranges (see section 2.2.3).

INRANGE_ SB: In the sugars on the backbone, bond length C4'–C3' and bond angles C5'–C4'–C3' and C4'–C3'–O3' in each nucleotide should be within the specified ranges.

INRANGE_SS: Where the sugars meet the base, bond lengths O4'–C1' and C2'–C1' and bond angles C4'–O4'–C1', O4'–C1'–C2', C1'–C2'–C3', O4'–C1'–N1/9, and C2'–C1'–N1/9 should be within the specified ranges.

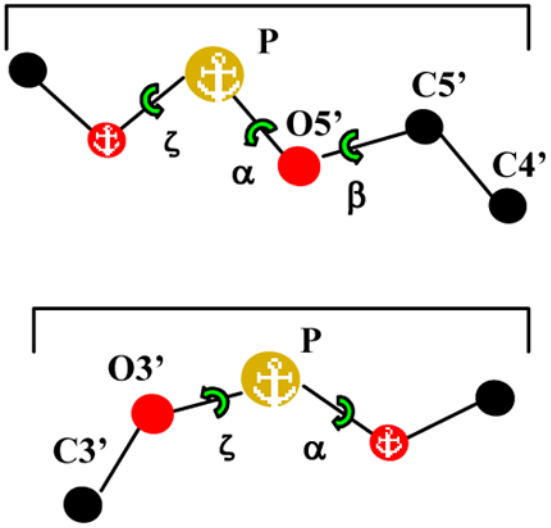
Our method applies these criteria in three steps: building backbone segments, building sugar geometry, and combining/clustering.

2.1.1. Step 1: building backbone segments

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



In the first step, we sample positions of 5 outer atoms in the dinucleotide backbone (O5', C5' & C4' in supplementary segment 1, and O3' & C3' in supplementary segment 2) by changing dihedral angles, and use forward kinematics to calculate allowable positions of these atoms. Given fixed phosphorus positions and the bond lengths and angles, we first calculate allowable positions of those 5 atoms and evaluate them using criterion INRANGE_BB, which relates them to the anchored atoms C1' and N1/9. To calculate the possible positions of atom C4' in supplementary segment 1, for example, with given positions of atoms P, O5' and C5', we rotate C4' around bond O5'–C5' (i.e. rotate dihedral angle β). In the current implementation, we begin with a coarse rotation in steps of 5° (default), followed by a finer rotation with default value of 1° in a ±2° span.

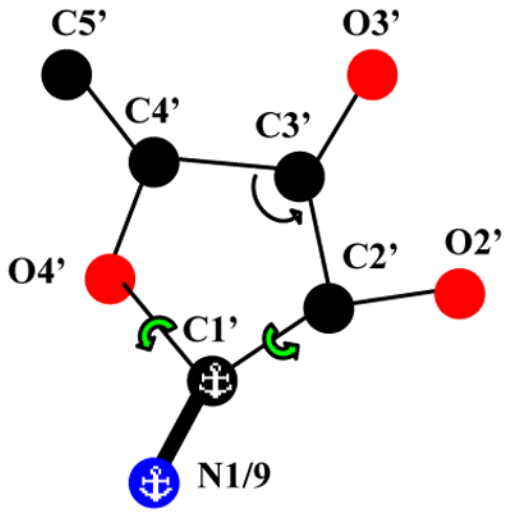
After calculating the allowed positions of atoms in the two supplementary segments, we calculate allowed positions of atoms C3', O3', O5', C5', C4' in the main segment and evaluate them using criteria INRANGE_BB, INRANGE_SB, and NOCLASH_M. The positions of atoms O5' and O3' are calculated from the anchored phosphorus by sampling three Euler angles, which represent the rotation of a 3D object by the angles of rotation around three chosen axes. This ensures that O5' and O3' are sampled from a sphere centered at P with angle O5'–P–O3' fixed. The positions of atoms C5', C4', and C3' are calculated from the positions of O5' and O3' and the relevant bond and dihedral angles.

2.1.2. Step 2: building sugar geometry

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



In the second step, we construct the two sugars in the suite from the coordinates of the two sextuples {C5', C4', C3', O3', C1', N1/9} around them — these are the atoms in the three bonds that join a sugar to the rest of the structure. The first sextuple has C5' and C4' from supplementary segment 1 and C3' and O3' from the main segment. The second sextuple has C5' and C4' from the main segment and C3' and O3' from supplementary segment 2. The positions of atoms C1' and N1/9 are anchored. We generate allowable sextuples by evaluating the combinations of main segment and two supplementary segments using criterion INRANGE_SB. For each allowable sextuple, we calculate the positions of atoms O1P, O2P, 1H5' and 2H5' in the main segment and evaluate them using criterion NOCLASH_M.

For each sextuple, we first calculate positions of O4' and C2' separately by varying the bond angles C5'–C4'–O4', C3'–C4'–O4', O3'–C3'–C2' and C4'–C3'–C2' within 3 standard deviations of canonical values; note that canonical bond angles for the sugars differ slightly between C3' and C2' ring puckers. The positions of O4' and C2' are chosen to satisfy criteria INRANGE_SS and NOCLASH_S.

Next, for each position of C2', we choose a position of O4' so that the sugar pucker constructed satisfies criterion PUCKERTYPE. Finally, we calculate the positions of O2' and hydrogens H1', 1H2', 2HO', H3' and H4' and evaluate them by criterion NOCLASH_S. To calculate O2', we vary bond angles C3'–C2'–O2' and C1'–C2'–O2' in certain ranges. For each sextuple, we output only one sugar that satisfies criteria INRANGE_SS, NOCLASH_S, and PUCKERTYPE.

2.1.3. Step 3: combining/clustering In the third step, after obtaining suite conformations (including two sugars for each suite) that satisfy all the criteria, we cluster similar conformations and output them. The basic idea is that two conformations are considered equivalent if the sum of the absolute differences of the corresponding dihedrals is less than a threshold value. The user can output all inequivalent conformations, or choose that these conformations may be further clustered or sampled.

Summing of the absolute differences of all dihedrals might either cluster dissimilar conformations or generate too many conformations; therefore we split the dihedrals into 5 groups:

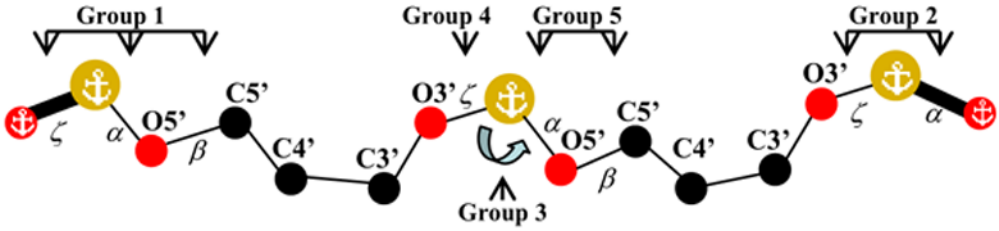
1. angles ζ, α and β for O5', C5' and C4' of supplementary segment 1,
2. angles ζ and α for O3' and C3' of supplementary segment 2,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

3. three Euler angles for positions O3’ and O5’ in the main segment,
4. angle ζ for C3’ in the main segment, and
5. angles α and β for C5’ and C4’ in the main segment.



We assign a threshold value for each group sum: two conformations are considered equivalent if and only if the sums of the angle differences for all five groups are less than the corresponding threshold values. As an approximation (since the effects of different angles are not equivalent) we use 6° for group 4, 11° for the 2-angle groups 2 and 5, and 14° for the 3-angle groups 1 and 3. In practice, we found these threshold values are small enough to generate all distinct conformations and large enough for avoiding similar conformations.

Output conformations are named with a character or number for each of the 7 dihedral angles in the suite, by the system of [Murray *et al*. 2003], where p, t, and m stand for gauche+, trans, and gauche-. For example, A-form helical conformation is named 3’emmtp3’. A hash table is used to assign these names from the specific angle values, while an angle outside of the defined allowable ranges is denoted by “X”. A full 7-angle conformation can be compared to recognized RNA backbone “rotamers” [Murray *et al*. 2003], but that comparison set will be revised when consensus backbone conformers are defined by the RNA Ontology Consortium [Leontis *et al*. 2006].

2.2. Implementation

The program RNABC (RNA Backbone Correction) implements our method in C++, producing a list of names and dihedrals for the proposed alternative conformations, kinemage display files [Richardson & Richardson 2001], and backbone coordinates in PDB format.

2.2.1. User-specifiable Parameters—Currently, each command line invocation of RNABC works on one specified suite. We provide a broad set of parameters, all with defaults but with the option of user specification. For example, sugar-pucker type or method is often set, geometrical parameters specified, or tolerances broadened. Table 1 shows some of the parameters that users can change by flags on the command line. (A fuller listing of flags, syntax & choices is given by typing RNABC -help.)

2.2.2. Output—For each run on a specified RNA suite, RNABC outputs a single text file containing both coordinates and kinemage graphics for zero (if no trials were successful) to 10 (the default maximum) new alternative conformations that satisfy the specified steric clash and covalent geometry conditions. The first half of the file consists of PDB-format coordinates for each output conformation (with its name and dihedral-angle values), while the second half is readable by the Mage and KiNG kinemage viewers [Richardson & Richardson 2001, Davis *et al*. 2004] for 3D display of the original and new conformations. Mage and KiNG can ignore the first half of the file, and do not need it to have a specific extension (e.g., *.kin).

Mage (C) and KiNG (Java), available at http://kinemage.biochem.duke.edu, are open-source software for multi-platform display and modeling of molecules. Both can display RNABC output, along with electron density maps and MolProbity validation kinemages of the original

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

structure. Mage can build a dockable dinucleotide with adjustable backbone rotamers, if further fitting is desired. KiNG reads more map formats, recontours and moves in them in real time, and can be used on-line in the MolProbity service of the above web site, by reading in the RNABC output file and the user's electron density map (or fetching a map from the Electron Density Server at http://eds.bmc.uu.se/eds/ [Kleywegt *et al*. 2004]. When the user has selected a preferred new conformation, the corresponding coordinates can then be cut-and-pasted from the RNABC output file into the PDB file for the overall structure, for submission to further crystallographic refinement.

2.2.3. Early rejection—Although forward kinematics generates each segment conformation quickly, sampling many configurations to find segments that satisfy closure constraints can make this method slow. For example, in the first step, in order to calculate the positions of C4' in the main segment, we need to calculate the positions of O5' and C5' first. The positions of O5' are decided by three Euler angles, and the positions of C5' and C4' are decided by dihedrals α and β. Even with a coarse sampling of angles, every 5°, the total of possible positions for C4' can be $(360/5)^5 > 10^9$. Next we describe the improvements we have made for acceleration.

We chose most of our criteria to be able to reject supplementary segments, main segment and sugar puckers that contain disallowed atom positions as soon as these are calculated. For example, for the supplementary segment P–O3'–C3' in residue 2, after calculating a position of O3', we check the distance from O3' to C1'. If the distance is not within a valid range, we reject O3' and need not calculate C3'.

For the criterion INRANGE_BB, the distances C5'–C1', C4'–N1/9, O3'–C1' and C3'–N1/9 depend on the angles C5'–C4'–C1', C4'–C1'–N1/9, O3'–C3'–C1' and C3'–C1'–N1/9. These angles depend on the pucker state of the sugars and cannot be obtained directly from the other bond lengths and angles. Also we introduce two dihedral angles C5'–C4'–C1'–N1/9 and O3'–C3'–C1'–N1/9, which are used to reject disallowed sugar poses, because the distance and angle criteria allow symmetric sugar poses but the β-D-ribose sugar in RNA has a fixed chirality at the C1' atom. To obtain these angles, we construct sugars from the given bond lengths and angles, then measure the range of these angles before we compute conformations for the dinucleotide. The construction takes a little extra time but we can obtain accurate ranges of these angles and reject configurations with disallowed positions for atoms O5', C5', C4', C3', and O3'.

Early rejection prevents disallowed positions for most backbone atoms. Table 2 shows a typical example, listing the numbers of possible and allowed positions for suite 32 (residue 31 and 32) of tr0002/1EVV using the default coarse and fine rotation angles 5° and 1°, and with ±3 standard deviations for each canonical bond length and angle, In the coarse step, early rejection can reduce the total calculations by a factor 3.2×106, while in the fine step, early rejection can reduce the total calculations by a further factor of 45.

2.2.4. Fast rejections of disallowed angle-dependent positions—There are two ways in which we construct atom positions from angles. In step 2, constructing the sugar for sextuple {C5', C4', C3', O3', C1', N1/9}, we calculate the positions of O4', C2' and O2' in the configuration of Figure 4:

Given an origin O, coordinates of atoms A, B, and ranges for angles ∠AOC and ∠BOC, the possible positions of atom C lie on the sphere with radius ||C|| in a patch bounded by four planes that satisfy $A \cdot C = \|A\| \, \|C\| \cos\angle AOC$, and $B \cdot C = \|B\| \, \|C\| \cos\angle BOC$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Each equation defines a range of planes containing origin *O* with the constraints of ∠*AOC* and ∠*BOC*; the atom *C* lies where their intersection line pierces the sphere and is above the plane of *AOB*.

Similarly, hydrogen atoms H1', 1H2', H3' and H4' form tetrahedral conformations with three heavy atoms, one of which has a four-arc region as in Figure 4.

Given an origin O and coordinates A, B, C, the atom D in tetrahedral conformation lies at distance ||D|| along the vector –(A/||A||+B/||B||+C/||C||); if C is confined to a four-arc region, then so is D.

In RNABC, we sample angles ∠*AOC* and ∠*BOC* by integer multiples of their standard deviations from their original values, because this gives a sufficiently dense set of possible positions of atom C. For example, if the original values of angles ∠*AOC* and ∠*BOC* are 110° and 108°, their standard deviations are 1.5° and 1.2° and we allow ±3 standard deviations, then the values of angle ∠*AOC* are {105.5°, 107°, 108.5°, 110°, 111.5°, 113°, 114.5°}, the values of angle ∠*BOC* are {104.4°, 105.6°, 106.8°, 108°, 109.2°, 110.4°, 111.6°}, and in total we obtain 49 positions for atom C in its *four-arc region*. To rapidly test whether any values of C are allowed by NOCLASH_S, the four-arc region can be approximated by a quadrilateral: for typical bond length 1.5Å, standard deviation = 1.5°, and ±4 standard deviations, the maximum error is < 0.01Å.

During the calculations of the sugar, the positions of O4' and C2' may be rejected by the criteria INRANGE_SS and NOCLASH_S. We do not know if a sugar is allowable until we calculate the positions of some or all of these atoms (including H1' and 2HO'). If all positions of one atom or the combination of positions of several atoms are disallowed, we try the next sextuple {C5', C4', C3', O3', C1', N1/9} until we find an allowable sugar. If we find an allowable sugar, we can cluster the conformations and avoid calculating similar sextuples, but if we fail, we may try to construct the sugar again and again since we don't know which sextuple could construct an allowable sugar beforehand.

The calculations can be accelerated if we can check and reject four corners of the four-arc regions of atoms O4', H4', C2', H3', O2' and 1H2' (see Figure 5) before we calculate and check all the possible positions within these regions. As all the criteria for distances, angles (or cosine values of angles) and steric clashes (measured by the distances between two atoms) are linear, we found it sufficient to test the corners of four-arc regions, eliminating those where all four corners are disallowed by some criterion, or where two criteria combine to eliminate a pair of triangles that cover the quadrilateral (see Figure 6). In Figure 5 only two adjacent corners of the four-arc region are disallowed; our program does not combine this case with others. For the example of section 2.2.3, the total calls to calculate atoms O4', C2' and O2' were 5,532,265 before and 984,696 after this optimization, and the total calls to calculate H1', 1H2', H3' and H4' were 422,720 before and 287,034 after, so we reduce the calls by 82% and 32%. In our experiments, we did not reject any allowable atom positions by these tests.

There are several places in which calculations are repeated, and where we have also optimized by preserving intermediate results. For example, both the first and the second steps use criterion INRANGE_SB. In the first step, a proposed set of atom positions in the main segment is kept only if there exists at least one compatible supplementary segment; the comparison stops when one is found, but we record its position for later use. When sextuples are constructed, main and supplementary segments are paired using the same INRANGE_SB criteria. Starting that search from the previously recorded first match rather than from the beginning saved 35% of the comparisons for the example in section 2.2.3.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

2.3. Running time performance

The program was compiled by Microsoft Visual C++ 6.0 and tested on a Dell machine with 1.8 GHz Centrino processor, 768MB memory and Windows XP operating system. Performance is compared before and after the optimizations discussed in section 2.2.4. (The running time of the program would be extremely slow if not optimized by the early rejection technique in section 2.2.3). To demonstrate the time that RNABC takes on a typical example, we chose suites 52, 75 and 41 of tr0002/1EVV (see Table 3), which exemplify three types of collisions that RNABC can resolve: a) sugar clashes with base, b) backbone clashes with base, and c) sugar/backbone clashes with sugar/backbone. We report running times and number of conformations for three standard deviation ranges (±3, ±4 and ±5σ) for all bond lengths and angles.

As we can see from the data, the optimizations described in section 2.2.4 make the program 2–9 times faster. Larger standard deviation ranges increase the running time because the numbers of sextuples {C5’, C4’, C3’, O3’, C1’, N1/9} increase. Collision case (a) takes more time than cases (b) or (c) because there is a steric clash of 1H2’ in the first residue with the second base, and we cannot know whether the position of 1H2’ is allowed until we have calculated the positions of O4’, C2’ and O2’. Still, we see significant improvement of performance even in case (a).

2.4. Methods for the practical tests

Coordinate files were downloaded either from the NDB (Nucleic acid Data Base [Berman *et al*. 1992]) or the PDB (Protein Data Bank [Berman *et al*. 2000]). In the text, files are described by both the 6-character NDB code and the 4-character PDB code (e.g., rr0082/1S72); here we list them by NDB code, for brevity, giving only the changing final number for codes with the same starting characters. For the S-motif test, files were: pr0015, 205; rr0009,16, 20–23, 28–30, 33, 42–45, 47, 49, 52, 54–61, 67, 71, 76–82; ur0002, 7, 26, 33–35. For the test on 154 non-redundant suites, files were: ar0002, 4, 24, 28; dr0008, 10; pr0005, 11, 18, 26, 32, 67, 73, 81, 85, 90; prv001; rr0005, 10, 16, 19, 33; trna12; ur0012, 19.

Hydrogen atoms were added and optimized by Reduce [Word *et al.* 1999b]. Residue numbers for S-motifs were obtained from the SCOR database [Klosterman *et al.* 2002]. Problem suites were identified in the MolProbity web service [Davis *et al*. 2004] as having suspect sugar puckers or serious all-atom clashes. Bond length and angle deviations were checked within RNABC. We define an all-atom steric clash when the distance of two atoms i and j (including hydrogens; i and j > three bonds apart) is less then $vdw_i + vdw_j - 0.4$Å, where vdw_i is the van der Waals radius for atom i from Probe [Word *et al*. 1999a]. Bad geometry is defined as a bond length or angle > 4 standard deviations away from canonical value [Parkinson *et al*. 1996].

The all-atom contact dots shown in Figures 1, 8, and 9 were calculated by Probe on the MolProbity web service. A 0.25Å radius probe sphere is rolled over the van der Waals surface of each atom, leaving a contact dot only when the probe touches another not-covalently-bonded atom. The dots are colored by the local gap width between the two atoms: blue when near maximum 0.5Å separation, shading to bright green near perfect van der Waals contact (0Å gap). When suitable H-bond donor and acceptor atoms overlap, the dots are shown in pale green, forming lens or pillow shapes. When incompatible atoms interpenetrate, their overlap is emphasized with spikes instead of dots, and with colors ranging from yellow for negligible overlaps to hot pink for serious clash overlaps >0.4Å.

RNABC was run on each problem suite, first with default parameter choices (see Table 1). If RNABC failed to find an allowable output conformation at that level, it was rerun with -PARAMETER7 (trying all 3 geometry references: canonical, original, and the average) and -

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

SIG4 allowable deviations. In the second test, adjacent suites were also run and their results combined, and explicit sugar puckers sometimes specified if needed. If RNABC still produced no output conformations, that example was considered a failure. Table 4 gives sample command lines used at each level of trial and the number of suites in test two that first gave output conformations at each level.

The output conformations (see section 2.2.2) were visualized in KiNG [Davis *et al*. 2004], along with a MolProbity multi-criterion kinemage of the starting structure and $2F_{obs} - F_{calc}$ electron density maps from the EDS server [Kleywegt *et al*. 2004], if structure factors had been deposited. Conformations were discarded if they were very close to the original or if they were clearly a poorer fit to the electron density. For numerical analysis, Excel spreadsheets were populated with data on initial conformations and their indiscretions, RNABC run parameters, and output conformations, including dihedral values and pucker parameters from Dang [Word *et al*. 2000]. For Figures 8–10, the selected output coordinates were edited into the PDB file and a new all-atom contact kinemage produced in MolProbity and displayed in KiNG. Such comparison kinemages were used in the second test to judge the level of improvement over the original structure (e.g. quantitative changes in clashes or hydrogen bonding). Any suggested conformations remaining after all these filtering steps were considered reliable options for improving the structure.

3. Results and discussion

3.1. Removing clashes in many similar S-motif structures

The S-motif (or sarcin-, S-turn-, bulged G-, or loop E-motif) is a distinctive and highly structured internal loop within an A-form RNA double helix, especially common in ribosomal RNAs; an example is shown in Figure 7. It includes several non-canonical base pairs and a base triple, and the backbone forms a pronounced S-shape on the primary strand and a small dent and a stack switch on the secondary strand. The S-motif is named for its occurrence in loop E of the 5S ribosomal RNA and especially in the highly conserved sarcin/ricin loop of the large ribosomal subunit, which binds essential translation factors. Toxins like sarcin, ricin, and restrictocin inactivate ribosomes by cleaving the sarcin loop; the S-motif is at the toxin binding site. Classic S-motifs and variants also occur elsewhere in ribosomal and other RNAs, so there are many similar but not identical examples in the structural database, including a few at very high resolution (e.g., ur0035/1Q9A at 1.04Å resolution [Correll *et al*. 2003] shown in Fig. 2a).

102 S-motifs in 42 crystal structures are listed by the SCOR database of RNA motifs [Klosterman *et al*. 2002]. One S-motif (ur0002/430D a8–a12) has a steric clash between the residue 12 C1’, whose position is held fixed by RNABC, and an out-of-suite N6 on residue 20, and was removed from the test set.

We studied the three distinctive non-A-form suites on the primary strand. The sugar puckers are typically C3’-C2’ for the first suite, C2’-C2’ for the second, and C2’-C3’ for the third. The backbone conformations differ in each suite; they are not easy to fit accurately, so they often show serious steric clashes and sometimes deviant geometry — out of 101 S-motifs, all but 13 contain either steric clashes or bad geometry, as defined in section 2.4 — making this dataset suitable for testing RNABC.

For the above 88 S-motifs, we ran RNABC on the suites containing either steric clashes or bad geometry, specifying clash-free output within ±4 standard deviations of canonical parameters. For example, for the S-motif with primary-strand residues 76–79 in chain 9 of rr0082/1S72 (5S ribosomal RNA) which is shown in Figure 8, residues 76 and 77 contain steric clashes so we ran RNABC on suites 76–77 and 77–78, but not on suite 78–79. Table 5 summarizes the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

results. Although adjusting neighboring suites can help in difficult cases, we have confined ourselves in this test to running only the suites with clashes.

For the 101 original S-motifs, 84 have at least one steric clash, and RNABC proposes at least one clash-free conformation for 71 of those (85%). In the 33 S-motifs with bad geometry, RNABC found conformations with good geometry for 30 of them (91%).

Electron density was available for 30 of the 42 structures (71 of the 101 S-motifs). The output conformations were checked for acceptable fit to the electron density where available (e.g. Figure 10), and two S-motif outputs were rejected at this stage. Combining both criteria, the overall success rate on this first test was 72 good new proposed conformations out of the 88 S-motifs originally having problems (82%). As an example of what can be accomplished, the RNABC refit shown in Figure 8c is very similar to the hand refit in Figure 8b, but took significantly less time and expertise.

3.2. Conformations: improving many dissimilar problem suites

Having shown the consistent usefulness of RNABC in correcting a specific backbone motif, a second test was conducted to determine the program's ability to handle severe local problems in a variety of contexts. A set of 25 diverse structures were chosen from the RNA database of Murray, *et al.* (2003), with representatives ranging from simple duplex RNA to the ribosomal subunits and tRNAs. For each of these structures, we used MolProbity and KiNG to identify suites with especially bad clashes and sugar-pucker outliers. RNABC was run on those suites, as well as suites immediately before and after. If an RNABC run with default parameters failed to yield results, parameters were relaxed in a sequential manner, ensuring that new conformations were found whenever feasible (see section 2.4).

RNABC suggested new conformations for 72 of the 154 suites tested. However, 8 of these new suites were later rejected (see below), 3 due to remaining steric overlaps and/or sugar pucker outliers, 2 because of poor fit to the electron density, and 3 for both of those reasons. Thus, RNABC produced new clash-free conformations and/or better sugar puckers, with satisfactory geometry and density fit, for 64 of the 154 suites tested (42%); 19 of those successes were obtained with default parameters.

Table 6 shows the most common problems identified among the original 72 suites, along with how well RNABC improved them. A given suite may have multiple problems, which are categorized into steric clashes (separated by specific pairs of clashing atoms), pucker outliers, and unfavorable ε dihedral values. Pucker and ε dihedral problems often occur together since distortion of ε is often the result of fitting a ribose into the wrong pucker state. RNABC does best at correcting steric clashes, as these were its central design emphasis. It can usually improve and sometimes correct sugar puckers that are misfit as 3' or 4' when they should be 2', as in the example of Figure 9. The "other" puckers are extreme distortions, which the program finds difficult to improve or correct. Each of the bad ε values was related to a bad sugar pucker; RNABC corrects 5 of them; the 14 ε values that remain unfavorable correspond to 14 sugar puckers that are improved but are not corrected completely. For all but three suites, when RNABC aggravated a problem in one category, it greatly improved the other two categories.

The final filter was to determine for the 10 structures (42 of the 72 suites) that had structure factors available, how well RNABC's proposed new conformations fit into the electron density. Although RNABC currently incorporates no constraints for electron density, the fit improved in almost every case — dramatically for some suites, as depicted in Figure 10. Five suites were exceptions; three conformations already targeted for elimination by other geometric offenses and two new cases were found that lay significantly outside the density compared to the initial

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

structure. Thus, 8 of the 72 outputs were rejected by these post filtering steps, with 89% of the suggested suite conformations deemed acceptable for future refinement. Overall, this test of RNABC on extreme structural deviations had a 42% success rate, with a fairly low rate of false positives.

We close with a look at how many different sets of conformations are output by RNABC, and how different these are from the original structure. In the 235 suites for which RNABC produced output conformations, the output dihedral angles differed from the original by 20° (±3°) RMSD across the 6-dihedral sets, with the extremes ranging from 2° (tiny wiggles) to 100° (large backbone shifts). Often a single dihedral undergoes a relatively large change while the other dihedrals adjust slightly to accommodate; sometimes two dihedrals change 30°–50° (usually α and γ in the long-recognized "crankshaft" motion). Cases in which 3 or more dihedrals change more than 35° were rare. Moreover, 30% of the time RNABC yields two conformations that are different from each other as well (dihedral RMSD > 20°); a further 5% yield 3 or more different conformations. Thus, RNABC is capable of giving the user significantly new and sometimes varied options with which to replace the original local conformation.

4. Conclusion

RNABC is the first piece of software that aims to correct identified local problems in the backbone conformation of RNA structures. Although its abilities will undoubtedly continue to develop, it has here been shown to produce new clash-free conformations with acceptable geometry for a large fraction of RNA suites with local backbone problems. RNABC is freely available on multiple platforms, straightforward to run, executes quickly, and is now suitable for routine crystallographic use.

Although we have performed our tests on correcting errors in completed structures, we believe that the best way to use RNABC is to incorporate it into the process of crystallographic refinement. By improving the geometry of RNA backbone earlier in the process of refinement and rebuilding, one can hope to improve the phases and map clarity at the next iteration, as has been done very successfully for protein backbone and sidechains [Arendall *et al*. 2005].

Sometimes RNABC fails to produce a permissible conformation and there is no guarantee that its output will always include the optimally correct answer. However, it seems highly probable that on-line diagnosis in the MolProbity validation site followed by RNABC calculations and then re-refinement could significantly improve backbone conformation in almost any RNA crystal structure. These changes are often sufficiently large, and in sufficiently critical positions, that they would affect structure/function conclusions about biologically important RNA molecules.

5. Future work

We plan to enhance the RNABC procedure in three major ways. The first addition is to incorporate both real-space measures of electron-density fit and also uni- and multi-dimensional dihedral-angle preferences. These would guide algorithms for small movements of the anchored atoms and provide scores for evaluating, clustering, and pruning the output conformations. The second addition is an empirical study of the patterns of shifted phosphate and base positions caused by know misfittings, in order to suggest efficient small shifts by the RNABC algorithms. The third addition is to follow the initial forward-kinematics step (which ensures thorough coverage of the conformational possibilities) with a step of cyclic coordinate descent [Canutescu & Dunbrack 2003] or of simple minimization of an overall scoring function, to optimize between points in the search grid. This should improve the contrast between acceptable and excellent alternatives. These changes would allow the few critical atom

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

positions and bond angles to be tightly restrained rather than completely fixed, which should greatly improve the ability to discover correct solutions from badly deviant starting conformations.

Although expert crystallographic evaluation will always be the final arbiter for which, if any, of the RNABC output conformations should be adopted, the provision of more extensive scoring information by the program will make that process more user friendly. Finally, both current and enhanced structure-improvement proposals from RNABC will be tested by our own and collaborative re-refinements of RNA crystal structures.

Acknowledgements

This research is supported by NIH grant GM-074127.

References

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Cryst D 2002;58:1948–1954. [PubMed: 12393927]

Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. Nature 2004;430(6995):45–50. [PubMed: 15175762]

Arendall WB III, Tempel W, Richardson JS, Zhou W, Wang S, Davis IW, Liu ZJ, Rose JP, Carson WM, Luo M, Richardson DC, Wang BC. A test of enhancing model accuracy in high-throughput crystallography. J Struct Funct Genomics 2005;6(1):1–11. [PubMed: 15965733]

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4Å resolution. Science 2000;289(5481):905–920. [PubMed: 10937989]

Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature 2004;432(7015):411–415. [PubMed: 15549109]

Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B. The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 1992;63(3):751–759. [PubMed: 1384741]

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28(1):235–242. [PubMed: 10592235]

Brunger AT. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 1992;355:472–475.

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Cryst D 1998;54:905–921. [PubMed: 9757107]

Canutescu AA, Dunbrack RL Jr. Cyclic coordinate descent: A robotics algorithm for protein loop closure. Protein Sci 2003;12(5):963–72. [PubMed: 12717019]

Chen JL, Greider CW. Telomerase RNA structure and function: implications for dyskeratosis congenita. Trends Biochem Sci 2004;29(4):183–192. [PubMed: 15082312]

Claverie JM. Fewer genes, more non-coding RNA. Science 2005;309(5740):1529–1530. [PubMed: 16141064]

Correll CC, Beneken J, Plantinga MJ, Lubbers M, Chan YL. The common and distinctive features of the bulged-G motif based on a 1.04Å resolution RNA structure. Nucleic Acids Res 2003;31(23):6806–6818. [PubMed: 14627814]

Crick F. Central dogma of molecular biology. Nature 1970;227(5258):561–563. [PubMed: 4913914]

Davis IW, Murray LW, Richardson JS, Richardson DC. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res 2004;32:W615–W619. [PubMed: 15215462]

Doudna JA, Cech TR. The chemical repertoire of natural ribozymes. Nature 2002;418(6894):222–228. [PubMed: 12110898]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Duarte CM, Wadley LM, Pyle AM. RNA structure comparison, motif search and discovery using a reduced representation of RNA conformational space. Nucleic Acids Res 2003;31(16):4755–4761. [PubMed: 12907716]

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D 2004;60:2126–2132. [PubMed: 15572765]

Ferre-D'Amare AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. Nature 1998;395(6702):567–74. [PubMed: 9783582]

Frank J. Electron microscopy of functional ribosome complexes. Bipolymers 2003;68(2):223–233.

Golden BL, Kim H, Chase E. Crystal structure of a phage Twort group I ribozyme-product complex. Nature Struct Mol Biol 2005;12(1):82–89. [PubMed: 15580277]

Hansen JL, Moore PB, Steitz TA. Structures of five antibiotics bound at the peptidyl transferase center of the large ribosomal subunit. J Mol Biol 2003;330(5):1061–1075. [PubMed: 12860128]

Huang DB, Vu D, Cassiday LA, Zimmerman JM, Maher LJ III, Ghosh G. Crystal structure of NF-kappaB $(p50)_2$ complexed to a high-affinity RNA aptamer. Proc Natl Acad Sci USA 2003;100(16):9268–9273. [PubMed: 12886018]

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron-density maps and the location of errors in these models. Acta Crystallogr A 1991;47:110–119. [PubMed: 2025413]

Jovine L, Djordjevic S, Rhodes D. The crystal structure of yeast phenylalanine tRNA at 2.0Å resolution: cleavage by Mg(2+) in 15-year-old crystals. J Mol Biol 2000;301(2):401–414. [PubMed: 10926517]

Klein DJ, Ferre-D'Amare AR. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. Science 2006;313(5794):1752–1756. [PubMed: 16990543]

Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. J Mol Biol 2004;340(1):141–177. [PubMed: 15184028]

Klein DJ, Schmeing TM, Moore PB, Steitz TA. The kink-turn: a new RNA secondary structure motif. EMBO J 2001;20(15):4214–4221. [PubMed: 11483524]

Kleywegt GJ, Harris MR, Zou JY, Taylor TC, Wahlby A, Jones TA. The Uppsala Electron-Density Server. Acta Cryst D 2004;60:2240–2249. [PubMed: 15572777]

Klosterman PS, Tamura M, Holbrook SR, Brenner SE. SCOR: a structural classification of RNA database. Nucleic Acids Res 2002;30:392–394. [PubMed: 11752346]

Kolk MH, van der Graaf M, Wijmenga SS, Pleij CW, Heus HA, Hilbers CW. NMR structure of a classical pseudoknot: interplay of single- and double-stranded RNA. Science 1998;280(5362):434–438. [PubMed: 9545221]

Krissinel, E. CCP4 Coordinate Library Project. 2004 p.http://www.ebi.ac.uk/~keb/cldoc/

Leontis NB, Altman RB, Berman HM, Brenner SE, Brown JW, Engelke DR, Harvey SC, Holbrook SR, Jossinet F, Lewis SE, Major F, Mathews DH, Richardson JS, Williamson JR, Westhof E. The RNA Ontology Consortium: an open invitation to the RNA community. RNA 2006;12(4):533–541. [PubMed: 16484377]

Lilley DM. Structure, folding and mechanisms of ribozymes. Curr Opin Struct Biol 2005;15(3):313–323. [PubMed: 15919196]

Lolle SJ, Victor JL, Young JM, Pruitt RE. Genome-wide non-Mendelian inheritance of extra-genomic information in Arabidopsis. Nature 2005;434(7032):505–509. [PubMed: 15785770]

Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: φ, ψ and Cβ deviation. Proteins: Structure, Function and Genetics 2003;50:437–450.

Lukavsky PJ, Kim I, Otto GA, Puglisi JD. Structure of HCV IRES domain II determined by NMR. Nature Struct Biol 2003;10(12):1033–1038. [PubMed: 14578934]

Martick M, Scott WG. Tertiary contacts distant from the active site prime a ribozyme for catalysis. Cell 2006;126(2):309–320. [PubMed: 16859740]

Mattick JS. Non-coding RNAs: the architects of eukaryotic complexity. EMBO Rep 2001;2:986–991. [PubMed: 11713189]

McCarthy, JM. Introduction to theoretical kinematics. MIT Press; Cambridge, MA: 1990.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

McRee DE. XtalView/Xfit - A Versatile Program for Manipulating Atomic Coordinates and Electron Density. J Struct Biol 1999;125:156–165. [PubMed: 10222271]

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins 1992;12:345–364. [PubMed: 1579569]

Murray HL, Jarrell KA. Flipping the switch to an active spliceosome. Cell 1999;96:599–602. [PubMed: 10089873]

Murray LJ, Arendall WB III, Richardson DC, Richardson JS. RNA backbone is rotameric. PNAS 2003;100:13904–13909. [PubMed: 14612579]

Murthy VL, Srinivasan R, Draper DE, Rose GD. A Complete Conformational Map for RNA. J Mol Biol 1999;291(2):313–327. [PubMed: 10438623]

Nielson H, Westhof E, Johansen S. An mRNA is capped by a 2', 5' lariat catalyzed by a Group I-like ribozyme. Science 2005;309(5740):1584–1587. [PubMed: 16141078]

Nilsen TW. RNA-RNA interactions in the spliceosome: unraveling the ties that bind. Cell 1994;78:1–4. [PubMed: 7518355]

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. Science 2000;289(5481):920–930. [PubMed: 10937990]

Oberstrass FC, Lee A, Stefl R, Janis M, Chanfreau G, Allain FH. Shape-specific recognition in the structure of the Vts1p SAM domain with RNA. Nat Struct Mol Biol 2006;13(2):160–167. [PubMed: 16429156]

Parkinson G, Vojtechovsky J, Clowney L, Brünger AT, Berman HM. New parameters for the refinement of nucleic acid containing structures. Acta Crystallogr D Biol Crystallogr 1996;52:57–64. [PubMed: 15299726]

Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nature Struct Biol 1999;6(5):458–463. [PubMed: 10331874]

Richardson, JS.; Richardson, DC. MAGE, PROBE, and Kinemages, chapter 25.2.8. In: Rossmann, MG.; Arnold, E., editors. International Tables for Crystallography. F. Kluwer Academic Publishers; Dordrecht, the Netherlands: 2001. p. 727-730.

Salehi-Ashtiani K, Luptak A, Litovchick A, Szostak JW. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. Science 2006;313(5794):1788–1792. [PubMed: 16990549]

Sasisekharan V, Lakshminarayanan AV. Stereochemistry of Nucleic Acids and Polynucleotides. VI Minimum Energy Conformations of Dimethyl Phosphate Biopolymers 1969;8:505–514.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. Cell 2000;102(5):615–623. [PubMed: 11007480]

Schneider B, Moravek Z, Berman HM. RNA conformational classes. Nucleic Acids Res 2004;32(5): 1666–1677. [PubMed: 15016910]

Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature 2006;441(7097):1167–1171. [PubMed: 16728979]

Soukup JK, Soukup GA. Riboswitches exert genetic control through metabolite-induced conformational change. Curr Opin Struct Biol 2004;14:344–349. [PubMed: 15193315]

Stahley MR, Strobel SA. Structural Evidence for a Two-Metal-Ion Mechanism of Group I Intron Splicing. Science 2005;309(5740):1587–1590. [PubMed: 16141079]

Sussman JL, Kim S. Three-dimensional structure of a transfer RNA in two crystal forms. Science 1976;192(4242):853–858. [PubMed: 775636]

Terwilliger TC. Automated structure solution, density modification and model building. Acta Cryst D 2002;58:1937–1940. [PubMed: 12393925]

Tomari Y, Zamore PD. Perspective: machines for RNAi. Genes Dev 2005;19(5):517–529. [PubMed: 15741316]

Torres-Larios A, Swinger KK, Krasilnikov AS, Pan T, Mondragon A. Crystal structure of the RNA component of bacterial ribonuclease P. Nature 2005;437(7058):584–587. [PubMed: 16113684]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Wimberly BT, Brodersen DE, Clemons WM Jr, Morgan-Warren RJ, Carter AP, Vonrhein C, Hartsch T, Ramakrishnan V. Structure of the 30S ribosomal subunit. Nature 2000;407(6802):327–339. [PubMed: 11014182]

Winkler W, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 2002;419(6910):952–956. [PubMed: 12410317]

Word, JM. All-atom small-probe contact surface analysis: An information-rich description of molecular goodness-of-fit. Ph.D thesis, Duke University; Durham, NC: 2000.

Word JM, Lovell SC, LaBean TH, Taylor HC, Zalis ME, Presley BK, Richardson JS, Richardson DC. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. J Mol Biol 1999a;285(4):1711–1733. [PubMed: 9917407]

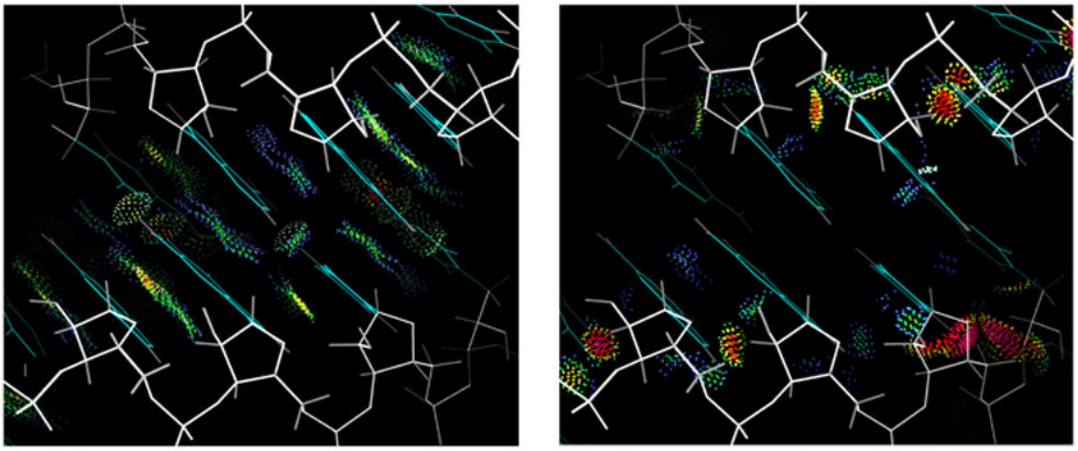
Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 1999b;285(4):1735–1947. [PubMed: 9917408]

Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF. Crystal structure of the ribosome at 5.5Å resolution. Science 2001;292(5518):883–896. [PubMed: 11283358]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) all-atom-contact dots within bases.

(b) all-atom-contact dots within backbone and between backbone and bases.

Figure 1.

Selected all-atom-contacts in tr0002/1EVV (yeast phenylalanine tRNA [Jovine *et al*. 2000]) at 2.0Å resolution (residues 28–32 and 40–44). The green and blue all-atom-contact dots in (a) show almost perfect van der Waals and H-bond contacts between the stacked and paired bases, while the red spikes in (b) show large steric clashes that indicate a locally misfit backbone.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

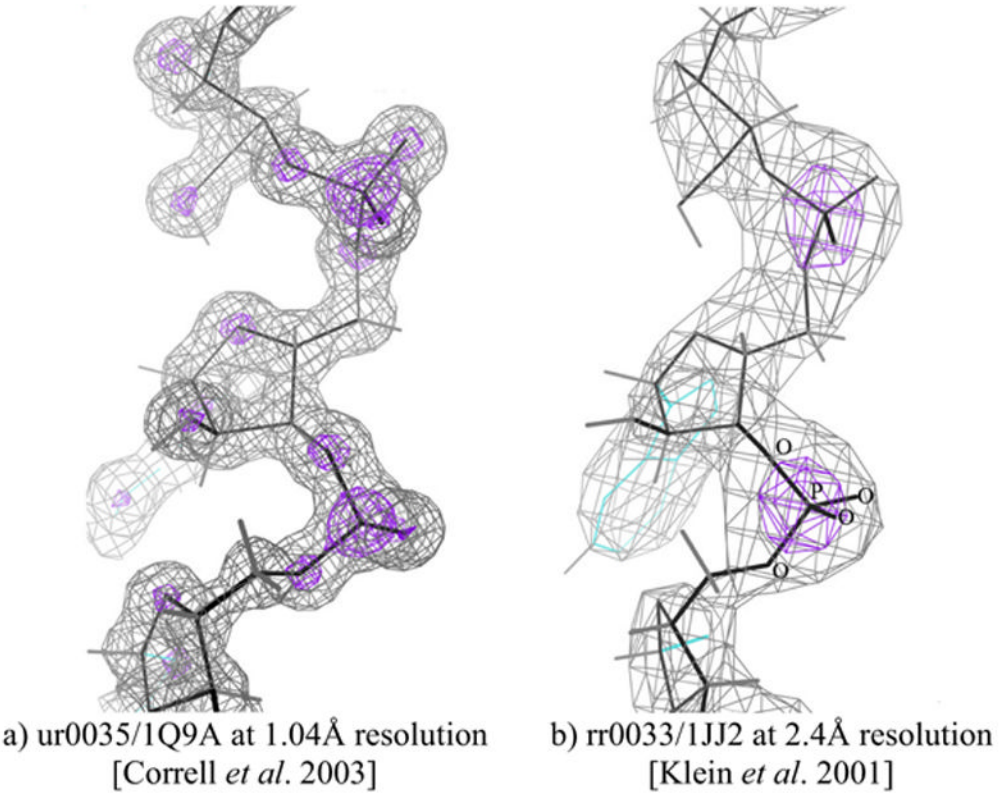


Figure 2.
Contoured electron density maps and atomic models for the same piece of ribosomal RNA structure (part of the "sarcin loop") solved at quite different resolutions.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

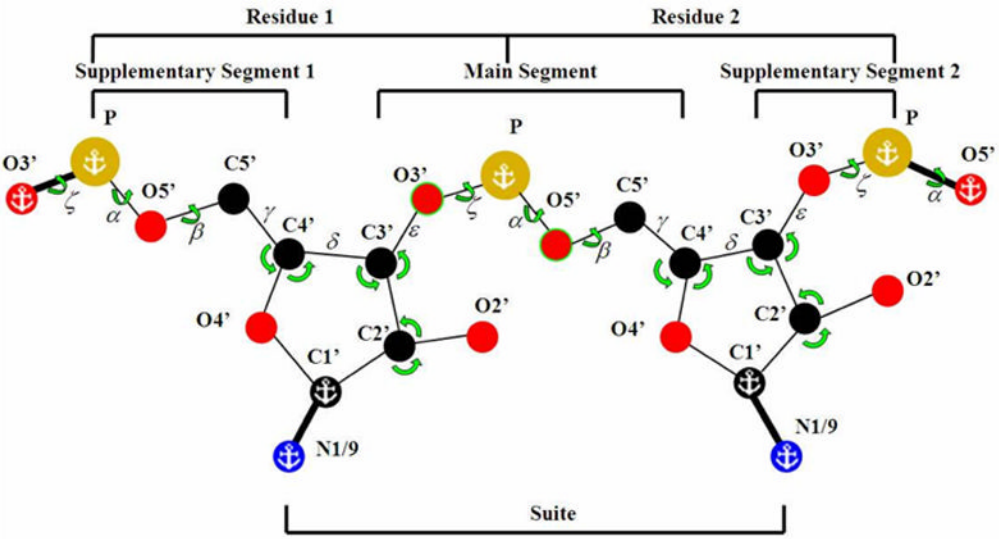


Figure 3.

Atom labeling and nomenclature for reconstructing a suite within a dinucleotide span. Anchors mark atoms with fixed positions; green arrows mark the conformational degrees of freedom that are explored directly: dihedrals α, β, and ζ, PO_4 orientation around the anchored P, and two of the three bond angles around C2', C3', and C4'. Hydrogens are not shown but are used extensively in RNABC.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

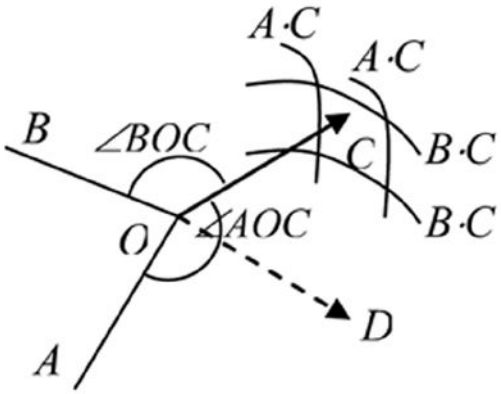


Figure 4.
The 4-arc region of C

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

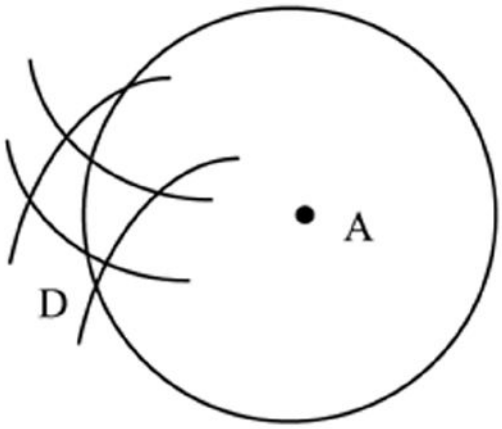


Figure 5.
Atom A intersects with part of the allowable region of D

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

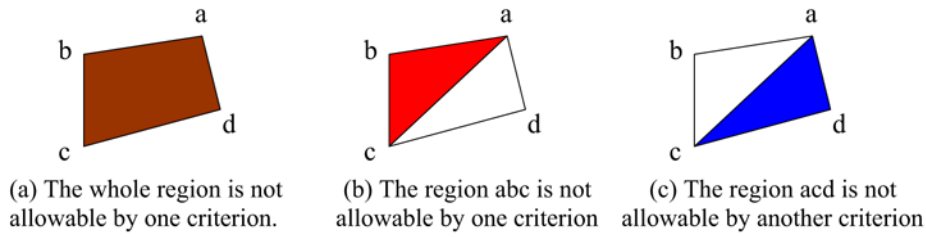


Figure 6.
In (a), the whole region is eliminated by one criterion, while in (b) and (c), each criterion eliminates half of the region, and their combination eliminates the whole.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

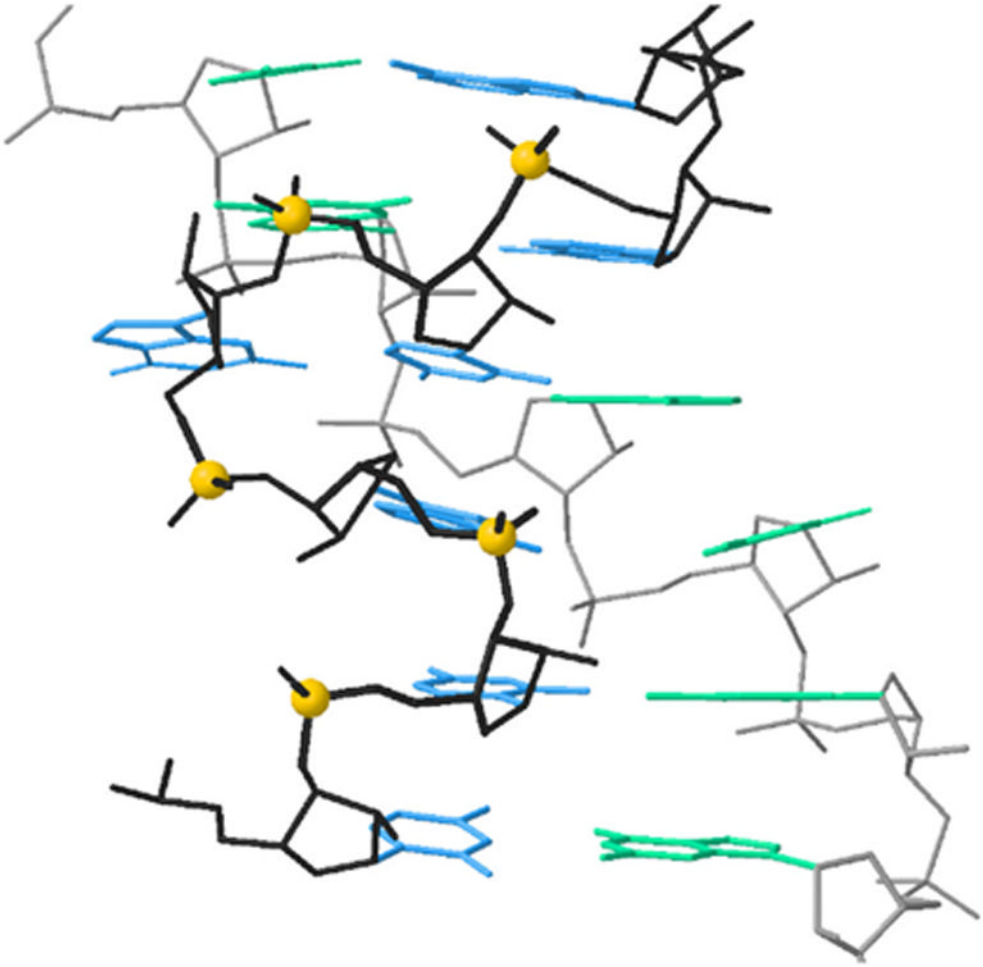


Figure 7.
S-motif 587–589 in rr0082/1S72; primary strand (front) has black backbone and blue bases. Gold P-atom balls mark the 3-suite, "S"-shaped region studied, but this example was clash-free and thus refit was unnecessary.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

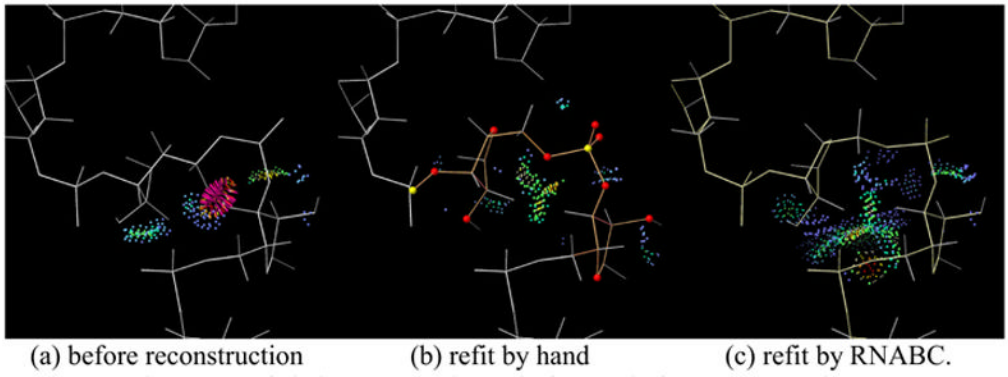
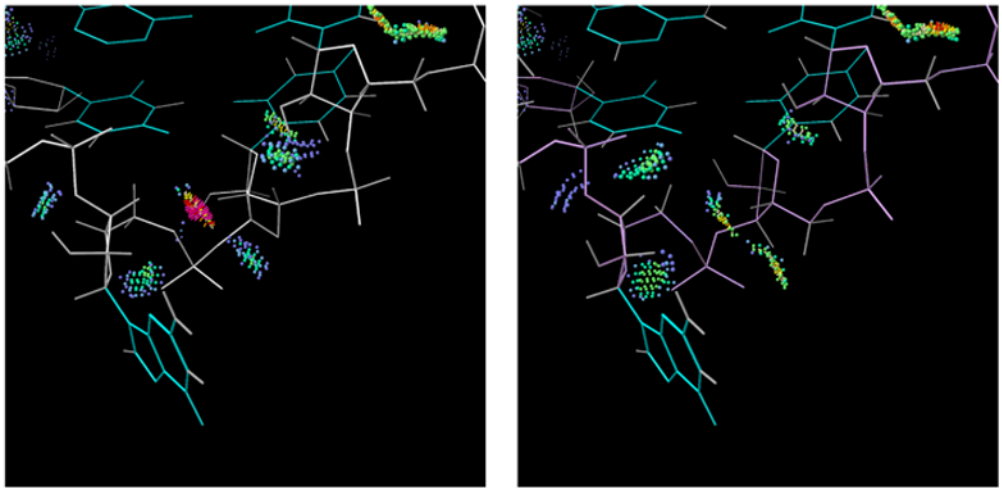


Figure 8.
Suite 76–77 of chain 9, rr0082/1S72 before and after reconstruction.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



(a) before refit: clash of 1H5' with O2', and C4' puckers for both sugars

(b) RNABC refit, in which puckers are improved (C2'-endo) and the clash has been removed

Figure 9.
pr0032/1FFY suite 33–34 before and after refit by RNABC.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

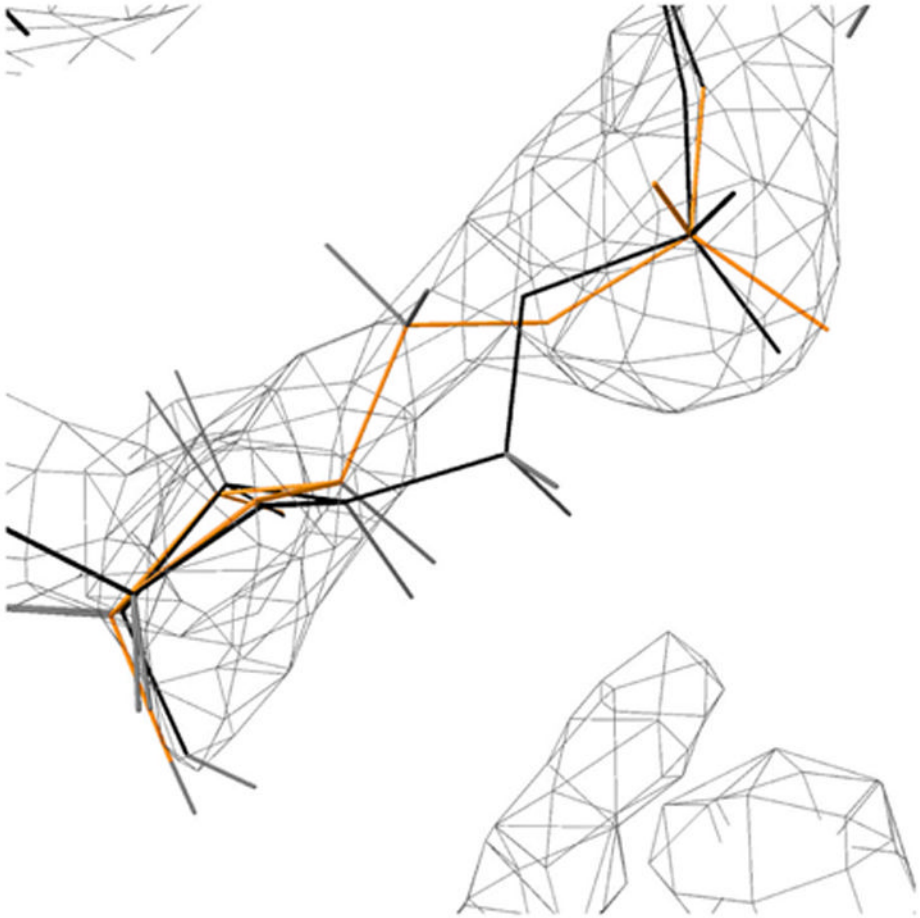


Figure 10.
rr0082/1S72 suite 1941–1942 refit . The original is in black, and the refit in orange; RNABC's conformation, chosen to avoid bad geometry and clashes, also fits the density better.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Parameters often specified by RNABC users

Flag	Parameter details
-RESNUM	Residue number of central P atom in suite to be analyzed.
-CHAIN	Chain ID character, default = first chain in file.
-PUCKER	Pucker type or method for first [second] sugar in suite, default = both determined by 3'P perpendicular to C1'–N1/9 vector.
-COARSESTEP	The step size for coarse rotation angles, default = 5°.
-FINESTEP	The step size for fine rotation angles, default = 1°.
-SIG	The allowable standard deviation of key bond lengths and angles, default = 3σ; 4σ sometimes useful.
-PARAMETER	Specifies reference bond lengths and angles. Users can choose canonical, original, average of canonical and original, or specify values in a file. Default = canonical.
-CLASHLEVEL	The overlap distance considered a steric clash, default = 0.4Å; 0.5Å is a more permissive option.
-WITHINCHAIN	Check collisions only with atoms on the local chain.
-CONFORM	The maximum number of conformations to be output (default = 10), or specifies method to choose output conformations by allowed angle ranges or by distance to preferred backbone conformers.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Comparison of total and allowed positions of backbone atoms found for suite 32 of tr0002/1EVV

		Coarse step (every 5°) Total positions	Coarse step (every 5°) Allowed positions	Ratio (total/allow)	Fine step (every 1°) Total positions*	Fine step (every 1°) Allowed positions (tot/allow)	Ratio
Supplementary segment 1	O5’	72	11	7	55	49	1.1
	C5’	5,200	59	88	1,475	690	2.1
	C4’	370,000	98	3.800	12,250	2,495	4.9
Supplementary segment 2	O3’	72	4	18	20	17	1.2
	C3’	5,200	10	518	250	123	2.0
Main segment†	O5’	370,000	44	8,500	5,500	689	8.0
	C5’	27,000,000	116	230,000	72,500	3,137	23.1
	C4’	1,900,000,000	159	12,000,000	496,875	4,012	123.8
	O3’	370,000	44	8,500	5,500	689	8.0
	C3’	27,000,000	75	360,000	46,875	2,330	20.1
Total		2,000,000,000	620	3,200,000	641,300	14,231	45.0

*The total positions in the fine step are derived from the allowed positions in the coarse step. In the fine step, each angle in a position allowed by the coarse step is sampled within ±2°, giving five refined positions for each. E.g., in supplementary segment 2, we sample the angles of O3’ and C3’ within ±2°, giving 25×10 = 250 total refined positions for C3’.

†The main-segment O3’ and O5’ are obtained by 3 Euler rotation angles around P, so the total number of positions of O3’ and O5’ could be $(360/5)^3 = 3.7\times105$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Running times (in seconds) for three types of collision in tr0002/1EVV.

	# std devs	# of sextuples	Before optimization (s)	After optimization (s)
a) sugar clashes with base	±3	63,000	19	5
	±4	430,000	243	25
	±5	5,600,000	1,435	339
b) backbone clashes with base	±3*	60,000	5	3
	±4	500,000	40	12
	±5	3,000,000	318	64
c) sugar/backbone clashes with sugar/backbone	±3*	6,500	2	2
	±4	140,000	15	9
	±5	880,000	140	38

*No allowable conformations found.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 4

Command lines at successive trial levels for test two. (Note that pucker parameter can be -PUCKER3-3, 3-2, 2-2, or 2–3.)

Sample Command	New cases output
RNABC -CHAIN[x]-RESNUM[n] [input.pdb] > [outputfile]	21
RNABC -CHAIN[x]-RESNUM[n]-PARAMETER7 [input.pdb] > [outputfile]	15
RNABC -CHAIN[x]-RESNUM[n]-PARAMETER7 -SIG4 [input.pdb] > [outputfile]	21
RNABC -CHAIN[x]-RESNUM[n]-PARAMETER7 -SIG4 -PUCKER2–3 [input.pdb] > [outputfile]	15

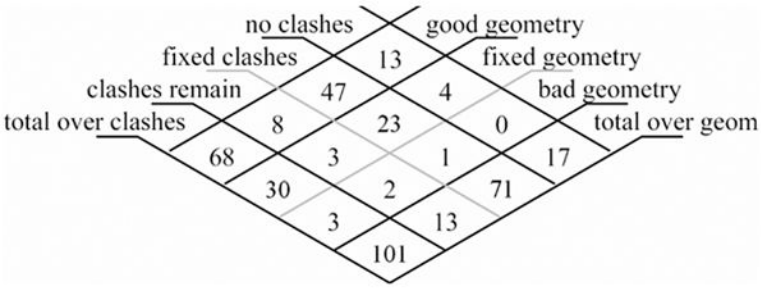
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 5

Performance on removing steric clashes and bad geometry for the 101 S-motifs



NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 6

Corrections: Instances of three categories of problems in the original structures for 72 suites, and how many were fixed, improved, unchanged, or worsened by RNABC. Configurations are deemed unchanged unless there is a difference of either 5 clash spikes, 10° δ dihedral, 0.5Å perpendicular-line length, or 40° ε dihedral. Note that the total number of clashes is greater than 72 — many suites contained several clashes.

Common problems	# of instances	# fixed completely	# improved	# unchanged	# worse	% fixed	% fixed or improved
Steric Clashes							
1H5'–O2'	29	17	6	3	3	59	79
2HO'–P	23	13	5	4	1	57	78
C5' or H5'–C2' or H2'	19	11	4	3	1	58	79
1H2'–O4'	16	10	2	2	2	63	75
Others	80	45	17	7	11	56	78
Pucker outliers							
C4'	12	2	8	2	0	17	83
C3' → C2'	11	2	7	1	1	18	82
Others	11	1	0	4	6	9	9
Unfavorable ε dihedrals (−45° to +155°)							
Bad ε	19	5	0	14	2	26	26