CHAPTER SEVEN

Computational Methods for RNA Structure Validation and Improvement

Swati Jain[*,†,‡], **David C. Richardson**[†,1], **Jane S. Richardson**[†]

[*] Program in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina, USA
[†] Department of Biochemistry, Duke University Medical Center, Durham, North Carolina, USA
[‡] Department of Computer Science, Duke University, Durham, North Carolina, USA
[1] Corresponding author: e-mail address: dcrjsr@kinemage.biochem.duke.edu

Contents

1. Introduction 182
2. General Validation Criteria 186
   2.1 Data validation 186
   2.2 Model-to-data match 186
   2.3 Model validation 188
3. Criteria Specific to RNA Conformation 189
   3.1 RNA bases 189
   3.2 Ribose pucker 191
   3.3 Backbone conformers 193
4. How to Interpret Validation Results 195
   4.1 As an end user 195
   4.2 As a journal referee 198
   4.3 As a structural biologist 200
5. Correcting the Problems 201
   5.1 Manual rebuilding 202
   5.2 RNABC 203
   5.3 Coot and RCrane 203
   5.4 PHENIX refinement with pucker-specific targets 204
   5.5 ERRASER, with examples 204
6. What's Coming Next 207
   6.1 ERRASER for RNA/protein 207
   6.2 Better ion identification 208
   6.3 Putting conformers into initial fitting 208
Acknowledgments 208
References 209

Methods in Enzymology, Volume 558
ISSN 0076-6879
http://dx.doi.org/10.1016/bs.mie.2015.01.007

Abstract

With increasing recognition of the roles RNA molecules and RNA/protein complexes play in an unexpected variety of biological processes, understanding of RNA structure–function relationships is of high current importance. To make clean biological interpretations from three-dimensional structures, it is imperative to have high-quality, accurate RNA crystal structures available, and the community has thoroughly embraced that goal. However, due to the many degrees of freedom inherent in RNA structure (especially for the backbone), it is a significant challenge to succeed in building accurate experimental models for RNA structures. This chapter describes the tools and techniques our research group and our collaborators have developed over the years to help RNA structural biologists both evaluate and achieve better accuracy. Expert analysis of large, high-resolution, quality-conscious RNA datasets provides the fundamental information that enables automated methods for robust and efficient error diagnosis in validating RNA structures at all resolutions. The even more crucial goal of correcting the diagnosed outliers has steadily developed toward highly effective, computationally based techniques. Automation enables solving complex issues in large RNA structures, but cannot circumvent the need for thoughtful examination of local details, and so we also provide some guidance for interpreting and acting on the results of current structure validation for RNA.

1. INTRODUCTION

As any reader of this volume appreciates, the discovery of new RNA biology has exploded in recent decades. Determination of 3D structures for large RNAs and ribonucleoprotein complexes (RNPs) has surged as well to provide the detailed molecular descriptions essential for mechanistic understanding of the diverse biological functions. The primary techniques used for large RNA structures are electron microscopy and X-ray crystallography. The developments described in this chapter have so far been applied almost exclusively to crystallography but should become useful for electron microscopy as it increasingly attains higher resolutions.

The methodology for RNA crystallography has been improving rapidly. However, the tools still lag far behind those for protein crystallography and the problems are inherently more difficult, both in crystallization and phasing and also in accurate model building and refinement. Base pairing can be predicted with quite good, if far from perfect, accuracy, and the expected double helices and their base pairs can be located with good reliability in electron density, even at the 2.5–4 Å resolution typical for attainable crystal structures of large RNAs and RNPs. Most phosphates are also identifiable as peaks of strong, nearly spherical density (see Fig. 1). In contrast, it is very difficult to fit an accurate atomic model for the sugar–phosphate backbone into low-resolution electron density (Murray, Arendall, Richardson, & Richardson, 2003). A nucleotide residue has six torsion-angle parameters (α–ζ), each of which frequently takes on a value widely different from that in A-form. At best, low-resolution density shows a blob for the ribose, with no shape to indicate the ring pucker, and a round tube for the backbone sections between ribose and phosphate, with no way to see which direction the atoms actually zigzag (Fig. 1). The result of this anisotropy in structural clarity is that RNA bases are usually well positioned in deposited crystal structures (e.g., Fig. 2A), while even structures done by expert and careful practitioners often have high rates of physically impossible backbone conformations (e.g., Fig. 2B).

Many important things can be learned from the overall 3D structures of large RNAs, but the details of the backbone conformation also matter, more than one might think, for mechanistic understanding of biological function. Backbone atoms are involved in most types of RNA catalysis (Ferré-D'Amaré & Scott, 2010; Lang, Erlacher, Wilson, Micura, & Polacek, 2008; Zaher, Shaw, Strobel, & Green, 2011). Particular non-A-form backbone conformations are critical to specificity for a wide variety of protein, ligand, aptamer, and drug interactions of RNAs (Klein, Schmeing, Moore, & Steitz, 2001; Warner et al., 2014), as for instance seen in Fig. 3, where one entire side of the specific drug binding is to non-A-form

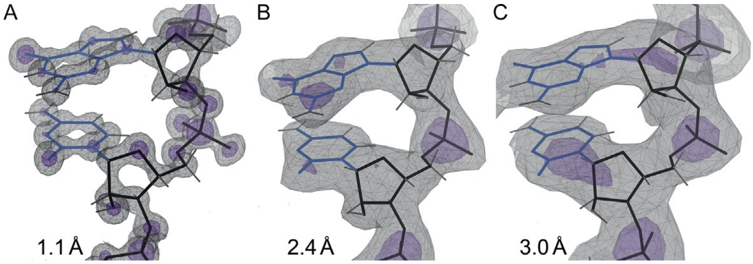


Figure 1 Resolution comparison: three examples, each a backbone conformer **1a** suite in regular A-form double helix. (A) From PDB 2Q1O at 1.1 Å (Li et al., 2007); (B) 3CC2 at 2.4 Å (Blaha, Gural, Schroeder, Moore, & Steitz, 2008), and (C) 3R8T at 3.0 Å (Dunkle et al., 2011). The $2F_o$-F_c density is shown at 1.8σ as pale gray mesh and at 5.3σ as purple (dark gray in the print version) mesh. Note that the phosphate density is high and round in all cases. Base type is unambiguous at 1.1 Å; the G is still identifiable in this case at 2.4 Å, but only purine-versus-pyrimidine at 3.0 Å. At 2.4 Å, there is a bump for the 2′O, but not at 3.0 Å. The zigzag of backbone atoms between P and ribose is clear at 1.1 Å, but that backbone density is just a round tube in (B) and (C).

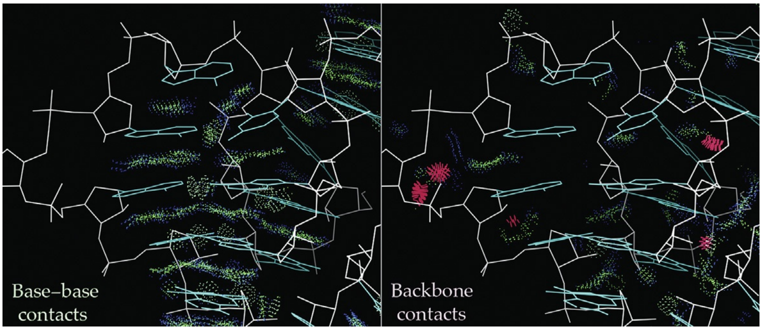


Figure 2 Base versus backbone accuracy as shown by all-atom steric contacts. (A) Well-fit base–base contacts, shown by favorable H-bond and van der Waals contact dots (green and blue). (B) Fairly frequent backbone problems shown by spike clusters of steric clash overlap ≥0.4 Å (hotpink). *From PDB 2HOJ at 2.5 Å (Edwards & Ferré-D'Amaré, 2006).*

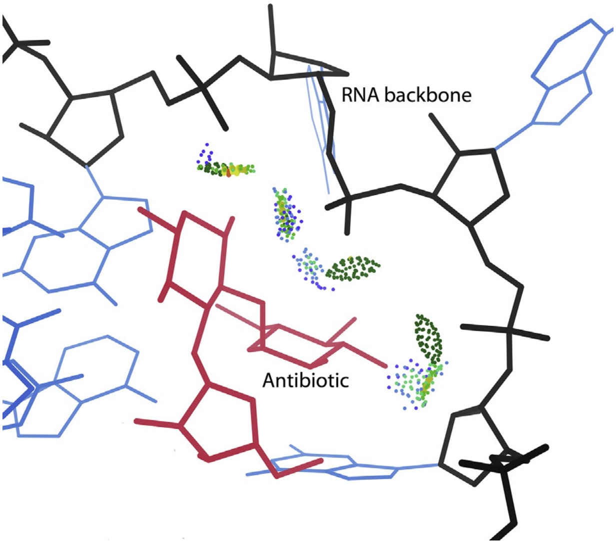


Figure 3 Specific ribostamycin binding by non-A-form RNA backbone. All-atom contacts are shown only between drug and RNA backbone. *From PDB 3C3Z at 1.5 Å (Freisz, Lang, Micura, Dumas, & Ennifar, 2008).*

backbone. RNA–RNA backbone interactions stabilize distant binding sites (Batey, Gilbert, & Montange, 2004) or in general enable the conformational changes that accompany RNA–protein recognition (Williamson, 2000). Other molecules feel the RNA backbone in its full-atomic detail, even if our experimental techniques cannot always see that detail. Therefore, it is important both to pursue better accuracy where feasible and to provide validation measures that allow end users of RNA structures to easily judge reliability of the local features most important to their work.

Fortunately, the addition of explicit hydrogen atoms and calculation of all-atom contacts (Word, Lovell, LaBean, et al., 1999; Word, Lovell, Richardson, & Richardson, 1999) provide sensitive local validation such as the markup shown in Fig. 2, positive for H-bonds and favorable van der Waals and negative for unfavorable steric clashes. Even more importantly, physical factors are now known to limit the accessible combinations of RNA backbone torsions and ribose puckers (Richardson et al., 2008), enabling inferences about detailed backbone conformation to be calculated from the better-observed base and phosphate positions. Complete versions of these tools and their results are available in user-friendly form both online at the MolProbity site (Chen et al., 2010; http://molprobity.biochem.duke.edu) and in the PHENIX crystallographic software system (Adams et al., 2010; http://phenix-online.org). Validation summaries are now available for each X-ray entry in the PDB (Protein Data Bank; Berman et al., 2000), including RNAs, at the RCSB, PDBe, and PDBj sites of the wwPDB (Berman, Henrick, & Nakamura, 2003; Gore, Velankar, & Kleywegt, 2012). Pucker diagnosis and pucker-specific target parameters aid RNA refinement in PHENIX. Highly effective correction of RNA backbone problems is now feasible with the new enumerative real-space refinement assisted by electron density under Rosetta (ERRASER) program (Adams et al., 2013; Chou, Sripakdeevong, Dibrov, Hermann, & Das, 2012), which uses Rosetta's RNA tools (Das & Baker, 2007) and density match (DiMaio, Tyka, Baker, Chiu, & Baker, 2009), MolProbity diagnosis, and crystallographic refinement to do exhaustive local conformational sampling.

A DNA side note: DNA structure can be validated by traditional model-to-data match and covalent geometry, and by all-atom contacts, but the RNA conformational tools are unfortunately not valid for DNA because it is more locally flexible, with broader torsion-angle preferences and more sugar-pucker states accessible (Svozil, Kalina, Omelka, & Schneider, 2008).

This chapter describes the basis and properties of currently available and forthcoming computational tools and visualizations that help structural biologists improve the accuracy of new experimental models for large RNAs. We also discuss how any user can readily interpret the validation results to judge the overall accuracy, or most importantly the local reliability, of those RNA crystal structures.

2. GENERAL VALIDATION CRITERIA

Crystallographic validation includes both global (whole structure) and local (residue level) criteria, and it has three logically distinct components: for the experimental data, for the 3D model, and for the model-to-data match.

2.1 Data validation

Validation of the X-ray data (reviewed in Read et al., 2011) is of concern to expert crystallographer reviewers and of course to the experimenters themselves, since it can determine whether or not a structure solution is possible at all. However, there are several aspects that are understandable and of relevance to an end user. The most obvious is the resolution, the higher the better (lower absolute number), which is the most important global, single-number criterion of structure quality. Better than 1.5 Å is called atomic resolution, since most nonhydrogen atoms are separately visualized in the electron density, and interatomic distances and conformational parameters are determined with reliably high accuracy. The caveat is that in disordered regions with multiple conformations or high B-factors, accuracy can become very poor at any resolution. Historically, and for proteins, a majority of macromolecular crystal structures are in the quite dependable, workhorse 1.5–2.5 Å resolution range. At 2.5–3 Å, the overall chain trace is nearly always correct, although some atomic groups will be misplaced into the wrong piece of density or the wrong orientation. Beyond 3 Å resolution, local details are not reliable, and the incidence of local sequence misalignment increases. Effective resolution is better than the nominal value if there is a lot of noncrystallographic symmetry (such as for viruses) and worse than nominal if the data are less than 85–90% complete in the highest resolution shell or especially if the crystal is "twinned" (with parts at different orientations). Twinning is noted in the wwPDB validation report; it can be dealt with fairly well in refinement (Yeates, 1997; Zwart, Grosse-Kunstleve, & Adams, 2005), but there is still less information content than for an untwinned crystal.

2.2 Model-to-data match

Model-to-data match evaluates how well the modeled atomic coordinates account for the observed diffraction data: the structure-factor amplitudes, or "*F*"s, measured for each X-ray reflection. The global criterion is called the R-factor, which is the residual disagreement between the F's observed and the F's back-calculated from the model. Even more revealing is the cross-validation residual, R_{free} (Brunger, 1992), calculated for 5–10% of the data kept out of the refinement process. A long-time rule of thumb was that near 2 Å resolution the R should be $\leq$20% and $R_{\mathrm{free}} - R$ should be $\leq$5%. Now, however, one need not rely on such rules, because wwPDB validation for each structure reports how its R_{free} compares, as a percentile relative to all other deposited structures at similar resolution (see open bar on top "slider" bar in Fig. 4).

Local match of an individual residue's model cannot be evaluated in the "reciprocal space" of the diffraction peaks, because the Fourier transform between the data and the electron density image relates every atom to every reflection and vice versa. Therefore, local match is evaluated in the "real space" of the model and in the electron density map. The usual criterion is either the real-space correlation coefficient or the real-space residual (RSR) measured as the agreement between the "experimental" electron density (usually $2\mathrm{mF_{obs}} - \mathrm{DF_{calc}}$; Read, 1986) and the purely model-calculated F_{calc} electron density, within a mask around the atoms in that residue (Jones, Zou, Cowan, & Kjeldgaard, 1991). The wwPDB reports percentile scores for RSR-Z, which is a version of the RSR normalized as a Z-score (essentially the number of standard deviations from the mean)

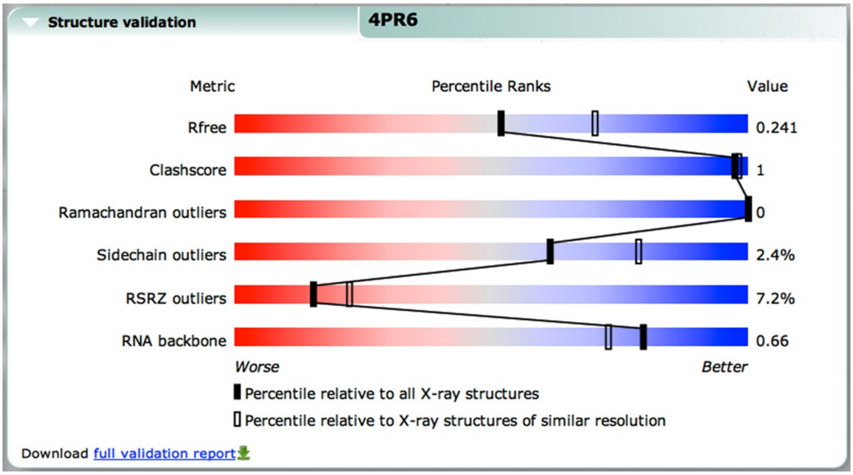


Figure 4 wwPDB summary "slider" validation plot for an RNA/protein complex (4PR6 at 2.3 Å; Kapral et al., 2014). Percentile scores on six validation criteria are plotted versus all PDB X-ray structures (filled bars) and versus the cohort at similar resolution (open bars).

compared to values at that resolution for the same amino acid or nucleotide (Gore et al., 2012; Kleywegt et al., 2004; next-to-last slider bar in Fig. 4).

2.3 Model validation

Model validation covers covalent geometry and steric interactions, both applicable to any molecule, and conformational parameters, which are quite distinct for protein and RNA and are therefore covered in detail in Section 3.

Geometry criteria include covalent bond lengths and angles, planarity, and chirality. Target values, including their estimated standard deviations or weights, are derived from the databases of small-molecule crystal structures (Allen, 2002; Grazulis et al., 2009), or perhaps from quantum calculations, especially for unusual bound ligands (e.g., Moriarty, Grosse-Kunstleve, & Adams, 2009). With some exceptions discussed below, geometry validation primarily serves as a sanity check for whether sensible restraints were used in the refinement.

Steric interactions include both favorable hydrogen bonds and van der Waals interactions and also unfavorable or even impossible atomic overlaps, or "clashes." Not every donor or acceptor is H-bonded and not every atom grouping is tightly packed, but that should nearly always be true in the molecule interior and especially in regular secondary structure. If there are two possible conformations consistent with the electron density and one of them has more of the good interactions, then it is much more likely to be correct. Bad steric clashes have been flagged as problems in just about every relevant analysis, but only for non-H atoms until our lab's development of all-atom contacts (Word, Lovell, LaBean, et al., 1999; Word, Lovell, Richardson, et al., 1999), which is the most distinctive contribution of MolProbity validation. Our Reduce program adds all H atoms, now by default in the electron cloud-center positions (Deis et al., 2013) that are most appropriate both for crystallography, where it is the electrons that diffract X-rays, and also for all-atom contact analysis, where van der Waals interactions are between the electron clouds not between the nuclei. Most H atoms lie in directions determined quite closely by the planar or tetrahedral geometry of their parent heavy atoms, and even methyl groups spend almost all their time very close to a staggered orientation. Reduce then optimizes the rotatable OH, SH, and NH_3 positions within entire H-bond networks, including any needed correction of the 180° "flip" orientation for side-chain amides and histidine rings (Word, Lovell, Richardson, et al., 1999). The Probe program analyzes all atom–atom contacts within 0.5 Å of touching van der Waals surfaces, assigning numerical scores and producing visualizations as paired patches of dot surface like those seen in Figs. 2 and 3. A cluster of hotpink clash spikes gives the most telling signal of a serious local problem in the model. Barring a misunderstood atom nomenclature, a flagged steric overlap >0.4 usually, and >0.5 Å nearly always, means that at least one of the clashing atoms must move away. The MolProbity "clashscore" is normalized as the number of clashes per 1000 atoms in the structure and is reported as percentile scores by the wwPDB (Fig. 4, second slider bar). As shown by the example in Fig. 2, all-atom clashes are a valuable diagnostic for RNA backbone conformation.

3. CRITERIA SPECIFIC TO RNA CONFORMATION

3.1 RNA bases

Pairing and stacking of the bases is the most obvious and the most energetically important aspect of RNA conformation. Nearly all bases in large RNAs or RNPs are both stacked and paired, but the stacking may be with intercalated small molecules or with protein aromatic side chains and the pairing is often not canonical Watson–Crick. Base pairs can deviate a fair amount from coplanarity but should maintain good H-bonding. There is simply not enough room for a purine–purine pair within an A-form helix, but occasionally a pyrimidine–pyrimidine singly H-bonded pair can be tolerated, such as the C–U that ends the helix below the S-motif in the rat sarcin–ricin loop (Correll et al., 1998) compared to the G–C pair in that position for *E. coli* (Correll, Beneken, Plantinga, Lubbers, & Chan, 2003). Comparing across phylogenetically related sequences, base pairing is very strongly conserved both in helices and in other structural motifs, as judged by "isosteric" replacements (Leontis & Westhof, 2001; Stombaugh, Zirbel, Westhof, & Leontis, 2009), but there are occasional surprises such as the hole in *E. coli* Phe tRNA next to the nonpaired position 26 base (Byrne, Konevega, Rodnina, & Antson, 2010; Dunkle et al., 2011). The H-bonds of base pairs and the broad, tight van der Waals contacts of base stacking are shown visually by all-atom contacts (as in Fig. 2), but neither MolProbity nor wwPDB validation evaluates them quantitatively. They can be checked out with the MC-Fold/MC-Sym system (Parisien & Major, 2008), for instance, and probably should be added into formal nucleic acid validation criteria.

The bond that connects the base to the ribose is called the glycosidic bond, and the dihedral angle across that bond is called the χ angle, defined by atoms O4′-C1′-N9-C4 for purines (bases A and G) and O4′-C1′-N1-C2 for pyrimidines (bases C and U). The χ angle is known to adopt two major conformations: in the much more common *anti* conformation, the bulkier side of the base (the six-membered ring for purines and the ring substituent oxygen for pyrimidines) points away from the sugar, putting χ near 180°, while in the less common *syn* conformation the bulky group points toward the sugar (Saenger, 1983). A variant of *anti* conformation called high-*anti* (near 270°) is also seen sometimes, which is awkward for the generic chemical definition of *syn* as $0 \pm 90°$ and *anti* as $180 \pm 90°$. An even more serious disconnect, now corrected, is that early parameters for χ were often based on DNA (B-form) rather than on the very different RNA distributions. It is now generally accepted that the spread of χ values is greater for purines than for pyrimidines.

To define updated parameters for RNA χ, we used the RNA11 dataset of low-redundancy ≤3 Å resolution chains and then filtered to omit any base with a steric clash, B-factor ≥40, chemical modification, or ribose-pucker outlier (see Section 3.2). The original 25,000-residue distribution populates all 360° of χ, but the high-quality 10,000-residue distribution is empty for two ranges of 110–150° and 320–360°. Both base type and pucker state make large differences in what values can occur, in patterns that are highly nonperiodic and not symmetric around zero. Comparing Fig. 5A

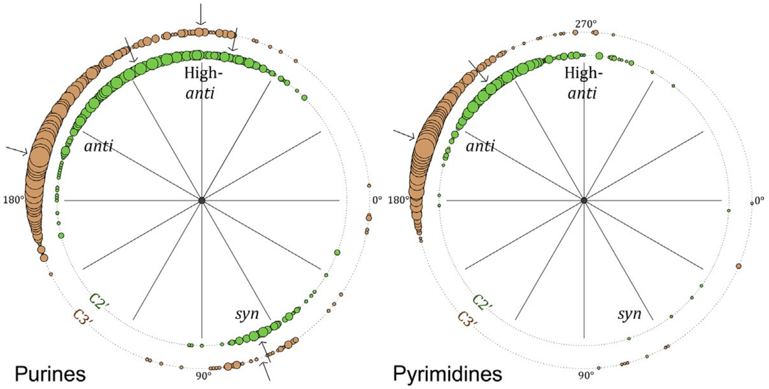


Figure 5 Empirical plot of preferred base χ angle values from the RNA11 dataset, separated by purine (left) versus pyrimidine (right) and by C3′-*endo* pucker (outer ring) versus C2′-*endo* pucker (inner ring). Occurrence frequency is shown by circle radius, on a log scale. Arrows mark the pucker-specific χ-angle parameters now used in PHENIX.

(purines) with B (pyrimidines) shows that pyrimidines have essentially no reliable instances of *syn* χ. Comparing the outer versus inner plots shows a very large shift of the *anti* peak between C3′-*endo* and C2′-*endo* ribose puckers, and also that C3′-*endo* pyrimidines almost never have high-*anti* conformations. Since purine bases are much larger, it seems counterintuitive that they have less conformational restriction. However, close interactions between base and backbone are with the larger six-membered ring for pyrimidines and with the smaller five-membered ring for purines. The end result of this update was a set of RNA preferred χ values for use in PHENIX, specific both for base type and for pucker type, as marked by the arrows in Fig. 5.

3.2 Ribose pucker

The ribose ring in RNA structures is known to adopt two main conformations, C3′-*endo* pucker and C2′-*endo* pucker. There is a clean bimodal distribution in the CSD small-molecule data (Allen, 2002;), echoed closely at high resolution and high quality in the RNA11 dataset, here plotted in Fig. 6 as a function of the closely correlated backbone dihedral angle δ (C5′-C4′-C3′-O3′). The mean value of δ is 84° for C3′-*endo* and 145°

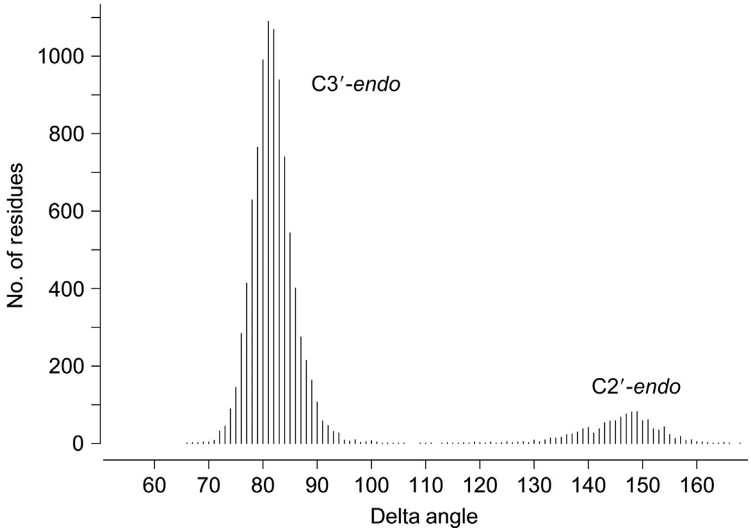


Figure 6 Occurrence frequency at each δ dihedral-angle value. *Data from the RNA11 dataset, with residue-level filters of backbone atom clash, B-factor ≥60 and geometry outliers applied. The distribution is bimodal, with δ angle range of 60°–105° for C3′-endo pucker and 125°–165° for C2′-endo pucker.*

for C2′-*endo* pucker, with acceptable ranges of 60°–110° and 125°–165°, respectively.

However, as described in Section 1, it is extremely difficult to place the backbone atoms correctly into the electron density and distinguish between the two puckers, especially at the resolutions lower than 2.5 Å that are typical for large RNA structures. Therefore, ribose puckers are often modeled incorrectly in RNA structures, with an understandable but too thorough bias toward the C3′-*endo* conformation found four times more often for RNA residues. The three large groups of atoms attached to the ribose (the two backbone directions and the base) are pointed in different directions by the different pucker states, which means that fitting and refining those groups around an incorrect pucker nearly always produces further problems such as clashes and bond-angle outliers. Pucker outliers occur even at fairly high resolution and may be in biologically important regions, such as the tRNA/synthetase example in Fig. 7.

Fortunately, there exists a simple and reliable way to detect the pucker of an RNA residue from several atoms of RNA structure models that in practice turn out to be relatively well placed in the electron density. Even at low resolution, the phosphorus atom is well centered in a nearly spherical high peak of the map (see Fig. 1) but flanked by round, featureless tubes of backbone density. The large blob of base density is offset to one side from the smaller ribose blob. Fortunately, initial model construction into this evident configuration and spacing of density usually places the P atom and the C1′–N1/9 glycosidic bond (the line connecting ribose and base) very close to their final refined positions. If a perpendicular is dropped from the 3′ P to

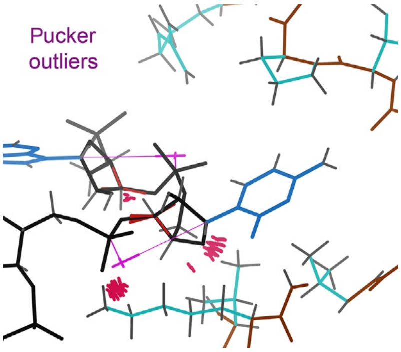


Figure 7 Example of pucker outliers (magenta (highlighted by circle and arrow in the print version) crosses, pointing along the glycosidic bond vector), at the contact of a tRNA CCA end with its synthetase enzyme. *From PDB 1N78 at 2.1 Å (Sekine et al., 2003).*

the extended line of the glycosidic bond vector, the length of that perpendicular (abbreviated as the "Pperp" distance) reliably indicates the pucker of the ribose ring: it is long for C3′-*endo* pucker and short for C2′-*endo* pucker.

This relationship is empirically confirmed by the data shown in Fig. 8, where an initial approximate correlation improves to a clean distinction as additional residue-level filters are applied to the residues in the RNA11 dataset. A simple Pperp distance cutoff can reliably distinguish between the two puckers: $\geq$2.9 Å for C3′-*endo* pucker and $<$2.9 Å for C2′-*endo* pucker. This relationship can be judged quite easily by eye, and is trivial to automate. It holds well even for misfit models, so validation can tell whether the ribose pucker of an RNA residue is correct or not by testing whether the Pperp distance is paired with the correct range of backbone dihedral angle δ. If not, it is flagged as a pucker outlier. Tools for correcting ribose-pucker outliers are described in Section 5 below.

3.3 Backbone conformers

The RNA backbone is highly flexible with 6 degrees of freedom (dihedral angles α–ζ) and, as described in Section 1, it is extremely difficult to place correctly in the electron density. Fortunately, there is order in this chaos. The individual dihedral angles sample a large range of angle values, but when analyzed together as high-dimensional clusters, they describe a discrete set of

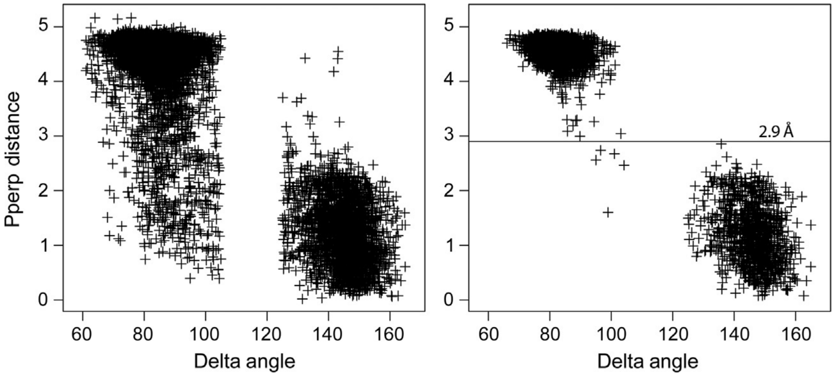


Figure 8 Distinguishing whether the ribose pucker should be C3′-*endo* or C2′-*endo*, from the "Pperp" length of the perpendicular dropped from the 3′ P to the glycosidic bond vector. On the left, Pperp versus δ angle for unfiltered RNA11 data. On the right, RNA11 data filtered on backbone atom clashes, on backbone atom B-factor $\geq$60, and on ε and geometry outliers. The cutoff used to distinguish between the two pucker conformations is 2.9 Å, with Pperp value $\geq$2.9 Å for C3′-*endo* pucker and $<$2.9 Å for C2′-*endo* pucker.

conformations that the RNA backbone can adopt. RNA backbone has been shown to be rotameric (Murray et al., 2003), a study that was extended by the RNA Ontology Consortium to describe 54 distinct conformations commonly adopted by the RNA backbone (Richardson et al., 2008).

These backbone conformers are described in terms of the sugar-to-sugar unit of the RNA backbone called a "suite", because the dihedral angles within a suite are better correlated than the dihedral angles in a traditional nucleotide (Fig. 9). Each suite i consists of seven dihedral angles: δ, ε, and ζ from the heminucleotide i-1, and α, β, γ, and δ from the heminucleotide i. The 54 backbone conformers were identified using the data from a quality-conscious, nonredundant dataset of RNA crystal structures (RNA05), with each distinct cluster in the seven dimensions of the suite identified as an individual conformer. Most conformers are well separated in the seven-dimensional space but some are not, such as the "satellites" around the big A-form cluster. Close cluster pairs can usually be distinguished on the

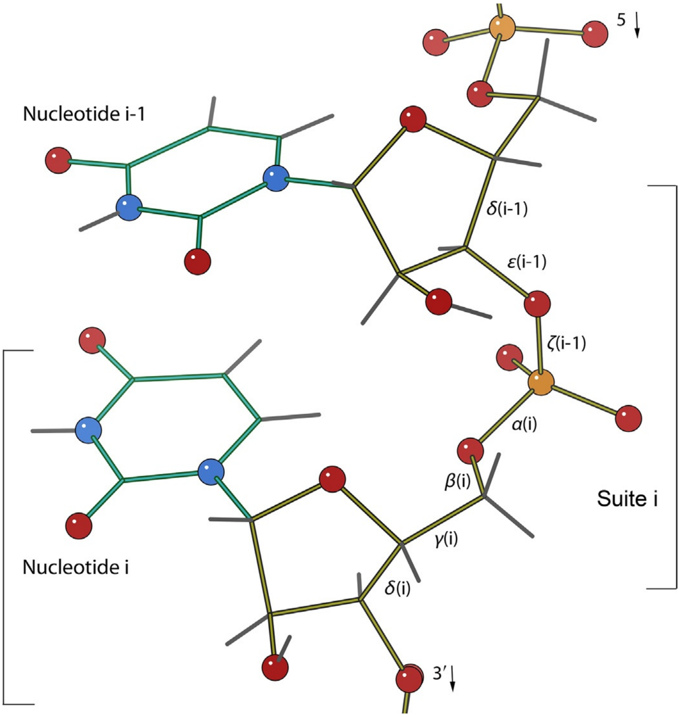


Figure 9 Definition of the "suite" division of RNA backbone that spans from sugar to sugar, with its seven backbone dihedral angles: δ, ε, and ζ from heminucleotide i-1, and α, β, γ, and δ from heminucleotide i.

basis of their different structural roles, such as the base-stacked **1b** versus the intercalated **1[**.

Each conformer is given a two-character name with first a number describing the conformation of heminucleotide i-1 and second a letter (or letter-related character) describing the conformation of heminucleotide i. For example, **1a** represents the RNA backbone conformation in a standard A-form helix, **1[** is the most common intercalation conformation, **1g** is the starting conformer for the GNRA tetraloop motif, and **5z** and **4s** are two of the three conformers that describe the backbone of an S-motif. For automated backbone conformer assignment in PHENIX and MolProbity, a program called Suitename was developed. Suitename takes as input the seven backbone dihedral angles, assigns the suite one of the 12 δ–γ–δ bins (two ranges for δ angles and three ranges for γ angle) and then assigns a backbone conformer to the suite within that bin by a rather complex process (Richardson et al., 2008). It also provides a value called the "suiteness" for each suite, which indicates how far from the center of the backbone cluster this particular suite is. Any suite which does not belong to the 54 backbone conformers is flagged as a backbone-conformer outlier, denoted by **!!**.

One of the advantages of using backbone conformers and their two-character names is that any three-dimensional RNA structure can be represented as a string of two-character backbone conformers' names, called a "suitestring." Many commonly occurring structural motifs have consensus suite strings, for example, kink-turns (**7r6p2[0a** on the kinked strand), S-motifs (**5z4s#a** on the "S" strand, **1e** opposite), or GNRA tetraloops (**1g1a1a1c** above a Watson–Crick pair). These suitestrings are useful to identify new occurrences of known structural motifs and to find new structural motifs. In the present context, they form the basis for validating RNA backbone conformation and are especially important to the correction process (described in Section 5). This is a two-way relationship, as the recognized suites are found by ERRASER even though they are not explicitly used in its scoring function, and we now use ERRASER to verify possible new conformers that have too few examples to define empirically.

4. HOW TO INTERPRET VALIDATION RESULTS

4.1 As an end user

For a general orientation to the 3D structure of your RNA of interest, resolution or validation scores do not matter, and primarily you just want to take into account the state of processing, binding partners, ligands, or alternative conformational forms for the molecule in a particular deposited structure as well as how closely related it is to your own research subject. However, when specific details matter, you can benefit greatly from considering validation information, both global and local.

If many relevant structures are available, you can choose the best one or ones to use overall by starting with the "slider plots" (such as the one in Fig. 4) on each structure's page at the wwPDB Web sites. Compare the absolute values of the various validation criteria, or just visually judge the positions of the filled bars on each slider, which show the percentile score within PDB entries at all resolutions. If the most relevant structure is low quality, then you should look both at that one and also at the best generic structure of the set.

Most important are local validation criteria in the regions you especially care about for your own work. Local accuracy and reliability often vary widely within a single structure, mostly because some parts are much more mobile than others. Where the map is unclear, the model can be influenced by subjective judgments or by software idiosyncrasies. Even high resolution and excellent quality scores cannot protect you from a serious local modeling error; conversely, even in a poor-quality structure, if the local region of interest has no validation outliers at all, then it is very probably correct.

At the wwPDB sites (RCSB, PDBe, PDBj), local validation is in the "full validation report" pdf for each entry; a download link is just above or below the summary "slider" report (see Fig. 4). In the "Overall quality at a glance" section, the last table lists any ligands whose geometry or electron density quality is an outlier; further details are in later sections. An outlier in electron density quality (LLDF) is certainly not bound at full occupancy and risks having a component of wishful thinking. In the "Residue property plots" section, you can see where there are concentrations of geometry or density match outliers along the sequence, in each RNA or protein chain. *Note*: the wwPDB uses "geometry" to include dihedral-angle conformation as well as covalent geometry. Look for the residue numbers you care about; if they show up, then you probably want to check the more detailed information sections, or better yet consult a representation of validation markup on the 3D structure.

The most powerful way to evaluate local model quality is interactively in the three-dimensional structure, where you can see clustering in space as well as in sequence, and even consult the electron density map (usually what you want is called a $2F_o$-F_c map). At the wwPDB and elsewhere, there are now many interactive 3D viewers for a quick and easy look at overall molecular structure, but they are not as good for comprehending local detail, and none of them show validation markup. The Electron Density Server (EDS; Kleywegt et al., 2004; eds.bmc.uu.se/eds) has two different online viewers available for looking at the model and electron density map in 3D. It also has excellent per-residue sequence plots of the various model-to-data validation metrics such as RSR-Z, with nothing omitted and mouse-over details, and it is the easiest source to download maps for viewing in other software. PyMol (www.pymol.org), still free to academics, is probably the most full featured of the nonexpert-accessible macromolecular viewers, both for interactive use and for making 2D presentation images. It can show electron density maps but does not have an easy way to import validation markup. The only user-friendly, high-comprehension software we know of for viewing models, maps, and validation markup in 3D is KiNG, which was used for producing most of the figures here. It is a central feature of the MolProbity Web site, for viewing the "multi-criterion kinemage," doable directly online without any installation needed as long as you can run Java, or KiNG can be run on your own computer. From the overview, either locate the worst clusters of problems visually and zoom in on them, or find and center on a residue of interest and see how good that region is. If it is in a cluster of problems, you should treat its information as unreliable in detail. On the other hand, if that residue and its neighbors are free of major outliers (true for most parts of most crystal structures), then the local model and inferences from it are trustworthy.

Single outliers need a more nuanced approach. A bad steric clash really does mean that one or both of the clashing atoms must move, although perhaps not by much. A ribose-pucker outlier is almost certainly wrong. A backbone-conformer outlier (!!), however, is fairly often valid, especially if the bases of the suite are far apart such as in a one-residue bulge or a helix junction. Bond-angle outliers are not necessarily serious for their own sake, but in RNA they are most often a symptom of a locally misfit conformation. As in any macromolecule, disorder at chain termini or elsewhere means there is more than one conformation, which makes fitting even more difficult and almost always produces outliers. Although that casts no doubt on the overall quality of the structure, it still means that you cannot know the local conformation well.

For the proteins in RNP structures at low resolution, keep in mind that there may be significant problems such as local out-of-register sequence, unjustified Ramachandran outliers or *cis* peptides, or mismodeled interactions at the protein/RNA interface. For instance, an Arg side chain ending near a phosphate but forming no H-bonds is very likely to have its rotamer fit backward. In such a case, you might benefit from consulting the version on the PDB Redo Web site (www.cmbi.ru.nl/pdb_redo; Joosten, Joosten, Murshudov, & Perrakis, 2012), which re-refines all PDB crystal structure entries, including automated side-chain and peptide-flip rebuilding for proteins but not yet local corrections for RNA.

In this process, when you have discovered a potential local problem of concern in the most relevant structure, it is often quite helpful to check out that same region in other related structures. See whether they agree or disagree, and consider that in the context of how they differ both in information content and in biological context.

4.2 As a journal referee

First of all, feel free to ask for coordinates and structure factors—you may well get them, and are then positioned to do the best job of fairly evaluating the structure. Failing that, however, you should certainly require access to the wwPDB validation report. Here are some suggestions on what to look for in such reports: what's good, or at least OK; what needs an explanation; and what is unacceptable.

In the context of journal review, the technical issue that matters most is how well the authors handled the data that could be collected from the system in question. Therefore, the percentile scores versus the resolution cohort (the open bars on the wwPDB summary slider plots, as in Fig. 4) matter much more than absolute or all-PDB scores. A high-resolution structure should be held to higher absolute standards of accuracy than a low-resolution one. The validation criteria have been chosen to be as independent as feasible, so they do not necessarily all have similar values; it is somewhat unusual for a structure to score near the 60th percentile, say, on all measures. It is not particularly disturbing, for instance, to have very good Ramachandran scores and quite poor rotamer scores; that presumably just means the crystallographer, or the software, concentrated much more on the backbone. Some combinations are disturbing, however. Top scores for R_{free} and/or RSR-Z coupled with near-zero percentiles for clash and conformation criteria might either indicate some poor methodology that produced extreme model bias (possible, for instance, if twinning is handled incorrectly) or even possible fraud (Janssen, Read, Brunger, & Gros, 2007), or else might just mean that the methodology was complex enough to be treated poorly by the automated routines that calculate those crystallographic scores. Such cases need to be looked at carefully, and probably questions should be asked of the authors. And of course the issue still remains of the near-zero conformational percentiles.

Note that RSR-Z is reported for values >2, which is only a 2σ outlier, compared to the 4σ level or more for most other outlier definitions. The overall RSR-Z score can look bad just from having some mobile regions in the structure (e.g., for the 4PR6 of Fig. 4), because the RSR is even more sensitive to low electron density than to the exact shape of that density. This effect will change dramatically depending on whether the depositor believes in making a complete model, or believes in leaving out anything not seen clearly; there is no good answer to that dilemma, so neither side should be penalized for their decision. RNA crystallographers are generally on the side of more complete models. Locally, poor (high) per-residue RSR-Z scores just mean that the density in that region was weak and that part of the molecule presumably mobile but still modeled, not that anything is wrong with the structure or the methodology in general. But if poor RSR-Z scores occur in a place important to the conclusions of the paper, then they must be explained or discounted, and in general necessitate explicit defense of any strong claims.

An especially important aspect of validating model-to-data match is whether or not there is good evidence in the electron density to support the modeled presence of a bound ligand, inhibitor, or drug. If a bound inhibitor, drug, or other ligand is discussed in the paper, then check its ligand geometry and especially its match to data in the detailed tables for the presence of yellow flags. Highly deviant ligand geometry is always suspect, especially at lower resolution or in weak density, and may mean something else is locally wrong besides that geometry. The "LLDF" score estimates the electron density quality of the ligand relative to that of the surrounding macromolecule. If that score is an outlier, the ligand is certainly not bound at full occupancy, and no strong conclusions should have been drawn about its conformation or even its binding. The request to make in such cases is for an omit map covering the region of the ligand, to see whether there is unbiased evidence supporting its presence and conformation.

Not every structure needs to be, or can be, in the top 10%, so scores that are on average typical for the resolution should be considered acceptable in most cases. A structure far worse than average should be questioned, however. If it is not an especially important structure, then it may not be worth reporting in an inaccurate state. If it is really important, then as a referee it would be worth pushing for improved quality to better support the presumed important uses of it; if it is not corrected now, then neither the depositor nor anyone else will be able to get funding to improve it later.

4.3 As a structural biologist

Besides a better report card from the wwPDB, you can benefit from better refinement behavior, clarify the map in other places through phase improvement, and especially can locate the places that are wrong in ways that can often be successfully corrected but will occasionally be valid and interesting, such as the well-established case in Fig. 10. In general, MolProbity validation aims for conservative identification of problems, to avoid wasting user effort on chasing false alarms.

Backbone clashes are especially diagnostic in RNA, and some types are easy to correct, such as when the 5′ H atoms are turned inward, putting them too close to the ribose (as for the manual correction example in Section 5.1). A ribose-pucker outlier is essentially always incorrect; they can sometimes be fixed in Coot, or even by refinement with pucker-specific targets, but their repair may require running ERRASER (see Section 5 below). Suite outliers are less serious, since not all valid conformers have yet been identified. Aim for 80–90% known conformers, but not for 100% unless your structure is pure A-form helix. The wwPDB slider reports percentile scores

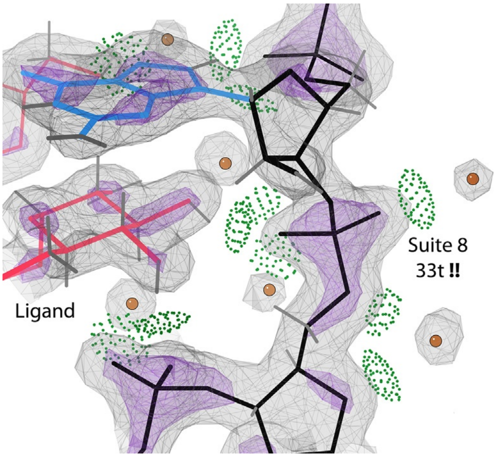


Figure 10 A valid !! "outlier" suite conformer, with excellent confirming electron density and favorable interactions. The $2F_o$-F_c density is shown at 1.2σ as pale gray mesh and at 3.2σ as purple. The water molecules are shown as separate balls and hydrogen bonds as pillows of dots. *From PDB 3C3Z, HIV-1 RNA/antibiotic complex at 1.5 Å (Freisz et al., 2008), near residue B 8.*

for the average "suiteness" per residue, which is useful but not the best indication of outright mistakes. Using MolProbity, or consulting the detailed tables in the wwPDB validation report, you can find specifically which nucleotides or suites are pucker or conformer outliers.

A few types of base modification, such as the dihydro-uridine common in tRNAs (Dalluge, Hashizume, Sopchik, McCloskey, & Davis, 1996), break aromaticity of the ring and should indeed be strongly nonplanar. For other bases, deviations from base-ring planarity great enough to notice by eye are a warning sign of missing or poorly balanced refinement restraints—although not necessarily of an otherwise inaccurate structure. In contrast, a spatial cluster of bad bond-angle outliers is usually caused by a serious misfitting of the local conformation and is worth worrying about, in either protein or RNA. Historically, achieving a low level of geometry outliers has been quite difficult for RNA structures, and simply tightening the restraints is not very effective. However, now that we have access to better tools for correcting local errors in RNA conformation, such as ERRASER, we find that geometry improves as an additional result and persists robustly in further refinement.

As a crystallographer, the most helpful form of validation and correction is to work within the PHENIX GUI (Echols et al., 2012), where MolProbity-style validation is reported at the end of each macrocycle in tables, plots, and a multi-criterion kinemage. Each entry in a table can with one click take you to the relevant place in Coot to work on the problem. Even if you prefer another program for refinement, it might well be worth an occasional excursion into PHENIX, to refine briefly with pucker-specific targets and then use the RNA-specific validation as coupled with Coot. Tools for RNA backbone correction and examples of their use will be found in Section 5.

The take-home message (Richardson & Richardson, 2013) is what we like to call:

The zen of model anomalies:

Consider each outlier and correct most.
Treasure the valid, meaningful few.
Do not fret over a small inscrutable remainder.

5. CORRECTING THE PROBLEMS

Detection of modeling errors in RNA and RNP structures is valuable, especially for the hard-to-fit backbone. However, it is more satisfying and very much more useful if that diagnosis can be followed by correction. This section describes the development of increasingly powerful and user-friendly tools to correct RNA modeling errors.

5.1 Manual rebuilding

Manual rebuilding of individual RNA backbone suites, even with graphics tools developed specifically for the task, is tedious and limited in scope, but quite educational. First the SuiteFit tool in Mage (Richardson & Richardson, 2001) graphics and then the RNA Rotator tool available in KiNG (Chen, Davis, & Richardson, 2009) were developed for manual rebuilding of individual RNA backbone suites. They allow the user to interactively adjust the seven backbone dihedral angles and two χ angles associated with the suite, or to select a new starting point from a list of all 54 backbone conformers. As well as the essential reference of the electron density contours, display of all-atom contact dots is updated in real time as the user changes dihedral angles, providing essential feedback. In Mage, the user controls overall rotation and translation of the suite explicitly. In KiNG, the user controls suite positioning by specifying a set of atoms to keep optimally superimposed on the original as angles are changed, and additional feedback is provided by real-time update of the "suiteness" conformer-quality parameter (see Section 3.3). This process is only sometimes successful even for expert users.

Figure 11 shows before and after states for a manual correction done using RNA Rotator on suite 542 in the 50S ribosomal subunit of

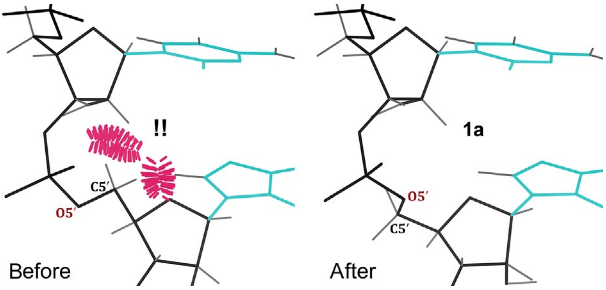


Figure 11 Example of a manual RNA backbone correction. At left, the originally modeled backbone of suite 542 clashes in both directions, with backbone of 541 and base of 542 (clusters of hotpink (dark gray in the print version) spikes), and suite 542 is a backbone-conformer outlier (**!!**). At right, after clash correction using RNA Rotator, suite 542 adopts a valid **1a** conformer. *From PDB 3CC2, Hm 50S ribosomal subunit at 2.4 Å (Blaha et al., 2008).*

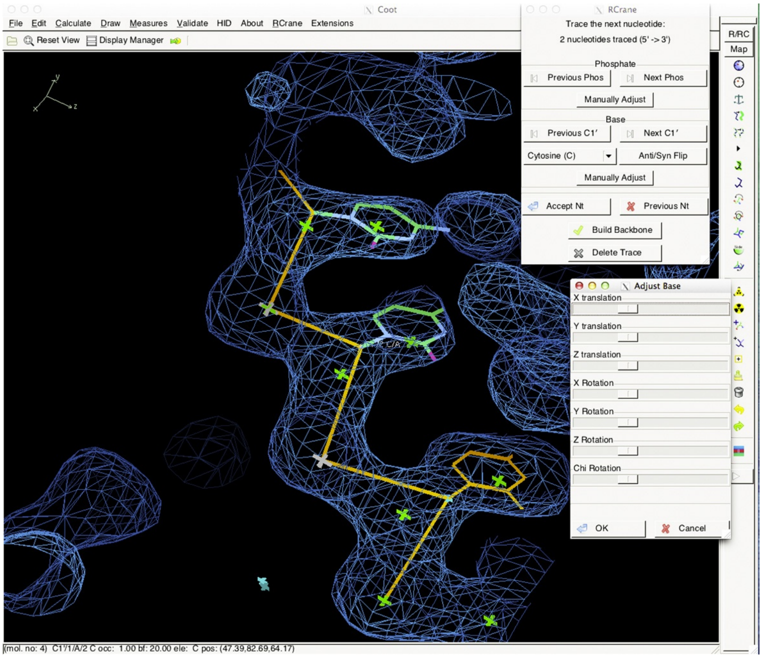
H. marismortui (PDB ID: 3CC2). Changing the α, β, and γ backbone dihedral angles and superimposing the new suite on the 5′ and 3′ ends and the base atoms, suite 542 was changed from a **!!** to a **1a** conformation, which then refined successfully.

5.2 RNABC

RNABC is an automated tool developed to help fix all-atom clashes in RNA backbone (Wang et al., 2008). As previously noted, the base and the phosphate are most easily visible in the density, hence RNABC keeps the base and P atom fixed and searches for a better rebuild of the rest of a specified dinucleotide using forward kinematics. The bond lengths and angles are held fixed, and the pucker of the two ribose rings in the dinucleotide is either specified by the user or determined by the Pperp test (described in Section 3.2). Possible conformations are scored based on steric clashes, pucker, and geometry terms, and the best-scoring set of nonredundant possible conformations are output for the user to choose from. However, RNABC does not take the electron density or the known RNA backbone conformers into account during the rebuilding process.

5.3 Coot and RCrane

RCrane (Keating & Pyle, 2012) is a plug-in available in the Coot model-building software (Emsley, Lohkamp, Scott, & Cowtan, 2010), to do semiautomated building of RNA structure into electron density. It is partially based on the coarse-grained RNA backbone parameters η and θ (the two pseudo-dihedral angles defined by the P and C4′ atoms), used successfully to describe and locate RNA structural motifs (Duarte & Pyle, 1998). For this purpose, they are modified to a θ', η' form that uses the more reliably fit C1′ atom rather than C4′, and the suite rather than the nucleotide division. These dihedrals are related to the all-dihedral RNA backbone suite conformers (see Section 3.3), though the relationship is not one-to-one. The tool takes as input the electron density and builds a trace for the user-selected nucleotides. This is done by marking the highest intensity peaks as crosses (see Fig. 12 screen shot) and allowing the user to choose the probable P and C1′ atoms from these density peaks and adjust their positions (within 10 Å) as needed. They are joined alternately into a proposed virtual backbone trace. An automated selection is then made of the few most probable backbone conformers (based on θ', η', Pperp, C1′–C1′, and P–P distances), followed by individual coordinate minimization within the



(mol. no: 4) C1'/1/A/2 C occ: 1.00 bf: 20.00 ele: C pos: (47.39,82.69,64.17)

Figure 12 Screen shot of the RCrane tool in Coot. Peaks of electron density are marked with crosses. RCrane allows the user to choose and manually adjust the position of the P and C1′ atoms. The resulting virtual-bond backbone trace is shown, from which probable backbone conformers will then be proposed.

electron density. The user chooses from the list of possible conformers and their scores, for each built nucleotide. This technique can also be used for correction of errors in the RNA structure, given the input model and the electron density, and rebuilding selected nucleotides.

5.4 PHENIX refinement with pucker-specific targets

Now that the right pucker state can be identified even for misfit nucleotides using the Pperp test, that simple function allows the use of pucker-specific dihedral-angle and bond-angle targets to be used in refinement, a functionality available in PHENIX (Adams et al., 2010).

5.5 ERRASER, with examples

All the correction and rebuilding tools described above require major input from the user, either in identifying errors or in the actual correction or building process, and none of them have nearly as high a success rate as outlier correction in protein structures achieved some time ago

(Arendall et al., 2005). Now, however, we have software that has proven to be truly effective in automated error correction of RNA backbone, called ERRASER (Adams et al., 2013; Chou et al., 2012). Figure 13 shows the quite revolutionary level of cleanup that ERRASER can achieve. It utilizes capabilities in PHENIX and MolProbity for identification of modeling errors, and a stepwise assembly (SWA) procedure to rebuild each residue by enumerating many conformations covering all build up paths, taking into account the fit of the model to the electron density. ERRASER can be used to rebuild whole RNA structures (where error detection is done automatically), or single residues as specified by the user, in which case ERRASER returns its top 10 distinct conformations for the user to choose among. One can also specify that a particular set of residues remain fixed.

The rebuilding process in ERRASER consists of three steps: First, ERRASER minimizes all torsion angles and all backbone bond lengths and bond angles using the Rosetta energy function for RNA (Das & Baker, 2007), including an electron density correlation score (DiMaio et al., 2009). Second, PHENIX's MolProbity-style RNA validation tools are used to identify errors (geometry, pucker, and unrecognized backbone conformations) in the minimized model. These residues, as well as residues with large rms deviation (>2 Å) between their original position and the

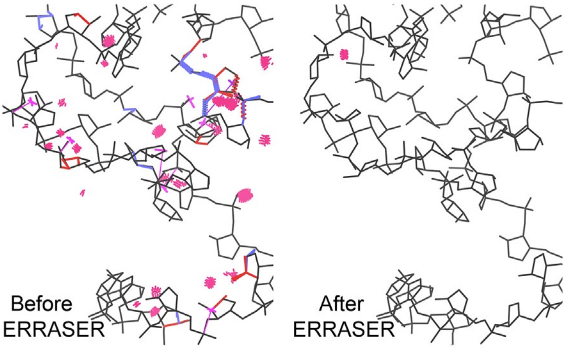


Figure 13 Overall correction efficiency of ERRASER, supplemented with other tools. The active-site region of the uncleaved HDV ribozyme structure, with the RNA backbone in black. At left, the original structure (PDB ID: 1VC7; Ke, Zhou, Ding, Cate, & Doudna, 2004), with many clashes (hotpink spikes), bond-length outliers (red and blue spirals), bond-angle outliers (red and blue fans), and ribose-pucker outliers (magenta crosses). At right, the rebuilt structure (PDB ID: 4PRF; Kapral et al., 2014), with essentially all validation outliers corrected.

minimized position, are identified as residues to be rebuilt. The second step is skipped if the user has specified a particular residue to be rebuilt. Third, the residues from step two are rebuilt one at a time with the SWA procedure, and then minimized again. This process is carried out usually for three cycles.

Figure 14 shows an example of a ribose-pucker correction done using ERRASER. Residue 152 in the uncleaved HDV ribozyme structure (PDB ID: 1VC7; Ke et al., 2004) is incorrectly modeled as a C3′-*endo* pucker, leading to deviations in geometry and steric clashes with the phosphate group of the next residue. O2′ is modeled out of the $2F_o$-F_c density, and there is a large peak of positive difference density (blue mesh) near the residue. ERRASER was run on the entire HDV ribozyme structure, which resulted in this residue being remodeled as C2′-*endo* pucker and correction of all other validation outliers (PDB ID: 4PRF; Kapral et al., 2014). The O2′ moves into $2F_o$-F_c density, getting rid of the positive difference density peak.

ERRASER can also be used to correct a single residue at a time. Figure 15 shows an incorrectly modeled kink-turn in the 50S ribosomal subunit of *H. marismortui* (PDB ID: 3CC2; Blaha et al., 2008), with the suite identities shown in the figure. The first suite in the kink-turn is a **!!** conformer outlier, its backbone clashing with other residues in the kink-turn. ERRASER was run on that residue (1603), and the top-scoring conformation it returned fixed the steric clash and changed suite 1603 from **!!** to **7r**,

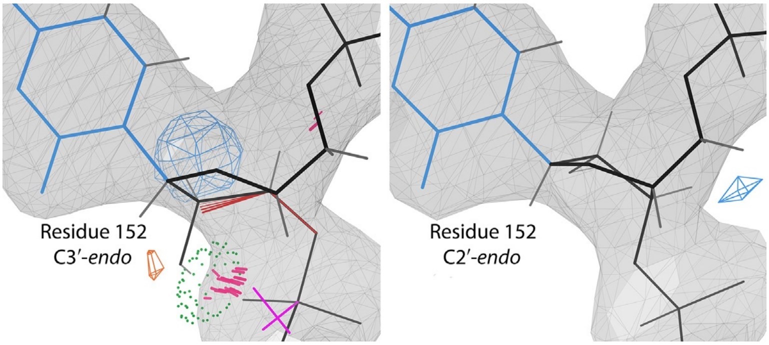


Figure 14 ERRASER correction of a ribose-pucker outlier, as part of a full-structure run. At left, residue 152 modeled incorrectly as $C3_0$-endo pucker (PDB ID: 1VC7), shown in $2F_o$-F_c density at 1.2σ (pale gray mesh). It has a too-tight, clashing hydrogen bond (hotpink (dark gray in the print version) spikes within the green (dark gray in the print version) dot pillow), a bond-angle outlier (red (dark gray in the print version) fan), ribose-pucker outlier (magenta (light gray in the print version) cross), and a large positive difference density peak at 3.5σ (blue (light gray in the print version) mesh). At right, the residue 152 rebuilt by ERRASER as $C2_0$-endo pucker, and all other outliers gone (PDB ID: 4PRF).

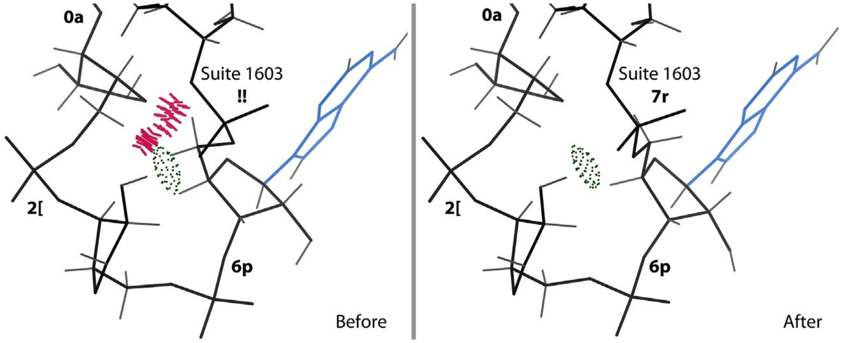


Figure 15 ERRASER single-residue correction of a kink-turn suite. Suites 1603–1606 form a kink-turn motif in the Hm 50S ribosomal subunit (PDB ID: 3CC2), but suite 1603 was modeled as a backbone-conformer outlier (**!!**), badly clashing with the 1605 backbone. At right, ERRASER has made large rotations in the backbone between ribose and 5′ P to resolve the clash and rebuild suite 1603 as a **7r** conformer.

making the kink-turn backbone conformation consistent with the consensus suitestring for a kink-turn motif (**7r6p2[0a**).

ERRASER has proven to be very powerful for correcting backbone errors in RNA structures. However, it requires installation of Rosetta, and because of its comprehensive search for different possible conformations, running ERRASER is compute intensive, especially for large RNA or RNP structures. As with any rebuilding procedure, crystallographic refinement should be run both before and after ERRASER. Compared with the time and effort required for any other RNA correction method, and with their relatively limited success rate, ERRASER is still a very worthwhile bargain.

6. WHAT'S COMING NEXT

6.1 ERRASER for RNA/protein

The current version of ERRASER recognizes only nonmodified RNA residues and removes all other residue types (protein residues, waters, ions, modified residues, DNA, ligands) from the PDB file before starting the structure optimization. This is a problem for rebuilding RNA residues that interact with the neglected residue types, such as in large RNA/protein complexes such as ribosomes. ERRASER can be misled into moving the new RNA conformation into protein or ligand density, which creates impossible steric clashes during the final step of merging the rebuilt structure with the original PDB file. We are working on the thorough rewrite of

ERRASER that will allow it to recognize the currently neglected residue types, and thus take into account their interaction with RNA residues during the rebuilding process. This should greatly improve modeling and correction of both RNA–small molecule and RNP structures.

6.2 Better ion identification

Because of the high negative charge on nucleic acid backbones, ions are a crucial factor for RNA folding, stability, and function. However, it is tricky to locate them in a structure, and even more so to identify their molecular type, since their coordination to the RNA surface is seldom neat and complete, and either partial occupancy or small ions such as Mg are hard to distinguish from waters. Tools are being developed (Echols et al., 2014) and will be expanded, to make use of anomalous-scattering X-ray data that can be diagnostic of ion presence and identity, and of contact information that distinguishes waters from + or − ions, or unmodeled alternate conformations, based on distances and nature of the interacting atoms (Headd & Richardson, 2013).

6.3 Putting conformers into initial fitting

The set of 54 recognized RNA backbone suite conformers (Richardson et al., 2008) are utilized for structure validation and correction, but could also be very valuable in initial fitting of RNA models, especially in non-helical regions. Some initial trials have been done, and work is in progress to take advantage of that possibility. The conformer set is similar in nature to protein side-chain rotamers, but perhaps even more useful, because they apply to the continuous backbone rather than just to an individual side branch. In implementing such a tool, the philosophy is to work from the features best seen in electron density, a P and its two flanking glycosidic bonds, but using at least seven parameters, so as to encapsulate the full information contained in those positions. Such a system would be integrated into the automated procedures in PHENIX.

ACKNOWLEDGMENTS

We would like to acknowledge the contributions of former lab members Laura Murray and Gary Kapral and of our collaborators Rhiju Das, Jack Snoeyink, Kevin Keating, the Phenix developers, and the validation team at the wwPDB. This work was supported by the National Institutes of Health (previously GM088674 and GM074127, currently GM063210 and GM073919).

REFERENCES

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., et al. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica*, *D66*, 213–221.

Adams, P. D., Baker, D., Brunger, A. T., Das, R., DiMaio, F., Read, R. J., et al. (2013). Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annual Review of Biophysics*, *42*, 265–287.

Allen, F. H. (2002). The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallographica*, *B58*, 380–388.

Arendall, B. W., III, Tempel, W., Richardson, J. S., Zhou, W., Wang, S., Davis, I. W., et al. (2005). A test of enhancing model accuracy in high-throughput crystallography. *Journal of Structural and Functional Genomics*, *6*, 1–11.

Batey, R. T., Gilbert, S. D., & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, *432*, 411–415, 1U8D.

Berman, H. M., Henrick, K., & Nakamura, H. (2003). Announcing the worldwide Protein Data Bank. *Nature Structural Biology*, *10*, 980.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*, 235–242.

Blaha, G., Gural, G., Schroeder, S. J., Moore, P. B., & Steitz, T. A. (2008). Mutations outside the anisomycin-binding site can make ribosomes drug-resistant. *Journal of Molecular Biology*, *379*, 505–519, 3CC2.

Brunger, A. T. (1992). Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, *355*, 472–475.

Byrne, R. T., Konevega, A. L., Rodnina, M. V., & Antson, A. A. (2010). The crystal structure of unmodified tRNA Phe from *Escherichia coli*. *Nucleic Acids Research*, *38*, 4154–4162, 4L0U.

Chen, V. B., Arendall, W. B., III, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., et al. (2010). MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallographica*, *D66*, 12–21.

Chen, V. B., Davis, I. W., & Richardson, D. C. (2009). KiNG (Kinemage, Next Generation): A versatile interactive molecular and scientific visualization program. *Protein Science*, *18*, 2403–2409.

Chou, F.-C., Sripakdeevong, P., Dibrov, S. M., Hermann, T., & Das, R. (2012). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods*, *10*, 74–76.

Correll, C. C., Beneken, J., Plantinga, M. J., Lubbers, M., & Chan, Y. L. (2003). The common and distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Research*, *31*, 6806–6818, 1Q9A.

Correll, C. C., Munishkin, A., Chan, Y. L., Ren, Z., Wool, I. G., & Steitz, T. A. (1998). Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proceedings of the National Academy of Sciences of the United States of America*, *95*, 13436–13441, 430D.

Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A., & Davis, D. R. (1996). Conformational flexibility in RNA: The role of dihydrouridine. *Nucleic Acids Research*, *24*, 1073–1079.

Das, R., & Baker, D. (2007). Automated *de novo* prediction of native-like RNA tertiary structures. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 14664–14669.

Deis, L. N., Verma, V., Videau, L. L., Prisant, M. G., Moriarty, N. W., Headd, J. J., et al. (2013). Phenix/MolProbity hydrogen parameter update. *The Computational Crystallography Newsletter*, *4*, 9–10.

DiMaio, F., Tyka, M., Baker, M., Chiu, W., & Baker, D. (2009). Refinement of protein structures into low-resolution density maps using Rosetta. *Journal of Molecular Biology*, *392*, 181–190.

Duarte, C. M., & Pyle, A. M. (1998). Stepping through an RNA structure: A novel approach to conformational analysis. *Journal of Molecular Biology*, *284*, 1465–1478.

Dunkle, J. A., Wang, L., Feldman, M. B., Pulk, A., Chen, V. B., Kapral, G. J., et al. (2011). Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science*, *332*, 981–984, 4GD1,2, 3R8S,T.

Echols, N., Grosse-Kunstleve, R. W., Afonine, P. V., Bunkoczi, G., Chen, V. B., Headd, J. J., et al. (2012). Graphical tools for macromolecular crystallography in Phenix. *Journal of Applied Crystallography*, *45*, 581–586.

Echols, N., Morshed, N., Afonine, P. V., McCoy, A. J., Miller, M. D., Read, R. J., et al. (2014). Automated identification of elemental ions in macromolecular crystal structures. *Acta Crystallographica*, *D70*, 1104–1114.

Edwards, T. E., & Ferré-D'Amaré, A. R. (2006). Crystal structures of the Thi-Box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure*, *14*, 1459–1468, 2HOJ.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica*, *D66*, 486–501.

Ferré-D'Amaré, A. R., & Scott, W. G. (2010). Small self-cleaving ribozymes. *Cold Spring Harbor Perspectives in Biology*, *2*, a003574.

Freisz, S., Lang, K., Micura, R., Dumas, P., & Ennifar, E. (2008). Binding of aminoglycoside antibiotics to the duplex form of the HIV-1 genomic RNA dimerization initiation site. *Angewandte Chemie (International Edition in English)*, *47*, 4110–4113, 3C3Z.

Gore, S., Velankar, S., & Kleywegt, G. J. (2012). Implementing an X-ray validation pipeline for the Protein Data Bank. *Acta Crystallographica*, *D68*, 478–483.

Grazulis, S., Chateigner, D., Downs, R. T., Yokochi, A. T., Quiros, M., Lutterotti, L., et al. (2009). Crystallography Open Database—An open-access collection of crystal structures. *Journal of Applied Crystallography*, *42*, 726–729.

Headd, J., & Richardson, J. (2013). Fitting Tips #5: What's with water? *The Computational Crystallography Newsletter*, *4*, 2–5.

Janssen, B. J. C., Read, R. J., Brunger, A. T., & Gros, P. (2007). Crystallographic evidence for deviating C3b structure? *Nature*, *448*, E1–E2.

Jones, T. A., Zou, J. Y., Cowan, S. W., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica*, *A47*, 110–119.

Joosten, R. P., Joosten, K., Murshudov, G. N., & Perrakis, A. (2012). PDB_REDO: Cobstructive validation, more than just looking for errors. *Acta Crystallographica*, *D68*, 484–496.

Kapral, G. J., Jain, S., Noeske, J., Doudna, J. A., Richardson, D. C., & Richardson, J. S. (2014). New tools provide a second look at HDV ribozyme structure, dynamics, and cleavage. *Nucleic Acids Research*, *42*, 12833–12846, 4PR6, 4PRF.

Ke, A., Zhou, K., Ding, F., Cate, J. H. D., & Doudna, J. A. (2004). A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature*, *429*, 201–205, 1VC7.

Keating, K. S., & Pyle, A. M. (2012). RCrane: Semi-automated RNA model building. *Acta Crystallographica*, *D68*, 985–995.

Klein, D. J., Schmeing, T. M., Moore, P. B., & Steitz, T. A. (2001). The kink-turn: A new RNA secondary structure motif. *The EMBO Journal*, *20*, 4214–4221.

Kleywegt, G. J., Harris, M. R., Zou, J. Y., Taylor, T. C., Wahlby, A., & Jones, T. A. (2004). The Uppsala Electron Density Server. *Acta Crystallographica*, *D60*, 2240–2249.

Lang, K., Erlacher, M., Wilson, D. N., Micura, R., & Polacek, N. (2008). The role of 23S ribosomal RNA residue A2451 in peptide bond synthesis revealed by atomic mutagenesis. *Chemistry & Biology*, *15*, 485–492.

Leontis, N. B., & Westhof, E. (2001). Geometric nomenclature and classification of RNA base pairs. *RNA*, *7*, 499–512.

Li, F., Pallan, P. S., Maier, M. A., Rajeev, K. G., Mathieu, S. L., Kreutz, C., et al. (2007). Crystal structure, stability, and in vitro RNAi activity of oligoribonucleotides containing the ribo-difluorotoluyl nucleotide: Insights into substrate requirements by the human RISC Ago2 enzyme. *Nucleic Acids Research*, *35*, 6424–6438, 2Q1O.

Moriarty, N. W., Grosse-Kunstleve, R. W., & Adams, P. D. (2009). electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallographica*, *D65*, 1074–1080.

Murray, L. W., Arendall, W. B., III, Richardson, D. C., & Richardson, J. S. (2003). RNA backbone is rotameric. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 13904–13909.

Parisien, M., & Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, *452*, 51–55.

Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica*, *A42*, 140–149.

Read, R. J., Adams, P. D., Arendall, W. B., III, Brunger, A. T., Emsley, P., Joosten, R. P., et al. (2011). A new generation of crystallographic validation tools for the protein data bank. *Structure*, *19*, 1395–1412.

Richardson, D. C., & Richardson, J. S. (2001). MAGE, PROBE, and Kinemages. In M. G. Rossmann & E. Arnold (Eds.), *Crystallography of biological macromolecules: Vol. F. IUCr's international tables of crystallography* (pp. 727–730). Dortrecht: Kluwer Academic Press, (chapter 25.2.8).

Richardson, J. S., & Richardson, D. C. (2013). Doing molecular biophysics: Finding, naming, and picturing signal within complexity. *Annual Review of Biophysics*, *42*, 1–28.

Richardson, J. S., Schneider, B., Murray, L. W., Kapral, G. J., Immormino, R. M., Headd, J. J., et al. (2008). RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA*, *14*, 465–481.

Saenger, W. (1983). *Principles of nucleic acid structure*. New York: Springer.

Sekine, S., Nureki, O., Dubois, D. Y., Bernier, S., Chenevert, R., Lapointe, J., et al. (2003). ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *The EMBO Journal*, *22*, 676–688, 1N78.

Stombaugh, J., Zirbel, C. L., Westhof, E., & Leontis, N. B. (2009). Frequency and isostericity of RNA base pairs. *Nucleic Acids Research*, *37*, 2294–2312.

Svozil, D., Kalina, J., Omelka, M., & Schneider, B. (2008). DNA conformations and their sequence preferences. *Nucleic Acids Research*, *36*, 3690–3706.

Wang, X., Kapral, G. J., Murray, L. W., Richardson, D. C., Richardson, J. S., & Snoeyink, J. (2008). RNABC: Forward kinematics to reduce all-atom steric clashes in RNA backbone. *Journal of Mathematical Biology*, *56*, 253–278.

Warner, K. D., Chen, M. C., Song, W., Straek, R. L., Thorn, A., Jaffery, S. R., & Ferre-D'Amare, A. R. (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature Structural & Molecular Biology*, *21*, 658–663.

Williamson, J. R. (2000). Induced fit in RNA-protein recognition. *Nature Structural Biology*, *7*, 834–837.

Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., et al. (1999). Visualizing and quantifying molecular goodness-of-fit: Small-probe contact dots with explicit hydrogen atoms. *Journal of Molecular Biology*, *285*, 1711–1733.

Word, J. M., Lovell, S. C., Richardson, J. S., & Richardson, D. C. (1999). Asparagine and glutamine: Using hydrogen atom contacts in the choice of sidechain amide orientation. *Journal of Molecular Biology*, *285*, 1735–1747.

Yeates, T. O. (1997). Detecting and overcoming crystal twinning. *Methods in Enzymology*, *276*, 344–358.

Zaher, H. S., Shaw, J. J., Strobel, S. A., & Green, R. (2011). The 2′-OH group of the peptidyl-tRNA stabilizes an active conformation of the ribosomal PTC. *The EMBO Journal*, *30*, 2445–2453.

Zwart, P. H., Grosse-Kunstleve, R. W., & Adams, P. D. (2005). Xtriage and Fest: Automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsletter*, *43*, Contribution 7.