



Published in final edited form as:

Crystals (Basel). 2021 August ; 11(8): . doi:10.3390/cryst11080952.

Engineering Crystal Packing in RNA Structures I: Past and Future Strategies for Engineering RNA Packing in Crystals

Narsimha Pujari¹, Stephanie L. Saundh¹, Francis A. Acquah², Blaine H. M. Mooers^{2,3}, Adrian R. Ferré-D'Amaré⁴, Adelaine Kwun-Wai Leung^{1,*}

¹Department of Veterinary Biomedical Sciences, University of Saskatchewan, Saskatoon, SK S7N 5B4, Canada

²Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

³Stephenson Cancer Center, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

⁴Biochemistry and Biophysics Center, National Heart, Lung and Blood Institute, Bethesda, MD 20892, USA

Abstract

X-ray crystallography remains a powerful method to gain atomistic insights into the catalytic and regulatory functions of RNA molecules. However, the technique requires the preparation of diffraction-quality crystals. This is often a resource- and time-consuming venture because RNA crystallization is hindered by the conformational heterogeneity of RNA, as well as the limited opportunities for stereospecific intermolecular interactions between RNA molecules. The limited success at crystallization explains in part the smaller number of RNA-only structures in the Protein Data Bank. Several approaches have been developed to aid the formation of well-ordered RNA crystals. The majority of these are construct-engineering techniques that aim to introduce crystal contacts to favor the formation of well-diffracting crystals. A typical example is the insertion of tetraloop–tetraloop receptor pairs into non-essential RNA segments to promote intermolecular association. Other methods of promoting crystallization involve chaperones and crystallization-friendly molecules that increase RNA stability and improve crystal packing. In this review, we discuss the various techniques that have been successfully used to facilitate crystal packing of RNA molecules, recent advances in construct engineering, and directions for future research in this vital aspect of RNA crystallography.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

*Correspondence: adelaine.leung@usask.ca.

Author Contributions: Original draft preparation, N.P., S.L.S., F.A.A., B.H.M.M., A.K.-W.L. review and editing A.R.F.-D. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Keywords

ribonucleic acid; RNA; X-ray crystallography; crystal packing design; RNA motifs; nucleic acid crystallography; RNA crystallization constructs; RNA crystallography; RNA crystallogenesis

1. Introduction

Structural biology of RNA molecules began in the 1960s with the stepwise resolution improvement of the yeast tRNA^{Phe} crystal structures [1–3]. The cloverleaf structure of the tRNA is likely the first introduction to the RNA structure encountered by many science students. The tRNA crystal structures unveiled how specific tertiary interactions maintain the three-dimensional fold of the RNA molecule, and how the three-dimensional architecture enables its role in protein translation. The next breakthrough came in the crystal structures of the ribozymes [4–6] and self-splicing introns [7], which demonstrated how RNA can fold into an architecture capable of carrying out enzymatic catalysis. These early RNA crystal structures set the stage for the structural investigation of large RNA-containing macromolecular complexes such as the ribosome and the spliceosome that are key players in the central dogma of molecular biology [8–10]. The millennium began with the discovery of many more types of non-coding RNAs (e.g., microRNAs, riboswitches, lncRNAs; see [11] for a comprehensive review). X-ray crystallography has remained a critical experimental approach to understanding the molecular mechanisms enabling these non-coding RNAs to perform extraordinarily diverse biological functions.

Crystallization of RNA molecules for structural determination and molecular interaction elucidation is often more challenging than the crystallization of soluble proteins [12,13]. This in part explains why RNA-only structures account for less than 0.7% of the biomolecular crystal structures deposited in the Protein Data Bank (PDB). As an equilibrium process, crystallization is governed by both the properties of the soluble molecule and the nature of its crystalline state. Although RNA duplexes are highly stable thermodynamically, the higher-order folding landscape of RNA is often complex, with multiple competing minima, as well as kinetic traps [14–17]. This results in conformational heterogeneity in the solution [18]. The molecular surface of folded RNAs is dominated by a regular array of negatively charged phosphates that can lead to packing (through counterions) into crystals that are disordered at the atomic level [19–21]. Strategies for improving conformational homogeneity of folded RNAs are generally useful for any biophysical or structural study of this biomolecule, while techniques for enhancing the formation of improved crystal contacts are often necessary for the successful application of X-ray, neutron, or electron crystallography [12,22–26].

The idea of engineering nucleic acid sequence to promote crystal contacts originated in the work of Schultz et al., where they successfully crystallized the *E. coli* catabolite gene activator protein complexed with its DNA-binding site after scanning through 26 different DNA sequences [27]. Over the years, several molecular engineering techniques have been developed to favor ordered crystal packing of RNAs (e.g., [28–30]). Furthermore, optimized crystallization conditions and sometimes post-crystallization treatments have also

been crucial for success in RNA crystallography [31,32]. In this review, we highlight new engineering strategies for RNA crystallization, as well as techniques that have proved successful in aiding the growth of diffraction quality crystals in the past [33]. In addition, we emphasize considerations that may affect their adoption and use in different RNA contexts beyond where they were initially developed [34]. This review may be of interest to nucleic acid crystallographers who seek to expand their repertoire of engineering tools to tackle the crystallization of challenging RNA targets.

2. The Propensity of RNA Helices to Form Intermolecular Stacks

The simplest RNA interactions are formed by Watson-Crick (WC) base paired duplexes. The duplex can form in *cis* between adjacent complementary sequences generating a loop structure or between two complementary strands far away in the primary sequence but brought together through more complicated intermediate RNA structures. Duplexes can also form between two RNA strands binding in *trans*, which is common in trypanosome RNA editing, RNA viruses, and RNA interference. The stacking of terminal base pairs between RNA helices is a common packing scheme observed in many crystals. Therefore, engineering helical ends to encourage inter-helical stacking is an excellent strategy to promote or improve crystal packing (for historical context, see, e.g., [35]). The length of the duplex may be the primary determinant of the angle between the long axes of the terminal base pairs at the junction between two stacked duplexes. This angle can be positive or negative, and the helical axes of the duplexes can be aligned or offset. When the axes are offset, the ribose ring of one terminal nucleotide may pack against the terminal base pair of the adjacent duplex and be correlated with buckling of the terminal base pair. AU base pairs may tolerate this buckling better than GC base pairs, so the identity of the terminal base pairs and chain length may be important determinants of helical stacking. Note that the effects of AU and UA base pairs in the terminal positions are probably not equivalent. The inclusion of terminal phosphates [30] and overhanging nucleotides can also influence success at crystallization. An example of the latter was illustrated in structural studies of the minimal hairpin ribozyme [25]. Packing analysis of a crystal that diffracted to 3.17 Å resolution revealed suboptimal intermolecular helical stacking at the crystal packing interface. The RNA construct was subsequently re-designed to have a single U “sticky” end that permitted a more optimal helical stacking and crystalline order, thereby increasing the diffraction limit to 2.05 Å.

There are three notes of caution when designing constructs to study RNA duplexes. First, the stability of duplex stacking can often lead to the crystallization of duplexes other than those intended. This result is most problematic when other conformations of RNA, such as stem-loops, are the subject of study, but crystallization selects non-desirable duplex forms [36,37]. Second, when duplexes are formed from two different RNA strands, the heteroduplexes are prone to an end-on-end disorder that can promote crystal twinning [20]. Because the phosphate backbone is symmetrical while the base pairings are not, the helices do not have to be packed consistently in the same orientation. A twinned data set will occur if a fraction of the helices is flipped while maintaining the same inter-helical packing. To avoid this crystal packing artifact, it is advisable to fuse two copies of the duplex as has been done successfully for fragments of the U-tail:pre-mRNA duplex from a trypanosome

RNA editing substrate [38,39]. The substrates were 8- and 16-base pairs long and formed duplexes with 16- and 32-base pairs, respectively. The longer construct actually gave 1.05 Å diffraction data suitable for ab initio structure determination by direct methods despite the presence of translational pseudosymmetry due to the three helical turns in the RNA duplex [40]. Third, because crystallization is only compatible with 1-, 2-, 3-, 4-, and 6-fold rotations [41], when there are duplexes whose helical pitches are not compatible with this requirement (i.e., their length does not correspond to a multiple of a full, 1/2, 1/3, 1/4, or 1/6th helical turn) crystals can form where the molecule is incommensurate with the crystallographic unit cell. An example is the crystallization of an 18-mer A-form duplex that resulted in the RNA packing in 36 different registers; any of its residues occupied any one helical position in the crystal. That is, the bases in the structure were disordered 36-fold, while the sugar-phosphate backbone of the RNA adopted apparently identical (at 1.2 Å-resolution) conformations at all positions [19]. This kind of disorder, a special form of twinning, is readily apparent in the distribution of scattering factor intensities [21].

It is not easy to predict accurately whether a short RNA will crystallize as a hairpin or a duplex. Many of the early crystal structures of RNA duplexes about one helical turn in length had mismatch bases in the middle with flanking Watson–Crick base pairs, for example, the RNA dodecamer duplex (r-GGACUUCGGUCC)₂ [42]. These RNA duplexes were failed attempts at crystallizing RNA hairpins with tetraloops. The high concentration of RNA in the crystallization drop and the stable crystalline packing of the duplex drove disproportionation. The undesired duplex formation was favored in the crystallization of a 27-nt RNA that was crystallized under solution conditions that favored hairpin formation [43]. The conformation of single-stranded RNAs is easily checked by native PAGE with the appropriate control RNA species of known duplex or hairpin conformation, but the oligomerization state of an RNA in solution may differ from that favored by crystallization [44,45].

3. Hairpin Loops and Their Utility in Crystal Packing Design

When adjacent complementary sequences hybridize to form an RNA duplex, a hairpin or stem loop structure is formed. Proteins can bind some loops (e.g., spliceosomal proteins U1A [46], U2A 'B' [47], the pseudouridine synthase TruB [48]). Other loops can serve as folding nucleation sites, confer stability, and participate in tertiary RNA interactions (e.g., the UUCG tetraloop [49]). Tetraloops and kissing loops are the most common hairpin loops that engage in tertiary RNA interactions. Replacing flexible structures between complementary sequences by tetraloops or kissing loops is a common strategy to promote the formation of well-packed crystals by removing undesirable elements that hinder crystallization and generating potential interfaces for crystal packing. We review several specific examples below to illustrate these principles in action.

3.1. Promoting Loop-Loop Crystal Contacts: The Kissing Loop Complex

The kissing loop is an RNA tertiary interaction motif that can be used for designing sequence-specific loop-loop contacts. Found in natural RNAs, the kissing loop formation is essential for the regulation of the ColE1 plasmid replication [50], and the dimerization

of the viral genome that occurs during the life cycle of HIV-1 [51] and the Hepatitis C virus [52]. The kissing loop complex is formed by Watson–Crick base pairing between the complementary palindromic sequences of the two unpaired stem loops. The first kissing loop complex was visualized in the crystal structure of the yeast tRNA^{ASP} where its anti-codon forms a 3-nt kissing loop complex with the anti-codon of a symmetry-related mate [53]. Since then, kissing loops with variable lengths have been observed in other structures. For example, the H3 stem loop of the Moloney murine leukemia virus forms a homodimeric kissing loop complex with just two intermolecular GC base pairs [54] (Figure 1A-i). On the other hand, the RNA I/II kissing loop complex from the ColE1 plasmid is formed between 7-nts from the two RNA stem loops [55,56] (Figure 1A-ii). The versatility of the kissing loop tertiary motif was exploited in the crystallization of the spliceosomal U1 snRNP [57–59].

3.2. Promoting Loop to Stem Crystal Contacts: Loop to Receptor Motifs

3.2.1. Tetraloop-Tetraloop Receptor Motifs—A tetraloop is a 4-nt loop that caps a stem. The GNRA and the UNCG (N = any nucleotides, R = either purine) families are the most common types found in natural RNAs. The GNRA tetraloop typically adopts a U-turn conformation, characterized by the first nucleobase forming a hydrogen bond with the backbone of the fourth nucleotide. This base to backbone interaction positions the first nucleobase to form a stacking (π -OP) interaction with the oxygen of the phosphate backbone of the third nucleotide. The UNCG tetraloop typically adopts a Z-turn conformation, characterized by the first and fourth nucleotides forming a *trans* Sugar/WC interaction and an unusual ribose backbone conformation that allows the formation of a stacking (π -O4') interaction between the fourth nucleobase and the ribose oxygen of the third nucleotide [64]. The GNRA tetraloops are more commonly found to engage in tertiary interactions. In contrast, the UNCG tetraloop has exceptional thermostability, and it is typically used to replace flexible regions undesirable for crystal packing. The thermal stability of the UNCG tetraloop is mainly contributed by hydrogen bonds between the 2'-hydroxyl groups of the first and second sugar moiety with the Hoogsteen edge of the base of the fourth nucleotide [65,66]. Interestingly, the core of the four-way junction is recently identified as a receptor for a UNCG tetraloop [67]. However, a four-way junction is not easy to implement in other structural contexts. We will consider tetraloops with their cognate receptors that are convenient to incorporate in different RNA structures to promote loop to stem contacts.

The GAAA Loop and Its 11-nts Receptors (GAAA-R): One of the most common GNRA tetraloops, the GAAA tetraloop, was first observed to interact with the ribose moiety in the minor groove of an A-form helix without much sequence specificity [5,68]. However, its long-range tertiary interaction with a sequence-specific region on the self-splicing group I intron was deduced and characterized by chemical probing [69] and mutagenesis [70]. The 11-nt sequence motif was identified as the GAAA receptor (hereafter, GAAA-R) by co-variation analysis [71]. The detailed interaction between the GAAA tetraloop and the 11-nt receptor (GAAA-R) was finally revealed in the crystal structure of the P4-P6 domain of the group I self-splicing intron [7]. In vitro selection of different types of GNRA tetraloops has identified other receptor sequences, showcasing the versatility of the tetraloop/receptor

motif [72,73]. The GAAA-R motif can be engineered anywhere on an RNA helix, thus providing a loop to duplex interaction design option (Figure 1C-i). The GAAA/GAAA-R interaction is used in a popular strategy [29] to promote RNA crystal formation and has found numerous applications, for instance, to promote crystallization of the spliceosomal U4 snRNP core domain [59,74,75]. Crystallization of the Ribonuclease P holoenzyme in complex with tRNA^{Phe} was achieved by introducing the GAAA tetraloop into the RNase P and the GAAA-R to the tRNA [76]. More recently, the GAAA and GAAA-R were engineered to facilitate the crystallization of a short RNA helix with CUG repeats, which occur in myotonic dystrophy when CTG in the 3' UTR of a gene is abnormally expanded. The transcribed CUG repeats can form a stable A-form RNA duplex that sequesters RNA binding proteins and thereby interferes with the normal functions of the sequestered proteins. The incorporation of GAAA/GAAA-R into the CUG repeat duplex has an added advantage in that it created a large solvent-accessible space within the crystal. As a result, the crystal form could become a convenient platform to screen potential molecules that can target the CUG repeat duplex for diagnostic or therapeutic purposes [28].

The GAAC Loop and Its 20-nts Receptor (GAAC-R): The GAAC is a non-GNRA tetraloop found in group II introns and has a folding geometry distinct from the GNRA fold [77,78]. Instead of a *trans* Sugar/Hoogsteen base pair between the first and fourth nucleotide, a *trans* Sugar/WC base pair was observed. As a result, the first base of the loop no longer stacks with the 5' WC base pair of the stem. A modular 20-nt receptor was identified by in vitro selection with a binding affinity (kD ~ 2.4 nM) comparable to that of the GAAA/GAAA-R interaction [79] (Figure 1C-ii). The GAAC/GAAC-R loop to stem interaction may have a different orientation compared to that formed by GAAA/GAAA-R. This is based on the relative catalytic efficiency observed in the in vitro selection system. The tetraloop and its receptor are engineered in a ribozyme structural fold, and the ribozyme catalysis depends on their interaction. Although the GAAC/GAAC-R motif resulted in catalysis, the efficiency was ~30% less than that observed for the GAAA/GAAA-R motif [80]. The authors attributed the catalytic difference to a variation in the orientation of the interaction motif. Thus, the GAAC/GAAC-R motif could create alternative packing geometry compared to the GAAA/GAAA-R motif. Unfortunately, no structural confirmation of the interaction is available to date. Nonetheless, a chemical probing experiment suggests that the WC edges of the middle two As interact with the receptor [79].

3.2.2. The C-Loop and Its 20-nts Receptor (C-loop-R)—The C-loop is a recurrent motif characterized by an asymmetric internal loop. As observed in ribosome structures, the main structural function of the C-loop is to increase the helical twist of the stem loop where it is embedded so that the hairpin loop can engage in optimal tertiary interaction [80]. Although it does not naturally participate in tertiary interaction, a 20-nts loop receptor (C-loop-R) in the form of a loop was identified by in vitro selection [81] (Figure 1C-iii). Therefore, a possible crystal packing design incorporates a C-loop motif into a stem, and the C-loop-R could be inserted as a stem loop. Unfortunately, no structural information of the interaction is available. However, chemical probing experiments suggest that the C-loop motif binds C-loop-R via non-WC interactions [81].

4. Designing Lateral Contacts between Duplexes

Terminal base pair stacking of intermolecular duplexes and loop-loop interaction of the kissing loops will promote crystal contact along the same direction of the helical axis. Lateral packing of helices frequently involves ribose zipper motifs that utilize two consecutive base pairs from two regions of the RNA stem running in an anti-parallel fashion. Different types of ribose zippers are characterized by their base-backbone and base-base interaction patterns in the minor grooves [82,83]. Although there is a certain degree of sequence specificity for the ribose zipper, its variability makes it difficult to use as a general motif for crystal packing design. In this section, we will discuss several design options that can promote lateral interaction of RNA helices.

4.1. Kissing Loop with Two Bulged Purines

In addition to the typical kissing loop complex introduced earlier, an unusual kissing loop complex with two bulged purines was observed in an RNA duplex fragment that contains the HIV-1 genome dimerization initiation site (DIS). The DIS contains a 9-nt loop with two conserved purines at the 5' end, followed by a 6-nt sequence that makes kissing loop contacts and an invariant adenine at the 3' end (Figure 1B-i). The crystal structure of the DIS complex contains a 7-bp WC stem capped by the DIS loop. As expected, the two hairpins form a head-to-head pseudo-continuous coaxial A-form duplex via WC base pairing of the 6-nt complementary sequence. Interestingly, the two conserved purines 5' to the kissing sequence are flipped out to form crystal contacts lateral to the duplex by stacking with their symmetry-equivalent mates, while the 3' conserved adenine remains unpaired within the helix [60]. While the biological relevance of this unusual kissing complex is controversial [84,85], the same bulged out residues were observed in three different sequences, crystallization conditions, and space groups [60]. Therefore, this is likely a recurrent motif that could be useful if a coaxial duplex with lateral stabilization is desired in the crystal packing design. This kissing loop with two bulged purines was utilized to improve the diffraction quality of the U1 snRNP crystal [57,59].

4.2. Paromomycin Binding Motif

Another motif that can be used to stabilize duplexes laterally is the paromomycin motif. Paromomycin is an antibiotic that binds to the A site of the 16S rRNA of the *E. coli* ribosome. Its interaction causes structural changes that render the ribosome unable to discriminate near-cognate tRNA [86]. The crystal structure of an 18-bp helix with two 13-nt paromomycin binding sites incorporated between WC base-pairs shows that the binding of paromomycin flips out two adenines, which then form critical A-minor motif interactions with the neighboring backbones [61,62] (Figure 1B-ii). In an attempt to improve the diffraction quality of the spliceosomal U4 snRNP core domain, the paromomycin motif was incorporated into the U4 snRNA. Indeed, crystals in different space groups were obtained only in the presence of paromomycin [59,74]. However, the diffraction limit did not improve, highlighting the reality that improving packing geometry is much more difficult than promoting crystal formation.

5. Introducing Crystal Contact Tags

The ends of the RNA could be another accessible point for engineering crystal packing motifs. Here we will suggest two possibilities: the G-quadruplex and the three-way junction module.

5.1. G-Quadruplex

The G-quadruplex is a natural structure found in G-rich regions of nucleic acids (e.g., telomeric DNA and RNA). The quartet adopts a square and planar structure where four guanine bases form hydrogen-bonding interactions through their WC and Hoogsteen edges (the structures of RNA G-quadruplexes have been reviewed recently [87]). In the crystal structure of the σ -subunit of RNA polymerase in complex with a 10-nt ssDNA, the 3' terminal GGG flips out into the solvent to form a pseudo-continuous G-quadruplex column, which forms the main crystal lattice [88]. At the crystal packing interface, a string of 3 G's flipped out from each complex to form three sets of *trans* Hoogsteen/WC base pairs with one symmetry-related complex and *trans* WC/Hoogsteen base pairs with another symmetry-related complex. In theory, a short single-stranded GGG RNA extension engineered at either end of the RNA structure could form a similar G-quadruplex crystal contact, potentially generating a four-fold symmetry (Figure 1D-i).

5.2. The Three-Way Junction (3WJ) Building Block

Recurrent RNA 3D motifs observed in RNA crystal structures have been utilized in RNA nanotechnology to build self-assembling platforms for various biomedical applications [89–92]. The self-assembly of RNA nanoparticles requires building blocks that can form long-range interactions, which is the same prerequisite as RNA modules that promote crystallization. The field of RNA nanotechnology had been exploiting RNA tertiary binding motifs that occur in natural RNAs to develop interesting ways to form symmetrical polygons known as origami [93]. These polygons can be engineered to the ends of the RNA to promote crystal packing, similar to what was envisioned by the pioneer of DNA nanotechnology, Nadrian Seeman, who first began to rationally design 3D DNA crystals with self-assembling oligonucleotides [94].

The three-way junction (3WJ) observed in biological RNAs has been developed and utilized as a versatile building block in RNA nanotechnology. The 3WJ connects three WC base-paired helices with three 1–3 nts single-stranded segments. Interestingly, the 3WJ motif can be assembled readily from three short RNA oligonucleotides in water at room temperature and the resulting complex is stable in 8 M urea (Figure 1D-ii) [95]. The lengths of the H1, H2, and H3 helices can be shortened to 6, 8, and 6 base pairs without affecting complex formation. The ability to self-assemble into a stable complex from short oligonucleotides makes this an attractive motif for engineering crystal contacts of the RNA structure. One possibility is to engineer the two strands (a (18nt) and c (16 nt)) into the 5' and 3' tail of the RNA structure, and the third strand (b (20nt)) can be added before crystallization. Thus, the tails of the RNA can contact each other in a head-to-tail manner, and the synthetic oligonucleotide can be used for fine-tuning the packing geometry. Because the orientation of the helices joined by 3WJ is dictated by the number of unpaired nucleotides on strand b, the

packing geometry could be adjusted by changing the number of nts on the single-stranded segment of the synthetic oligonucleotide rather than reengineering the whole RNA construct [33].

6. Introducing RNA Binding Proteins

The molecular surface of RNA is dominated by repetitive anionic backbone phosphates. The limited chemical diversity and charge repulsion reduce the chances that the RNA molecule can pack in a unique and consistent manner. RNA also tends to be more conformationally heterogeneous. These factors impede the crystallizability of RNAs compared to proteins. To promote favorable crystal packing of RNAs, one strategy is to introduce peptides or proteins to bridge the RNA molecules.

6.1. RNA Binding Protein—U1A

The spliceosomal protein U1A is the first and most widely used protein module for crystallizing RNAs. The N-terminal 98-residue RNA recognition motif of U1A (U1A-RRM) recognizes stem loop II of the U1 snRNA. The 1.9 Å crystal structure of U1A bound to a fragment of stem loop II of U1 snRNA shows how the U1A-RRM interacts with the 10 unpaired nucleotides (AUUGCACUCC) on the stem loop [46]. By engineering the 10-nt U1A binding loop onto a functionally dispensable RNA stem and incorporating the mutant U1A-RBD previously engineered to improve the surface properties for crystal packing, a multitude of RNA structures have been determined. A detailed review of successful cases using the U1A module has been documented previously [96]. More recently, the U1A module was used to promote crystal packing of a fusion construct of U170K bound to the stem loop I of U1 snRNA. Like U1A, U170K also has an RRM, but it uses a different binding mechanism to interact with its stem loop [57,59].

6.2. Antibody Fragment

Fragments of antibody (Fabs) selected to bind RNA by phage display have been utilized to create an interface for crystal contacts. Compared to U1A, Fabs are more efficient at creating crystal contacts as they are larger (50 kDa compared to 11 kDa) and thus have more surface area available for potential crystal contacts [97]. In addition, the β -sheets in antibody structure are predisposed to make good crystal contacts [98]. Fab-assisted crystallization was successfully applied to the crystal structure of the P4-P6 domain of the Tetrahymena group I intron [99] and the artificial ribozyme, the class I ligase [97]. Interestingly, the latter structure revealed that the stem loop sequence GAAACAC is the minimal antigen for the Fab fragment developed in the study [97]. Thus, this loop sequence and its Fab fragment can be exploited as a general crystallization tag for any RNA structure. The loop can be engineered on top of any A-form helix, and crystallization can be attempted with the same antibody fragment, which can be expressed and purified efficiently from *E. coli* culture [97]. In addition to promoting crystal packing, the Fab fragment can facilitate phase determination by molecular replacement and Cryo-EM map interpretation.

6.3. Peptide Nucleic Acid

If introducing a large artificial protein is undesirable, one could consider utilizing a synthetic molecule to form specific WC-like base pairs with the RNA. Peptide nucleic acid (PNA) is a synthetic molecule comprising of repeating N-(2-aminoethyl) glycine units, essentially resembling a peptide backbone attached to nucleobases [100]. PNA was developed into potential therapies that target abnormal nucleic acids. However, because PNA has an overall neutral charge that eliminates the intrinsic electrostatic repulsion of the phosphate backbone, it could facilitate the crystallization of RNA molecules. PNA synthesized with complementary bases can hybridize to an RNA to form a PNA/RNA duplex with slight geometric deviation from an RNA A-form helix [101]. The PNA backbone can be further modified to enhance rigidity and incorporate other functional groups [102]. Although PNA has not been utilized as a crystallization tool yet, it may be worth exploring in parallel to the crystal contact tag strategy suggested earlier.

7. Post-Crystallization Treatment

When incorporating interacting modules to promote crystal packing, it is important to try to place them along with several WC bp insertions to fine-tune the orientation. Each bp insertion will add a rotation of 32.7° and a rise of 2.6 \AA along the helical axis [103]. If a crystal is obtained but diffraction is poor, it is vital to not give up on the construct before sufficient optimization is performed. First, the crystal needs to be of sufficient size. For the $\sim 100 \text{ kDa}$ U4 snRNP core domain crystal, the best diffraction data ($\sim 3.6 \text{ \AA}$) were not obtained unless the crystal was at least $300\text{--}400 \text{ \mu m}$ in size [74]. Also, it is important to perform diffraction tests at room temperature to ensure that the cryoprotection protocol is not damaging the crystal [104]. As with protein crystals [105], controlled dehydration of RNA crystals can improve diffraction limits (e.g., dehydration of *glmS* ribozyme-riboswitch crystals improved diffraction from $\sim 3.0 \text{ \AA}$ to 1.7 \AA) [106]. Finally, cation replacement coupled with dehydration should be attempted to promote repacking of the RNA [31]. Replacing 20 mM MgCl_2 with 40 mM SrCl_2 in the presence of a high concentration of precipitant successfully improved the diffraction limit from ~ 8 to $\sim 3 \text{ \AA}$ of the crystals of the ternary complex of the Stem I domain of a T-box riboswitch, its cognate tRNA, and the RNA-binding protein YbxF [107]. Structural analysis showed that the more flexibly coordinated Sr^+ metal cations allowed the RNA components to shift as rigid bodies and repack more optimally, thereby contributing to the remarkable improvement in diffraction quality [31]. Barium has a high coordination number like strontium and may be a suitable alternative. Finally, if all else fails, determining a low to medium resolution structure to understand the packing could inspire new designs to improve crystal packing.

8. Future Directions of RNA Crystallography

When prior knowledge crystallization modules and rationale design approaches fail, a stochastic process might succeed. One example is the exciting “in crystallo” selection in which error-prone PCR generates a pool of 10 million mutant DNA templates (each DNA containing 0–2 mutants). Next, the templates are transcribed with Phage T7 RNA polymerase, gel purified, and then used in crystallization experiments using conditions

known to give crystals of the wild-type RNA [13]. The largest crystals are isolated, dissolved, reverse transcribed, and sequenced. In principle, the selection process is repeated for several cycles to obtain better crystals. When the second round of selection was applied to the P4-P6 domain of Tetrahymena, smaller crystals were obtained, so the selection process was abandoned. The P4-P6 domain crystallizes with two molecules in the asymmetric unit. In the first cycle, wild-type molecules were present; they may have acted as crystallization chaperones for the mutant molecules because a wild-type molecule paired with a mutant molecule in the asymmetric unit. The mutant RNAs were made and crystallized individually, and structures were determined from four of the mutants. Some of the mutant RNAs gave inferior diffraction in the absence of the wild-type RNA. The authors suggest that the selection process may be more stringent in selecting mutants that give crystals when only one RNA molecule can occupy the asymmetric unit. The authors did not explore the combination of two or more mutations to determine if the multi-mutant RNA could give crystals that diffract to a higher resolution than the wild-type RNA. The authors found improvement in the electron density maps around the mutant sites compared to the wild-type structure by local structural rearrangements in loops and new intermolecular contacts [13]. Their results suggested that bulged residues could be walked along the chain in either direction without disrupting the core structure. They also suggest that “bulge engineering” could be applied to any unpaired surface regions. The “in crystallo” selection experiments lead to the discovery of one potentially universal RNA crystal engineering tactic, and it is reasonable to expect that additional principles will be suggested by the structures of other mutant RNAs from future “in crystallo” selection experiments. While this study failed to improve diffraction quality, it demonstrated that “in crystallo” selection can be applied to RNA, and it opened a new approach to promoting RNA crystallization.

Another under-exploited area is the enhancement of the RNA structure stability by introducing X bonds between halogen atoms in bases and the backbone oxygen atoms. These interactions have been better characterized in DNA [108]. Bromine atoms are routinely introduced in synthetic RNAs to obtain experimental phases for structure determination. The position of the halogenated base within an RNA fragment can shift the equilibrium between duplex formation and hairpin formation, so several constructs may have to be tested [109]. Constructs could be prescreened for their conformation by native gel electrophoresis.

Another rapidly emerging structure determination technique of possible relevance to RNA crystallography is microcrystal electron diffraction (MicroED). These experiments are conducted with a transmission electron microscope in diffraction mode. The advantage of this method is that high-resolution diffraction data can often be obtained from nanocrystals (at least one dimension smaller than 100 nm) that are a billionth of the volume considered suitable for X-ray diffraction studies [110]. The disadvantages are the very limited access to instruments that can rotate the sample during data collection. The early successes with peptides and small proteins suggest that it might work well with smaller RNAs. Further technological advancements may be required for success with crystals of larger RNAs.

9. Future Relevance of Engineering Crystal Packing in RNA Structures

With the remarkable advancement in structural prediction algorithms and Cryo-EM, one may ask if there is still relevance in designing crystal contacts for nucleic acid structure determination. While the accuracy of AlphaFold2 is remarkable for predicting protein structures [111,112], it may still take some time for nucleic acids structure prediction due to the sparsity of structures available to train a neural network model for RNA structure prediction. Suppose the deep learning approach eventually succeeds at making accurate predictions of nucleic acid structure. In that case, can the deep learning models be harnessed to assist in engineering crystal contacts? Alternatively, other computational approaches from the protein re-design field, like the dead-end elimination theorem [113], could be used to perform the above-mentioned bulge walking in-silico in the presence of all of the surrounding symmetry mates. Such a computational approach to crystal lattice engineering should now be within reach thanks to the increased computational power available from GPUs. For instance, molecular docking of large libraries of small molecules against ensembles of protein structures is now practical thanks to GPUs. Ultimately, predicted structures will never replace the need for experimentally-determined structures designed with biological functions in mind.

With the recent “resolution revolution” in Cryo-EM, Cryo-EM has yielded several low-resolution structures of single-stranded RNAs as small as 40 kDa, suggesting that Cryo-EM will generally apply to large and medium-sized RNA structures [114,115]. It remains to be shown if such low-resolution structures are accurate enough to inform the design of constructs for crystallization to obtain high-resolution structures. Cryo-EM has great promise in the study of RNAs that have conformational heterogeneity, especially if the alternate conformations can be clustered in a modest number of conformations. With suitable clustering algorithms, most of the conformations of a small ensemble can be characterized from a single sample. Thus, Cryo-EM can give multiple structures from one sample. If the conformational heterogeneity cannot be parsed into distinct clusters, it can be addressed computationally at the expense of reduced resolution of the Cryo-EM map.

The crystal packing strategies described in this review can reduce the flexibility of the interacting regions. Some of these crystal packing modules generate symmetry, which should promote crystallization because proteins with molecular symmetry are known to crystallize more readily than those without molecular symmetry [116]. For example, the kissing loop complex generates two-fold symmetry, the G-quadruplex generates a four-fold symmetry, and the 3WJ junction has been further engineered to form a stable planar triangle, square, and pentagon using oligonucleotides [117]. These polygons are formed using different external 48-nt strands that form the side and one internal strand that base pairs with each external strand like a tape (Figure 2). The affinity of the strands to form the polygon is in the order of ~20 nM, and the formation is highly efficient. An RNA structure can potentially be engineered at the 3' end of each external strand, allowing the RNA to assemble into a higher-order complex with the polygon in the middle. The addition of these polygons could facilitate Cryo-EM studies by increasing the molecular weight of the RNA and overcoming preferred orientation. If the attached RNAs of interest have consistent

orientations with regards to the polygon, they can also provide a rotational symmetry during 3D reconstruction.

10. Conclusions

Over 98% of the human genome codes for noncoding RNAs [118]. In vivo genome-wide probing of RNA structures shows the dynamic structural characteristics of mRNA that correlate to cellular physiology [119]. Thus, there are likely many RNA structures with important functions waiting to be discovered. Nucleic acid crystallography will remain a vital structural approach in the years to come. Crystal packing design will continue to be an essential prerequisite for crystallization success and potentially also facilitate Cryo-EM investigations.

Acknowledgments:

A.R.F. thanks Kiyoshi Nagai (in memoriam) for encouraging the development of U1A as an RNA crystallization tool.

Funding:

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) (RGPIN-2014-06012) and internal funding support from the Western College of Medicine and College of Medicine, University of Saskatchewan, NIH R01 CA242845, NIH R01 AI088011, NIH P20 GM103640, NIH P30 CA225520, and the intramural program of the National Heart, Lung and Blood Institute, NIH.

References

1. Kim SH; Quigley G; Suddath FL; McPherson A; Sneden D; Kim JJ; Weinzierl J; Blattmann P; Rich A The three-dimensional structure of yeast phenylalanine transfer RNA: Shape of the molecule at 5.5-Å resolution. *Proc. Natl. Acad. Sci. USA* 1972, 69, 3746–3750. [PubMed: 4566458]
2. Kim SH; Quigley GJ; Suddath FL; McPherson A; Sneden D; Kim JJ; Weinzierl J; Rich A Three-dimensional structure of yeast phenylalanine transfer RNA: Folding of the polynucleotide chain. *Science* 1973, 179, 285–288. [PubMed: 4566654]
3. Kim SH; Suddath FL; Quigley GJ; McPherson A; Sussman JL; Wang AH; Seeman NC; Rich A Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 1974, 185, 435–440. [PubMed: 4601792]
4. Pley HW; Flaherty KM; McKay DB Three-dimensional structure of a hammerhead ribozyme. *Nature* 1994, 372, 68–74. [PubMed: 7969422]
5. Scott WG; Finch JT; Klug A The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* 1995, 81, 991–1002. [PubMed: 7541315]
6. Ferre-D'Amare AR; Zhou K; Doudna JA Crystal structure of a hepatitis delta virus ribozyme. *Nature* 1998, 395, 567–574. [PubMed: 9783582]
7. Cate JH; Gooding AR; Podell E; Zhou K; Golden BL; Kundrot CE; Cech TR; Doudna JA Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 1996, 273, 1678–1685. [PubMed: 8781224]
8. Kastner B; Will CL; Stark H; Luhrmann R Structural Insights into Nuclear pre-mRNA Splicing in Higher Eukaryotes. *Cold Spring Harb. Perspect. Biol* 2019, 11.
9. Plaschka C; Newman AJ; Nagai K Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. *Cold Spring Harb. Perspect. Biol* 2019, 11.
10. Yan C; Wan R; Shi Y Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. *Cold Spring Harb. Perspect. Biol* 2019, 11.
11. Cech TR; Steitz JA The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 2014, 157, 77–94. [PubMed: 24679528]

12. Ke A; Doudna JA Crystallization of RNA and RNA-protein complexes. *Methods* 2004, 34, 408–414. [PubMed: 15325657]
13. Shoffner GM; Wang R; Podell E; Cech TR; Guo F In Crystallo Selection to Establish New RNA Crystal Contacts. *Structure* 2018, 26, 1275–1283.e3. [PubMed: 29910185]
14. Russell R; Zhuang X; Babcock HP; Millett IS; Doniach S; Chu S; Herschlag D Exploring the folding landscape of a structured RNA. *Proc. Natl. Acad. Sci. USA* 2002, 99, 155–160. [PubMed: 11756689]
15. Chen SJ; Dill KA RNA folding energy landscapes. *Proc. Natl. Acad. Sci. USA* 2000, 97, 646–651. [PubMed: 10639133]
16. Cordero P; Das R Rich RNA Structure Landscapes Revealed by Mutate-and-Map Analysis. *PLoS Comput. Biol* 2015, 11, e1004473. [PubMed: 26566145]
17. Lynch DC; Schimmel PR Cooperative binding of magnesium to transfer ribonucleic acid studied by a fluorescent probe. *Biochemistry* 1974, 13, 1841–1852. [PubMed: 4601164]
18. Bernetti M; Hall KB; Bussi G Reweighting of molecular simulations with explicit-solvent SAXS restraints elucidates ion-dependent RNA ensembles. *Nucleic Acids Res.* 2021.
19. Klosterman PS; Shah SA; Steitz TA Crystal structures of two plasmid copy control related RNA duplexes: An 18 base pair duplex at 1.20 Å resolution and a 19 base pair duplex at 1.55 Å resolution. *Biochemistry* 1999, 38, 14784–14792. [PubMed: 10555960]
20. Mueller U; Muller YA; Herbst-Irmer R; Sprinzl M; Heinemann U Disorder and twin refinement of RNA heptamer double helices. *Acta Crystallogr. D Biol. Crystallogr* 1999, 55, 1405–1413. [PubMed: 10417408]
21. Shah SA; Brunger AT The 1.8 Å crystal structure of a statically disordered 17 base-pair RNA duplex: Principles of RNA crystal packing and its effect on nucleic acid structure. *J. Mol. Biol* 1999, 285, 1577–1588. [PubMed: 9917398]
22. Batey RT; Kieft JS Improved native affinity purification of RNA. *RNA* 2007, 13, 1384–1389. [PubMed: 17548432]
23. Golden BL; Podell ER; Gooding AR; Cech TR Crystals by design: A strategy for crystallization of a ribozyme derived from the Tetrahymena group I intron. *J. Mol. Biol* 1997, 270, 711–723. [PubMed: 9245599]
24. Lippa GM; Liberman JA; Jenkins JL; Krucinska J; Salim M; Wedekind JE Crystallographic analysis of small ribozymes and riboswitches. *Methods Mol. Biol* 2012, 848, 159–184. [PubMed: 22315069]
25. MacElrevey C; Spitale RC; Krucinska J; Wedekind JE A posteriori design of crystal contacts to improve the X-ray diffraction properties of a small RNA enzyme. *Acta Crystallogr. D Biol. Crystallogr* 2007, 63, 812–825. [PubMed: 17582172]
26. Reyes FE; Garst AD; Batey RT Strategies in RNA crystallography. *Methods Enzymol.* 2009, 469, 119–139. [PubMed: 20946787]
27. Schultz SC; Shields GC; Steitz TA Crystallization of Escherichia coli catabolite gene activator protein with its DNA binding site. The use of modular DNA. *J. Mol. Biol* 1990, 213, 159–166. [PubMed: 2160019]
28. Coonrod LA; Lohman JR; Berglund JA Utilizing the GAAA tetraloop/receptor to facilitate crystal packing and determination of the structure of a CUG RNA helix. *Biochemistry* 2012, 51, 8330–8337. [PubMed: 23025897]
29. Ferré-D’Amaré AR; Zhou K; Doudna JA A general module for RNA crystallization. *J. Mol. Biol* 1998, 279, 621–631. [PubMed: 9641982]
30. Tamjar J; Katorcha E; Popov A; Malinina L Structural dynamics of double-helical RNAs composed of CUG/CUG- and CUG/CGG-repeats. *J. Biomol. Struct. Dyn* 2012, 30, 505–523. [PubMed: 22731704]
31. Zhang J; Ferré-D’Amare AR Dramatic improvement of crystals of large RNAs by cation replacement and dehydration. *Structure* 2014, 22, 1363–1371. [PubMed: 25185828]
32. Zhang J; Ferré-D’Amare AR Improving RNA Crystal Diffraction Quality by Postcrystallization Treatment. *Methods Mol. Biol* 2021, 2323, 25–37. [PubMed: 34086271]
33. Chen Y; Varani G Engineering RNA-binding proteins for biology. *FEBS J.* 2013, 280, 3734–3754. [PubMed: 23742071]

34. Khisamutdinov EF; Jasinski DL; Guo P RNA as a boiling-resistant anionic polymer material to build robust structures with defined shape and stoichiometry. *ACS Nano* 2014, 8, 4771–4781. [PubMed: 24694194]
35. Ferre-D'Amare AR; Doudna JA Methods to crystallize RNA. *Curr. Protoc. Nucleic Acid Chem* 2001.
36. Holbrook SR RNA structure: The long and the short of it. *Curr. Opin. Struct. Biol* 2005, 15, 302–308. [PubMed: 15963891]
37. Dibrov SM; Parker MA; Bergdahl BM; Hermann T Crystal structure of a benzimidazole hepatitis C virus inhibitor free and in complex with the viral RNA target. *J. Chem. Crystallogr* 2013, 43, 235–239. [PubMed: 23750099]
38. Mooers BH; Singh A The crystal structure of an oligo(U):pre-mRNA duplex from a trypanosome RNA editing substrate. *RNA* 2011, 17, 1870–1883. [PubMed: 21878548]
39. Mooers BH Fusion RNAs in crystallographic studies of double-stranded RNA from trypanosome RNA editing. *Methods Mol. Biol* 2015, 1240, 191–216. [PubMed: 25352146]
40. Mooers BH Direct-methods structure determination of a trypanosome RNA-editing substrate fragment with translational pseudosymmetry. *Acta Crystallogr. D Struct. Biol* 2016, 72, 477–487. [PubMed: 27050127]
41. Drenth J Principles of Protein X-Ray Crystallography, 3rd ed.; Springer-Verlag: New York, NY, USA, 2007.
42. Holbrook SR; Cheong C; Tinoco I Jr.; Kim SH Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature* 1991, 353, 579–581. [PubMed: 1922368]
43. Dibrov S; McLean J; Hermann T Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA. *Acta Crystallogr. D Biol. Crystallogr* 2011, 67, 97–104. [PubMed: 21245530]
44. Beuning PJ; Tessmer MR; Baumann CG; Kallick DA; Musier-Forsyth K Sequence-dependent conformational differences of small RNAs revealed by native gel electrophoresis. *Anal. Biochem* 1999, 273, 284–290. [PubMed: 10469499]
45. Woodson SA; Koculi E Analysis of RNA folding by native polyacrylamide gel electrophoresis. *Methods Enzymol.* 2009, 469, 189–208. [PubMed: 20946790]
46. Oubridge C; Ito N; Evans PR; Teo CH; Nagai K Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 1994, 372, 432–438. [PubMed: 7984237]
47. Price SR; Evans PR; Nagai K Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 1998, 394, 645–650. [PubMed: 9716128]
48. Hoang C; Ferre-D'Amare AR Cocrystal structure of a tRNA^{Psi55} pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell* 2001, 107, 929–939. [PubMed: 11779468]
49. Ennifar E; Nikulin A; Tishchenko S; Serganov A; Nevskaya N; Garber M; Ehresmann B; Ehresmann C; Nikonov S; Dumas P The crystal structure of UUCG tetraloop. *J. Mol. Biol* 2000, 304, 35–42. [PubMed: 11071808]
50. Tomizawa J Control of ColE1 plasmid replication: The process of binding of RNA I to the primer transcript. *Cell* 1984, 38, 861–870. [PubMed: 6207934]
51. Paillart JC; Skripkin E; Ehresmann B; Ehresmann C; Marquet R A loop-loop “kissing” complex is the essential part of the dimer linkage of genomic HIV-1 RNA. *Proc. Natl. Acad. Sci. USA* 1996, 93, 5572–5577. [PubMed: 8643617]
52. Shetty S; Kim S; Shimakami T; Lemon SM; Mihailescu MR Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. *RNA* 2010, 16, 913–925. [PubMed: 20360391]
53. Moras D; Comarmond MB; Fischer J; Weiss R; Thierry JC; Ebel JP; Giege R Crystal structure of yeast tRNA^{Asp}. *Nature* 1980, 288, 669–674. [PubMed: 7005687]
54. Kim CH; Tinoco I Jr. A retroviral RNA kissing complex containing only two G.C base pairs. *Proc. Natl. Acad. Sci. USA* 2000, 97, 9396–9401. [PubMed: 10931958]

55. Lee AJ; Crothers DM The solution structure of an RNA loop-loop complex: The ColE1 inverted loop sequence. *Structure* 1998, 6, 993–1005. [PubMed: 9739090]
56. Marino JP; Gregorian RS Jr.; Csankovszki G; Crothers DM Bent helix formation between RNA hairpins with complementary loops. *Science* 1995, 268, 1448–1454. [PubMed: 7539549]
57. Kondo Y; Oubridge C; van Roon AM; Nagai K Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5' splice site recognition. *eLife* 2015, 4, e04986.
58. Pomeranz Krummel DA; Oubridge C; Leung AK; Li J; Nagai K Crystal structure of human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* 2009, 458, 475–480. [PubMed: 19325628]
59. Leung AKW; Kondo Y; Krummel DAP; Li J; Price SR; van Roon A-MM Engineering Crystal Packing in RNA-Protein Complexes II: A Historical Perspective from the Structural Studies of the Spliceosome. *Crystals* 2021, in press.
60. Ennifar E; Dumas P Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J. Mol. Biol* 2006, 356, 771–782. [PubMed: 16403527]
61. Kondo J; Pachamuthu K; Francois B; Szychowski J; Hanessian S; Westhof E Crystal structure of the bacterial ribosomal decoding site complexed with a synthetic doubly functionalized paromomycin derivative: A new specific binding mode to an a-minor motif enhances in vitro antibacterial activity. *ChemMedChem* 2007, 2, 1631–1638. [PubMed: 17722211]
62. Vicens Q; Westhof E Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site. *Structure* 2001, 9, 647–658. [PubMed: 11587639]
63. Leontis NB; Westhof E Geometric nomenclature and classification of RNA base pairs. *RNA* 2001, 7, 499–512. [PubMed: 11345429]
64. D'Ascenzo L; Leonarski F; Vicens Q; Auffinger P Revisiting GNRA and UNCG folds: U-turns versus Z-turns in RNA hairpin loops. *RNA* 2017, 23, 259–269. [PubMed: 27999116]
65. Nozinovic S; Furtig B; Jonker HR; Richter C; Schwalbe H High-resolution NMR structure of an RNA model system: The 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res.* 2010, 38, 683–694. [PubMed: 19906714]
66. Allain FH; Varani G Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol* 1995, 250, 333–353. [PubMed: 7608979]
67. D'Ascenzo L; Vicens Q; Auffinger P Identification of receptors for UNCG and GNRA Z-turns and their occurrence in rRNA. *Nucleic Acids Res.* 2018, 46, 7989–7997. [PubMed: 29986118]
68. Pley HW; Flaherty KM; McKay DB Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 1994, 372, 111–113. [PubMed: 7526219]
69. Murphy FL; Cech TR GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J. Mol. Biol* 1994, 236, 49–63. [PubMed: 8107125]
70. Jaeger L; Michel F; Westhof E Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J. Mol. Biol* 1994, 236, 1271–1276. [PubMed: 7510342]
71. Costa M; Michel F Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* 1995, 14, 1276–1285. [PubMed: 7720718]
72. Costa M; Michel F Rules for RNA recognition of GNRA tetraloops deduced by in vitro selection: Comparison with in vivo evolution. *EMBO J.* 1997, 16, 3289–3302. [PubMed: 9214644]
73. Geary C; Baudrey S; Jaeger L Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res.* 2008, 36, 1138–1152. [PubMed: 18158305]
74. Leung AK; Kambach C; Kondo Y; Kampmann M; Jinek M; Nagai K Use of RNA tertiary interaction modules for the crystallisation of the spliceosomal snRNP core domain. *J. Mol. Biol* 2010, 402, 154–164. [PubMed: 20643141]
75. Leung AK; Nagai K; Li J Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* 2011, 473, 536–539. [PubMed: 21516107]
76. Reiter NJ; Osterman A; Torres-Larios A; Swinger KK; Pan T; Mondragon A Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* 2010, 468, 784–789. [PubMed: 21076397]

77. Toor N; Keating KS; Taylor SD; Pyle AM Crystal structure of a self-spliced group II intron. *Science* 2008, 320, 77–82. [PubMed: 18388288]
78. Keating KS; Toor N; Pyle AM The GANC tetraloop: A novel motif in the group IIC intron structure. *J. Mol. Biol* 2008, 383, 475–481. [PubMed: 18773908]
79. Ishikawa J; Furuta H; Ikawa Y An in vitro-selected RNA receptor for the GAAC loop: Modular receptor for non-GNRA-type tetraloop. *Nucleic Acids Res.* 2013, 41, 3748–3759. [PubMed: 23382175]
80. Lescoute A; Leontis NB; Massire C; Westhof E Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res.* 2005, 33, 2395–2409. [PubMed: 15860776]
81. Ohuchi SP; Ikawa Y; Nakamura Y Selection of a novel class of RNA-RNA interaction motifs based on the ligase ribozyme with defined modular architecture. *Nucleic Acids Res.* 2008, 36, 3600–3607. [PubMed: 18460545]
82. Tamura M; Holbrook SR Sequence and structural conservation in RNA ribose zippers. *J. Mol. Biol* 2002, 320, 455–474. [PubMed: 12096903]
83. Kondo J; Dock-Bregeon AC; Willkomm DK; Hartmann RK; Westhof E Structure of an A-form RNA duplex obtained by degradation of 6S RNA in a crystallization droplet. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun* 2013, 69, 634–639.
84. Reblova K; Fadrna E; Sarzynska J; Kulinski T; Kulhanek P; Ennifar E; Koca J; Sponer J Conformations of flanking bases in HIV-1 RNA DIS kissing complexes studied by molecular dynamics. *Biophys. J* 2007, 93, 3932–3949. [PubMed: 17704156]
85. Sarzynska J; Reblova K; Sponer J; Kulinski T Conformational transitions of flanking purines in HIV-1 RNA dimerization initiation site kissing complexes studied by CHARMM explicit solvent molecular dynamics. *Biopolymers* 2008, 89, 732–746. [PubMed: 18412127]
86. Ogle JM; Brodersen DE; Clemons WM Jr.; Tarry MJ; Carter AP; Ramakrishnan V Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 2001, 292, 897–902. [PubMed: 11340196]
87. Banco MT; Ferre-D'Amare AR The emerging structural complexity of G-quadruplex RNAs. *RNA* 2021, 27, 390–402. [PubMed: 33483368]
88. Feklistov A; Darst SA Crystallographic analysis of an RNA polymerase σ -subunit fragment complexed with –10 promoter element ssDNA: Quadruplex formation as a possible tool for engineering crystal contacts in protein-ssDNA complexes. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun* 2013, 69, 950–955.
89. Yoshizawa S *Nanotechnology Tools for the Study of RNA*, 1st ed.; Academic Press: Cambridge, MA, USA, 2016; Volume 139.
90. Guo P; Haque F *RNA Nanotechnology and Therapeutics: Methods and Protocols*; Springer: Berlin/Heidelberg, Germany, 2015.
91. Grabow WW; Jaeger L RNA self-assembly and RNA nanotechnology. *Acc. Chem. Res* 2014, 47, 1871–1880. [PubMed: 24856178]
92. Jasinski D; Haque F; Binzel DW; Guo P Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* 2017, 11, 1142–1164. [PubMed: 28045501]
93. Xia K; Shen J; Li Q; Fan C; Gu H Near-Atomic Fabrication with Nucleic Acids. *ACS Nano* 2020, 14, 1319–1337. [PubMed: 31976651]
94. Seeman NC Nucleic acid junctions and lattices. *J. Theor. Biol* 1982, 99, 237–247. [PubMed: 6188926]
95. Shu D; Shu Y; Haque F; Abdelmawla S; Guo P Thermodynamically stable RNA three-way junction for constructing multifunctional nanoparticles for delivery of therapeutics. *Nat. Nanotechnol* 2011, 6, 658–667. [PubMed: 21909084]
96. Ferre-D'Amare AR Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods* 2010, 52, 159–167. [PubMed: 20554048]
97. Koldobskaya Y; Duguid EM; Shechner DM; Suslov NB; Ye J; Sidhu SS; Bartel DP; Koide S; Kossiakov AA; Piccirilli JA A portable RNA sequence whose recognition by a synthetic antibody facilitates structural determination. *Nat. Struct. Mol. Biol* 2011, 18, 100–106. [PubMed: 21151117]

98. Koide S Engineering of recombinant crystallization chaperones. *Curr. Opin. Struct. Biol* 2009, 19, 449–457. [PubMed: 19477632]
99. Ye JD; Tereshko V; Frederiksen JK; Koide A; Fellouse FA; Sidhu SS; Koide S; Kossiakoff AA; Piccirilli JA Synthetic antibodies for specific recognition and crystallization of structured RNA. *Proc. Natl. Acad. Sci. USA* 2008, 105, 82–87. [PubMed: 18162543]
100. Nielsen PE; Egholm M; Berg RH; Buchardt O Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 1991, 254, 1497–1500. [PubMed: 1962210]
101. Kiliszek A; Banaszak K; Dauter Z; Rypniewski W The first crystal structures of RNA-PNA duplexes and a PNA-PNA duplex containing mismatches—toward anti-sense therapy against TREDs. *Nucleic Acids Res.* 2016, 44, 1937–1943. [PubMed: 26717983]
102. Verona MD; Verdolino V; Palazzesi F; Corradini R Focus on PNA Flexibility and RNA Binding using Molecular Dynamics and Metadynamics. *Sci. Rep* 2017, 7, 42799. [PubMed: 28211525]
103. Arnott S; Hukins DW; Dover SD Optimised parameters for RNA double-helices. *Biochem. Biophys. Res. Commun* 1972, 48, 1392–1399. [PubMed: 5077825]
104. Pflugrath JW Practical macromolecular cryocrystallography. *Acta Crystallogr. F Struct. Biol. Commun* 2015, 71, 622–642. [PubMed: 26057787]
105. Russo Krauss I; Sica F; Mattia CA; Merlino A Increasing the X-ray diffraction power of protein crystals by dehydration: The case of bovine serum albumin and a survey of literature data. *Int. J. Mol. Sci* 2012, 13, 3782–3800. [PubMed: 22489183]
106. Klein DJ; Ferre-D’Amare AR Crystallization of the glmS ribozyme-riboswitch. *Methods Mol. Biol* 2009, 540, 129–139. [PubMed: 19381557]
107. Zhang J; Ferre-D’Amare AR Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* 2013, 500, 363–366. [PubMed: 23892783]
108. Voth AR; Hays FA; Ho PS Directing macromolecular conformation through halogen bonds. *Proc. Natl. Acad. Sci. USA* 2007, 104, 6188–6193. [PubMed: 17379665]
109. Ennifar E; Bernacchi S; Wolff P; Dumas P Influence of C-5 halogenation of uridines on hairpin versus duplex RNA folding. *RNA* 2007, 13, 1445–1452. [PubMed: 17630326]
110. Shi D; Nannenga BL; Iadanza MG; Gonen T Three-dimensional electron crystallography of protein microcrystals. *eLife* 2013, 2, e01345. [PubMed: 24252878]
111. Senior AW; Evans R; Jumper J; Kirkpatrick J; Sifre L; Green T; Qin C; Zidek A; Nelson AWR; Bridgland A; et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020, 577, 706–710. [PubMed: 31942072]
112. Jumper J; Evans R; Pritzel A; Green T; Figurnov M; Ronneberger O; Tunyasuvunakool K; Bates R; Zidek A; Potapenko A; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021.
113. Desmet J; De Maeyer M; Hazes B; Lasters I The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* 1992, 356, 539–542. [PubMed: 21488406]
114. Kappel K; Zhang K; Su Z; Watkins AM; Kladwang W; Li S; Pintilie G; Topkar VV; Rangan R; Zheludev IN; et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat. Methods* 2020, 17, 699–707. [PubMed: 32616928]
115. Su Z; Zhang K; Kappel K; Li S; Palo MZ; Pintilie GD; Rangan R; Luo B; Wei Y; Das R; et al. Cryo-EM structures of full-length Tetrahymena ribozyme at 3.1 Å resolution. *Nature* 2021.
116. Banatao DR; Cascio D; Crowley CS; Fleissner MR; Tienson HL; Yeates TO An approach to crystallizing proteins by synthetic symmetrization. *Proc. Natl. Acad. Sci. USA* 2006, 103, 16230–16235. [PubMed: 17050682]
117. Khisamutdinov EF; Li H; Jasinski DL; Chen J; Fu J; Guo P Enhancing immunomodulation on innate immunity by shape transition among RNA triangle, square and pentagon nanovehicles. *Nucleic Acids Res.* 2014, 42, 9996–10004. [PubMed: 25092921]
118. Claverie JM Fewer genes, more noncoding RNA. *Science* 2005, 309, 1529–1530. [PubMed: 16141064]
119. Rouskin S; Zubradt M; Washietl S; Kellis M; Weissman JS Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 2014, 505, 701–705. [PubMed: 24336214]

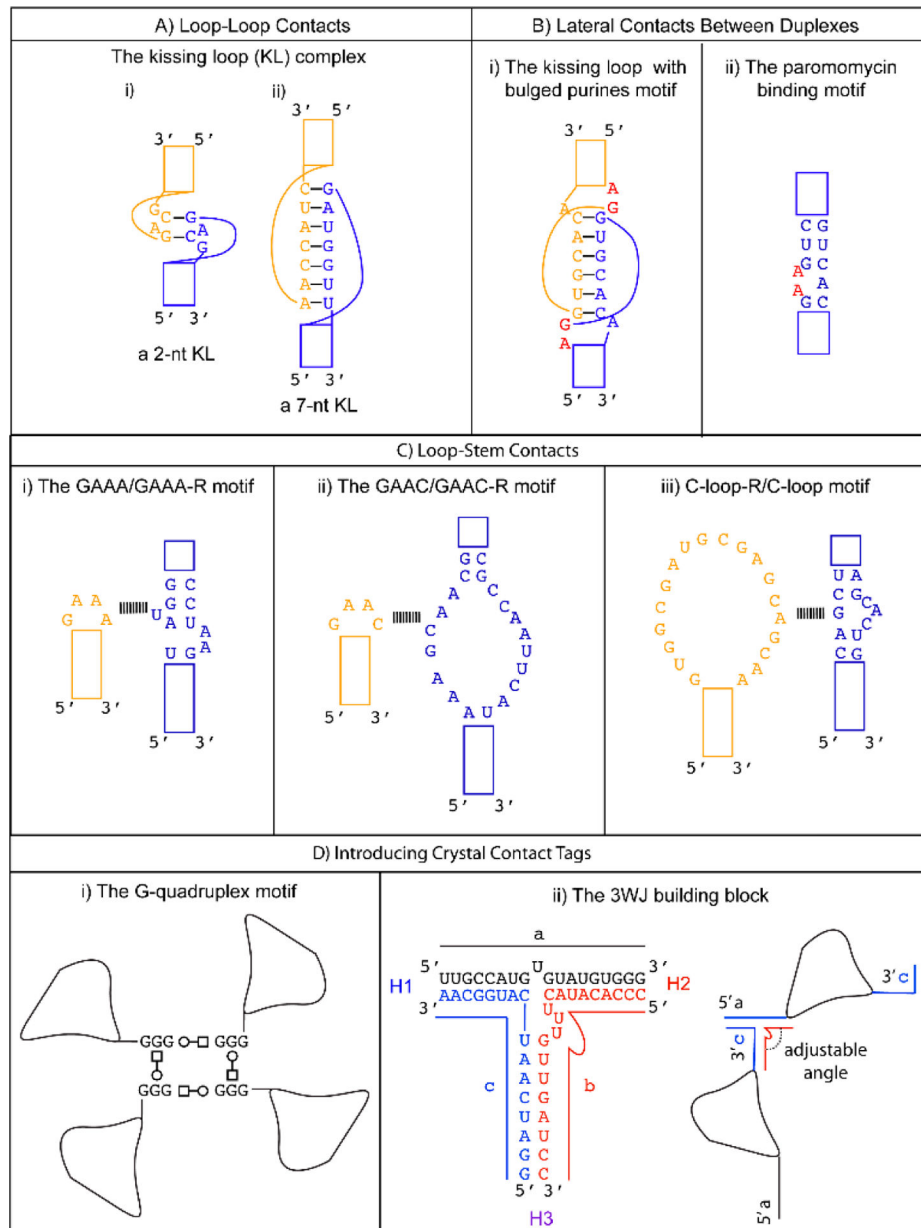


Figure 1. Summary of motifs discussed in the review. **(A)** Two examples of the kissing loop (KL) complex. **(i)** A 2-nt KL complex is formed between the H3 stem loop of the Moloney Murine leukemia virus [54]. **(ii)** A 7-nt KL complex is formed between stem loops I and II of the ColE1 plasmid [55,56]. **(B)** Design options to induce lateral contacts between duplexes. **(i)** A kissing loop complex with flipped-out purines from the dimerization initiation site of the HIV-1 genome [60]. **(ii)** The paramomycin binding site from the 16S rRNA of the *E. coli* ribosome can be utilized to introduce bulged adenines that can form tertiary interactions with a neighboring backbone [61,62]. **(C)** Design options to induce loop to stem contacts. **(i)** The classical GNRA tetraloop (GAAA) interacting with its 11-nt receptor GAAA-R. **(ii)** The non-GNRA tetraloop (GAAC) interacting with its 20-nt receptor

GAAC-R identified by in vitro selection. **(iii)** The asymmetrical C-loop interacting with its 20-nt receptor C-loop-R identified by in vitro selection. **(D)** Design option to introduce a crystal contact tag. **(i)** A G-rich tag can be introduced to the ends of a nucleic acid structure to induce the formation of a G-quadruplex. The squares and circles indicate the tertiary interactions following the Leontis and Westhof nomenclature [63]. A circle indicates a Watson–Crick edge, and a square indicates a Hoogsteen edge. Open or white color indicates *trans* orientation of the glycosidic bond. **(ii)** The three-way junction motif (3WJ) can be self-assembled from three oligonucleotides (strands a, b, c). One way to utilize this motif is to engineer strands a and c into the ends of the RNA. The third strand b can be supplied in the crystallization drop. If a crystal is formed, the packing geometry can be fine-tuned by simply changing strand b with a different number of unpaired nucleotides at the junction.

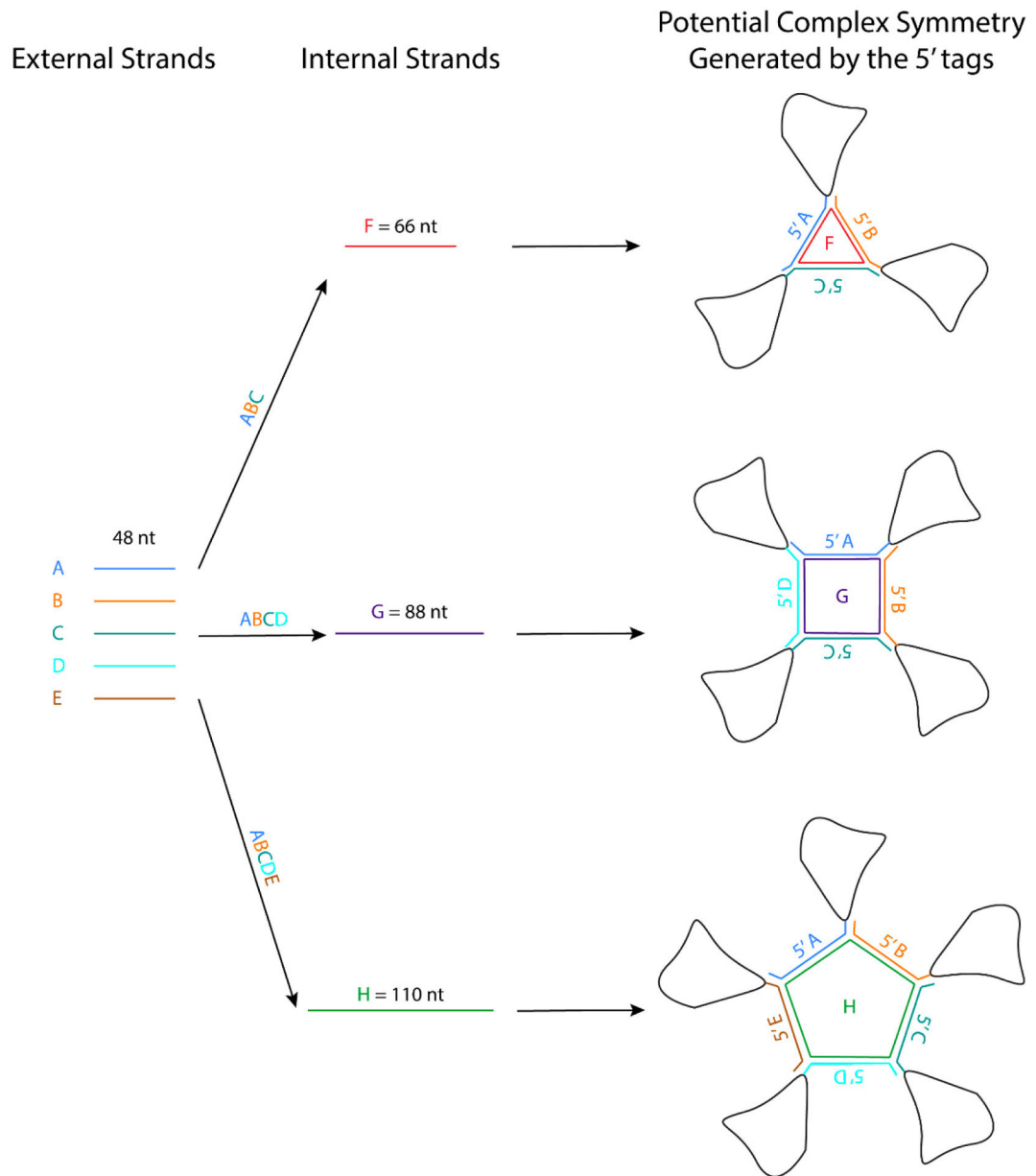


Figure 2. Potential symmetry generation using 3WJ junction forming oligonucleotides (Adapted from Khisamutdinov et al., 2014 [117]). External strands can be cloned into the 5' end of a complex. Internal strands can be supplied into the sample to induce polygon formation.