Ribonucleic acid (RNA) design offers unique opportunities for engineering genetic networks and nanostructures that self-assemble within living cells. Recent years have seen the creation of increasingly complex RNA devices, including proof-of-concept applications for in vivo three-dimensional scaffolding, imaging, computing, and control of biological behaviors. Expert intuition and simple design rules — the stability of double helices, the modularity of noncanonical RNA motifs, and geometric closure — have enabled these successful applications. Going beyond heuristics, emerging algorithms may enable automated design of RNAs with nucleotide-level accuracy but, as illustrated on a recent RNA square design, are not yet fully predictive. Looking ahead, technological advances in RNA synthesis and interrogation are poised to radically accelerate the discovery and stringent testing of design methods.

Addresses
1 Department of Biochemistry, Stanford University, Stanford, CA 94305, USA
2 Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
3 Department of Physics, Stanford University, Stanford, CA 94305, USA

Corresponding author: Das, R (rhiju@stanford.edu)

Ribonucleic acid (RNA) is, in many respects, an ideal polymer for biomolecular design. Natural and designed RNA chains can be exquisitely functional: they code for genetic information, assemble into intricate three-dimensional structures, catalyze chemical reactions, and engage in nearly every essential biological process in living cells (extensively reviewed in [1]). Most appealingly, these RNA behaviors can be explained and designed by rules that appear strikingly simple and, in favorable cases, are quantitatively predictive. Indeed, our motivation for studying RNA design puzzles is that they offer opportunities to discover and rigorously test such predictive rules, which might, in turn, enable the modeling of any RNA ‘from scratch’.

This review focuses on current and missing rules for RNA design, starting with a description of three basic heuristics in universal use. The three subsequent sections review recent work to expand each rule into predictive theories and tools, highlighting current mismatches between theory and practice. We conclude with a perspective on trends that may radically accelerate the discovery of new RNA design theories and devices.

### Three heuristic design rules

**Rule 1. Watson–Crick base pairs generate stable double helices.** Every RNA design made to date has taken advantage of helical stems formed as RNA strands double back on themselves or associate with other strands to form Watson–Crick base pairs. For the RNA square, its edges are four helices, each involving association of four strands (Figure 1b).

**Rule 2. RNA motifs can preserve their behavior when copied and pasted into new contexts.** Explorations of RNAs in living systems and elegant in vitro selections have revealed a plethora of natural RNA catalysts, sensors, and structures. Many of these molecules’ functions are due to small (4–15 nucleotide) RNA motifs with non-canonical structure. These motifs can be grafted into if care is taken to avoid mispairings with flanking sequences. For example, the four ‘nanocorners’ of the RNA square (Figure 1c) are copies of a five-nucleotide bulge motif, drawn from a right-angle-forming structure in the Hepatitis C virus genome [5].

**Rule 3. Geometric closure ensures correct three-dimensional structure.** For 3D structures, the geometry of helical stems must successfully interconnect with noncanonical motifs — a stringent requirement for RNA designs that encompass closed ‘ring’ topologies (Figure 1d). For the RNA square, the choice of 10 base pairs for the edges leads to a well-ordered square assembly with no detectable alternative species. Extension to 11 base pairs precludes the closure of the square and gives higher-order assemblies [2**].

Can these intuitive rules be explained and expanded into quantitatively predictive theories for RNA design?

### Expanding Rule 1: theories for RNA secondary structure design

Physical theories underlying RNA secondary structure formation are the most developed models in RNA science
2 Engineering and design

A square illustrates RNA design. (a) Sequences (inner and outer) determined ‘by hand’ by Hermann group [2**]; four copies of the two sequences form the square. (b) Multi-strand secondary structure involves each inner and outer strand forming a nanocorner, and four nanocorners pairing through three-nucleotide sticky ends. (c) Inspiration for the nanocorner, a 5-nt bulge that forms a 90 °C bend in the Hepatitis C virus internal ribosomal entry site (crystallized in PDB ID 2P4N). (d) Crystallographic analysis for full RNA square (PDB ID 3P95). (e) Secondary structure predictions for square design are inaccurate (NUPACK models shown [26]). (f) The ab initio modeling problem of rebuilding the nanocorner’s 5-nt bulge — even given the rest of the coordinates — is not solvable at atomic resolution. (g) An RNA-Puzzle [70**]; blind prediction of the nanosquare conformation by Rosetta methods is not atomically accurate, even when given the inner strand coordinates as a constraint.

and arguably amongst the most predictive theories available in biophysics. Several decades of melt experiments on thousands of RNA sequences have been distilled into models that parameterize canonical base pair formation into two dozen ‘nearest-neighbor’ parameters, where the thermodynamic stability of a given base pair depends on adjacent base pairs and temperature [6,7]. Further modeling and measurements have provided approximate energetic rules (Figure 2a) for many hairpin loops, base pair mismatches, and more complex inter-helical motifs such as three-way junctions [7–13]. Given such energetic models, dynamic programming algorithms implemented in packages such as mfold/UNAfold [14], RNAstructure [15], and ViennaRNA [16] permit the comprehensive statistical mechanical description of RNA secondary structure ensembles for arbitrary sequences (Figure 2b) with the ability to recover ~70% of phylogenetically determined base pairs [7]. Modeling methods for secondary structure formation kinetics are also under exploration in packages such as Kinefold [17].

Please cite this article in press as: Bida JP, Das R. Squaring theory with practice in RNA design, Curr Opin Struct Biol (2012), http://dx.doi.org/10.1016/j.sbi.2012.06.003
Figure 2

Theory and practice in RNA secondary structure design. (a) The free energy of a sequence in a given secondary structure conformation is calculated by summing up the free energies of nearest neighbor terms derived from extensive empirical measurements (green) [6–13]. (b) Dynamic programming algorithms can calculate the minimal free energy structure, base-pair probabilities, and partition function [14,15]. These methods recursively calculate...
4 Engineering and design

Most relevant for RNA design are recent approaches for predicting sequences that fold into given target secondary structures, the “inverse” folding problem. Methods range from simple Monte Carlo sequence searches (InverseRNA [18]) to faster hierarchical schemes (RNA-SSD [19] and INFO-RNA [20]) that attempt to solve the design problem for substructures or with simplifying assumptions before merging or optimizing solutions (Figure 2c). Most efforts have focused on finding sequences whose minimum free energy conformations recover the desired structure or, in some cases, multiple structures [16,21–23,24].

The NUPACK software has presented optimization algorithms for a novel and intuitive target metric, the ‘ensemble defect’, which parameterizes the fluctuations of each nucleotide away from its desired configuration [25*], [26]. Systematic experimental benchmarks of these design algorithms on novel secondary structure targets, perhaps using chemical accessibility mapping [27], would be valuable but are not yet available.

RNA secondary structure prediction and design algorithms are being widely used to help develop novel molecules. Design of RNA thermometers for temperature-controlled gene expression has made use of the nearest neighbor rules [28] (Figure 2d). Isambert, Schwab, and other labs have designed model systems with appealing simplicity to study cotranscriptional folding and two-state switching [29–32]. Carothers and colleagues have calculated the properties of RNA devices taking into account the folding rates of ribozymes and aptamers calculated in Kinefold [33]. Pierce and colleagues have imported the ‘hybridization chain reaction’ first developed in DNA engineering to create multiplexed amplifiers for in situ hybridization to mRNA targets in zebrafish embryos [34*] (Figure 2e and f) and to induce apoptosis of cultured human cells in response to cancer marker RNAs [35*].

While making use of RNA secondary structure modeling algorithms, these successful design efforts have still required significant insight from experts and would be challenging to generate automatically. For example, the RNA square, which was designed ‘by hand’ [2*], is a problem case for current secondary structure calculation methods, which predict alternative structures (Figure 1e). The squares’ nanocorners are especially stable motif sequences that form noncanonical hydrogen bonds and specifically coordinate multivalent ions — features that are not treated in current RNA thermodynamic models. In principle, quantum chemical and molecular dynamics approaches should enable the ab initio calculation of these ‘missing’ parameters in arbitrary solution conditions but are still under calibration [36–43]. Further, most secondary structure prediction packages do not yet model ring-like or pseudoknot structures, although new extensions are attempting to tackle this issue [44,45,46*].

Expanding Rule 2: new RNA motifs and new combinations

A large arsenal of functional motifs is available for RNA design (Figure 3a). Quite broadly, RNA sequences that form specific three-dimensional structures, bind to small molecules or proteins, or mediate cellular localization, degradation, transcription termination, splicing, editing, or other cellular RNA processing events are widely used in biological inquiry (reviewed recently in [47,48,49]). Several groups are compiling databases with sequences and structures of existing motifs derived from natural functional RNAs or in vitro selections [50–54].

RNA designs have demonstrated the power of utilizing multiple interacting motifs, often coupled allosterically via secondary structure switches or kissing loops. Motifs embedded in such devices include small-molecule-binding ‘aptamers’ such as the theophylline-binding motif, catalytic modules such as the hammerhead ribozyme three-way junction, and protein-binding motifs [55–57,58*], [59]. The resulting devices have been expressed in bacteria, yeast, and mammalian cells to create circuits of growing complexity, including on/off switches [60,61] (Figure 3a), RNA-based logic gates [56,62–64], cooperative behaviors [56], and bacteria that swim toward and ingest the herbicide atrazine [65]. Creative ways to discover and select new functional motifs are also being pursued, for example, through juxtaposition of randomized sequences by tertiary scaffolds [66,67] (see also below). While all the designs have leveraged the modularity of functional motifs, many successes have required trial-and-error or selection in vitro [57,65]. Presentation of failure cases and their in-depth dissections, as are now carried out in the protein design field [68,69], would be valuable.

Even with the growing database of known functional motifs, there may always be components for RNA design that are not naturally available, such as aptamers for specific moieties in large macromolecules. A general solution to such de novo design problems would be useful but has not been demonstrated. Indeed, when the RNA square crystallographic structure was offered as a blind trial to RNA modelers, no algorithm reached atomic accuracy predictions of the nanocorners even when given the coordinates of four of the eight internal strands of the motifs (Figure 1g) [70*]. Nevertheless, a recent
explosion of activity in RNA 3D classification [71–73] and motif modeling [70*,74] (Figure 1f and g), with some algorithms achieving atomic accuracy in favorable cases [39,75*,76], suggests that de novo design may become a reality. As an early step, we adapted the Rosetta design method for fixed-backbone biopolymer design [77] to successfully thermostabilize a non-canonical motif in the signal recognition particle RNA through mutations previously unseen in Nature (Figure 3b) [75*].

Expanding Rule 3: theories for RNA tertiary structure design

The successful design of highly complex 3D RNA structures continues to involve connecting RNA motifs and

Theory and practice in RNA tertiary structure design. (a) Cutting and pasting motifs: kissing loops, a theophylline aptamer, and the IS10 ncRNA were rationally combined to engineer trans-acting ncRNAs that regulates gene translation that can be turned on or off with theophylline [57]. (b) New tertiary motifs generated with Rosetta fixed backbone design for RNA, based on in silico energy minimization. The resulting designs for the most conserved domain of the signal recognition particle RNA are distinct from natural RNAs and turn out to be more stable experimentally [75*]. (c) RNA tectonics: combination of tertiary components and helices in three dimensions, optimized by hand or computationally [106]. (d) Nanoring assembled from RNA/I/II inverse kissing complexes, validated by atomic force microscopy (middle) and non-denaturing gel electrophoresis [88].
helical stems, typically with additional tertiary motifs and cross-strand interactions that generate ring-like topologies (Figure 3d). This procedure, sometimes called RNA ‘tectonics’, is accomplished with manually directed programs such as NanoTiler [78] or Assemble [79] to put together RNA components in silico.

Visually compelling examples of designed 3D RNA structures have proliferated since the first ‘TectoRNA’, an RNA dimer stabilized by docking of tetraloops and their cognate receptors [80]. Several groups have designed and validated RNA tiles [81,82], multimers [83], cubes [24*], prisms [84], triangles [85], fabrics [81,82*], fibrils [86,87], and rings (Figure 3d) [88,89], often using different components (tetraloop/receptors, kissing loops, tRNA-based junctions) to make similar objects [90*].

Many of these designs echo work in DNA nanotechnology [4], but RNA design permits, in principle, the deployment of nanostructures in living cells. Notably, Delebecque and colleagues [82*] have demonstrated self-assembly of nanofabrics in vivo and used the resulting structures to accelerate hydrogen production in bacteria.

Figure 4

A massively parallel design cycle should be possible for RNA. Thousands of sequences (a) designed by experts, citizen scientists, and automated design tools are sent for high-throughput oligonucleotide template synthesis (b) [96,97], (figure adopted from LC sciences) transcribed into RNA (c), and then falsified or validated at nucleotide resolution through multiplexed chemical mapping (d) utilizing-next generation-sequencing methods [99].
through the scaffolding of protein enzymes. Other applications include scaffolding gold particles for nanowires [91], encaging other molecules [84], and protecting duplexes intended for RNA silencing from degradation [88, 92].

Atomic-scale validation of 3D design — as was achieved quite early for protein design [93, 94] — has not yet been demonstrated for RNA nanostructures. For example, the crystallographic analysis of the four-fold sequence-symmetric RNA square showed the four nanocorners forming different structures (Figure 1d). In addition to structure modeling puzzles, an unaddressed challenge is to model the thermodynamic stabilities of RNA tertiary folds with accuracies comparable to the nearest-neighbor rules for secondary structure energetics. Improving tools for measuring helix and junction flexibility (see, e.g., [95]) should aid greatly in such efforts.

**Perspective: an impending acceleration**

In recent years, RNA designers have leveraged simple rules for secondary structure formation, the modularity of small RNA motifs, and three-dimensional closure to produce devices and nanostructures of growing complexity. It is particularly exciting to see these designs deployed into biological systems, suggesting future routes to biomedical devices that sense and perhaps correct cellular dysfunction. Nevertheless, challenges remain, particularly in attaining predictive theories for RNA assembly at nucleotide resolution, much less atomic resolution. Multi-strand topologies, non-canonical interactions, and energies of 3D assemblies remain unsolved challenges in RNA design theory. Both our physical models and our ability to make predictions from these models need improvement.

We are optimistic about the development of more quantitative and predictive theories for RNA structure design and eventually RNA functional design due to the explosion of RNA data expected in upcoming years. The cycle of RNA design and testing can be short, especially compared to protein engineering — arbitrary RNA sequences up to hundreds of nucleotides in length are experimentally straightforward to synthesize, purify, interrogate, and evolve on the timescale of days. Current technologies offer the parallel synthesis of thousands of arbitrary DNA templates (Figure 4a and b) [96, 97], which can be transcribed into RNA in **vitro** and, in **vivo**, (Figure 4c). Further, multiplexed single-nucleotide-resolution chemical mapping [98, 99] and elegant selection strategies [100] can provide information-rich data on all of these molecules’ structures and functions, using deep-sequencing platforms with turn-around times of hours (Figure 4d). The key challenges will then be to distribute [101] and accurately analyze these data [102–105], to update RNA modeling algorithms, and to feed back these insights into the next rounds of synthesis and mapping. The result, a massively parallel design cycle, would offer unprecedented opportunities for squaring theory with practice in RNA design.

**Acknowledgments**

We thank EteRNA players for sequences in Figure 4. We thank members of the Das laboratory, EteRNA players Quaspecies and Eli Fisker, J. Lee, and J. Lucks for discussions and manuscript comments. We acknowledge funding from the Burroughs-Wellcome Foundation (CASI to RD) and a Stanford Bio-X Games Fellowship (to JPB).

**References and recommended reading**

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


Fascinating demonstration of the hybridization chain reaction of designed small RNAs in vivo. The small RNAs are specifically triggered by cancer marker mRNAs to polymerize into long duplexes, which initiate an innate immune response that leads to cell death.


44. Cao S, Chen SJ: Predicting structures and stabilities for H-type pseudoknots with interhexil loops. RNA 2009, 15:696-706.


46. Bindewald E, Afonin K, Jaeger L, Shapiro BA: Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots. ACS Nano 2011, 5:9542-9551. Intriguing proposal that simple, heuristic design algorithms may be powerful for designing complex 3D topologies, even if they sacrifice the thermodynamic rigor and applicability to natural RNAs of classic modeling approaches.


59. RNA loop-loop interactions, a new adition to the in vivo design toolkit, can be multiplexed, work in tandem, and propagate signals via transcriptional cascades. A highlight is the quantitative and fine-grained recapitulation of a tandem RNA gate’s in vivo properties from the careful characterization of its components.


72. First instance of a community-wide blind trial in RNA structure prediction – a compelling synopsis of what can and cannot be currently modeled.


78. Application of core ideas in high-resolution protein structure prediction and design to RNA motifs, including a first step to ‘de novo’ design of new motifs.


86. First demonstration of RNA nanostucture creation in vivo, based on tile sequences from an automated design algorithm. The resulting scaffolds engage with metabolic enzymes in bacteria to enhance hydrogen production.


95. Uniquely systematic exploration of a classic puzzle in RNA design, the formation of tiles or squares.

10 Engineering and design


