PERSPECTIVES

Principles for targeting RNA with drug-like small molecules

Katherine Deigan Warner, Christine E. Hajdin and Kevin M. Weeks

Abstract | RNA molecules are essential for cellular information transfer and gene regulation, and RNAs have been implicated in many human diseases. Messenger and non-coding RNAs contain highly structured elements, and evidence suggests that many of these structures are important for function. Targeting these RNAs with small molecules offers opportunities to therapeutically modulate numerous cellular processes, including those linked to 'undruggable' protein targets. Despite this promise, there is currently only a single class of human-designed small molecules that target RNA used clinically — the linezolid antibiotics. However, a growing number of small-molecule RNA ligands are being identified, leading to burgeoning interest in the field. Here, we discuss principles for discovering small-molecule drugs that target RNA and argue that the overarching challenge is to identify appropriate target structures — namely, in disease-causing RNAs that have high information content and, consequently, appropriate ligand-binding pockets. If focus is placed on such druggable binding sites in RNA, extensive knowledge of the typical physicochemical properties of drug-like small molecules could then enable small-molecule drug discovery for RNA targets to become (only) roughly as difficult as for protein targets.

The vast majority of small-molecule therapeutics in clinical use target proteins. Successful protein-targeted drugs include small molecules that bind proteins expressed by humans (FIG. 1) or those expressed by bacteria, viruses and other infectious organisms. If small-molecule therapeutics could be extended to modulate RNA, the landscape of targetable macromolecules would be expanded by more than an order of magnitude. For example, only roughly 1.5% of the human genome encodes proteins[1,2]. Within the roughly 20,000 expressed human proteins, 10–15% of these are thought to be disease-related[3–5], meaning that disrupting or altering their activity is likely to have a therapeutically useful consequence. Within the group of disease-related proteins, many are termed 'undruggable', meaning that these molecules lack distinctive cleft-like motifs into which small molecules can bind with high specificity and affinity[5]. The net implication of this analysis is that de novo human-designed small-molecule drugs target a sliver (the 3.5% of drugged proteins) of a sliver (1.5% of the human genome that encodes proteins) of the human genome. Currently approved protein-targeted drugs interact therapeutically with fewer than 700 gene products[6], meaning that only 0.05% of the human genome has been drugged (FIG. 1).

RNA functions as a central conduit of information transfer in all biological systems[7,8] and, in principle, there are numerous opportunities for creating small-molecule therapeutics targeting RNA. If mRNAs could be targeted, protein gene products could be modulated by upregulating or downregulating translation efficiency or by altering mRNA abundance or stability. By directly targeting mRNAs, activities of proteins that are very difficult to drug or undruggable might be modulated before or during their biogenesis. In addition, a large fraction of the human genome (~70%) is transcribed into non-coding RNAs (RNAs that function directly as RNA)[9] (FIG. 1). One class of non-coding RNAs, microRNAs, are now well-validated therapeutic targets[10,11]. Humans probably produce more than 15,000 long non-coding RNAs[12], and a subset of these may eventually prove to be good drug targets. In summary, if a fraction of the tens of thousands of mRNAs and non-coding RNAs were targetable, the extent of the druggable human genome could increase considerably.

Strong proof of principle for RNA-targeted drugs has been provided by experiments and clinical trials using antisense oligonucleotides (several of which have gained regulatory approval), as well as preclinical studies with synthetic RNAs that redirect the cellular RNA interference (RNAi) machinery or that activate CRISPR-based systems[10,11,13–16]. Depending on the approach employed, it is possible to inhibit or to upregulate expression[13,14] of particular mRNAs and to inhibit non-coding RNA function by targeting specific RNA sequences. However, these nucleic-acid-based approaches involve large, often highly charged molecules and present delivery challenges. Given the choice between an oligonucleotide-based therapy and a (currently hypothetical) small-molecule-based therapy that targets RNA, the small-molecule therapy would be preferred in many cases.

A number of small molecules that alter RNA function have been identified, providing encouraging evidence that RNA-targeted small-molecule therapeutics could be developed. Some of these studies are based on molecules identified in natural systems. For example, many RNA riboswitch regulatory elements that bind diverse small-molecule metabolites and regulate gene expression have been characterized[17], and small-molecule antibiotics produced by numerous bacterial species bind ribosomal RNA and interfere with translation[18]. Furthermore, a growing number of synthetic monovalent and multivalent molecules that bind to secondary structure elements or repeat-containing RNAs have been shown to modulate biological function[19]. This progress is helping to address scepticism that RNA could be a viable small-molecule drug target, but key challenges remain to be addressed.
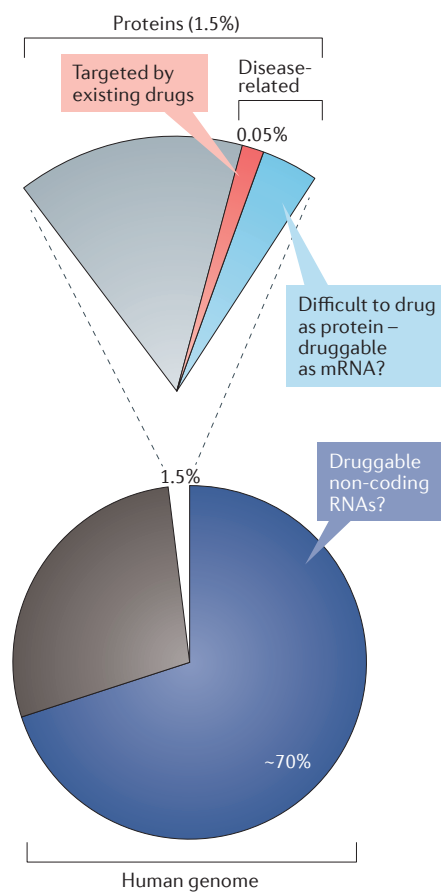
Figure 1 | **The potential RNA-targeted druggable genome.** Only a small fraction of the human genome has been successfully drugged to date. As shown in the pie chart, only ~1.5% of the genome encodes proteins (corresponding to ~20,000 proteins)[1,2]. As shown in the expanded pie segment above, an estimated 10–15% of proteins are thought to be disease-related[3–5] (~2,000–3,000 proteins; encoded by 0.2% of the genome). Currently approved drugs interact therapeutically with <700 of these proteins (encoded by 0.05% of the human genome)[6]. Targeting RNAs could expand on the proportion of the human genome that could be therapeutically targeted. Possible RNA targets include mRNAs that encode disease-related proteins that have been characterized as undruggable or difficult to drug (shown in light blue in the pie segment) and also non-coding RNAs that influence disease (corresponding to an unknown proportion of the ~70% of the genome that encodes non-coding RNAs).

Any discussion of how to target RNA with small molecules needs to immediately acknowledge that no one currently knows how to create drug-like, RNA-targeted small molecules in a repeatable and scalable way. There is currently only a single human-designed and approved drug class that functions by binding RNA

alone — the linezolid antibiotics, linezolid and tedizolid (FIG. 2). In our view, successful targeting of RNA in therapeutically useful ways with small molecules requires three distinct components: (1) identification of RNAs and RNA motifs with disease-related function, (2) use of screening approaches and libraries likely to identify drug-like molecules with appropriate pharmacological properties and (3) identification of and focus on RNA motifs with sufficient structural sophistication that make it likely that high affinity and specificity in small-molecule binding can be achieved. To date, most work has focused on therapeutically well-validated targets (component 1), and there has been a notable recent focus on screening against RNA targets with drug-like lead compounds (component 2). However — and critically in our view — considerations of target complexity and achievable selectivity (component 3) have been largely neglected. With this in mind, after overviewing the current state of small-molecule ligand discovery for RNA targets and some instructional examples, we present a set of hypotheses regarding how to target RNA with small molecules and guidelines for achieving this goal. This article is intended to be provocative, and we look forward to the feedback and responses that our ideas will engender.

Current status of RNA-targeted ligands

Inspection of well-researched classic reviews on RNA-targeted small molecules reveals that, up to 2012, efforts to find small-molecule ligands for RNAs mostly identified highly basic (and thus positively charged under physiological conditions) and planar molecules capable of intercalation between and stacking on RNA bases[20–22]. These molecules tend to bind to RNA with high affinity but low selectivity. These kinds of molecules often have poor pharmacological properties, including low cellular uptake and high toxicity. Thus, much of the early work was focused on molecules with decidedly undrug-like properties. The field of RNA-targeted ligand discovery is now changing rapidly, and many groups are now focusing on approaches that use strategies likely to identify molecules with plausible drug-like properties[23–26] (BOX 1).

We have compiled and analysed a collection of instructional molecules that target RNA and are active in a cell culture or animal model or in humans, focusing on molecules with reasonable drug-like properties, high potencies or both (FIGS. 2,3). We de-emphasized aminoglycosides

and molecules with likely nonspecific intercalation behaviour. In cases where a research group developed a series of related compounds, we included a single representative example. Compounds are grouped first by the type of RNA structure they target: tertiary structures involving multiple closely packed helices (FIG. 2a), irregular and usually bulge-containing secondary structures (FIG. 3a) or triplet repeats (FIG. 3b). Within each of these structure-based divisions, compounds are listed by their quantitative estimate of drug-likeness (QED) score[27] (BOX 1), with the most drug-like first (reading left to right and then down the figures). Overall, there are a limited number of examples in this collection, and several of the molecules have the same targets, such as survival motor neuron protein (SMN2) pre-mRNA, the HIV-1 transactivation response (TAR) bulged helix and the myotonic dystrophy type 1 (DM1) CUG triplet repeat. Nonetheless, this collection represents the current state of the field and provides a reasonable starting point for thinking about principles for targeting RNA specifically and potently with small molecules.

The four named molecules in FIG. 2a — linezolid, ribocil, branaplam and SMA-C5 — represent the current gold standards in biologically effective small molecules that target RNA. Each molecule targets a complex RNA structural motif, and each has a high QED score, but notably, each is a special case. Moreover, each of these molecules was identified in a phenotypic screen, and only after the screen was completed was it discovered that these molecules function by interaction with RNA. Thus, none of these examples provide clear guidance on how to intentionally target RNA with small molecules.

Linezolid is a broad-spectrum antibacterial agent that binds to the large subunit RNA of the ribosome and appears to interfere with correct tRNA positioning[28,29]. Linezolid is a special case because ribosomes are highly abundant in cells and because small molecules targeting the ribosome need to achieve only modest binding affinity; the linezolid–ribosome equilibrium dissociation constant (K_d) is approximately 20 μM. Merck's ribocil binds to riboflavin riboswitches, which function as RNA-mediated regulators of gene expression in bacteria. Ribocil has many attractive features but is unlikely to be pursued further because bacteria rapidly develop resistance to this drug[30,31]. Ribocil bears no chemical similarity to the natural metabolite ligand

and represents a major accomplishment in human-initiated RNA-targeted ligand development. Nonetheless, ribocil binding to a riboswitch is a special case because the drug binds in the same pocket as the natural flavin mononucleotide ligand, and thus this RNA is predisposed to bind a small molecule. The binding sites for Novartis's branaplam (originally called NVS-SM1) and PTC Therapeutics' chemotypically similar SMA-C5 have been publicly shared in a qualitative way[32,33]. These molecules both appear to bind at the RNA duplex formed between the U1 RNA and the target (SMN2) pre-mRNA and are probably stabilized either indirectly or directly by protein components

of the U1 small nuclear ribonucleo-protein particle (snRNP)[32–34]. Compounds from these programmes have been tested in clinical trials[35]. Discoveries of these molecules were impressive achievements, but the target site pocket includes protein components, so neither is strictly an RNA-targeted drug.

The rest of the instructional compounds show distinct and interesting features (FIGS. 2,3). Notably, these molecules were discovered using a wide variety of strategies that were intentionally focused on identifying RNA targets, and a subset are broadly drug-like, a feature emphasized by other recent

reviewers[24–26]. Discovery strategies included high-throughput screening[30,32,36–39], use of focused libraries[40–42], structure-inspired design[43–45], fragment-based approaches[46] and computational modelling[47,48]. Both compact and multivalent molecules are represented. Together, the molecules in FIGS. 2 and 3 strike us as reasonable validation of the potential of RNA as a small-molecule drug target. However, with the exceptions of the four named compounds in FIG. 2a, these molecules show deficiencies in one or more features of a conventional drug-like molecule.

The relationship between drug-likeness and potency for these instructional ligands has two intriguing characteristics. First, most (five out of six) of the molecules targeting higher-order tertiary structures are drug-like (as inferred by comparing their QED score with those of orally available drugs), independent of their potency (FIG. 2b, circles and green shading). This observation suggests that the ability to identify RNA-targeting small molecules with good drug-like properties is interrelated with the choice of RNA target. We explore this idea in depth below. Second, for compounds targeting RNA secondary structures or repeat sequences, it has been difficult to achieve good drug-likeness and high potency simultaneously. Indeed, for molecules targeting secondary structures, drug-likeness and RNA binding affinity are anticorrelated: high potencies are achieved only with the most undrug-like molecules (FIG. 2b, squares, diamonds and dashed line). Moreover, in the relatively few cases where authors have examined a substantial number of molecules in a ligand discovery programme, the observed structure–activity relationship (SAR) is often relatively flat[39,40,46,49], indicating that it has proved difficult to improve the potency of a lead molecule. Thus, despite considerable effort and creativity, most compounds identified to date do not have characteristics of known potent bioavailable small-molecule drugs and do not target complex RNA sites as do known high-specificity drugs.

We emphasize that we are focusing on targeting RNA with small molecules, which we think requires the identification of a cleft-containing or pocket-containing RNA (see below). Our analysis does not rule out the possibility that low-complexity sites or large multivalent ligands might prove therapeutically useful. In the protein-targeting field, these compounds would be roughly the analogues of those drugs that target protein–protein interfaces[50]
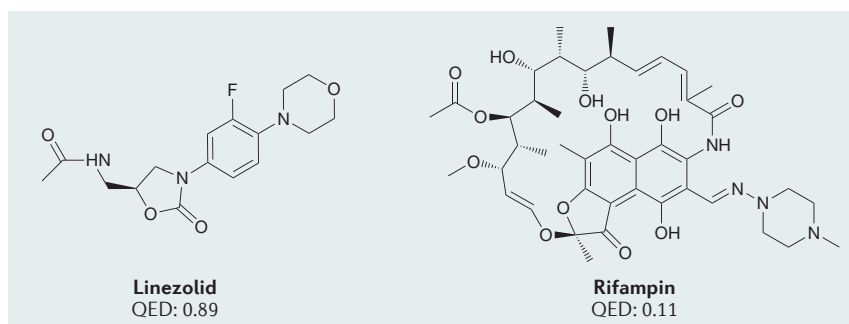

Box 1 | Drug-likeness

Clinically useful small-molecule drugs have physicochemical properties that tend to fall within a narrow range of possible values. Compounds that have molecular features consistent with these properties are said to be drug-like. Emphasizing drug-likeness is important because this characteristic provides a useful shorthand for molecules that have properties that are desirable in a drug to be given orally to humans, including solubility, cell and tissue permeability, metabolic stability and lack of toxicity.

Lipinski's rule of five is a well-known set of criteria used to evaluate the drug-likeness of a molecule. These rules were derived from the retrospective analysis of known drugs and drug candidates[96]. Lipinski's rules emphasize that molecules intended to become orally bioavailable drugs should not be too large (molecular mass less than 500 Da), too polar or too hydrophobic (fewer than five hydrogen-bond donors and a logP, the octanol–water partition coefficient, less than five). The emphasis on five serves as a mnemonic by which to remember the rules.

The idea of a privileged chemical space has recently been extended to quantify, in a single term, the overall drug-likeness for an orally bioavailable molecule. This approach allows some cut-offs (for example, one or more of Lipinski's rules) to be broken if the overall molecular quality of a compound is high. In this work, we emphasize the quantitative estimation of drug-likeness (QED) scores[27]. QED scores range from zero (all properties unfavourable) to one (all properties favourable). QED is a continuous function that allows the overall drug-likeness of a molecule to be summarized in a single term and that shows good discriminatory power and correlation with intuitive chemical aesthetics cultivated by medicinal chemists[27,97]. Bioactive molecules with high and low QED scores are shown in the figure. The median QED score for current orally bioavailable human drugs is approximately 0.65, and 75% of orally available drugs have QED scores ≥0.5.

The concept of drug-likeness and the specific parameterization of QED are, of course, primarily based on small-molecule drugs that target proteins. Some successful drugs, especially antibiotics and viral protease inhibitors, have quite low QED scores[27,97], so a low drug-likeness profile does not rule out the potential usefulness of a small-molecule drug candidate. It is also possible that RNA-targeted drugs will follow a different set of principles from those that target proteins. There is some evidence that molecules that bind RNA have different structural[24] and kinetic[98] properties, on average, than small molecules that bind proteins. Nonetheless, drug-likeness is strongly influenced by bioavailability and human physiology, which are the same for all drugs. Thus, we suggest that RNA-targeted drugs will ultimately follow roughly the same rules for drug-likeness as their currently better-characterized protein-targeting cousins.
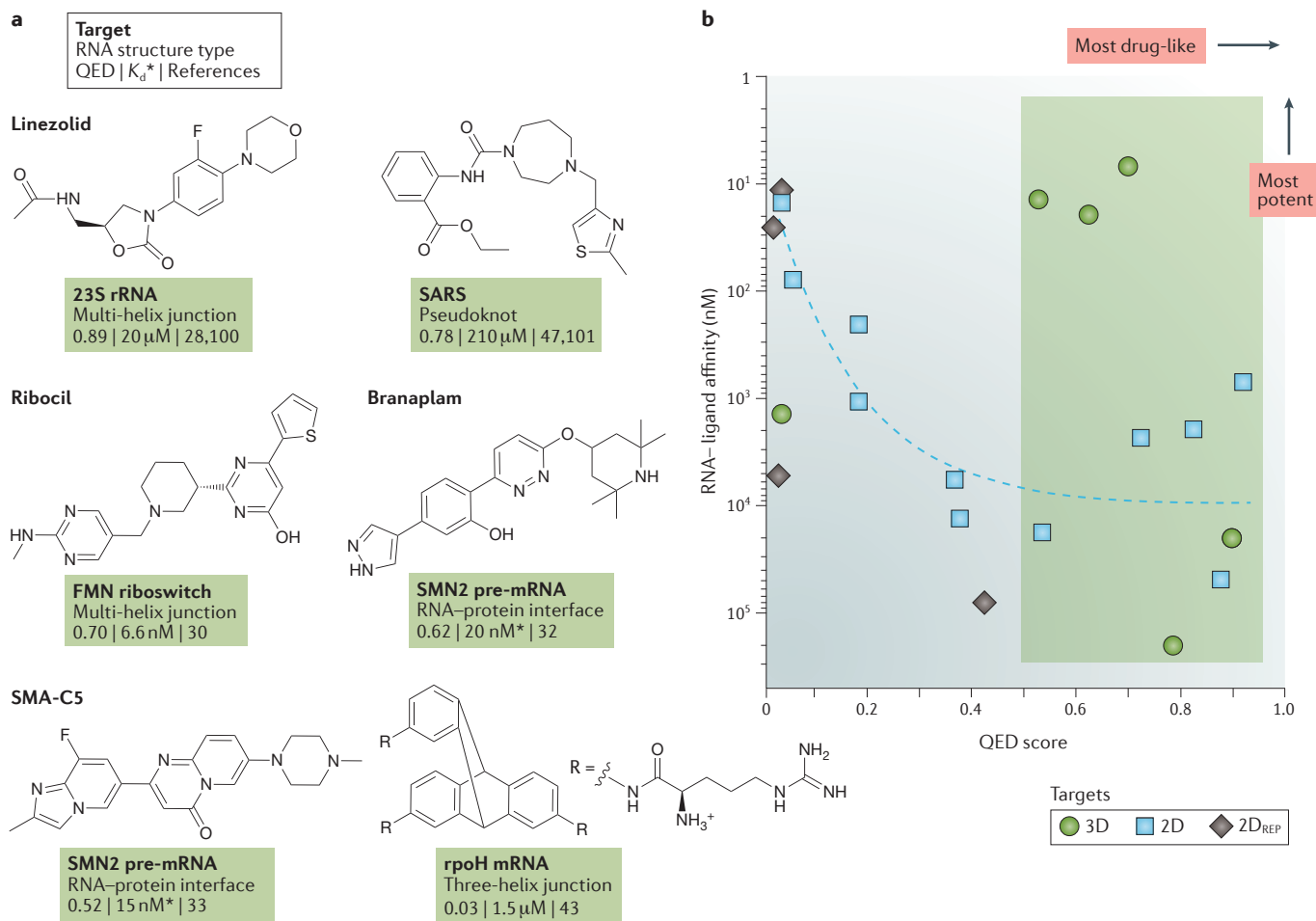

Linezolid
QED: 0.89

Rifampin
QED: 0.11

Figure 2 | **Instructional examples of bioactive small molecules that bind to RNA. a** | Molecules that bind RNA tertiary structures. Molecules are listed by their quantitative estimate of drug-likeness (QED) score (BOX 1), with the most drug-like molecules towards the top and towards the left if the molecules are on the same row. The equilibrium dissociation constant (K_d) is listed for most molecules; if no K_d was available, the median inhibitory concentration (IC_{50}) is provided instead (indicated by an asterisk). Molecular properties were calculated with SilicosIT. **b** | Relationship between QED scores and RNA binding affinity. Molecules targeting tertiary structures (shown in this figure; 3D targets) and those targeting secondary and repeat structures (from FIG. 3; 2D and 2D$_{REP}$ targets) are shown as circles, squares and diamonds, respectively. The green box highlights the region where most (five out of six) molecules that target RNA tertiary structures fall. The dashed blue curve highlights a trend for molecules targeting RNA secondary structures and repeat sequences: molecules with greater potency tend to have lower drug-likeness. FMN, flavin mononucleotide; rRNA, ribosomal RNA; SMN2, survival motor neuron protein.

and constitute 'beyond rule-of-five' (BOX 1) molecules[51,52], respectively. Targeting pathogenic repeat expansion sequences (for example, the CUG and CAG repeats characteristic of DM1 and Huntington disease, respectively) might represent an intriguing special case. Although trinucleotide repeat sequences have low complexity and do not appear to contain an obvious ligandable pocket (see below), it is possible that cooperative binding or in situ assembly of ligands driven by the supra-secondary structure of the repeating sequence might afford critical specificity and selectivity[19,53]. Any of these strategies may prove fruitful but can generally be expected to have a notably increased degree of difficulty.

Hypotheses for RNA-targeted drugs

RNA-targeted drugs should probably look like conventional protein-targeted drugs.
One question that arises immediately when thinking about targeting RNA with small molecules is whether the structural differences between RNA and proteins render RNA less druggable than proteins. RNA is composed of only four primary nucleotide building blocks and is much more highly charged and more hydrophilic than a typical protein. We suggest that focusing on these global differences misses the point and that the key issue in drugging RNA is whether RNA can form specific binding sites capable of binding drug-like molecules. Current evidence suggests that bacterial[17,54], viral[55–59] and mammalian (both coding

and non-coding)[60–66] RNAs fold back on themselves to form complex structures. We argue below that RNAs with complex structures also tend to contain pockets with sufficient structural sophistication to allow specific and high-affinity binding by small molecules.

That RNA binders can be drug-like is demonstrated by the analysis of linezolid[28,29] and ribocil[30,31] (FIG. 2a). Linezolid and ribocil are both unremarkable molecules from the point of view of conventional medicinal chemistry. Both are consistent with Lipinski's rule of five (BOX 1), both have low total polar surface areas (tPSAs) consistent with good membrane permeability and neither contains any red flags for toxicity. Most critically, both bind to structural 'pockets'
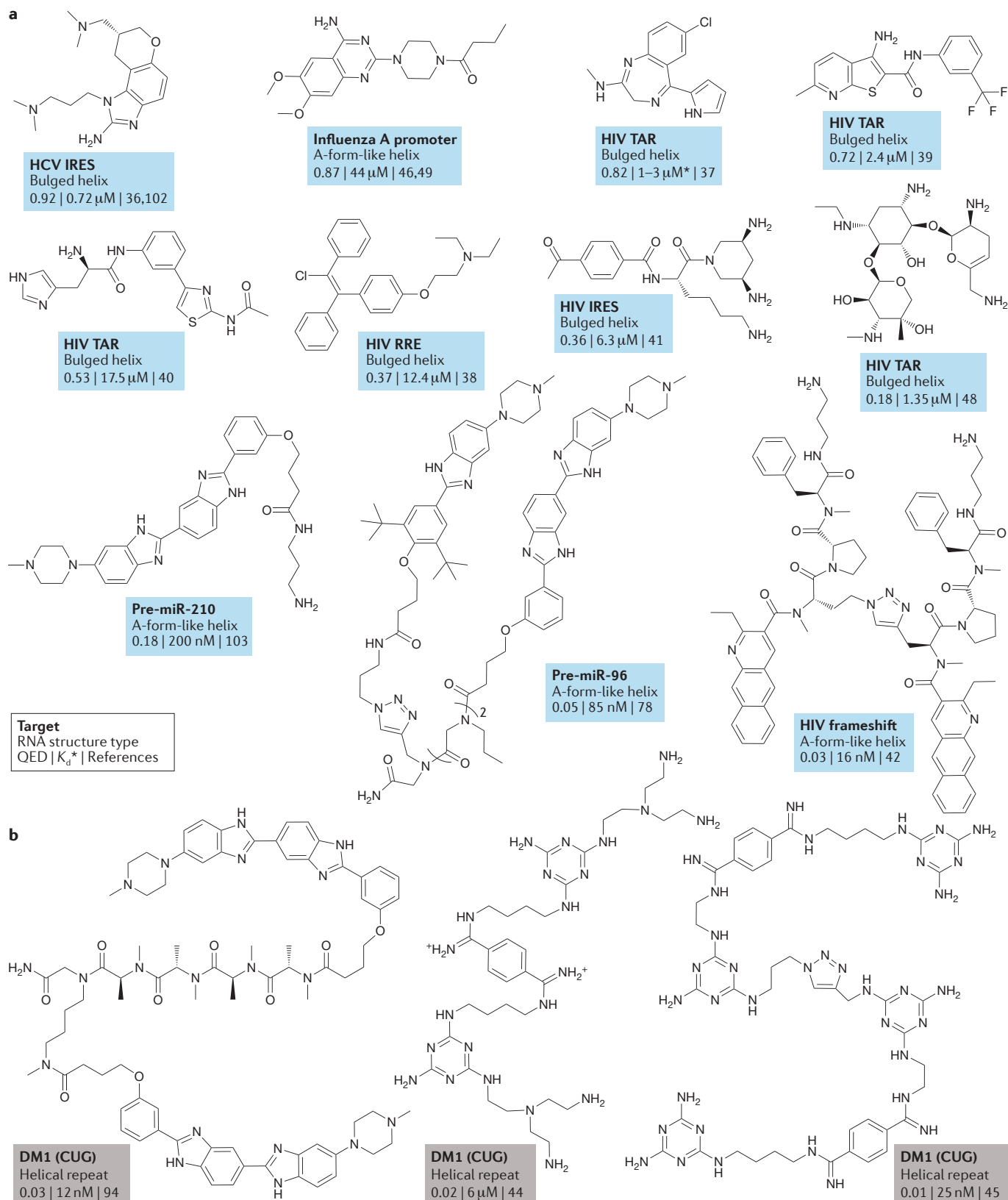
a

HCV IRES
Bulged helix
0.92 | 0.72 µM | 36,102

Influenza A promoter
A-form-like helix
0.87 | 44 µM | 46,49

HIV TAR
Bulged helix
0.82 | 1–3 µM* | 37

HIV TAR
Bulged helix
0.72 | 2.4 µM | 39

HIV TAR
Bulged helix
0.53 | 17.5 µM | 40

HIV RRE
Bulged helix
0.37 | 12.4 µM | 38

HIV IRES
Bulged helix
0.36 | 6.3 µM | 41

HIV TAR
Bulged helix
0.18 | 1.35 µM | 48

Pre-miR-210
A-form-like helix
0.18 | 200 nM | 103

Pre-miR-96
A-form-like helix
0.05 | 85 nM | 78

HIV frameshift
A-form-like helix
0.03 | 16 nM | 42

Target
RNA structure type
QED | K_d* | References

b

DM1 (CUG)
Helical repeat
0.03 | 12 nM | 94

DM1 (CUG)
Helical repeat
0.02 | 6 µM | 44

DM1 (CUG)
Helical repeat
0.01 | 25 nM | 45

Figure 3 | **Instructional examples of bioactive small molecules that bind RNA secondary structures.** Molecules that bind RNA secondary structures (part **a**) and trinucleotide repeats (part **b**). Molecules are listed by their quantitative estimate of drug-likeness (QED) score, with the most drug-like molecules towards the top and towards the left if the molecules are on the same row. The equilibrium dissociation constant (K_d) is listed for most molecules; if no K_d was available, the median inhibitory concentration (IC$_{50}$) is provided instead (indicated by an asterisk). HCV IRES, hepatitis C virus internal ribosome entry site; K_d, equilibrium dissociation constant; pre-miR, pre-microRNA; RRE, Rev response element; TAR, transactivation response.
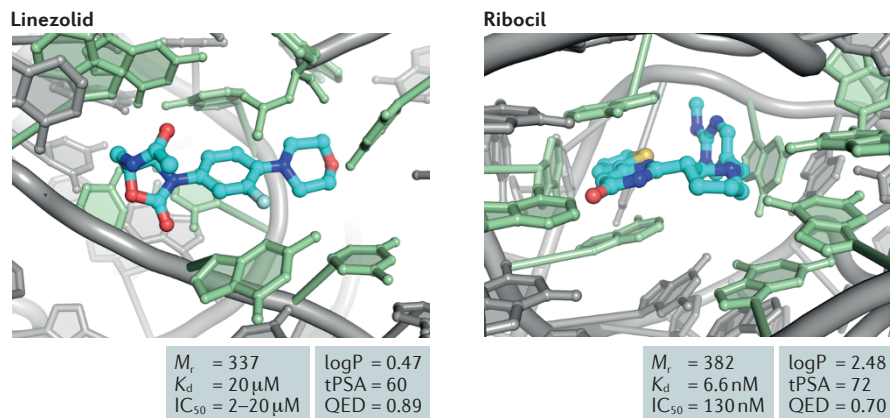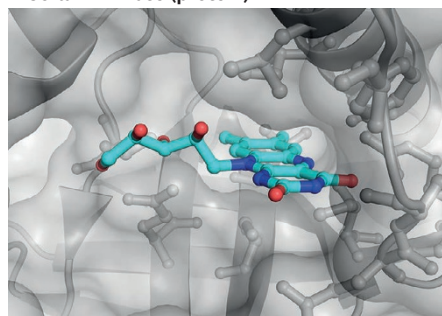
Linezolid

M_r = 337	logP = 0.47
K_d = 20 µM	tPSA = 60
IC_{50} = 2–20 µM	QED = 0.89

Ribocil

M_r = 382	logP = 2.48
K_d = 6.6 nM	tPSA = 72
IC_{50} = 130 nM	QED = 0.70

Figure 4 | **Drug-like molecules specifically targeting RNA.** Both linezolid and ribocil have high quantitative estimate of drug-likeness (QED) scores[27] and are bound within cleft-like sites[29,30] composed of first-sphere nucleotides in contact with the ligand (green) supported by second-sphere nucleotides (grey). These motifs have high information content. IC_{50}, median inhibitory concentration; K_d, equilibrium dissociation constant; logP, octanol–water partition coefficient; M_r, relative molecular mass; tPSA, total polar surface area.

in their target RNAs — the large subunit bacterial ribosomal RNA and a bacterial flavin mononucleotide (FMN) riboswitch, respectively (FIG. 4). Both ligand-binding sites are complex in the sense that they feature both inner-sphere nucleobases that are in direct contact with the ligand and a set of outer-sphere nucleotides that further rigidify and support the structure of the ligand-binding site[29–31,67] (FIG. 4). These factors — direct intimate contact with inner-sphere nucleotides and support by a complex 3D network — make these binding sites similar to the many pockets in protein targets that bind therapeutically useful small molecules.

Comparison of how a single ligand, riboflavin, can be recognized by both protein and RNA further supports the ability of RNA to recognize diverse physical features of a ligand (FIG. 5). Although riboflavin has a tPSA value that is high and other features that are moderate outliers relative to most drugs, it is, in fact, clinically useful[68]. Riboflavin contains three distinctive molecular entities: a nitrogen-rich pteridine-2,4-dione two-ring system, a hydrophobic dimethylbenzene ring and a ribose group. Riboflavin kinase protein[69] and a flavin RNA riboswitch[70] recognize riboflavin in similar but not identical ways. Both protein and RNA form multiple hydrogen bonds with the pteridine ring system, both make multiple van der Waals contacts with the dimethylbenzene ring, both form hydrophobic or stacking interactions from above and below the plane of the three-ring system, and both form hydrogen bonds with a hydroxyl group in the ribose chain (FIG. 5). This analysis shows

that, in principle, RNA is capable of making specific molecular interactions with a wide variety of functional groups and ligand surfaces. The most consistent feature of the examples discussed thus far is the ability of the RNA target to form a pocket into which a ligand fits intimately and forms multiple specific interactions.

Good RNA targets should have high information content. An overarching concern in targeting RNA is how to achieve both high binding affinity and high selectivity in the cellular context in which there are many similar RNA motifs. To date, a substantial fraction of the effort devoted to targeting RNA has focused on simple RNA structures consisting of base-paired helices or irregular and bulge-containing duplexes (FIG. 3). By contrast, RNAs that interact with drug-like RNA-targeting molecules, such as linezolid, ribocil and riboflavin, bind to complex structures and are encapsulated in cleft-like sites reminiscent of ligand-binding sites in proteins (FIGS. 2,4,5).

The complexity of an RNA can be characterized in terms of its information content, which is measured in bits[71–73] (BOX 2). One bit is the amount of information required to distinguish between two possibilities. RNA secondary structures can now be modelled with high accuracy on the basis of sequence covariation information (when many alignable sequences are available)[74] and using chemical probing strategies, such as selective 2′-hydroxyl acylation analysed by primer extension (SHAPE), enabling a useful estimation of information content for individual RNAs[75,76].

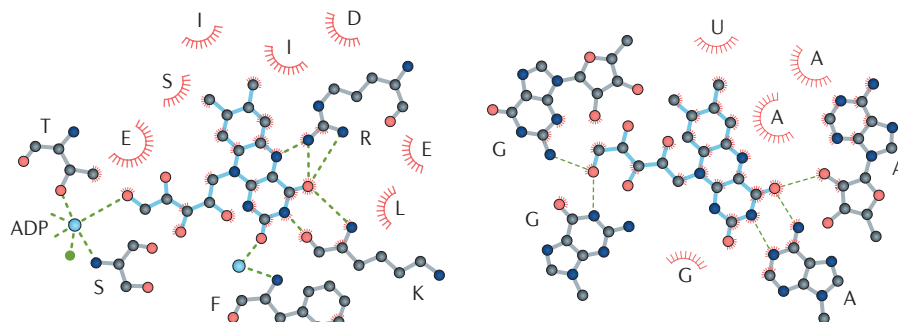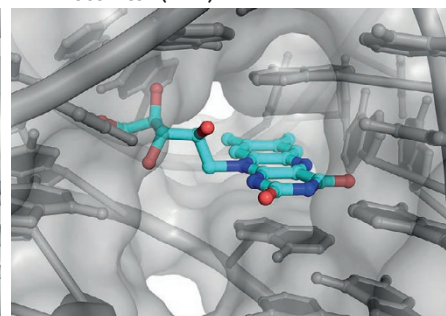**Riboflavin kinase (protein)**

FMN riboswitch (RNA)

Figure 5 | **RNA and protein recognition of the same riboflavin ligand.** Molecular interactions of riboflavin (blue) with riboflavin kinase protein (Protein Data Bank (PDB) identifier: inb9)[69] (left) and flavin mononucleotide (FMN) riboswitch RNA (PDB ID: 3f4g)[70] (right). Hydrogen bonds are shown with green dashed lines, and van der Waals and stacking interactions are emphasized with red arcs. Water molecules and metal ions are shown as cyan and green spheres, respectively. Riboflavin has a molecular mass of 376 Da, an octanol–water partition coefficient of –0.76, a total polar surface area of 162 and a quantitative estimate of drug-likeness score of 0.33. Bottom images were initially drafted with LigPlot+.

The number of bits in an RNA target is approximated by summing the information in individual structure elements (BOX 2). Analysis of the information content is especially attractive for RNA, as this metric is relatively easy to calculate, given a secondary structure model and modest additional biochemical information, and provides an impartial way to evaluate current and future targets. A generic six base pair RNA helix has an information content of roughly 9 bits, whereas the linezolid, ribocil and branaplam binding sites probably exceed 50 bits. The information content of a ligand-binding site reflects both inner-sphere and outer-sphere interactions (see FIG. 4), such that nucleotides do not necessarily have to touch an RNA ligand directly to affect the information content of a site[31,67,77].

Perhaps the most instructive illustration to date of interrelationships between RNA complexity and ligand binding comes from prior studies focusing on examining RNA sequences that have been identified by in vitro selection to bind GTP[73]. Interactions between GTP and these RNAs probably involve primarily the guanosine entity[77] (molecular mass = 283 Da; QED score = 0.40). The GTP binding affinity shows a strong correlation with the complexity and information content of the RNA target (FIG. 6, circles). For GTP, increasing the information content of the target RNA by 10 bits has been experimentally observed to yield a roughly 10-fold increase in binding affinity[73], which we term the 10→10 relationship for brevity. This 10→10 relationship also holds for another ligand for which affinity versus target site complexity data are available, targaprimir-96 (FIG. 6, diamonds; data from REF. [78]). The targaprimir-96 line is offset from the GTP line because targaprimir-96 has a higher affinity for a given information content than GTP, probably because of its larger size.

The approximate 10→10 relationship identified in prior studies[73] provides a framework for interpreting recent efforts to target RNA with small molecules. For example, ribocil and branaplam both have low-nanomolar affinities and bind to RNA and RNA–protein sites, respectively, with high information content (FIG. 6, triangles). Similarly, diverse efforts focused on identifying ligands against targets with low-to-moderate information content — including a stem-loop (20 bits), the TAR hairpin (22 bits) and an asymmetric loop in the hepatitis C virus internal ribosome entry site (HCV IRES) element (26 bits) — have

resulted in molecules with affinities in the 1 to 100 μM range. These micromolar binding affinities for low-molecular-mass ligands agree closely with expectations based on the 10→10 relationship (FIG. 6, squares). All these low-to-moderate-affinity target sites fall below the 30 bit threshold, which is roughly the amount of information required to uniquely specify an RNA target in the context of the human transcriptome (BOX 2).

The critical conclusion from this analysis of information content is that current efforts at targeting RNA with drug-like molecules are doing about as well as expected. Efforts to target relatively simple structures have achieved modest potency


Box 2 | Information content of RNA targets

The amount of information required to specify an RNA target can be defined by its structural complexity, measured in bits[71,72]. One bit corresponds to the information required to distinguish between two possibilities — for example, between purine and pyrimidine bases. Two bits are therefore required to define an invariant or absolutely required nucleotide in an RNA element. Nucleotide positions with no sequence constraints correspond to zero bits. The information required to define a conserved base pair can be subtle because a degree of mismatched or wobble pairing is often allowed[73]. A canonical Watson–Crick base pair, allowing G–U, corresponds to roughly 2 bits. For many RNA motifs, a short helix is required to stabilize a given structure, but the specific sequence, length and degree of allowed mismatches can vary. We estimate that a generic helix requires the equivalent of six base pairs, half of which could be wobble or mismatch pairs, corresponding to $(3 \times 2) + (3 \times 1) = 9$ bits.

The total information content for a given RNA motif or drug target is given by summing the information content for every position in the target element (see below). RNA structural motifs can be modelled with good accuracy based on chemical probing data, especially using the selective 2′-hydroxyl acylation analysed by primer extension (SHAPE) strategy. SHAPE experiments use reagents that form covalent adducts with the 2′-hydroxyl group in RNA, such that the degree of reactivity reports local nucleotide flexibility[55,75,76]. These SHAPE reactivities can then be interpreted using RNA folding software to define the secondary structure of an RNA target, where regions with low versus high reactivities tend to be base-paired or single-stranded, respectively. The total number of bits in an RNA target site can be estimated from this secondary structure when taken together with additional biochemical data from, for example, comparative sequence analysis[71,73], biochemical experiments[72] or inspection of interactions observed in high-resolution studies. Applying these rules to the recognition of the HIV-1 transactivation response (TAR) RNA element by short transactivator of transcription (Tat)-derived peptides[99], scored based on the information content of structural elements recognized by the peptides, reveals that this interaction site contains about 22 ± 5 bits of information, as shown in the figure. Boxed nucleotides show sequence specificity; brackets indicate elements where structure, but not sequence, is the primary constraint.

There are a few useful reference points to help calibrate the distinctiveness of an RNA motif with a given amount of information content. For example, 10 bits is equivalent to an expected occurrence of 1 in 1,024, whereas 30 bits corresponds to a motif that will occur roughly once in a billion nucleotides (see figure). Thus, 30 bits are required to be reasonably confident that an RNA target is unique in the human transcriptome.
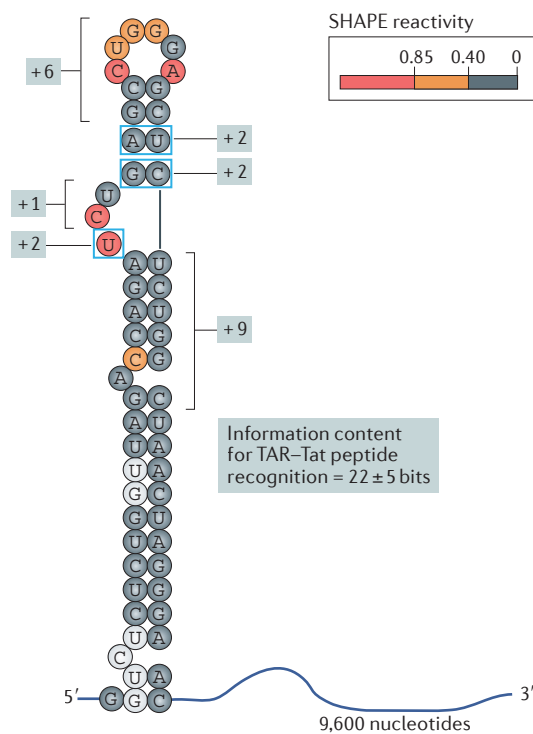
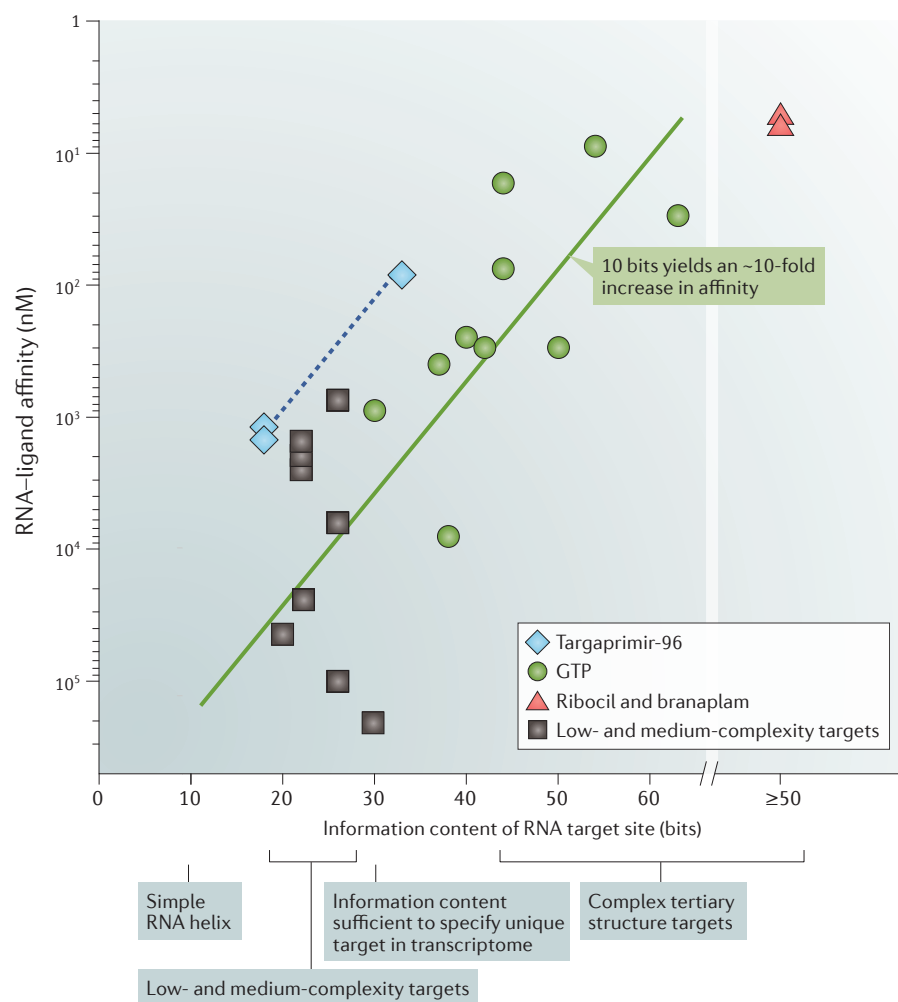| Information content in RNA motif	Expected occurrence (in nucleotides)
10 bits	Once in 1,024 (2^{10})
20 bits	Once in ~1 million (2^{20})
30 bits	Once in ~1 billion (2^{30})

Figure 6 | **Relationship between information content and specificity and binding affinity.** The solid and dashed lines show the relationship between RNA target information content and binding affinity for GTP (circles) and targaprimir-96 (diamonds) (based on data from REFS [73,78]). Simple helices are assigned 9 ± 2 bits of information. Ribocil[30] and branaplam[32] are estimated to have ≥50 bits of information and are placed on a separate scale on the right. Binding data for low- and medium-complexity targets (squares) are from the following: influenza A promoter stem-loop (20 bits)[49], transactivation response (TAR) (22 bits)[37,39,40,48], hepatitis C virus internal ribosome entry site (HCV IRES) bulged-stem (26 bits)[36,41,102] and severe acute respiratory syndrome (SARS) pseudoknot (30 bits)[47,101].

and selectivity. Thus, it is not necessarily the methods used or the small-molecule molecular frameworks that are limiting the RNA-targeted drug discovery field. We argue that the field has been limited by the choice of targets without sufficient complexity and information content.

Relationship between RNA information content and quality of the ligand pocket.

Our hypothesis is that focusing on RNA target sites with high information content is important both in order to identify sites that can be targeted with the roughly one in a billion (30 bits) selectivity necessary to identify a site uniquely in a complex transcriptome and in order to achieve high

potency. Here, we discuss a third critical advantage of targeting complex sites: complex sites contain high-quality pockets into which ligands and drugs can bind.

We examined the ligandablity of representative targets in terms of their pocket quality. Pocket prediction algorithms use a combination of sequence, structure and ligand information to provide estimates of potential binding sites within a target[79,80]. We used an algorithm, PocketFinder[81], to estimate the quality of potential ligand-binding pockets formed from a representative set of RNAs. The pocket-finding algorithm defines a ligand-binding envelope based on the transformation of the Lennard-Jones

potential determined from the target surface and does not require pre-existing knowledge of a putative ligand. This pocket-finding strategy predicts the location of a pocket and estimates its shape and size (FIG. 7). For example, the pockets in riboflavin kinase and the FMN riboswitch are distinguished by both their volumes (>160 Å³) and their 'buriedness' (>0.75; where 0.5 indicates the surface is flat and 1.0 corresponds to a completely buried pocket); these values indicate large cleft-like regions (FIG. 7a). Similarly, the pocket in the 23S ribosomal RNA into which linezolid binds is large and highly buried (FIG. 7b).

In strong contrast to these characteristics, many RNA motifs targeted by small molecules to date do not contain what would be considered either intuitively or quantitatively notable pockets. For example, helices containing CAG or CUG trinucleotide repeats (and almost certainly other repeat sequences) do not contain features that meet common definitions of a ligand-binding pocket. RNA motifs composed primarily of secondary structures, including regulatory stem-loops, microRNAs and the HIV TAR and Rev response element (RRE) stem-loops, contain small or shallow pockets (FIG. 7c). It is possible that these kinds of motifs can be targeted, but doing so will probably prove challenging in much the same way as has targeting the shallow grooves that characterize many protein–protein interactions[51].

It is straightforward to identify RNAs that adopt high-quality and probably targetable structures. The simplest motifs that have complex, high-information-content structures are multi-helix junctions. Analysis of the structures of three-helix junctions[82,83] reveals that these motifs can have information contents exceeding 30 bits and contain high-quality pockets (FIG. 7d). These pockets are simply consequences of having a complex structure. Multi-helix junctions occur widely in large RNAs, many are known to overlap or occur within functional motifs, and a subset forms well-defined, stable structural elements. Pseudoknots are a simple motif formed when unpaired nucleotides in a loop pair with a region outside the stem that closes the loop[84]. Many pseudoknots contain high-quality pockets, as illustrated by the structure of the simian retrovirus type 1 (SRV-1) pseudoknot[85], which forms the common H-type pseudoknot (FIG. 7e). Pseudoknots are thought to be relatively rare in large RNAs but appear to be over-represented in functionally important regions[86].
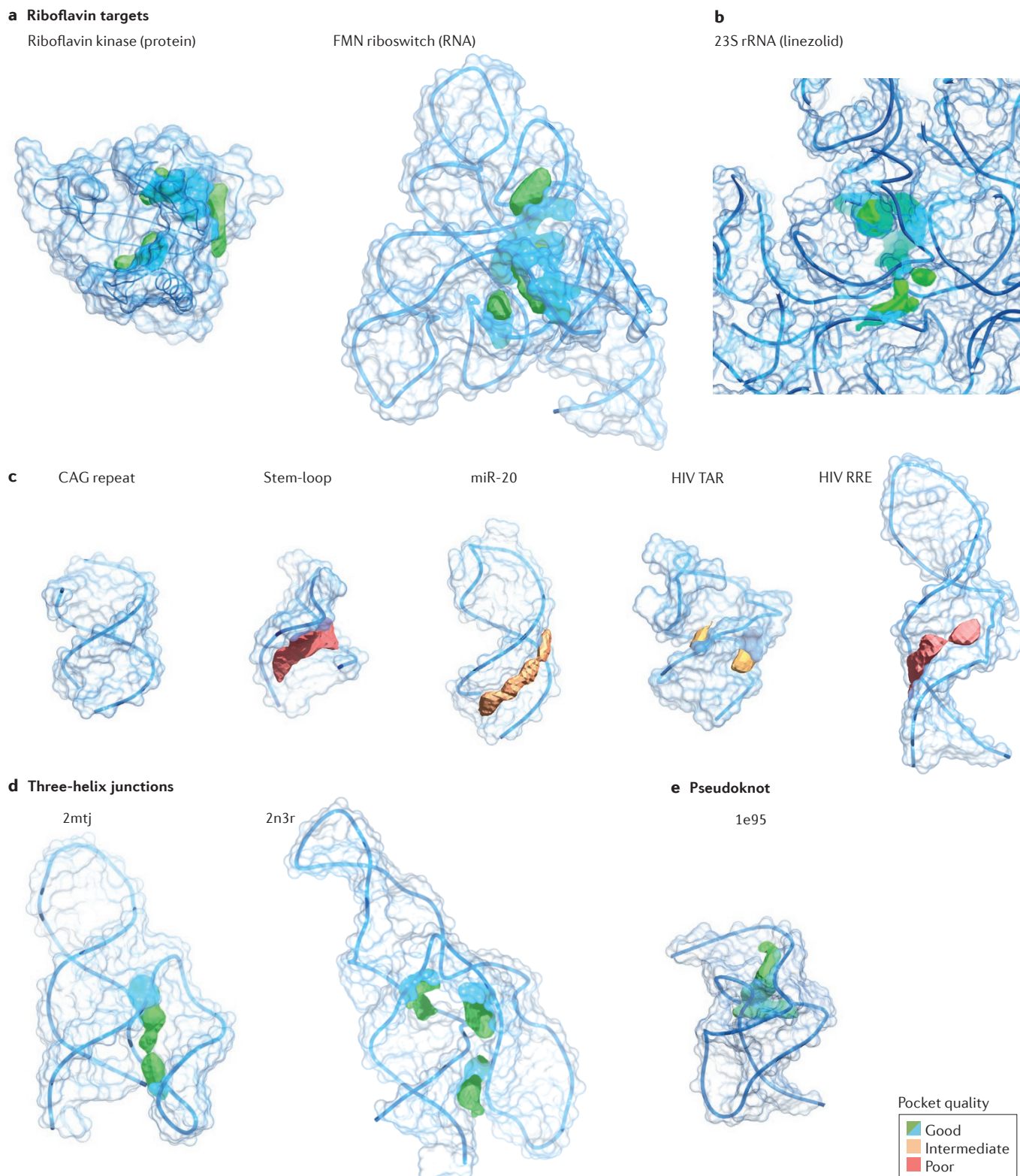
a **Riboflavin targets**

Riboflavin kinase (protein)   FMN riboswitch (RNA)

b   23S rRNA (linezolid)

c   CAG repeat   Stem-loop   miR-20   HIV TAR   HIV RRE

d **Three-helix junctions**

2mtj   2n3r

e **Pseudoknot**

1e95

Pocket quality

■ Good
■ Intermediate
■ Poor

Figure 7 | **Pocket analysis of current and aspirational RNA targets.** Structures are coloured by pocket quality, with larger and more buried pockets ranking higher. **a** | Riboflavin binding pockets for protein (Protein Data Bank (PDB) identifier: 1nb9)[69] and RNA (PDB ID: 3f4g)[70] targets. Macromolecular targets are the same as those shown in FIG. 5. **b** | Pocket for linezolid in the *Escherichia coli* 23S ribosomal RNA (rRNA) (PDB ID: 3dll)[29]. **c** | Ligand-binding pockets in representative low-to-medium-complexity targets including a CAG helical repeat (PDB ID: 4j50)[104], stem-loop (PDB ID: 1oq0)[105], microRNA (miR; PDB ID: 2n7x)[106] and HIV transactivation response (TAR) (PDB ID: 1qd3)[107] and Rev response element (RRE) (PDB ID: 1i9f)[108]. **d,e** | Potential ligand-binding pockets in high complexity sites: three-helix junctions (PDB ID: 2mtj and 2n3r)[82,83] and the simian retrovirus type 1 (SRV-1) pseudoknot (PDB ID: 1e95)[85]. Pocket qualities were calculated using Pocket-Finder[81], as part of the MolSoft package. FMN, flavin mononucleotide.

Provocative guidelines

We first reiterate that no one currently knows how to target RNA with drug-like small molecules in a scalable and reproducible way. Proof-of-principle examples exist (FIGS. 2,3), but the most drug-like of these molecules were discovered based on phenotypic assays and were only later found to target RNA. The molecules discovered to date via strategies intentionally designed to target RNA generally suffer from trade-offs in drug-likeness and potency due to low information content and low pocket quality. Nonetheless, current examples are supportive of the fundamental targetability of RNA. We suspect that a wide variety of screening strategies — including phenotypic assays, high-throughput screening, fragment-based screening, small-molecule microarrays, other biophysical partitioning approaches, and structure-inspired and computationally assisted design — will eventually prove useful. In this light, we would like to propose the following guidelines for future efforts directed at RNA-targeted drug discovery.

- Focus on RNA motifs of sufficient complexity that they will be unique, able to form high-quality pockets and capable of binding small-molecule ligands with high affinity. We think that good RNA targets should have 30 bits and a pocket. Such motifs are likely to be common in large RNAs. We posit that few good targetable pockets exist in the absence of high information content.

- Be cautious of early attempts to define specialized privileged chemical scaffolds for targeting RNA. These might exist, but a lot more information is likely to be required before promising scaffolds can be defined. Instead, emphasize rules derived from decades of development of protein-targeted drugs. Most are likely to apply to RNA-targeted drugs as well.

- De-emphasize 'lessons' from the ribosome. Due to their uniquely high concentration in cells, ribosomal RNAs are a special case. The only engagement expected for a ligand with micromolar affinity is a target that is present at micromolar concentration. Drugs targeting ribosomal RNAs thus achieve a unique kind of specificity by binding relatively weakly to the ribosome.

- Be cautious in interpreting the results of studies focused on compounds that are highly basic, intercalating, strongly stacking and/or highly hydrophobic. These kinds of compounds may bind RNA with high affinity but are unlikely to show strong discrimination between target and off-target sites and are likely to be difficult to improve in a medicinal chemistry campaign.

- Target RNA–protein interactions or RNA elements stabilized by protein binding, as branaplam and SMA-C5 appear to do[32,33]. It is easy to imagine that recognition of a target-specific RNA element in conjunction with a general protein component of cellular metabolism or regulatory network might be a highly productive strategy for achieving both specificity (via a selective RNA-targeted component) and affinity (owing to the high information content of a protein interface).

- Use methods that enable rapid identification of RNA motifs with high-information-content structures and high-quality pockets. A clear advantage of RNA over proteins is that tools, such as quantitative chemical probing approaches[55,75,87], already exist that make it possible to identify high-information-content[72,73] and structurally well-determined (low-entropy) target sites[54,55], to examine the influence of the cellular environment on a specific RNA structure and to ensure that simplified screening constructs recapitulate the native RNA motif.

- Develop additional tools and new refinements of current tools, including advanced structural biology and modelling strategies for RNAs with complex structures[88–90]. These technologies can then be leveraged to identify and rapidly characterize the subset of complex RNA structures with high-quality ligand-binding pockets and to rapidly assess the impact of ligand binding on the functions of these sites.

- Implement reproducible strategies for connecting ligand binding to complex targets with clear therapeutic mechanisms. There is modest-to-good evidence that disrupting or stabilizing RNA structures near translation initiation sites can modulate gene expression in therapeutically useful ways[13,64,91]. There is some evidence that the binding of small molecules to mRNAs can modulate the efficiency of translation[92,93]. Both disrupting and stabilizing RNA–protein interactions are therapeutically promising. Covalent drugs are gaining acceptance in the protein-targeting field; analogously, ligands that can form covalent linkages or induce cleavage at an RNA target site[44,94] also hold promise.

Perspective

Three key components are necessary to enable effective, repeatable and scalable RNA-targeted drug discovery: (1) a therapeutically compelling RNA target, (2) a screening approach that will identify drug-like lead molecules with appropriate pharmacological properties and (3) the identification of RNA motifs with sufficient information content such that high-specificity and high-potency binding to a high-quality pocket is achieved. We are arguing for a broad change in strategy for creating RNA-targeted drugs. Much greater emphasis needs to be placed on the quality of the RNA target (component 3). Biologically well-validated RNA targets are important, of course, but much greater effort needs to be placed on ensuring the 'ligandability' of these targets. Ultimately, as we learn more about what constitutes a high-quality RNA target, RNA might prove to be no more difficult to drug than proteins. Indeed, given that targetable RNA motifs appear to be ubiquitous (REFS [17,54–66] and FIG. 7d,e) and that the principles of RNA structure and folding are simpler than those for proteins[95], we wonder if RNA might eventually prove to be more generically targetable than proteins. Success of RNA-targeted therapeutic discovery efforts would open up vast opportunities (FIG. 1) for modulating the functions of currently undruggable protein-mediated pathways and the non-coding transcriptome.

Katherine Deigan Warner[1], Christine E. Hajdin[1] and Kevin M. Weeks[2]*

[1]*Ribometrix, Durham, NC, USA.*

[2]*Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA.*

*e-mail: weeks@unc.edu

1. Clamp, M. et al. Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 19428–19433 (2007).
2. Ezkurdia, I. et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* **23**, 5866–5878 (2014).
3. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
4. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).
5. Dixon, S. J. & Stockwell, B. R. Identifying druggable disease-modifying gene products. *Curr. Opin. Chem. Biol.* **13**, 549–555 (2009).
6. Santos, R. et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16**, 19–34 (2017).
7. Sharp, P. A. The centrality of RNA. *Cell* **136**, 577–580 (2009).
8. Cech, T. R. & Steitz, J. A. The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
9. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).

10. Matsui, M. & Corey, D. R. Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.* **16**, 167–179 (2017).

11. Adams, B. D., Parsons, C., Walker, L., Zhang, W. C. & Slack, F. J. Targeting noncoding RNAs in disease. *J. Clin. Invest.* **127**, 761–771 (2017).

12. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).

13. Liang, X.-H. et al. Translation efficiency of mRNAs is increased by antisense oligonucleotides targeting upstream open reading frames. *Nat. Biotechnol.* **34**, 875–880 (2016).

14. Liang, X.-H. et al. Antisense oligonucleotides targeting translation inhibitory elements in 5′ UTRs can selectively increase protein levels. *Nucleic Acids Res.* **45**, 9528–9546 (2017).

15. Crooke, S. T., Witztum, J. L., Bennett, C. F. & Baker, B. F. RNA-targeted therapeutics. *Cell Metab.* **27**, 714–739 (2018).

16. Fellmann, C., Gowen, B. G., Lin, P.-C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nat. Rev. Drug Discov.* **16**, 89–100 (2017).

17. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).

18. Wilson, D. N. Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat. Rev. Microbiol.* **12**, 35–48 (2014).

19. Childs-Disney, J. L. & Disney, M. D. Approaches to validate and manipulate RNA targets with small molecules in cells. *Annu. Rev. Pharmacol. Toxicol.* **56**, 123–140 (2016).

20. Thomas, J. R. & Hergenrother, P. J. Targeting RNA with small molecules. *Chem. Rev.* **108**, 1171–1224 (2008).

21. Aboul-ela, F. Strategies for the design of RNA-binding small molecules. *Future Med. Chem.* **2**, 93–119 (2010).

22. Guan, L. & Disney, M. D. Recent advances in developing small molecules targeting RNA. *ACS Chem. Biol.* **7**, 73–86 (2012).

23. Connelly, C. M., Moon, M. H. & Schneekloth, J. S. The emerging role of RNA as a therapeutic target for small molecules. *Cell Chem. Biol.* **23**, 1077–1090 (2016).

24. Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew. Chem. Int. Ed. Engl.* **56**, 13498–13502 (2017).

25. Hermann, T. Small molecules targeting viral RNA. *Wiley Interdiscip. Rev. RNA* **7**, 726–743 (2016).

26. Rizvi, N. F. & Smith, G. F. RNA as a small molecule druggable target. *Bioorg. Med. Chem. Lett.* **27**, 5083–5088 (2017).

27. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **4**, 90–98 (2012).

28. Moellering, R. C. Linezolid: the first oxazolidinone antimicrobial. *Ann. Intern. Med.* **138**, 135–142 (2003).

29. Wilson, D. N. et al. The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning. *Proc. Natl Acad. Sci. USA* **105**, 13339–13344 (2008).

30. Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).

31. Howe, J. A. et al. Atomic resolution mechanistic studies of ribocil: a highly selective unnatural ligand mimic of the *E. coli* FMN riboswitch. *RNA Biol.* **13**, 946–954 (2016).

32. Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. *Nat. Chem. Biol.* **11**, 511–517 (2015).

33. Sivaramakrishnan, M. et al. Binding to SMN2 pre-mRNA-protein complex elicits specificity for small molecule splicing modifiers. *Nat. Commun.* **8**, 1476 (2017).

34. Kondo, Y., Oubridge, C., van Roon, A.-M. M. & Nagai, K. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5′ splice site recognition. *eLife* **4**, 360 (2015).

35. Calder, A. N., Androphy, E. J. & Hodgetts, K. J. Small molecules in development for the treatment of spinal muscular atrophy. *J. Med. Chem.* **59**, 10067–10083 (2016).

36. Seth, P. P. et al. SAR by MS: discovery of a new class of RNA-binding small molecules for the hepatitis C virus: internal ribosome entry site IIA subdomain. *J. Med. Chem.* **48**, 7099–7102 (2005).

37. Mei, H. Y. et al. Discovery of selective, small-molecule inhibitors of RNA complexes—I. The Tat protein/TAR RNA complexes required for HIV-1 transcription. *Bioorg. Med. Chem.* **5**, 1173–1184 (1997).

38. Prado, S. et al. Bioavailable inhibitors of HIV-1 RNA biogenesis identified through a Rev-based screen. *Biochem. Pharmacol.* **107**, 14–28 (2016).

39. Sztuba-Solinska, J. et al. Identification of biologically active, HIV TAR RNA-binding small molecules using small molecule microarrays. *J. Am. Chem. Soc.* **136**, 8402–8410 (2014).

40. Joly, J.-P. et al. Artificial nucleobase-amino acid conjugates: a new class of TAR RNA binding agents. *Chemistry* **20**, 2071–2079 (2014).

41. Carnevali, M., Parsons, J., Wyles, D. L. & Hermann, T. A modular approach to synthetic RNA binders of the hepatitis C virus internal ribosome entry site. *Chembiochem* **11**, 1364–1367 (2010).

42. Hilimire, T. A. et al. HIV-1 frameshift RNA-targeted triazoles inhibit propagation of replication-competent and multi-drug-resistant HIV in human cells. *ACS Chem. Biol.* **12**, 1674–1682 (2017).

43. Barros, S. A., Yoon, I. & Chenoweth, D. M. Modulation of the *E. coli* rpoH Temperature sensor with triptycene-based small molecules. *Angew. Chem. Int. Ed. Engl.* **55**, 8258–8261 (2016).

44. Nguyen, L. et al. Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1. *J. Am. Chem. Soc.* **137**, 14180–14189 (2015).

45. Luu, L. M. et al. A potent inhibitor of protein sequestration by expanded triplet (CUG) repeats that shows phenotypic improvements in a *Drosophila* model of myotonic dystrophy. *ChemMedChem* **11**, 1428–1435 (2016).

46. Lee, M.-K. et al. A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem. Commun. (Camb.)* **50**, 368–370 (2014).

47. Park, S.-J., Kim, Y.-G. & Park, H.-J. Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J. Am. Chem. Soc.* **133**, 10094–10100 (2011).

48. Stelzer, A. C. et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat. Chem. Biol.* **7**, 553–559 (2011).

49. Bottini, A. et al. Targeting influenza A virus RNA promoter. *Chem. Biol. Drug Des.* **86**, 663–673 (2015).

50. Jin, L., Wang, W. & Fang, G. Targeting protein-protein interaction by small molecules. *Annu. Rev. Pharmacol. Toxicol.* **54**, 435–456 (2014).

51. Doak, B. C., Over, B., Giordanetto, F. & Kihlberg, J. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. *Chem. Biol.* **21**, 1115–1142 (2014).

52. Matsson, P., Doak, B. C., Over, B. & Kihlberg, J. Cell permeability beyond the rule of 5. *Adv. Drug Deliv. Rev.* **101**, 42–61 (2016).

53. Zimmerman, S. C. A journey in bioinspired supramolecular chemistry: from molecular tweezers to small molecules that target myotonic dystrophy. *Beilstein J. Org. Chem.* **12**, 125–138 (2016).

54. Mustoe, A. M. et al. Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* **173**, 181–195.e18 (2018).

55. Siegfried, N. A., Busan, S., Rice, G. M., Nelson, J. A. E. & Weeks, K. M. RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat. Methods* **11**, 959–965 (2014).

56. Mauger, D. M. et al. Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl Acad. Sci. USA* **112**, 3692–3697 (2015).

57. Pirakitikulr, N., Kohlway, A., Lindenbach, B. D. & Pyle, A. M. The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol. Cell* **62**, 111–120 (2016).

58. Villordo, S. M., Carballeda, J. M., Filomatori, C. V. & Gamarnik, A. V. RNA structure duplications and flavivirus host adaptation. *Trends Microbiol.* **24**, 270–283 (2016).

59. Kutchko, K. M. et al. Structural divergence creates new functional features in alphavirus genomes. *Nucleic Acids Res.* **46**, 3657–3670 (2018).

60. Cencig, S. et al. Mapping and characterization of the minimal internal ribosome entry segment in the human c-myc mRNA 5′ untranslated region. *Oncogene* **23**, 267–277 (2004).

61. Chakraborty, S., Mehtab, S., Patwardhan, A. & Krishnan, Y. Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *RNA* **18**, 1014–1028 (2012).

62. Smola, M. J. et al. SHAPE reveals transcript-wide interactions, complex structural domains, and protein interactions across the Xist lncRNA in living cells. *Proc. Natl Acad. Sci. USA* **113**, 10322–10327 (2016).

63. Ball, C. B., Solem, A. C., Meganck, R. M., Laederach, A. & Ramos, S. B. V. Impact of RNA structure on ZFP36L2 interaction with luteinizing hormone receptor mRNA. *RNA* **23**, 1209–1223 (2017).

64. Corley, M. et al. An RNA structure-mediated, posttranscriptional model of human α-1-antitrypsin expression. *Proc. Natl Acad. Sci. USA* **114**, E10244–E10253 (2017).

65. Lee, A. S. Y., Kranzusch, P. J. & Cate, J. H. D. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**, 111–114 (2015).

66. Xue, Z. et al. A G-rich motif in the lncRNA Braveheart interacts with a zinc-finger transcription factor to specify the cardiovascular lineage. *Mol. Cell* **64**, 37–50 (2016).

67. Long, K. S. & Vester, B. Resistance to linezolid caused by modifications at its binding site on the ribosome. *Antimicrob. Agents Chemother.* **56**, 603–612 (2012).

68. World Health Organization. *WHO Model Lists of Essential Medicines, March 2017* (WHO, 2017).

69. Karthikeyan, S. et al. Crystal structure of human riboflavin kinase reveals a beta barrel fold and a novel active site arch. *Structure* **11**, 265–273 (2003).

70. Serganov, A., Huang, L. & Patel, D. J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233–237 (2009).

71. Schneider, T. D., Stormo, G. D., Gold, L. & Ehrenfeucht, A. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.* **188**, 415–431 (1986).

72. Witherell, G. W. & Uhlenbeck, O. C. Specific RNA binding by Q beta coat protein. *Biochemistry* **28**, 71–76 (1989).

73. Carothers, J. M., Oestreich, S. C., Davis, J. H. & Szostak, J. W. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **126**, 5130–5137 (2004).

74. Shang, L., Xu, W., Ozer, S. & Gutell, R. R. Structural constraints identified with covariation analysis in ribosomal RNA. *PLoS ONE* **7**, e39383 (2012).

75. Weeks, K. M. & Mauger, D. M. Exploring RNA structural codes with SHAPE chemistry. *Acc. Chem. Res.* **44**, 1280–1291 (2011).

76. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **10**, 1643–1669 (2015).

77. Carothers, J. M., Davis, J. H., Chou, J. J. & Szostak, J. W. Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *RNA* **12**, 567–579 (2006).

78. Velagapudi, S. P. et al. Design of a small molecule against an oncogenic noncoding RNA. *Proc. Natl Acad. Sci. USA* **113**, 5898–5903 (2016).

79. Pérot, S., Sperandio, O., Miteva, M. A., Camproux, A.-C. & Villoutreix, B. O. Druggable pockets and binding site centric chemical space: a paradigm shift in drug discovery. *Drug Discov. Today* **15**, 656–667 (2010).

80. Fauman, E. B., Rai, B. K. & Huang, E. S. Structure-based druggability assessment — identifying suitable targets for small molecule therapeutics. *Curr. Opin. Chem. Biol.* **15**, 463–468 (2011).

81. An, J., Totrov, M. & Abagyan, R. Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol. Cell. Proteomics* **4**, 752–761 (2005).

82. Bonneau, E. & Legault, P. Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. *Biochemistry* **53**, 6264–6275 (2014).

83. Bonneau, E., Girard, N., Lemieux, S. & Legault, P. The NMR structure of the II-III-VI three-way junction from the *Neurospora* VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure. *RNA* **21**, 1621–1632 (2015).

84. Staple, D. W. & Butcher, S. E. Pseudoknots: RNA structures with diverse functions. *PLoS Biol.* **3**, e213 (2005).

85. Michiels, P. J. et al. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J. Mol. Biol.* **310**, 1109–1123 (2001).

86. Hajdin, C. E. et al. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl Acad. Sci. USA* **110**, 5498–5503 (2013).

87. Smola, M. J., Calabrese, J. M. & Weeks, K. M. Detection of RNA-protein interactions in living cells with SHAPE. *Biochemistry* **54**, 6867–6875 (2015).

88. Gilbert, S. D., Reyes, F. E., Edwards, A. L. & Batey, R. T. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **17**, 857–868 (2009).

89. Warner, K. D. et al. Validating fragment-based drug discovery for biological RNAs: lead fragments bind and remodel the TPP riboswitch specifically. *Chem. Biol.* **21**, 591–595 (2014).

90. Miao, Z. & Westhof, E. RNA structure: advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **46**, 483–503 (2017).

91. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361**, 13–37 (2005).

92. Harvey, I., Garneau, P. & Pelletier, J. Inhibition of translation by RNA-small molecule interactions. *RNA* **8**, 452–463 (2002).

93. Bandyopadhyay, S. et al. Novel 5′ untranslated region directed blockers of iron-regulatory protein-1 dependent amyloid precursor protein translation: implications for down syndrome and Alzheimer's disease. *PLoS ONE* **8**, e65978 (2013).

94. Rzuczek, S. G. et al. Precise small-molecule recognition of a toxic CUG RNA repeat expansion. *Nat. Chem. Biol.* **13**, 188–193 (2017).

95. Tinoco, I. & Bustamante, C. How RNA folds. *J. Mol. Biol.* **293**, 271–281 (1999).

96. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).

97. Ritchie, T. J. & Macdonald, S. J. F. How drug-like are 'ugly' drugs: do drug-likeness metrics predict ADME behaviour in humans? *Drug Discov. Today* **19**, 489–495 (2014).

98. Gleitsman, K. R., Sengupta, R. N. & Herschlag, D. Slow molecular recognition by RNA. *RNA* **23**, 1745–1753 (2017).

99. Weeks, K. M. & Crothers, D. M. RNA recognition by Tat-derived peptides: interaction in the major groove? *Cell* **66**, 577–588 (1991).

100. Lin, A. H., Murray, R. W., Vidmar, T. J. & Marotti, K. R. The oxazolidinone eperezolid binds to the 50S ribosomal subunit and competes with binding of chloramphenicol and lincomycin. *Antimicrob. Agents Chemother.* **41**, 2127–2131 (1997).

101. Ritchie, D. B., Soong, J., Sikkema, W. K. A. & Woodside, M. T. Anti-frameshifting ligand reduces the conformational plasticity of the SARS virus pseudoknot. *J. Am. Chem. Soc.* **136**, 2196–2199 (2014).

102. Parsons, J. et al. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nat. Chem. Biol.* **5**, 823–825 (2009).

103. Costales, M. G. et al. Small molecule inhibition of microRNA-210 reprograms an oncogenic hypoxic circuit. *J. Am. Chem. Soc.* **139**, 3446–3455 (2017).

104. Yildirim, I., Park, H., Disney, M. D. & Schatz, G. C. A dynamic structural model of expanded RNA CAG repeats: a refined X-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. *J. Am. Chem. Soc.* **135**, 3528–3538 (2013).

105. Leeper, T., Leulliot, N. & Varani, G. The solution structure of an essential stem-loop of human telomerase RNA. *Nucleic Acids Res.* **31**, 2614–2621 (2003).

106. Chen, Y. et al. Rbfox proteins regulate microRNA biogenesis by sequence-specific binding to their precursors and target downstream Dicer. *Nucleic Acids Res.* **44**, 4381–4395 (2016).

107. Faber, C., Sticht, H., Schweimer, K. & Rösch, P. Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J. Biol. Chem.* **275**, 20660–20666 (2000).

108. Zhang, Q., Harada, K., Cho, H. S., Frankel, A. D. & Wemmer, D. E. Structural characterization of the complex of the Rev response element RNA with a selected peptide. *Chem. Biol.* **8**, 511–520 (2001).