Published in final edited form as:

Drug Discov Today. 2013 December ; 18(0): . doi:10.1016/j.drudis.2013.07.024.

Rational design of chemical genetic probes of RNA function and lead therapeutics targeting repeating transcripts

Matthew D. Disney

Department of Chemistry, The Scripps Research Institute, 130 Scripps Way #3A1, Jupiter, FL 33458, USA

Matthew D. Disney: Disney@scripps.edu

Abstract

RNA is an important yet vastly underexploited target for small molecule chemical probes or lead therapeutics. Small molecules have been used successfully to modulate the function of the bacterial ribosome, viral RNAs and riboswitches. These RNAs are either highly expressed or can be targeted using substrate mimicry, a mainstay in the design of enzyme inhibitors. However, most cellular RNAs are neither highly expressed nor have a lead small molecule inhibitor, a significant challenge for drug discovery efforts. Herein, I describe the design of small molecules targeting expanded repeating transcripts that cause myotonic muscular dystrophy (DM). These test cases illustrate the challenges of designing small molecules that target RNA and the advantages of targeting repeating transcripts. Lastly, I discuss how small molecules might be more advantageous than oligonucleotides for targeting RNA.

The development of chemical probes of function or lead therapeutics is an arduous task replete with many failures along the way. Commonly, these investigations are initiated with lead compounds [1] that are identified by screening small molecule libraries [2] or by using information about transition states in enzymatic pathways [3]. These approaches have been used successfully to identify and optimize small molecules targeting proteins, and the results of these studies are often applicable to other protein targets. For example, collections of small molecules that inhibit protein–protein interactions identified features of inhibitors that engender potency, leading to the design of focused chemical libraries specifically for this class of targets [4].

Natural products have provided a rich source of lead small molecules to target DNA, and detailed information about their binding modes has enabled the design of improved compounds [5–7]. The *tour de force* example is the exemplary work of the Dervan group [7]. They used information on the molecular recognition of natural products such as distamycin (Fig. 1a) and in-depth synthetic and biophysical studies to design modularly assembled small molecules that essentially can recognize the minor groove of any DNA sequence [7].

There are generally fewer lead small molecules available to initiate drug discovery efforts for RNA. Although there are examples of natural products that bind RNA and modulate function, they are mostly restricted to targeting the bacterial ribosome [8,9]. Typically, antibacterials bind essential and/or dynamic regions in rRNA and affect mRNA decoding [10,11]. The classic example is the binding of aminoglycosides to the A-site (Fig. 1b), which causes recognition of both cognate and noncognate tRNAs [12,13]. The bacterial ribosome is the most exploited RNA drug target for many reasons, including: (i) rRNAs are the most highly expressed RNAs in the transcriptome, comprising as much as 90% of total cellular RNA [14]; (ii) compounds that are not exquisitely specific for the ribosome can be inhibitory because of the high expression level of the ribosome relative to other RNAs; and

(iii) the ribosome has an essential role in cellular function, and relatively minor effects in translation can have drastic effects on cellular homeostasis.

Challenges and opportunities for targeting RNA with small molecules

The challenge

Compared with rRNA, other RNA targets in the transcriptome are expressed at low levels. Thus, a small molecule modulator of function must: (i) be selective because it must recognize its target preferentially over other cellular RNAs that are expressed at similar or higher levels; (ii) bind with high affinity; and (iii) localize to the correct region of the cell (i.e. the same location as the RNA target, which is variable unlike DNA).

As described above, information on the molecular recognition of the DNA minor groove enabled the design of small molecules that can decode any DNA sequence. Although small molecules are known to bind to and modulate the cellular function of the ribosome, there is no distamycin equivalent for RNA. Moreover, it would not provide sufficient information to develop generalizable methods to design specific small molecules targeting RNA because of its structural diversity (Fig. 1b,c). To design effectively a modularly assembled compound to target RNA, small molecule binding modules would be required that bind base pairs, internal loops, hairpins, bulges, multibranch loops and pseudoknots (Fig. 1b,c) [16–18]. Thus, a larger set of small molecules is required to design small molecules that target RNA than DNA, more than even several ‘DNA-distamycin equivalents’ [8].

The opportunity

Despite the challenges discussed above, targeting RNA with small molecules could be advantageous. Its structural diversity suggests that RNA forms binding pockets that could be inherently specific for small molecule binding. However, there is currently little information about the structural motifs in RNA that are privileged binding sites for small molecules or the small molecules that bind them.

Identifying RNA structural motifs that provide small molecule binding sites

Fortuitously, a broadly applicable technology termed systematic evolution of ligands by exponential enrichment (SELEX) has been developed that enables vast libraries of RNAs to be tested for binding a small molecule [19]. After multiple rounds of selection, the RNAs that bind with the highest affinity (aptamers) are isolated and identified by sequencing. Aptamers can recognize ligands with affinities and selectivities that exceed antibodies [20,21]. The only disadvantage of SELEX is that aptamers are relatively large (20–40 nt long) and, thus, are unlikely to be found in a transcriptome. Despite the large size of the randomized regions, portions of aptamers identified via SELEX have been found in biological RNAs [22,23].

To identify smaller motifs that bind small molecules with high affinity and selectivity, an approach was developed called two-dimensional combinatorial screening (2DCS), a library-versus-library screen. Small molecules are immobilized on an agarose-coated microarray and incubated with an RNA library that displays discrete (small) secondary structural elements (motifs) that are likely to be found in cellular RNAs. Selections are completed under conditions of high oligonucleotide stringency to ensure that compounds only bind the randomized region [24]. Bound RNAs are harvested from the array by manual excision and sequenced to identify privileged motifs that bind each small molecule. These motifs can be searched for in disease-associated RNAs, and the corresponding small molecule serves as a lead inhibitor. Potentially, 2DCS could enable the acquisition of a wealth of information about RNA motif–small molecule interactions (i.e. distamycin–DNA equivalents) that can

be used to design modularly assembled small molecules that target RNA (Figs. 2 and 3), analogous to the modular assembly of polyamides that target DNA [7]. This approach has been successfully employed, providing bioactive compounds against the RNA that causes myotonic dystrophy type 1 (DM1) [25,26].

Identifying small molecules that bind RNA

The most common technique for identifying small molecules that bind an RNA target of interest is screening. A validated RNA target or oligonucleotide mimic is screened for binding to a library of small molecules, such as those available from the National Cancer Institute (NCI) or eMolecules. Unlike their protein counterparts, low hit rates from standard small molecule libraries for RNA targets are often observed. These results suggest that a yet undefined chemical space(s) is unique for RNA.

Structure-aided design and docking

Structure-aided design and docking are methods commonly used to design lead small molecules. Small molecule docking simulations were first developed for protein targets [27–31] and the simulation parameters were then modified for RNA. Briefly, the interactions of a small molecule and an RNA target are simulated, governed by potential electrostatic interactions, hydrogen bonding, van der Waals interactions, solvation and so on. The corresponding energetics of the interaction is calculated; small molecules with the most favorable interactions serve as lead compounds. Most docking software allows for flexibility within the structure of the small molecule; more recently, programs have been developed that also allow flexibility in the structure of the RNA [32–39]. A recent report used a dynamic ensemble of the structure of HIV TAR (Fig. 1b) to dock a library of small molecules, affording a bioactive small molecule that inhibits HIV replication [34].

Other computational methods

Although structure-aided design and docking are powerful tools, they are computationally intensive. Other computational methods are available to identify potentially improved compounds based on the structure of a small molecule lead, including chemical similarity searching and quantitative structure–activity relationship (QSAR) modeling. Chemical similarity searching identifies compounds that are structurally similar to a small molecule lead but may have improved properties. The 3D shape of the lead compound, including the positioning of functional groups, is determined by energy minimization and then compared with the 3D shapes of compounds within small molecule databases [40]. The degree of similarity is quantified by the shape Tanimoto coefficient, which ranges in value from 0 (no similarity) to 1 (complete similarity) [41]. QSAR modeling defines relations between the chemical properties of small molecules and their bioactivities and/or toxicities [42,43]. Those relations are then used to predict the activity or toxicity of other compounds.

Targeting repeating transcripts that cause DM

Herein, I present how the methods described above have been applied to identify lead small molecule inhibitors and bioactive compounds that target the RNAs that cause DM. There are two types of DM: 1 (DM1) and 2 (DM2), both caused by expanded RNA repeats located in a noncoding region [44–46]. In the case of DM1, the toxic RNA is an expanded r(CUG) repeat [r(CUG)^{exp}; 50 repeats; Fig. 2a] present in the 3' untranslated region (UTR) of the dystrophin myotonia protein kinase (*DMPK*) mRNA [45,46]. r(CUG)^{exp} folds into a hairpin structure that binds and inactivates muscleblind-like 1 protein (MBNL1) (Fig. 2a). MBNL1 regulates alternative pre-mRNA splicing [46]; thus, dysregulation of splicing is observed in DM1-affected muscle cells [47–51]. DM2 is the second most common cause of adult-onset DM. Disease is caused by an expanded r(CCUG) repeat [r(CCUG)^{exp}; Fig. 2b]

present in intron 1 of the zinc finger 9 (*ZNF9*) pre-mRNA [44,46]. As observed in DM1, r(CCUG) repeats of a pathogenic length [75 to 11,000 repeats] fold into a hairpin that binds and inactivates MBNL1 (Fig. 2b), causing alternative pre-mRNA splicing defects [52,53]. Interestingly, splicing defects are also observed in the brains of DM patients, caused by inactivation of MBNL2 by r(CUG)^{exp} and r(CCUG)^{exp} [54]. There are no current treatments available for DM1 or DM2.

Targeting repeating transcripts, such as r(CUG)^{exp} and r(CCU-G)^{exp}, with small molecules could be advantageous. Positive cooperativity could enable the specific recognition of long r(CUG) or r(CCUG) repeats over bystander targets. Interestingly, there are no other RNAs in the mouse or human transcriptome that have stretches of r(CUG) repeats that are as long as r(CUG)^{exp}. Moreover, repeating transcripts provide ideal targets for a modular assembly and/or multivalency approach; that is, an optimal compound would display RNA-binding modules at a distance that mimics the distance between repeating motifs. A modular approach is likely to be advantageous for any RNA target. There are many single copies of a targetable motif in a transcriptome, but few RNAs that have two or more targetable motifs separated by a specific distance (e.g., two GC base pairs) [15].

Design of small molecules that target r(CUG)^{exp}

Two approaches have been used to design compounds that bind r(CUG) repeats: structure-aided design and using information about RNA motif–small molecule interactions from 2DCS selections.

Structure-aided design of compounds that bind r(CUG) repeats

Structure-aided design was employed to design small molecules that bind r(CUG)^{exp} [55]. A crystal structure of r(CUG) repeats shows that it adopts an overall A-form geometry. However, the 1 × 1 nt UU internal loops adopt various conformations that could be engaged by a small molecule [15,56–58]. Using this information, triazine compounds with an acridine linker were designed (Tri-Ac, Fig. 4) [55]. The triazine moiety is proposed to bind the 1 × 1 nt UU internal loops via a Janus wedge-type interaction, whereas the acridine stacks on the adjacent GC base pairs. These compounds have binding affinities for r(CUG)^{exp} in the nanomolar range (~ 400 nM) and are 13–169-fold selective for r(CUG) repeats over RNAs with other mismatches [55]. Tri-Ac also inhibits formation of the r(CUG)^{exp}–MBNL1 complex *in vitro* (IC₅₀ = ~ 50 μM) [55]. By completing additional studies on the Janus wedge compounds, small molecules that target r(CCUG)^{exp} (causative agent of DM2) were developed (see below) [59].

Using RNA motif–small molecule interactions: upside-down lead drug and probe discovery

Given the repeating nature of the structure of r(CUG)^{exp}, a modular approach could be employed, affording a multivalent compound that binds multiple elements within the structure of the RNA simultaneously, namely the 1 × 1 nt UU internal loops (Fig. 2a). However, an RNA-binding module that binds the loops was required. It was previously identified that the *bis*-benzimidazole Hoechst 33258 binds the repeating 5'CUG/3'GUC motif found in r(CUG)^{exp} with nanomolar affinity [60]. Thus, a derivative of this compound, **H**, was designed and modularly assembled onto a peptoid scaffold (Fig. 3).

A peptoid scaffold was used to display multiple copies of **H** because peptoid synthesis is facile [61] and because the peptoid scaffold enables precise control of the number of RNA-binding modules displayed (valency) and the distance between them. Valency was controlled by installation of alkyne submonomers to which **H** was conjugated via a Cu-

catalyzed Huisgen dipolar cycloaddition reaction [62]. The distance between each RNA-binding module was controlled by the number of propylamine spacing modules conjugated between RNA-binding modules (Fig. 3). The general nomenclature for the modularly assembled compounds is **nL-m**, where **n** indicates the number of RNA-binding modules (valency), **L** indicates the RNA-binding module, and **m** indicates the number of propylamine spacing modules, or the distance between RNA-binding modules.

A library of compounds was synthesized with different distances between **H** RNA-binding modules and screened for inhibiting the r(CUG)^{exp}-MBNL1 complex. The most potent inhibitors *in vitro* were **nH-4** compounds (**2H-4**, **3H-4** and **4H-4**) (Fig. 3) [63]. Affinities of the **nH-4** compounds range from 13 nM to 100 nM. They are up to 54-fold selective for r(CUG)^{exp} over herring sperm DNA. It is noteworthy that designed, modularly assembled compounds recognize r(CUG)^{exp} with higher affinity and selectivity than MBNL1 does [63].

H and the **nH-4** compounds were studied for improving DM1-associated defects in cell-based model systems [25]. DM1-associated defects include: (i) dysregulation of alternative pre-mRNA splicing due to sequestration and inactivation of MBNL1 [49,50,64]; (ii) decreased translation of the *DMPK* mRNA owing to its poor nucleocytoplasmic transport [65,66] and (iii) formation of nuclear foci caused by r(CUG)^{exp}-protein complexes [65,67–69]. Although the **H** monomer did not improve DM1-associated defects in any of the model systems, modularly assembled compounds improved all three defects to varying extents. For example, **2H-4** improves dysregulation of alternative pre-mRNA splicing to levels observed in the absence of r(CUG)^{exp} at low micromolar concentrations [25]. The compound also disrupts the formation of nuclear foci and improves the DM1-associated translational defect as determined using a cell-based luciferase reporter system. Improvement of the translational defect is especially interesting as it requires the small molecule to traverse the nuclear envelope, disrupt the r(CUG)^{exp}-MBNL1 complex, and allow the r(CUG)^{exp} target to be transported to the cytoplasm for translation. Thus, **2H-4** alters the subcellular localization of the target RNA, suggesting an additional use of the compound to study and modulate RNA localization.

Chemical similarity searching

Although various methods are available for optimization of a small molecule lead, chemical similarity searching is particularly facile. Thus, compounds were searched for that are chemically similar to **H** in the e-Molecules and the NCI databases [70]. These studies identified **H1** (Fig. 4) as a potent *in vitro* and *in vivo* inhibitor of the r(CUG)^{exp}-MBNL1 complex. **H1** improves DM1-associated pre-mRNA splicing defects in cell culture models and in a DM1 mouse model [70]. Binding studies revealed that the small molecule preferentially binds r(CUG)^{exp} over other bystander RNAs because of positive cooperativity [K_d = 70 nM; 2-fold selective for r(CUG) repeats over fully paired RNAs]; that is, **H1** specifically recognizes RNAs with large numbers of 1 × 1 nt UU internal loops over ones with fewer loops, such as the human rRNA A-site [70].

Screening to identify compounds that reduce r(CUG)^{exp}-associated toxicity

Several approaches have been used to identify small molecules from chemical libraries that bind r(CUG)^{exp}. In the first example, a small molecule library of compounds known to bind nucleic acids was probed for inhibition of the r(CUG)^{exp}-MBNL1 complex. These studies identified that pentamidine (Fig. 4) disrupts complex formation *in vitro* and improves several defects associated with DM1 *in vivo* [71]. Specifically, it improves pre-mRNA splicing defects and reduces the number of nuclear foci in cell culture models and improves pre-mRNA splicing defects in a DM1 mouse model [71].

Dynamic combinatorial chemistry (DCC) was used to identify compounds that bind r(CUG) repeats *in vitro* [72]. DCC uses reversible chemistry to enable the assembly of compounds onto a target, thus shifting the equilibrium toward binders [72]. Disulfide bond formation was used to identify lead compounds *in vitro*, which afforded four lead compounds that bind with affinities of $\sim 2 \mu\text{M}$ and IC_{50}s of $\sim 20 \mu\text{M}$. A second-generation compound was developed from these leads (DCC-1, Fig. 4) that improves the DM1-associated translational defect in cell culture models and, importantly, improves alternative pre-mRNA splicing defects in a DM1 mouse model [73].

Another study created a high-throughput screening assay to identify *in vitro* inhibitors of the r(CUG)–MBNL1 complex [74]. The assay was then used to screen 1280 pharmacologically active compounds (LOPAC¹²⁸⁰, Sigma–Aldrich). Six compounds were identified with IC_{50}s ranging from ~ 20 to $\sim 55 \mu\text{M}$ [74]. Small molecule leads were evaluated in cell culture models for ameliorating DM1-associated defects. Interestingly, one of the leads, compound **2** (Fig. 4), improves alternative pre-mRNA splicing defects by binding r(CUG)^{exp} [$K_d = 125 \text{ nM}$; >10 -fold selective for r(CUG) repeats over RNAs with other mismatches or fully paired RNAs] [75]. Another small molecule lead, compound **1** (Fig. 4), exacerbated the DM1 phenotype in cell culture models by binding and inhibiting MBNL1 ($K_d = 470 \text{ nM}$) [75]. These studies demonstrate that lead compounds can be identified by screening small molecule libraries; however, care must be exercised if a small molecule can be identified as a hit if it binds to the RNA or to the protein.

A library of D-amino acid containing hexapeptides was screened in a *Drosophila* model of DM1 to identify compounds that suppress r(CUG)^{exp} lethality [76]. The best compound identified in those studies improves DM1-associated defects in the *Drosophila* model and in a DM1 mouse model (D-hexapeptide, Fig. 4) [76]. These studies illustrate that *Drosophila* disease models can enable the identification of compounds that reduce RNA toxicity, and it is likely that this approach can be applied to other diseases.

Design of small molecules that target r(CCUG)^{exp}

In a selection of the RNA motifs that bind kanamycin A, it was found that the highest affinity loops are similar to those present in r(CCUG)^{exp}, the causative agent of DM2 (Fig. 2b,c) [24,77]. Therefore, a kanamycin A derivative (6'-N-5-hexynoate kanamycin A; **K**) was used that could be modularly assembled as a lead small molecule (Figs. 2c and 3). The binding of 6'-N-5-hexynoate kanamycin A (**K**) to RNAs with varied numbers of the repeating motif present in r(CCUG)^{exp}, or 5'CCUG/3'GUCC, was studied (Fig. 2a) [77]. Interestingly, **K** binds mimics of r(CCUG)^{exp} with high affinity (400 nM), similar to the RNAs selected from a resin-based screen (Fig. 2c) [77]. Thus, **K** was suitable as a lead small molecule for a modular assembly approach.

Akin to the studies described above for targeting r(CUG)^{exp}, the optimal distance between **K** RNA-binding modules was determined by studying a library of compounds (with different distances between **K**s) for inhibiting formation of the r(CCUG)^{exp}–MBNL1 complex *in vitro*. The most potent compounds identified were **2K-4** and **3K-4** (Fig. 3), which have low nanomolar IC_{50}s . **3K-4** binds r(CCUG)^{exp} with a 15-fold higher affinity ($K_d = 8 \text{ nM}$) and greater specificity than MBNL1 does (~ 20 -fold over similar RNAs) [77]. Notably, these designed small molecules are cell permeable despite their relatively high molecular weights [77,78]. Unfortunately, when **2K-4** and **3K-4** were developed, there were no cell culture models of DM2 to evaluate their bioactivity.

Derivatizing a DM1 inhibitor to develop a small molecule that binds r(CCUG) repeats

Baranger and Zimmerman used their designed Tri-Ac small molecule that binds r(CUG) repeats to develop a compound that binds r(CCUG) repeats [59]. Given that both RNAs contain pyrimidine mismatches closed by GC base pairs, the authors hypothesized that compounds similar to triaminotriazine might bind r(CCUG) repeats and inhibit binding of MBNL1. They synthesized one triaminotriazine- and two triaminopyrimidine-based small molecules and tested them for binding to r(CCUG) repeats. Interestingly, the triaminotriazine did not bind, whereas the triaminopyrimidine-based compounds bind with low micromolar affinity ($K_d \sim 0.1\text{--}3.6 \mu\text{M}$) and good selectivity (0.2-fold to 93-fold over related nucleic acids; 10-fold over herring sperm DNA; and 670-fold over tRNA) [59]. Both triaminopyrimidines inhibit the r(CCUG)–MBNL1 interaction, with IC_{50} s ranging from ~ 35 to $60 \mu\text{M}$ [59].

Summary and perspectives: exploiting other RNA targets rationally

How can we exploit other RNA targets in the transcriptome with small molecules?

Each example described above serves as a case study on the development of small molecules that target RNA. One of the far-reaching difficulties in this area has been developing selective compounds owing to: (i) the low expression levels of most RNA targets; (ii) the degeneracy of single RNA motifs in the transcriptome. For example, the human ribosome has a 1×1 nt UU internal loop that is similar to the 5'CUG/3'GUC motif found in r(CUG)^{exp} [15]; and (iii) the currently embryonic understanding of RNA structure in a cell.

Various methods have been used to target the RNAs that cause myotonic dystrophy that are likely to be applicable to other targets. In the absence of a 3D structure of the RNA target, information about RNA–small molecule interactions could be used. Selection methods, such as 2DCS, identify optimal targets and provide information about potential off-targets [15]. Synergistically, modular assembly provides a facile means to optimize compounds for specifically recognizing a target of interest. Peptoids provide an almost ideal modular assembly scaffold because they are easy to synthesize [61,79], and impart cellular permeability despite their high molecular weight [78,80,81]. In many instances, larger assembled compounds that display RNA-binding modules are more cell permeable than the modules themselves [78]. Moreover, lead small molecules with little or no activity *in vivo* can be optimized computationally via chemical similarity searching or QSAR.

Screening small molecule libraries generally fails to provide high-quality lead compounds for targeting RNA. This is perhaps not surprising because these libraries were constructed for screening proteins and are inherently biased as such. Studies are being completed to identify chemotypes in small molecules that predispose them for binding RNA [82], which may enable the design of RNA-focused small molecule libraries that have higher hit rates.

Importantly, compounds targeting RNA specifically are likely to violate the Lipinski guidelines for drug development [83]. More hydrogen bond donors and acceptors will be required to recognize selectively RNA than these guidelines permit. Furthermore, high molecular weight compounds are likely to provide potent modulators of function because they sequester larger amounts of the target's surface area than traditional compounds.

Why bother? Just use oligonucleotides

Oligonucleotide-based targeting of RNA has proven effective for target validation, but is perhaps suboptimal for the development of therapeutic modalities. Only a few oligonucleotide-based therapeutics have been approved for use in humans, although several

more are moving through the approval process. However, oligonucleotide-based therapeutics improve DM1-associated defects in cellular and animal models, including improvement of myotonia [84,85].

There are major advantages of using cell-permeable small molecules for targeting RNA rather than oligonucleotides. Small molecules are often orally available. If toxicity or other issues arise, derivatives can be synthesized to mitigate them. Given that only a limited set of oligonucleotide derivatives are available such that bioactivity is retained, lead optimization using traditional medicinal chemistry approaches is less tractable for oligonucleotides.

An important and perhaps less recognized benefit of using small molecules is that they recognize RNA structure not sequence. Highly structured RNAs could, in some cases, be too thermodynamically stable to be effectively targeted with oligonucleotides; that is, strand invasion would be required. For example, r(CUG)^{exp} folds into a stable RNA hairpin; its complementary oligonucleotide, (CAG)_n could also fold into a stable hairpin structure. For sequence recognition (duplex formation) to occur, both the RNA target and the oligonucleotide must be unfolded. The interplay of these structures likely influences bioactivity. Breaking of the target's structure would not be required for designer small molecules that recognize specific motifs.

Acknowledgments

I thank my students who worked on the projects described herein. Funding for research in the Disney Lab was provided by the National Institutes of Health, the Muscular Dystrophy Association, the Camille and Henry Dreyfus Foundation, Research Corporation and NYSTAR.

References

1. Schreiber SL. Organic synthesis toward small-molecule probes and drugs. *Proc Natl Acad Sci U S A*. 2011; 108:6699–6702. [PubMed: 21464328]
2. Nielsen TE, Schreiber SL. Towards the optimal screening collection: a synthesis strategy. *Angew Chem Int Ed Engl*. 2008; 47:48–56. [PubMed: 18080276]
3. Schramm VL. Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu Rev Biochem*. 2011; 80:703–732. [PubMed: 21675920]
4. Yin H, Hamilton AD. Strategies for targeting protein–protein interactions with synthetic agents. *Angew Chem Int Ed Engl*. 2005; 44:4130–4163. [PubMed: 15954154]
5. Dervan PB, Edelson BS. Recognition of the DNA minor groove by pyrrole–imidazole polyamides. *Curr Opin Struct Biol*. 2003; 13:284–299. [PubMed: 12831879]
6. Hamilton PL, Arya DP. Natural product DNA major groove binders. *Nat Prod Rep*. 2012; 29:134–143. [PubMed: 22183179]
7. Dervan PB. Molecular recognition of DNA by small molecules. *Bioorg Med Chem*. 2001; 9:2215–2235. [PubMed: 11553460]
8. Guan L, Disney MD. Recent advances in developing small molecules targeting RNA. *ACS Chem Biol*. 2012; 7:73–86. [PubMed: 22185671]
9. Thomas JR, Hergenrother PJ. Targeting RNA with small molecules. *Chem Rev*. 2008; 108:1171–1224. [PubMed: 18361529]
10. Kaul M, et al. Fluorescence-based approach for detecting and characterizing antibiotic-induced conformational changes in ribosomal RNA: comparing aminoglycoside binding to prokaryotic and eukaryotic ribosomal RNA sequences. *J Am Chem Soc*. 2004; 126:3447–3453. [PubMed: 15025471]
11. Kaul M, et al. Aminoglycoside-induced reduction in nucleotide mobility at the ribosomal RNA A-site as a potentially key determinant of antibacterial activity. *J Am Chem Soc*. 2006; 128:1261–1271. [PubMed: 16433544]

12. Yoshizawa S, et al. Recognition of the codon–anticodon helix by ribosomal RNA. *Science*. 1999; 285:1722–1725. [PubMed: 10481006]
13. Carter AP, et al. Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature*. 2000; 407:340–348. [PubMed: 11014183]
14. Berg, JM., et al. *Biochemistry*. 6. W.H. Freeman and Co; 2007.
15. Kumar A, et al. Myotonic dystrophy type 1 RNA crystal structures reveal heterogeneous 1×1 nucleotide UU internal loop conformations. *Biochemistry*. 2011; 50:9928–9935. [PubMed: 21988728]
16. Doudna JA. Structural genomics of RNA. *Nat Struct Biol*. 2000; 7 (Suppl):954–956. [PubMed: 11103998]
17. Batey RT, et al. Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed Engl*. 1999; 38:2326–2343. [PubMed: 10458781]
18. Mathews DH. Revolutions in RNA secondary structure prediction. *J Mol Biol*. 2006; 359:526–532. [PubMed: 16500677]
19. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*. 1990; 249:505–510. [PubMed: 2200121]
20. Joyce GF. *In vitro* evolution of nucleic acids. *Curr Opin Struct Biol*. 1994; 4:331–336. [PubMed: 11539574]
21. Klug SJ, Famulok M. All you wanted to know about SELEX. *Mol Biol Rep*. 1994; 20:97–107. [PubMed: 7536299]
22. Carlson CB, et al. Preferred RNA binding sites for a threading intercalator revealed by *in vitro* evolution. *Chem Biol*. 2003; 10:663–672. [PubMed: 12890540]
23. Tao J, Frankel AD. Arginine-binding RNAs resembling TAR identified by *in vitro* selection. *Biochemistry*. 1996; 35:2229–2238. [PubMed: 8652564]
24. Disney MD, Childs-Disney JL. Using selection to identify and chemical microarray to study the RNA internal loops recognized by 6'-N-acylated kanamycin A. *Chem Bio Chem*. 2007; 8:649–656.
25. Childs-Disney JL, et al. Rationally designed small molecules targeting the RNA that causes myotonic dystrophy type 1 are potentially bioactive. *ACS Chem Biol*. 2012; 7:856–862. [PubMed: 22332923]
26. Childs-Disney JL, et al. Rational design of bioactive, modularly assembled aminoglycosides targeting the RNA that causes myotonic dystrophy type 1. *ACS Chem Biol*. 2012; 7:1984–1993. [PubMed: 23130637]
27. Kuntz ID, et al. A geometric approach to macromolecule–ligand interactions. *J Mol Biol*. 1982; 161:269–288. [PubMed: 7154081]
28. Goodsell DS, Olson AJ. Automated docking of substrates to proteins by simulated annealing. *Proteins*. 1990; 8:195–202. [PubMed: 2281083]
29. Morris GM, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009; 30:2785–2791. [PubMed: 19399780]
30. Gohlke H, et al. Knowledge-based scoring function to predict protein–ligand interactions. *J Mol Biol*. 2000; 295:337–356. [PubMed: 10623530]
31. Velec HF, et al. Drug Score(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem*. 2005; 48:6296–6303. [PubMed: 16190756]
32. Daldrop P, et al. Novel ligands for a purine riboswitch discovered by RNA–ligand docking. *Chem Biol*. 2011; 18:324–335. [PubMed: 21439477]
33. Pfeffer P, Gohlke H. Drug Score RNA: knowledge-based scoring function to predict RNA–ligand interactions. *J Chem Inf Model*. 2007; 47:1868–1876. [PubMed: 17705464]
34. Stelzer AC, et al. Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol*. 2011; 7:553–559. [PubMed: 21706033]
35. Filikov AV, et al. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J Comput Aided Mol Des*. 2000; 14:593–610. [PubMed: 10921774]

36. Lind KE, et al. Structure-based computational database screening, *in vitro* assay, and NMR assessment of compounds that target TAR RNA. *Chem Biol.* 2002; 9:185–193. [PubMed: 11880033]
37. Pinto IG, et al. Discovery of ligands for a novel target, the human telomerase RNA, based on flexible-target virtual screening and NMR. *J Med Chem.* 2008; 51:7205–7215. [PubMed: 18950148]
38. Tuccinardi T. Binding-interaction prediction of RNA-binding ligands. *Future Med Chem.* 2011; 3:723–733. [PubMed: 21554078]
39. Fulle S, Gohlke H. Molecular recognition of RNA: challenges for modelling interactions and plasticity. *J Mol Recognit.* 2010; 23:220–231. [PubMed: 19941322]
40. Willett P. Similarity searching using 2D structural fingerprints. *Methods Mol Biol.* 2011; 672:133–158. [PubMed: 20838967]
41. Haigh JA, et al. Small molecule shape-fingerprints. *J Chem Inf Model.* 2005; 45:673–684. [PubMed: 15921457]
42. Kortagere S, et al. Role of computational methods in pharmaceutical sciences. *Methods Mol Biol.* 2012; 929:21–48. [PubMed: 23007425]
43. Valerio LG Jr, Choudhuri S. Chemoinformatics and chemical genomics: potential utility of *in silico* methods. *J Appl Toxicol.* 2012; 32:880–889. [PubMed: 22886396]
44. Liquori CL, et al. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of *ZNF9*. *Science.* 2001; 293:864–867. [PubMed: 11486088]
45. Orengo JP, et al. Alternative splicing dysregulation secondary to skeletal muscle regeneration. *Ann Neurol.* 2011; 69:681–690. [PubMed: 21400563]
46. Day JW, Ranum LP. RNA pathogenesis of the myotonic dystrophies. *Neuromuscul Disord.* 2005; 15:5–16. [PubMed: 15639115]
47. Dansithong W, et al. MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. *J Biol Chem.* 2005; 280:5773–5780. [PubMed: 15546872]
48. Ho TH, et al. Muscleblind proteins regulate alternative splicing. *EMBO J.* 2004; 23:3103–3112. [PubMed: 15257297]
49. Savkur RS, et al. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet.* 2001; 29:40–47. [PubMed: 11528389]
50. Philips AV, et al. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science.* 1998; 280:737–741. [PubMed: 9563950]
51. Wang ET, et al. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell.* 2012; 150:710–724. [PubMed: 22901804]
52. Fugier C, et al. Misregulated alternative splicing of *BIN1* is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat Med.* 2011; 17:720–725. [PubMed: 21623381]
53. Savkur RS, et al. Insulin receptor splicing alteration in myotonic dystrophy type 2. *Am J Hum Genet.* 2004; 74:1309–1313. [PubMed: 15114529]
54. Charizanis K, et al. Muscleblind-like 2-mediated alternative splicing in the developing brain and dysregulation in myotonic dystrophy. *Neuron.* 2012; 75:437–450. [PubMed: 22884328]
55. Arambula JF, et al. A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. *Proc Natl Acad Sci U S A.* 2009; 106:16068–16073. [PubMed: 19805260]
56. Warf MB, Berglund JA. MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA.* 2007; 13:2238–2251. [PubMed: 17942744]
57. Parkesh R, et al. NMR spectroscopy and molecular dynamics simulation of r(CCGCUGCGG)₂ reveal a dynamic UU internal loop found in myotonic dystrophy type 1. *Biochemistry.* 2011; 50:599–601. [PubMed: 21204525]
58. Kiliszek A, et al. Structural insights into CUG repeats containing the ‘stretched U–U wobble’: implications for myotonic dystrophy. *Nucleic Acids Res.* 2009; 37:4149–4156. [PubMed: 19433512]

59. Wong CH, et al. Selective inhibition of MBNL1–CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2). *Nucleic Acids Res.* 2011; 39:8881–8890. [PubMed: 21768123]
60. Cho J, Rando RR. Specific binding of Hoechst 33258 to site 1 thymidylate synthase mRNA. *Nucleic Acids Res.* 2000; 28:2158–2163. [PubMed: 10773086]
61. Simon RJ, et al. Peptoids: a modular approach to drug discovery. *Proc Natl Acad Sci U S A.* 1992; 89:9367–9371. [PubMed: 1409642]
62. Kolb HC, et al. Click chemistry: diverse chemical function from a few good reactions. *Angew Chem Int Ed Engl.* 2001; 40:2004–2021. [PubMed: 11433435]
63. Pushechnikov A, et al. Rational design of ligands targeting triplet repeating transcripts that cause RNA dominant disease: application to myotonic muscular dystrophy type 1 and spinocerebellar ataxia type 3. *J Am Chem Soc.* 2009; 131:9767–9779. [PubMed: 19552411]
64. Kimura T, et al. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca^{2+} -ATPase in myotonic dystrophy type 1. *Hum Mol Genet.* 2005; 14:2189–2200. [PubMed: 15972723]
65. Taneja KL, et al. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol.* 1995; 128:995–1002. [PubMed: 7896884]
66. Fu YH, et al. Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science.* 1993; 260:235–238. [PubMed: 8469976]
67. Jiang H, et al. Myotonic dystrophy type 1 is associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins and deregulated alternative splicing in neurons. *Hum Mol Genet.* 2004; 13:3079–3088. [PubMed: 15496431]
68. Mankodi A, et al. Nuclear RNA foci in the heart in myotonic dystrophy. *Circ Res.* 2005; 97:1152–1155. [PubMed: 16254211]
69. Wojciechowska M, Krzyzosiak WJ. Cellular toxicity of expanded RNA repeats: focus on RNA foci. *Hum Mol Genet.* 2011; 39:8938–8951.
70. Parkesh R, et al. Design of a bioactive small molecule that targets the myotonic dystrophy type 1 RNA via an RNA motif–ligand database & chemical similarity searching. *J Am Chem Soc.* 2012; 134:4731–4742. [PubMed: 22300544]
71. Warf MB, et al. Pentamidine reverses the splicing defects associated with myotonic dystrophy. *Proc Natl Acad Sci U S A.* 2009; 106:18551–18556. [PubMed: 19822739]
72. Gareiss PC, et al. Dynamic combinatorial selection of molecules capable of inhibiting the (CUG) repeat RNA–MBNL1 interaction *in vitro*: discovery of lead compounds targeting myotonic dystrophy(DM1). *J Am Chem Soc.* 2008; 130:16254–16261. [PubMed: 18998634]
73. Ofori LO, et al. From dynamic combinatorial ‘hit’ to lead: *in vitro* and *in vivo* activity of compounds targeting the pathogenic RNAs that cause myotonic dystrophy. *Nucleic Acids Res.* 2012; 40:6380–6390. [PubMed: 22492623]
74. Chen CZ, et al. Two high-throughput screening assays for aberrant RNA–protein interactions in myotonic dystrophy type 1. *Anal Bioanal Chem.* 2012; 402:1889–1898. [PubMed: 22218462]
75. Childs-Disney JL, et al. Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. *Nat Commun.* 2013; 4:2044. [PubMed: 23806903]
76. Garcia-Lopez A, et al. *In vivo* discovery of a peptide that prevents CUG-RNA hairpin formation and reverses RNA toxicity in myotonic dystrophy models. *Proc Natl Acad Sci U S A.* 2011; 108:11866–11871. [PubMed: 21730182]
77. Lee MM, et al. Rational and modular design of potent ligands targeting the RNA that causes myotonic dystrophy 2. *ACS Chem Biol.* 2009; 4:345–355. [PubMed: 19348464]
78. Lee MM, et al. Influencing uptake and localization of aminoglycoside-functionalized peptoids. *Mol Biosyst.* 2011; 7:2441–2451. [PubMed: 21611644]
79. Zuckermann RN, Kodadek T. Peptoids as potential therapeutics. *Curr Opin Mol Ther.* 2009; 11:299–307. [PubMed: 19479663]
80. Kwon YU, Kodadek T. Quantitative evaluation of the relative cell permeability of peptoids and peptides. *J Am Chem Soc.* 2007; 129:1508–1509. [PubMed: 17283989]

81. Unciti-Broceta A, et al. Synthesis, penetrability and intracellular targeting of fluorescein-tagged peptoids and peptide-peptoid hybrids. *Bioorg Med Chem*. 2009; 17:959–966. [PubMed: 18343124]
82. Tran T, Disney MD. Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat Commun*. 2012; 3:1125. [PubMed: 23047683]
83. Walters WP, et al. Recognizing molecules with drug-like properties. *Curr Opin Chem Biol*. 1999; 3:384–387. [PubMed: 10419858]
84. Mulders SA, et al. Triplet-repeat oligonucleotide-mediated reversal of RNA toxicity in myotonic dystrophy. *Proc Natl Acad Sci U S A*. 2009; 106:13915–13920. [PubMed: 19667189]
85. Wheeler TM, et al. Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. *Science*. 2009; 325:336–339. [PubMed: 19608921]

FIGURE 1.

Designing small molecules targeting RNA is challenging because there are few lead compounds. A large data set of leads is likely required to provide sufficient information to reveal general routes to target RNA because of its structural diversity. (a) Targeting approaches used for DNA were enabled, in part, by the structures of natural products, such as distamycin, bound to the minor groove. (b) Examples of the structural diversity of RNA in validated drug targets. (c) RNA forms many different structures that can be utilized as binding sites for small molecules, as is illustrated in the structures of RNA base pairs and examples of 1 × 1 nt RNA internal loops. *Abbreviation:* TAR, *trans*-activation response.

FIGURE 2.

Development of small molecules that target the RNAs that cause myotonic dystrophy (DM). (a) Myotonic dystrophy type 1 (DM1) is caused by a r(CUG) repeat expansion [r(CUG)^{exp}] that inactivates muscleblind-like 1 protein (MBNL1). Inactivation of MBNL1 causes dysregulation of pre-mRNA splicing. A modularly assembled compound that displays a *bis*-benzimidazole (**H**, Fig. 3) improves DM1-associated defects in cell culture models of disease [25,63]. (b) DM2 is caused by an expanded r(CCUG) repeat [r(CCUG)^{exp}], which also binds and inactivates MBNL1. A modularly assembled compound that displays **K** inhibits the r(CCUG)^{exp}-MBNL1 complex (Fig. 3) [77]. (c) Identification of a lead small molecule to target the r(CCUG)^{exp}. In two-dimensional combinatorial screening (2DCS) selections, the randomized regions of the RNA libraries are kept intentionally small such that they are found as components of cellular RNAs. The RNA motif libraries are then screened for binding small molecules. Collectively, these studies identified a set of RNA motif–small molecule interactions that can be used to design compounds against RNAs that cause or contribute to disease, such as DM2. In a resin-based selection, kanamycin A was used, which led to the development of the derivative **K** to target r(CCUG) repeats by using a modular assembly approach [24,77].

FIGURE 3.

Rationally designed, modularly assembled small molecules targeting the causative agents myotonic dystrophy type 1 (DM1) and type 2 (DM2), $r(\text{CUG})^{\text{exp}}$ and $r(\text{CCUG})^{\text{exp}}$, respectively [24,63,77]. RNA-binding modules (**H** or **K**) were modularly assembled onto a peptoid backbone and are potent inhibitors of the $r(\text{CUG})^{\text{exp}}$ –MBNL1 complex (**nH-4** compounds) or the $r(\text{CCUG})^{\text{exp}}$ –MBNL1 complex (**nK-4** compounds). **nH-4** compounds improved DM1-associated defects in cell culture models [25].

Drug Discovery Today

FIGURE 4.

Small molecules that improve myotonic dystrophy type 1 (DM1)-associated defects in cellular and/or mouse models of disease. **H1** was designed using chemical similarity searching [70], whereas Tri-Ac was designed using structure-aided design [55]. DCC-1 [72], pentamidine [71], **1** [75] and **2** [75] were identified via *in vitro* screening methods. D-Hexapeptide was identified via a screen in a *Drosophila* model of DM1 [76].