Published in final edited form as:

Bioorg Med Chem Lett. 2016 December 01; 26(23): 5792–5796. doi:10.1016/j.bmcl.2016.10.037.

Development of Pharmacophore Models for Small Molecules Targeting RNA: Application to the RNA Repeat Expansion in Myotonic Dystrophy Type 1

Alicia J. Angelbello^{a,c}, Àlex L. González^{b,c}, Suzanne G. Rzuczek^a, and Matthew D. Disney^a

^aDepartment of Chemistry, The Scripps Research Institute, Scripps Florida, 130 Scripps Way, Jupiter, Florida 33458, United States

^bGrup d'Enginyeria Molecular (GEM), Institut Químic de Sarrià (IQS) – Universitat Ramon Llull (URL), Barcelona, Catalonia, 08017, Spain

Abstract

RNA is an important drug target, but current approaches to identify bioactive small molecules have been engineered primarily for protein targets. Moreover, the identification of small molecules that bind a specific RNA target with sufficient potency remains a challenge. Computer-aided drug design (CADD) and, in particular, ligand-based drug design provide a myriad of tools to identify rapidly new chemical entities for modulating a target based on previous knowledge of active compounds without relying on a ligand complex. Herein we describe pharmacophore virtual screening based on previously reported active molecules that target the toxic RNA that causes myotonic dystrophy type 1 (DM1). DM1-associated defects are caused by sequestration of muscleblind-like 1 protein (MBNL1), an alternative splicing regulator, by expanded CUG repeats (r(CUG)^{exp}). Several small molecules have been found to disrupt the MBNL1-r(CUG)^{exp} complex, ameliorating DM1 defects. Our pharmacophore model identified a number of potential lead compounds from which we selected 11 compounds to evaluate. Of the 11 compounds, several improved DM1 defects both *in vitro* and in cells.

Graphical Abstract

To create your abstract, type over the instructions in the template box below. Fonts or abstract dimensions should not be changed or altered.

Correspondence to: Matthew D. Disney.

^cThese authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

RNA; Chemical biology; Pharmacophore; Virtual screening; Chemotype

The ENCODE project and other efforts are discovering disease-causing RNAs at a rapid pace. Of great interest is the development of lead therapeutics and chemical probes of function in order to target these RNAs with small molecules.^{1–3} In particular, microsatellite expansion disorders are a class of RNA-mediated diseases where the toxic RNA folds into an extended hairpin structure with regularly repeating internal loops.⁴ One example is myotonic dystrophy type 1 (DM1), which is an autosomal dominant, multisystemic disease with a core pattern of clinical presentation including myotonia, muscle weakness, cardiac conduction defects, posterior iridescent cataracts, and endocrine disorders.⁵ DM1 is caused by an expansion of an unstable CTG trinucleotide repeat (r(CUG)^{exp}) in the 3' untranslated region (UTR) of the dystrophin myotonia protein kinase (*DMPK*) gene.⁶ r(CUG)^{exp} causes disease via a gain-of-function mechanism by binding all three paralogues of MBNL.^{7, 8} MBNL1 is a key regulator of alternative splicing of several transcripts such as insulin receptor (*IR*),^{9, 10} cardiac troponin T (*cTNT*)¹¹ and muscle-specific chloride ion channel (*CLCN1*)¹².

Traditionally, structure and ligand-based drug design have been widely used in medicinal chemistry for the design and optimization of protein targeting agents. Recently, these methods have been successfully applied to RNA targets^{1, 13–17} and have been employed to identify small molecules that selectively bind r(CUG)^{exp} and improve DM1-associated defects both *in vitro* and in cells.^{2, 18–21} Our goal in this work was to identify novel scaffolds that favor binding to r(CUG)^{exp} by using a rational chemical selection. Among the ligand-based methods, pharmacophore screening is one of the most applied techniques for screening and selection of new compounds.^{22–25} Briefly, a pharmacophore is constituted by an ensemble of steric and electronic features necessary to optimize the supramolecular interactions between a ligand and a biological target. The main assumption is that a family of active compounds shares a set of common features that interact with a set of complementary sites on the biological target. Although pharmacophoric screenings have found widespread use in medicinal chemistry, they have not been extensively applied to nucleic acid-targeting strategies. In the course of identifying novel small molecule r(CUG)^{exp}-binders, we were particularly interested in substituted benzimidazoles, whose bioactivity and selectivity for r(CUG)^{exp} has been previously reported.²⁶ Pharmacophore modeling enables selection of small molecules with similar features to known binders; however, the screening is not limited to a particular molecular shape or scaffold. Thus, it can allow for rational identification of new chemotypes targeting a specific biomolecule.

First we proposed a training set containing a total of ten small molecules, that included a previously reported *bis*-benzimidazole (H1)² and nine small molecules from The Scripps Research Institute's small molecule library, which have been identified to have favorable features for binding RNA and improve splicing defects.²⁶ The main hypothesis is that if the ten active compounds share the same pharmacophoric features, other compounds with the same features should be active.

We then proposed several pharmacophore models by combining pairs of active compounds. The pharmacophore models were generated using LigandScout 3.03b and by systematically combining common features between pairs of compounds, resulting in 15 possible pharmacophore combinations (Table 1). Each pharmacophore contained specific information about the disposition in space of hydrogen bond donors and acceptors, and hydrophobic aromatic regions. Ideally, each pharmacophore is built from an active compound, hence should represent an 'active' pharmacophoric disposition. In order to assess the discriminatory capability of each pharmacophore model we included 310 inactive compounds (decoys) to the original training set. All of these molecules were filtered against each model in order to analyze both their sensitivity (ability to select active ligands) and specificity (ability to discard decoys). In this analysis, the best models have a good rate of true positives (molecules experimentally active and in compliance with the pharmacophore model) and a low rate of false positives (experimentally inactive and in good compliance with the pharmacophore model). Even though the best pharmacophore models had a reasonable sensitivity, a high number of false positives was always present. For this reason, we decided to build a consensus pharmacophore model which consisted of merging only those features that the models had in common. Models with good retrieval of active compounds, reasonable enrichment factor ($EF_{100\%} > 5$, hence early recovery of active molecules during the screening) and small hit overlap were combined into a consensus pharmacophore that comprised four pharmacophoric models (Figure 1A, models **1**, **3**, **10**, and **15**). The most observed common features among the pharmacophore models were hydrogen-bond acceptors and donors within the benzimidazole rings and two or three aromatic rings. Moreover, all the pharmacophores were required to be planar. The resulting consensus pharmacophore retrieved a 0.96 area under the curve after 100% of screened database ($AUC_{100\%}$) and a 7.2 of $EF_{100\%}$. In addition, the consensus model reported correctly 10 out of 10 active compounds, identified as shown in the receiver-operating characteristic (ROC) curve (Figure 1B), but 34 decoys were identified as false positives. Twelve of these false positive compounds, however, showed moderate $r(CUG)_{12}$ -MBNL1 complex disruption in *in vitro* assays (data not shown) confirming the comprehensive covering of potential hits by the pharmacophore.

Next, novel chemical entities were searched in a ~4.3 million lead-like subset ZINC database (250 MW 350, xlog *P* 3.5, rotatable bonds 7).^{27, 28} The selected small molecules had to fulfill the following requirements: (i) they must fit into the pharmacophoric keys; (ii) their scaffolds must be as different as possible; and (iii) their scaffold must be as different as possible from the training set. The reasoning behind requirements 2 and 3 was to obtain a diverse set of molecules, not a focused collection. Prospective pharmacophore virtual screening was achieved by selecting those molecules contained in the ZINC subset

that were the most dissimilar to the known active molecules. Thus, we ensured that the new chemical entities were not too similar in terms of structure, and presumably activity, to the previously tested compounds. At the same time, these molecules should match the consensus pharmacophore query. A total of 86 molecules were identified as potentially active candidates. Finally, the candidates were clustered according to scaffold similarity, and 11 molecules were selected by means of a diversity-based selection (Figure 2A). Not surprisingly, the benzimidazole scaffold appeared as a representative fragment among the candidates. Pyridine and furan carboxamides were present in eight compounds and were observed to be common fragments among the most active predicted molecules. By inspection of the compound structures, each has favorable properties to bind RNA by formation of hydrogen bonds and/or stacking interactions. Thus possible binding modes include groove binding, intercalation, or a combination thereof.

The ability of the selected candidates to disrupt a r(CUG)₁₂-MBNL1 complex was studied using a previously described time-resolved fluorescence resonance energy transfer (TR-FRET) assay.²⁹ Small molecules were screened at 50 μ M concentration (Figure 2B). The results of the screen showed that compounds **p1**, **p2**, **p4**, **p5**, **p7**, **p8**, and **p9** disrupted >20% of complex while the other three compounds did not disrupt the complex at this concentration (Figure 2B). Compound **p3** was not soluble under screening conditions and thus was not studied. We then assessed the bioactivity of compounds with greater than 20% complex disruption *in cellulis*.

We evaluated the ability of these compounds to stimulate nucleocytoplasmic transport of nuclearly retained r(CUG)^{exp} (sequestered in foci as RNA-protein aggregates) using C2C12 cells expressing a luciferase-encoding mRNA with 800 CUG repeats in the 3' UTR.¹⁸ Compounds disrupting the RNA-protein complex should improve nucleocytoplasmic transport of the mRNA and hence its translation, as measured by increased luciferase activity; compounds that cannot disrupt the complex will not exhibit turnover of a pro-luminescent substrate by luciferase. As a control, the same cell line expressing no CUG repeats was used to eliminate compounds that non-specifically enhance luciferase translation. The seven compounds that were active *in vitro* were tested in cells at 50, 25, and 12.5 μ M (Figure 3A). Of these seven candidates, six increased luciferase expression (**p8** was inactive), suggesting that they could increase nuclear transport. Three compounds (**p1**, **p2**, and **p7**) selectively increased luciferase in cells expressing the reporter containing CUG repeats. The other three compounds (**p4**, **p5**, and **p9**) increased luciferase signal in cells expressing 800 and 0 repeats by the same amount, suggesting off-target effects. The three compounds that were active in the luciferase assay (**p1**, **p2**, and **p7**) were used in additional *in cellulis* experiments to determine if they could improve DM1-associated splicing defects.

Bioactive compounds were tested for improving DM1-associated pre-mRNA splicing defects caused by sequestration of MBNL1. To do so, we co-transfected a DM1 mini-gene containing 960 CUG repeats and a mini-gene reporter for the alternative splicing of cardiac troponin T (*cTNT*) exon 5, which is deregulated in DM1.¹¹ In the absence of r(CUG)₉₆₀ the amount of mature mRNA containing exon 5 is about 60%, but in the presence r(CUG)₉₆₀ the amount of mRNA containing exon 5 is about 90% (Figure 3B). If small molecules bind to r(CUG)₉₆₀ and displace MBNL1, then the amount of mature mRNA with exon 5 should be

reduced. Small molecules that were selectively active in the luciferase assay (**p1**, **p2**, and **p7**) were tested for their ability to improve *cTNT* alternative splicing defects (Figure 3C). Each compound improved *cTNT* alternative splicing at 100 μM to varying extents. **p1** and **p7** were the most active compounds and both retained modest activity at 10 μM . Importantly, none of the compounds affect *cTNT* exon 5 alternative splicing in cells that do not express $\text{r(CUG)}^{\text{exp}}$ (Figure S2). A WST-1 assay was used to assess each compound's effect on cell viability. None of the compounds significantly affected cell viability (Figure S3). The two compounds that most potently improved *cTNT* alternative splicing defects (**p1** and **p7**) were evaluated for their effects on *DMPK* levels using RT-qPCR. While **p7** had no effect on *DMPK* levels, **p1** slightly increased *DMPK* levels (Figure S4).

Finally, the selectivity of **p7** was probed with *in vitro* binding assays. The K_d of **p7** to an RNA with one 5'CUG/3'GUC internal loop was $4.6 \pm 0.5 \mu\text{M}$ (Figure S5). The K_d to a GC paired RNA is $>>30 \mu\text{M}$. Likewise, the K_d to tRNA and AT hairpin DNA is $>>50 \mu\text{M}$ (Figure S5). Thus, **p7** was determined to be a selective binder of CUG that has the ability to improve DM1-associated splicing defects.

Analysis of the physicochemical properties of active and inactive compounds was completed in order to compare the compounds to the previously studied RNA-focused library (Table S1).²⁶ Collectively, the properties of the compound library in this study were not statistically different from the starting RNA-focused library. Additionally, the physicochemical properties of new bioactive compounds discovered using pharmacophore screening were not statistically different from previously discovered bioactive compounds. Importantly, although the properties were not statistically different, the scaffolds in the new bioactive compounds are novel.

In summary, ligand-based pharmacophore modeling provides an excellent tool for identifying novel chemotypes, even for targeting complex biological targets such as RNA. Herein we presented a rational chemical selection approach that allowed us to identify unreported scaffolds that improve DM1-associated defects such as reduced nucleocytoplasmic transport and deregulated alternative splicing. This study demonstrates the viability of conventional computer-aided drug design techniques to screen correctly and rapidly identify compounds from large chemical libraries for targeting RNA. As more information on the types of small molecules that target RNA is developed using methods such as Inforna^{14, 30}, these screening techniques may provide useful data for lead optimization of such compounds.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the National Institutes of Health (NIH; DP1NS096787 to MDD) and the Muscular Dystrophy Association (Grant #380467 to MDD) for funding.

References

1. Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, Kuntz ID. RNA. 2009; 15:1219–1230. [PubMed: 19369428]
2. Parkesh R, Childs-Disney JL, Nakamori M, Kumar A, Wang E, Wang T, Hoskins J, Tran T, Housman D, Thornton CA, Disney MD. J. Am. Chem. Soc. 2012; 134:4731–4742. [PubMed: 22300544]
3. Disney MD, Yildirim I, Childs-Disney JL. Org. Biomol. Chem. 2014; 12:1029–1039. [PubMed: 24357181]
4. Pearson CE, Nichol Edamura K, Cleary JD. Nat. Rev. Genet. 2005; 6:729–742. [PubMed: 16205713]
5. Mathieu J, Allard P, Potvin L, Prevost C, Begin P. Neurology. 1999; 52:1658–1662. [PubMed: 10331695]
6. Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, et al. Cell. 1992; 68:799–808. [PubMed: 1310900]
7. Kino Y, Mori D, Oma Y, Takeshita Y, Sasagawa N, Ishiura S. Hum. Mol. Genet. 2004; 13:495–507. [PubMed: 14722159]
8. Ho TH, Savkur RS, Poulos MG, Mancini MA, Swanson MS, Cooper TA. J. Cell Sci. 2005; 118:2923–2933. [PubMed: 15961406]
9. Lee JE, Cooper TA. Biochem. Soc. Trans. 2009; 37:1281–1286. [PubMed: 19909263]
10. Foff EP, Mahadevan MS. Muscle & Nerve. 2011; 44:160–169. [PubMed: 21607985]
11. Philips AV, Timchenko LT, Cooper TA. Science. 1998; 280:737–741. [PubMed: 9563950]
12. Charlet BN, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA. Mol. Cell. 2002; 10:45–53. [PubMed: 12150906]
13. Pushechnikov A, Lee MM, Childs-Disney JL, Sobczak K, French JM, Thornton CA, Disney MD. J. Am. Chem. Soc. 2009; 131:9767–9779. [PubMed: 19552411]
14. Velagapudi SP, Gallo SM, Disney MD. Nat. Chem. Biol. 2014; 10:291–297. [PubMed: 24509821]
15. Aboulela F. Future Med. Chem. 2010; 2:93–119. [PubMed: 21426048]
16. Guan L, Disney MD. ACS Chem. Biol. 2012; 7:73–86. [PubMed: 22185671]
17. Tran T, Disney MD. Nat. Commun. 2012; 3:1125. [PubMed: 23047683]
18. Childs-Disney JL, Hoskins J, Rzuczek SG, Thornton CA, Disney MD. ACS Chem. Biol. 2012; 7:856–862. [PubMed: 22332923]
19. Coonrod LA, Nakamori M, Wang W, Carrell S, Hilton CL, Bodner MJ, Siboni RB, Docter AG, Haley MM, Thornton CA, Berglund JA. ACS Chem. Biol. 2013; 8:2528–2537. [PubMed: 24028068]
20. Lee MM, Childs-Disney JL, Pushechnikov A, French JM, Sobczak K, Thornton CA, Disney MD. J. Am. Chem. Soc. 2009; 131:17464–17472. [PubMed: 19904940]
21. Childs-Disney JL, Stepniak-Konieczna E, Tran T, Yildirim I, Park H, Chen CZ, Hoskins J, Southall N, Marugan JJ, Patnaik S, Zheng W, Austin CP, Schatz GC, Sobczak K, Thornton CA, Disney MD. Nat. Commun. 2013; 4:2044. [PubMed: 23806903]
22. Dror O, Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. J. Chem. Inf. Model. 2009; 49:2333–2343. [PubMed: 19803502]
23. Spitzer GM, Wellenzohn B, Laggner C, Langer T, Liedl KR. J. Chem. Inf. Model. 2007; 47:1580–1589. [PubMed: 17518460]
24. Krautscheid Y, Senning CJ, Sartori SB, Singewald N, Schuster D, Stuppner H. J. Chem. Inf. Model. 2014; 54:1747–1757. [PubMed: 24849814]
25. Langer T, Wolber G. Drug Discov. Today Technol. 2004; 1:203–207. [PubMed: 24981486]
26. Rzuczek SG, Southern MR, Disney MD. ACS Chem. Biol. 2015; 10:2706–2715. [PubMed: 26414664]
27. Irwin JJ, Shoichet BK. J. Chem. Inf. Model. 2005; 45:177–182. [PubMed: 15667143]
28. Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. J. Chem. Inf. Model. 2012; 52:1757–1768. [PubMed: 22587354]

29. Chen CZ, Sobczak K, Hoskins J, Southall N, Marugan JJ, Zheng W, Thornton CA, Austin CP. *Anal. Bioanal. Chem.* 2012; 402:1889–1898. [PubMed: 22218462]
30. Disney MD, Winkelsas AM, Velagapudi SP, Southern M, Fallahi M, Childs-Disney JL. *ACS Chem. Biol.* 2016; 11:1720–1728. [PubMed: 27097021]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Pharmacophore screening and selection. (A) Pharmacophore hypothesis for models 1, 3, 10 and 15. LigandScout color codes were used to represent the pharmacophore features: hydrogen bond donor (green arrow), hydrogen bond acceptor (red sphere), and hydrophobic aromatic region (yellow sphere) (B) ROC curve of consensus pharmacophore after screening the 320 compound library. Area under the curve (AUC) and enrichment factor (EF) at 1, 5, 10 and 100% of screened database are presented. A total of ten positive and 34 false positive compounds were found as potentially active molecules after the screening.

A **B** **Figure 2.**

Compound selection and initial Screening. (A) Structures of the 11 compounds selected from the ZINC lead-like subset database, classified as potentially bioactive by the consensus pharmacophore. (B) TR-FRET screening of compounds for disruption of a $r(\text{CUG})_{12}$ -MBNL1 complex. Compounds were screened at 50 μM . Hits were defined as compounds that disrupted greater than 20% of the complex.

Figure 3.

Cellular activity of hit compounds. (A) Luciferase activity of **p1**, **p2**, **p4**, **p5**, **p7**, **p8**, and **p9** in reporters with 800 CUG repeats and no CUG repeats. Three compounds (**p1**, **p2**, and **p7**) were bioactive, increasing luciferase. (B) Schematic representation of the *cTNT* premRNA splicing pattern in the presence and absence of r(CUG)₉₆₀. (C) Quantification of *cTNT* splicing analysis and representative gel images of the two most active compounds (**p1** and **p7**) tested at 100, 10 and 1 μM .

Table 1

Number of true positives (*TP*) and false positives (*FP*) obtained for each pharmacophore. Enrichment factor (EF) and area under the curve (AUC) were used as metrics in order to assess each model's quality. The EF defines the early recovery of active molecules during the screening; the AUC varies between 0 and 1 and quantifies the predictive capacity of the model. Both parameters were assessed at 1, 5, 10 and 100% of screened database. The hit rate states the percentage of *TP* found after screening the whole database.^a

Model	TP	FP	AUC _{1%}	AUC _{5%}	AUC _{10%}	AUC _{100%}	EF _{1%}	EF _{5%}	EF _{10%}	EF _{100%}	Hit rate (%)
1	3	7	0.68	0.94	0.97	0.64	10.6	9.5	9.5	9.5	30
2	2	51	0	0	0.38	0.51	0	0	1	1.2	20
3	3	5	0.68	0.94	0.97	0.64	21.2	11.9	11.9	11.9	30
4	7	52	0	0.74	0.87	0.8	0	10.6	6.2	3.8	70
5	2	4	0.68	0.94	0.97	0.59	21.2	10.6	10.6	10.6	20
6	6	103	0	0.16	0.58	0.67	0	2.1	1	1.8	60
7	9	134	0.68	0.94	0.97	0.68	10.6	2.1	1	2	90
8	8	66	0	0.68	0.84	0.81	0	4.2	5.1	3.4	80
9	3	13	0.68	0.94	0.97	0.63	21.2	6.4	6	6	30
10	5	21	0	0.74	0.87	0.72	0	8.5	6.1	6.1	50
11	3	2	1	1	1	0.65	31.8	19.1	19.1	19.1	30
12	1	6	0.68	0.94	0.97	0.54	10.6	4.5	4.5	4.5	10
13	4	34	0	0.55	0.77	0.65	0	6.4	4.1	3.3	40
14	3	22	0	0.74	0.87	0.62	0	4.2	3.8	3.8	30
15	5	15	1	1	1	0.73	31.8	10.6	7.9	7.9	50

^aPlease see supporting for full definitions of the parameters