DEVELOPING POTENTIAL DRUGS FOR THE TREATMENT OF MYOTONIC DYSTROPHY: FROM RATIONAL DESIGN TO LEAD DEVELOPMENT

BY

CHUN-HO WONG

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Doctoral Committee:

Professor Steven C. Zimmerman, Chair and Director of Research
Professor Paul J. Hergenrother
Assistant Professor Douglas A. Mitchell
Adjunct Professor Anne M. Baranger, University of California, Berkeley
ABSTRACT

Myotonic dystrophy (DM) is currently an incurable genetic disease that affects 1 in 8,000 humans worldwide. Although extensive efforts have been made to understand its pathogenesis, the mechanism by which DM causes its symptoms is not fully understood. Nevertheless, it is known that the alternative splicing regulator, MBNL1, is sequestered by two types of abnormally long RNAs. Specifically there are the tri- and tetra-nucleotide repeats, CUG and CCUG, in DM1 and DM2, respectively. This key discovery inspired the development of ligands that inhibit MBNL1-RNA complex formation, allowing MBNL1 to resume its biological functions. The background to the DM disease will be given in Chapter 1.

In 2009, our group reported a rationally designed ligand (11) that recognizes the base mismatches in CUG repeats found in type 1 myotonic dystrophy (DM1). Although little was known about the binding mode, this lead compound disrupts the MBNL1-RNA complex \textit{in vitro}. Attempts to obtain a crystal of the RNA-ligand complex for X-ray analysis were unsuccessful. As part of an alternative strategy, we have synthesized a small library of lead-like compounds to better understand the structure-activity relationship. In Chapter 2, possible binding modes of the lead ligands are elucidated using a combination of synthesis, gel shift assay, and molecular dynamics simulation. During the course of study, a lead compound for the type 2 myotonic dystrophy (DM2) was discovered. The synthesis, biophysical studies, and \textit{in vitro} activity of this compound is described in Chapter 3. A new class of potential DM drugs has been discovered using a reported structure of a HIV RNA-ligand complex, combined with the knowledge gained in studying ligand 11. These ligands possess better aqueous solubility and lower toxicity. More importantly, they are shown to be active in a DM1 \textit{Drosophila} model. The design, synthesis, and evaluation of these ligands are described in Chapter 4.
To my wife and parents
ACKNOWLEDGEMENTS

My sincere thanks first and foremost go to my advisor – Prof. Steven C. Zimmerman who provided guidance, encouragement, and resources that sustained me throughout my Ph.D. studies. He is not just a great scientific advisor but has also been my great mentor in past five years. I would also like to express my special thanks to my original doctoral committee members Prof. Scott Denmark and Prof. Wilfred van der Donk for providing their valuable suggestions and guidance from year to year. I want to gratefully acknowledge Prof. Anne Baranger for her continuous encouragement and support even after leaving Illinois. I would also like to thank Prof. Paul Hergenrother and Prof. Douglas Mitchell for their interest in my project and for serving on my Ph.D. committee.

I would like to thank Eric Yen-Jun Ho, Dr. Stacie Richardson, Dr. Sreenivasa Ramisetty, Dr. Yuan Fu, and Prof. Tiziano Tuccinardi for their insightful discussions and experimental support over the years. I am also deeply indebted to a group of talented young scientists, Alex Lucas, Long Luu, Jeannette Sanchez, Lien Nguyen, and Jessie Peh, that I have worked alongside. It has been a pleasure moving this Myotonic Dystrophy project forward with them. I would also like to thank Prof. Edwin Chan, Prof. Woody Chan, and Prof. Sherlock Lam at the Chinese University of Hong Kong for their collaborative efforts and trust in me. I especially thank Dr. Si-Kyung Yang for his moral support and thank other members of Prof. Zimmerman’s group for helpful discussions and creating a nice atmosphere for research.

Lastly, I wish to thank my parents and my wife Joyce for their love and encouragement throughout my stay in US and their unwavering support whenever there is a tough decision to be made.
# TABLE OF CONTENTS

**Chapter 1**  Introduction .................................................................................................................. 1  
  1.1 Background to Myotonic Dystrophy ......................................................................................... 1  
  1.2 Disease Mechanism and RNA Gain-of-Function Toxicity ......................................................... 2  
  1.3 RNA as a Viable Target for Treating DM .................................................................................. 4  
  1.4 Antisense Strategy ..................................................................................................................... 5  
  1.5 Inhibition by Small Molecules Discovered from Screening .................................................... 7  
  1.6 Oligomerization of CUG-binders ............................................................................................... 10  
  1.7 Lead Development in Our Laboratory ....................................................................................... 11  

**Chapter 2**  Investigating the Binding Mode of the Ligand-RNA Complex .......................... 14  
  2.1 Introduction ............................................................................................................................... 14  
  2.2 Structures of the CUG Repeats .............................................................................................. 16  
  2.3 Elucidating the Modes of Binding ........................................................................................... 18  
  2.4 Methyl Scanning Approach .................................................................................................... 21  
  2.5 Conclusions .............................................................................................................................. 28  

**Chapter 3**  Lead Discovery for DM2 ..................................................................................... 30  
  3.1 Introduction ............................................................................................................................... 30  
  3.2 Targeting C-T and C-U Mismatches ....................................................................................... 31  
  3.3 Discovery of DM2 Lead Compounds ....................................................................................... 38  
  3.4 MBNL1-CCUG Inhibition by DM2 Lead Compounds .............................................................. 41  
  3.5 Biophysical Studies of the DM2 Lead Compounds ................................................................. 43  
  3.6 Conclusions .............................................................................................................................. 48  

**Chapter 4**  Bisamidinium-Based Inhibitors ........................................................................... 50  
  4.1 Introduction ............................................................................................................................... 50  
  4.2 Rational Design of the Bisamidinium Inhibitors ...................................................................... 51  
  4.3 Synthesis of Bisamidinium Inhibitors ..................................................................................... 54  
  4.4 Foci Disruption and Splicing Correction by Bisamidinium Inhibitors ..................................... 58  
  4.5 Biophysical Studies of the Bisamidinium Inhibitors ................................................................. 63  
  4.6 Activity in a DM1 Drosophila Model ....................................................................................... 65  
  4.7 Conclusions and Outlook ......................................................................................................... 67  

**Chapter 5**  Materials and Methods ......................................................................................... 69  

**Chapter 6**  References ................................................................................................................... 79  

Appendix: Publications ......................................................................................................................... 85
Chapter 1 Introduction

1.1 Background to Myotonic Dystrophy

Myotonic dystrophy (dystrophia myotonica or DM) is currently an incurable, inherited neuromuscular disease with a prevalence of 1 in 8,500 worldwide.¹ DM is characterized by its highly variable age of disease onset with a wide range of multisystemic symptoms (Table 1). Medical managements such as physiotherapy and supportive devices are available for only some of the symptoms. In some cases, surgery is required for example for drooping eyelids (ptosis) and cataracts. There are a limited number of drugs available for relieving symptoms; mexiletine² was recently found to be effective in reducing myotonia but none of them cures the disease (Chart 1). Currently, two major types of DM have been identified: type 1 (DM1) and type 2 (DM2). One major difference between the two types is the lack of congenital form in DM2. The symptoms for DM2 are

Table 1. Common clinical features of DM1 and DM2 and their management.

<table>
<thead>
<tr>
<th>Symptom</th>
<th>DM1</th>
<th>DM2</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muscle weakness</td>
<td>++</td>
<td>+</td>
<td>Drugs for symptoms; occupational therapy; physiotherapy</td>
</tr>
<tr>
<td>Myotonia</td>
<td>++</td>
<td>+</td>
<td>Mexiletine; heated gloves</td>
</tr>
<tr>
<td>Calf hypertrophy</td>
<td>−</td>
<td>+</td>
<td>No specific management</td>
</tr>
<tr>
<td>Cataract</td>
<td>+</td>
<td>+</td>
<td>Eye assessment; cataract removal</td>
</tr>
<tr>
<td>Diabetes</td>
<td>+</td>
<td>+</td>
<td>Glucose measurement; diet modification</td>
</tr>
<tr>
<td>Cardiac problems</td>
<td>++</td>
<td>±</td>
<td>Annually to hourly electrocardiography (ECG); pacemakers</td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>++</td>
<td>−</td>
<td>Respiratory assessment; nocturnal non-invasive ventilation</td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>+</td>
<td>−</td>
<td>Cholestyramine; gastroenterology assessment</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>+</td>
<td>−</td>
<td>Speech therapy</td>
</tr>
<tr>
<td>Developmental disability</td>
<td>+</td>
<td>−</td>
<td>Physiotherapy; occupational therapy; pre-natal testing</td>
</tr>
</tbody>
</table>

+ +: present and prominent; +: present; −: not present; ±: minor
often milder compared to those for DM1 and the disease progresses more slowly. Unlike DM1 which is present in all most countries, DM2 was mainly found in Europe. It was further suggested that the prevalence of DM2 in Germany could be as high as that of DM1. Like DM1, DM2 is also incurable.

1.2 Disease Mechanism and RNA Gain-of-Function Toxicity

The discovery of the mutated genes for DM1 and DM2, in 1992 and 2001, respectively, has led to a burgeoning research effort focused on understanding the disease mechanism and development of therapeutic strategies. It was discovered that DM1 and DM2 are caused by unstable nucleotide repeat expansions in DNA. DM1 is caused by the trinucleotide CTG expansion, from a normal length of <37 up to 2,000 or higher in those with the disease. The expansion is located in the 3'-untranslated region (3'-UTR) of the dystrophia myotonica protein kinase (DMPK) gene (Figure 1a), whereas DM2 is associated with a tetranucleotide CCTG repeat expansion (from <30 to 11,000 repeats) in the intron 1 of the zinc-finger 9 (ZNF9) gene (Figure 1b). It was shown that the severity of the disease increases and the age of onset decreases with increasing number of repeats of the genes. The expansion also grows progressively longer during the patients lifetime. Several disease mechanisms have been proposed; 

Chart 1. Drugs for relieving some of the DM symptoms.
currently, the generally accepted hypothesis is a toxic RNA gain-of-function model\textsuperscript{14,15} for the non-coding RNA transcripts of the expanded DNA (Figure 1c). These mutant RNA sequences, (CUG)$_n$ and (CCUG)$_n$ for DM1 and DM2, respectively, form nuclear inclusions which in turn sequester an alternative splicing protein, muscleblind-like 1 (MBNL1) (Figure 1d). These MBNL1-RNA complexes were observed as nuclear foci in both DM1 and DM2 cells using immunofluorescence and fluorescence \textit{in situ}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{(a) Location of trinucleotide CTG repeats in the 3'-UTR of the \textit{DMPK} gene on chromosome 19. (b) Location of tetranucleotide CCTG repeats within intron 1 of the \textit{ZNF9} gene on chromosome 3. (c) Expanded transcripts from these genes form stable stem-loop structures with U-U mismatch in DM1 and C-U mismatch in DM2. (d) MBNL1 proteins were sequestered by the expanded RNAs inside the nucleus and the resulting foci can be seen as green spots under confocal microscope using immunofluorescence staining. (e) MBNL1 sequestration leads to misregulated alternative splicings of > 100 pre-mRNAs.}
\end{figure}
hybridization (FISH) techniques. The depletion of MBNL proteins leads to global mis-regulated alternative splicings of >100 pre-mRNAs, which in turn leads to the multisystemic symptoms of the DM diseases (Figure 1e).

What is the mechanism for repeat expansion? Most studies suggest the formation of unusual DNA structures during DNA metabolism that are the key determinants of the expansion. The exact mechanisms of how these unusual structures lead to repeat expansion beyond the normal range (< 30–40 repeats) remains controversial; different or even contradictory results have been obtained across different organisms, such as yeast, mouse, and human models. The repeat expansion is also found to be depending on the specific tissue- and cell-type within an individual as well as the stages of development. This complexity is further increased by the number of proteins and pathways involved during DNA replication, repair, recombination, and transcription. Investigations of the expansion mechanism are ongoing but, at this writing, targeting the expanded DNA with the hope of slowing, or even reversing the CTG expansion, is yet to be a viable strategy.

1.3 RNA as a Viable Target for Treating DM

As described previously, although the DNA expansion mechanism is not known, most results suggest a toxic gain-of-function of the non-coding, expanded RNA as the cause of multisystemic symptoms of DM. There are multiple strategies that can be applied at the RNA level including degrading or repairing the expanded RNA. Among these strategies, inhibiting the MBNL1-RNA interaction, thereby releasing MBNL1 to resume its biological functions, is perhaps one the of most promising approaches. In 2006,
Swanson and co-workers reported a significant reversal of missplicing of a major skeletal muscle chloride channel (Clcn1) by overexpression of the Mbnl1 protein in mouse muscle.\textsuperscript{21} A substantial reduction of myotonia was observed using electromyography. This result clearly indicated the importance of having a significant level of free MBNL1 inside the cells and suggested the idea of inhibiting the MBNL1-RNA interaction as a therapeutic strategy.

1.4 Antisense Strategy

One of the appealing approaches is the use of antisense oligonucleotides (AONs) to hybridize with the mutant RNA to release MBNL1 from the foci.\textsuperscript{22,23} Whereas common antisense strategies involve RNase H or other nuclease-mediated pathways to cleave the target RNA, Thornton and co-workers reported the use of the antisense morpholino oligonucleotide agent, CAG25, to inhibit MBNL1-CUG interaction without triggering the

![Figure 2](image_url)

**Figure 2.** Comparative structure of CAG antisense oligonucleotide: (a) phosphorodiamidate morpholino oligonucleotide (PMO), (b) 2'-O-methylated phosphorothiolate (2'-OMe-PT), (c) r(CAG) segment in hU7-snRNA, and (d) locked nucleic acid (LNA) with phosphorothiolate linkage.
cleavage of the expanded RNA, thereby lowering the risk of degrading other CUG-containing transcripts (Note: CAG25 is a 25-nt antisense morpholino oligonucleotides containing 7 CAG repeats). CAG25 (Figure 2a) not only dispersed the nuclear foci but also restored the expression of ClC-1 channels in tibialis anterior (TA) muscle of HSA LR transgenic mice. Similar effects were also observed by Wansink and co-workers using an 2'-O-methyl phosphorothioate (PT) (CAG)7 antisense oligonucleotide (Figure 2b). It is noteworthy that both PMO and 2'-OMe-PT AONs are not expected to mediate RNase-based degradation because of their structures. Therefore, the mechanism for the unpredicted reduction of the expanded transcripts using these AONs remains unclear. Furling and co-workers reported the transduction of an engineered human U7 small nuclear RNAs (hU7-snRNAs) containing a (CAG)15 segment into human DM1 muscle cell (Figure 2c). The transduced DM cells expressed the hU7-(CAG)15 which in turn inhibited the MBNL1-CUG interaction, thereby removing the need of repeated administration of AONs. Recently, Thornton and co-workers showed that locked-nucleic acid (LNA)-based AONs (Figure 2d) not only reduced the amount of nuclear foci but also reduced the average length of CTG repeats. However, the mechanism that led to the contraction is still unknown.

Although antisense technology yielded promising results in cell and animal models, there are still several limitations that need to be addressed before entering clinical trials, namely, delivery, distribution, and toxicity. The current results with AONs have relied on local injection into a particular muscle, so reversal of the phenotype in an organism as a whole has yet to be demonstrated. It is also observed that the inter-tissue spreading is minimal for 2'-OMe-PT-(CAG)7 suggesting the difficulty of achieving a
global phenotype reversal through single injection. This low dispersion rate is not surprising given the high molecular weights of these AONs (6,400–7,500 Da). These challenges are not unique to the treatment of DM, given the history of issues associated with the safe and selective delivery of AONs in the treatment of other diseases. Oral administration has yet to be shown as a viable strategy. Another major concern is the long-term toxicity profile of these AONs. Commonly observed toxicities are resulted from off-targeting binding and undesirable immune activation. Like other drugs, the toxicity of AONs was observed at higher doses. The question remains whether one can treat DM with antisense technology with selective delivery at a safe dose range.

1.5 Inhibition by Small Molecules Discovered from Screening

While this project was ongoing, Miller and co-workers reported the first small molecule inhibitors for treating myotonic dystrophy in 2008. By screening a resin-bound dynamic combinatorial library (RBDCL) of 11,325 members, the authors discovered three lead compounds capable of inhibiting MBNL1-CUG interaction in vitro with low micromolar $K_d$ and $K_i$ values (Chart 2). In 2012, the same group reported a new series of compounds from RBDCL and found that derivatives 4 and 5 are stronger CUG binders and partially restore missplicing in a DM1 mouse model.

By a screening of 26 known nucleic acid binders, Berglund and co-workers discovered that both pentamidine (6) and neomycin B (7) disrupt the MBNL1-CUG interaction in vitro but only pentamidine reversed the missplicing of cardiac troponin T (cTNT) and insulin receptor (IR) pre-mRNAs in HeLa cells (Chart 3). The authors also found that pentamidine partially reversed the missplicing of the chloride-1 (Clc-1) and
sarco/endoplasmic reticulum Ca\textsuperscript{2+}-ATPase-1 (Serca1) pre-mRNAs in DM1 mouse model and, thus, pentamidine became the first small molecule that was shown to be active in animal model. Unfortunately, the therapeutic window was narrow because of pentamidine’s high toxicity.\textsuperscript{32} For example, a treatment of 30 mg/kg dosage twice a day was lethal to mice while 40 mg/kg per day partially reversed the missplicing of the Clc-1
and Serca1 mRNAs. Pentamidine is also known to bind to various protein\textsuperscript{33,34} and nucleic acid targets.\textsuperscript{35-38} Although the selectivity of pentamidine is poor, the report indicated that small molecule inhibition is a viable approach for the treatment of DM1 in animal model.

Artero and co-workers identified a D-amino acid hexapeptide (8; Ac-ppyawe-NH\textsubscript{2}) from a combinatorial peptide library that suppressed DM1 phenotypes in both Drosophila and mouse models (Chart 4).\textsuperscript{39} It was observed that the hexapeptide and Drosophila muscleblind (Mbl) aggregate with the r(CUG)\textsubscript{23} which prevented the authors from quantifying the binding and inhibiting activities \textit{in vitro}. Based on the circular dichroism (CD) studies, it was suggested that the hexapeptide destabilized the duplex

\textbf{Chart 3.} Active compounds discovered by Berglund and co-workers.

\textbf{Chart 4.} Structure of the D-amino acid hexapeptide (8) discovered by Artero and co-workers.
region of the \(r(CUG)_{60}\) hairpin thereby reducing the affinity of the RNA towards the Mbl protein.

### 1.6 Oligomerization of CUG-binders

In principle, an oligomeric ligand has a higher affinity toward the target because of the multivalent binding.\(^{40}\) Oligomeric ligands are also expected to provide better inhibition because of a larger steric coverage of the target. By taking advantage of the repetitive nature of the toxic RNA, Disney and co-workers reported a repertoire of peptoid-based oligomeric ligands for inhibiting MBNL1-RNA interactions for DM1\(^{41}\) and DM2\(^{42}\) (Figure 3). As expected, these oligomers bind much more strongly to the toxic RNA than the

![Chemical structures](image)

**Figure 3.** Disney's oligomeric binders for (a) CUG repeats in DM1 and (b) CCUG repeats in DM2.
corresponding monomers. For example, pentamer 9 and hexamer 10 bind to CUG and CCUG repeats, respectively, with low nanomolar $K_d$ values. These tighter RNA binders are also better inhibitors of the MBNL1-RNA interaction with IC$_{50}$ in the low nanomolar range. Although these oligomers have high molecular weights (MW = 5,122 for 9 and 3,438 for 10), they are shown to be cell-permeable to mouse myoblasts.$^{41,42}$ In both series of oligomers, the $K_d$ and IC$_{50}$ values improves with increasing number of repeating units. To further improve the bioavailability of these oligomers, the authors reported the use of d-Arg$_9$-conjugated peptoids to facilitate the cellular uptake.$^{43}$ Substantial improvements in IC$_{50}$ and splicing correction were observed with the d-Arg$_9$-conjugated oligomers.$^{43}$ In Chapter 4, we proposed an “in situ” ligand oligomerization approach that takes advantages of both the higher bioavailability of monomers and the high potency of the resulting oligomers.

1.7 Lead Development in Our Laboratory

Based on an X-ray structure (PDB: 1ZEV) reported by Berglund in 2005 (Figure 4a),$^{44}$ we reported a class of ligands consisting of an intercalator (11) linked to a “Janus-wedge”$^{45}$ recognition unit that showed low micromolar affinity to CTG and CUG repeats (Figure 4b).$^{46}$ By linking a recognition unit (a triaminotriazine in this system) to the acridine unit by a short tetramethylene linker with the hope of reducing non-selective intercalation by the formation of a π-stacked conformation in aqueous medium (Figure 5a). The triaminotriazine moiety is designed to function as a Janus-wedge being capable of binding two thymine or uracil units simultaneously from the major or minor groove if the hydrogen bonding complementarity is the only consideration (Figure 5b).
The π-stacked acridine ring intercalates between the mismatch and the neighbor Watson-Crick base pair (Figures 4b and 5c). The study of binding mode of ligand 11 to CTG and CUG repeats will be presented in Chapter 2. Using isothermal titration calorimetry (ITC), ligand 11 was shown to bind to CTG and CUG repeats with low-micromolar affinity (0.5–0.7 $\mu$M).\(^{46}\) In addition, ligand 11 exhibited moderate to high selectivity (8-fold weaker CCG and >70-fold weaker to CAG and CGG). More importantly, it was shown by electrophoretic mobility shift assay (EMSA) that the ligand inhibited MBNL1 protein from binding to r(CUG)\(_{12}\) with low-micromolar \(K_i\) (\(~7\) $\mu$M). However, moderate toxicity and non-selectivity intercalation were observed in various cellular studies.

**Figure 4.** (a) X-ray structure r(CUG)\(_6\) duplex (PDB: 1ZEV) showing the characteristic U-U mismatches (in white) sandwiched between CG steps (in green). (b) Molecular modeling of ligand 11 (in blue) binding to r(CUG)\(_6\) duplex.

**Figure 5.** (a) Ligand 11 contains a triaminotriazine unit designed for selective targeting T-T and U-U mismatches. (b) Proposed Janus-wedge binding via complementary hydrogen bonds. (c) Proposed binding mode of ligand 11 to CTG or CUG mismatches.
conducted in the Hergenrother group at UIUC. Solubility problems were also encountered in several ITC and EMSA assays. Nonetheless, ligand 11 is a good lead compound because of its high affinity and merits further structural modifications to achieve better properties. In fact, we discovered two lead ligands that are structurally similar to ligand 11 for potential treatment of DM2 and the results will be presented in Chapter 3. In 2011, a promising lead was designed based on the knowledge and results generated in the past. In collaborations with other research groups in the United States and Hong Kong, the biological activity of the new lead compounds has been extensively studied. Many of these results will be presented in Chapter 4.
Chapter 2  Investigating the Binding Mode of the Ligand-RNA Complex

2.1 Introduction

As mentioned in Section 1.6, our first lead compound (11) was designed based on the crystal structure of r(CUG)$_6$ duplex reported by Berglund in 2005. We believe that we can design better compounds if we have a better understanding of how the ligands interact with the CUG repeats. After our first report of ligand 11 in 2009, we initiated a collaboration with the Berglund lab attempting to grow a crystal for solving the structure

Chart 5. Derivatives of ligand 11 with improved aqueous solubility.

---

of the RNA-ligand complex. Unfortunately, we cannot obtain any crystal and we believe that one of the reasons for the failure is the high percentage (5%) of DMSO used in the crystallization settings. Although the in vitro properties of ligand 11 are promising, we need to prepare derivatives with better aqueous solubility not only for preparing crystal structures but also for cellular and animal studies. Thus, after modifying the acridine ring of 11 to carry an ammonium ion, we are able to obtain several derivatives (e.g., 12 and 13) with improved aqueous solubilities (~0.5 vs. <0.1 mg/mL). These compounds have activities comparable to that of ligand 11 in terms of RNA binding and inhibition of the MBNL1-CUG complexation. Only 1% (v/v) of DMSO was used in those assays (vs 5–10% DMSO for ligand 11). In collaboration with Professor Raven Huang at UIUC, we attempted to grow co-crystals of the ligand 11 with CTG duplexes (i.e., DNA sequences) along with its aqueous soluble derivatives 12 and 13. We were able to obtain crystals with ligand 12 but the quality was not suitable for X-ray crystallography. We are currently solving an NMR solution structure with compounds 11–13 in collaboration with Professor Sherlock Lam at the Chinese University of Hong Kong.

In the absence of structural information, we set out to establish the structure-activity relationship using an indirect approach. Using a combination of biophysical and molecular dynamics (MD) simulations, this study supports distinct binding modes for the recognition of CUG sites in RNA and CTG sites in DNA. More importantly, we discovered a number of ligands that are CTG-selective but show no measurable affinity for RNA, potentially opening an avenue for a DNA-targeted molecular therapy for DM1.
2.2 Structures of the CUG Repeats

At this writing, there are a total of 6 reports on the structures of CUG repeats (5 reports on X-ray\textsuperscript{44,48-51} and 1 on NMR structures).\textsuperscript{52} All CUG-containing structures generally adopted standard A-form RNA duplex conformations with a shallow and wide minor groove and a deep and narrow major groove (Figure 6).\textsuperscript{53} Three of the representative CUG structures are shown in Figures 6b–d showing the characteristic A-form features. All of them contain U-U mismatches sandwiched between CG steps. The U-U mismatches have a C1'-C1' distance (~10.4 Å) that is close to the average value of a standard A-form duplex.\textsuperscript{51} Interestingly, this type of U-U mismatch appears to be an intermediate structure between the cis-U-U\textsuperscript{54} and trans-U-U pairs (Figure 7).\textsuperscript{55,56} The U-U mismatch in CUG repeats has more hydrogen bonding sites available for recognition.

\textbf{Figure 6.} (a) Structure of a typical A-form RNA duplex (PDB: 1QC0). Three representative CUG repeats structures adopting A-form RNA conformation (b)–(d).
than a cis-U-U pair but the C1'-C1' distance is not so lengthened that will distort the helical backbone. Thus, the overall structure of CUG repeats do not deviate from a typical A-form helix even with the recurring U-U mismatches.

The dynamics of the U-U pairs was reported by Disney and co-workers in 2011.\textsuperscript{52} Combined with NMR and molecular dynamics (MD) simulations, the authors suggested that these U-U pairs are weakly paired and the majority (relative populations = 76.5\%) of the mismatches are singly hydrogen bonded. This result is in good agreement with the crystal structures described above (see Figure 7b; PDB: 3GM7). Very few doubly hydrogen bonded U-U pairs were observed (7.6\%) and the pairs showed relatively short C1'-C1' distance (8.9 Å) signifying a distortion in the RNA helical backbone. Interestingly, a substantial number (15.9\%) of the observed U-U pairs are not hydrogen bonded. They are water-bridged and have a similar C1'-C1' distance to standard A-helix. Similar results were also seen in the MD simulations performed by our collaborator, Professor Tiziano Tuccinardi at the University of Pisa, Italy (\textit{vide infra}).
2.3 Elucidating the Modes of Binding

Although there is a lack of X-ray or NMR structure of the RNA-ligand complex, we have evidence of the intercalative mode of binding from a circular dichroism (CD) titration of ligand \(11\) to \(r(CUG)_{12}\) (Figure 8). The CD spectra showed an increase in CD signal from negative to positive from 280–300 nm and a decrease in signal <260 nm to more negative indicating of an intercalation of the acridine moiety to an A-form duplex.\(^{57}\) An intriguing question how the triaminotriazine moiety interact with the RNA remains unanswered. As reported in 2009, the mismatch in the DNA duplex is important for binding suggesting the triaminotriazine is interacting with the mismatch.\(^{46}\) In addition, in a thermal denaturation study the binding of ligand \(11\) is substantially weakened when the triaminotriazine moiety is replaced by a 2-aminopyridine unit.\(^{46}\) Very recently, one of our collaborators, Professor Sherlock Lam at the Chinese University of Hong Kong, observed a perturbation in NOE signal of the imino proton of a T-T mismatch upon

\[\text{Conditions:} \]
\[
[r(CUG)_{12}] = 25 \mu M \\
[MOPS] = 20 mM \\
[NaCl] = 300 mM \\
\text{Temperature} = 25 ^\circ C \\
\text{pH} = 7
\]

**Figure 8.** CD titration of ligand \(11\) to \(r(CUG)_{12}\).
titration of ligand 11 (unpublished result). We expect that the triaminotriazine unit is likely interacting the U-U pairs in CUG repeats.

With the reported CUG repeat structures (Figure 6), we considered two possible triplet modes of binding of ligand 11 to CUG repeats (Figures 9a,b). Simple modeling suggested that binding of ligand from the minor groove would require a lengthening of the C1'-C1' distance of the U-U pair from 10.4 to 13.8 Å. Although the RNA minor groove is readily accessible by small molecules, 3.4 Å is a significant increase in the C1'-C1' distance (Figure 9a) and may signal a distortion in the DNA or RNA backbone that is energetically unfavorable.58,59 This distance is even longer than that of a G-A purine-purine pair (12.5 Å) in a DNA duplex.60 In fact, repeated attempts to model [r(CUG)₆]₂⁻·11 with the ligand in the minor groove in failed to yield a stable complex. On the other hand, modeling of a major groove base-triplet suggested a C1'-C1' distance fairly close to that in the X-ray structure (10.1 vs 10.4 Å; Figure 9b). Indeed, an unconstrained 10 ns MD simulation showed ligand 1 could fit satisfactorily in the major groove (Figure 10a). The average C1'-C1' distance of the complexed U-U pair was also

Figure 9. Three potential modes of binding: (a) Minor groove triplet binding. (b) Major groove triplet binding. (c) Minor groove stretched wobble binding. (d) A ChemDraw representation of Figure 7b (PDB: 3GM7). R = CH₃ (DNA) or H (RNA).
analyzed (Figure 10b). Clearly, this binding mode does not introduce significant deviation in the C1'-C1' distance.

The de-twinned CUG repeat structure (PDB: 3MG7) with high-resolution hydration details (Figure 7b) led us to consider another possible binding mode (Figure 9c). Modeling showed that the water molecule bridging the two uracil bases in a stretched wobble pair (Figure 9d) could be replaced by the triazine unit in the minor groove (Figure 9c). The C1'-C1' distance of the binding mode is moderately lengthened by 1.4 Å (11.8 vs 10.4 Å). The MD simulation of this minor groove stretched wobble
binding was also performed (Figure 10c). Although fewer hydrogen bonds are formed, the unconstrained simulation led to a stable structure. The average C1’-C1’ distance of the U-U mismatch is moderately lengthened by 1.4 Å (11.8 vs 10.4 Å, see Figure 10d).

2.4 Methyl Scanning Approach

To distinguish between these models, an extended “methyl scanning” approach appeared to be an ideal approach because the number of hydrogen bonds differ significantly in the base triplet (Figures 9a,b) and the stretched wobble binding modes (Figure 9c). In particular, the wobble binding mode has two, possibly three free N–H group whereas only a single N–H group is free in the triplet binding modes. Thus, five N-methylated analogs 14–18 were synthesized (Chart 6) and tested the binding using ITC with both d(CTG)_2 and r(CUG)_2 to establish a structure-activity relationship (SAR).

Ligands 14–18 were prepared using the same synthetic sequence for preparing ligand 11. Thus, a linker is first attached to the Janus wedge followed by the acridine moiety. For example, ligands 15 and 16 were prepared from the corresponding methylated wedges (19 and 20), which in turn, were prepared from

![Chart 6. A series of N-methylated versions (14–18) of ligand 11, and RNA and DNA duplexes containing a r(CUG)_2 and d(CTG)_2 motifs, respectively, for SAR studies.](image-url)
aminodichlorotriazine 21 (Scheme 1). The linker was attached to the wedge by heating 19 or 20 with an excess of 1,4-diaminobutane to afford intermediates 22 and 23, respectively. The intercalator unit was attached to the linker using an activated acridine 24 with catalytic amount of trifluoroacetic acid (TFA) to yield ligands 15 and 16. Ligands 14, 17, and 18 containing a methyl group at nitrogen position of the linker
required the corresponding precursor 25 that is prepared from tetrahydrofuran (THF) in 6 steps (Scheme 2). Similarly, intermediates 34–36 were prepared from the corresponding wedges using precursor 25 followed by Pd-catalyzed hydrogenolysis (Scheme 3). Attempts to purify these intermediates (34–36) resulted in low yields (20–40%). For example, compound 35 was isolated in only 39% yield. They were used directly in the subsequent steps to afford ligands 14, 17, and 18, respectively.

**Table 2.** $K_d$ values (in $\mu$M) of ligands 11, 14–18 with T-T (DNA) and U-U (RNA) mismatches.

<table>
<thead>
<tr>
<th>Target</th>
<th>11</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-T</td>
<td>0.39 ± 0.08</td>
<td>0.8 ± 0.4</td>
<td>0.12 ± 0.04</td>
<td>2.2 ± 0.5</td>
<td>50 ± 30</td>
<td>nb</td>
</tr>
<tr>
<td>U-U</td>
<td>2.1 ± 0.2</td>
<td>nb</td>
<td>nb</td>
<td>14 ± 5</td>
<td>nb</td>
<td>nb</td>
</tr>
</tbody>
</table>

nb: no measurable binding and a lower limit of $K_d > 200 \mu$M is assigned.

**Figure 11.** ITC profiles for complexation of d(CTG)$_2$ with (a) ligand 14 and (b) ligand 15.
The d(CTG)₂ recognition by 14–18 was measured by ITC and all \( K_d \) values are collected in Table 2. Ligand 15 was found to bind >3-fold stronger than ligand 11 (\( K_d = 0.12 \mu M \) for 15 vs 0.39 \( \mu M \) for 11), whereas ligands 14, 16, and 17 each bound d(CTG)₂ progressively more weakly than 11. The fact that the affinities of these methylated ligands binding to d(CTG)₂ compared to that of 11 ruled out the possibility of a triplet formation, irrespective of the groove preference. Figure 11 showed the ITC profiles of ligands 14 and 15 binding to a (CTG)₂-containing duplex. Ligand 18, carrying three methyl groups, showed no measurable binding to d(CTG)₂. Because of the limited aqueous solubility, we were only able to assign a lower limit of \( K_d > 200 \mu M \). The combined ITC and modeling results is consistent with Watson-Crick-type recognition of the stretched wobble pair (Figure 12). This binding mode also explains the decreased binding to the T-T pair by 17 (\( K_d = 50 \mu M \) for 17 vs 0.39 \( \mu M \) for 11) because of the presence of rotamers that interrupt the hydrogen bonding interaction. An unconstrained 10 ns MD simulation further suggested that the weakly bound thymine unit could be
flipped-out from the dsDNA (Figure 13). Base flipping is often observed with DNA modifying and error repairing enzymes and there are precedents for base-flipping induced by small molecules. Yen-Jun Ho in our group attempted to use potassium permanganate (KMnO₄) as a probe for flipped-out thymines and observe by electrophoresis but unsuccessful.

The most interesting finding came from the ITC results of the ligand-RNA binding. In contrast to the results with d(CTG)₂, ligand 15 did not bind under similar conditions to r(CUG)₂ (Figure 14a, Table 2) and, indeed, the single N-methyl group in 14 was enough to eliminate the ligand binding (Kₐ > 200 µM). The only CUG-binding N-methylated

![Figure 13. Snapshot from a 10 ns MD simulation of a complex of d(CTG)₁₂⁻¹⁵ showing one of the thymines was flipped-out from the helix.](image)

![Figure 14. ITC profiles for complexation of the complexation of r(CUG)₂ to (a) ligand 15 and (b) ligand 16.](image)
ligand was 16 and even its $K_d$ was 7-fold weaker in comparison to 11 (Figure 14b, Table 2). One of the possible explanations is that only ligands 11 and 16 allow the formation of triplet with a U-U mismatch (cf. Figures 9a,b). Again, the weaker binding of 16 to U-U mismatch can be explained by the presence of a rotamer that disfavors the triplet formation.

The ability of ligands 11, 14–18 to inhibit MBNL1 binding to r(CUG)$_{12}$ was measured by EMSA and the results paralleled the ITC data. Thus, only ligands 11 and 16 showed inhibition (Figure 15). Although nothing can be said about the groove preference from this experiment, the MD simulations favor major groove triplet formation (vide supra), and it is also known that loop and mismatch structures can enhance the accessibility of the major groove.70

Figure 15. EMSA screening of ligands 11, 14–18 showing only ligands 11 and 16 inhibit MBNL1-CUG interaction. The structure of r(CUG)$_{12}$ used in the assay is shown on the right.

To get a complete picture of the binding selectivity, the affinity of the ligand to other mismatches was also measured (Table 3). In general, these ligands exhibited high selectivity (7- to 1,600-fold) for DNA over RNA mismatches. The ligands were also found to bind weakly to purine-purine mismatches. Interestingly, all the methylated ligands
bind quite strongly \( (K_d = 0.13-2.0 \mu M) \) to \( d(CCG)_2 \) (sequence shown in Table 3), a trinucleotide repeat sequence associated with Fragile XE syndrome\textsuperscript{71,72} and chronic lymphocytic leukemia,\textsuperscript{73} suggesting a potential molecular therapeutic approach to these diseases that warrants additional study. More importantly, all methylated ligands showed reduced affinity toward normally paired duplex DNA (i.e., no mismatch), including herring sperm DNA. We believe that the ability of the recognition unit to stack on the
intercalator is enhanced by methylation which in turn reduced the non-specific intercalative binding to duplex DNA.\textsuperscript{74,75}

The combined experimental and computational approach described herein suggests that this class of ligands bind DNA and RNA by significantly different modes. Thus, it is proposed that ligands \textbf{11} and \textbf{16} recognize the U-U mismatch in RNA through formation of a major groove base triplet whereas ligands \textbf{11}, \textbf{14}–\textbf{17} bind the T-T mismatch in DNA via the stretched wobble pair (Figure 16). These binding modes also explain the decreased binding to U-U by \textbf{16} ($K_d = 14$ vs $2.1 \mu M$ for \textbf{11}) and to T-T by \textbf{17} ($K_d = 50 \mu M$ for \textbf{17} vs $0.39 \mu M$ for \textbf{11}). Thus, the C–NH(Me) bond rotation leads to unfavorable binding for one or more of the rotamers.\textsuperscript{76}

\textbf{2.5 Conclusions}

We have used the methyl scanning method combined with MD simulations to indirectly investigate the possible binding modes by which ligands \textbf{11}, \textbf{14}–\textbf{17} recognize CUG and CTG sites in RNA and DNA, respectively. Beyond informing on the binding mode, the ability to substitute the amino groups of ligand \textbf{11} suggests these as sites for further
modifications that may enhance the selectivity and efficacy of these lead compounds in treating DM1.

More significant was the unexpected discovery of DNA-selective ligands for CTG. Compared to the ligand 11, which binds both CUG and CTG with similar strength, ligand 15 showed a >3-fold increase in affinity to CTG relative to ligand 11 with negligible binding to RNA. It was also found that these ligands showed reduced binding toward DNA and RNA duplexes (i.e., without mismatches) upon methylation, potentially open an avenue for a DNA-targeted molecular therapy of DM1.8,18,77 Competitive binding to the corresponding CUG transcript would be expected to reduce the effectiveness of this approach and could also complicate efforts to assess the DNA-targeted approach. More broadly, this study increases our knowledge of how small molecules can selectively recognize nucleic acids.
Chapter 3  Lead Discovery for DM2

“I and others with DM2 appreciate the research being done to develop a potential treatment in DM2 by you and other researchers at the University of Illinois. Although DM2 is very rare in the States, in Germany it is much more common. Thank you.”

A DM2 patient in Germany

3.1 Introduction

DM2 is a second form of myotonic dystrophy, one that caused by the expansion of a CCTG repeat of the ZNF9 gene on chromosome 3.6 Like DM1, DM2 is an incurable neuromuscular disease. As we can see in Table 1, the severity of DM2 is considered to be milder than that of DM1. Although DM2 patients do not suffer from the severe congenital form of disorder that occurs in DM1, the prevalence of DM2 is predicted to be at least as high as DM1 depending on the geographical location (1 in 8,000).4 Our preliminary results on targeting CTG/CUG repeats suggested the use of the stacked intercalator approach to target CCTG/CCUG repeats in DM2. Targeting CCTG (or CCUG) repeats represents a challenge because their secondary structures are not well-studied.78,79 In addition, structural changes in DNA and RNA upon binding of small

---

2 This chapter includes previously published material. The other contributors were Yuan Fu, Sreenivasa R. Ramisetty, Anne M. Baranger, and Steven C. Zimmerman. Reused with permission from: Wong, C.-H.; Fu, Y.; Ramisetty, S. R.; Baranger, A. M.; Zimmerman, S. C. Nucleic Acids Res. 2011, 39, 8881–8890.
molecules are always possible.\textsuperscript{80,81} Indeed, such structural changes are usually substrate-dependent as is observed in riboswitches.\textsuperscript{81} Thus, conformations (a) and (b) are possible with different ligands. There are two limiting conformations for CCTG repeats that may be considered upon binding (Figure 17). As both CUG and CCUG repeats are able to sequester MBNL proteins with high affinities ($K_d = 0.26–0.07 \mu$M),\textsuperscript{82} this suggests the structural similarities between the two repeats in which conformation (b) with a T-T mismatch might be operating. Thus, we set out to examine a number of $\pi$-stacked ligands with the goal of therapeutic agents discovering of DM2 lead.

3.2 Targeting C-T and C-U Mismatches

In comparison to CUG repeats, targeting CCUG repeats represents a particular challenge because neither an NMR nor X-ray structure of RNA containing CCUG repeats is available. However, the encouraging results with DM1 prompted us to design analogous $\pi$-stacked ligands that can bind to CCTG (or CCUG) repeats in DM2. Thus, we designed the 2,6-diaminopyrimidinone-based intercalator (37) to recognize the C-T mismatch with complementary hydrogen bonding on both edges forming a base triplet (Figure 18). Indeed, both C-U and C-T mismatches have been successfully targeted by 2,6-diaminopyrimidinone-based compounds by Lehn\textsuperscript{45} and McLaughlin,\textsuperscript{83} respectively. It should be

\textbf{Figure 17.} Possible CCTG repeat structures with (a) two C-T mismatches or (b) one T-T mismatch and two C-bulges.

\textbf{Figure 18.} Structure of ligand 37 and the proposed binding mode to a C-T mismatch in CCTG repeats.
noted that Lehn used cytosine and uracil nucleobases instead of nucleotides in their studies. Nevertheless, all reported ligands were based on a 2,6-diaminopyrimidinone moiety. Ligand 37 was prepared to determine whether it can inhibit the MBNL1-RNA interactions (Scheme 4). Removal of a t-butyl group of known compound 38 by TFA gave compound 39 in 72% yield. Coupling of compounds 39 and 40 using EDCI provided ligand 37 in 83% yield. It is well known that DNA predominantly adopts a B-form conformation in solutions, which has wide major groove, it is worth to test the idea whether an aminoisocytosine-containing intercalator will bind to a C-T mismatch (Figure 18). Thus, preliminary binding studies were performed on duplex containing two complimentary strands containing a d(CCTG)₂ unit (Figure 19a). It was found that ligand 37 bound to the sequence in an expected 2:1 ratio with \( K_d(1) = 3 \mu M \) and \( K_d(2) = 8 \mu M \). The third binding (\( K_d(3) = 0.3 \) mM) was attributed to non-specific binding. These results also suggest that \( \pi \)-stacking may still occur even if there is an amide bond in the linker. Unfortunately, the strong binding of ligand 37 for CCTG-containing DNA did not translate into strong binding for an RNA oligonucleotide containing a CCUG repeat (Figure 19b). The reason of this distinct difference on binding to RNA and DNA is still unknown. In a gel-shift assay performed by Dr. Yuan Fu in the Baranger group, ligand 37 also did not show any inhibition in good agreement with the ITC result (Figure 20).
Nevertheless, in the absence of an NMR or X-ray structure, the binding mode remains unknown.

In principle, inhibition may still be observed if a ligand such as 11 or its derivatives can recognize the uracil unit in a C-U mismatch. Previous results indicated that ligand 11 is capable to bind to U-U mismatch in r(CUG)₄ with low micromolar affinity ($K_d = 0.43 \mu M$) and bind to C-C mismatch with moderately reduced affinity ($K_d = 13.5 \mu M$).

**Figure 19.** ITC binding studies of ligand 37 using (a) a DNA sequence containing CCTG repeat and (b) a RNA sequence containing a CCUG repeat.

Nevertheless, in the absence of an NMR or X-ray structure, the binding mode remains unknown.

In principle, inhibition may still be observed if a ligand such as 11 or its derivatives can recognize the uracil unit in a C-U mismatch. Previous results indicated that ligand 11 is capable to bind to U-U mismatch in r(CUG)₄ with low micromolar affinity ($K_d = 0.43 \mu M$) and bind to C-C mismatch with moderately reduced affinity ($K_d = 13.5 \mu M$).

**Figure 20.** Gel-shift assay showing ligand 37 could not interrupt the MBNL1-RNA complex. Lane 1: CCUG RNA only; Lane 2: RNA + MBNL without ligand 37.
µM). Compared to the $K_d$ values measured for A-A and G-G mismatches ($K_d > 300$ µM), ligand 11 showed strong preference toward pyrimidine-pyrimidine mismatches. Therefore, we attempted to use ligand 11 to target CCUG repeat for DM2.

The biophysical studies were first carried out on the same CCTG sequence with ligand 11 (Figure 21). A Job plot analysis at a total concentration of 60 µM indicated a stoichiometry of 2:1 (ligand to DNA). When the concentration is increased to 120 µM, 2.5:1 was obtained suggesting a third binding site with a $K_d$ in the µM range. Such a shift in Job plot was not observed at these concentrations with CTG repeats indicating the observed shift was not a result of non-specific intercalation.

The thermodynamic aspects of binding were studied using isothermal titration calorimetry (ITC). A 3:1 stoichiometry was also observed at the equivalence point of the isotherm (Figure 22). The
stoichiometry determined by ITC is usually more accurate than those obtained from other methods because of the precise computer-controlled injections. The ITC data were therefore fit using a three-sequential binding sites model. The ITC data suggests a relative weak first binding (14 µM) followed by a strong binding (108 nM) and a third weak binding (37 µM) which is consistent with the Job analysis. The positive cooperativity could result from the first ligand inducing formation of two C-bulges and a central T-T mismatch (see Figure 17b), the latter site providing the same tight binding observed at CTG sites, whereas the C-bulge recognition would be comparatively weaker. This C-bulge and T-T mismatch structure, which is suggested to be a potential drug target, was recently observed by NMR spectroscopy for CCTG repeats. A CD titration was also performed (Figure 23). The increase in the CD signal at about 290 nm and the small positive induced CD (ICD) signal indicates intercalative binding of ligand 11 to the d(CCTG)$_2$-containing duplex (Figure 23a).

![Figure 23](image)

**Figure 23.** (a) CD spectra showing the change in CD signal upon ligand binding (Conditions: [duplex] = 25 µM, [MOPS] = 20 mM, [NaCl] = 300 mM, temperature = 25 °C, pH = 7). (b) A plot of CD signal at 288 nm versus the mole ratio of ligand to duplex indicating the saturation of binding sites at about 3 equivalents of ligand 11.
versus mole ratio of ligand indicated that the stoichiometry is about 3:1 (Figure 23b). Again, if non-specific intercalation would have occurred, the CD signal would not be saturated at 3:1 stoichiometry. However, a closer look at the CD titration curve suggests a strong binding event and relatively weaker second and third binding events as the slope is steeper until the first equivalence is reached. Overall, Job plot, CD, and ITC studies suggested a 3:1 binding stoichiometry of ligand 11 to the CCTG duplex but the disagreement in the cooperativity (ITC vs. CD) needs to be reconciled with additional experiments including ESI-MS or fluorescence titration that might be capable of

Figure 24. ITC binding studies of ligand 11 using a RNA sequence containing (a) a r(CCUG)$_2$ and (b) a slipped-CCUG unit.

...versus mole ratio of ligand indicated that the stoichiometry is about 3:1 (Figure 23b). Again, if non-specific intercalation would have occurred, the CD signal would not be saturated at 3:1 stoichiometry. However, a closer look at the CD titration curve suggests a strong binding event and relatively weaker second and third binding events as the slope is steeper until the first equivalence is reached. Overall, Job plot, CD, and ITC studies suggested a 3:1 binding stoichiometry of ligand 11 to the CCTG duplex but the disagreement in the cooperativity (ITC vs. CD) needs to be reconciled with additional experiments including ESI-MS or fluorescence titration that might be capable of

Figure 25. Two possible stem–loop structures for CCUG repeats with two consecutive C–U mismatches (left) or alternating C–C and U–U mismatches in the slipped form (right).
establishing the correct binding sequence. The binding to RNA was examined using a RNA duplex containing a r(CCUG)$_2$ unit as shown in Figure 24a. In contrast to the results with the DNA oligonucleotide, ligand 11 binds CCUG comparatively weakly; a single $K_d = 79 \mu M$ was obtained. This CCUG repeat structure is predicted by m-fold to be comprised of two C-G base pairs and two C-U mismatches, although alternative structures cannot be ruled out.\(^87\) The slipped CCUG structure contains U-U and C-C mismatches separated by C-G base-pairs as shown in Figure 25.\(^82\) Although a less stable form, it was found that MBNL1 bound this slipped structure with ~1.5-fold higher affinity than to the structure containing two C-U mismatches.\(^82\) For this reason, the binding of ligand 11 to the slipped sequence was also investigated by ITC using the 11-mer shown in Figure 24b. A 2:1 complex was observed with low-micromolar $K_d$ values: $K_d(1) = 3.5 \mu M$ and $K_d(2) = 7.0 \mu M$.

**Figure 26.** (a) Gel electrophoretic mobility shift assay of MBNL1 binding to r(CCUG)$_6$ RNA. (b) MBNL1 binding to r(CCUG)$_6$ in the presence of 100 nM tRNA. Control lane 1 (C1): RNA only. Control lane 2 (C2): RNA + MBNL1. (c) Plots of fractions of r(CCUG)$_6$ bound as a function of MBNL1 concentration in the presence (green) and absence (blue) of 100 nM tRNA.
To determine whether ligand 11 can inhibit the MBNL1-CCUG interaction, an electrophoretic mobility shift assay (EMSA) was performed by Dr. Sreeni Ramisetty in the Baranger group using a r(CCUG)$_6$ RNA. The $K_d$ values for the complexes formed between r(CCUG)$_6$ and MBNL1 were determined in the absence and presence of 100 nM of competitor tRNA to be 15 ± 2 nM and 30 ± 4 nM, respectively (Figure 26). These values are in good agreement with literature values.$^{82}$ No significant inhibition of the MBNL1-RNA interaction was observed with ligand 11 up to a concentration of 250 µM (Figure 27). Therefore, an accurate IC$_{50}$ cannot be obtained and a upper limit of 250 µM was assigned.

### 3.3 Discovery of DM2 Lead Compounds

The discovery of compounds 41 and 42 turned out to be serendipitous. During a screening of a series of compounds for inhibition of MBNL1-CCUG interaction, we found that ligands 11, 41, and 42 can inhibit the toxic protein-RNA interaction (Figure 28). The ability of simple triaminopyrimidine ligands 45 and 46 and triaminotriazine ligands 47 and 48 to inhibit MBNL1 binding was examined (Figure 28a). Consistent with our
previous results for DM1, wedges without tethered intercalators did not destabilize the MBNL1-CCUG complexes (Figure 28b). Again, simple intercalators such as 40 do show inhibition. Originally, ligands 41 and 42 were synthesized as structural derivatives of 11 (Scheme 5). Ligand 41 is identical to 11, but with a triaminopyrimidine moiety that
forms a C-linked Janus-wedge. It was an appealing recognition unit to explore because 5-(4-aminobutyl)-2,4,6-triaminopyrimidine\textsuperscript{88} was a known compound, allowing \textit{41} to be easily synthesized by the same route as used for ligand \textit{11}. The triaminopyrimidine ring also has an extra N–H group for post-synthesis modification. Unfortunately and somewhat surprisingly, ligand \textit{41} exhibited very limited water solubility compared to the structurally similar ligand \textit{11}. The limited solubility of \textit{41} prevented it from being studied by ITC with less than 10% DMSO. Ligand \textit{42} was prepared with a more water-soluble intercalator, \textit{N}-(2-(dimethylamino)ethyl)acridine-4-carboxamide (DACA)\textsuperscript{89} in place of the 9-amino-6-chloro-2-methoxyacridine unit and this allowed the triaminopyrimidine moiety to be examined (Scheme 5).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig29.png}
\caption{(a) Gel electrophoretic mobility shift assay of ligand \textit{41} with r(CCUG)\textsubscript{6} RNA in the presence of 100 nM tRNA. Control lane 1 (C1): RNA only. Control lane 2 (C2): RNA+MBNL1. Conditions: [MBNL1] = 100 nM, [RNA] = 0.1 nM, [NaCl] = 175 mM, [MgCl\textsubscript{2}] = 5 mM, [BME] = 1.25 mM. (b) Gel electrophoretic mobility shift assay of ligand \textit{42} with (CCUG)\textsubscript{6} RNA in the presence of 100 nM tRNA. Control lane 1 (C1): RNA only. Control lane 2 (C2): RNA+MBNL1. (c) Plots illustrating inhibition of MBNL1-(CCUG)\textsubscript{6} complex with ligand \textit{41} (green) and \textit{42} (blue).}
\end{figure}
3.4 MBNL1-CCUG Inhibition by DM2 Lead Compounds

Using EMSA, both ligands 41 and 42 showed good inhibition of the MBNL1-CCUG interaction in concentration-dependent studies (Figure 29; see Figure 27 for 11). Despite the limited solubility of ligand 41, its ability to inhibit MBNL1 binding to RNA could be determined because of the lower concentrations required for electrophoretic mobility shift assays. Both ligands 41 and 42 were found to significantly destabilize the MBNL1-CCUG complex. The IC\textsubscript{50} of ligands 41 and 42 were determined by incubating varying concentrations of the ligands with a constant concentration of MBNL1-CCUG complex and evaluating the results using gel shift assays. The apparent inhibition constant (K\textsubscript{i}) of these ligands was determined using the equation K\textsubscript{i} = IC\textsubscript{50} × (K\textsubscript{d}/[MBNL1]\textsubscript{total}), in which K\textsubscript{d} is the dissociation constant of the MBNL1-RNA complex and the concentration of MBNL1 is in large excess of the K\textsubscript{d} value. The comparison of K\textsubscript{i} values, rather than the IC\textsubscript{50} values, under different conditions is important in order to compare the effects of the conditions on the behavior of the ligands independently of

<table>
<thead>
<tr>
<th>RNA</th>
<th>tRNA (nM)</th>
<th>K\textsubscript{d} (nM)</th>
<th>[MBNL1] \textsubscript{total} (nM)</th>
<th>41</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>-----</td>
<td>-----------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>r(CCUG)\textsubscript{6}</td>
<td>–</td>
<td>15 ± 2</td>
<td>350</td>
<td>59 ± 5</td>
<td>2.5 ± 0.2</td>
</tr>
<tr>
<td>100</td>
<td>30 ± 4</td>
<td>350</td>
<td>40 ± 11</td>
<td>3.4 ± 0.9</td>
<td>35 ± 3</td>
</tr>
<tr>
<td>r(CUG)\textsubscript{12}</td>
<td>–</td>
<td>150 ± 20</td>
<td>1000</td>
<td>37 ± 11</td>
<td>5.5 ± 1.4</td>
</tr>
<tr>
<td>cTNT\textsuperscript{c}</td>
<td>–</td>
<td>1.0 ± 0.2</td>
<td>15</td>
<td>101 ± 26</td>
<td>6.7 ± 1.6</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Conditions: [MBNL1] = 100 nM, [RNA] = 0.1 nM, [NaCl] = 175 mM, [MgCl\textsubscript{2}] = 5 mM, [BME] = 1.25 mM. Errors are reported as the standard deviation (SD) of 3 or more trials. \textsuperscript{b} RNA concentrations were 0.1 nM. \textsuperscript{c} 18-nt fragment of the cTNT pre-mRNA.
the effects of the varying conditions on the $K_d$ of the MBNL1-CCUG complex. Importantly, ligands 41 and 42 do not bind to MBNL1 as measured by ITC.

To evaluate the specificity of ligands 41 and 42 for CCUG RNA, the inhibition experiments were performed in the presence of competitor RNA sequences. For ligand 41, the $K_i$ value is increased only slightly to 3.4 ± 0.9 µM in the presence of 100 nM tRNA (Figure 29a,c). Evaluation of ligand 42 provided IC$_{50}$ values of 52 ± 8 µM and 35 ± 3 µM in the absence and presence of 100 nM tRNA, respectively, with a $K_i$ of 2.2 ± 0.3 µM and 2.8 ± 0.3 µM, respectively (Figure 29b,c). The selectivity of ligand 41 and 42 for the target sequences r(CCUG)$_6$ relative to r(CUG)$_{12}$ (i.e., DM2 vs. DM1) and their relative abilities to inhibit MBNL1 binding to the same sequences were also evaluated. Ligand 41 inhibited binding to r(CCUG)$_6$ and r(CUG)$_{12}$ with similar $K_i$ values of 2.4 µM and 4.8 µM, respectively, whereas ligand 42 exhibited a ~7-fold higher inhibition of the MBNL1-CCUG interaction compared to the MBNL1-CUG interaction. The ability of ligand 42 to destabilize complexes formed between MBNL1 and cTNT (RNA I, Figure 30).

**Figure 30.** Nucleic acids studied in ITC experiments (tRNA and herring sperm DNA are not shown). The structures shown represent the most stable structures predicted by m-fold. Note that RNA I is most likely single stranded under the conditions used to perform the binding or complex inhibition experiments.
30) was also evaluated. cTNT is the 18-nt fragment of the human cardiac troponin T (hcTNT) pre-mRNA, which is a natural target of MBNL1. Ligand 42 only weakly inhibited the MBNL1-cTNT interaction. Table 4 summarizes the inhibition studies of ligands 41 and 42.

### 3.5 Biophysical Studies of the DM2 Lead Compounds

The ability of ligands 41 and 42 to bind CCUG and CCTG repeats was examined by ITC. However, the studies with 41 were hampered by its poor water-solubility. To achieve the high ligand concentration needed for ITC studies, ligand 41 required more than 10% (v/v) DMSO. With as little as 1% (v/v) DMSO, a 500 µM solution of ligand 42 could be prepared in a 300 mM NaCl solution (MOPS buffer; pH 7), presumably

<table>
<thead>
<tr>
<th>Nucleic acids</th>
<th>Specific target</th>
<th>$T_m$ (°C)$^b$</th>
<th>$K_d(1)$</th>
<th>$K_d(2)$</th>
<th>$K_d(3)$</th>
<th>Selectivity$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>r(CCUG)$_6$</td>
<td>C-U pairs$^d$</td>
<td>61</td>
<td>5.2 ± 0.6</td>
<td>478 ± 85</td>
<td>–</td>
<td>1.4</td>
</tr>
<tr>
<td>A</td>
<td>C-U pairs</td>
<td>53</td>
<td>3.6 ± 0.8</td>
<td>258 ± 37</td>
<td>–</td>
<td>1.0</td>
</tr>
<tr>
<td>B</td>
<td>C-C and U-U</td>
<td>54</td>
<td>0.13 ± 0.04</td>
<td>10 ± 5</td>
<td>202 ± 14</td>
<td>0.04</td>
</tr>
<tr>
<td>C</td>
<td>Single C-U pair</td>
<td>62</td>
<td>85 ± 30</td>
<td>231 ± 55</td>
<td>–</td>
<td>24</td>
</tr>
<tr>
<td>D</td>
<td>U-U</td>
<td>63</td>
<td>25 ± 13</td>
<td>271 ± 83</td>
<td>–</td>
<td>6.9</td>
</tr>
<tr>
<td>E</td>
<td>C-C</td>
<td>62</td>
<td>0.77 ± 0.28</td>
<td>162 ± 8</td>
<td>–</td>
<td>0.2</td>
</tr>
<tr>
<td>F</td>
<td>A-A</td>
<td>62</td>
<td>118 ± 8</td>
<td>2270 ± 360</td>
<td>–</td>
<td>33</td>
</tr>
<tr>
<td>G</td>
<td>G-G</td>
<td>67</td>
<td>103 ± 26</td>
<td>7690 ± 280</td>
<td>–</td>
<td>29</td>
</tr>
<tr>
<td>H</td>
<td>T-T</td>
<td>51</td>
<td>1.4 ± 1.5</td>
<td>102 ± 25</td>
<td>–</td>
<td>0.4</td>
</tr>
<tr>
<td>I</td>
<td>–</td>
<td>–</td>
<td>334 ± 11</td>
<td>–</td>
<td>–</td>
<td>93</td>
</tr>
<tr>
<td>tRNA</td>
<td>–</td>
<td>–</td>
<td>2400 ± 2100</td>
<td>–</td>
<td>–</td>
<td>670</td>
</tr>
<tr>
<td>hsDNA</td>
<td>–</td>
<td>–</td>
<td>36 ± 7$^e$</td>
<td>–</td>
<td>–</td>
<td>10</td>
</tr>
</tbody>
</table>

$^a$ $K_d$ values are reported in µM. Errors are reported as the standard deviation (SD) of 3 or more trials.

$^b$ Concentrations of duplexes are 10 µM; additional 1% (v/v) DMSO has negligible effect on the stability of the duplexes ($\Delta T_m < 2$ °C). $^c$ Selectivity is the ratio of $K_d(1)$ of the particular nucleic acid to that of RNA A.

$^d$ mfold predicts two 2x2 nucleotide internal loops. $^e$ Ligand 42 bound to about every 10 base pairs.
because it contains a water-solubilizing substituent (Scheme 5). Thus, only the affinity and selectivity of ligand 42 for binding to CCUG RNA was evaluated by measuring the affinity of ligand 42 for several oligonucleotides (Figure 30). The stabilities of these RNAs were evaluated by melting curves obtained by UV spectroscopy and are reported in Table 5. The duplexes containing C-U (RNA C), U-U (RNA D), C-C (RNA E), A-A (RNA F), and G-G (RNA G) were designed to evaluate the ability of ligand 42 to selectively recognize the mismatches that are most likely found in the poly(CCUG) RNA compared to related mismatched sequences. The binding of ligand 42 to the DNA analog of the DM1 sequence containing a T-T mismatch (DNA H) and the 18-nt fragment of the human cardiac troponin T (hcTNT) pre-mRNA, a natural target of MBNL1 (RNA I) was also evaluated. To investigate the non-specific binding of ligand 42

![Figure 31](image_url)

**Figure 31.** ITC profiles of ligand 42 binding to (a) r(CCUG)₆ and (b) an RNA duplex containing a single r(CCUG)₂ unit.
to nucleic acids in general, its affinity toward tRNA and herring sperm DNA (hsDNA) was measured (Table 5).

ITC was used to measure the affinity of ligand 42 for the duplexes (Table 5). Ligand 42 bound to r(CCUG)₆ and RNA A with similar affinity and stoichiometry. Specifically, both sequences fit a binding model with a single, reasonably tight binding site (3–5 µM) and a second, significantly weaker binding site (250–480 µM). The tighter binding is presumably associated with one of the two CCUG sites in (CCUG)₆ and the single CCUG site in RNA A (Figure 31a). Potential weaker sites in r(CCUG)₆ include the remaining CCUG site, (CCUG)₂ loop region, and GC stem regions. In fact, titration with RNA A, lacking the second CCUG site and loop, give an ITC curve with better curvature (Figure 31b). These results suggest that the residual heat at the end of the titration for r(CCUG)₆ likely arises from a general, nonspecific binding of the weaker sites. Indeed, the affinity is similar to that observed for ligand 42 binding to the single stranded human cardiac troponin T (hcTNT) pre-mRNA (RNA I).

Overall the data are consistent with ligand 42 binding to only one of the two C-U mismatches and a single CCUG site when two sites are neighboring. The origin of this latter effect is unclear.
Similar to the preference exhibited by MBNL1, ligand 42 binds with ca. 40-fold higher affinity to the slipped-CCUG structure (RNA B; Figure 32) than to the duplex containing adjacent C-U mismatches (RNA A). These results are in good agreement with the $K_d$ values for the duplexes containing the single U-U (RNA D) and C-C mismatches (RNA E), suggesting that ligand 42 binds strongly to the C-C mismatch in the slipped structure and to the U-U with a reduced affinity.

Surprisingly, ligand 42 does not bind strongly ($K_d = 85 \pm 30 \, \mu\text{M}$) to RNA C, which contains a single C-U step with two adjacent CG steps. Ligand 42 also binds weakly ($K_d > 100 \, \mu\text{M}$) to the purine-purine mismatches in duplexes F and G. The binding of ligand 42 to DNA containing a CTG sequence (DNA H), which is one of the targets of DM1, occurs with low micromolar affinity ($K_d = 1.4 \pm 1.5 \, \mu\text{M}$). The difference in binding affinity of ligand 42 toward analogous RNA and DNA duplexes (RNA D, U-U; $K_d = 25 \, \mu\text{M}$) and (DNA H, T-T; $K_d = 1.4 \, \mu\text{M}$) may result from structural differences between B-form DNA and A-form RNA.

Importantly, ligand 42 not only binds CCUG, it inhibits the formation of MBNL1-CCUG complex with low micromolar $K_i$ values. Together with the ITC results, these data suggest that ligand 42 is a good lead ligand for targeting either the CCUG or CCUG-slipped repeat structures implicated as the causative agent of DM2. Finding agents that complex CCUG and inhibit MBNL1 binding is a challenge, but finding agents that are highly selective is perhaps even more difficult and likely more important given the difficulty in developing highly selective agents targeted to nucleic acids. The selectivity profile of ligand 42 has been examined in several different contexts and these data can
be found in Tables 4 and 5. For example, it is notable that ligand 42 binds r(CCUG)$_6$ >600-fold tighter than it does tRNA. Thus, even in the presence of tRNA, ligand 42 is capable of disrupting the MBNL1-CCUG interaction with minimally diminished $K_i$ values. It is also significant that ligand 42 does not inhibit the interaction between MBNL1 and its natural target, the 18-nt fragment of human cardiac troponin T (hcTNT) pre-mRNA (IC$_{50}$ > 250 µM).

The RNA hydrolytic properties of ligands 11, 41, and 42 were also investigated. The stability of the RNA was found to be identical within experimental error in the presence of 1–10% DMSO. In addition, the $T_m$ of r(CCUG)$_6$ was concentration-independent suggesting the formation of a hairpin not a duplex. The hydrolytic stability of r(CCUG)$_6$ RNA in the presence of ligands 41 and 42 was also evaluated by Dr. Yuan Fu from the Baranger Group. The r(CCUG)$_6$ RNA was incubated with ligands 41 and 42 for at least 4 hours at room temperature and the RNA was evaluated by denaturing polyacrylamide gel electrophoresis (Figure 33). No cleavage of the RNA was observed under these conditions. Therefore, the RNA is stable to hydrolytic cleavage in all of the binding and inhibition experiments reported herein.

![Figure 33](image)

**Figure 33.** The EMSA studies showed that ligands 11, 41, and 42 were not cleaving the target r(CCUG)$_6$ RNA.
3.6 Conclusions

Although we did not establish a systematic structure-activity relationship for this DM2 study, but the data from the gel electrophoretic mobility shift assays using simple wedge recognition compounds 11, 41, and 42 indicate that the acridine unit is critically important. These results agree with the previously proposed $\pi$-stacked intercalator model wherein the acridine is responsible for affinity whereas the selectivity is dictated by the nature of the wedge unit. In this regard, it is intriguing that the small structural difference between triaminotriazine and triaminopyrimidine units leads to the preferential binding to CUG for the former and to CCUG for the latter. It is likely that the basicity difference between triaminotriazine ($pK_a \sim 5$)\textsuperscript{90} and triaminopyrimidine ($pK_a \sim 6.7$)\textsuperscript{91} plays an important role and this is an aspect that merits additional study (Figure 34). At pH 7, a substantial amount of protonated triaminopyrimidine may present. As we have seen in the previous chapter, this may allow favorable interaction with the C-U mismatch from the major groove without serious distortion in the RNA backbone (Figure 35).

![Figure 34. The difference in pK$_a$ values of triaminotriazine (in ligand 11) and triaminopyrimidine (in ligands 41 and 42).](image-url)
Finally, the broadest finding from the current study is that ligands 41 and 42 exhibit among the highest documented selectivity for CCUG and, thus, represent an excellent starting point for the development more selective and potent inhibitors of MBNL1 binding. Because of the presence of multiple repeating units in the toxic RNAs that cause DM1 and DM2, an obvious approach involves oligomerization of these or analogous ligands.41,42 This is an approach that has already proven effective. For example, Disney reported that the oligomerization of a DNA dye (Hoechst 33258) and kanamycin A, ligands that exhibit only moderate selectivity for the CUG and CCUG sequence, respectively, result in oligomers of better affinity and selectivity.41,42,92 The use of intercalators presents an inherent limitation on selectivity because an unstacked intercalator can nonselectively complex a wide range of duplex DNAs and RNAs. Developing new ligands where the recognition wedge remains stacked or in close proximity to the intercalator, or where the intercalator is replaced by a different unit that provides affinity, may obviate this potential source of poor selectivity. Regarding to the latter, we have developed a new generation of ligands without intercalating moiety which is described in the following chapter.
Chapter 4  Bisamidinium-Based Inhibitors

4.1 Introduction

Inhibitors 11, 42, and their analogs offered promising activities as were described in the previous chapters. After the extensive in vitro studies of these ligands, we began to ask if these ligands are also bioactive in DM1 cell models or even in animal models. As pointed out in Chapter 1, a reversal of DM phenotype may be achieved by disrupting the toxic MBNL1-RNA foci, which in turn, rescues the mis-regulated splicing. The goal is to develop a small molecule that is of low-toxicity and can enter the nucleus readily to disrupt the toxic MBNL1-RNA aggregates. Sulforhodamine B (SRB) toxicity studies indicated a relatively high level of cytotoxicity of ligand 11 (IC₅₀ = 20.4 ± 0.5 µM in a 24-hour assay). The IC₅₀ value is further reduced to 4.2 ± 0.2 µM in a 3-day SRB assay. Aggregation of 11 was also observed in cell culture media as a result of its relatively low solubility. In cell culture assays performed by Jessie Peh using confocal microscopy, it appeared that ligand 11 concentrated in the cytoplasm of HeLa cells (Figure 36). Intercalators that are inherently charged, such as phenanthridinium and TO-PRO dyes, or have solubilizing groups are known to enter the nucleus. So it is possible that

---

3 This chapter contains materials that are being patented. The other contributors were Jessie Peh, Lien T. T. Nguyen, Long M. Luu, Jeannette S. Sanchez, Stacie L. Richardson, Tiziano Tuccinardi, Ho Tsoi, Wood-Yee Chan, Edwin H. Y. Chan, Anne M. Baranger, Paul J. Hergenrother, and Steven C. Zimmerman.
ligand 11 can be further developed for example ligands 12 and 13 are much more aqueous-soluble and deserve further investigations. Very recently, Amin Jahromi has developed derivatives of 12 and 13 that disrupt MBNL1-CUG foci and partially restore misregulated splicing. Incidentally we have set out to search for new scaffolds that provide satisfactory not only good aqueous solubility but also interact selectively with the A-form RNA duplex.

4.2 Rational Design of the Bisamidinium Inhibitors

In 2011, Butcher and co-workers reported an NMR solution structure of an RNA frameshift site stimulator, ligand 49, which was labeled DB213. Ligand 49 was found to reside in the stem-loop region of the HIV-1 frameshift site (FS) RNA (Figure 37). Similar to CUG repeats, the stem-loop region of HIV-1 FS RNA is essentially an A-form helix. Figure 38a shows the superposition of the [r(CUG)$_6$]$_2$ structure (PDB: 3GM7) with the
stem-loop region of the HIV-1 FS RNA (PDB: 2L94). It was observed that the two ammonium ions in ligand 49 spanned a distance of about 7 base pairs; i.e., from U$_{17}$-A$_{45}$ to U$_{23}$-A$_{40}$, Figure 38b, which is the distance between every other U-U mismatch in a CUG repeat sequence (Figure 38c). Although ligand 49 bound only weakly to the HIV-1 FS RNA ($K_d \sim 360 \mu$M by NMR), we believed that the bis-amidine moiety would serve as a ds-RNA-binding scaffold and also provide satisfactory aqueous solubility because the two amidinium groups would be positively-charged at physiological pH. Thus, we designed ligand 50 by replacing the two dimethylamino groups of ligand 49 with two triaminotriazine-based recognition units (Figure 38c). In comparison to ligand 11, the additional triaminotriazine

![Figure 37](image)

**Figure 37.** NMR solution structure of DB213 (49) binding to the HIV-1 FS RNA in the major groove. Copyright (2011) American Chemical Society.

![Figure 38](image)

**Figure 38.** (a) Superposition of the structures of the [r(CUG)$_3$]$_2$ duplex (blue; PDB: 3GM7) and the HIV-1 FS RNA (gray; PDB: 2L94). (b) The binding site of the HIV-1 FS RNA for ligand 49. (c) The proposed binding site of the CUG repeats for the bis-amidinium ligands (50–52).
unit was designed to provide enhanced affinity and selectivity by the simultaneous recognition of the two U-U mismatches. Molecular modeling provided insight into the optimum linker length for ligand binding (Figure 39). It is clear that ligand 50 containing two tri-methylene (C3) linkers is too short to allow the triaminotriazine units to form base triplets with the U-U mismatches (Figure 39a). Thus, the triaminotriazine units are almost perpendicular to the U-U pairs. On the other hand, ligand 51, which has two C4 linkers, can interact with the U-U pair to form a base triplet. In this binding mode, ligand 51 spans a distance of three r(CUG)$_2$ repeats (Figure 38c). One triaminotriazine binds a U-U mismatch as a base triplet while the bisamidine linker covers the adjacent CUG site and the other triaminotriazine binds the third U-U mismatch (Figures 39c,d). The C5 derivative (52) also showed similar results in both molecular modeling and MD simulations which were carried out by Prof. Tiziano Tuccinardi at University of Pisa.

Figure 39. Energy minimized structure (Amber) showing (a) ligand 50 and (b) ligand 51 binding to a r(CUG)$_6$ duplex in the major groove and the recognition of two of the U-U pairs of ligand 51 through hydrogen bonding (c and d). Triaminotriazine-interacting U-U pairs in green and the internal U-U pair in gray.
Thus, these results suggested that the one or two extra methylene units in ligands 51 and 52, respectively, are necessary to allow effective recognition of the U-U pairs.

4.3 Synthesis of Bisamidinium Inhibitors

The synthesis of the bisamidinium-based inhibitors was thought to be straightforward based on the reported synthesis of ligand 49 (Scheme 6). The synthesis requires a bisimino-ester intermediate (53), which was prepared from the treatment of 1,4-dicyanobenzene with dry HCl gas. This intermediate was reacted with the corresponding amine (54) to afford ligand 49 as a free base. The ligand was subsequently converted to a tetra-HCl salt upon treatment of hydrogen chloride gas in ethanol. The resulting solid was simply filtered off to afford ligand 49 (as tetra-HCl salt) of high purity based on \(^1\)H NMR spectroscopy. Unfortunately, the first few attempts to synthesize the bisamidine ligands using the same strategy turned out to be problematic. The required amine (55), the same intermediates for preparing inhibitor 11, was prepared by refluxing diaminochlorotriazine in an excess of 1,4-diaminobutane (Scheme 7).
7). A detailed analysis of the this synthetic step revealed the formation of a small amount (5–10%) of dimeric compound 56. During the synthesis and purification of ligand 11, this dimeric compound can be effectively removed by column chromatography. Compound 56 was prepared independently and was found to be highly insoluble in water. It is anticipated that the differences in solubility and polarity of ligands 51 and 56 will allow simple purification of ligand 51, for example by filtration. Unfortunately, after experimenting with different purification techniques, such as HPLC, sephadex, ion exchange, and recrystallization, we were unable to obtain a satisfactory $^1$H NMR spectrum of ligand 51. From both the ESI-MS and NMR spectrum, we identified compound 56 as one of the impurities in the sample. We set out to dissolve the sample in minimum amount of water and filter the insoluble compound 56. Surprisingly, a sample containing up to 10% of compound 56 is still fully water soluble and this is attributed to the formation of supramolecular interactions between compounds 51 and 56 (Figure 40). Not only could this contribute to the difficulty in purification, some identified mono-functionalized intermediates (such as 57) can also form this kind of donor-acceptor-donor-acceptor (DA·DA) hydrogen bonding with ligand 51. To solve this purification problem, a pure form of compound 55 is required (i.e., without carrying 56.

*Figure 40.* (a) Proposed supramolecular interactions between compounds 51 and 56. (b) Structure of a mono-functionalized intermediate (57).
into the subsequent step). Secondly, a technique that allows the separation of any mono-functionalized intermediates is needed. Removing 56 from 55 turned out to be relatively straightforward by careful column chromatography in silica gel using 1–5% of ammonium hydroxide in a methanol/dichloromethane eluent. As expected, a relatively pure sample of ligand 51 can be obtained with pure amine 55. Careful examination of the HPLC trace reveals that there are still impurities appeared as a shoulder peak (Figure 41a). Past experience in purifying supramolecular polymers suggests the use of highly polar solvents. In particular, one must prevent the hydrogen-bond mediated self-assembly by using an acidic eluent. Using acidic eluent in silica gel chromatography for purifying highly polar compounds is not uncommon. After experimenting with many combinations of eluent mixture and acids (e.g., HCl and acetic acid), we found the best eluent to be a mixture of MeOH, CH₂Cl₂, and HCl (in the form of 4 M HCl in dioxane; see experimental section for details). With this procedure, a pure sample of

![HPLC traces](image-url)

**Figure 41.** HPLC traces of (a) crude ligand 51 and (b) purified 51. Conditions: 95% water (0.1% formic acid) to 95% acetonitrile (0.1% formic acid) over 10 min using a reverse-phase HPLC column.
ligand 51 can be obtained (Figure 41b). It is noteworthy that there is a change in retention time for ligand 51 which is typical for supramolecular aggregates. The pure sample also gave a narrower peak width at half height ($w_h \approx 0.3$ min for crude vs 0.04 min for pure 51). The reaction conditions were also optimized and is shown in Scheme

Figure 42. $^1$H NMR spectrum (500 MHz) of ligand 51 in DMSO-$d_6$. 
8. The reaction yield is typical between 70–80%. The reaction has also been scaled up from milligram to gram scale. Figure 42 shows the $^1$H NMR spectrum of ligand 51. Using the same experimental conditions, ligands 50 and 52, containing C3 and C5 linkers, respectively, were also prepared.

### 4.4 Foci Disruption and Splicing Correction by Bisamidinium Inhibitors

To test whether these new ligands can disrupt MBNL1-CUG foci in cell culture, Jessie Peh and Lien Nguyen transfected HeLa cells with a DMPK plasmid containing a

![Figure 43](image)

**Figure 43.** Confocal images of untreated DM1 HeLa cells (a) and a time-course assay with treatment of ligand 49 (100 µM) from 24–72 h (c–d). No reduction of foci number was observed up to 72 h. Scale bar = 10 µm. Nuclei was stained by Hoechst 33342; MBNL1 was immunofluorescent (IF) stained by mouse anti-MBNL1 antibody followed by goat anti-mouse AlexaFluor 488 antibody; r(CUG)$_{960}$ was visualized with Cy3-(CAG)$_{10}$. Data obtained by Jessie Peh.
d(CTG)\textsubscript{960} repeat which produced expanded CUG transcripts that in turn colocalized with MBNL1 (Figure 43a).\textsuperscript{31} Treatment of the control ligand \textit{49} at 100 $\mu$M does not disrupt foci appreciably even up to a 72 h incubation (Figures 43b–d). On the other hand, significant reduction in number of foci with treatment of ligand \textit{51} (100 $\mu$M) in the first 24 h (Figure 44a). Indeed, most foci were dissolved from 48–72 h (Figures 44b,c). These results highlighted the importance of the recognition units in ligand \textit{51}. Ligand \textit{52}, containing penta-methylene linkers, also showed disruption but not as effectively as \textit{51}. In a dose-dependent study, ligand \textit{51} was shown to be effective in reducing the number of foci with as little as 5 $\mu$M (Figure 45). In agreement with the time-course experiment, ligand \textit{51} is more effective then control \textit{49}.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure44.png}
\caption{Confocal images from a time-course assay with treatment of ligand \textit{51} (100 $\mu$M) from 24–72 h (a–c). Significant reduction of foci number was observed in the first 24 h (a) when compared to untreated cells in Figure 43a. Disruption of most foci from 48–72 h (c,d). Scale bar = 10 $\mu$m. Nuclei was stained by Hoechst 33342; MBNL1 was immunofluorescent (IF) stained by mouse anti-MBNL1 antibody followed by goat anti-mouse AlexaFluor 488 antibody; \textit{r(CUG)}\textsubscript{960} was visualized with Cy3-(CAG)\textsubscript{10}. Data obtained by Jessie Peh.}
\end{figure}
The next important step was to test the potential of ligand 51 to restore misregulated splicing in DM1 cells. The splicing assays of cardiac troponin T (cTNT) and insulin receptor (IR) pre-mRNA were studied by Jessie Peh and Lien Nguyen, respectively. The required DM1 cell culture models were generated by transfecting HeLa cells with CTG<sub>960</sub> and cTNT (or IR) plasmids. Untransfected cells (i.e., CUG<sub>0</sub>) were used as controls for normal cells. The cTNT, (CUG)<sub>0</sub> and (CUG)<sub>960</sub> plasmids were obtained from Prof. Thomas Cooper (Baylor College of Medicine) and the IR plasmid from Prof. Nicholas Webster (University of California, San Diego). In a splicing assay, one determines the ratio of the two isoforms of the mRNA of interest. This ratio is different between a healthy adult and a DM1 patient and the amount of free MBNL1 is a key determinant of this ratio. DM patients generally produce a larger number of fetal isoform mRNAs. For example in the alternative cTNT splicing, the production of the

![Figure 45. Dose-dependent foci disruption (48 h treatment) by ligand 51 and comparison of % cell with foci for ligands 51 and 49 at 100 µM. P values reaching statistical significance are marked (2 tailed t-test with unequal variance). Error bars represent standard errors of the mean. Data obtained by Jessie Peh.](image)
embryonic mRNA (exon 5 included) is enhanced in DM patients (Figure 46a). It is suggested that the correct splicing ratio can be restored if the MBNL1 protein is released from the foci. Thus we treated the transfected HeLa cells with ligand 51 at 100 µM and the ratio of the two isoforms were observed over 3 days (Figure 46b). Although there is no observable difference in the ratio from 24–48 h, a significant amount of adult isoform (i.e., without exon 5) was observed after treatment of ligand 51 for 72 h. On the other hand, control 49 did not show any restoration during the same period of time. In addition, the amount of exon 5-excluded mRNA increases with increasing ligand 51 concentration in a dose-dependent study (Figure 46c).

Similar results were obtained with HeLa cells transfected with IR plasmid (Figure 47). In DM, the production of IR-B isoform (exon 11 included) is limited (Figure 47a). Results from both time- and dose-dependent assays suggested that ligand 51 enhanced the production of IR-B mRNA (Figures 47b,c). Again, control 49 showed no

---

**Figure 46.** (a) Alternative splicing of cardiac troponin (cTNT) pre-mRNA. (b) Time-dependent splicing correction for cTNT pre-mRNA by treatment of ligand 51 (100 µM) for 24–72 h and the comparison with ligand 49 at 100 µM (c) Dose-dependent treatment of ligand 51 for 72 h at the indicated concentrations. Data obtained by Jessie Peh.
effect on splicing rescue (Figure 47b). In conclusions, these splicing assays suggest that ligand 51 can partially restore the mis-regulated splicing processes in DM.

These results above also pointed out the low cytotoxicity profile of ligand 51. All HeLa cells remained healthy with dosing at 100 µM for more than 3 days. In fact, the cytotoxicity of ligand 51 was also found to be quite low (IC$_{50}$ > 100 µM) with two other cell lines, DM1 fibroblasts and 3T3 mouse fibroblasts, using SRB or MTT assays. With the assistance from Rachel Botham in the Hergenrother group, we also studied the toxicity by measuring the maximum tolerated dose (MTD) in C57BL6 mice. The compound was dissolved in water and administrated by intraperitoneal (IP) injection. The MTD was found to be between 50 and 100 mg/kg. At 50 mg/kg, the mice showed mild neurotoxicity and fatigue in hind limb; improvement in motion was seen after 1 h. All mice appeared to be fully recovered by the next day. At 100 mg/kg, mice were sacrificed because of the severe seizures.
4.5 Biophysical Studies of the Bisamidinium Inhibitors

Isothermal titration calorimetry (ITC) was used to study the affinity and selectivity of these ligands binding to various targets. It was found that ligand **51** binds to r(CUG)\(_{12}\) with low micromolar affinity (\(K_d \sim 8 \, \mu\text{M}\)) (Figure 48a). On the other hand, the binding affinity of **51** to bulk tRNA and GST-tagged MBNL1 was too low to be measured. Thus, an upper limited of \(K_d > 200 \, \mu\text{M}\) is assigned (Figures 48b,c). These results indicated that **51** inhibits the MBNL1-RNA complex by targeting the CUG not by binding the protein. The importance of the triaminotriazine units in **51** is also highlighted when comparing the ITC profiles for control **49** and pentamidine (**58**) (Figure 49). Again, only an upper limit of \(K_d > 200 \, \mu\text{M}\) can be assigned for both **49** and **58**. This is also consistent with the \(K_d\) value (~ 360 \(\mu\text{M}\)) of control **49** binding to A-form RNA duplexes reported by Butcher.\(^97\) An electrophoretic mobility shift assay (EMSA) was performed to

\[ K_d = 8 \pm 2 \, \mu\text{M} \]

![ITC profiles](image)

**Figure 48.** ITC profiles of ligand **51** binding to (a) r(CUG)\(_{12}\), (b) bulk tRNA, and (c) GST-tagged MBNL1.
measure the strength by which 51 inhibits the MBNL1-CUG interaction (Figure 50). The EMSA was performed under stringent conditions with $K_d \sim 7$ nM for the MBNL1-CUG complex. Thus, the measured IC$50$ was $\sim 115 \mu$M, corresponding to an inhibition constant

\[ K_i = 2.6 \pm 0.1 \mu \text{M} \]

Figure 49. ITC profiles of (a) control 49 and (b) pentamidine (58) binding to r(CUG)$_{12}$.

Figure 50. A electrophoretic gel showing the inhibition of MBNL1-(CUG)$_{12}$ by ligand 51. Conditions: [MBNL1] = 100 nM, [RNA] = 0.2 nM, [NaCl] = 175 mM, [MgCl$_2$] = 5 mM, [BME] = 1.25 mM, 0.05% Triton-X. IC$50$ and $K_i$ were presented as mean ± SD of 4 trials. Data obtained by Lien Nguyen and Stacie Richardson.
\( K_i = 2.6 \, \mu M \) for ligand 51, which is comparable to its \( K_d = 8 \, \mu M \) for r(CUG)\(_{12} \). In addition, ligand 50, containing C3 linkers, inhibits MBNL1-CUG interaction very weakly and thus an upper limit of IC\(_{50} \) > 6.8 mM.

### 4.6 Activity in a DM1 Drosophila Model

*Drosophila* has been a valuable animal model to study neurodegenerative diseases including human trinucleotide repeat disorders.\(^{103}\) Several *Drosophila* models of DM1 have been reported.\(^{39,104-106}\) Cataract in humans is a common symptom of DM. In *Drosophila* genetically modified with CTG repeats eye defects develop that are easily visualized by microscopy. The *Drosophila* models are advantageous for multiple reasons. In particular, the invertebrates are fast-growing and provide a rapid way to study the phenotypic change upon therapeutic treatment.

Given the promising *in vivo* results with ligand 51, we were interested in pursuing their application in the *Drosophila* models. Thus, in collaboration with Prof. Woody Chan and Prof. Edwin Chan at the Chinese University of Hong Kong, we tested ligand 51 with a *Drosophila* model of DM1 expressing 480 interrupted CUG repeats i(CUG)\(_{480} \).\(^{39,104}\) To our delight, ligand 51 reduced the toxic effect from the expanded CUG repeats as seen in the improvement in the DM1 eye histology under light microscope (Figure 51a). The glossy appearance of untreated flies is visible representing a severe phenotype. Upon treatment of ligand 51 (from 200 to 800 \( \mu M \)), the glossy phenotype is suppressed, presenting a less severe rough eye phenotype. Preliminary scanning electron microscope (SEM) images of these eyes provided further details on the eye structure (Figures 51b,c). Untreated DM1 fly eyes showed substantial fusion and disorganization.
of ommatidia with a loss of inter-ommatidial bristles (Figure 51c). Upon treatment of ligand 51 at 200 µM, the number of fused ommatidia is significantly reduced. At 400 and 800 µM, the individual ommatidium are further patterned in a more organized way with more increased number of inter-ommatidial bristles indicating a protective role for ligand 51 in preventing the DM phenotype.

Figure 51. Phenotypic change in eyes of DM1 Drosophila model upon treatment of ligand 51 from 200 to 800 µM observed under (a) light microscope (× 230), (b) SEM (× 230), and (c) SEM (× 700). Data obtained from Dr. Ho Tsoi, Prof. Edwin Chan, and Prof. Woody Chan.
4.7 Conclusions and Outlook

Through rational design, we developed a new generation of ligands (50–52) that has a very low toxicity profile and high aqueous solubility. In particular, ligand 51, containing C4 linkers, showed the most promising activity among others. Ligand 51 inhibits foci formation in DM1 cell culture and partially corrects splicing defects in a DM1 Drosophila model. All these results suggested that ligand 51 is a promising lead ligand for developing new approach to better inhibit the MBNL1-CUG interaction.

In this regard, we propose a new therapeutic strategy using template-assisted “click chemistry.” As explained in the Background (see also Figure 1), the toxic RNA involved in DM is an expanded CUG or CCUG repeat range from 100 to >10,000 repeating units (i.e., up to 30,000–40,000 nt in length). Thus, the use of oligomeric ligands should provide better inhibition because of a larger steric coverage of the RNA. In principle, an oligomeric ligand also has a higher affinity toward the RNA because of the multivalent binding.40 As discussed in Section 1.5, Disney and co-workers have put this concept forward to target both CUG and CCUG repeats using oligomeric peptoids.41-43,92,107 We believe that this in vivo click approach is benefited by both the repetitive nature of the RNA and high selectivity of the lead ligands. In essence, the expanded RNA will not only be the substrate of the ligands but also a template for “in situ” ligand

Scheme 9
oligomerization. This approach will bypass the difficulty of cellular uptake of high MW ligands yet providing a stronger binding and inhibition. Ligand 59 was thus synthesized for this purpose (Scheme 9). Preliminary studies indicated that ligand 59 possesses RNA-binding affinity and inhibitory activity comparable to that of 51. The two acetylene groups in 59 would allow oligomerization using in situ click chemistry.

Finally, it is hoped that the compounds and concepts developed in this dissertation will contribute to the development of therapeutic strategy for curing the myotonic dystrophy disease.
Chapter 5  Materials and Methods

Materials. Single-stranded oligodeoxynucleotides (purified by standard desalting) were purchased from Integrated DNA Technologies (Coralville, IA). The DNA sample solutions were prepared using TE buffer (pH 7.6) and the concentrations were determined by performing absorbance measurements at 25 °C on a Shimadzu UV-2501PC spectrophotometer (Kyoto, Japan). The concentration of each single-stranded DNA was calculated using Beer’s law with the extinction coefficient ($\varepsilon_{260}$) provided by the supplier. The double-stranded DNA solution was freshly prepared by mixing required volumes of the corresponding ssDNA. The dsDNA was then annealed by heating in a water bath at 90 °C for 5 min and slow cooling to room temperature. MOPS buffer solution (100 mM), NaCl solution (3.0 M), and Millipore water were added to make up a DNA solution with 20 mM MOPS (pH 7.0 ± 0.2) and 300 mM NaCl.

Thermal denaturation studies. The melting temperatures of the duplexes were measured on a Shimadzu UV-2501PC spectrometer equipped with a temperature-controller. The absorbance of each DNA solution (12 $\mu$M duplex) with 20 mM MOPS (pH 7.0 ± 0.2), 300 mM NaCl, and 10 mM EDTA was annealed and cooled to 0 °C. The absorbance of each sample was monitored at 260 nm from 0 °C to 90 °C at a ramp rate 1 °C/min. The melting temperature ($T_m$) of each sample was determined from the maximum point of the first derivative of the melting curve with Origin 7.0 (Northampton, MA).

Isothermal titration calorimetry (ITC) studies. ITC measurements were performed at 25 °C on a MicroCal VP-ITC calorimeter (Northampton, MA). A typical experiment
consisted of titrating 10 µL of a ligand solution (500 µM) from a 250 µL syringe (stirred at 300 rpm) into a sample cell containing 1.42 mL of a DNA solution (10–20 µM) with total 28 injections. The initial delay prior to the first injection was 60 s. The duration of each injection was 24 s and the delay between injections was 300 s. The heat of dilution was determined using a blank solution containing no DNA. The ITC data was then subtracted from the blank. All experiments were repeated at least twice. Data analysis was carried out with Origin 5.0 software (MicroCal). Binding parameters such as the dissociation constant \(K_d\), enthalpy change \(\Delta H\), and entropy change \(\Delta S\) were determined by fitting the experimental binding isotherms with appropriate models. The ligand stock solution was 10 mM in DMSO. The buffer solution for ITC experiments was MOPS (20 mM; pH 7.0 ± 0.2), NaCl (300 mM) and 5–10% (v/v) DMSO to balance the residual DMSO in the ligand solution.

**Job Plot Analysis.** Fluorescence measurements were carried out at 25 °C on a HORIBA Jobin Yvon (Edison, NJ) FluoroMax-3 spectrofluorometer using a 250 µL quartz cell with 0.1 cm path length. The concentrations of the DNA and ligand were varied in each sample whereas the sum of the concentrations \(c_0\) was kept constant. All solution contained MOPS (20 mM; pH 7.0 ± 0.2) and NaCl (300 mM). The blanks consisted of the same solutions with no DNA were used to correct the dilution effect. The fluorescence signal was measured with an excitation wavelength of 310 nm and emission wavelength was collected from 470 to 530 nm. The fluorescence intensity of the sample \(F\) and the blank \(F_0\) at 495 nm for each mole fraction of ligand \(\chi_L\) was used for the Job plot analysis. A Job plot was constructed by plotting the change in
fluorescence ($\Delta F = F_0 - F$) versus the mole fraction. The stoichiometry was determined from the intercept of the two tangents aligned at the positions $\chi_L = 0.0$ and 1.0.

**Circular dichroism (CD) studies.** The CD measurements were carried at 25 °C on a Jasco J-715 spectropolarimeter (Easton, MD) using a cylindrical quartz cell with 1 mm path length. A solution of DNA (50 $\mu$M duplex, 20 mM MOPS, pH 7.0 ± 0.2, and 300 mM NaCl) was titrated with a concentrated ligand solution (10 mM in DMSO). The titration was carried out with injection of 5 $\mu$L of the ligand solution until a saturation of CD signal was reached. The CD signals were collected from 400 to 240 nm. A titration curve was then constructed by plotting the CD signal at 288 nm versus the mole ratio of ligand to duplex.

**Synthetic procedures.** $^1$H and $^{13}$C NMR spectra for structural characterization were recorded on a Varian Unity Inova 500 spectrometer ($^1$H: 500 MHz; $^{13}$C: 125 MHz). Unless otherwise stated, all NMR measurements were carried out in CDCl$_3$ at room temperature. Chemical shifts were reported as parts per million (ppm) in scale using residual peaks of CDCl$_3$ ($^1$H: 7.26; $^{13}$C: 77.16) or DMSO-$d_6$ ($^1$H: 2.50; $^{13}$C: 39.52) as internal standards. Coupling constants ($J$) were reported in hertz. Mass spectra were obtained by the Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois. High performance liquid chromatography (HPLC) was performed by a Dynamax SD-200 system with a UV, detector set at 254 nm, using an Alltech Denali C-18 column (250 x 10 mm) using a dual solvent system of 0.1% (v/v) formic acid/H$_2$O (Solvent A) and 0.1% (v/v) formic acid/MeCN (Solvent B). Melting points were measured on a Thomas Hoover capillary melting point apparatus (Philadelphia, PA) and
Elemental analyses were performed in the Microanalysis Laboratory, School of Chemical Sciences, University of Illinois.

All non-aqueous reactions were carried out under dry N\textsubscript{2} atmosphere with oven-dried (115 °C) glassware. Unless otherwise stated, all solvents and reagents were of reagent quality, purchased commercially, and used without further purification. Tetrahydrofuran (THF) was freshly distilled from Na/benzophenone ketyl under N\textsubscript{2}. Toluene was freshly distilled from Na sphere under N\textsubscript{2}. Dichloromethane (CH\textsubscript{2}Cl\textsubscript{2}) was freshly distilled from CaH\textsubscript{2} under N\textsubscript{2}. The progress of reaction was monitored by thin layer chromatography (TLC) performed on Merck pre-coated silica gel 60F\textsubscript{254} plates, and compounds were visualized with a spray of 5% (w/v) phosphomolybdic acid hydrate in ethanol with subsequent heating. Flash chromatography was carried out on columns of Macherey-Nagel MN Kieselgel 60 M (230–400 mesh) silica gel.

The synthetic procedures for unpublished compounds (i.e., 37, 43, 44, 50, 51, 52, 59, and 63) published will be reported below:

**Ligand 37:**

![Chemical structure of ligand 37]

To a stirred mixture of 91 mg (0.32 mmol) of compound 39, 98 mg (0.32 mmol) of compound 40, 81 mg (0.42 mmol) of EDCI, and 25 mg (0.19 mmol) of HOBt in 10 mL of DMF was added 0.14 mL of DIPEA (0.80 mmol) at 0 °C. The temperature was raised to 25 °C after the addition. The reaction mixture was stirred for 24 h. The solvent was removed in vacuo and the residue was dissolved in MeOH and dry-packed for flash column chromatography (MeOH:CH\textsubscript{2}Cl\textsubscript{2}:NH\textsubscript{4}OH) to give 90 mg (60%) of ligand 37 as a
yellow solid. $^1$H NMR (DMSO-$d_6$): $\delta$ 9.90 (br s, 1 H, NH), 8.59 (d, $J = 9.5$ Hz, 1 H, ArH), 8.50 (s, 1 H, ArH), 8.01 (s, 1 H, ArH), 7.90 (d, $J = 4.5$ Hz, 1 H, ArH) 7.89 (br s, 1 H, NH), 7.65 (dd, $J = 1.5$ Hz and 9 Hz, 1 H, ArH), 7.46 (d, $J = 9$ Hz, 1 H, ArH), 7.06 (br s, 4 H, NH$_2$), 4.17 (br s, 2 H, CH$_2$), 3.96 (s, 3 H, OCH$_3$), 3.58 (m, 2 H, CH$_2$), 3.17 (s, 2 H, CH$_2$CO).

$N$-(6-aminopentyl)melamine (63):

\[
\begin{align*}
\text{Cl} & \quad \text{H}_2\text{N(CH}_2\text{)NH}_2 \\
\text{H}_2\text{N} & \quad \text{H}_2\text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{NH} & \quad \text{NH} \\
\text{NH} & \quad \text{NH} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\end{align*}
\]

A mixture of 1.24 g (8.52 mmol) of chlorodiaminotriazine and 4.39 g (42.9 mmol) of 1,5-diaminopentane was heated in an oil bath at 130 °C for 5 h. The reaction mixture was cooled to room temperature. A solution of 10 mL of 35% NH$_4$OH aqueous solution was added. The precipitated was filtered and triturated with 50 mL of EtOH. The suspension was filtered and dried in vacuo to give 0.402 g (22%) of compound 63 as a pale yellow solid. $^1$H NMR (DMSO-$d_6$): $\delta$ 6.43 (t, $J = 5.5$ Hz, 1 H, MelNHCH$_2$), 6.10 (br s, 2 H, NH$_2$), 5.93 (br s, 2 H, NH$_2$), 4.84 (br s, 2 H, CH$_2$NH$_2$), 3.15 (q, $J = 6.5$ Hz, 2 H, MelNHCH$_2$), 2.62–2.59 (m, 2 H, CH$_2$), 1.45–1.40 (m, 4 H, CH$_2$), 1.06–1.03 (m, 2 H, CH$_2$).

Ligand 43:

\[
\begin{align*}
\text{Cl} & \quad \text{H}_2\text{N(CH}_2\text{)NH}_2 \\
\text{H}_2\text{N} & \quad \text{H}_2\text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{N} & \quad \text{N} \\
\text{NH} & \quad \text{NH} \\
\text{NH} & \quad \text{NH} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\text{H} & \quad \text{H} \\
\end{align*}
\]

A mixture of 0.402 g (1.89 mmol) of compound 63, 0.667 g (1.99 mmol) of compound 24, and 0.20 mL of TFA in 20 mL of acetonitrile was refluxed to reflux for 6 h. The reaction mixture was cooled to room temperature and concentrated in vacuo. The residue was dry-packed with silica gel and purified by flash column chromatography (10% MeOH in CH$_2$Cl$_2$) to afford a yellow solid. The solid was treated with excess TFA and the dried in vacuo to afford 0.36 g (28%) compound 43 as a TFA salt. $^1$H NMR of the free base (DMSO-$d_6$): $\delta$ 8.31 (d, $J = 9$ Hz, 1 H, ArH), 7.87–7.82 (m, 2 H, ArH), 7.60
(s, 1 H, ArH), 7.40 (d, J = 8.5 Hz, 1 H, ArH), 7.30 (d, J = 8.5 Hz, 1 H, ArH), 6.80 (br s, 1 H, AcrNH), 6.42 (t, J = 6 Hz, 1 H, MeINH), 6.11 (br s, 2 H, NH$_2$), 5.95 (br s, 2 H, NH$_2$), 3.92 (s, 3 H, OCH$_3$), 3.72 (t, J = 7 Hz, 2 H, AcrNCH$_2$), 3.12 (q, J = 6 Hz, 2 H, MelNCH$_2$), 1.75–1.69 (m, 2 H, CH$_2$), 1.46–1.41 (m, 2 H, CH$_2$), 1.35–1.30 (m, 2 H, CH$_2$).

Anal. calcd for C$_{26}$H$_{27}$N$_8$O$_5$F$_6$Cl: C, 45.86; H, 4.00; N, 16.45. Found: C, 46.12; H, 3.86; N, 16.30.

**Ligand 44:**

![diagram](image)

A mixture of 0.367 g (1.86 mmol) of compound 55, 0.367 g (1.85 mmol) of compound 64, and 1.10 g (11.7 mmol) of phenol was heated in an oil bath at 140 °C for 12 h. The reaction mixture was cooled to room temperature and dissolved in 20 mL of MeOH. The solution was dry-packed with silica gel and purified by flash column chromatography (MeOH:CH$_2$Cl$_2$:NH$_4$OH = 2:10:0.1) to give 0.602 g (90%) of compound 44 as a pale yellow solid. $^1$H NMR (DMSO-$d_6$): δ 8.37 (d, J = 5.5 Hz, 1 H, ArH), 8.25 (d, J = 9 Hz, 1 H, ArH), 7.77 (d, J = 2 Hz, 1 H, ArH), 7.42 (dd, 2.5 Hz and 9 Hz, ArH, 1H), 7.29 (t, J = 5 Hz, 1 H, QCNHCH$_2$), 6.48 (t, 6 Hz, 1 H, MeINHCH$_2$), 6.45 (d, J = 5.5 Hz, 1 H, ArH, 1H), 6.10 (br s, 2 H, NH$_2$), 5.93 (br s, 2 H, NH$_2$), 3.29–3.22 (m, 4 H, CH$_2$), 1.68–1.63 (m, 2 H, CH$_2$), 1.60–1.56 (m, 2 H, CH$_2$). ESI-MS ($m/z$): 359.1 ([M + H]$^+$, 60%). ESI-HRMS ($m/z$): [M + H]$^+$ calcd for C$_{16}$H$_{20}$N$_8$Cl, 359.1499, found, 359.1501.

**Ligand 50:**

![diagram](image)
A 25-mL, two-necked, round-bottomed flask equipped with a magnetic stir bar containing a white suspension of 0.124 g (0.423 mmol) of compound 53 in 10 mL of anhydrous EtOH is cooled in an ice-water bath under a nitrogen atmosphere. To the suspension was added 0.30 mL (1.722 mmol) of DIPEA to produce a colorless clear solution. A solution of 0.175 g (0.953 mmol) of compound 65 in 5 mL of ethanol was added dropwise using a pipette. The reaction mixture was allowed to warm to 25 °C slowly and stirred for 18 h. The reaction was monitor by TLC (n-BuOH:AcOH:H₂O = 3:1:4; Rf = 0.15). The white suspension was filtered and washed with EtOH (20 mL). The white solid was dissolved in 15 mL of 1:9 (v/v) H₂O/MeOH mixture and chromatographed on short silica gel (~200 mL) column starting with CH₂Cl₂:n-BuOH:MeOH mixture (gradient from 1:5:4 to 1:3:6). Once the impurities were removed, the eluent was acidified with a gradient of 0.10–0.15 mL of 4 M HCl in dioxane (per liter of eluent) to afford the desired product. The product-containing fractions were combined, filtered, and concentrated in vacuo to afford 0.180 g (86%) of product as a white tetra·HCl salt. Rf(AcOH:H₂O:MeOH = 3:6:1) = 0.25. ¹H NMR (DMSO-d₆): 10.20 (s, 2 H, NH), 9.76 (s, 2 H, NH), 9.36 (s, 2 H, NH), 7.99 (s, 4 H, ArH), 7.34 (br s, 2 H, Het-NH), 6.77 (br s, 8 H, Het-NH), 3.49 (br s, 4 H, N=CNHC₂H₅), 3.32 (overlap with H₂O, 4 H, Het-NHC₂H₅), 1.89 (q, 4 H, J = 10 Hz, CH₂). ESI-MS [M+H]+:10% [M+2H]²⁺: 100%.

**Ligand 51:**

![Diagram of ligand 51](image)

A 50-mL, two-necked, round-bottomed flask equipped with a magnetic stir bar containing a white suspension of 1.11 g (3.79 mmol) of compound 53 in 50 mL of anhydrous EtOH is cooled in an ice-water bath under a nitrogen atmosphere. To the suspension was added 1.10 mL (7.89 mmol) of Et₃N to produce a colorless clear solution. A solution of 1.54 g (7.81 mmol) of compound 55 in 5 mL of ethylene glycol was added dropwise using a pipette. The reaction mixture was allowed to warm to 25
°C slowly and stirred for 18 h. The reaction was monitored by TLC (\(n\)-BuOH:AcOH:H\(_2\)O = 3:1:4; \(R_f = 0.1\)). The white suspension was filtered and washed with EtOH (20 mL). The white solid was dissolved in 15 mL of 1:9 (v/v) H\(_2\)O/MeOH mixture and chromatographed on short silica gel (~200 mL) column starting with CH\(_2\)Cl\(_2\):n-BuOH:MeOH mixture (gradient from 1:5:4 to 1:3:6). Once the impurities were removed, the eluent was acidified with a gradient of 0.10–0.15 mL of 4 M HCl in dioxane (per liter of eluent) to afford the desired product. The product-containing fractions were combined, filtered, and concentrated in vacuo to afford 1.95 g (77\%) of product as a white tetra-HCl salt (m.p. > 230 °C (decomp.)). \(R_f (\text{AcOH}:\text{H}_2\text{O}:\text{MeOH} = 3:6:1) = 0.30\). 1H NMR (DMSO-\(d_6\)): 10.25 (s, 2 H, NH), 9.81 (s, 2 H, NH), 9.43 (s, 2 H, NH), 7.98 (s, 4 H, ArH), 6.61 (s, 2 H, Het-NH), 6.19 (s, 4 H, Het-NH), 6.03 (s, 4 H, Het-NH), 3.47 (m, 4 H, N=CNHCH\(_2\)), 3.22 (q, 4 H, \(J = 10\) Hz, Het-NHC\(_2\)), 1.66 (q, 4 H, \(J = 10\) Hz, CH\(_2\)), 1.56 (q, 4 H, \(J = 10\) Hz, CH\(_2\)). 13C NMR (DMSO-\(d_6\)): 167.2, 166.9, 166.4, 161.5, 133.0, 128.6, 42.7, 26.6, 24.8 (one overlapped). HR-ESI-MS for \(\text{C}_{22}\text{H}_{35}\text{N}_{16}\): 523.3231 found: 523.3233. ESI-MS \([\text{M+H}]^+\):30% \([\text{M+2H}]^{2+}\): 100%. Elemental analysis for \(\text{C}_{22}\text{H}_{34}\text{N}_{16}\cdot4\text{HCl}\): C:39.53, H:5.73, N:33.53, Cl:21.21 found: C:39.32, H:5.89, N:31.96, Cl:21.77.

**Ligand 52:**

A 25-mL, two-necked, round-bottomed flask equipped with a magnetic stir bar containing a white suspension of 0.435 g (1.48 mmol) of compound 53 in 10 mL of anhydrous EtOH is cooled in an ice-water bath under a nitrogen atmosphere. To the suspension was added 0.60 mL (3.44 mmol) of DIPEA to produce a colorless clear solution. A solution of 0.658 g (3.11 mmol) of compound 63 in 10 mL of ethanol was added dropwise using a pipette. The reaction mixture was allowed to warm to 25 °C slowly and stirred for 18 h. The reaction was monitored by TLC (\(n\)-BuOH:AcOH:H\(_2\)O =
The white suspension was filtered and washed with EtOH (20 mL). The white solid was dissolved in 15 mL of 1:9 (v/v) H₂O/MeOH mixture and chromatographed on short silica gel (~200 mL) column starting with CH₂Cl₂:n-BuOH:MeOH mixture (gradient from 1:5:4 to 1:3:6). Once the impurities were removed, the eluent was acidified with a gradient of 0.10–0.15 mL of 4 M HCl in dioxane (per liter of eluent) to afford the desired product. The product-containing fractions were combined, filtered, and concentrated in vacuo to afford 0.751 g (73%) of product as a white tetra·HCl salt (m.p. > 260 °C (decomp.)). Rf (AcOH:H₂O:MeOH = 3:6:1) = 0.35. ¹H NMR (DMSO-d₆): 10.20 (s, 1 H, NH), 9.74 (s, 1 H, NH), 9.36 (s, 1 H, NH), 8.21 (br s, 2 H, Het-NH), 7.97 (s, 4 H, ArH), 7.89 (s, 8 H, Het-NH), 3.45 (q, 4 H, J = 5 Hz, N=CNHCH₂), 3.29 (q, 4 H, J = 10 Hz, Het-NHCH₂), 1.68 (q, 4 H, J = 10 Hz, CH₂), 1.54 (q, 4 H, J = 10 Hz, CH₂), 1.40 (br s, 4 H, CH₂). HR-ESI-MS for C₂₄H₃₉N₁₆: 551.3544 found: 5551.3549. ESI-MS [M+H]+:30% [M+2H]²⁺: 100%.

Bis-acetylene ligand 59:

A 25-mL, two-necked, round-bottomed flask equipped with a magnetic stir bar containing a white suspension of 0.402 g (1.37 mmol) of compound 53 in 10 mL of anhydrous EtOH is cooled in an ice-water bath under a nitrogen atmosphere. To the suspension was added 0.45 mL (3.23 mmol) of Et₃N to produce a colorless clear solution. A solution of 0.711 g (3.02 mmol) of compound 66 (prepared by Long Luu) in 2 mL of EtOH was added dropwise using a pipette. The reaction mixture was allowed to warm to 25 °C slowly and stirred for 24 h. The reaction was monitor by TLC (n-BuOH:AcOH:H₂O = 3:1:4; Rf = 0.2). The white suspension was filtered and washed with EtOH (20 mL). The white solid was dissolved in 5 mL of MeOH and chromatographed on short silica gel (~200 mL) column starting with 1:9 (v/v) CH₂Cl₂:MeOH mixture. Once the impurities were removed, the eluent was acidified with a gradient of 0.10–0.15 mL of
4 M HCl in dioxane (per liter of eluent) to afford the desired product. The product-containing fractions were combined, filtered, and concentrated *in vacuo* to afford 0.561 g (55%) of product as a white tetra·HCl salt (m.p. > 200 °C (decomp.)).

$R_f(AcOH:HO:MeOH = 3:6:1) = 0.45$. $^1$H NMR (DMSO-$d_6$): 10.32 (s, 2 H, NH), 9.86 (s, 2 H, NH), 9.48 (s, 2 H, NH), 7.99 (s, 4 H, ArH), 6.78 (br t, 4 H, Het-NH, rotamers), 6.23 (br t, 4 H, Het-NH, rotamers), 3.97 (br s, 4 H, CH$_2$C≡C), 3.49 (m, 4 H, N=CNHCH$_2$), 3.23 (br s, 4 H, Het-NHCH$_2$), 3.02 (s, 2 H, C≡CH), 1.66 (br s, 4 H, CH$_2$), 1.57 (br s, 4 H, CH$_2$). $^{13}$C NMR (DMSO-$d_6$): 166.6, 166.2, 165.6, 165.3, 161.5, 132.8, 128.6, 82.5, 72.2, 42.6, 29.2, 26.5, 24.8. ESI-MS [M+H]$^+$: 5% [M+2H]$^{2+}$: 100%.
Chapter 6 References


(14) O'Rourke, J. R.; Swanson, M. S. *J. Biol. Chem.* 2009, 284, 7419–7423.


(107) Tran, T.; Disney, M. D. Biochemistry 2011, 50, 962–969.
Appendix: Publications


Selective inhibition of MBNL1–CCUG interaction by small molecules toward potential therapeutic agents for myotonic dystrophy type 2 (DM2)

Chun-Ho Wong, Yuan Fu, Sreenivasa Rao Ramisetty, Anne M. Baranger* and Steven C. Zimmerman*

Department of Chemistry, University of Illinois at Urbana-Champaign, 600 S. Mathews Avenue, Urbana, Illinois 61801

Received January 1, 2011; Revised April 17, 2011; Accepted May 10, 2011

ABSTRACT

Myotonic dystrophy type 2 (DM2) is an incurable neuromuscular disease caused by expanded CCUG repeats that may exhibit toxicity by sequestering the splicing regulator MBNL1. A series of triaminotriazine- and triaminopyrimidine-based small molecules (ligands 1–3) were designed, synthesized and tested as inhibitors of the MBNL1–CCUG interaction. Despite the structural similarities of the triaminotriazine and triaminopyrimidine units, the triaminopyrimidine-based ligands bind with low micromolar affinity to CCUG repeats ($K_d \sim 0.1–3.6 \mu M$) whereas the triaminotriazine ligands do not bind CCUG repeats. Importantly, these simple and small triaminopyrimidine ligands exhibit both strong inhibition ($K_i \sim 2 \mu M$) of the MBNL1–CCUG interaction and high selectivity for CCUG repeats over other RNA targets. These experiments suggest these compounds are potential lead agents for the treatment of DM2.

INTRODUCTION

Myotonic dystrophy (DM) is an RNA-mediated disorder that affects both skeletal and smooth muscles as well as internal organs such as the heart, lung and eye lens (1). Two major types of DM have been identified: types 1 (DM1) and 2 (DM2). It was discovered that both DM1 and DM2, respectively, form inclusion complexes that sequester muscleblind-like (MBNL) proteins in the nucleus. A member of a family of structurally and functionally similar proteins, MBNL is expressed at high level in skeletal muscles and is known to be directly involved in causing DM. MBNL1 is involved in the regulation of RNA splicing, editing and translation (4). The sequestration of MBNL1 proteins results in misregulated, alternative splicing of several pre-mRNAs, including the cardiac troponin T (cTNT), fast skeletal troponin T (Tnnt3), insulin receptor (IR) and chloride channel (ClC-1) pre-mRNAs. It is proposed that these errors in alternative splicing, in turn, cause the multisystemic clinical features of DM (5). It has been suggested that agents that bind selectively to the pathogenic RNAs might reverse the phenotype by inhibiting MBNL1 protein from binding, thereby allowing the protein and mRNAs to function normally (6,7). Indeed, antisense oligonucleotides that target the CUG repeat were shown to reverse the phenotype of DM1 in a mouse model (8,9).

With the knowledge that disrupting the MBNL1–CUG repeat complex might provide a therapeutic strategy to treat DM1, recent efforts have focused on small molecules that target CUG repeats (10–13). An ideal candidate should also reverse the splicing defects observed in DM. For example, pentamidine is capable of rescuing four different pre-mRNAs affected by CUG repeats as well as displacing MBNL1 from CUG ribonuclear foci (10). In this regard, we have developed a simple, low molecular weight ligand (ligand 1) that exhibits high selectivity in binding the CUG base triplet and, further, inhibits the MBNL1–CUG interaction with micromolar IC50 values (Figures 1 and 2) (12). As an alternative-splicing regulator, 11 000 repeats) in intron 1 of the ZNF9 (zinc-finger protein 9) gene on chromosome 3 (3). The corresponding expanded transcripts, r(CUG)$_n$ and r(CCUG)$_n$ for DM1 and DM2, respectively, form inclusion complexes that sequester muscleblind-like (MBNL) proteins in the nucleus.
MBNL1 binds to other stem-loop structures containing pyrimidine–pyrimidine mismatches, including the pre-mRNAs discussed above. Therefore, it is essential to develop ligands that bind selectively to the pathogenic RNAs in preference to the natural targets of the MBNL1 and other nucleic acids. Indeed, achieving RNA and sequence selectivity is particularly difficult given the presence of a large amount of genomic DNA in the nucleus. The majority of work to date has focused on DM1, however recently, Disney and coworkers (14) reported oligomeric ligands that target expanded CCUG repeats in DM2. Specifically, a trimeric ligand with three kanamycin A units was shown to bind strongly to CCUG RNA and inhibit the MBNL1–CCUG interaction. A subsequent report showed that the flexibility of the ligand backbone plays a critical role in determining affinity and selectivity (15). Despite these important advances, alternative approaches are likely needed given the challenges in developing clinically useful therapeutic agents.

In examining the interaction of CUG repeats with ligand 1, the triaminotriazine moiety was found to be important for the selective binding of U–U mismatches, whereas the acridine unit provided binding affinity for the CUG RNA (12). Stacking of the two heterocyclic units, which was observed by 1H NMR, is likely to increase the mismatch selectivity by significantly reducing the ability of the acridine to function as an unselective RNA intercalator. Although the binding mode is not known, ligand 1 was designed to intercalate between the U–U and G–C base pairs and simultaneously form a base triplet with the triaminotriazine unit inserted between the U–U mismatch (Figure 1). Therefore, small molecule ligands structurally similar to ligand 1 (IC50 = 43 μM for MBNL1–CUG) were explored. Herein, we report on efforts toward a similar approach, to develop π-stacked intercalators that specifically recognize CCUG repeats and inhibit the MBNL1–CCUG interaction.

**MATERIALS AND METHODS**

**Materials**

RNA oligonucleotides were purchased from Dharmacon RNAi Technologies, Inc. with 20-deprotection, desalting and HPLC purification. DNA sequences were purchased from Integrated DNA Technologies, Inc. with standard desalting. The RNA oligonucleotides were dissolved in Ambion RNA Storage Solution (pH 6.4) to give 1–2 mM stock solutions. DNA sequences were dissolved in TE buffer (pH 7.6) to give 1–2 mM stock solutions. The oligonucleotide concentration was determined by its absorbance at λmax = 260 nm at 25°C on a Shimadzu UV-2501PC spectrophotometer. The concentration of each single-stranded sequence was calculated using Beer’s law with extinction coefficient (ε260) provided by the supplier.

RNA and DNA duplexes were freshly prepared by mixing the required volumes of the corresponding ssRNA and ssDNA, respectively. The solution was annealed by heating in a water bath at 95°C for 2 min and slow cooling to 25°C. Final duplex solutions were prepared by adding MOPS buffer (100 mM; pH 7.0), NaCl solution (5.0 M) and diluting to the required concentrations by adding AccuGENE Molecular Biology Grade Water and 1 or 10% (v/v) DMSO to give 20 mM MOPS and 300 mM NaCl solutions.

**Protein expression and purification**

An expression vector for a truncated MBNL1 comprised of amino acids 1–272 was obtained from Maurice S. Swanson (University of Florida, College of Medicine, Gainesville, FL, USA). This MBNL1 construct is comprised of the four zinc-finger motifs of MBNL1 and a hexahistidine tag (C-terminus) and binds RNA with similar affinity as the full-length MBNL1. The protein was expressed and purified as described previously (16).
The molecular weight was confirmed by MALDI-TOF mass spectrometry, the concentration was determined by Bradford protein assay (Bio-Rad), and the purity determined by silver-stained SDS-PAGE.

Electrophoretic mobility shift assay
RNAs were labeled with [γ-32P]ATP using T4 polynucleotide kinase (New England Biolabs) and were purified by phenol extraction and precipitation from ethanol. Labeled RNA (5 μl, 5 nM) was heated at 95°C for 2 min, placed on ice for 10 min and diluted to 125 μl in RNA storage buffer [66 mM NaCl, 6.7 mM MgCl2, 27 mM Tris-HCl (pH 7.5)]. For assays with competitor rRNA, bulk yeast rRNA was added to the RNA solution to give a final concentration of ~100 nM. MBNL1 was serially diluted in binding buffer [175 mM NaCl, 5 mM MgCl2, 20 mM Tris-HCl (pH 7.5), 1.25 mM 2-mercaptoethanol (BME), 12.5% glycerol, 0.4 mg/ml bovine serum albumin (BSA), 0.1 mg/ml heparin]. Protein (5 μl) and RNA (5 μl) solutions were combined. The reaction mixture was incubated at room temperature for 25 min and loaded onto a 6% polyacrylamide gel (80:1) at 4°C. The gels were run for 1 h at 360 V in 22.5 mM Tris-borate buffer (pH 8). Gels were visualized on a Molecular Dynamics Storm PhosphorImager. The apparent molecular weight was confirmed by MALDI-TOF mass spectrometry.

Inhibition assays using gel electrophoretic mobility shift assay
The inhibition of the MBNL1–RNA complex was investigated using the above procedure, but with addition of the small molecule to the RNA–protein complex after 25 min of incubation. The reaction mixture was incubated for an additional 10–15 min at room temperature. The inhibition assays were performed in the presence of 1 or 10% (v/v) DMSO (depending on the solubility of ligand). The IC50 values were determined by fitting to the equation: \( B = \Delta B \exp(-0.69[\text{ligand}]/[\text{IC}50]) + B_{\text{initial}} \), where \( B \) is the radioactivity (cpm) of the RNA, \( \Delta B = B_{\text{final}} - B_{\text{initial}} \), [ligand] is the ligand concentration, IC50 is the concentration of ligand, at which \( (F - 0.5\Delta B + B_{\text{initial}}) \). The apparent inhibition constant (Ki) was determined using the equation: \( K_i = \text{IC}50 \times (K_d/\text{protein}_{\text{total}}) \), where \( K_d \) is the dissociation constant of the MBNL1–RNA complex and [protein] is at least 7-fold greater than the \( K_i \).

Thermal denaturation studies
The melting temperatures of the duplexes were measured on a Shimadzu UV-2501PC spectrophotometer equipped with a temperature controller. The absorbance of each RNA or DNA solution (10 μM duplex) with 20 mM MOPS (pH 7.0), 300 mM NaCl and 10 mM EDTA was annealed and cooled to 0°C. The absorbance of each sample was monitored at 260 nm from 0°C to 90°C at a ramp rate of 1°C/min. The melting temperature \( (T_m) \) of each sample was determined from the maximum point of the first derivative of the melting curve with Origin 7.0.

Isothermal titration calorimetry studies
Isothermal titration calorimetry (ITC) measurements were performed at 25°C on a MicroCal VP-ITC calorimeter. A typical experiment consisted of titrating 10 μl of a ligand solution (500 μM) from a 250 μl syringe (stirred at 300 rpm) into a sample cell containing 1.42 ml of a RNA or DNA solution (10–20 μM) with a total of 30 injections (1 μl for the first injection and 10 μl for the remaining injections). The initial delay prior to the first injection was 60 s. The duration of each injection was 24 s and the delay between injections was 400 s. All experiments were performed in triplicate. Data analysis was carried out with Origin 5.0 software (MicroCal). Binding parameters, such as the dissociation constant (Kd), enthalpy change (ΔH) and entropy change (ΔS), were determined by fitting the experiments to the one binding isotherms with appropriate models. The ligand stock solution was 50 mM in DMSO. The buffer solution for ITC experiments was MOPS (20 mM; pH 7.0), NaCl (300 mM) and 1% (v/v) DMSO to balance the DMSO in the ligand solution.

Synthetic procedures
1H and 13C NMR spectra for structural characterization were recorded on a Varian Unity Inova 500 spectrometer (1H: 500 MHz; 13C: 125 MHz). Unless otherwise stated, all NMR measurements were carried out in CDCl3 at room temperature. Chemical shifts were reported as parts per million (ppm) in scale using residual peaks of DMSO-d6 (1H: 2.50, 13C: 39.52) as internal standard. Coupling constants (J) were reported in Hertz. Mass spectra were obtained by the Mass Spectrometry Laboratory, School of Chemical Sciences, University of Illinois.

All non-aqueous reactions were carried out under dry N2 atm atmosphere with oven-dried (115°C) glassware. Unless otherwise stated, all solvents and reagents were of reagent quality, purchased commercially and used without further purification. The progress of reaction was monitored by thin layer chromatography (TLC) performed on Merck precoated silica gel 60F254 plates, and compounds were visualized by UV and subsequently with a spray of 5% (w/v) phosphomolybdic acid hydrate in ethanol with subsequent heating. Flash chromatography was carried out on columns of Macherey-Nagel MN Kieselgel 60 M (230–400 mesh) silica gel or Sigma-Aldrich Brockmann I, standard grade (~150 mesh) alumina.

Ligand 2
A mixture of 163 mg (0.830 mmol) of 5-(4-aminobutylyl)-2,4,6-triaminopyrimidine (17), 334 mg (0.996 mmol) of 6-chloro-2-methoxy-9-phenoxacyridine and 0.17 ml (0.98 mmol) of DIPEA in 6 ml of DMF was heated to 90°C in an oil bath for 48 h. The reaction mixture was cooled to room temperature and concentrated in vacuo. The residue was purified by flash chromatography on alumina using CH2Cl2: MeOH mixture (9:1 gradient to 4:1) to afford 279 mg (0.638 mmol, 77%) of ligand 2 as a yellow solid.
7.87 (d, J = 2 Hz, ArH, 1H), 7.83 (d, J = 9.5 Hz, ArH, 1H), 7.60 (d, J = 2 Hz, ArH, 1H), 7.40 (dd, 2.5 and 9.5 Hz, ArH, 1H), 7.32 (dd, 2 and 9 Hz, ArH, 1H), 6.91 (t, J = 5.5 Hz, ArCH, 1H), 5.40 (s, NH2, 4H), 5.11 (s, NH2, 2H), 5.39 (s, OCH3, 3H), 5.37 (q, J = 7 Hz, NHCH2, 2H), 2.21 (t, J = 7.5 Hz, PymCH2, 2H), 7.45–7.37 (m, ArH, 2H), 5.43 (s, NH2, 4H), 5.13 (s, NH2, 2H), 3.86 (t, J = 6.5 Hz, CONH, 1H), 2.57 (t, J = 5.5 Hz, ArCH2, 2H), 2.55 (t, J = 6 Hz, PymCH2, 2H), 2.31 (s, N(CH3)2, 6H), 2.21 (quin, J = 7.5 Hz, 2H). 

Results

Binding of ligand 1 to CCTG and CCUG sequences

Binding studies of ligand 1 were initially performed with a DNA duplex containing a single (CCTG) step (Figure 3a). Using ITC, it was found that ligand 1 binds strongly to the CCTG duplex with a 3 : 1 stoichiometry and $K_d(1) = 14 \mu M$. $K_d(2) = 0.11 \mu M$ and $K_d(3) = 37 \mu M$. Job analysis and circular dichroism titration studies support the 3 : 1 stoichiometry (see Supplementary Data). The binding to RNA was examined using the 11-mer shown in Figure 3b. In contrast to the results with the DNA oligonucleotide, ligand 1 binds CCUG comparatively weakly; a single $K_d = 79 \mu M$ was obtained. The CCUG repeat is capable of forming a slipped structure containing U-U and C-C mismatches separated by C-G base pairs as shown in Figure 4 (19). For this reason, the binding of ligand 1 to the slipped sequence was also investigated by ITC using the 11-mer shown in Figure 3c. A 2 : 1 complex was observed with low-micromolar $K_d$ values: $K_d(1) = 3.5 \mu M$ and $K_d(2) = 7.0 \mu M$ (Figure 3c).

To determine whether ligand 1 can inhibit the MBNL1–CCUG interaction, an electrophoretic mobility shift assay using (a) a DNA sequence containing a CCTG repeat, (b) an RNA sequence containing a CCUG repeat (RNA A) and (c) an RNA containing a slipped-CCUG motif (RNA B).
was performed using a (CCUG)₆ RNA. The Kᵩ values for the complexes formed between (CCUG)₆ and MBNL1 were determined in the absence and presence of 100 nM of competitor tRNA to be 15 ± 2 nM and 30 ± 4 nM, respectively (Figure 5). These values are in good agreement with literature values (19). No significant inhibition of the MBNL1–RNA interaction was observed with ligand 1 up to a concentration of 250 μM (see Supplementary Data).

Development of ligands 2 and 3

The high affinity of ligand 1 for U–U mismatches and moderate affinity for C–C mismatches led us to explore similar ligands that recognize C–U mismatches in CCUG repeats. To this end, a structurally similar analog 2 was synthesized and studied (Scheme 1). The ability of ligand 2 to bind CCUG and CCTG repeats was examined by ITC, however, these studies were hampered by poor water solubility. To achieve the high ligand concentration needed for ITC studies, ligand 2 required more than 10% (v/v) DMSO. With as little as 1% (v/v) DMSO, a 500 μM solution of ligand 3 could be prepared in a 300 mM NaCl solution (MOPS buffer; pH 7), presumably because it contains a water-solubilizing substituent. Because experiments with ligand 2 were performed in the presence of 10% DMSO, the melting temperature (Tₘ) of (CCUG)₆ was measured in the presence and absence of 10% DMSO. The stability of the RNA was found to be identical within experimental error in the presence of 1–10% DMSO. In addition, the Tₘ of (CCUG)₆ was concentration independent suggesting the formation of a hairpin not a duplex. The hydrolytic stability of (CCUG)₆ RNA in the presence of ligands 2 and 3 was also evaluated. The (CCUG)₆ RNA was incubated with ligands 2 and 3 for at least 4 h at room temperature and the RNA was evaluated by denaturing polyacrylamide gel electrophoresis (see Supplementary Data). No cleavage of the RNA was observed under these conditions. Therefore, the RNA is stable to hydrolytic cleavage in all of the binding and inhibition experiments reported herein.

Affinity and selectivity of ligand 3 for RNA

The affinity and selectivity of ligand 3 for binding to CCUG RNA was evaluated by measuring the affinity of ligand 3 for several oligonucleotides (Figure 6).
The stabilities of these RNAs were evaluated by melting curves obtained by UV spectroscopy and are reported in Table 1. The duplexes containing C–U (RNA C), U–U (RNA D), C–C (RNA E), A–A (RNA F) and G–G (RNA G) were designed to evaluate the ability of ligand 3 to selectively recognize the mismatches that are most likely found in the poly(CCCUG) RNA compared to related mismatched sequences. The binding of ligand 3 to the DNA analog of the DM1 sequence containing a T–T mismatch (DNA H) and the 18-nt fragment of the human cTNT (htTNT) pre-mRNA, a natural target of MBNL1 (RNA I) was also evaluated. To investigate the non-specific binding of ligand 3 to nucleic acids in general, its affinity toward tRNA and herring sperm DNA (hsDNA) was measured (Table 1).

ITC was used to measure the affinity of ligand 3 for the duplexes (Table 1). For (CCUG)_6 RNA, ligand 3 bound with low micromolar affinity \( K_d(1) = 5.2 \pm 0.6 \) M and a significantly weaker \( K_d(2) = 478 \pm 85 \) M. Ligand 3 binds to RNA A, which contains a single CCUG site, i.e. two proximal U–C mismatches (RNA A), similarly, exhibiting a low micromolar affinity \( K_d(1) = 3.6 \pm 0.8 \) M and \( K_d(2) = 258 \pm 37 \) M. The affinity for the slipped-CCUG structure (RNA B) was significantly higher, with first and second \( K_d \) values of \( 0.13 \pm 0.04 \) M and \( 10 \pm 5 \) M, respectively. For comparison purposes, the \( K_d \) values for the duplexes containing a single C–U (RNA C, \( K_d = 85 \pm 30 \) M), U–U (RNA D, \( K_d = 25 \pm 13 \) M) and C–C mismatches (RNA E, \( K_d = 0.77 \pm 0.28 \) M) were measured.

Completing the picture, the binding of ligand 3 was measured to duplexes F and G containing A–A and G–G mismatches, respectively, and to DNA H with a CTG site relevant in DM1 (20). The purine–purine mismatches were bound weakly by ligand 3 (\( K_d > 100 \) M) whereas the CTG sequence (DNA H), was bound with a low micromolar affinity (\( K_d = 1.4 \pm 1.5 \) M). Ligand 3 was found to bind weakly (\( K_d = 334 \pm 11 \) M) to the 18-nt fragment of \( h_tTNT \) pre-mRNA (RNA I) and tRNA (\( K_d = 2400 \pm 2100 \) M). Finally, ligand 3 bound herring sperm DNA with \( K_d = 36 \pm 7 \) M.

**Table 1.** Dissociation constants (\( K_d \)) and \( T_m \) values for all tested nucleic acids with ligand 3*.

<table>
<thead>
<tr>
<th>Nucleic acids</th>
<th>Specific target</th>
<th>( T_m ) (°C)</th>
<th>( K_d(1) )</th>
<th>( K_d(2) )</th>
<th>( K_d(3) )</th>
<th>Selectivity*</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CCUG)_6</td>
<td>C–U pair**</td>
<td>61</td>
<td>5.2 ± 0.6</td>
<td>478 ± 85</td>
<td></td>
<td>1.4</td>
</tr>
<tr>
<td>A</td>
<td>C–U pairs</td>
<td>53</td>
<td>3.6 ± 0.8</td>
<td>258 ± 37</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>B</td>
<td>C–C and U–U</td>
<td>54</td>
<td>0.13 ± 0.04</td>
<td>10 ± 5</td>
<td>202 ± 14</td>
<td>0.04</td>
</tr>
<tr>
<td>C</td>
<td>Single C–U pair</td>
<td>62</td>
<td>85 ± 30</td>
<td>231 ± 55</td>
<td></td>
<td>24.0</td>
</tr>
<tr>
<td>D</td>
<td>U–U</td>
<td>63</td>
<td>25 ± 13</td>
<td>271 ± 83</td>
<td></td>
<td>6.9</td>
</tr>
<tr>
<td>E</td>
<td>C–C</td>
<td>62</td>
<td>0.77 ± 0.28</td>
<td>162 ± 8</td>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>F</td>
<td>A–A</td>
<td>62</td>
<td>118 ± 8</td>
<td>2270 ± 360</td>
<td></td>
<td>33.0</td>
</tr>
<tr>
<td>G</td>
<td>G–G</td>
<td>67</td>
<td>103 ± 26</td>
<td>7690 ± 280</td>
<td></td>
<td>29.0</td>
</tr>
<tr>
<td>H</td>
<td>T–T</td>
<td>51</td>
<td>14 ± 1.5</td>
<td>102 ± 25</td>
<td></td>
<td>0.4</td>
</tr>
<tr>
<td>I</td>
<td></td>
<td>334 ± 11</td>
<td></td>
<td></td>
<td></td>
<td>93.0</td>
</tr>
<tr>
<td>tRNA</td>
<td></td>
<td>2400 ± 2100</td>
<td></td>
<td></td>
<td></td>
<td>670.0</td>
</tr>
<tr>
<td>hsDNA</td>
<td></td>
<td>36 ± 7*</td>
<td></td>
<td></td>
<td></td>
<td>10.0</td>
</tr>
</tbody>
</table>

*\( K_d \) values are reported in M. **Concentrations of duplexes are 10 μM; additional 1% (v/v) DMSO has negligible effect on the stability of the duplexes (\( \Delta T_m < 2 \) °C). Selectivity is the ratio of \( K_d(1) \) of the particular nucleic acid to that of RNA A. *m-fold predicts two 2 x 2 nt internal loops.

**Figure 6.** Nucleic acids studied in ITC experiments (tRNA and herring sperm DNA are not shown). The structures shown represent the most stable structures predicted by m-fold. Note that RNA I is most likely single stranded under the conditions used to perform the binding or complex inhibition experiments.
Both ligands 2 and 3 were found to significantly destabilize the MBNL1–CCUG complex. The IC\textsubscript{50} of ligands 2 and 3 were determined by incubating varying concentrations of the ligands with a constant concentration of MBNL1–CCUG complex and evaluating the results using gel shift assays. The IC\textsubscript{50} of these ligands was determined using the equation $K_i = IC_{50} \times (K_a/\text{protein})_{	ext{total}}$, in which $K_i$ is the dissociation constant of the MBNL1–RNA complex and the concentration of MBNL1 is in large excess of the $K_i$ value. The comparison of $K_i$ values, rather than the IC\textsubscript{50} values, under different conditions is important in order to compare the effects of the conditions on the behavior of the ligands independently of the effects of the varying conditions on the $K_i$ of the MBNL1–CCUG complex. Importantly, ligands 2 and 3 do not bind to MBNL1 as measured by ITC.

To evaluate the specificity of ligands 2 and 3 for CCUG RNA, the inhibition experiments were performed in the presence of competitor RNA sequences. For ligand 2, the $K_i$ value is increased only slightly to 3.4 ± 0.9 μM in the presence of 100 nM tRNA (Figure 7a and c). Evaluation of ligand 3 provided IC\textsubscript{50} values of 52 ± 8 μM and 35 ± 3 μM in the absence and presence of 100 nM tRNA, respectively, with a $K_i$ of 2.2 ± 0.3 μM and 2.8 ± 0.3 μM, respectively (Figure 7b and c). The selectivity of ligands 2 and 3 for the target sequences (CCUG)$_6$ relative to (CUG)$_{12}$ (i.e. DM2 versus DM1) and their relative abilities to inhibit MBNL1 binding to the same sequences were also evaluated. Ligand 2 inhibited binding to (CCUG)$_6$ and (CUG)$_{12}$ with similar $K_i$ values of 2.4 and 4.8 μM, respectively, whereas ligand 3 exhibited a ~7-fold higher inhibition of the MBNL1–CCUG interaction compared to the MBNL1–CUG interaction. The ability of ligand 3 to destabilize complexes formed between MBNL1 and c\textit{TNT} (RNA I, Figure 6) was also evaluated. c\textit{TNT} is the 18-nt fragment (19) of the \textit{hcTNT} pre-mRNA, which is a natural target of MBNL1. Ligand 3 only weakly inhibited the MBNL1–c\textit{TNT} interaction. Table 2 summarizes the inhibition studies of ligands 2 and 3.

Finally, the ability of simple triaminopyrimidine (ligands 4 and 5) and triaminotriazine (ligands 6 and 7) recognition wedges to inhibit MBNL1 binding was examined (Figure 8). Consistent with our previous results for DM1, wedges without tethered intercalators did not destabilize the MBNL1–CCUG complexes (Figure 8).

**DISCUSSION**

The multisystemic clinical features of DM2 resemble those of DM1, although patients with DM2 do not suffer from the severe congenital form of disorder that occurs in DM1 (21). The prevalence of DM2 can be as high as in DM1 (1 in 8000) depending on the ethnicity of the population (22). As noted in the introduction, the phenotype of

![Figure 7](image-url)

**Table 2.** Summary of inhibition studies of various RNAs with ligands 2 and 3

<table>
<thead>
<tr>
<th>RNA</th>
<th>tRNA (nM)</th>
<th>$K_i$ (nM)</th>
<th>[MBNL1] (nM)</th>
<th>IC\textsubscript{50} (μM)</th>
<th>$K_i$ (μM)</th>
<th>IC\textsubscript{50} (μM)</th>
<th>$K_i$ (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CCUG)$_6$</td>
<td>100</td>
<td>15 ± 2</td>
<td>350</td>
<td>59 ± 5</td>
<td>2.5 ± 0.2</td>
<td>52 ± 8</td>
<td>2.2 ± 0.3</td>
</tr>
<tr>
<td>(CUG)$_{12}$</td>
<td>150 ± 20</td>
<td>1000</td>
<td>37 ± 11</td>
<td>3.4 ± 0.9</td>
<td>35 ± 3</td>
<td>3.0 ± 0.3</td>
<td></td>
</tr>
<tr>
<td>c\textit{TNT}</td>
<td>1.0 ± 0.2</td>
<td>15</td>
<td>101 ± 26</td>
<td>6.7 ± 1.6</td>
<td>&gt;250</td>
<td>&gt;17</td>
<td></td>
</tr>
</tbody>
</table>

*RNA concentrations were 0.1 nM.

*18-nt fragment of the c\textit{TNT} pre-mRNA (RNA I in Figure 6).
the use of ligand containing CTG repeats, there was a definite negative cooperativity for adjacent sites. In contrast, ligand 1 appears by ITC fitting, CD titration studies and Job analysis to bind the CCTG sequence in 1:3 stoichiometry. Although care must be taken not to overinterpret the $K_d$ values obtained from ITC fitting, the tight binding reflected in the $K_d(2)$ value may indicate a conformational change with formation of a 1:1 complex. Thus, the positive cooperativity could result from the first ligand inducing formation of two C-bulges and a central T–T mismatch, the latter site providing the same tight binding observed at CTG sites, whereas the C-bulge recognition would be comparatively weaker (25). This C-bulge and T–T mismatch structure, which is suggested to be a potential drug target, is recently observed by NMR spectroscopy for CCTG repeats (26).

The strong binding of ligand 1 for CCTG-containing DNA did not translate into strong binding for an RNA oligonucleotide containing a CCUG repeat (Ligand 1 binds CTG and CUG with similar affinity; see ref. 12). Furthermore, no inhibition of MBNL1 binding to (CCUG)$_b$ was observed. Although the detailed structure of the CCUG repeat is currently unknown, it has been proposed to adopt an A-form helical conformation, similar to that of the CUG repeat (6,27). At this time it is not known why ligand 1 is unable to recognize strongly the CCUG repeat. Nonetheless, the fact that ligand 1 is capable of strong and selective CUG binding in RNA and is able to inhibit binding of MBNL1 suggested that it should be possible to discover and develop ligands exhibiting similar behavior with CCUG sequences. This was the impetus for exploring ligands 2 and 3.

To improve the ability of ligand 1 to bind to CUG RNA, structurally similar ligands such as 2 were pursued. Ligand 2 is identical to 1, but with a triminopyrimidine moiety that forms a C-linked Janus-wedge. It was an appealing recognition unit to explore because 5-(4-amino-butyl)-2,4,6-triaminopyrimidine (17) was a known compound, allowing 2 to be easily synthesized by the same route as used for ligand 1. Unfortunately and somewhat surprisingly, ligand 2 exhibited very limited water solubility compared to the structurally similar ligand 1. The limited solubility of 2 prevented it from being studied by

Figure 8. (a) Chemical structures of wedge molecules. (b) Gel electrophoretic mobility shift assay showing that only π-stacked intercalators inhibit the MBNL1–CCUG complex. Control lane 1 (C1): RNA only. Control lane 2 (C2): RNA + MBNL1. Control lane 3 (C3): RNA + MBNL1 with 10% DMSO. Except when otherwise noted, all lanes contain 350 nM MBNL1, 100 μM ligands and 10% DMSO.
ITC with <10% DMSO. Ligand 3 was prepared with a more water-soluble intercalator, N,N′,N′′-[3-(dimethylamino)ethyl]acridine-4-carboxamide (DACA) in place of the 9-aminoo-6-chloro-2-methoxoacridine unit and this allowed the triaminopyrimidine motif to be examined.

Ligand 3 binds to (CCUG)m and RNA A with similar affinity and stoichiometry. Specifically, both sequences fit a binding model with a single, reasonably tight binding site (3–5 μM) and a second, significantly weaker binding site (250–480 μM). The tighter binding is presumably associated with one of the two CCUG sites in (CCUG)m and the single CCUG site in RNA A (Supplementary Data). Potential weaker sites in (CCUG)m include the remaining CCUG site, (CCUG)2, loop region and GC stem regions. In fact, titration with RNA A, lacking the second CCUG site and loop, give an ITC curve with better curvature. This suggests the residual heat at the end of the titration for (CCUG)m likely arises from a general, non-specific binding of the weaker sites. Indeed, the affinity is similar to that observed for ligand 3 binding to the single-stranded hCTN% pre-mRNA (RNA I). Overall, the data are consistent with ligand 3 binding to only one of the two C–U mismatches and a single CCUG site when two sites are neighboring. The origin of this latter effect is unclear.

Similar to the preference exhibited by MBNL1 (19), ligand 3 binds with ~40-fold higher affinity to the slipped-CCUG structure (RNA B) than to the duplex containing adjacent C–U mismatches (RNA A). These results are in good agreement with the Kd values for the duplexes containing the single U–U (RNA D) and C–C mismatches (RNA E), suggesting that ligand 3 binds strongly to the C–C mismatch in the slipped structure and to the U–U with a reduced affinity.

Surprisingly, ligand 3 does not bind strongly (Kd ~ 85 ± 30 μM) to RNA C, which contains a single C–U step with two adjacent CG steps. Ligand 3 also binds weakly (Kd ~ 100 μM) to the purine-purine mismatches in duplexes F and G. The binding of ligand 3 to DNA containing a CTG sequence (DNA H), which is one of the targets of DM1 (20), occurs with low micromolar affinity (Kd = 1.4 ± 1.5 μM). The difference in binding affinity of ligand 3 toward analogous RNA and DNA duplexes (RNA D, U–U; Kd = 25 μM) and (DNA H, T–T; Kd = 1.4 μM) may result from structural differences between B-form DNA and A-form RNA.

Importantly, ligand 3 not only binds CCUG, it inhibits MBNL1 binding to CCUG with low micromolar Kd values. Together with the ITC results, these data suggest that ligand 3 is a good lead ligand for targeting either the CCUG or CCUG-slipped repeat structures implicated as the causative agent of DM2. Finding agents that complex CCUG and inhibit MBNL1 binding is a challenge, but finding agents that are highly selective is perhaps even more difficult and likely more important given the difficulty in developing highly selective agents targeted to nucleic acids. The selectivity profile of ligand 3 has been examined in several different contexts and can be found in Tables 1 and 2. For example, it is notable that ligand 3 binds (CCUG)m >600-fold tighter than it does tRNA. Thus, even in the presence of tRNA, ligand 3 is capable of disrupting the MBNL1–CCUG interaction with minimally diminished Kd values. It is also significant that ligand 3 does not inhibit the interaction between MBNL1 and its natural target, the 18-nt fragment of hCTN% pre-mRNA (IC50 > 250 μM).

Developing a systematic structure-activity relationship is beyond the scope of the current study, but the data from the gel electrophoretic mobility shift assays using simple wedge recognition compounds 4–7 indicate that the acridine unit is critically important. These results agree with the previously proposed π-stacked intercalator model wherein the acridine is responsible for affinity whereas the selectivity is dictated by the nature of the wedge motif. In this regard, it is intriguing that the small structural difference between triaminotriazine and triaminopyrimidine units leads to the preferential binding to CCUG for the former and to CCUG for the latter. It is likely that the basicity difference between triaminotriazine (pKa ~ 5) (28) and triaminopyrimidine (pKa ~ 6.7) (29) plays an important role and this is an aspect that merits additional study.

Finally, the broadest finding from the current study is that ligands 2 and 3 exhibit among the highest documented selectivity for CCUG and, thus, represent an excellent starting point for the development more selective and potent inhibitors of MBNL1 binding. Because of the presence of multiple repeating units in the toxic RNAs that cause DM1 and DM2, an obvious approach involves oligomerization of these or analogous ligands (11,14). This is an approach that has already proven effective. For example, Disney reported that the oligomerization of a DNA dye (Hoechst 33258) and kanamycin A, ligands that exhibit only moderate selectivity for the CUG and CCUG sequence, respectively, result in oligomers of better affinity and selectivity (11,14,15). The use of intercalators presents an inherent limitation on selectivity because an unstacked intercalator can non-selectively complex a wide range of duplex DNAs and RNAs. Developing new ligands where the recognition wedge remains stacked or in close proximity to the intercalator, or where the intercalator is replaced by a different unit that provides affinity, may obviate this potential source of poor selectivity. These and other approaches to optimize this class of RNA binding agents are under active investigation and will be reported in due course.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**FUNDING**

National Institutes of Health (RO1AR058361). C.H.W. is a Croucher Foundation Scholar and an Illinois Department of Chemistry Ullyot Fellow. Y.F. is a recipient of a Department of Chemistry Fellowship. Funding for open access charge: National Institutes of Health (RO1AR058361).

Conflict of interest statement. None declared.
REFERENCES


Investigating the Binding Mode of an Inhibitor of the MBNL1-RNA Complex in Myotonic Dystrophy Type 1 (DM1) Leads to the Unexpected Discovery of a DNA-Selective Binder


The development of small molecules that recognize specific DNA and RNA sequences or structures remains a critical challenge. If cell-permeable, such agents might allow the regulation of genes for therapeutic purposes or the targeting of DNA or RNA sites known to play some other roles in causing disease. A few general approaches to the recognition of double-helical DNA,[1] base mismatches,[2] G-quartets,[3] and other targets[4] have emerged over the past two decades. However, many difficulties remain. For example, it is often difficult to obtain high selectivity for DNA over RNA or vice versa.[5] We reported that ligand 1 binds CUG and CTG sites with similar high affinity and selectivity in RNA and DNA oligonucleotides, respectively.[6] Herein we show that ligands 2 and 3, with one or two N-methyl groups, respectively, selectively abolish RNA binding, with 3 giving a more than threefold increase in DNA affinity that could allow DNA-targeted gene therapy for myotonic dystrophy type 1 (DM1).[7]

This work arose from the well-supported hypothesis that DM1 originates in the sequestration of the alternate splicing regulator protein, muscleblind-like 1 (MBNL1), by abnormally long CUG triplet repeats. It has been suggested that ligands able to bind selectively to these pathogenic triplet repeats might reverse the DM1 phenotype by inhibiting the MBNL1-RNA complex, allowing the splicing regulator to resume its normal function.[8] Much less attention has focused on agents that might shorten or halt the expansion of the CTG repeats found in the DMPK gene but this DNA sequence has been identified as a potential target for molecular therapy.[9] Thus, we developed ligand 1,[6] which selectively binds CTG and CUG sites and inhibits the MBNL1-CUG complex.[10]

The design of ligand 1 was based on the X-ray crystal structure of r(CUG)₆ determined by Berglund,[11] which showed the CUG repeat in a standard A-form helical structure with the U/C0U mismatch flanked by G/C0C pairs. The triaminotriazine unit was selected as a Janus wedge[12–14] to form a base triplet with U/C0U (Figure 1 A). Critical to the design was the assumption that the two heterocycles would p-stack, reducing the nonspecific intercalation (Figure 1 B). Insertions from the minor

Figure 1. A) Ligand 1 contains a triaminotriazine unit designed to selectively target T/C0T and U/C0U mismatches. B) Proposed binding mode of ligand 1 to CUG (or CTG) repeats.

Figure 2. Three potential modes of binding: A) Minor groove triplet binding. B) Major groove triplet binding. C) A stretched U–U wobble pair with a bridging water molecule. D) Minor groove stretched wobble binding. R=CH₃ (DNA) or H (RNA).
major groove were considered, but the former was in line with the intercalation preference of 9-aminoacridine.\[15\]

The data collected for ligand 1 was consistent with the binding model shown in Figures 1 and 2A. However, a de-twinned CUG repeat structure with high-resolution hydration details\[16\] led us to consider other possible binding modes. In particular, there was concern that the minor groove base-triplet model (Figure 2A) required the lengthening of the C1’/C0-C1’ distance of the U/C0-U from 10.4 to 13.8 Å. Although the RNA minor groove is wide and shallow, 3.4 Å is a significant increase in the C1’/C0-C1’ distance (Figure 2A) and may signal a distortion in the DNA or RNA backbone that is energetically unfavorable.\[17\] This distance is even longer than that of a G/C0-A purine-purine

Table 1. Dissociation constants (μM) of ligands 1-6 with r(CUG)\(_2\) and d(CTG)\(_2\) containing U-U and T-T mismatches, respectively.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>r(CUG)(_2)</th>
<th>d(CTG)(_2)</th>
<th>hsDNA(^{\text{[b]}})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1 ± 0.2</td>
<td>0.39 ± 0.08</td>
<td>55 ± 16</td>
</tr>
<tr>
<td>2</td>
<td>n.b.(^{\text{[a]}})</td>
<td>0.8 ± 0.4</td>
<td>n.b.(^{\text{[a]}})</td>
</tr>
<tr>
<td>3</td>
<td>n.b.(^{\text{[a]}})</td>
<td>0.12 ± 0.04</td>
<td>n.b.(^{\text{[a]}})</td>
</tr>
<tr>
<td>4</td>
<td>14 ± 5</td>
<td>2.2 ± 0.5</td>
<td>n.b.(^{\text{[a]}})</td>
</tr>
<tr>
<td>5</td>
<td>n.b.(^{\text{[a]}})</td>
<td>50 ± 30</td>
<td>n.b.(^{\text{[a]}})</td>
</tr>
<tr>
<td>6</td>
<td>n.b.(^{\text{[a]}})</td>
<td>n.b.(^{\text{[a]}})</td>
<td>n.b.(^{\text{[a]}})</td>
</tr>
</tbody>
</table>

\(^{[a]}\) n.b.: no measurable binding and a lower limit of \(K_d > 200 \mu M\) is assigned. \(^{[b]}\) Herring sperm DNA.
pair (12.5 $\text{Al}$)$^\text{[19]}$. In fact, repeated attempts to model r(CUG)$_2$-T with the ligand in the minor groove failed to yield a stable complex.

Modeling showed that the water molecule bridging the two uracil bases in a stretched wobble pair (Figure 2C) could be replaced by the triazine unit in the minor groove (Figure 2D).$^\text{[19]}$ This led to a smaller structural reorganization of the original pair (Figure 2D). Further, modeling of a major groove base triplet suggested a C$^\text{1'}$–C$^\text{1'}$ distance even closer to that in the X-ray structure (Figure 2B). In absence of direct structural information, an extended "methyl scanning" approach$^\text{[20]}$ appeared to be ideal to distinguish between models in Figure 2A and 2D. Thus, N-methylated analogues 2–6 were synthesized and tested using isothermal titration calorimetry (ITC) with both d(CTG)$_2$ and r(CUG)$_2$ (Figure 3). The ability to inhibit the MBNL1·CUG interaction was examined by electrophoretic mobility shift assays (EMSA). These methods were supplemented with molecular dynamics (MD) simulations.

The methyl scanning method was possible because the number of hydrogen bonds differ significantly in the base triplet (Figure 2D). In particular, the wobble binding mode has two, possibly three, free N–H groups, whereas only a single N–H group is free in the triplet binding modes.

The possibility of base-triplet recognition through the major groove was also considered, because simple modeling indicated the C$^\text{1'}$–C$^\text{1'}$ distance to be minimally adjusted from 10.4 to 10.1 Å (Figure 2B). Indeed, an unconstrained 10 ns MD simulation showed ligand 1 could fit satisfactorily in the major groove (Figure 4A and Video S1). The MD simulation of the minor groove stretched wobble binding was also performed (Figure 4B and Video S2). Although fewer hydrogen bonds are formed, the unconstrained simulation led to a stable structure. The average C$^\text{1'}$–C$^\text{1'}$ distance of the U–U mismatch is moderately lengthened by 0.4 Å (11.8 vs. 10.4 Å, see Figure 5).

The d(CTG)$_2$ recognition by 1–6 was measured by ITC and all $K_\text{d}$ values are collected in Table 1. Ligand 3 was found to bind more than threefold stronger than ligand 1 ($K_\text{d} = 0.12 \text{µM}$ for 3 vs. 0.39 µM for 1), whereas ligands 2, 4, and 5 each bound d(CTG)$_2$ but progressively more weakly than 1.$^\text{[20]}$ Ligand 6, carrying three methyl groups, showed no measurable binding to d(CTG)$_2$. Because of the limited aqueous solubility, we were only able to assign a lower limit of $K_\text{d} > 200 \text{µM}$ for the combined ITC and modeling results are consistent with Watson–Crick-type recognition of the stretched wobble pair (Figure 2C). An unconstrained 10 ns MD simulation further suggested the weakly bound thymine could be flipped-out from the dsDNA (Video S3 and Figure S3). In contrast to the results with d(CTG)$_2$, ligand 3 did not bind under similar conditions to r(CUG)$_2$ (Figure 6, Table 1) and, indeed, the single N-methyl group in 2 was enough to eliminate ligand binding. The only CUG-binding, N-methylated ligand was 4, and even its $K_\text{d}$ was seven times weaker than that of 1. All methylated ligands also showed reduced affinity towards herring sperm DNA (hsDNA). The ability of the recognition unit to stack on the intercalator is enhanced by methylation which in turn reduced the nonspecific intercalative binding to duplex DNA.$^\text{[21]}$

Figure 6. ITC profiles for complexation of ligand 3 with duplex containing the A) T–T and B) U–U mismatches.

Figure 7. EMSA screening of ligands 1–6 showing only ligands 1 and 4 inhibit MBNL1·CUG interaction. [r(CUG)$_2$]$_0$ = 0.2 nM, [MBNL1] = 460 nM, (ligand) = 100 µM, 10% DMSO (except lane 2), Tris borate (pH 8.3). The structure of r(CUG)$_2$ used in the assay is shown on the right.
To assess the potential of these ligands for drug development, their binding affinity to various DNA and RNA duplexes was studied by ITC (Table S1). All methylated ligands showed reduced affinity toward the duplexes, including hsdDNA. The ability to complex other mismatches was also measured. In general, these ligands exhibited high selectivity (up to 1600-fold) for DNA over RNA mismatches. The ligands were also found to bind weakly to purine-purine mismatches. Interestingly, all the methylated ligands bind quite strongly ([Kd] = 0.13–2.0 μM) to d(CCG)n, a trinucleotide repeat sequence associated with Fragile X syndrome and chronic lymphocytic leukemia, thus suggesting a potential molecular therapeutic approach to these diseases that warrants additional study.

The ability of ligands 1–6 to inhibit MBLN1 binding to r(CUG)n was measured by EMSA, and the results paralleled the ITC data. Thus, only ligands 1 and 4 showed inhibition (Figure 7). Although nothing can be said about the groove preference from these experiments, the MD simulations favor major groove triplet formation (see above), and it is known that loop and mismatch structures can enhance the accessibility of the major groove.[19]

The combined experimental and computational approach described herein suggests that this class of ligands bind DNA and RNA by significantly different modes. Thus, it is proposed that ligands 1 and 4 recognize the U–U mismatch in RNA through formation of a major groove base triplet (Figure 8 A) whereas ligands 1–5 bind the T–T mismatch in DNA through the stretched wobble pair (Figure 8 B). These binding modes also explain the decreased binding to U–U by 4 ([Kd] = 14 μM vs. 2.1 μM for 1) and to T–T by 5 ([Kd] = 50 μM for 5 vs. 0.39 μM for 1). Thus, the C–NH(Me) bond rotation[20] leads to unfavorable binding for one or more of the rotamers.[21]

In conclusion, we have used the methyl scanning method combined with MD simulations to indirectly investigate the possible binding modes by which ligands 1–5 recognize CUG and CTG sites in RNA and DNA, respectively. Beyond informing on the binding mode, the ability to substitute the amino groups of ligand 1 suggests these as sites for further modifications that might enhance the selectivity and efficacy of these lead compounds in treating DM1.

More significant was the unexpected discovery of DNA-selective ligands for CTG. Whereas ligand 1 binds both CUG and CTG with similar strength, ligand 3 showed a more than threefold increase in affinity to CTG with negligible binding to RNA. It was also found that these ligands showed reduced binding toward DNA and RNA duplexes (i.e., without mismatches) upon methylation, potentially opening an avenue for a DNA-targeted molecular therapy of DM1.[20] Competitive binding to the corresponding CUG transcript would be expected to reduce the effectiveness of this approach and could also complement efforts to assess the DNA-targeted approach. More broadly, this study increases our knowledge of how small molecules can selectively recognize nucleic acids.

Acknowledgements
This work was supported by the National Institutes of Health (RO1AR058361). C.H.W. is a recipient of a Croucher Foundation Scholarship (HK) and a Pines Graduate Fellowship.

Keywords: DNA · drug design · inhibitors · recognition · RNA
Replacing the triazine by pyrimidine unit led to a change in mismatch selectivity, presumably due to the difference in $pK_a$ values, see C.-H. Wong, Y. Fu, S. R. Ramisetty, A. M. Baranger, S. C. Zimmerman, Nucleic Acids Res. 2011, 39, 8881 – 8890.

Repeating the triazine by pyrimidine unit led to a change in mismatch selectivity, presumably due to the difference in $pK_a$ values, see C.-H. Wong, Y. Fu, S. R. Ramisetty, A. M. Baranger, S. C. Zimmerman, Nucleic Acids Res. 2011, 39, 8881 – 8890.


[19] This type of singly hydrogen-bonded U-U mismatches is also observed in a recent NMR study, see R. Parkesh, M. Fountain, M. D. Disney, Biochemistry 2011, 50, 599 – 601.


[21] The result here is different from the observation in ref. [13d] that a N,N-dimethylated triazine peptide inhibited the triplex formation.


[26] Rotamers were observed for those ligands in both $^1$H and $^{13}$C NMR spectroscopy (see the Supporting Information).


Received: September 18, 2012
Published online on October 24, 2012
FEATURE ARTICLE
Chun-Ho Wong and Steven C. Zimmerman
Orthogonality in organic, polymer, and supramolecular chemistry: from Merrifield to click chemistry

Orthogonality in organic, polymer, and supramolecular chemistry: from Merrifield to click chemistry

Chun-Ho Wong and Steven C. Zimmerman*

The concept of orthogonality has been applied to many areas of chemistry, ranging from wave functions to chromatography. But it was Barany and Merrifield’s orthogonal protecting group strategy that paved the way for solid phase peptide syntheses, other important classes of biomaterials such as oligosaccharides and oligonucleotides, and ultimately to a term in widespread usage that is focused on chemical reactivity and binding selectivity. The orthogonal protection strategy has been extended to the development of orthogonal activation, and recently the click reaction, for streamlining organic synthesis. The click reaction and its variants are considered orthogonal as the components react together in high yield and in the presence of many other functional groups. Likewise, supramolecular building blocks can also be orthogonal, thereby enabling programmed self-assembly, a superb strategy to create complex architectures. Overall, orthogonal reactions and supramolecular interactions have dramatically improved the syntheses, the preparation of functional materials, and the self-assembly of nanoscale structures.

1. Introduction

As with many terms in chemistry, the term orthogonality, as applied to protecting groups, began with a very precise definition in a limited setting. Thus, in the context of his revolutionary advance in peptide synthesis, Merrifield first reported the term in 1977 as a protecting group removal strategy. It was a simple but powerful concept that one of multiple protecting groups could be removed in the presence of all others by using a cleavage reaction with a different mechanism.

Since then the usage of the term orthogonality has broadened dramatically. Indeed, the term is now used in supramolecular chemistry for systems where two non-covalent interactions occur with no crosstalk as well as in accelerated...
polymer and dendrimer syntheses. Recently, the widely-used, copper-catalyzed alkyne–azide coupling (CuAAC) has been described as a highly orthogonal reaction and in biological contexts as a bioorthogonal conjugation method. Although the click reaction by itself does not fit the classical Barany and Merrifield’s definition of orthogonality (vide infra), its use in this context is pervasive and the clearest indication that as a chemical term orthogonality has achieved polysemic status.²

This feature article traces the history of orthogonal chemistry and summarizes its current application in several areas of chemistry with a focus on levels of orthogonal complexity.

2. Chemoselectivity

...“The ability to discriminate among the reactive sites is referred to as chemoselectivity.”... Trost (1983)³

Although orthogonality predates chemoselectivity, it is helpful to discuss the latter term first. Chemoselectivity was coined by Trost in 1983 to describe the ability of a chemical reagent to discriminate among reactive sites.⁴ For example, in compound 1 the reaction with the sodium salt of dimethyl malonate in DMF occurs exclusively with displacement of the bromide to give 2 (Fig. 1).³⁴ But in the presence of tetrakis(triphenylphosphine)-palladium(0), the allylic acetate is activated and undergoes displacement to 3 with no formation of 2. The high chemoselectivity of these two reactions originates from their quite different mechanisms.

Chemoselectivity may also be achieved with a single mechanism and two structurally similar sites, provided that there is precise control of the reaction conditions. An example would be two acetal groups that cleave in acid, but have graduated reactivity; so increasing acidity entirely removes one and then the other. This strategy commonly employed in protecting group chemistry is referred to as the modulated lability approach (Fig. 2a).⁵ This approach takes advantage of the reactivity difference of functional groups A and B toward a mild and more forcing reagent in the same type of reaction. Indeed, the early peptide synthesis developed by Merrifield used the 1000 : 1 reactivity difference of the tert-butyloxycarbonyl (Boc) and benzyl groups toward TFA to allow differential deprotection of the growing peptide at the amino terminus and without deprotecting the peptide side chains.⁶ The use of a stronger acid (e.g., HF) removes both groups and cleaves the peptide from the resin. In this and the previous example, it is not possible to reverse the order of the deprotection steps. Thus, the sequence of the reaction is controlled by the different reactivity of the two functional groups.

3. Orthogonal protecting group chemistry

...“An orthogonal system is defined as a set of completely independent classes of protecting groups. In a system of this kind, each class of groups can be removed in any order and in the presence of all other classes.”... Barany and Merrifield (1977)³⁸

3.1 Orthogonality vs. chemoselectivity

The orthogonal reactions developed by Barany and Merrifield involve two protecting group removal steps that occur with high chemoselectivity (Fig. 2b). In practice, each of the deprotection reactions occurs by a different mechanism designed so that each is able to effect removal of one protecting group exclusively. Thus, each reaction is chemoselective so each may occur in any order. As shown in Fig. 2b, the use of either sequence affords the same product. Orthogonality can be considered as a subset of chemoselectivity. However, the terms “orthogonal” and “chemoselective” are often used interchangeably in the literature. In addition, terms such as partial or quasi-orthogonal are occasionally seen, sometimes referring to partial chemoselectivity.

3.2 Increasing complexity: multiple dimensions

The discovery of the 9-fluorenylmethoxycarbonyl (Fmoc) group by Carpano meant that for the first time an amino protecting group could be removed under mildly basic conditions.⁷ Because the conditions are orthogonal to the acidic tert-Bu removal and high yielding, the Fmoc–t-Bu strategy (orthogonality-based; Fig. 3) has largely replaced Merrifield’s classic peptide
synthesis using the Boc–benzyl strategy (modulated lability-based). The Fmoc and t-Bu groups represent an excellent pair of orthogonal protecting groups.

The number of “dimensions” of an orthogonal system can be increased. For example, in his Nobel Lecture, Merrifield described a pentapeptide containing three orthogonal protecting groups (Fig. 4). The peptide is linked to the resin by an o-nitrobenzyl (ONb) group and protected by two additional functional groups (t-Bu and Dts) that can each be cleaved without affecting the other.

Another way in which the complexity of a set of orthogonal reactions can be increased is by conducting their operation simultaneously in a single flask (Fig. 2c). This scenario represents a more stringent example of chemical orthogonality because beyond the reagents being selective for their respective functionality they must not affect one another. This type of transformation, illustrated in Fig. 2c, can offer higher efficiency, with reduced time and waste; allowing us to envision a multistep synthesis where all the reagents are added at once. The simultaneous, multi-functionalization of macromolecules and the dual labeling of cellular targets for imaging are already possible and will be discussed below (vide infra).

Higher dimensional orthogonal systems are possible. Indeed, some synthetic targets benefit from multiple orthogonal protecting group strategies. For example, the challenges in oligosaccharide synthesis are significant because the many hydroxyl groups have similar reactivity. A wider range of orthogonal protecting groups is now available and this enabled Wong and co-workers to report the synthesis of a library of oligosaccharides containing 38 416 members. The four dimensional system consisted of chloroacetyl, p-methoxybenzyl, levulinyl, and t-butyldiphenylsilyl protecting groups that were selectively deprotected using four different mechanisms: basic hydrolysis, acidic hydrolysis, hydrazinolysis, and fluoride-based deprotection, respectively (Fig. 5).

Boons and co-workers reported another set of orthogonal protecting groups for oligosaccharides that are chemoselectively removed using DDQ oxidation, Pd-catalysis, hydrazinolysis, and fluoride-based desilylation. It was also found that the selected protecting group pattern on the mannose rings is important for obtaining high yields and selective β-mannosylation.

Reviews on protecting group strategies and the influence on stereoselectivity in oligosaccharide synthesis are available.

### 3.3 Chromatic orthogonality

“Orthogonality is defined as the possibility of making one functional group react selectively in the presence of others under specific conditions. A differentiation based on the color of a light beam could be named chromatic orthogonality.”

Bochet (2004)

A powerful approach to spatiotemporal orthogonal deprotection avoids reagent-based reactions, using photolytic cleavage reactions instead to achieve the desired chemoselectivity. Deprotection can be performed using a wide variety of photolabile groups (e.g., o-nitrobenzyl, p-hydroxyphenacyl, benzophenone) in the presence of other reagent-based protecting groups. In general, photolabile groups form an orthogonal system with a wide range of chemical-based protecting groups.
(e.g., silyl and allyl groups) because light can be absorbed with considerable selectivity. Indeed, many functional groups do not absorb light from common laboratory light sources and many that do are not photoreactive. It is further possible to create a set of photolabile groups that each react with light of a different wavelength—a type of orthogonality referred to as chromatic orthogonality.\footnote{Baran’s seminal report of a protecting-group-free synthesis of hapalindole Q in 2004.} For example, the 3',3'-dimethoxybenzoin and 2-nitroveratryl esters shown in Fig. 6a are photodeprotected at 254 and 420 nm, respectively. The approach was later refined with a pair of α-nitrobenzyl-based photolabile groups that can be photo-deprotected orthogonally by fine-tuning using both substituent- and isotopic effects on quantum yields (Fig. 6b).\footnote{The use of orthogonal protecting group strategies is now commonplace in the synthesis of a range of complex molecules. These include natural products and biomolecules such as oligosaccharides, glycoproteins, and nucleic acids. The high level of dimensionality is most evident in Kocienski’s classification of a large number of protecting groups into 13 orthogonal sets; members of each set are removable by a unique mechanism.\footnote{Yu and co-workers combined the Horner–Wadsworth–Emmons and Heck reactions as two orthogonal coupling reactions so that the linkages prepared are all vinylene-type.}}

The use of orthogonal protecting group strategies is now commonplace in the synthesis of a range of complex molecules. Orthogonal glycosylation developed by Ogawa and co-workers.\footnote{The second reported example of an orthogonal coupling approach to prepare a complex molecule came in the context of polymer synthesis. Our group disclosed the first orthogonal dendrimer synthesis featuring two AB\textsubscript{2}-type monomers (i.e., AB\textsubscript{i} and CD\textsubscript{j}) that couple in a convergent approach utilizing alternating Mitsunobu esterification and Sonogashira coupling steps. Because these two reactions are orthogonal, deprotection or activation steps could be avoided for the first time. Thus, each coupling reaction added one layer to the dendrimer. The approach was extended with the use of AB\textsubscript{i} and CD\textsubscript{j} monomers (Frechet’s branched-monomer approach), so that each orthogonal coupling reaction adds two layers to the dendrimer (Fig. 9). This remains the fastest dendrimer synthesis to date in terms of number of steps from the monomer units. Indeed, the MW of 4 and 5 are just 392 and 416, respectively, but they combine in three synthetic steps to produce a sixth generation dendrimers with a MW = 20 896. This is a rate of growth (MW per step) not even observed in nature. By using the Mitsunobu esterification and Sonogashira coupling reactions, the layers of the dendrimer alternate. Similarly, Majoral and co-workers reported an orthogonal dendrimer synthesis in a divergent approach using alternating hydrazone formation and Staudinger reaction. To produce a homogeneous dendrimer, where the layers are chemically identical, Yu and co-workers combined the Horner–Wadsworth–Emmons and Heck reactions as two orthogonal coupling reactions so that the linkages prepared are all vinylene-type.} The potential of this approach was made particularly apparent in Baran’s seminal report of a protecting-group-free synthesis of hapalindole Q in 2004.\footnote{Although the orthogonality aspect of click reactions will be discussed in Section 4.5, it should be noted here that click approach has also been adopted in solid phase oligosaccharide syntheses and in the production of a combinatorial library of oligosaccharides.}

The chosen set of reactions is therefore shown to be orthogonal.\footnote{Zeng and Zimmerman (1996)\footnote{The use of orthogonal coupling reactions, the layers of the dendrimer alternate. Similarly, Majoral and co-workers reported an orthogonal dendrimer synthesis in a divergent approach using alternating hydrazone formation and Staudinger reaction. To produce a homogeneous dendrimer, where the layers are chemically identical, Yu and co-workers combined the Horner–Wadsworth–Emmons and Heck reactions as two orthogonal coupling reactions so that the linkages prepared are all vinylene-type.} “It was anticipated that both pairs of functional groups and their resulting coupling products would be inert to the conditions of the other coupling reaction, orthogonality that would allow 4 and 5 to be employed consecutively in either order.”}

To our knowledge, the earliest example of an orthogonal coupling strategy is the orthogonal glycosylation reported by Ogawa and co-workers in 1994 (Fig. 7).\footnote{This approach was later refined with a pair of α-nitrobenzyl-based photolabile groups that can be photo-deprotected orthogonally by fine-tuning using both substituent- and isotopic effects on quantum yields (Fig. 6b).} In this approach, two distinct anemic groups (phenethylthio and fluoride) can be orthogradly activated without affecting one another. This
chemistry is emerging as an efficient method to construct dendritic structures. Hawker et al. reported several approaches involving the use of CuAAC and thiol–ene reactions coupled to "non-click" reactions to construct dendritic structures. They also reported the use of two orthogonal click reactions (CuAAC and thiol–ene reactions) to accelerate the dendrimer synthesis such that a G6 dendrimer can be prepared in a single day (Fig. 10). Monteiro demonstrated the use of two orthogonal click reactions (CuAAC and nitroxide radical coupling (NRC)) to prepare dendrimers in either convergent or divergent approaches. Under appropriate conditions, it is possible to synthesize G3 dendrimers with a variety of building blocks in a one-pot setting. Other orthogonal click syntheses of dendrimers are also reported.

4.3 Chemical ligation of peptides

Although peptides containing up to 100 amino acids are now synthesized routinely, larger peptides and proteins (i.e., >150 amino acids) cannot be readily prepared in high yield and purity because the peptide coupling efficiency decreases with sequence length. Chemical ligation provides a convenient method to link two peptide segments covalently. In particular, the native chemical ligation has been widely adopted because it produces a natural peptide bond (Fig. 11). The method uses a Cys–thioester ligation with an entropically favored intramolecular transacylation. At pH 7–8, the thioester is reactive toward thiols, not hydroxyl nucleophiles. Internal cysteine residues can reversibly react with the thioester but this reaction is unproductive. As a result, the C-terminal thioester
will permanently connect only to the N-terminal cysteine, forming a native peptide linkage. The reaction therefore is highly chemoselective. As with other highly chemoselective reactions (see click reactions; Section 4.5), the chemical ligation of peptides has been termed orthogonal ligation.36

4.4 Orthogonal catalysis

The concept of orthogonal reactivity can also be applied to catalysis and catalytic cycles. For example, Buchwald and co-workers developed Pd- and Cu-based catalysts that effect chemoselective C- or N-arylation reactions, respectively (Fig. 12).39 Likewise, switching metal catalysts from Pd- to Cu-based allowed the selective N- and O-arylation, respectively, of various aminophenols.40 The site of arylation of aminoalcohols can also be controlled by the choice of ligands with high selectivities from 15:1 to >50:1.41 In each of these systems the two chemoselective transformations are expected to be orthogonal even if orthogonality was not demonstrated experimentally.

In orthogonal catalysis there are also different levels of complexity from the sequential catalytic transformations just described to systems where two or more catalytic cycles operate simultaneously. The latter arguably represents the most demanding of all orthogonal systems because each entire catalytic cycle must be fully orthogonal. Thus, the substrate and product and each reactive intermediate and catalytic species must be compatible unless the kinetics of the cycles are vastly different in which case a few of the species may not be present at the same time. Chung and co-workers reported a three-step, one-pot synthesis of fenestranes using Co nanoparticles and Pd(II) catalysts in which the two catalysts operate simultaneously (Fig. 13).42 The first step is carried out prior to the addition of the palladium catalyst. Nonetheless, this is an example of one-pot, orthogonal catalysis made possible by the mutual independence of the Pd(II) catalyst for the allylic alkylation in the second step and the Co catalyst for the last Pauson–Khand reaction. Fogg and dos Santos have termed such processes orthogonal tandem catalysis.43 Similarly, Hidai, Uemura and co-workers reported the use of ruthenium and platinum as orthogonal tandem catalysts to synthesize a series of substituted furans.44 Another impressive example is the synthesis of branched polyethylenes using three catalysts that operate simultaneously in a one-pot procedure (Fig. 14).45 Polyethylenes of different properties can be prepared simply by varying the mole ratios of the catalysts.

4.5 Click reactions, orthogonal click reactions, and bioorthogonality

The objective of this section is not to provide a comprehensive review of click chemistry but to focus on its relationship to orthogonality. In discussing click reactions generally, Sharpless noted that “although click reaction components are necessarily highly reactive, their chemoselectivity profiles are quite narrowly defined, that is, the reactions are “orthogonal” to an unusually broad range of reagents, solvents, and other functional groups”.46 Thus, he correctly noted the implicit orthogonality of click reactions as originating in the high reactivity of the coupling components with one another but broad tolerance of other reagents and functional groups. The prototypical click reaction, the copper-catalyzed alkyne–azide cycloaddition (CuAAC), illustrates these ideas well. It is a highly chemoselective reaction that proceeds in high yield under mild conditions47 yet the azide and alkyne groups are relatively unreactive.
The low reactivity profile of organic azides\(^{47,48}\) and terminal alkyne groups is illustrated in Fig. 15a and b, respectively. Thus, Lin and Walsh prepared a series of glycopeptide antibiotics by the chemoselective click reaction of various azido sugars with an alkyne-containing cyclic peptide (Tyc3PG; Fig. 15a).\(^{49}\) The azide group is carried through several steps prior to reaction with the alkyne group. Analogously, Gin reported the synthesis of a cyclodextrin analog by double click reaction of a trisaccharide carrying both click partners.\(^{50}\) In this example, the terminal alkyne is carried through six steps prior to the cyclization reaction (Fig. 15b).

### 4.5.1 Orthogonal functionalization of polymers using click chemistry

The orthogonality of the click reaction is well illustrated in the post-functionalization of polymers. Indeed, click methodology has the ability to create new polymers by introducing functionality at polymer chain ends or along the backbone.\(^{51}\) Preparation of alkyne- or azide-containing polymers and their subsequent click modifications will not be covered here; readers are directed to several recent reviews.\(^{52}\) This section focuses on orthogonal, simultaneous, one-pot post-polymerization modifications of polymers using click reactions. This approach minimizes the number of synthetic steps and reduces the number of work-up and purification operations.\(^{53,54}\)

Hawker reported the one-pot, simultaneous and cascade functionalization of polymers.\(^{55}\) Both strategies combined CuAAC and an esterification reaction as a highly orthogonal pair. For example, a simultaneous functionalization was performed on a water-soluble polymer containing acetylene- and hydroxyl-moieties for CuAAC and esterification, respectively (Fig. 16).\(^{55}\) NMR and IR analysis of the product polymer indicated >95% conversion for both reactions. The cascade-type functionalization occurs when a side-chain modification adds new functionality that can undergo a further transformation. The authors described the reaction of a polymer carrying active ester groups with propargylamine and an organic azide under CuAAC conditions. The amide formation reaction and alkyne click can occur in any order to furnish the cascade-functionalized polymer.

Lyon described a related approach to the formation of multi-responsive gels.\(^{56}\) Thus, EDC-based amidation and the CuAAC reaction were the orthogonal reactions used to link fluorescein and rhodamine dyes simultaneously to methacrylate-based microgels. Both FT-IR and epifluorescence microscopy demonstrated the synthesis of the desired materials.

Yang and Weck reported the one-pot, simultaneous polymer modification using two orthogonal click reactions — CuAAC and hydrazone formation.\(^{57}\) The polymer modification is highly modular with a wide range of substrates including those with biological interest (Fig. 17). Quantitative functionalization was observed by NMR and IR analysis.

Tunca and co-workers reported an analogous process using the CuAAC and Diels–Alder reaction.\(^{58}\) In this report, two end-functionalized polymers (maleimide–PMMA and acetylene–PEG) were attached to polystyrene polymers containing anthracene and azide moieties with >90% grafting efficiency based on NMR analysis. In an approach that has been reviewed recently, the combination of Diels–Alder and CuAAC has been used to prepare...
polymers of different architectures, including star polymers, cyclic polymers, and dendrimers.\textsuperscript{59} The one-pot, cascade functionalization of polystyrene using a combination of CuAAC and Diels–Alder reactions was also reported by Yagci and co-workers.\textsuperscript{60}

A final example of polymer-based click chemistry comes from Tunca who reported a combination of three orthogonal reactions (CuAAC, Diels–Alder, and NRC) in a one-pot preparation of linear tetrablock copolymers (Fig. 18).\textsuperscript{61} The isolated yields of the polymers were 50–55\%, with GPC and NMR analysis showing the \(M_n\) values of the isolated polymers to be in good agreement with the sum of the \(M_n\) of the reactant polymers.

### 4.5.2 Click chemistry beyond CuAAC.

As seen above, CuAAC is a powerful reaction with a wide range of applications. The high chemoselectivity of the CuAAC allows it to form a highly orthogonal reaction pair with many other reactions. Indeed, there are only a few reactions reported not to be fully compatible with CuAAC and these often involve the instability of organic azides toward heat, light, or phosphine.\textsuperscript{62} It has been reported that the copper catalyst may be partially poisoned by nucleophilic thiol species\textsuperscript{63} and that it is not fully compatible with oxime formation.\textsuperscript{64} Residual copper can also be problematic in certain applications.\textsuperscript{65}

The radical-mediated addition of thiols to terminal alkene groups, called the “thio-click” by Schlaad in 2007;\textsuperscript{66} has emerged as a useful complement or alternative to the CuAAC reaction. Radical “thio-click” reactions can be initiated thermally and photochemically where the latter has a higher reaction. Radical “thio-click” reactions can be initiated thermally and photochemically where the latter has a higher efficiency.\textsuperscript{67} The utility of the thiol–ene, thiol–yne, and thiol–Michael addition reactions is evident in their widespread applications, including in polymer and materials synthesis and as a tool in chemical biology studies.\textsuperscript{68,69} Nitroxide radical coupling is also considered a click-like reaction.\textsuperscript{70–72} As noted previously, CuAAC and NRC form an orthogonal reaction pair allowing the preparation a variety of polymeric G3 dendrimers with \(M_n > 20\,000\) in a one-pot fashion.\textsuperscript{73}

### 4.5.3 Orthogonal dynamic combinatorial libraries using click-type chemistry.

The resulting double-level dynamic libraries are therefore named orthogonal. The two processes, in principle, do not interfere, and analogously to the orthogonal protecting groups in organic chemistry”\textsuperscript{74} Lehn \textit{et al.} (2001)\textsuperscript{75} Together with the CuAAC reaction, certain imine-forming condensations\textsuperscript{76} have been classified by Sharpless as fulfilling the criteria of a click reaction. Similar to the CuAAC process, these condensation reactions can be orthogonal to a range of other reactions, but a key difference is that imine formation is reversible under some conditions. With this orthogonal and reversible nature, Eliseev and Lehn reported the first double-level, dynamic combinatorial library (DCL) based on the orthogonality of two reversible processes: imine formation and metal–ligand exchange.\textsuperscript{77} Other pairs of orthogonal exchange reactions have been reported, for example hydrazone formation–disulfide exchange\textsuperscript{78} and imine formation–nitro-based exchange.\textsuperscript{79} The complexity of DCLs has also been raised to a triple-level with the use of three orthogonal exchange processes. Thus, Otto and Nitschke reported a DCL with imine, disulfide, and metal–ligand exchanges.\textsuperscript{80} Furlan and co-workers reported a DCL with hydrazone, disulfide, and thioester exchanges.\textsuperscript{81} For more details, readers are directed to several recent reviews.\textsuperscript{82}

### 4.5.4 Bioorthogonal ligation using click chemistry.

“Selective chemical reactions that are orthogonal to the diverse functionality of biological systems are now recognized as important tools in chemical biology. These bioorthogonal reactions have inspired new strategies”\textsuperscript{83,84} Bertozzi \textit{et al.} (2004)\textsuperscript{85}

The past decade has seen an increasing focus on developing highly chemoselective reactions in biological settings. These click reactions were termed bioorthogonal and have proven useful new tools for studying biological processes by functionalizing biomolecules selectively.\textsuperscript{86} Beyond working in aqueous medium, bioorthogonal reactions may need to occur on or within living cells and ultimately in organisms.\textsuperscript{87} One example is the modified Staudinger reaction reported by Saxon and Bertozzi in 2000,\textsuperscript{88} which involves the selective functionalization of cell surface glycoproteins (Fig. 19). Thus, Jurkat cells were incubated with \(N\)-azido-acyethylmannosamine so that they displayed azido groups on their surface. The Staudinger–Bertozzi ligation, as it is now known,\textsuperscript{89} shows no cross-reactivity between the phosphate and cellular disulfide bonds. Although the Staudinger–Bertozzi ligation is relatively slow (\(k_{\text{obs}} \approx 10^{-7} \text{ M}^{-1} \text{ s}^{-1}\)),\textsuperscript{90} the reaction has shown to be applicable in bacterial cultures\textsuperscript{91} and in mice.\textsuperscript{92}

The CuAAC reaction was also applied to cellular bio-labeling.\textsuperscript{93} The issue of copper cytotoxicity can be circumvented by using polyazolate ligands\textsuperscript{94} to accelerate the reaction and protect the cells from reactive oxygen species (ROS) generated by the Cu-catalyzed reduction of oxygen. Nonetheless, considerable
cell death would be expected when the copper concentration is above the maximum tolerated level of ca. 500 μM.78

To overcome the toxicity of copper ions, Bertozzi and co-workers developed a number of cyclooctynes for the copper-free, strain-promoted alkyne–azide cycloaddition (SPAAC).79 The triple bond in the cyclooctyne (OCT) ring is more reactive (strain-promoted) than common terminal alkyne groups and allows the cycloaddition to occur under physiological conditions without the presence of copper catalyst. Boon’s dibenzo-cyclooctynes (DBCO) were also found to undergo SPAAC and were applied successfully to live cell labeling.87 The rates of the first generation cyclooctyne derivatives were found to be comparable to that of Staudinger ligation, and about 70–80-fold slower than the standard CuAAC reaction.78,79

Computational studies88 provided insights into the origin of the rate enhancement (e.g., electronic vs. strain effects) for a range of click partners (Fig. 20).95 With a difluoro substituted cyclooctyne moiety (DIFO), the rate of SPAAC is found to be comparable to CuAAC under pseudo-first-order conditions.90 Similarly, the DBCO systems reported by Boons and Popik were found to have similar reaction rates.77 Apparent drawbacks of these systems are their lengthy syntheses and hydrophobicity. The latter can limit aqueous solubility and lead to membrane or serum proteins localization. Sletten and Bertozzi reported the dimethoxyazacyclooctyne (DIMAC) system as having an improved log S value (−2.7 vs. −3.1 for OCT) and hence showing higher water-solubility.92 DIMAC also showed much less non-specific background labeling compared to its parent compounds. The system does require a 9-step sequence from a known pyranoside (overall yield ~6%) and the rate is comparable to that of cyclooctyne.92

The tetrazine-based inverse-electron-demand cycloaddition is receiving considerable attention as a new member in the bioorthogonal reaction toolbox because of its faster reaction rate (see Fig. 20).93 Schultz and Lemke showed the incorporation of trans-cyclooctene (TCO) containing lysine into proteins and the subsequent Diels–Alder reaction with fluorogenic tetrazine-functionalized dyes in living cells.94 The strain-promoted inverse-electron-demand Diels–Alder cycloaddition (SPIEDAC) occurs with a rate constant as high as 35 000 M⁻¹ s⁻¹. An even higher rate constant (380 000 M⁻¹ s⁻¹) was recently reported in a study on in vitro DNA labeling using (E)-bicyclo[6.1.0]non-4-ene95 derivatives. In general, tetrazine cycloadditions with cyclooctynes94 and norbornenes97 also proceed more rapidly compared to the same reaction with organic azides.98

The broad range of rate constants in Fig. 20 (note logarithmic scale) suggests that some of the bioorthogonal reactions can be used simultaneously without cross talk. Indeed, Jaeschke recently reported a one-pot, in vitro dual labeling of DNA using CuAAC and tetrazine-based cycloaddition as an orthogonal reaction pair.96 Also very recently, Hilderbrand reported the use of two copper-free cycloadditions (DBCO–azide and TCO–tetrazine) as an orthogonal reaction pair for selective multi-target imaging of cancer cells (Fig. 21).99

The development of "photo-click" chemistry offers spatio-temporal control over the reaction initiation.78,79b Thus, Lin and co-workers developed the photo-initiated high efficiency bioconjugation of proteins using 2,5-diaryltetrazoles.100

Fig. 19 Bioorthogonal Staudinger–Bertozzi ligation.80

Fig. 20 Comparison of range of rate constants of bioorthogonal reactions. Numbers in parenthesis are the largest rate constants reported in the series. Rate constants depend on the actual conditions and substituents on the substrates; the ranges are served as ballpark comparison only. Structures represented the best in the series. Adapted with permission from ref. 89.

Fig. 21 Two orthogonal click cyclization reactions can be used simultaneously in a biological setting.89
The diaryltetrazole undergoes nitrogen extrusion to form a nitrile–imine dipole that undergoes a [3 + 2]-cycloaddition with alkene dipolarophiles (Fig. 22). Boons and Popik also developed a photo-inducible system in which the diheterocyclooctyne ring is masked by a cyclopropenone motif. The photo-triggered decarboxylation offers the ability to label living organisms in a spatio-temporally controlled manner. This system has also been applied successfully in the modification of surfaces (vide infra).

**5. Orthogonality (fidelity) in supramolecular chemistry**

**5.1 Metal–ligand coordination combined with hydrogen bonding**

...“From a synthetic point of view it is important that the two types of interactions [metal–ligand coordination and hydrogen bonding] are “orthogonal”, that is, mutually compatible.”

Reinhoudt et al. (1997)

The orthogonality discussed thus far is based on chemical reactions, i.e., two highly chemoselective functional group inter-conversions. The concept of orthogonality can be extended to non-covalent interactions. Indeed, the remarkable complexity of biological systems is built upon a series of intricate non-covalent interactions that are an exceptional example of a high-dimension orthogonal system. Because supramolecular systems are at equilibrium, the orthogonality is continually being tested, unlike the case of an irreversible chemoselective reaction. Although this possible complication can be a challenge, it also represents an opportunity because unwanted supramolecular interactions may self-correct. So there is a proofreading mechanism possible in self-assembling systems in general.

Reflecting on some of these ideas, Reinhoudt reported in 1997 the first discrete nanostructure assembled by two different supramolecular interactions (Fig. 23). Metallo-dendrimers as large as 28 kDa were characterized making these among the largest discrete self-assembled polymers (dendrimers) known.

The importance of the orthogonality between the metal–ligand coordination and hydrogen bonding was explicitly recognized by Reinhoudt (see the above quote). The pincer coordination seen in Fig. 23 is now extensively used in combination with other hydrogen bonding recognition modules by the Weck group.

Other commonly used coordination complexes include 2,2'-bipyridine (bpy) and 2,2' : 6',2''-terpyridine (tpy) metal complexes that are orthogonal to many of the same hydrogen bonding modules used with the Pd–pincer complex. These units have been particularly useful in the design of main-chain and side-chain modified supramolecular polymers. For example, Schubert et al. reported the formation of hydrogen-bonded, metallo-supramolecular polymers containing the UPy hydrogen bonding module and terpyridine as end-groups in the presence of the Zn(II) ion. The same concept was applied to telechelic polymers to yield high molecular weight supramolecular polymers. Noting that the UPy dimer is stable only in non-polar solvents, Schmuck prepared a monomer containing a terpyridine and a guanidinoacarbonyl pyrrole carboxylate zwitterion. The zwitterion is known to form a highly stable dimer in DMSO ($K > 10^4$ M$^{-1}$) and the resulting polymer is observed in both DMSO and aqueous solutions displaying typical ring-chain equilibrium. The self-assembly of the hydrogen-bonding and metal-coordination is found to be fully reversible. This orthogonal approach to supramolecular polymers has been reviewed recently.

**5.2 Orthogonal hydrogen bonding interactions**

...“These non-standard nucleotides and the pairs that they form have had particular value as ‘orthogonal binders’, recognition elements that bind with DNA-like specificity, but without interference by natural DNA.”

Benner et al. (2007)

The very high fidelity of DNA replication is often associated with highly selective hydrogen bonding. In fact, DNA replication is not perfect and the high fidelity originates not only in base-pair selectivity but also in shape complementarity and subsequent proofreading processes. The issue of selectivity in the pairing of DNA bases and their analogs has been examined in multiple contexts. Benner and co-workers have developed base analogs as a way to expand the “genetic alphabet,”

---

**Fig. 22** Photo-triggered functionalization of D-allyl-tyrosine-containing Z-domain protein by photo-click chemistry.

**Fig. 23** Reinhoudt’s self-assembled dendrimer featuring orthogonal Pd–pincer coordination and cyanuric acid (CA)–melamine hydrogen bonded rosette motifs.
to create a new base that is orthogonal to A/T and G/C. Jørgensen used computational methods to explore the origin of base-pair stability and proposed that, in addition to the number of hydrogen bonds, the stability could be explained by the arrangement of the donor and acceptor groups and particularly the resultant differences in “secondary electrostatic interactions” (Fig. 24).

We reported in 1992 the first experimental data that supported the Jørgensen proposal and have since collected considerable additional data and this knowledge has been used in the design of more highly orthogonal and more stable DNA base-pair analogs. For example, the UG/DAN heterocomplex and DeAP dimer represent two hydrogen-bonded pairs that are highly orthogonal and significantly more stable than DNA base-pairs (Fig. 25).

Minakawa, Matsuda and co-workers found that these quadruply hydrogen bonded base pairs, depending on the sequence arrangement, are capable of increasing the thermal stability of oligonucleotides. Leigh and co-workers reported in 2011 a highly stable AAAA/DDDD quadruple hydrogen bonded complex with $K_a > 10^{12}$ M$^{-1}$ in CH$_2$Cl$_2$.

One area that has not been fully explored is the degree of orthogonality between the various hydrogen bonding modules represented by the arrays in Fig. 24. Nonetheless, these modules have proven quite useful in a wide range of orthogonal assembly studies. For example, we reported the orthogonal self-assembly of two dendritic systems consisting of bis-ureidodeazapterin (6) and pyrimido-naphthyridine (7) units (Fig. 26). Strikingly no mixing of the two heterocycles was observed even though the AADD and DDA arrays in 7 could interact with the AADD array in 6 with the formation of three hydrogen bonds.

Related areas where these high-affinity and high-fidelity hydrogen-bonded complexes have been applied are, e.g., in the formation of main-chain supramolecular polymers, in the noncovalent side-chain modification of polymers, and the formation of supramolecular polymer blends. An example of the latter involves two immiscible polymers (PS and PBMA), prepared with a few mol% of UG and DAN units on the polymer backbone of PBMA and PS, respectively. The properties (such as $T_g$ and viscosity) of the supramolecular blend were found to be tunable and thermoreversible. The same recognition pair was reported to form supramolecular alternating block copolymers with PEG and PBMA blocks. Again, the properties of the copolymers and the polymeric structures are found to be reversible because of the supramolecular nature of the recognition process.

Weck and co-workers demonstrated the orthogonal nature of UG-DAN and Hamilton receptor cyanuric acid (Hamilton receptor CA) and applied the units for the preparation of supramolecular ABC triblock copolymers. The triblock copolymer can be prepared by stepwise additions of each block or by a...
one-pot self-assembly. Recently, Lüning reported the use of Hamilton receptor CA and complementary ADDA ADP units for orthogonal assembly of a G2 supramolecular dendrimer.128 Weck reported the supramolecular side-chain functionalization of copolymers using a combination of Hamilton receptor CA and T-DAP motifs (Fig. 27) but noted non-specific interactions between the Hamilton receptor and T.127 Similarly, hydrogen-bonding and metal coordination might not be fully orthogonal, depending on the particular system.129,130 Reviews on side-chain supramolecular polymers can be referred to for more details.131,132

5.3 Quantifying orthogonality in supramolecular systems

In the case of orthogonal reactions, the extent of orthogonality can be measured by the degree of chemoselectivity, which, in turn, is measured by the chemical yield of desired and undesired products. In supramolecular systems, the extent of orthogonality can be directly observed. For example, in a remarkable example of "self-sorting" Isaacs has reported the 1H NMR spectrum of a multicomponent supramolecular mixture.133 The self-assembling systems, including Rebek's calixarene capsule, Meijer's ureido-pyrimidinone dimer, Reinhoudt's rosette, and Isaacs's cucurbituril dimer, all faithfully formed their designed complex without interference from the other components. The same group has defined different types of self-sorting and examined quantitatively the process under thermodynamic control using a combination of simulation and experiment.134-136

As a framework for discussing and quantifying supramolecular orthogonality, we defined the fidelity (F) of a supramolecular system as the mole fraction of desired complexes (Fig. 28).119,117,118 If each of the possible equilibrium constants can be measured or estimated, F can be calculated analytically at all concentrations and stoichiometries of interest. Importantly, a three-dimensional fidelity plot allows these systems to be understood in considerable detail (Fig. 28). To better measure the inherent orthogonality of a system, F at a 1:1 ratio of the compounds or complexes of interest was measured and this was termed intrinsic fidelity (intrinsic orthogonality).137

Two fidelity plots that examine the orthogonality of a pair of complexes are illustrated in Fig. 28. First, using a mixture of triply hydrogen-bonded complexes G-C and T-DAP, high fidelity (F > 0.96) is observed over a broad concentration range (1.0 × 10^{-6} M), indicating that the hydrogen-bonding arrays in G-C and T-DAP are capable of leading to highly selective binding. This is consistent with the fact that the homodimerizations and other undesired complexations are relatively weak (<10 M^{-1}). Similarly, a quadruply hydrogen-bonded system consisting of UG-DAN and (DeAP)_2 forms a very high fidelity system (F ≥ 0.99) over a wide range of concentrations and stoichiometries.119 Again, although both complexes are quadruply hydrogen-bonded, the difference in the arrays (DAAD vs ADDA and DDAA) contributes significantly to the orthogonality in the recognition process.

6. Orthogonal surface modifications

The modification of surfaces is important in many areas, including the fabrication of microelectronics, optoelectronics, and sensors.139 Modified surfaces also play a key role in various aspects of biological engineering, for example as scaffolds for tissue engineering, stem-cell differentiation, and cell culture.140

![Fig. 27](image_url) Weck’s supramolecular polymer side-chain functionalization displaying orthogonal self-assembly of Hamilton receptor CA and T-DAP motifs.127

![Fig. 28](image_url) Fidelity plots for mixtures of (a) G-C and T-DAP and (b) UG-DAN and (DeAP)_2 complexes. Concentrations are expressed in molar (M). Adapted with permission from ref. 119. Copyright (2005) American Chemical Society.

This journal is © The Royal Society of Chemistry 2013
Various lithographic techniques allow the creation of patterned surfaces with micron to nanometer scale features. The ability to orthogonally alter the surface chemistry allows precise control over the properties and patterns that are displayed.

6.1 Surface modifications by self-assembly

In 1989 Whitesides and co-workers reported the formation of patterned self-assembled monolayers, employing an orthogonal set of hard–soft acid–base interactions. In particular, the selective adsorption of alkanethiols on gold and organic acids on alumina surfaces was observed on gold surfaces with a micro-patterned silver overlay (Fig. 29). Two additional orthogonal self-assembly systems were reported, one using alkanethiols on gold and carboxylic acids or alkane isocyanides on platinum. The latter system did not achieve full orthogonality as it requires a non-equimolar solution to achieve full orthogonality as it requires a non-equimolar solution of the disulfide and isocyanide (40 : 1) to achieve selective surface functionalization because of the undesired competitive adsorption of isocyanide on a gold surface.

Recently, Hoeppeiner reported the self-assembly of Si and Au nanoparticles (NPs) and hydroxyl-containing micelles on a multifunctionalized silicon substrate with −NH₂, −SH, or −CO₂H groups, with potential selectivity based on their chemical reactivity. Full orthogonality was not possible because undesired interactions such as adsorption of Au NP onto the ammonium surface occur. To avoid cross reactivity, the functionalization was done in a sequential manner: (1) self-assembly of negative charged Si NP on the ammonium functionalized surface, (2) Au NP on the thiol surface, and (3) hydroxyl-containing micelles on the carboxylic acid surface.

6.2 Surface modifications via click chemistry

Not surprisingly CuAAC has emerged as a powerful method to covalently attach various entities to surfaces. By combining with other orthogonal reactions, CuAAC has further allowed selective surface functionalization to give patterned surfaces. For example, Gleason used capillary force lithography (CFL) to create a nanopattern of amine and alkyn groups from a bilayer consisting of 50 nm thick non-crosslinked poly(propargyl methacrylate) (PPMA) film laid by initiated chemical vapor deposition (iCVD) on top of a 100 nm thick poly(allylamine) (PAAm) film using plasma enhanced chemical vapor deposition (PECVD) on top of a 100 nm thick poly(allylamine) (PAAm) film using plasma enhanced chemical vapor deposition.

Fig. 29 Orthogonal surface self-assembly based on hard–soft acid–base principle.

Fig. 30 Simultaneous dual surface functionalization using a combination of CuAAC and carbodiimide chemistry. Reproduced with permission from ref. 145. Copyright (2008) American Chemical Society.

Fig. 31 (a) Dual protein patterning using oxime formation followed by CuAAC. (b) Overlay of myoglobin (green) and ubiquitin (red) fluorescence images. (c) Ubiquitin reacts with alkyn pattern. (d) Myoglobin reacts with aminooxy patterns. Scale bar = 5 μm. Adapted with permission from ref. 146.

Hawker reported a novel set of orthogonal copper-free, thermal reactions for selective surface modification through thermal microcontact printing (μCP). Three different dyes were anchored to the surface using orthogonal thermal azide–alkyne cycloaddition (TAAC) and acyl ketene formation followed by nucleophile trapping. It was found that acyl ketene was formed at 150 °C from its precursor and reacted with amines but not azides; the azide was found to be stable even at 150 °C.

Another example of surface modification using a copper-free, dibenzocyclooctyne-based SPAAC was reported by Popik.
and co-workers. This DIBO-based immobilization has also been applied successfully to functionalize micro bead surfaces. One of the appealing features of the DIBO system is that the reactive triple bond can be masked as a cyclopropenone motif (Fig. 32a). Photo-decarbonylation can be easily performed upon irradiation at 350 nm, thereby revealing the reactive triple bond for copper-free cycloaddition (Fig. 32b). Spatially controlled immobilization can be achieved simply using a shadow mask. Areas protected by the mask will not undergo cycloaddition, thus allowing spatial functionalization on a surface substrate (Fig. 32c–e). Bowman and co-workers also reported the spatiotemporally controlled alkyne–azide cycloaddition using the photochemical reduction of Cu(n) to Cu(0) with photolithographic techniques.

6.3 Surface modifications via supramolecular interactions

An extremely active area of research applies the principles of Section 5 – orthogonal supramolecular connections – to the patterning of surfaces and the formation of functional mono- and multi-layers. This area has been reviewed very recently by Yilmaz and Huskens and, thus, will not be covered here.

7. Conclusions

The concept of orthogonality has come a long way from the protecting group strategy first described by Barany and Merrifield in 1977. Indeed that very same notion of two chemoselective reactions operating with a complete lack of interference from one another has become pervasive in organic chemistry. Indeed, the complexity has multidimensional protecting group strategies available and one can envision a similar development in orthogonal coupling reactions and orthogonal catalytic cycles. Ultimately, this would facilitate more efficient syntheses and perhaps even lead to a day when multi-step synthesis can be carried out in a single pot with simultaneous addition of the reagents.

The orthogonal approach has moved well beyond the realm of synthetic organic chemistry into materials and polymer chemistry and also become a powerful tool for chemical labeling of biomolecules within living cells. The remarkable rate at which click chemistry and bioorthogonal reactions have been adopted is testimony to their powerful ability to link two entities in high yield and without interference even in complex systems. The expansion of the click chemistry toolbox by having additional orthogonal click reactions will undoubtedly further accelerate progress in many of the areas described herein. Where will the next big advance come? Predictions along these lines are always challenging, but more sophisticated strategies for controlling the spatial and temporal aspects of orthogonal processes, especially in synthetic organic and biological contexts, would represent a major advance.

Acknowledgements

We gratefully acknowledge the National Science Foundation (CHE-1012212) for funding.

Notes and references

2. From the Greek polys, “of many senses,” i.e., a term with different but related meanings.