HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2018 August 14.

Published in final edited form as:

Biochemistry. 2018 April 10; 57(14): 2094–2108. doi:10.1021/acs.biochem.8b00062.

Design of Bivalent Nucleic Acid Ligands for Recognition of RNA-Repeated Expansion Associated with Huntington's Disease

Shivaji A. Thadke[*,†,‡,§], **J. Dinithi R. Perera**[†,‡,§], **V. M. Hridya**[⊥], **Kirti Bhatt**[#], **Ashif Y. Shaikh**[†,‡,§], **Wei-Che Hsieh**[†,‡,§], **Mengshen Chen**[‖], **Chakicherla Gayathri**[†], **Roberto R. Gil**[†], **Gordon S. Rule**[‖], **Arnab Mukherjee**[⊥], **Charles A. Thornton**[#], and **Danith H. Ly**[*,†,‡,§]

[†]Department of Chemistry, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States

[‡]Institute for Biomolecular Design and Discovery (IBD), Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States

[§]Center for Nucleic Acids Science and Technology (CNAST), Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States

[‖]Department of Biological Sciences, Carnegie Mellon University, 4400 Fifth Avenue, Pittsburgh, Pennsylvania 15213, United States

[⊥]Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune, Maharashtra 411008, India

[#]Department of Neurology, Box 645, University of Rochester Medical Center, 601 Elmwood Avenue, Rochester, New York 14642, United States

Abstract

We report the development of a new class of nucleic acid ligands that is comprised of Janus bases and the MPγPNA backbone and is capable of binding rCAG repeats in a sequence-specific and selective manner via, inference, bivalent H-bonding interactions. Individually, the interactions between ligands and RNA are weak and transient. However, upon the installation of a C-terminal thioester and an N-terminal cystine and the reduction of disulfide bond, they undergo template-directed native chemical ligation to form concatenated oligomeric products that bind tightly to the RNA template. In the absence of an RNA target, they self-deactivate by undergoing an intramolecular reaction to form cyclic products, rendering them inactive for further binding. The work has implications for the design of ultrashort nucleic acid ligands for targeting rCAG-repeat expansion associated with Huntington's disease and a number of other related neuromuscular and neurodegenerative disorders.

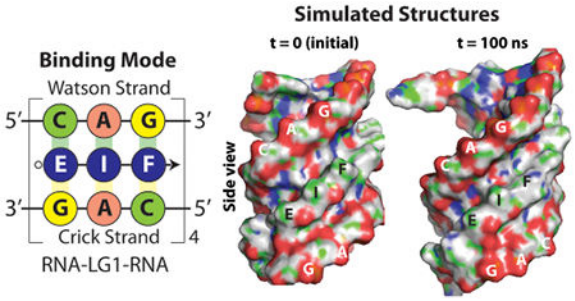
[*]**Corresponding Authors**: sathadke@andrew.cmu.edu. Telephone: (412) 268-4010. Fax: (412) 268-5579.; dly@andrew.cmu.edu.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.8b00062.

Spectroscopic data, MD simulations, and procedures for monomer and ligand syntheses (PDF)

The authors declare no competing financial interest.



RNA-repeated expansions are prevalent in neuromuscular disorders.[1–6] One such example is Huntington's disease (HD), an autosomal dominant disorder that affects muscle coordination and leads to behavioral changes, cognitive decline, and dementia.[7,8] HD typically manifests in midlife, and death ensues 10–20 years after onset. It is caused by an expansion of CAG repeats in the first exon of the *huntingtin* (*htt*) gene, from a normal range of 6–29 to a pathogenic range of 40–180.[9] The length of the CAG-repeat expansion (CAG^{exp}) is inversely related to the age of onset.[10,11] *htt* encodes a 348 kDa protein that is expressed ubiquitously but most predominantly in the neurons of central nervous system (CNS), with diverse physiological roles, including embryonic development and neuroprotection.[12–17] The CAG repeats are translated into a polyglutamine (polyQ) sequence in the N-terminal region of Htt. Despite the vast knowledge of molecular dysfunctions and clinical manifestations, the exact mechanism by which CAG^{exp} causes HD is not yet fully understood; however, emerging evidence suggests that it is a multivariate disorder.

A loss of protein function could contribute to the etiology of HD, but it is unlikely the main culprit because heterozygous and homozygous patients with CAG^{exp} have similar clinical features.[18] Furthermore, an individual who showed no abnormal phenotype despite a 50% reduction in the normal Htt protein level due to deletion in one of the *htt* alleles has been identified.[19] Likewise, mice heterozygous for the *htt* null mutation do not show clinical features of HD, and homozygotes die in early embryogenesis.[13–15] These findings underscore the importance of Htt in embryonic development, but not in the pathogenesis of HD. Emerging evidence points to deleterious gain of function as a more plausible cause of the disease. This suggestion is substantiated by the observations that CAG repeats exceeding a similar threshold (30–40 units) in unrelated genes are responsible for several other neuromuscular disorders, including Machado-Joseph disease (MJD), spinobulbar muscular atrophy (SBMA), dentatorubral-pallidoluysian atrophy (DRPLA), and the various forms of spinocerebellar ataxias (SCAs).[1,3,4,6,20] These disorders, known as polyQ diseases, have many traits, including subcortical and cortical atrophy, and nuclear aggregates that are in common with HD.

Three leading cytotoxic mechanisms have been proposed for HD. The first is polyQ toxicity.[21] PolyQ has a propensity to aggregate and interact with other polyQ-containing proteins to form large amyloid-like structures.[22–25] The binding and sequestration of these key proteins would lead to the losses of their physiological functions, resulting in dysregulation of a cascade of molecular and cellular events.[1] The second is toxic gain of RNA function.26 Upon transcription, $rCAG^{exp}$ adopts an imperfect hairpin structure that sequesters

muscleblind-like protein 1 (MBNL1), an alternative RNA-splicing regulator,[27] and other key proteins.[28–30] While the normal rCAG repeats can also adopt a hairpin motif, this repeat does so in a sequence context different from that of the expanded version.[31] Association of $rCAG^{exp}$ with MBNL1 results in a complex that is trapped in the nucleus as nuclear foci, precluding its export to the cytoplasm for the production of Htt protein. The gene transcripts that are misspliced as a result of the loss of MBNL1 are diverse.[32] The third pathogenic mechanism is protein toxicity.[33] Zu and Ranum[34] have shown that rCAG repeats beyond a certain length (>42 units) can be translated in the absence of an ATG starting codon via repeat-associated non-ATG (RAN) translation, leading to the production of toxic polyQ and polyalanine (polyA) proteins, along with polyserine (polyS).[33,35] The latter two mechanisms are further supported by the findings that expression of a long tract of untranslated rCAG repeats is deleterious in animal models, with abnormal behavioral phenotypes similar to that of HD.[35–37] Collectively, these findings point to HD as a multivariate disorder, caused by toxic gain of RNA and protein functions.

As such, one approach to remedy HD would be to selectively target the expanded transcript, as this would interfere with all three disease pathways. There is optimism that such an approach would prevail based on the findings by Yamamoto and Hen,[38] who showed in a conditional-knockout mouse model that the progression of HD is reliant on a constitutive expression of the expanded *htt* allele, and that its ablation not only halted the progression of the disease but also reversed the disease phenotypes. The challenge, however, is in how to design molecules that would be able to bind the expanded but not the wild-type *htt* transcript, because Htt plays crucial roles in normal development and neuroprotection.

At present, there are no effective treatments or cures for HD, even with the latest Food and Drug Administration-approved drug, deutetrabenazine.[39] The existing remedies are designed to alleviate the symptoms but not to halt or reverse progression of the disease. More recently, gene-specific targeting approaches, including small-molecule ligands,[40,41] siRNA,[42,43] shRNA,[44,45] antisense,[46,47] and gene editing strategies such as TALEN[48] and CRISPR/Cas9,[49] have been explored as potential remedies for HD. Notwithstanding recent advances in molecular design and chemical synthesis, the issues of binding specificity and/or selectivity, and cellular delivery, to a certain extent, remain for many of these classes of molecules, in particular antisense agents. A gene editing approach, although conceptually appealing because of its ability to permanently remove or disable the mutant allele, is unlikely to produce long-term therapeutic effects for HD due to the perpetual replication slippage of the normal allele, as a result of defects in the DNA repair machinery of the host patients.[5,50] Herein, we report a proof-of-concept study of the design of bivalent nucleic acid ligands for recognition of rCAG repeats, with a potential for generalization to other RNA-repeated sequences.

MATERIALS AND EXPERIMENTAL DETAILS

Ultraviolet (UV) Melting Analyses.

All UV melting samples were prepared by mixing ligands with RNA targets at the indicated concentrations in 0.1× PBS buffer and annealed by incubation at 90 °C for 5 min followed by a gradual cooling to room temperature. UV melting curves were collected using an

Agilent Cary UV–vis 300 spectrometer equipped with a thermoelectrically controlled multicell holder. UV melting spectra were recorded by monitoring UV absorption at 260 nm from 25 to 95 °C in the heating runs and from 95 to 25 °C in the cooling runs, both at a rate of 1 °C/min. The cooling and heating curves were nearly identical, indicating that the hybridization process was reversible. The recorded spectra were smoothed using a 20-point adjacent averaging algorithm. The first derivatives of the melting curves were taken to determine the melting temperatures of the complex.

Circular Dichroism (CD) Analyses.

The samples were prepared in 0.1× PBS buffer. All spectra represent an average of at least 15 scans collected at a rate of 100 nm/min between 200 and 375 nm, in a 1 cm path-length cuvette at 25 °C. The CD spectrum of the buffer solution was subtracted from the sample spectra, which were then smoothed via a five-point adjacent averaging algorithm.

Steady-State Fluorescent Measurements.

All steady-state fluorescence samples were prepared by mixing ligands with RNA at the indicated concentrations in 0.1× PBS buffer and annealed by being incubated at 90 °C for 5 min followed by a gradual cooling to 37 °C. The samples were incubated at 37 °C for 1 h before the measurements. Steady-state fluorescence data were collected at 25 °C with a Cary Eclipse fluorescence spectrometer (ε_{ex} = 330 nm, and ε_{em} = 340–600 nm).

Isothermal Titration Calorimetry (ITC) Measurements.

Isothermal titration calorimetry was performed using a MicroCal VP-ITC unit (MicroCal) following the protocol reported by Zimmerman et al.[1] Prior to the experiment, 1.7 mL of 5 μM R11A in 0.1× PBS buffer and 400 μL of 1.2 mM LG2 in 0.1× PBS buffer were annealed at 95 °C and cooled to room temperature. The two samples were then degassed for 1 min using a ThermoVac (MicroCal) while being stirred. A standard ITC experiment was performed at 25 °C in which 10 μL aliquots of the degassed LG2 solution were injected from a 287 μL rotating syringe (307 rpm) into the isothermal sample chamber containing 1.4616 mL of the degassed R11A solution. The duration of each injection was 20 s, and the delay time between the injections was 300 s. The initial delay before the first injection was 60 s. Each injection generated a heat burst curve (microcalories per second vs time). The area under each heat burst curve (isotherm) was determined by integration using the Origin 7.0 software (MicroCal) to give the amount of heat produced or absorbed per injection. The most accurate data (best fits) were obtained by fitting to the sequential binding model of 9 and 10 binding sites. The K_d values were calculated by taking the reciprocal of the K_n values provided by the nine-binding site model. The 0.1× PBS buffer consisted of 1 mM NaP_i, 13.7 mM NaCl, and 0.27 mM KCl (pH 7.4).

Competitive Binding Assays.

R24A was purchased from IDT. R127A [$r(CAG)_{127}$] and R96U [$r(CUG)_{96}$] were prepared as previously described.[13] R24A, R126A, and R96U were prepared in 0.1× PBS buffer and annealed by being heated to 90 °C for 5 min, followed by a gradual cooling to room temperature. Ligand and RNA were mixed at the indicated concentrations and incubated at

37 °C for 24 h. The samples were then loaded onto a 2% agarose gel with 1× Tris-borate buffer and electrophoretically separated at 100 V for 25 min. The gel was stained with SYBR-Gold and visualized with a UV-Transilluminator.

RESULTS

Rationale.

Nuclear magnetic resonance (NMR),[51] X-ray,[52,53] and biochemical[31,54] studies revealed that the rCAG hairpin structures are relatively dynamic compared to that of the canonical RNA duplex, with the A-A internal bulge exhibiting a large amplitude of motion. Such a molecular scaffold, comprising an internal A-A mismatch at every two canonical G-C/C-G base pairs (Figure 1A), akin to a "pothole" in the road, presents a distinct and viable receptor-like binding site for exogenous ligands. Given the relative ease of accessibility of the H-bond donors and acceptors of nucleobases and the relaxed thermodynamic requirement for establishing H-bonding interactions with such nucleobase targets, along with a recent development of ultra-high-affinity miniPEG-containing γ peptide nucleic acid (MPγPNA),[55,56] we posited that it may be feasible to develop relatively short nucleic acid ligands for targeting rCAG hairpin structures (Figure 1B). We envisioned that the binding free energy required to invade the transiently folded RNA hairpin could be attained by incorporating Janus bases (or JBs), E, I, and F (Figure 1C), that are capable of forming bivalent H-bonding interactions with nucleobases in both strands of the RNA double helix, with the conformationally preorganized MPγPNA backbone (Figure 1D,E). Several "Janus wedges" have been developed, notably by Lehn,[57] McLaughlin,[58,59] Zimmerman,[60] He,[61] Tor,[62] Bong,[63] Marchan,[64] and Micklefield,[65] for recognition of nucleic acid biopolymers; however, they are few in number and have different shapes and sizes (Figure S1) and, as such, cannot effectively be combined in a modular format for recognition of a nonhomogeneous sequence. The Janus bases reported in this study are part of a larger set of bifacial nucleic acid recognition elements, 16 in total, designed to bind to all 16 possible RNA base pair combinations (Figure S2). They differ from the other Janus bases, including "Janus wedges", in that they are uniform in shape, size, and chemical functionality and, as such, could be designed to bind to any RNA base pair combinations (Figure S3). This proof-of-concept study sought to determine whether a subset of Janus bases, namely, E, F, and I, could be chemically synthesized and whether the corresponding ligands could recognize a biomedically relevant RNA target, namely rCAG repeats.

Molecular Dynamics (MD) Simulations.

To assess the feasibility of Janus base recognition, we performed MD simulations of ligand LG1 bound to an RNA duplex containing four rCAG repeats. The C-terminal lysine residue was omitted, and the MP side chain was replaced with a methyl group (MeγPNA)56 to simplify the computational modeling. The CEG, AIA, and GFC triads were built and optimized by the 631G* basis set and grafted onto the respective RNA and MeγPNA backbone. The structure of the bound RNA–LG1–RNA complex was created using the NAB module of Ambertools. The final structure was solvated with water molecules and ions, energy minimized, and simulated for 100 ns. The resulting complex remained fairly stable throughout the simulation, with the four separate LG1 ligands fitting snuggly between the

two RNA strands (Figure 2A). The number of H-bonds, five for each of the CEG and GFC triads and four for AIA (Figure 2B), remained intact throughout the simulation (Figure 2C). Attempts to simulate binding with fewer than four LG1 ligands, however, were unsuccessful. The complexes unraveled, with some forming contorted structures, due to fraying of the terminal base pairs (Figure 2D,E). This finding suggests binding cooperativity between ligands and RNA, presumably because of intermolecular base-stacking interactions.

Synthesis of Chemical Building Blocks and Ligands.

Encouraged by the result of MD simulations, we synthesized Janus bases E and F, along with corresponding JB-MPγPNA monomers **1–3** (Chart 1). Janus base I was not prepared because it was commercially available. E and F were synthesized according to Schemes 1 and 2, respectively. The Boc protection/deprotection sequence in compounds **5–9** was found to be necessary to improve the chemical yield of NBS reaction and to suppress side reactions in the subsequent condensation and cyclization steps. It was discovered that further protection of **12** was necessary for Stille coupling to proceed smoothly. This was accomplished with Boc-anhydride, performed at an elevated temperature. Despite the apparent bulkiness, we had no issue coupling E to the MeγPNA backbone or in the assembly of the corresponding monomer on MBHA-resin. The core structure of F, 4-amino-2-(methylthio)-pyridine-5-carbonitrile (**17**), was prepared according to the published protocols.[66] Subsequent conversion of the cyano group to amidine followed by Boc protection yielded **21**. Attempts to fully protect the exocyclic amines with a large excess of Boc-anhydride under various conditions, however, were unsuccessful, as this led to the formation of multiple spots on TLC with different numbers of Boc groups that were difficult to separate by column chromatography. To address this challenge, we performed Boc protection in two steps. Subjecting **22** to oxidation and hydrolysis, followed by alkylation and hydrogenolysis, yielded F. Once prepared, E (**15**), F (**26**), and I (**27**) were coupled to the MeγPNA backbone (Scheme 3), synthesized with slight modifications of the published procedures.[67] Removal of the Alloc protecting group yielded desired monomers **1–3**. Ligands LG1–LG3, along with LG2P, which is in an opposite (parallel) orientation of LG2, were prepared on MBHA-resin using a PAL linker and HBTU as a coupling reagent. LG2P was included as a negative control, because the parent LG2 ligand showed the most promise for binding rCAG repeats. A lysine residue was incorporated at the C-terminus to improve water solubility. Upon completion of the last monomer coupling, ligands were cleaved from resin, precipitated with diethyl ether, purified by reverse phase high-performance liquid chromatography (RP-HPLC) (Figure S4), and confirmed by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (Figure S5 and Table S1).

Target Selection and Sample Preparations.

A series of model RNA targets were chosen for a binding study (Chart 2). The R11X series comprised 24 rCXG repeats (Chart 2A), which, upon adopting a secondary hairpin structure, provided 11 binding sites for ligands (Chart 2B). R11A contained the perfect match (A-A), while R11U and R11C contained the respective U-U and C-C mismatches for ligand binding. R11G (X = G) was not ascertained because it has been shown to be capable of forming G-quadruplex as well as hairpin structures,[68] which could confound the interpretation of the experimental results. Single-stranded RNA targets, WS (Watson strand)

and CS (Crick strand), with the LG2 binding site underlined (Chart 2C), as well as double-stranded hairpin HP containing a single ligand binding site (Chart 2D), were also examined. The samples were prepared by incubating the preannealed RNAs with respective ligands in 0.1 × PBS buffer [1 mM NaP_i, 13.7 mM NaCl, and 0.27 mM KCl (pH 7.4)] at 37 °C. This particular buffer was chosen on the basis of our initial screen, in which we found that R11A adopted a stable hairpin structure (Figure S6)[69] and that ligands were able to bind.

Spectroscopic Characterization of Ligand Binding.

A combination of UV–visible (UV–vis) and circular dichroism (CD) was employed to determine the binding properties of ligands. UV–vis measurements revealed that all three ligands were able to bind R11A (Figures S7-S9). The evidence of their interaction can be gleaned from the hypochromicity and bathochromic shift in the absorption of [R11A +ligand] at 252 and 330 nm. The UV absorption in the 275–375 nm regimes corresponds to the π-π^* transition of the E base. This finding was corroborated by CD data, which revealed a marked increase and red-shift in the signals at ~270 nm (Figure S10). In this series, LG2 exhibited the largest differential in CD amplitude. We took this as evidence of its effectiveness in being able to bind R11A. A plausible explanation for the preferential binding of LG2 could be the fact that its binding site has a C-terminal A-A mismatch (Figure 1A). Generally, a terminal mismatch is preferred over one in the middle, because it provides greater access for ligand to initiate “toehold” binding followed by chain migration. However, the reverse, N-terminal A-A mismatch, as in the binding site of LG3, is not preferred because it is in a reverse direction of the helical induction of the ligand, which is from C- to N-terminus.[70] CD titration of LG2 with R11A reached a saturation point at a 11:1 ratio, consistent with the predicted number of ligand binding sites of R11A (Figure 3). We made the assumption that R11A adopted a uniform hairpin structure, because it was thermally annealed prior to the addition of ligands; however, other combinations of intra- and intermolecular folds are also possible. In contrast, no significant differences in the CD signals were observed following the incubation of LG2 with the mismatched R11U (Figure 4A) and R11C (data not shown), single-stranded WS and CS (Figure 4B), double-stranded HP containing a single binding site (Figure 4C), or the mismatched orientation LG2P with R11A (Figure 4D). Because the results of the R11U and R11C mismatches were nearly identical in every respect, all the discussions of base mismatch hereafter are centered on R11U. Taken together, these results show that the interactions between ligand and RNA occurred in a sequence- and orientation-specific manner, with a preference for a double-stranded over a single-stranded RNA target, and in favor of that with multiple consecutive binding sites over one in isolation.

Confirmation of Ligand Binding by Fluorescent Measurements.

To further corroborate the UV–vis and CD findings, we measured the fluorescent signals of ligands with and without RNA targets following an excitation at 330 nm (λ_{max} of E base). We observed a significant fluorescent quenching of LG3 compared to that of LG1 and LG2 (Figure S11). Such a drastic reduction in the intensity of the fluorescent signals could be due to photoinduced electron transfer quenching, which is expected to be most efficient for LG3 because the fluorophore (E) is stacked between two other Janus bases. This interpretation is consistent with the observation that the fluorescent signals of LG3 were fully recovered

upon heating (data not shown). With all three ligands, the emission signals were further reduced upon the addition of R11A, but most dramatically with LG2 (Figure S11, inset): 60% reduction with LG2 as compared to 35 and 33% with LG1 and LG3, respectively. A photograph of the samples under UV illumination is shown in Figure S12. Unexpectedly, we observed a similar fluorescent quenching pattern upon the incubation of LG2 with R11U, [WS+CS], HP, and LG2P with R11A (Figure 5), although it was less dramatic than that of LG2 with R11A. A possible explanation for the fluorescent decay of LG2 with these off-targets could be due to nonspecific electrostatic and/or hydrophobic interactions between ligand and RNA, particularly in the groove(s) of a hairpin duplex, as in the case of R11U. A similar observation was made with a related class of bicyclic pyrimidine analogues,[71] indicating that the fluorescent yields of these fused-ring fluorophores are highly sensitive to the local environments. To test the hypothesis that the nonspecific binding of ligand to RNA can lead to fluorescent quenching, we incubated the respective RNA targets with pentamidine prior to adding LG2 and vice versa. The rationale was that because pentamidine is known to bind RNA in the groove(s) through a combination of electrostatic and hydrophobic interactions with relatively high affinity,[72] its inclusion should be able to displace ligands that bind weakly in the grooves of RNA through these nonspecific interactions and restore the fluorescent signals, except in a case in which ligand binding occurred via a different mode. In agreement with this prediction, the data showed that the fluorescent signals of LG2 with R11U and [WS+CS], as well as that of LG2P with R11A, were fully recovered, except in the case of LG2 with R11A, which remained suppressed (Figure 5, inset). This result is consistent with LG2 binding R11A via a mode different from that of pentamidine, presumably via the prescribed bivalent H-bonding interactions (Figure 1B,C).

Isothermal Titration Calorimetry (ITC) Analyses.

The binding stoichiometry and affinity of LG2 to R11A were estimated by ITC. The raw ITC profile is shown in Figure 6A, with each heat burst corresponding to a single injection of LG2 into the R11A sample. The thermogram is biphasic in character, with a combination of enthalpic and entropic contributions, suggesting two possible modes for binding of LG2 to R11A. The first binding mode is exothermic, which is apparent from the evolution of heat, likely as the result of H-bond formation between the Janus bases of the ligand and the nucleobases of RNA. This binding mode is enthalpically favorable because of a large differential in the net gain of H-bonds formed (+8 per ligand binding). Once the bivalent binding sites are fully occupied, further addition of ligand would result in its binding in the grooves through electrostatic and/or hydrophobic interactions, resulting in the ejection of counterions and hydrated water molecules into the bulk solution. Conformational changes of RNA could also contribute to ligand binding. This is evident in the endothermic pattern in the second half of the thermogram. This interpretation is consistent with the fluorescence data, suggesting two possible binding modes: one that is sequence-specific and another that is not. The binding isotherm was constructed by plotting the heat evolved/absorbed as a function of the ligand:RNA molar ratio (Figure 6B). The molar ratio at half-maximum height, corresponding to the stoichiometry of binding, is determined to be 10. This value is in good agreement with that obtained by CD titration. The most accurate fitting for the data was obtained from a sequential binding model of nine binding sites, which yielded the

following K_d values: 46.3 ± 10.3, 9.7 ± 2.3, 35.2 ± 5.7, 47.6 ± 2.7, 43.3 ± 3.9, 34.8 ± 3.4, 111 ± 5.9, 235 ± 11.6, and 67.1 ± 2.8 μM. Because of the complexity of the system (RNA with 11 consecutive binding sites) and the novel binding mode of the ligand, we suggested caution be exercised in the interpretation of these experimental findings. However, from these ITC data, we can estimate the K_d values to be in the low micromolar range.

NMR Titration.

To gain further insight into the mode of binding of LG2, we performed a series of multinuclear and multidimensional NMR experiments with LG2, R11A, and both. The samples were prepared in the same 0.1× PBS buffer as before, containing a 9:1 H_2O:D_2O volume ratio. The NOESY data were collected at 300, 200, and 100 ms mixing times to obtain proton–proton distances, while COSY experiments were performed to map the proton spin system of each residue. The expectation was that H-bonding interactions between LG2 and R11A would result in line broadening and downfield chemical shifts of the imino proton signals of the canonical G-C/C-G pairs due to base pair opening and in the emergence of a new set of imino proton signals due to the formation of H-bonds between ligand and RNA.

Consistent with the CD data, ^{1}H NMR experiments revealed that R11A adopted a stable hairpin structure in 0.1× PBS buffer, which is evident by the sharp imino proton signal at 12.35 ppm (Figure S13A). Variable-temperature measurements further supported this chemical-shift assignment, showing a gradual decay in peak intensity with an increasing temperature, while that of the nonexchangeable protons remained fairly constant (Figure S14). Partial assignments of the nucleobase protons were made by NOESY experiments (Figure S15). As predicted, the imino proton signal of R11A was gradually broadened and downfield-shifted upon the addition of LG2 (Figure 7). Moreover, the chemical shifts of C4-NH and G2-NH were significantly affected, an indication of their interactions with ligand. Despite the efforts, we were not able to assign or thus monitor the chemical shift of A6-NH upon titrating LG2 into R11A due to the degeneracy in the repeated sequence. Nonetheless, the result indicates that the G-C/C-G base pair openings are mediated by ligand. However, we did not observe any new imino proton signals in the 10–20 ppm regimes, as would be expected for formation of H-bonds between ligand and RNA. A plausible explanation for their absence could be due to the weak and transient interactions, as is evident from the ITC data, whereby the newly formed imino protons were in fast exchange with water molecules.[73] A second possibility, although unlikely based on the spectroscopic evidence, could be that LG2 failed to bind R11A via the prescribed bivalent H-bonding mode but instead associated with RNA through nonspecific electrostatic and/or hydrophobic interactions.

To delineate the two possibilities, we performed NMR titration of LG2P with R11A under identical conditions. LG2P contained the same sequence as LG2 but in the opposite orientation. The rationale was that if electrostatic and/or hydrophobic interactions were responsible for LG2 binding, such nonspecific interactions would be unlikely to be significantly affected by the orientation of ligand. Thus, we would expect a similar chemical-shift profile for [LG2P+R11A] and [LG2+R11A]. On the other hand, if LG2 interacts with R11A via a different mode, we would expect a drastically different spectral pattern. To our surprise, we found that the [LG2P+R11A] mixtures precipitated at a 6:1

molar ratio. The precipitation, in the form of a white sticky material, was not due to the poor solubility of LG2P because its stock solution was prepared at a concentration of 10 mM, significantly higher than that employed in the titration experiment, but rather due to its combination with R11A. The decay in the imino proton signal at 12.35 ppm upon addition of LG2P was not due to ligand binding (Figure S16), which has a distinct characteristic (Figure 7), but rather due to the reduction in the overall concentrations of R11A and LG2P. This is evident by the lack of a downfield shift in the imino proton signal, together with the reduction of the nonexchangeable proton signals of nucleobases (Figure S16, solid vertical lines). We attributed the precipitation to the electrostatic and/or hydrophobic interactions of LG2P with R11A, which resulted in the formation of larger complexes with reduced overall net charges, hence, the precipitation. While the NMR data alone are inconclusive, in combination with the spectroscopic findings, they provide supporting evidence of the formation of a binding mode that is different from electrostatic and hydrophobic interactions, presumably via the expected bivalent H-bonding; albeit, it was weak and transient.

Template-Directed Native Chemical Ligation (NCL).

The weak and transient interactions of ligand with RNA, as observed with LG2 and R11A, are unlikely to produce a meaningful biological response. To further improve the binding affinity of LG2, we synthesized a second LG2 derivative, LG2N (Scheme 4A), which contained a C-terminal thioester and an N-terminal cystine. This ligand was prepared using a second-generation *N*-acylurea linker developed by Blanco-Canosa and Dawson.[74] The dual-function probe design exploits the conformational preorganization of MPγPNA in slowing the two functional groups from spontaneously reacting with each other upon the reduction of the disulfide bond. The cystine group was employed to provide greater control in probe handling. Prior studies revealed that a ligand of the same length comprising all natural nucleobases has a reduced (acyclic) halflife of ~1 h at a physiological temperature (37 °C).[75] We expected a similar half-life for LG2N*. Scheme 4B depicts the reaction pathways of LG2N following the reduction of the disulfide bond, the expected chemical state of LG2N in the reducing intracellular environments.[76] In the presence of an RNA target, ligand was expected to form transient bivalent H-bonding interactions with nucleobases of the RNA target next to one another as the result of intermolecular base stacking (step 2). Upon the cleavage of the disulfide bond, ligand would undergo template-directed NCL to form concatenated products that bind an RNA template more tightly (step 3). In the absence of an RNA target, ligand would self-deactivate by undergoing an intramolecular NCL reaction to form cyclic products (step 4).

To confirm the prediction that LG2N would remain stable for a finite period following the reduction of the disulfide bond before undergoing a cyclization reaction, we monitored the reaction progress of the parent compound M from which LG2N* was derived *in situ* (Figure 8A). We selected MALDI-TOF over HPLC and other analytical techniques in the quantification of the reaction products because of the relatively short time scale of the intramolecular NCL reaction, which can be accomplished with the former in <5 min. While HPLC is a common method for the quantification of reaction products, it is not sufficiently robust for the system under investigation due to the competing intra- and intermolecular

reactions of ligand. Attempts were made to quench the reaction with an electrophile prior to HPLC analysis; however, such an effort was not successful because of the rapid intramolecular reaction of ligand. 4-Mercaptophenol (4MP) and 2-mercaptoethanol (2ME) were investigated as possible reducing agents. Both yielded similar kinetic profiles for the parent compound M (Figures S17 and S18); however, the former yielded an intermediate with an overlapping mass-to-charge ratio (m/z) like that of the hydrolyzed parent compound, making it difficult to differentiate one from the other. For this reason, we selected 4MP in the reaction monitoring by MALDI-TOF and in the gel-shift assay, while 2ME was employed in the subsequent melting experiments because of its transparency in the nucleobase absorption regions. Our data revealed that upon addition of 4MP to the parent compound M, two reactive species corresponding to that of the transesterification ($M^{\#}$) and the N-terminal cysteine ($M^{\#\#}$) were initially formed (Figure 8A). They persisted for a relatively short period (<15 min) before converting into the reactive LG2N* intermediate, and finally into the cyclic product (cLG2N) (Figure 8B). LG2N* was stable at a physiological temperature. It existed as a major product at a 1 h reduction time point, accounting for ~75% of the total products in the mixture, with the remaining 25% being mostly cLG2N. LG2N* has a half-life of ~3 h, which is ~3 times longer than that of the natural counterpart.[75] We attributed the chemical stability of LG2N* to the expanded aromatic ring size of the E base, which provides better base-stacking interaction and, thus, makes ligand less conformation-ally flexible.

Next, we performed UV melting experiments to determine the effect of template-directed NCL on the thermal stability of RNA. The samples were prepared in a 0.1 × PBS buffer at a 2:1 ratio of LG2N to the number of binding sites of R11A and incubated at 37 °C for 16 h prior to UV melting analyses. As expected, without the reducing agent, LG2N had a minimal effect on the melting transition (T_m) of R11A, with a AT_m of approximately +1 °C at most (Figure S19). A similar finding was made with LG2, which also did not contain a C-terminal thioester or an N-terminal cysteine (Figure S20). However, in the presence of 2ME, whether the incubation of [LG2N +R11A] was performed at ambient temperature or at 37 °C, the results were similar. A significant enhancement in the T_m of R11A appeared in both cases, in the range of 59–68 °C. The derivative of the melting curves showed broad sigmoidal distributions, suggesting the presence of a range of concatenated products being formed and bound to the RNA template. This finding was expected for template-directed synthesis. Despite the similarity in the T_m, the two melting curves are distinct in pattern, with the one incubated at an elevated temperature showing an inverse trend in optical absorption in the 25–50 °C range. This melting behavior is characteristic of hydrophobic–aromatic interaction, a phenomenon that has been previously observed with thermophilic foldamers[77] and other aromatic systems, such as perylene[78] and pyrene.[79] We attributed the inverse intensity distribution to the interactions of Janus bases of the concatenated products, which, presumably, were formed in greater quantities at 37 °C than at ambient temperature. When the compounds were heated, the hydrophobic interaction became more pronounced up to a certain point (~50 °C), beyond which repulsion ensued. The enhancement of the thermal stability of R11A is consistent with the formation of NCL products and their binding to the RNA template.

To confirm that the enhancement in the T_m of R11A was due to template-directed concatenation and binding of the resultant products to the RNA template, we performed MALDI-TOF analyses of the UV melting samples prior to heating. Formation of the concatenated products is apparent from the emergence of new peaks with progressively larger *m/z* values with a step size of ~1319 Da, corresponding to the mass of ligand (Figure 9A). These concatenated products, however, were not observed with the mismatched R11U (Figure 9B) or LG2N alone without the target (Figure 9C). Likewise, there was no evidence of the concatenated products being formed with single-stranded [WS+CS] or double-stranded HP (data not shown). This result is consistent with the occurrence of template-directed synthesis, mediated by the R11A hairpin structure.

Selective Binding of the Concatenated Ligand.

To determine whether LG2N can discriminate $rCAG^{exp}$ from the normal repeats, we performed a competitive binding assay. Two RNA targets containing $r(CAG)_n$ repeats in the normal (R24A; $n = 24$, the same as R11A) and pathogenic (R127A; $n = 127$) range along with the mismatched $r(CUG)_{96}$ control (R96U) were employed. Upon adopting the respective hairpin motif, these RNA transcripts would provide 11, 62, and 47 binding sites for LG2N. Two sets of samples were prepared, one in 0.1 × PBS and another in a physiologically relevant buffer [10 mM NaP_i, 150 mM KCl, and 2 mM $MgCl_2$ (pH 7.4)].[69] In both sets, equimolar mixtures of R24A (100 nM strand concentration and 1.1 μM binding sites) and R127A (18 nM strand concentration and 1.1 μM binding sites) were incubated with various concentrations of LG2N, along with 4MP and TCEP, at 37 °C for 24 h. TCEP was included to ensure that the disulfide bond was completely reduced. The reaction mixtures were analyzed by agarose gel and stained with SYBR-Gold for visualization. Inspection of Figure 10 revealed that LG2N bound preferentially to R127A over R24A, which is evident by the faster rate of disappearance of the top two bands compared to that of the bottom band in lanes 2–5. Such a binding event occurred only in the presence of reducing agents and with the perfectly matched RNA target, as no evidence of binding was observed in the absence of 4MP and TCEP (compare lane 6 to lane 1) and with the mismatched R96U (compare lane 8 to lane 7). Interestingly, we did not observe any shifted bands being formed, as would be expected for the complexation of ligand with RNA, suggesting that SYBR-Gold was not able to intercalate the RNA–ligand complex. This observation was not surprising considering that ligand binding is known to cause a dramatic change in the conformation of RNA (Figure 2). In contrast, we did not observe any evidence of ligand binding taking place at a physiologically relevant ionic strength (Figure S21), indicating that the RNA hairpin was not able to mediate template-directed ligand oligomerization under such a condition. This outcome is unlikely due to the inability of ligand to access the H-bond donors and acceptors of RNA nucleobases, as we have demonstrated in the invasion of the DNA double helix by MP γPNA,[80] but rather due to the lack of binding free energy. There are several ways in which the binding affinity of ligand can be improved. Nonetheless, the data show that LG2N can discriminate the expanded (pathogenic) from the normal rCAG repeats and that it occurred in a sequence-specific manner, although binding in this case is restricted to a relatively low ionic strength.

DISCUSSION

A large number of neuromuscular disorders, more than 20 in total, including HD, myotonic dystrophy type 1 (DM1) and type 2 (DM2), spinocerebella ataxias (SCAs), fragile X syndrome (FXS), Friedreich's ataxia (FRDA), and a subpopulation of amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease),[81,82] manifest as a result of unstable repeat expansions. An expansion in the coding region of a gene can lead to an altered protein function, whereas that occurring in the noncoding region can cause a disease without interfering with a protein sequence through toxic gain of RNA function and/or inadvertent production of deleterious polypeptides via RAN translation. Many of these genetic disorders are autosomal dominant, requiring an expansion (or mutation) in only one allele to cause a disease. As such, one way to interfere with such a disease pathway would be to target the affected allele or the corresponding gene transcript. Between the two, the latter provides greater accessibility for exogenous molecules and greater recognition specificity for bivalent ligands, such as JB-MPγPNAs, because of its propensity to adopt an imperfect hairpin structure.[51–53] The challenge, however, is in the design of molecules that would be able to discriminate the expanded (diseased) from the wild-type transcript. Because Htt protein, like most native gene products associated with repeat expansions, is known to play key roles in normal development and neuroprotection, perturbation of its expression level could compound the severity of the disease and lead to other adverse effects.[83] So far, several classes of molecules have been developed and have been shown to be effective in targeting rCAG repeats. The most notable were reported by Disney,[84–86] Zimmerman,[60,87] Thornton,[88,89] Corey,[46] and Cleveland.[47] However, it remains to be seen whether they can be translated into safe and effective molecular therapies for HD. The design of molecules with a novel binding mode, such as JB-γPNA ligands, would complement the ongoing efforts to develop remedies for these debilitating genetic disorders.

We concentrated our effort on rCAG repeats because of their expansive genetic penetrance in HD, as well as in a number of other neurodegenerative disorders, including MJD, SBMA, DRPLA, and SCAs.[4] In principle, a molecule that selectively binds the mutant *htt* transcript and interferes with the aforementioned disease pathways could be used to treat not only HD but also a number of other related genetic disorders. The nucleic acid ligands reported in this study differ from the conventional antisense agents in several respects.[90] First, they bind the CAG-RNA hairpin motif through bivalent H-bonding interactions. Second, they are relatively small in size, a single triplet-repeating unit in length. As such, they are easier to chemically synthesize, structurally modify, and scale up, via solution phase as opposed to solid phase chemistry. Third, with a molecular weight on the outer fringe of those of small molecules, such a "millamolecular" system is generally more pharmacokinetically favorable than a typical antisense agent.[91] Fourth, Janus base recognition is more sequence-specific and selective. It is more specific because a single-base mismatch that would normally occur on one face of a natural nucleobase would occur on both faces of a Janus base, and it is more selective, favoring a double-stranded over a single-stranded RNA target, because of the large differential in binding free energy of the two resultant products. However, unlike that of small molecules,[84,92] or riboswitches,[93] the recognition of Janus bases follows a defined set of rules, via bivalent Watson–Crick H-bonding interactions, as opposed to a more diverse

combination of attractive forces. Furthermore, JB ligand design is modular. As such, ligands could easily be prepared and modified to bind to any sequence of RNA repeats. The weak and transient interactions of JB ligands with normal-length rCAG repeats at low salt concentrations are desirable at the initial design stage, as this will allow for further improvement in the recognition selectivity by tuning the binding affinity of ligands and by adjusting the ionic strength of buffers toward that of the physiological conditions.

There are several ways in which the avidity of JB ligands could be improved. One approach involves combining the recognition event with the concatenation process, as illustrated in the binding and concatenation of LG2N in the reducing environments. Individually, the binding of such a ligand to its RNA target is relatively weak. However, upon the reduction of the disulfide bond, the expected chemical state of ligand in the reducing intracellular environments, the resultant intermediate would undergo template-directed synthesis, with the resulting products binding more tightly to the RNA template. In the absence of an RNA target, ligand would undergo intramolecular NCL to form cyclic products, rendering them inactive for further binding. Template-directed synthesis is exploited by Nature in the synthesis of biomolecules for tight and selective binding of their cognate partners.[94] Such a concept has long been recognized by chemists[95–97] and has been successfully applied in the synthesis of organic molecules 98 and in the organization and assembly of a broad range of macro- and supramolecular systems.[99,100] Despite the appeal, such an approach has rarely been applied to the recognition of a biomedically relevant target. This is largely due to the fact that natural biopolymers, such as nucleic acids and proteins, and the majority of the synthetic analogues that have been developed to date, are highly flexible in conformation.[101] As such, they are generally not suitable for RNA repeat-templated NCL, due to the propensity of the C-terminal thioester and N-terminal cysteine groups to undergo spontaneous cyclization reaction. The extended helical conformation of MPγPNA offers a unique opportunity to explore template-directed synthesis in the recognition of RNA repeats. A demonstration that the rCAG-RNA hairpin can function as a template for ligand concatenation via NCL suggests the possibility of intracellular applications. The ability of LG2N to self-deactivate in the absence of an RNA target may provide a mechanism for further reducing off-target binding and cytotoxic effects.

The present LG2N design is not yet optimal for NCL reaction or effective binding of the concatenated products with RNA. Instead of being a part of ligand, which would produce a seamless backbone in the conjugated products, the N-terminal cysteine residue was appended as an add-on. As such, it would introduce a spacer between each ligand unit in the concatenated products, which could negatively impact their binding affinity.[102] Such an issue could be circumvented by installing a cysteine side chain (CH_2SH) at the γ backbone of MPγPNA so that upon oligomerization a native backbone would be generated. A second approach involves covalent attachment of aromatic pendant groups at the termini of ligand, as this would facilitate intermolecular π–π interaction, promote cooperative binding, and improve binding selectivity, in favor of the expanded over the normal-length repeats.[103] A third approach involves implementation of next-generation Janus bases, such as those shown in Chart 3, which are capable of forming six H-bonds each with the respective C-G and G-C pairs, as compared to five in the first-generation design, along with the improvement in base-stacking interaction as a result of the expanded aromatic ring size. Work is currently

underway to explore these various design strategies in attempts to further improve the binding affinity, specificity, and selectivity of JB ligand, along with cellular uptake,[104–106] so that its biological effects can be assessed in the native intracellular environments.

In addition to the aforementioned strategies to discriminate the diseased from the normal *htt* transcript, it should be pointed out that contrary to the traditionally held view that both of these transcripts adopt a hairpin motif, Weeks and co-workers[31] have shown that formation of the r(CAG) hairpin structure is dependent on sequence context. In the *htt* transcript of a healthy allele, the previously observed r(CAG) hairpin structure was either absent or short, while that of a diseased allele was progressively extended with an increasing r(CAG) repeat length. A large proportion of the wild-type r(CAG) sequence is base-paired with the conserved down-stream sequence, including a r(CCG) repeat region. This finding is intriguing in light of a recent phylogenic study by Cattaneo,[107] which showed that the Htt ancestor possesses a protein with a single Q or no Q at the 18th amino acid residue, the starting site of polyQ in humans. PolyQ has emerged in only fish and gradually expanded in mammals and become the longest in humans, presumably due to replication slippage. Interestingly, polyproline (polyP), located downstream of polyQ is present only in invertebrates, specifically marsupials and mammals. This indicates that polyP, whose coding sequence is a CCG repeat, is a recent and sudden acquisition in *htt* evolution, perhaps as a way to counter the deleterious effect of CAG repeats of the normal allele by means of base pairing. Such an occurrence would explain why CAG repeats in *htt* beyond a certain length are pathogenic. The subtle structural difference in *htt* transcripts, between those above and those below the premutation (30–39 repeats), provides another mechanistic insight into how one might be able to discriminate the expanded transcript from that of the normal allele. The Janus bases reported in this study, along with the remaining 13 (Figure S2), which were uniformly designed to bind to all 16 possible base pair combinations, will be valuable as molecular tools for addressing mechanistic questions and emerging hypotheses, such as that pertaining to the folding of rCAG repeats in the expanded and in the normal *htt* transcript, as well as potential therapeutic agents for the treatment of neuromuscular disorders associated with repeat expansions.

In summary, we have shown that relatively short nucleic ligands, three nucleotides in length, comprising Janus bases and the MPγPNA backbone, could be prepared and are able to bind rCAG repeats in a sequence-specific and selective manner. While the binding mode has not yet been conclusively determined, spectroscopic evidence points to H-bonding interaction via strand invasion as a likely scenario. The potential benefits of such ligands for recognition of RNA-repeated expansions, such as $rCAG^{exp}$, are numerous, including specific and selective binding; ease of chemical synthesis, structural modification, and large-scale production; improved pharmacological properties; and flexibility in sequence design. Another unique and novel feature of JB ligand is the possibility for template-directed synthesis and recognition of biomedically relevant RNA targets in the intracellular environments. The added benefit of ligand undergoing self-deactivation in the absence of an RNA target may offer another dimension for further improving binding selectivity and for reducing cytotoxicity. If successfully developed, the Janus bases and the corresponding ligands could, in principle, be designed to bind to any sequence of RNA-repeated expansions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Nuclear magnetic resonance and MALDI-TOF MS instrumentation at Carnegie Mellon University was partially supported by the National Science Foundation (CHE-9808188, CHE0130903, and CHE-1039870).

Funding

This work was supported in part by the DSF Charitable Foundation, the National Institutes of Health (R21NS098102), and the National Science Foundation (CHE-1609159). AM. thanks SERB (Grant EMR/2016/001069), Department of Science and Technology, Government of India.

REFERENCES

(1). Gatchel JR , and Zoghbi H (2005) Diseases of unstable repeat expansion: mechanisms and common principles. Nat. Rev. Genet 6, 743–755.16205714

(2). Orr HT , and Zoghbi H (2007) Trinucleotide Repeat Disorders. Annu. Rev. Neurosci 30, 575–621.17417937

(3). Mirkin SM (2007) Expandable DNA repeats and human disease. Nature 447, 932–940.17581576

(4). La Spada AR , and Taylor JP (2010) Repeat expansion disease: progress and puzzles in disease pathogenesis. Nat. Rev. Genet 11, 247–258.20177426

(5). Lee D-Y , and McMurray CT (2014) Trinucleotide expansion in disease: why is there a length threshold? Curr. Opin. Genet. Dev 26, 131–140.25282113

(6). Lopez Castel A , Cleary JD , and Pearson CE (2010) Repeat instability as the basis for human diseases and a a potential target for therapy. Nat. Rev. Mol. Cell Biol 11, 165–170.20177394

(7). Cattaneo E , Zuccato C , and Tartari M (2005) Normal huntingtin function: An alternative approach to Huntington's disease. Nat. Rev. Neurosci 6, 919–930.16288298

(8). Zuccato C , Valenza M , and Cattaneo E (2010) Molecular mechanisms and potential therapeutic targets in Huntington's disease. Physiol. Rev 90, 905–981.20664076

(9). MacDonald M , et al. (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell 72, 971–983.8458085

(10). Duyao M , Ambrose C , Myers R , Novelletto A , Persichetti F , Frontali M , Folstein S , Ross C , Franz M , Abbott M , et al. (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat. Genet 4, 387–392.8401587

(11). Stine OC , Pleasant N , Franz ML , Abbott MH , Folstein SE , and Ross CA (1993) Correlation between the onset age of Huntington's disease and length of the trinucleotide repeat in IT-15. Hum. Mol. Genet 2, 1547–1549.8268907

(12). Trottier Y , Lutz Y , Stevanin G , Imbert G , Devys D , Cancel G , Saudou F , Weber C , David G , Tora L , Agid Y , Brice A , and Mandel J-L (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. Nature 378, 403–406.7477379

(13). Nasir J , Floresco SB , O'Kusky JR , Diewert VM , Richman JM , Zeisler J , Borowski A , Marth JD , Phillips AG , and Hayden MR (1995) Targeted disruption of the Huntington's disease gene results in embryonic lethality and behavioral and morphological changes in heterozygoes. Cell 81 , 811–823.7774020

(14). Duyao MP , Auerbach AB , Ryan A , Persichetti F , Barnes GT , McNeil SM , Ge P , Vonsattel JP , Gusella JF , Joyner AL , et al. (1995) Inactivation of the mouse Huntington's disease gene homolog Hdh. Science 269, 407–410.7618107

(15). Zeitlin S , Liu JP , Chapman DL , Papaioannou VE , and Efstratiadis A (1995) Increased apoptosis and early embryonic lethality in mice nullizygous for the Huntington's disease gene homologue. Nat. Genet 11, 155–163.7550343

(16). Ferrante RJ , Gutekunst C-A , Persichetti F , McNeil SM , Kowall NW , Gusella JF , MacDonald ME , Beal MF , and Hersch SM (1997) Heterogeneous topographic and cellular distribution of huntingtin expression in the normal human neostriatum. J. Neurosci 17, 3052–3063.9096140

(17). Rigamonti D , Bauer JH , De-Fraja C , Conti L , Sipione S , Sciorati C , Clementi E , Hackam A , Hayden MR , Li Y , et al. (2000) Wild-type huntingtin protects from apoptosis upstream of caspase-3. J. Neurosci 20, 3705–3713.10804212

(18). Wexler NS , Young AB , Tanzi RE , Travers H , Starosta-Rubinstein S , Penney JB , Snodgrass SR , Shoulson I , Gomez F , Ramos-Arroyo MA , et al. (1987) Homozygotes for Huntington's disease. Nature 326, 194–197.2881213

(19). Ambrose CM , Duyao MP , Barnes G , Bates GP , Lin CS , Srinidhi J , Baxendale S , Hummerich H , Lehrach H , Altherr M , et al. (1994) Structure and expression of the Huntington's disease gene: Evidence against simple inactivation due to an expanded CAG repeat. Somatic Cell Mol. Genet 20, 27–38.

(20). McMurray CT (2010) Mechanism of trinucleotide repeat instability during human development. Nat. Rev. Genet 11 , 786–799.20953213

(21). Ross CA , and Poirier MA (2004) Protein aggregation and neurodegenerative disease. Nat. Med 10, S10–S17.15272267

(22). Nucifora FC , Sasaki M , Peters MF , Huang H , Cooper JK , Yamada M , Takahashi H , Tsuji S , Toncoso J , Dawson VL , et al. (2001) Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. Science 291, 2423–2428.11264541

(23). Sakahira H , Breuer P , Hayer-Hartl MK , and Hartl FU (2002) Molecular chaperones as modulators of polyglutamine protein aggregation and toxicity. Proc. Natl. Acad. Sci. U. S. A 99, 16412–16418.12189209

(24). Yano H , Baranov SV , Baranova OV , Kim J , Pan Y , Yablonska S , Carlisle DL , Ferrante RJ , Kim AH , and Friedlander RM (2014) Inhibition of mitochondrial protein import by mutant huntingtin. Nat. Neurosci 17, 822–831.24836077

(25). Ciechanover A , and Brundin P (2003) The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. Neuron 40, 427–446.14556719

(26). Nalavade R , Griesche N , Ryan DP , Hildebrand S , and Krauß S (2013) Mechanisms of RNA-induced toxicity in CAG repeat disorders. Cell Death Dis 4, e752.23907466

(27). Mykowska A , Sobczak K , Wojciechowska M , Kozlowski P , and Krzyzosiak WJ (2011) CAG repeats mimic CUG repeats in the misregulation of alternative splicing. Nucleic Acids Res 39, 8938–8951.21795378

(28). Tsoi H , Lau TC , Tsang SY , Lau KF , and Chan HY (2012) CAG expansion induces nucleolar stress in polyglutamine diseases. Proc. Natl. Acad. Sci U. S. A 109, 13428–13433.22847428

(29). van Eyk CL , O'Keefe LV , Lawlor KT , Samaraweera SE , McLeod CJ , Price GR , Venter DJ , and Richards RI (2011) Perturbation of the Akt/Gsk3-b signalling pathway is common to Drosophila expressing expanded untranslated CAG, CUG and AUUCU repeat RNAs. Hum. Mol. Genet 20, 2783–2794.21518731

(30). Krol J , Fiszer A , Mykowska A , Sobczak K , de Mezer M , and Krzyzosiak WJ (2007) Ribonuclease dicer cleaves triplet repeat hairpins into shorter repeats that silence specific targets. Mol. Cell 25, 575–586.17317629

(31). Busan S , and Weeks KM (2013) Role of context in RNA structure: Flanking sequences reconfigure CAG motif folding in huntingtin exon 1 transcripts. Biochemistry 52, 8219–8225.24199621

(32). Du H , Cline MS , Osborne RJ , Tuttle DL , Clark TA , Donohue JP , Hall MP , Shiue L , Swanson MS , Thornton CA , et al. (2010) Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. Nat. Struct. Mol. Biol 17, 187–193.20098426

(33). Pearson CE (2011) Repeat associated non-ATG translation initiation: One DNA, two transcripts, seven reading frames, potentially nine toxic entities. PLoS Genet 7, e1002018.21423665

(34). Zu T , Gibbens B , Doty NS , Gomes-Pereira M , Huguet A , Stone MD , Margolis J , Peterson M , Markowski TW , Ingram MAC , et al. (2011) Non-ATG-initiated translation directed by microsatellite expansions. Proc. Natl. Acad. Sci. U. S. A 108, 260–265.21173221

(35). Wojciechowska M , Olejniczak M , Galka-Marciniak P , Jazurek M , and Krzyzosiak WJ (2014) RAN translation and frameshifting as translational challenges at simple repeats of human neurodegenerative disorders. Nucleic Acids Res 42, 11849–11864.25217582

(36). Li L-B , Yu Z , Teng X , and Bonini NM (2008) RNA toxicity is a component of ataxin-3 degeneration in Drosophila. Nature 453, 1107–1111.18449188

(37). Hsu R-J , Hsiao K-M , Lin M-J , Li C-Y , Wang L-C , Chen L-K , and Pan H (2011) Long tract of untranslated CAG repeats is deleterious in transgenic mice. PLoS One 6, e16417.21283659

(38). Yamamoto A , Lucas JJ , and Hen R (2000) Reversal of neuropathology and motor dysfunction in a conditional model of Huntington's disease. Cell 101, 57–66.10778856

(39). Reilmann R (2016) Deutetrabenazine-not revolution but welcome evolution for treating chorea in Huntington's disease. JAMA Neurology 73, 1404–1406.27749952

(40). Sarkar S , Perlstein EO , Imarisio S , Pineau S , Cordenier A , Maglathlin RL , Webster JA , Lewis TA , O'Kane CJ , Schreiber SL , Rubinsztein DC , et al. (2007) Small molecules enhance autophagy and reduce toxicity in Huntington's disease models. Nat. Chem. Biol 3, 331–338.17486044

(41). Kumar A , Parkesh R , Sznajder LJ , Childs-Disney JL , Sobczak K , and Disney MD (2012) Chemical correction of pre-mRNA splicing defects associated with sequestration of muscleblind-like 1 protein by expanded r(CAG)-containing transcripts. ACS Chem. Biol 7, 496–505.22252896

(42). Wang YL , Liu W , Wada E , Murata M , Wada K , and Kanazawa I (2005) Clinico-pathological rescue of a model mouse of Huntington's disease by siRNA. Neurosci. Res 53, 241–249.16095740

(43). DiFiglia M , Sena-Esteves M , Chase K , Sapp E , Pfister E , Sass M , Yoder J , Reeves P , Pandey RK , Rajeev KG , et al. (2007) Therapeutic silencing of mutant huntingtin with siRNA attenuates striatal and cortical neuropathology and behavioral deficits. Proc. Natl. Acad. Sci. U. S. A 104, 17204–17209.17940007

(44). Harper SQ , Staber PD , He X , Eliason SL , Martins IH , Mao Q Yang L , Kotin RM , Paulson HL , and Davidson BL (2005) RNA interference improves motor and neuropathological abnormalities in a Huntington's disease mouse model. Proc. Natl. Acad. Sci. U. S. A 102, 5820–5825.15811941

(45). Rodriguez-Lebron E , Denovan-Wright EM , Nash K , Lewin AS , and Mandel RJ (2005) Intrastriatal rAAV-mediated delivery of anti-huntingtin shRNAs induces partial reversal of disease progression in R6/1 Huntington's disease transgenic mice. Mol. Ther 12, 618–633.16019264

(46). Hu J , Matsui M , Gagnon KT , Schwartz JC , Gabillet S , Arar K , Wu J , Bezprozvanny I , and Corey DR (2009) Allele-specific silencing of mutant huntington and ataxia-3 genes by targeting expanded CAG repeats in mRNAs. Nat. Biotechnol 27, 478–484.19412185

(47). Kordasiewicz HB , Stanek LM , Wancewicz EV , Mazur C , McAlonis MM , Pytel KA , Artates JW , Weiss A , Cheng SH , Shihabuddin LS , et al. (2012) Sustained therapeutic reversal of Huntington's disease by transient repression of huntingtin synthesis. Neuron 74, 1031–1044.22726834

(48). Richard G-F , Viterbo D , Khanna V , Mosbach V , Castelain L , and Dujon B (2014) Highly specific contraction of a single CAG/CTG trinucleotide repeat by TALEN in yeast. PLoS One 9, e95611.24748175

(49). Shin JW , Kim K-H , Chao MJ , Atwal RS , Gillis T , MacDonald ME , Gusella JF , and Lee J-M (2016) Permanent inactivation of Huntington's disease mutation by personalized allele-specific CRISPR/Cas9. Hum. Mol. Genet 25, 4566–4576.28172889

(50). M0llersen L , Rowe AD , Larsen E , Rognes T , and Klungland A (2010) Continuous and periodic expansion of CAG repeats in Huntington's Disease R6/1 Mice. PLoS Genet 6, e1001242.21170307

(51). Tawani A , and Kumar A (2015) Structural insights reveal the dynamics of the repeating r(CAG) transcript found in Huntington's disease (HD) and spinocerebellar ataxias (SCAs). PLoS One 10, e0131788.26148061

(52). Kiliszek A , Kierzek R , Krzyzosiak WJ , and Rypniewski W (2010) Atomic resolution structure of CAG RNA repeats: Structural insights and implications for the trinucleotide repeat expansion diseases. Nucleic Acids Res 38, 8370–8376.20702420

(53). Yildirim I , Park H , Disney MD , and Schatz GC (2013) A dynamic structural model of expanded RNA CAG repeats: A refined x-ray structure and computational investigations using molecular dynamics and umbrella sampling simulations. J. Am. Chem. Soc 135, 3528–3538.23441937

(54). Sobczak K , de Mezer M , Michlewski G , Krol J , and Krzyzosiak WJ (2003) RNA structure of trinucleotide repeats associated with human neurological diseases. Nucleic Acids Res 31, 5469–5482.14500809

(55). Sahu B , Sacui I , Rapireddy S , Zanotti KJ , Bahal R , Armitage BA , and Ly DH (2011) Synthesis and characterization of conformationally preorganized, (R)-diethylene glycol-containing g-peptide nucleic acids with superior hybridization properties and water solubility. J. Org. Chem 76, 5614–5627.21619025

(56). Yeh JI , Shivachev B , Rapireddy S , Crawford MJ , Gil RR , Du S , Madrid M , and Ly DH (2010) Crystal structure of chiral gammaPNA with complementary DNA strand: Insights into the stability and specificity of recognition and conformational preorganization. J. Am. Chem. Soc 132, 10717–10727.20681704

(57). Branda N , Kurz G , and Lehn J-M (1996) JANUS WEDGES: A new approach towards nucleobase-pair recognition. Chem. Commun, 2443–2444.

(58). Chen D , Meena , Sharma SK , and McLaughlin LW (2004) Formation and stability of a janus-wedge type of DNA triplex. J. Am. Chem. Soc 126, 70–71.14709064

(59). Chen H , Meena , and McLaughlin LM (2008) A janus-wedge DNA triplex with A-W1-T and G-W2-C base triplets. J. Am. Chem. Soc 130, 13190–13191.18783217

(60). Arambula JF , Ramisetty SR , Baranger AM , and Zimmerman SC (2009) A simple ligand that selectively targets CUG trinucleotide repeats and inhibits MBNL protein binding. Proc. Natl. Acad. Sci. U. S. A 106, 16068–16073.19805260

(61). Zhao H , Huang W , Wu X , Qing Y , Xing Z , and He Y (2011) Synthesis of a complete Janus-type guanosine-cytosine base and its 2′-deoxyribonucleoside. Chem. Lett 40, 684–686.

(62). Shin D , and Tor Y (2011) Bifacial nucleoside as a surrogate for both T and A in duplex DNA J. Am. Chem. Soc 133, 6926–6929.

(63). Zeng Y , Pratumyot Y , Piao X , and Bong D (2012) Discrete assembly of synthetic peptide-DNA triplex structures from polyvalent melamine-thymine bifacial recognition. J. Am. Chem. Soc 134, 832–835.22201288

(64). Artigas G , and Marchan V (2013) Synthesis of janus compounds for the recognition of G-U mismatched nucleobase pairs. J. Org. Chem 78, 10666–10677.24087986

(65). Robinson CJ , Vincent HA , Wu M-C , Lowe PT , Dunstan MS , Leys D , and Micklefield J (2014) Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. J. Am. Chem. Soc 136, 10615–10624.24971878

(66). Urleb U , Stanovnik B , and Tisler M (1986) Transformations of 4-amino-5-cyanopyrimidines. Synthesis and transformations of pyrimido[4,5-d]pyrimidines. Croat. Chem. Acta 59, 79–87.

(67). Manna A , Rapireddy S , Sureshkumar G , and Ly DH (2015) Synthesis of optically-pure yPNA monomers: A comparative study. Tetrahedron 71, 3507–3514.

(68). Fry M , and Loeb LA (1994) The fragile X syndrome d(CGG)n nucleotide repeats from a stable tetrahelical structure. Proc. Natl. Acad. Sci. U. S. A 91, 4950–4954.8197163

(69). Santoro SW , and Joyce GF (1997) A general purpose RNA-cleaving DNA enzyme. Proc. Natl. Acad. Sci. U. S. A 94, 4262–4266.9113977

(70). Dragulescu-Andrasi A , Rapireddy S , Frezza BM , Gayathri C , Gil RR , and Ly DH (2006) A simple gamma-backbone modification preorganizes peptide nucleic acid into a helical structure. J. Am. Chem. Soc 128, 10258–10267.16881656

(71). Greco NJ , and Tor Y (2005) Simple fluorescent pyrimidine analogues detect the presence of DNA abasic sites. J. Am. Chem. Soc 127, 10784.16076156

(72). Sun T , and Zhang Y (2008) Pentamidine binds to tRNA through non-specific hydrophobic interactions and inhibits amino-acylation and translation. Nucleic Acids Res 36, 1654–1664.18263620

(73). Bothe JR , Nikolova EN , Eichhorn CD , Chugh J , Hansen AL , and Al-Hashimi HM (2011) Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. Nat. Methods 8, 919–931.22036746

(74). Blanco-Canosa JB , Nardone B , Albericio F , and Dawson PE (2015) Chemical protein synthesis using a second-generation N-acylurea linker for the preparation of peptide-thioester precursors. J. Am. Chem. Soc 137, 7197–7209.25978693

(75). Bahal R , Manna A , Hsieh W-C , Thadke SA , Sureshkumar G , and Ly DH (2018) RNA-templated concatenation of triplet nucleic acid probe. ChemBioChem 19, n/a.

(76). Maher P (2005) The effects of stress and aging on glutathione metabolism. Ageing Res. Rev 4, 288–314.15936251

(77). Urry DW (1993) Molecular machines: how motion and other functions of living organisms can result from reversible chemical changes. Angew. Chem., Int. Ed. Engl 32, 819–841.

(78). Wang W , Wan W , Zhou H-H , Niu S , and Li ADQ (2003) Alternating DNA and pi-conjugated sequenes. Thermophilic foldable polymers. J. Am. Chem. Soc 125, 5248–5249.12720416

(79). Siu H , and Duhamel J (2008) Molar absorption coefficient of pyrene aggregates in water. J. Phys. Chem. B 112, 15301–15312.18989917

(80). Rapireddy S , Bahal R , and Ly DH (2011) Strand invasion of mixed-sequence, double-helical B-DNA by g-peptide nucleic acis containing G-clamp nucleobases under physiological conditions. Biochemistry 50, 3913–3918.21476606

(81). Renton AE , Majounie E , Waite A , Simon-Sanchez J , Rollinson S , Gibbs JR , Schymick JC , Laaksoirta H , van Sieten JC , Laaksovirta H , et al. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72, 257–268.21944779

(82). DeJesus-Hernandez M , Mackenzie IR , Boeve BF , Boxer AL , Baker M , Rutherford NJ , Nicholson AM , Finch NA , Flynn G , Adamson J , et al. (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72, 245–256.21944778

(83). Zuccato C , Ciammola A , Rigamonti D , Leavitt BR , Goffredo D , Conti L , MacDonald ME , Friedlander RM , Silani V , Hayden MR , et al. (2001) Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. Science 293, 493–498.11408619

(84). Guan L , and Disney MD (2012) Recent advances in developing small molecules targeting RNA. ACS Chem. Biol 7, 73–86.22185671

(85). Childs-Disney JL , Stepniak-Konieczna E , Tran T , Yildirim G , Park H , Chen CZ , Hoskins J , Southall N , Marugan JJ , Patnaik S , et al. (2013) Induction and reversal of myotonic dystrophy type 1 pre-mRNA splicing defects by small molecules. Nat. Commun 4, 2044.23806903

(86). Rzuczek SG , Colgan LA , Nakai Y , Cameron MD , Furling D , Yasuda R , and Disney MD (2017) Precise small-molecule recognition of a toxic CUG RNA repeat expansion. Nat. Chem. Biol 13, 188–193.27941760

(87). Nguyen L , Luu LM , Peng S , Serrano JF , Chan HYE , and Zimmerman SC (2015) Rationally designed small molecules that target both the DNA and RNA causing myotonic dystrophy type 1. J. Am. Chem. Soc 137, 14180–14189.26473464

(88). Wheeler TM , Sobczak K , Lueck JD , Osborne RJ , Lin X , Dirksen RT , and Thornton CA (2009) Reversal of RNA dominance by displacement of protein sequestered on triplet repeat RNA. Science 325, 336–339.19608921

(89). Wheeler TM , Leger AJ , Pandey SK , MacLeod AR , Nakamori M , Cheng SH , Wentworth BM , Bennett CF , and Thornton CA (2012) Targeting nuclear RNA for in vivo correction of myotonic dystrophy. Nature 488, 111–115.22859208

(90). Sahu NK , Shilakari G , Nayak A , and Kohli DV (2007) Antisense technology: A selective tool for gene expression regulaton and gene targeting. Curr. Pharm. Biotechnol 8, 291–304.17979727

(91). Fosgerau K , and Hoffmann T (2015) Peptide therapeutics: current status and future directions. Drug Discovery Today 20, 122–128.25450771

(92). Thomas JR , and Hergenrother PJ (2008) Targeting RNA with small molecules. Chem. Rev 108, 1171–1224.18361529

(93). Mandal M , and Breaker RR (2004) Gene regulation by riboswitches. Nat. Rev. Mol. Cell Biol 5, 451–463.15173824

(94). Li X , and Liu DR (2004) DNA-templated organic synthesis: Nature's strategy for controlling chemical reactivity applied to synthetic molecules. Angew. Chem., Int. Ed 43, 4848–4870.

(95). Pedersen CJ (1967) Cyclic polyethers and their complexes with metal salts. J. Am. Chem. Soc 89, 2495–2496.

(96). Cram DJ (1988) The design of molecular hosts, guests, and their complexes. Science 240, 760–767.3283937

(97). Dietrich B , Lehn J-M , and Sauvage J-P (1969) Diaza-polyoxamacrocycles et macrobicles. Tetrahedron Lett 10, 2885–2888.

(98). Gartner ZJ , Tse BN , Grubina R , Doyon JB , Snyder TM , and Liu DR (2004) DNA-templated organic synthesis and selection of a library of macrocyles. Science 305, 1601–1605.15319493

(99). Zhang ZJ , and Zaworotko MJ (2014) Template-directed synthesis of metal-organic materials. Chem. Soc. Rev 43, 5444–5455.24831589

(100). Zhao HP , Zhou M , Wen LY , and Lei Y (2015) Template-directed construction of nanostructure arrays for highly-efficient energy storage and conversion. Nano Energy 13, 790–813.

(101). Khorkova O , and Wahlestedt C (2017) Oligonucleotide therapies for disorders of the nervous system. Nat. Biotechnol 35, 249–263.28244991

(102). Kool ET (1997) Preorganization of DNA: Design principles for improving nucleic acid recognition by synthetic oligonucleotides. Chem. Rev 97, 1473–1487.11851456

(103). Hsieh W-C , Bahal R , Thadke SA , Bhatt K , Sobczak K , Thornton CA , and Ly DH (2018) Design of a "mini" nucleic acid probe for cooperative binding of an RNA-repeated transcript assocaited with myotonic dystrophy type 1. Biochemistry 57, 907–911.29334465

(104). Zhou P , Wang M , Du L , Fisher GW , Waggoner A , and Ly DH (2003) Novel binding and efficient cellular uptake of guanidine-based peptide nucleic acids (GPNA). J. Am. Chem. Soc 125, 6878–6879.12783535

(105). Sahu B , Chenna V , Lathrop KL , Thomas SM , Zon G , Livak KJ , and Ly DH (2009) Synthesis of conformationally preorganized and cell-permeable guanidine-based gamma-peptide nucleic acids (gammaGPNAs). J. Org. Chem 74, 1509–1516.19161276

(106). Thomas SM , Sahu B , Rapireddy S , Bahal R , Wheeler SE , Procopio EM , Kim J , Joyce SC , Contrucci S , Wang Y , et al. (2013) Antitumor effects of EGFR antisense guanidine-based peptide nucleic acids in cancer models. ACS Chem. Biol 8, 345–352.23113581

(107). Tartari M , Gissi C , Lo Sardo V , Zuccato C , Picardi E , Pesole G , and Cattaneo E (2008) Phylogenic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. Mol. Biol. Evol 25, 330–338.18048403

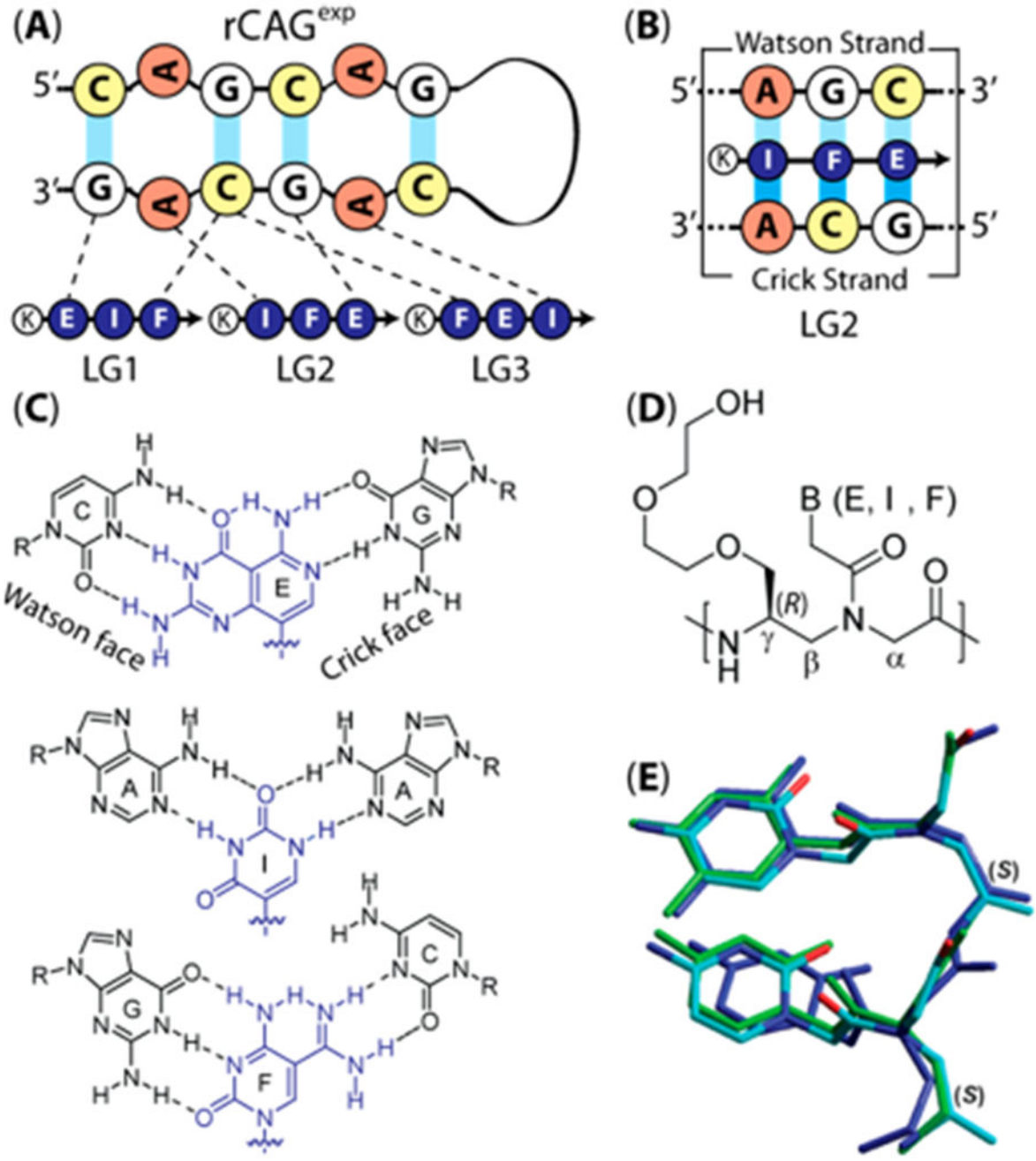


Figure 1.

(A) Design of JB-MPγPNA ligands LG1–LG3 for targeting rCAG^{exp} hairpin structure. K denotes L-lysine, and the arrow denotes the N-terminus. (B) Representative binding mode of LG2 in the preferred antiparallel orientation (N-terminus facing the 3′-end of the Watson strand). (C) H-Bonding interactions of E, I, and F with the respective C-G, A-A, and G-C base pairs of RNA. (D) Chemical structure of the MPγPNA backbone to which Janus bases were covalently attached. (E) Superposition of γPNAs [as a single strand (blue, determined

by NMR) and as a γPNA–DNA duplex (cyan, determined by X-ray diffraction)] and PNA [as a PNA–DNA duplex (green, determined by X-ray)].

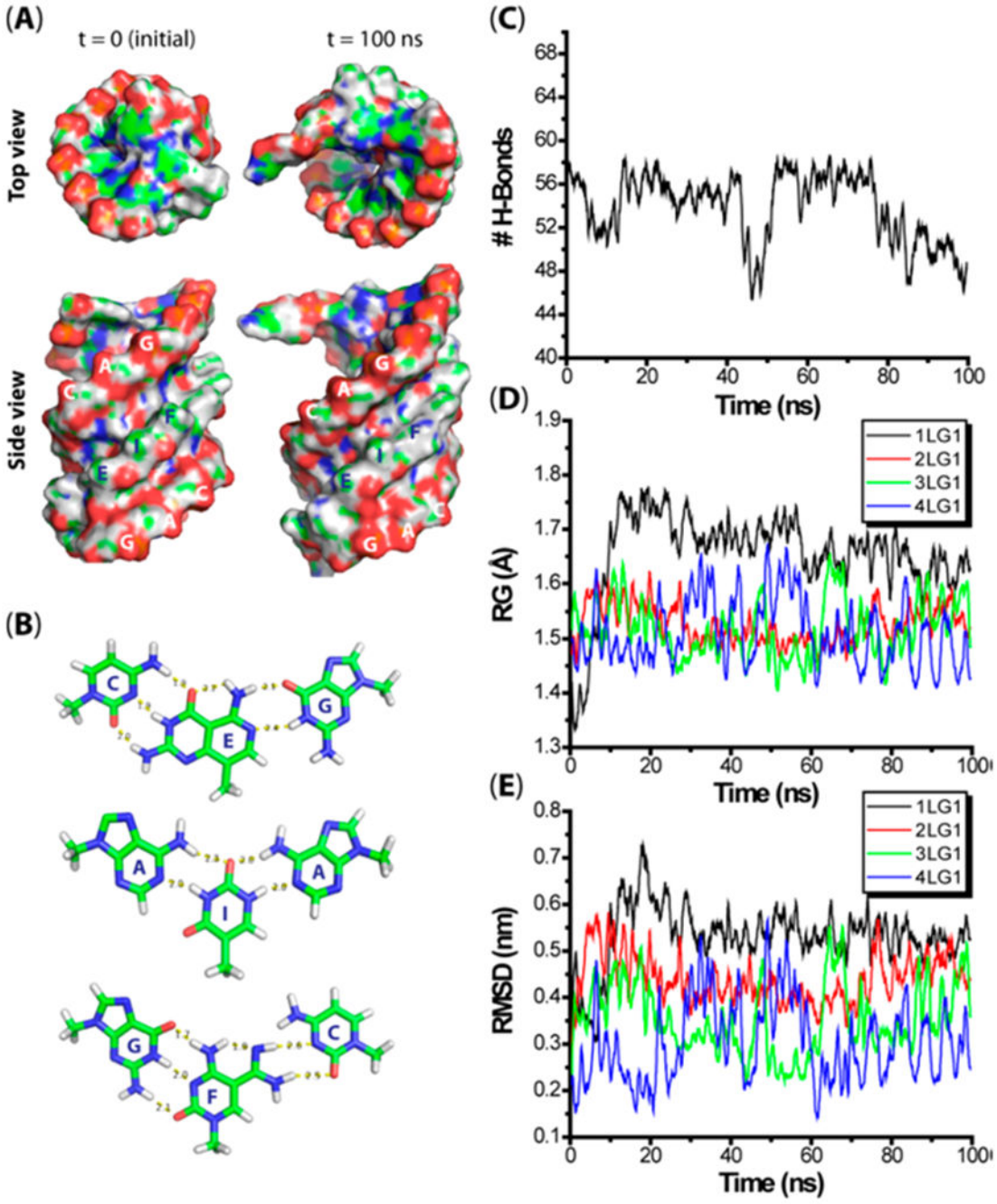


Figure 2.

Results of MD simulations of binding of LG1 to an RNA duplex containing the sequence r(CAG)$_4$/r(CAG)$_4$. LG1 is NH_2-EIF-H, with the γ-side chain being a methyl group. (A) Surface representation of the bound complex with four separate LG1 ligands and an RNA duplex containing four CAG repeats before ($t = 0$) and after ($t = 100$ ns) the simulation. (B) H-Bonding interaction of the CEG, AIA, and GFC triads after the simulation. (C) Number of H-bonds per triad between the RNA bases and Janus bases of LG1. (D) Radii of gyration of

the LG1–RNA complex. (E) Root-mean square deviation (RMSD) of the LG1–RNA complex with respect to the initial structure.

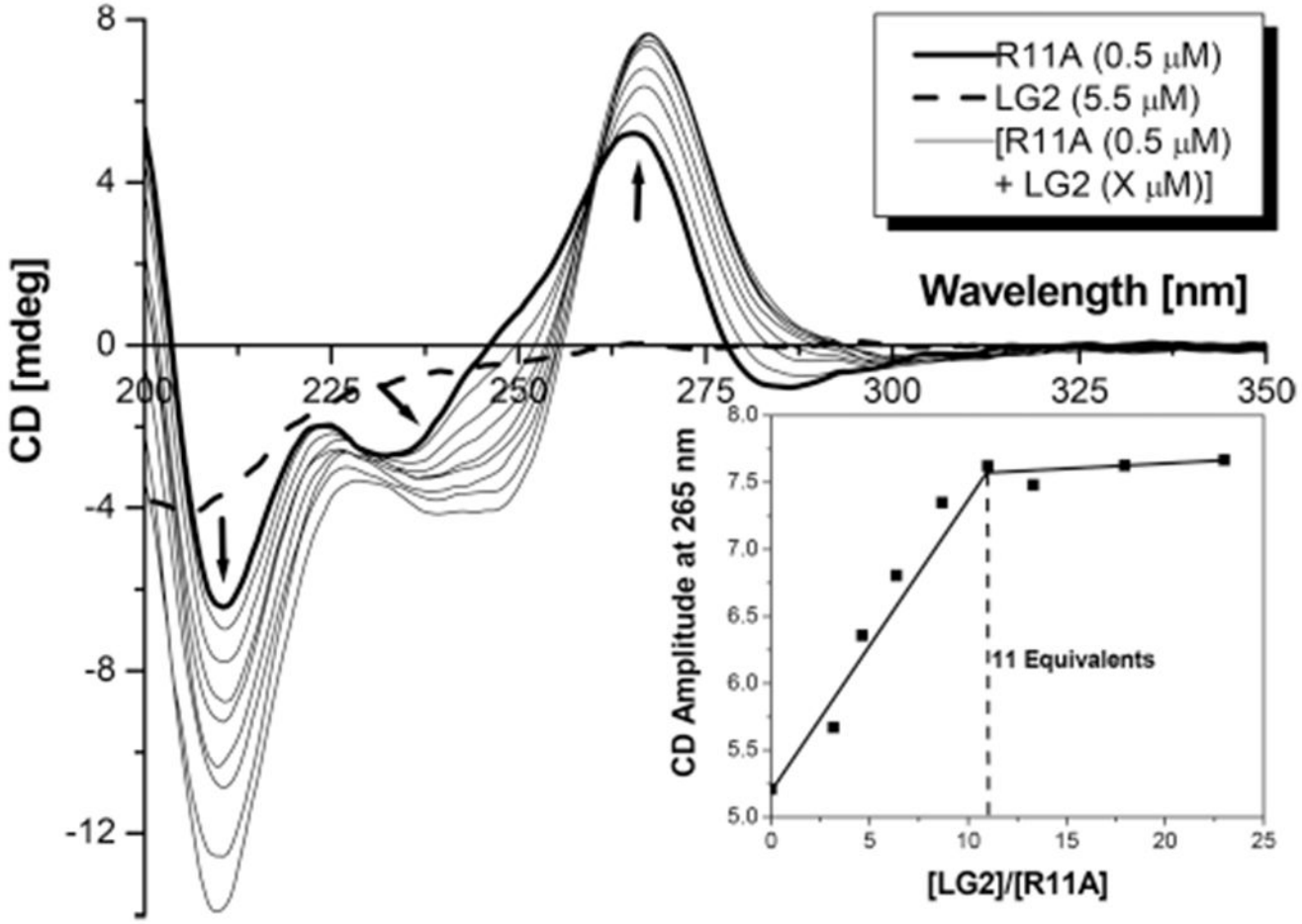


Figure 3.

CD titration of LG2 with R11A. The concentration of R11A was 0.5 μM, and those of LG2 were 0, 1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 8.5, and 11.5 μM. After the addition of LG2 to R11A, the mixtures were incubated at 37 °C for 30 min prior to the acquisition of data at 25 °C.

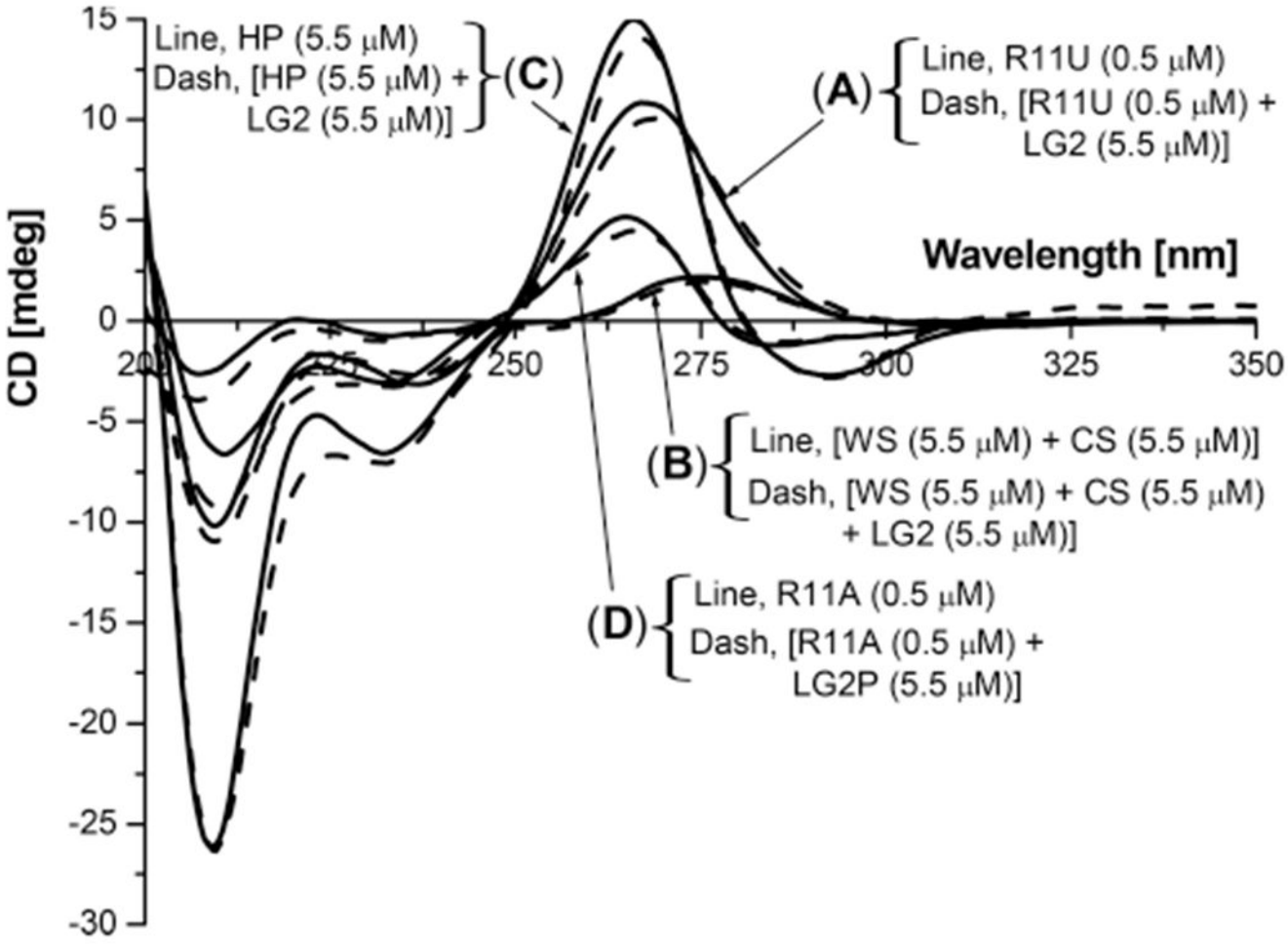


Figure 4.

CD spectra of LG2 with the (A) mismatched R11U, (B) single-stranded WS and CS, and (C) double-stranded HP containing a single binding site and (D) the spectrum of LG2P with R11A. CD spectra of RNAs (solid lines) and that of the [RNA+ligand] mixtures (dashed lines).

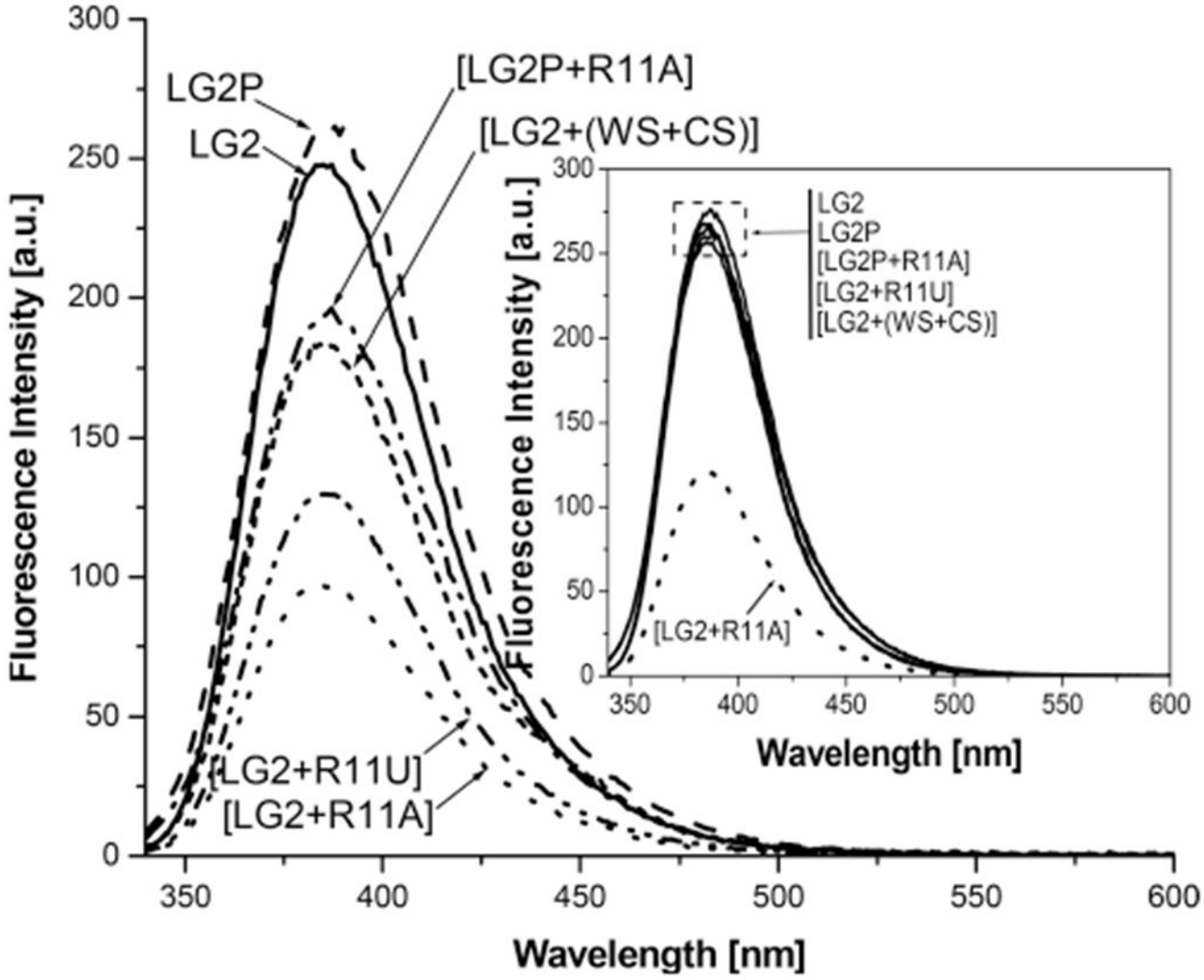


Figure 5.

Effects of ligand orientation (LG2P), mismatched R11U, and single-stranded targets (WS +CS). The inset shows the fluorescent spectra of the corresponding samples after the addition of 2.5 μM pentamidine.

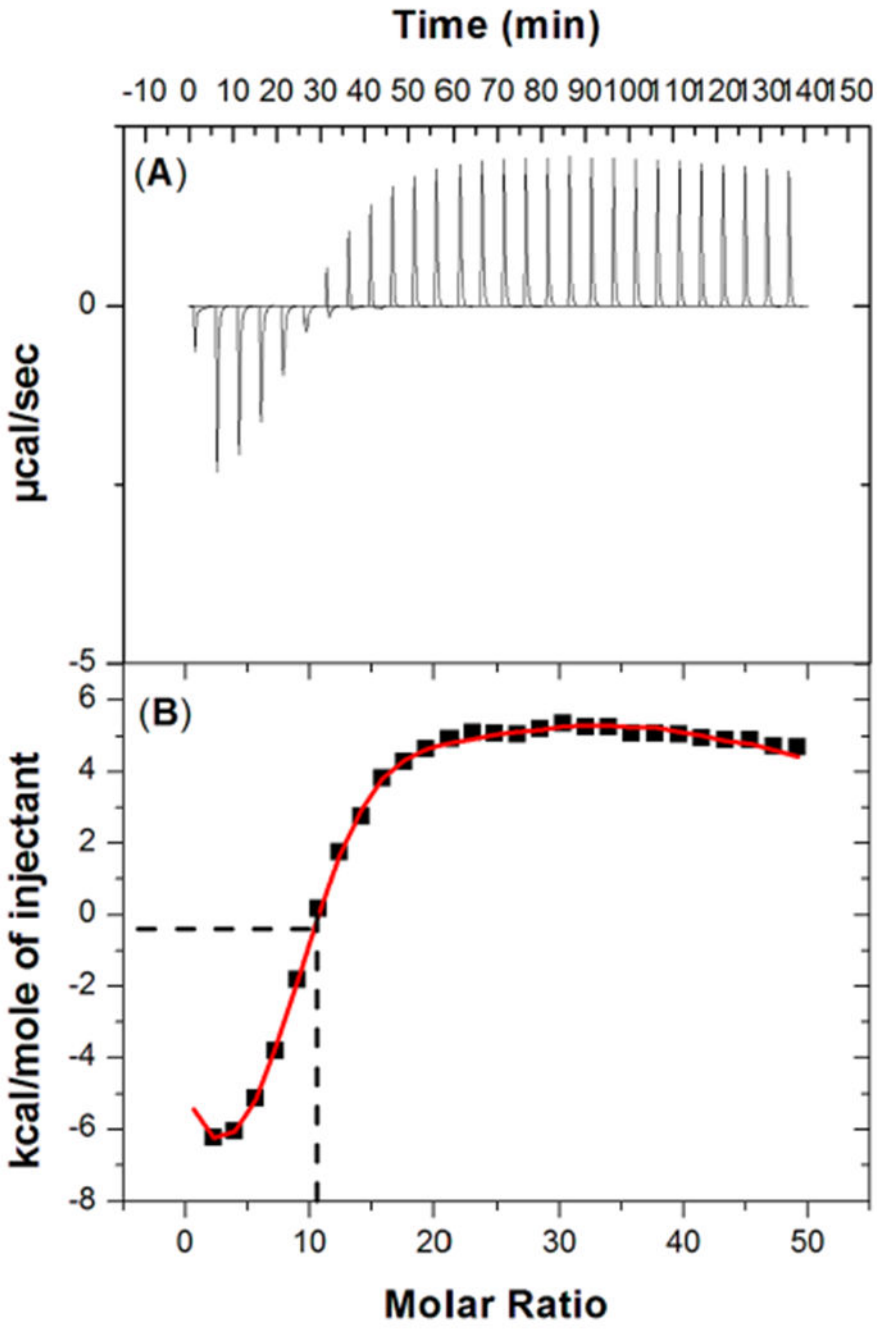


Figure 6.
(A) ITC thermogram of binding of LG2 to R11A and (B) the corresponding binding isotherm.

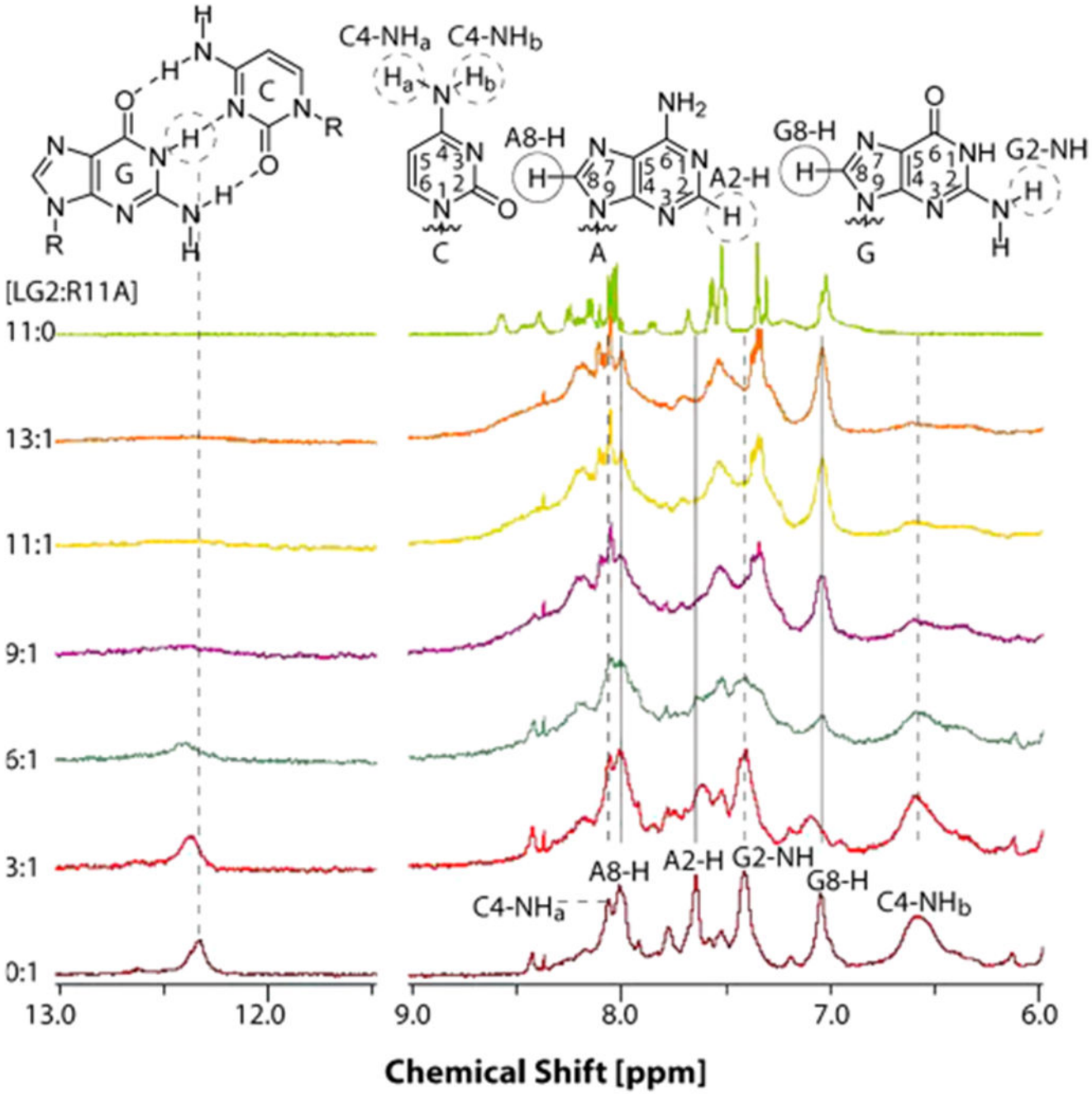


Figure 7.

NMR titration of LG2 with R11A The concentration of R11A was 0.1 mM. Aliquots of LG2 at the indicated molar ratios were added to R11A and incubated at 37 °C for 15 min prior to the acquisition of data. Solvent-exchangeable protons are indicated by dashed lines and nonexchangeable protons by vertical solid lines.

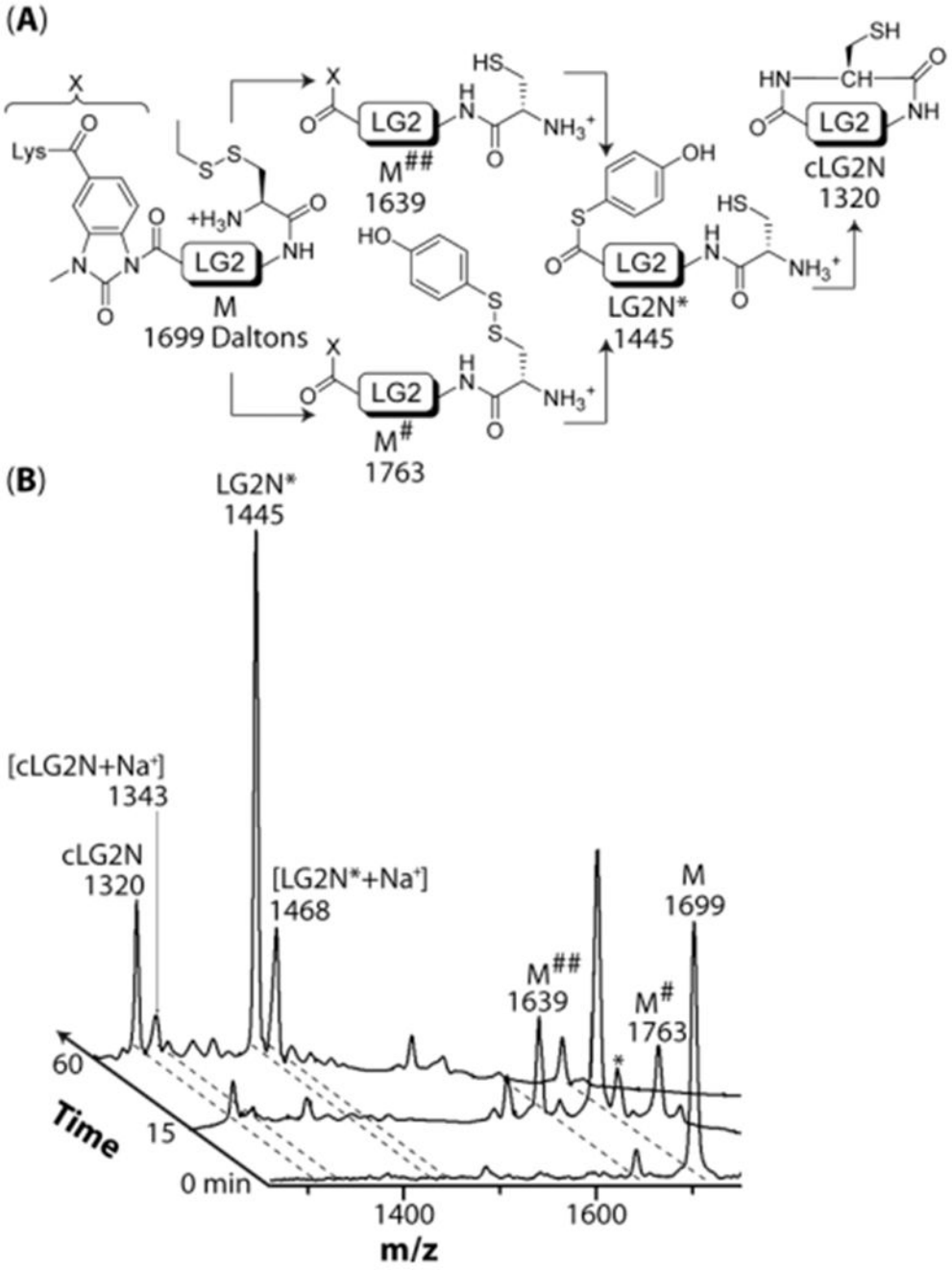


Figure 8.

Reaction progress of the parent compound M following the addition of 4MP at 37 °C. (A) An outline of the reaction sequence. The $M^{\#}$ and $M^{\#\#}$ intermediates were initially formed and then converted into the reactive LG2N* intermediate and finally to the cyclic product cLG2N. (B) MALDI-TOF MS spectra of M following the addition of 4MP and incubation at 37 °C for 0, 15, and 60 min. The samples were prepared in 0.1× PBS buffer.

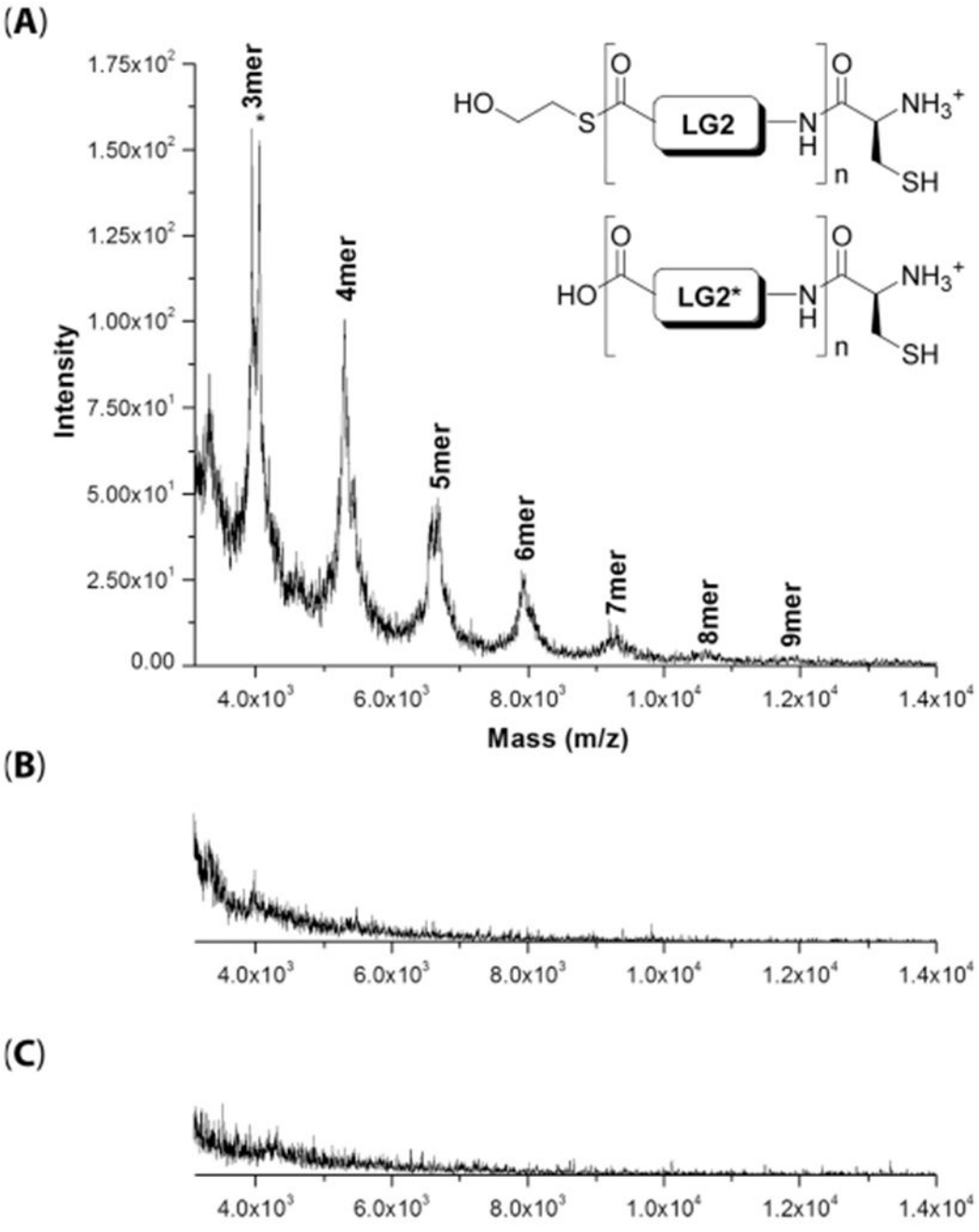


Figure 9.

MALDI-TOF spectra of (A) [LG2N+R11A], (B) [LG2N+R11U], and (C) LG2N alone following the incubation with 2ME at 37 °C for 16 h. Prior to MS analysis, the samples were reconstituted with 4 mM guanidinium chloride and heat-denatured at 95 °C for 5 min. The concentrations of LG2N, R11A, R11U, and 2ME were 22, 1, 1, and 500 μM, respectively. In the positive mode, RNA molecules were not observed in the MALDI-TOF spectra. The *Y*-axes in panels B and C are the same as that of panel A; they were omitted to save space.

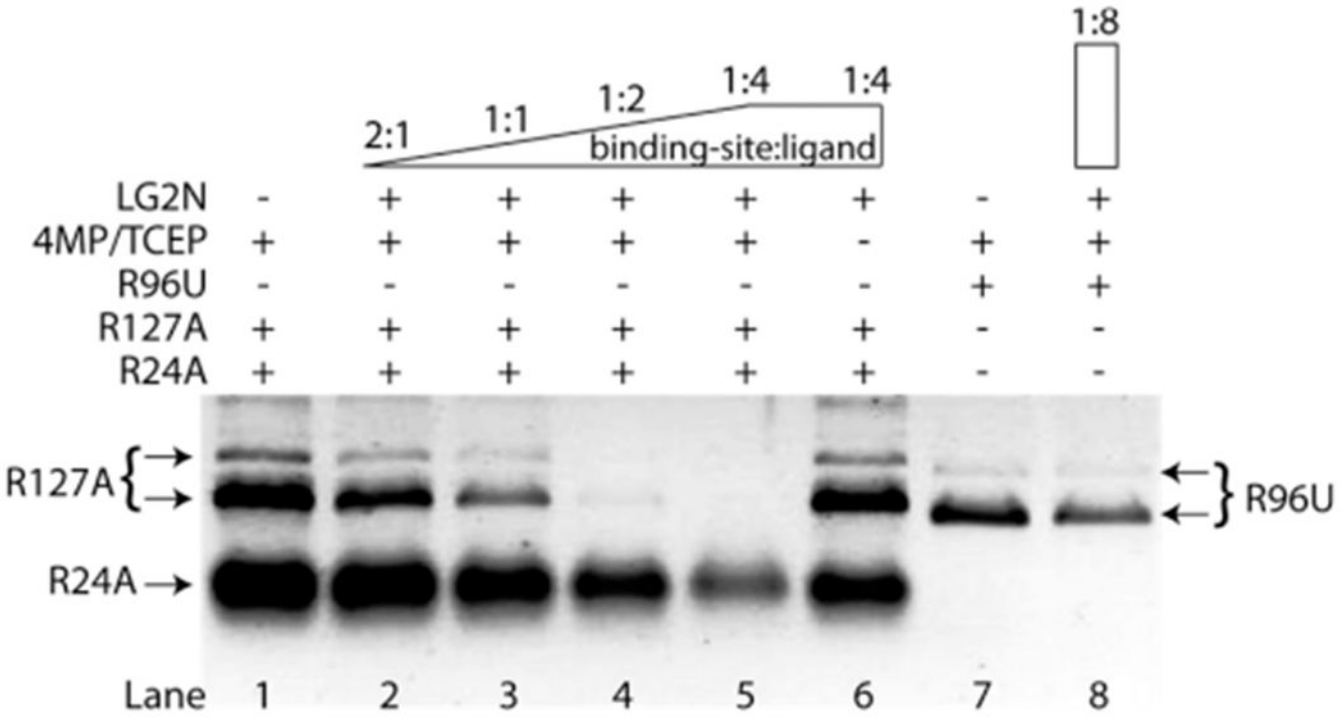
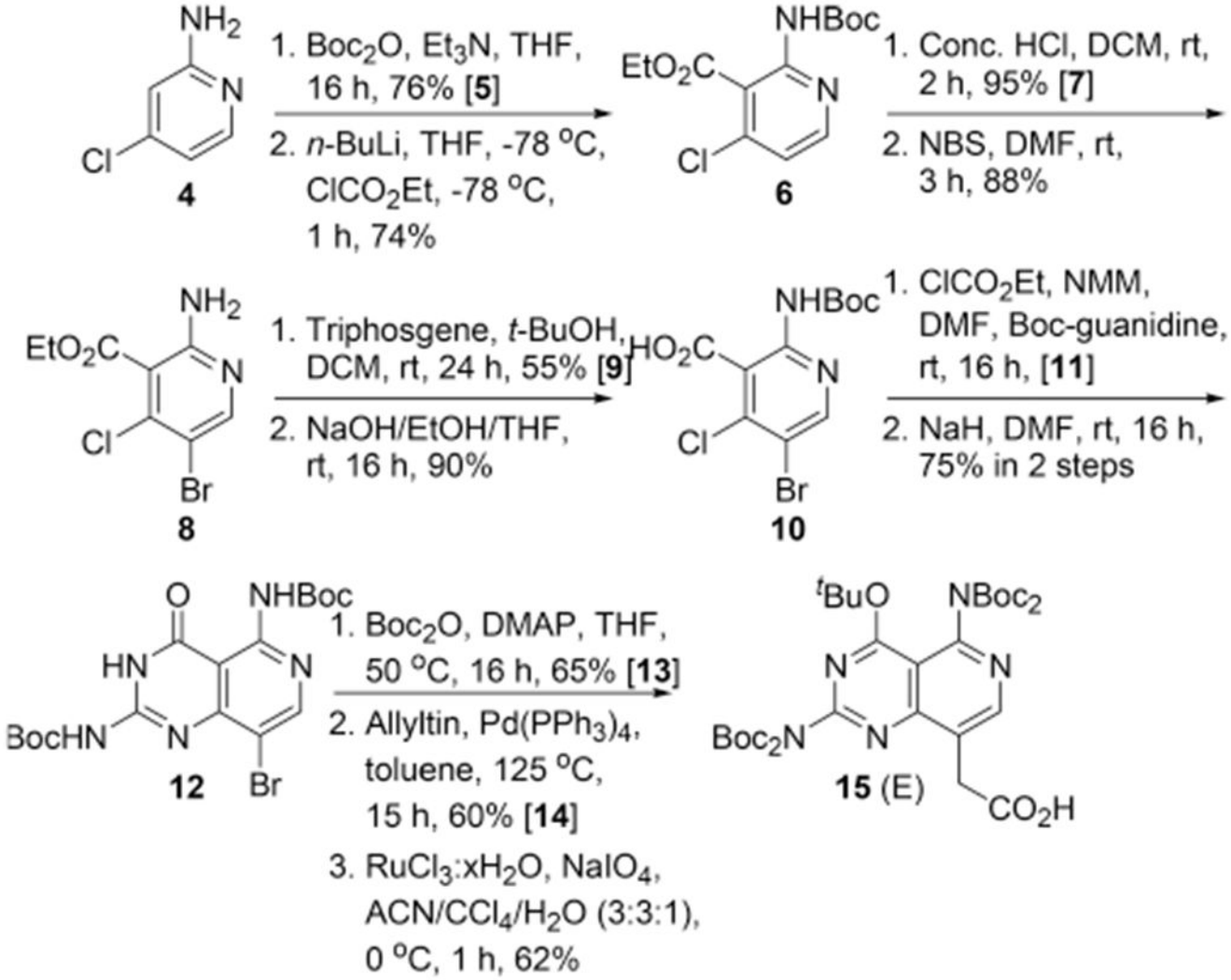
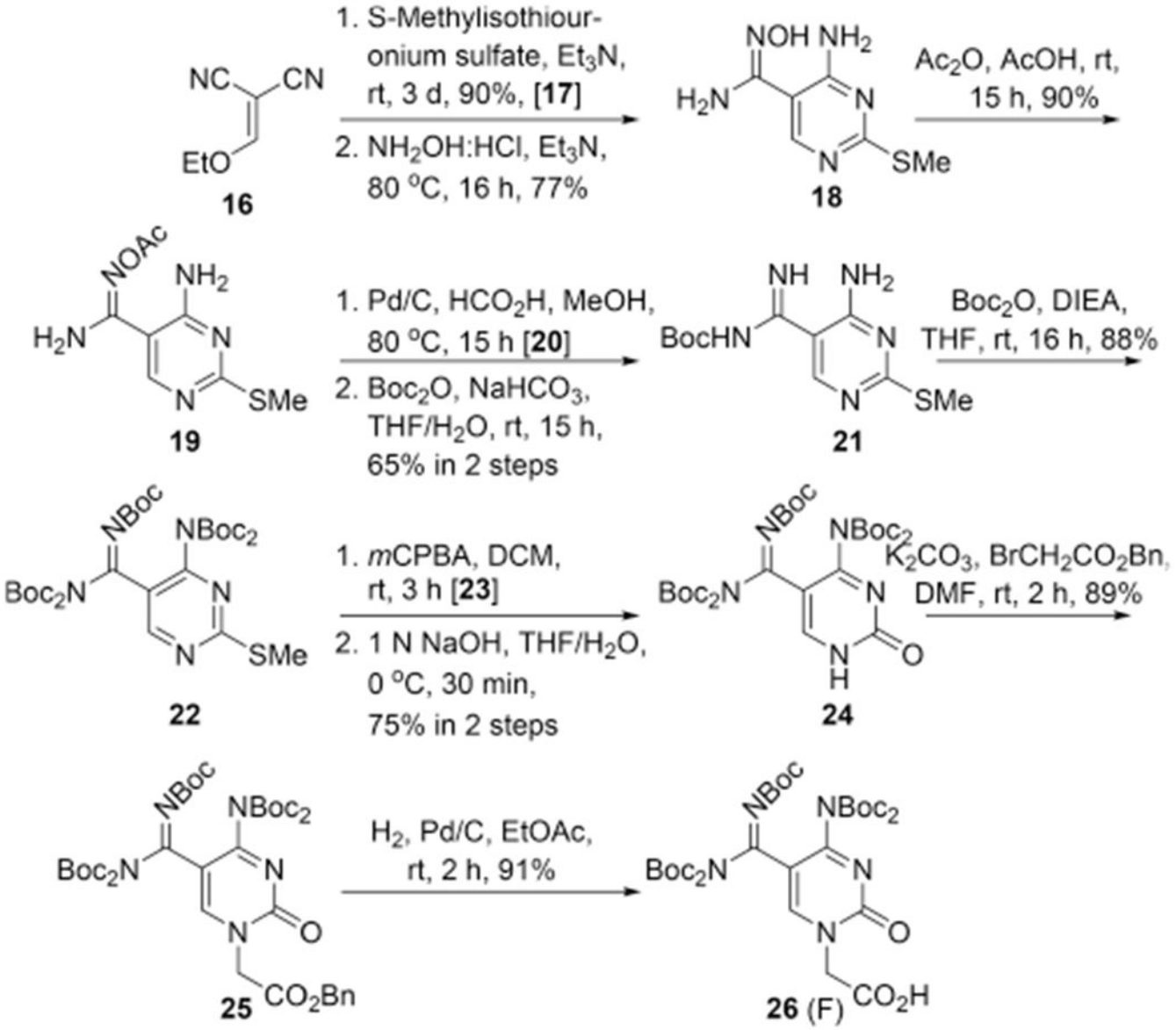


Figure 10.

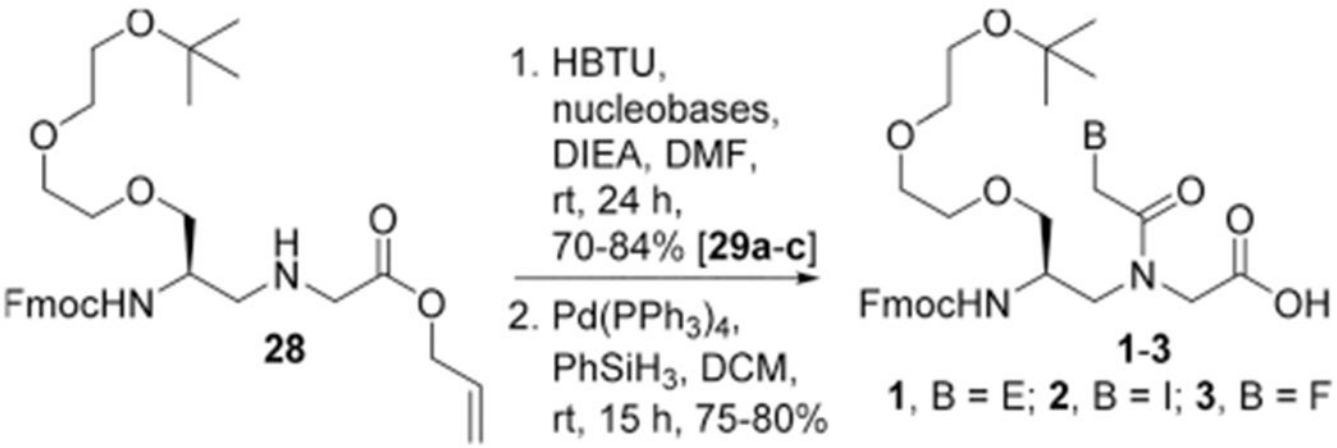
Competitive binding assay. The concentrations of R24A, R127A, and R96U were 100, 18, and 22 nM, respectively, with each containing 1.1 μM binding sites; the concentrations of LG2N in lanes 2–6 and 8 were 11, 22, 44, 88, 88, and 88 μM, respectively, and those of 4MP and TCEP were 500 and 100 μM, respectively. The ratios of binding site to ligand were as indicated. The samples were prepared in 0.1× PBS buffer and incubated at 37 °C for 24 h prior to separation on a 2% agarose gel and stained with SYBR-Gold.



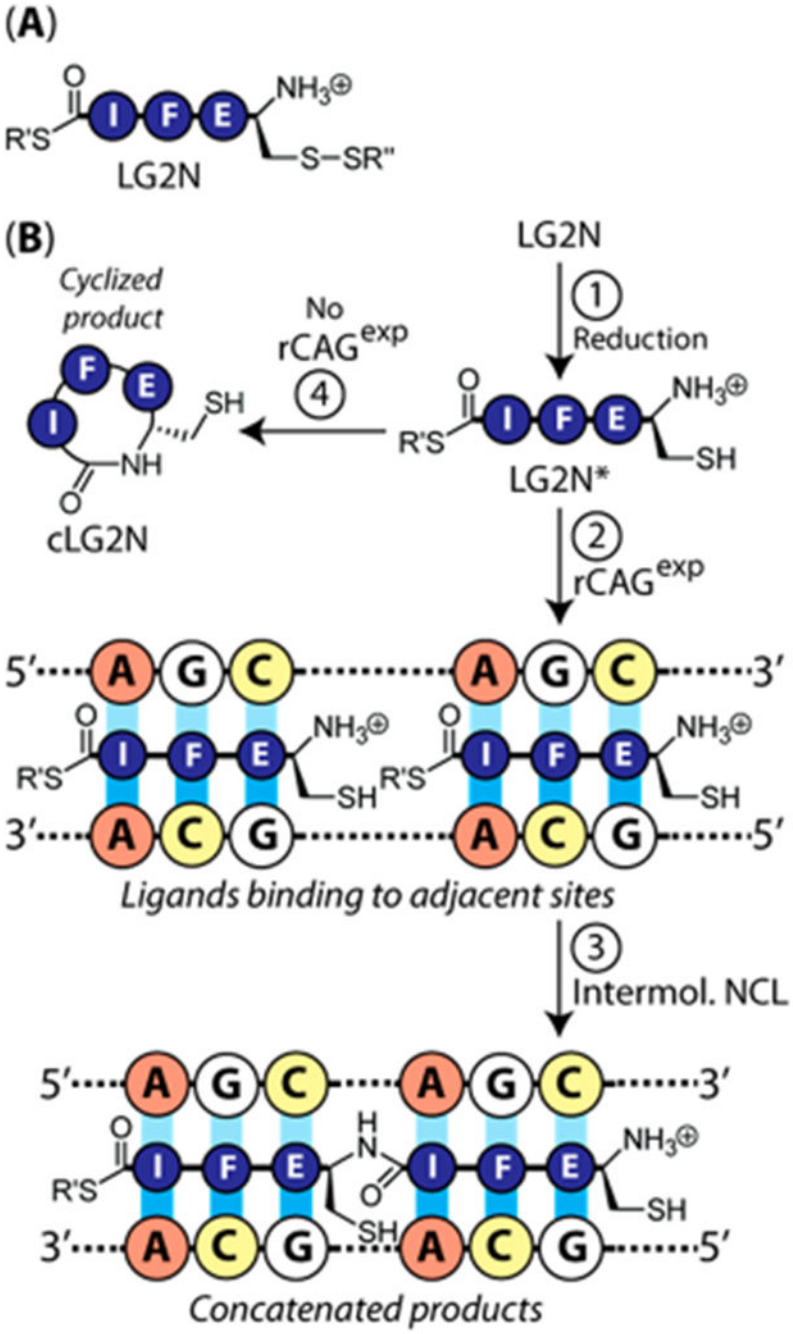
Scheme 1.



Scheme 2.



Scheme 3.



Scheme 4.

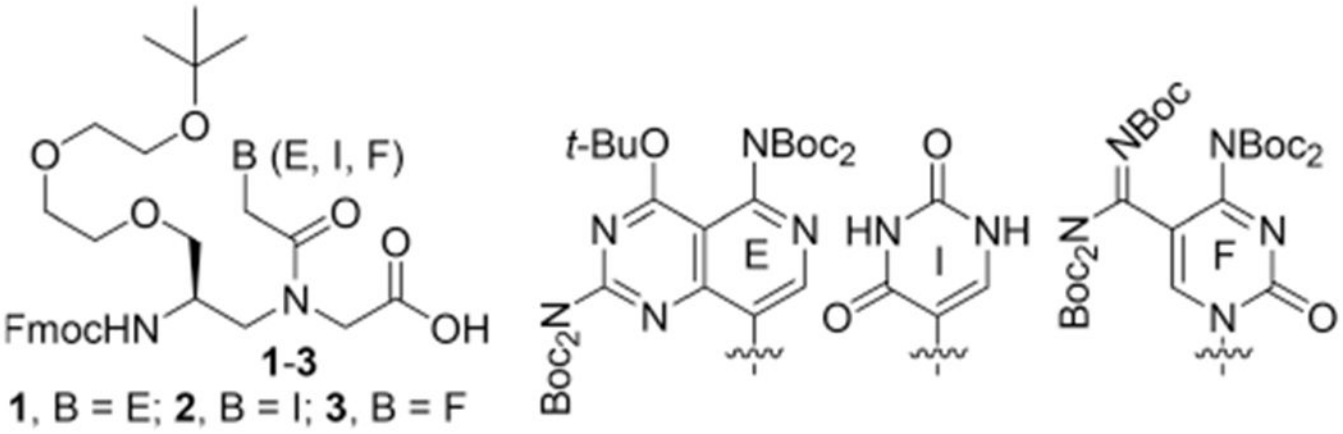


Chart 1.

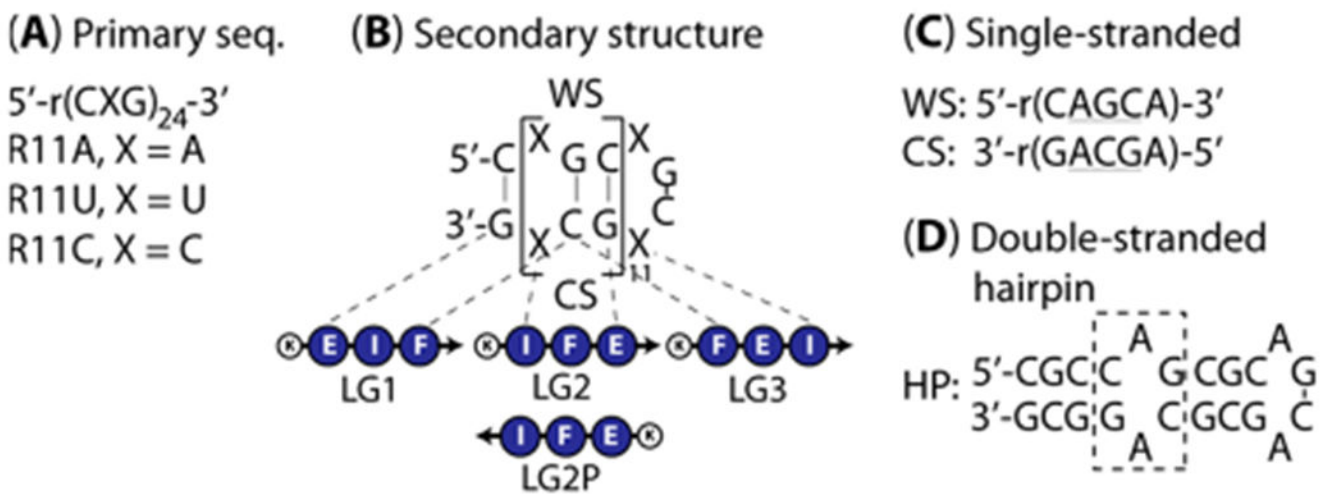


Chart 2.

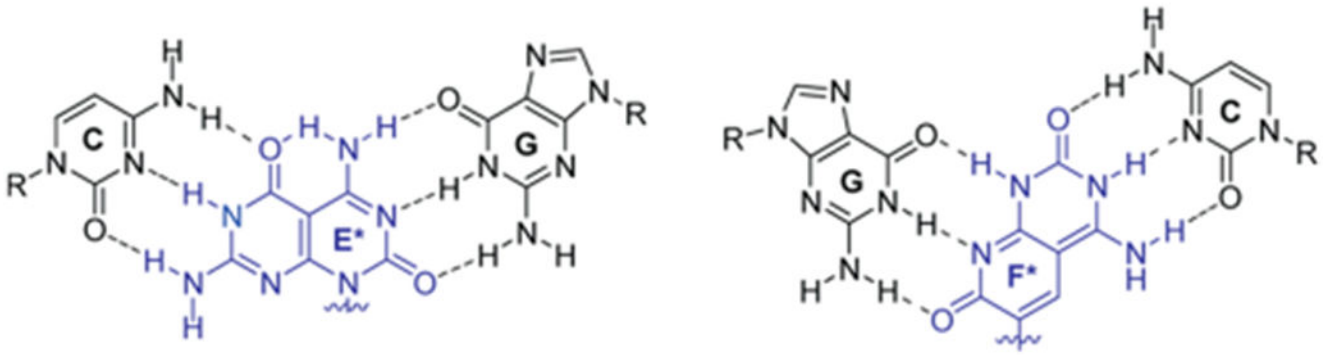


Chart 3.