Cyclic mismatch binding ligands interact with disease-associated CGG trinucleotide repeats in RNA and suppress their translation

Patryk Konieczny[1,2]*, Sanjukta Mukherjee[3,4]*, Ewa Stepniak-Konieczna[1], Katarzyna Taylor[1], Daria Niewiadomska[1], Agnieszka Piasecka[1], Agnieszka Walczak[1], Anna Baud[1], Chikara Dohno[3], Kazuhiko Nakatani[3] and Krzysztof Sobczak[1]

[1] Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland

[2] Institute of Human Biology and Evolution, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland

[3] Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, 567-0047, Japan

[4] National Centre for Biological Sciences (NCBS), Tata Institute of Fundamental Research (TIFR), Bellary Road, Bangalore 560065, Karnataka, India

* equally contributed authors

Correspondence should be addressed to KS. Tel: +48 61 829 5950; Fax: +48 61 829 5949; Email: ksobczak@amu.edu.pl

Key words: fragile X-associated tremor/ataxia syndrome, FXTAS, repeat-associated non-AUG translation, RAN translation, CGG repeats, cyclic mismatch binding ligands

Abstract

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset neurodegenerative disorder caused by a limited expansion of CGG repeats in the *FMR1* gene. Degeneration of neurons in FXTAS cell models can be triggered by accumulation of polyglycine protein (FMRpolyG), a by-product of translation initiated upstream to the repeats. Specific aims of our work included testing if naphthyridine-based molecules could (1) block FMRpolyG synthesis by binding to CGG repeats in RNA, (2) reverse pathological alterations in affected cells and (3) preserve the content of FMRP, translated from the same *FMR1* mRNA. We demonstrate that cyclic mismatch binding ligand CMBL4c binds to RNA structure formed by CGG repeats and attenuates translation of FMRpolyG and formation of nuclear inclusions in cells transfected with vectors expressing RNA with expanded CGG repeats. Moreover, our results indicate that CMBL4c delivery can reduce FMRpolyG-mediated cytotoxicity and apoptosis. Importantly, its therapeutic potential is also observed once the inclusions are already formed. We also show that CMBL4c-driven FMRpolyG loss is accompanied by partial FMRP reduction. As complete loss of FMRP induces FXS in children, future experiments should aim at evaluation of CMBL4c therapeutic intervention in differentiated tissues, in which FMRpolyG translation inhibition might outweigh adverse effects related to FMRP depletion.

Introduction

Fragile X-associated tremor/ataxia syndrome (FXTAS) is a late onset neurodegenerative disorder caused by a limited expansion (premutation) of the polymorphic CGG repeat tract in the 5' untranslated region (5'UTR) of the *FMR1* gene (1, 2). The number of CGG repeats in the mRNA from premutation alleles ranges from 55 to 200. Typical FXTAS clinical symptoms manifest in patients older than 50 years of age and include intention tremor, gait ataxia, parkinsonism and cognitive deficits. Magnetic resonance imaging and neuropathological examination revealed the underlying white matter disease and the overall brain atrophy (3–6), while further post-mortem brain analyses showed intra-nuclear eosinophilic and ubiquitin-positive inclusions in neurons and astrocytes throughout the cerebrum and brainstem, predominantly in the hippocampus. Despite the apparent loss of Purkinje cells, no or rare inclusions were observed in the remaining Purkinje cell population (3, 7), raising the question about the toxicity versus protective function of the inclusions. Premutation has relatively high incidence in humans that ranges from 1:430-850 in men to 1:150-300 in women. However, the penetrance is incomplete, with the current estimate of 40-75% males and 16-20% females to develop the full-blown FXTAS symptoms (8–10). Additionally, about 20% of the female carriers develop fragile X-associated primary ovarian insufficiency [FXPOI, (11)].

While *FMR1* mRNA content is elevated in FXTAS patients, FMRP translated from *FMR1* is usually slightly decreased (4, 5, 7, 12–14). Based mainly on the cellular toxicity and accumulation of the RNA from premutation alleles in the nucleus (15–18), a concept was put forward, in which neurodegeneration was attributed to sequestration of specific proteins onto the *FMR1* mRNA containing expanded CGG repeats. Indeed, over the years many proteins such as hnRNP A2/B1, PURα, SAM68, DROSHA and DGCR8 were found to be specifically trapped on expanded CGG repeats (CGG_{exp}) interacting either directly or indirectly with the RNA (16, 17, 19, 20). Counteracting this phenomenon with overexpression of some of the proteins was shown to amend the apparent toxicity observed in FXTAS cell models (17, 19, 20). A second concept of neurodegeneration in FXTAS is based on the fact that CGG_{exp} within the 5'UTR of *FMR1* transcript give rise to biosynthesis of homopolymeric proteins that can induce cell death (21–23). Translation of toxic peptides takes place in FXTAS because the mRNA from premutation alleles is efficiently transported to the cytoplasm, especially once the repeats are placed within their natural *FMR1* context (21, 23, 24). Competitive translation of FMRP and other peptides and proteins from the same *FMR1* transcript might also explain slight decreases of the FMRP content despite the elevated *FMR1* mRNA levels in FXTAS cells (23, 25).

Synthesis of homopolymeric proteins from *FMR1* and *ASFMR1* antisense transcripts can be carried out by: 1) repeat-associated non-AUG (RAN) translation initiated directly from CGG_{exp} (21, 22, 26), 2) RAN translation initiated with near-cognate start codons located upstream to the repeats as a result of 43S preinitiation complex stalling at CGG_{exp}, and 3) repeat expansion-independent translation of proteins containing either long or short polyglycine tracts (FMRpolyG) from a near-cognate ACG start codon embedded in the Kozak sequence upstream to the CGG repeats (23, 27). The latter is supported by the fact that once a GFP-stabilizing tag (see for example (28)) is fused to FMRpolyG, the protein is readily detectable, even when the number of CGG repeats is within the normal range (21). This indicates that shorter, presumably non-aggregating homopolymeric peptides can be efficiently degraded by the cell without any adverse effect (23). On the contrary, results from model systems indicate that the stable aggregated form of FMRpolyG translated from the mRNA from premutation alleles is exceptionally toxic (21–23). Particularly, Sellier et al. showed that FMRpolyG interaction with LAP2β triggers nuclear envelope disruption and cell death (23).

Although the contribution of FMRpolyG to pathology in FXTAS is still debatable (18, 29), there is increasing evidence that its high concentration in different model systems leads to cell toxicity (21–23). By the same token, downregulation of FMRpolyG could be a vital therapeutic goal as well as a read-out for successful treatment. Small molecules and modified antisense oligonucleotides targeting RNA with expansion of triplet repeats stand out as prospective therapeutic agents in many genetic diseases (30–32). Importantly, their mode of action could be designed to either block specific sites on the target mRNA or to induce transcript degradation. Similarly, in FXTAS they could be used to bind specific sites within *FMR1* mRNA from premutation alleles to induce the transcript degradation or block FMRpolyG translation (18, 25, 33–36). Some of these approaches would also result in freeing trapped molecules from the RNA repeats and releasing them to their natural environments, additionally contributing to the restoration of the cellular homeostasis.

One putative caveat of the strategies based on small molecule and antisense oligonucleotide delivery to FXTAS cells is that both FMRpolyG and FMRP are expressed from the same transcript and that downregulation of FMRpolyG might result in a parallel downregulation of FMRP protein (25, 34, 36). Specifically, targeted cleavage and removal of *FMR1* mRNA would result in equivalent loss of both proteins. This is of particular importance as FMRP loss is a hallmark of fragile X syndrome (FXS), an early-onset disorder characterized by alterations in physical appearance, autism and intellectual disabilities (37). In contrast to

FXTAS-related increases in transcription and the *FMR1* mRNA content, synthesis of *FMR1* mRNA in FXS patients is shut down due to CGG triplet cassette expansion to over 200 repeats and the consequential promoter region hypermethylation. The complete absence of FMRP results in dendritic spine changes and alterations in synaptic plasticity as well as changes in connective tissue due to the FMRP involvement in site specific translation of proteins in neurons (37). It is important to note, however, that such severe FXS symptoms might not be observed in established tissues of adult patients.

Targeting RNA structure by small molecules is still challenging due to the lack of fundamental understanding of how these interactions occur. Disney et al. reported molecules targeting RNA CGG repeat that could regulate RAN translation and RNA toxic gain of function (33, 34). Their bis-benzimidazole derivatives, known as typical minor groove binders of DNA, inhibit synthesis of FMRpolyG without affecting the downstream canonical translation (34). Recently, a G-quadruplex binder showed an inhibitory effect on FMRpolyG synthesis (35, 38), and some different classes of RAN translation inhibitors, including non-repeat binders, were identified by high-throughput screening (35). Our group has been focusing on a rational design approach to develop mismatch-binding ligands (MBLs) that target specific noncanonical structure of nucleic acids through complementary hydrogen bonding (39–42). Compounds used in this study were designed to recognize an array of Gs in the 5'-CGG-3'/5'-CGG-3' unit by complementary hydrogen bonding to the Watson-Crick surface (41–43). The different modes of binding alter the structure of RNA CGG repeat and its thermodynamic/kinetic stability, which could have significant effects on interactions with RNA binding proteins and the resulting biological events.

A series of small molecules designed to recognize an array of Gs have been synthesized, including naphthyridine carbamate dimer (NCD; (39)), a naphthyridine tetramer containing two NCD molecules (Z-NCTS; (40, 44)), and cyclic mismatch binding ligands (CMBLs) that consist of cyclophane containing bis(2-amino-1,8-naphthyridine) moieties connected by variable linkers (41, 42). Importantly, Z-NCTS was shown to bind to 5'-CGG-3'/5'-CGG-3' DNA sequence *in vitro* with 1:1 binding stoichiometry and a dissociation constant (K_d) of ca. 100 nM (40). We have demonstrated that CMBLs could bind to DNA repeats, including $(TGG)_8$, $(CCTG)_9$ and $(CAG)_9$ (41), and CAG RNA repeats (43). Synthesis of other CMBLs differing in molecular composition of the variable linker allowed to obtain ligands showing selective orthogonal binding relationships with DNA and RNA repeats. In the present study, we estimated therapeutic potential of NCD, Z-NCTS and 21 CMBL compounds.

Particularly, we screened their binding affinity to RNA CGG repeats *in vitro* and tested their ability to inhibit FMRpolyG translation in FXTAS cell line models.

Material and Methods

Surface Plasmon Resonance (SPR) assay

Surface plasmon resonance (SPR) single cycle kinetics assay was performed using Biacore T200 platform (GE Healthcare, Life Science). 5'-Biotin-TEG DNA or RNA sequences were immobilized on the Series S sensor chip SA surface using avidin-biotin coupling in HBS-N running buffer (HBS-N; 0.01 M, HEPES, 0.15 M NaCl, pH 7.4). Five sensor chips containing immobilized RNA [r(CGG)$_9$; SA chip-1, -2, -3, -4] or DNA [d(CGG)$_9$; SA chip-5] (Supplementary Table S1) were used to perform SPR assays for NCD, Z-NCTS and 21 CMBL compounds. Another five sensor chips (SA chip-6, -7 -8, -9 and -10) with immobilized rSS$_{30}$ (5'-biotin-TEG-CUUAGUACCAUUAUAGAUUUACCAUGAUUC-3'), rDS$_{30}$ (5'-biotin-TEG-CAGUAGUAGUAGUUUUUACUACUACUACUG-3'), or other repeat DNA or RNA sequences (Supplementary Table S2) were used for SPR assays for CMBL4c.

Amounts of immobilized RNA and DNA oligomers on each sensor surface are summarized in Supplementary Table S1 and S2. 5'-Biotinylated oligonucleotides were diluted to 0.2 μM in 10 mM HEPES-500 mM NaCl and injected to reach the response of around 500 RU. Blank immobilization was performed in the flow cell 1 for reference subtraction. Ligand solution was diluted using HBS-EP+ buffer (0.01 M HEPES, 0.15 M NaCl, 3.0 mM EDTA, pH 7.4, 0.005% v/v Surfactant P20). 1.0 mM CMBL stock solutions in DMSO were diluted to obtain the final ligand solutions that contained 5% DMSO. 1xHBS-EP+ (for NCD and Z-NCTS) and 5% DMSO containing 1xHBS-EP+ (for CMBLs) were used as running buffers for binding assays. Sensorgrams were obtained with the concentration range of 0.5 to 8.0 μM for NCD and 0.25 to 4.0 μM for Z-NCTS and 21 CMBL analogues. All sensorgrams were corrected by reference subtraction of blank flow cell response and buffer injection response. SPR sensorgram obtained for compounds using SA chip-1, -2, -4 and -5 were further standardized with respect to the immobilization amount (RU538) of (CGG)$_9$ on the SA chip-3. Regeneration of the sensor surface was carried out using 50% aqueous solution of DMSO and 50 mM aqueous solution of NaOH. Apparent binding constants of NCD, Z-NCTS and CMBL4a were determined from SPR curves applying 1:1 fitting model (Supplementary Table S3).

Filter Binding Assay (FBA)

The r(CGG)$_{20}$, r(AGG)$_{20}$ and r(CUG)$_{20}$ transcripts were gifts from W. Krzyżosiak (Polish Academy of Science). For 5' radiolabeling, 2 pmol of transcripts was incubated with 2 pmol (12 mCi) of [γ-P^{32}] ATP, 1 U RNasin®Plus RNase Inhibitor, 10 U OPTI Kinase (Affymetrix), 1x reaction buffer (Affymetrix) and dd$H_2$0 up to 10 µl, at 37°C for 30 min. The labeled RNA was run on a 8 % PA gel (29:1) in 0.5 x TBE, at 100 V for 60 min. The band of RNA was visualized on IP through FLA-1500 (FujiFilm), cut out followed by ethanol precipitation and resuspended in 20 µl ddH_2O. FBA was carried out by incubating 5' radiolabeled RNA (1 nM) with small compounds of indicated concentrations (ranging from 0 to 20 µM). Reactions were performed in a volume of 30 µl, in 1x buffer FBA (50 mM NaCl, 50 mM KCl, 50 mM TRIS-HCl, 1 mM $MgCl_2$, pH 8.0) and incubated at 37°C for 15 min. Next, 25 µl of each sample was loaded under vacuum induced pressure followed by a wash step with 100 ul of buffer FBA onto two membranes, previously wetted in buffer FBA and placed between the apparatus frames (dot blotter). Positively charged complexes of RNA and small compounds remained on the top nitrocellulose membrane (Protran BA 85, Whatman®) whereas free and negatively charged RNA, which went through the first membrane was subsequently caught on a nylon membrane (Hybond™ N+, Amersham) placed underneath. The signal from membranes was detected O/N, visualized on IP through FLA-1500 (FujiFilm) and quantified using Multi Gauge software (FujiFilm). The dissociation constant (Kd) of the RNA/small compounds interaction was calculated in the GraphPad program using the following equation: one site specific binding curve (Y = Bmax*X/(Kd + X)).

Electrophoretic Mobility Shift Assay (EMSA)

5' radiolabeled r(CGG)$_{20}$, r(AGG)$_{20}$ and r(CUG)$_{20}$ (see FBA protocol) were first subjected to 1 min denaturation step at 90°C followed by 10 min of renaturation on ice. EMSA was carried out by incubating 5' radiolabeled RNA (1 nM) with small compounds of indicated concentrations (ranging from 0 to 40 µM). Reactions were performed in a volume of 10 µl, in 1x buffer FBA (50 mM NaCl, 50 mM KCl, 50 mM TRIS-HCl, 1 mM $MgCl_2$, pH 8.0) and incubated at 37°C for 15 min. The samples were run on a native 8% PA gel in 0.5 x TBE at 100 V for 1 h. The gel was subsequently dried and the signal was detected O/N, and visualized on IP through FLA-1500 (FujiFilm). The differentiation between free RNA and RNA in complexes with small compounds was possible due to their differences in mobility. Free and negatively charged RNA migrated faster from RNA:small compound complexes which migrated much slower, most likely due to gained positive charge of tested compounds, size and structural

conformation of the complex. The dissociation constant (Kd) of the RNA/small compounds interaction was calculated for a signal of RNA bound to small compounds in the GraphPad program using the following equation: one site specific binding curve (Y = Bmax*X/(Kd + X)). The half-life of RNA/small compounds interaction was calculated for a signal of free RNA using the following equation: one phase decay (Y=(Y0 - Plateau)*exp(-K*X) + Plateau)).

Cell culture

COS7, HeLa, SH-SY5Y and HEK293T cells were grown in a high glucose DMEM medium with L-Glutamine (Lonza) supplemented with 10% fetal bovine serum (Sigma) and 1% antibiotic/antimycotic (Sigma). FXTAS (1022-07 (P3), XY, $(CGG)_{81}$; 1044-07 (F3), XY, $(CGG)_{97}$; WC26, XX, $(CGG)_{60}$/$(CGG)_{90}$) and control (1028-07 (C4), XY, $(CGG)_{22}$; C0603, XY, $(CGG)_{31}$) as well as DM1 (GM04033; $(CTG)_{1000}$) and control (GM07492) fibroblasts were cultured in EMEM medium (Lonza) supplemented with 15% fetal bovine serum (Sigma), 1% MEM non-essential amino acids (Thermo Fisher Scientific) and 1% antibiotic/antimycotic (Sigma). All cells were grown at 37°C in a humidified incubator containing 5% CO_2. FXTAS 1022-07 and 1044-07 and control 1028-07 fibroblast lines were a kind gift from P. Hagerman (see also (45)), while FXTAS WC26 and control C0603 fibroblast lines were obtained from A. Bhattacharyya (see also (46)). DM1 GM04033 and control GM07492 fibroblasts as well as HD GM04281 fibroblasts were purchased from the Coriell Institute for Medical Research.

Constructs

ATG$(CGG)_{99}$-*luc2*(+1) and 5′$(CGG)_{99}$-*luc2*(0) vectors were described previously (47). To generate 5′$(CGG)_{16}$-*luc2*(0), 5′$(CGG)_{34}$-*luc2*(0) and 5′$(CGG)_{44}$-*luc2*(0) constructs, 5′$(CGG)_{99}$-*GFP*(+1) was plated and screened for colonies containing shorter number of the repeats. After restriction digest, the number of the repeats was confirmed by Sanger sequencing. The repeats were cut out from the vector by *NheI* and pasted into the *NheI*-digested and dephosphorylated (CIAP, Invitrogen) pmirGLO modified vector containing the multicloning site in front of the *luc2* sequence (ATG-*luc2*; (28, 47)). 5′$(CGG)_{99}$ lacking a tag and containing the whole FMRpolyG sequence was generated first by cutting out *GFP* sequence from the 5′$(CGG)_{99}$-*GFP*(+1) vector with *AvrII* and *EagI* and ligating an insert constructed from two complementary phosphorylated oligonucleotides (F1/R1, Supplementary Table S4). The obtained vector was then used as a template to amplify a PCR product with F2/R2 primers

(Supplementary Table S4). The product was subsequently digested with *XhoI* and *EagI* and cloned into the digested template vector. The control plasmid containing 16 CGG repeats (5′(CGG)$_{16}$) was generated by cutting out the repeat cassette from the 5′(CGG)$_{16}$-*luc2*(0) plasmid with *KasI* and *XhoI* and cloning it into the digested 5′(CGG)$_{99}$ vector.

Plasmid and small compound delivery

For average fluorescence and luminescence signal quantification, COS7 cells were seeded in 125 µl DMEM medium on a 96-well plate one day before transfection. 12.5 µl of either p*EGFP*-C1, 5′(CGG)$_{99}$-*GFP*(+1), ATG(CGG)$_{99}$-*GFP*(+1), ATG(CGG)$_{99}$-*luc2*(+1), 5′(CGG)$_{16}$, 5′(CGG)$_{99}$, *mCherry*-NLS (48), 5′(CGG)$_{16}$-*luc2*(0), 5′(CGG)$_{34}$-*luc2*(0), 5′(CGG)$_{44}$-*luc2*(0) or 5′(CGG)$_{99}$-*luc2*(0) constructs were delivered with X-tremeGENE HP DNA Transfection Reagent (Roche; 125 ng DNA/0.25 µl X-tremeGENE HP) when cells reached 40% confluency. All small compounds, except for NCD, Z-NCTS and CMBL4a, were dissolved in dimethyl sulfoxide (DMSO) to 1 mM concentration, aliquoted and stored at 4°C. NCD and Z-NCTS were dissolved in H_2O and CMBL4a in H_2O or DMSO. No obvious differences in the activity of CMBL4a was observed in either solvent. Prior to the experiment, the compounds were brought to the desired concentration in DMEM and added to the cells to the total 150 µl. Solvent mixed with DMEM was delivered to control wells.

Fluorescence and confocal microscopy

Images were taken with Axio Observer.Z1 microscope equipped with AxioCam MRm camera, filter set 09 or 10 (GFP), 49 (Hoechst 33342) and 31 (mCherry), A-Plan 10×/0.25 Ph1 objective (Zeiss), and AxioVs40 module or a Nikon A1R confocal microscopy (20x objective). Total fluorescence signal, mCherry-NLS positive cell numbers, inclusion numbers, inclusion areas and nuclear circularity were calculated using ImageJ. Total fluorescence was estimated based on background subtraction (rolling ball algorithm) and mean intensity quantification. Inclusion areas and numbers were analyzed following image thresholding and using ‘Analyze Particles’ function. Nuclei were stained with Hoechst 33342 (Invitrogen) per manufacturer’s protocol. Briefly, cell culture media was removed and cells were washed in PBS. Next, sufficient volume of Hoechst staining solution (prepared by diluting the Hoechst stock 1:2000 in PBS) was added to cover the cells. Upon 10-minute incubation in a humidified CO2 incubator at 37°C, protected from light, staining solution was removed and cell were washed once in PBS prior to addition

of fresh PBS and fluorescence microscopy. For confocal microscopy, COS7 cells were fixed for 15 min with 4% PFA 48 h post transfection and blocked for 1 h in 1% BSA diluted in PBS-Tween (0.1%; PBS-T). Incubation with mouse FMRpolyG 8FM (1:50; (23, 49)) and rabbit Lamin B1 – Nuclear Envelope Marker (1:200; Abcam, ab16048) primary antibodies was conducted O/N at 4°C in the blocking solution. Secondary goat anti-mouse FITC-labeled (1:400; Jackson ImmunoResearch Laboratories) and goat anti-rabbit Alexa Fluor 546 (1:200; Invitrogen) secondary antibodies were applied for 1 h at RT in PBS-T.

Luciferase assay

For luciferase assay, cells were harvested at different timepoints as indicated in figure legends. Following lysis, cells were transferred to a Nunc F96 MicroWell Black Polystyrene Plate (137101, Thermo Scientific) and luminescence of Firefly and Renilla luciferases were measured consecutively using Dual Luciferase Assay System (Promega), infinite F200 PRO, and i-control 1.8 SP1 microplate reader software (Tecan).

Cell Viability Assays

HEK293T and COS7 cells seeded in a 96-well plate (Figure 7E and Supplementary Figure S16B) were transfected with either 5'$(CGG)_{16}$ or 5'$(CGG)_{99}$ and treated with indicated amount of CMBL4c or an equal volume of DMSO (control). Cell viability, toxicity and apoptosis were analyzed at indicated time-points post plasmid delivery using ApoToxGlo^{TM} Triplex Assay (Promega) according to the instructions provided by the manufacturer. Triplicate repeats were used for each experiment. Fluorescence at two distinct wavelength sets ($400_{EX}/505_{EM}$ for viability, $485_{EX}/520_{EX}$ for cytotoxicity) as well as luminescence (for apoptosis) were recorded on Tecan Spark microplate reader.

HeLa cells (5000 cells/well; Supplementary Figure S7) were seeded in a 96-well plate in a final volume of 90 mL in DMEM containing 10% FBS. After 24 hours of incubation, cells were treated with 10 mL of seven different concentrations (0, 0.1, 0.3, 1.0, 3.0, 10, 30, 100 mM) of corresponding CMBLs (CMBL1a, 1b, 1c, 3a, 3b, 4a, 4b, 4c and 5c) prepared in DMEM containing 10% FBS and 10% DMSO. Final concentration of 1% DMSO was maintained in each well and each concentration was treated in experimental sextuplicate. After 24 h of incubation, 5 mL of cell proliferation reagent (WST-8) in DMEM containing 10% FBS was added to each well and incubated for 1 h. Readings of the absorbance were taken at 450 nm

wavelength, after 10 sec of shaking time in Spectramax M5. Graphs were plotted using the Graphpad Prism8 version 8.0.2. Each experiment was performed at least in triplicate.

Cell viability assay in SH-SY5Y cells (Supplementary Figure S8) was performed using the RealTime-Glo MT Cell Viability Assay according to the manufacturer's protocol (Promega). Cells were plated in 20% confluency in a 96-well in 50 µL of culture medium. The next day the medium was removed and replaced with 25µL medium containing 2 × RealTime-Glo reagents. The CMBL titration was prepared in medium at 2 × concentrations and added to the plate at an equal volume. For time-zero measurements, cells were incubated with RealTime-Glo MT Cell Viability reagent for 1h at 37 °C. Luminescence was then read every 8h for 48 h on Tecan M200. The quantification of the luminescence was based on four experimental replicas. P-values were calculated using GraphPad Prism software, 8.3.0 (https://www.graphpad.com/).

Lethality in mice

The experiments in mice were performed with the permission of the local ethical committee (NCBS-IAE-2017/05(N)). BALB/c mice (4–6-week-old) were injected subcutaneously with a gauge needle either with 1% DMSO in PBS (n=4) or one of the five concentrations of CMBL4c (25, 50, 100, 250, 500 mg/kg body weight of mice; n=6 mice for each treatment group) prepared from the DMSO stock solution and diluted with PBS to obtain the desired CMBL4c concentration and 1% of DMSO. Mice were monitored closely for 48 h. As no signs of morbidity or toxicity were observed during this time period, the treatment was continued at regular intervals of 48 h for additional 12 days. No signs of morbidity or toxicity were observed in the course of the two-week experiment in either treatment group.

RNA isolation and RT-PCR

For semi-quantitative multiplex RT-PCR (Figure 6A and B), CMBL compounds were added 18 h post plasmid delivery to COS7 cells. Following additional 22 h, RNA was isolated with TRI Reagent (Sigma) according to the manufacturer's instructions. Total RNA (1 µg) was reverse-transcribed with GoScript™ Reverse Transcriptase and oligo(dT) primers (Promega). PCR was performed with either *GFP*/anchored oligo$(dT)_{20}$ (Invitrogen) and *GAPDH* primers at 50.3°C annealing temperature or with *FMR1*- and *GAPDH*-specific primers (Supplementary Table S4).

For quantitative PCR (qPCR) and splicing analyses (Figure 6C, Supplementary Figure S10 and S11A), fibroblasts (GM07492 and GM04033) were harvested in TRIzol 48 h following delivery of DMSO or CMBLs. Each experiment was performed in triplicate repeats. RNA was isolated using Total RNA Zol-Out™ kit (A&A Biotechnology) per manufacturers' instructions. cDNA was synthesized from 500 ng total RNA using SuperScript™ IV Reverse Transcriptase (Invitrogen) according to manufacturers' protocol. qPCRs were performed in a QuantStudio™ 7 Flex System (Thermo Fisher Scientific) using Maxima SYBR Green/ROX qPCR Master Mix (Thermo Fisher Scientific) according to the manufactures' instructions. Targets containing a various number of CGG repeats were amplified with primers listed in Supplementary Table S4 at 58-60°C annealing temperature. Ct values were normalized against *GAPDH*. Fold differences in expression level were calculated according to the $2^{-\Delta\Delta Ct}$ method (50). Splicing of *MBNL1* e1, *MBNL1* e5, *MBNL2* e7, *NCOR2* e45, *NFIX* e7 and *PHKA* e19 was analyzed following standard PCR reactions using GoTaq® Flexi DNA polymerase (Promega). Sequences of primers are listed elsewhere (51, 52). PCR images were captured using G:Box EF2 (Syngene) and analyzed using GeneTools image analysis software (Syngene). Results were plotted on bar graphs as % exon spliced in (ΔPSI).

Western blot

DMSO and CMBL compounds were added to FXTAS and non-FXTAS fibroblast lines at ~80% confluency to 2 μM (0.2% DMSO) final concentration and incubated for 48 h. Fibroblasts were harvested with trypsin and lysed with RIPA buffer. Protein extracts were centrifuged at 10,000 RPM for 10 min at 4°C, measured with Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) and heat-denatured for 10 min at 70°C with the addition of Bolt LDS buffer (Invitrogen; Figure 8C and Supplementary Figure S17A**)** or 5 min at 95°C following addition of Laemmli buffer (Supplementary Figure S17B). 20-25 μg of protein per well was separated in Bolt™ 4-12% Bis-Tris Plus gel (Invitrogen, NW04120BOX) and Bolt™ MES SDS Running Buffer. Proteins were transferred to PVDF transfer membrane (0.45 μM, GE Healthcare) for 1 h at 100 V in Leammli buffer with 20% methanol. Following transfer, membranes were blocked and incubated with antibodies in SNAP id Protein Detection System (Merck Millipore). Blocking was performed for 20 min with either 0.125% non-fat dry milk (NFDM) in TBS with 0.1% Tween 20 (TBS-T; Figure 8C and Supplementary Figure S17A) or 1% BSA in TBS-T (Supplementary Figure S17B). Correspondingly, primary antibodies were diluted in either 0.125% NFDM/TBS-T or 1% BSA/TBS-T. The following concentrations and incubation times

were used: rabbit anti-FMRP (ab17722, Abcam), 1:500, 1 h 15 min; mouse anti-GAPDH (sc-47724, Santa Cruz), 1:10,000, 20 min; rabbit anti-SLC40A1 (NBP1-21502, Novus Biologicals), 1:1000, 1 h; and rabbit anti-QKI (ab126742, Abcam), 1:1000, 1 h. Membranes were washed in TBS-T and incubated with horseradish peroxidase conjugated secondary antibodies: anti-rabbit (A9169, Sigma), 1:20,000 for 15 min or anti-mouse (A9044, Sigma), 1:20,000 for 15 min and washed with TBS. Antibody-antigen complexes were visualized by enhanced chemiluminescence (ECL) using Luminata Forte HRP Substrate (Merck Millipore, WBLUF0500) and detected with G:Box System (Syngene).

The western blot analysis for HTT protein (17/68Q tract; Supplementary Figure S11B) was performed as previously described (53). Briefly, 25 µg of total protein was run on a Tris-acetate sodium dodecyl sulphate (SDS)-polyacrylamide gel (1.5 cm, 4% stacking gel/4.5 cm, 5% resolving gel, acrylamide:bis-acrylamide ratio of 35:1) in XT Tricine buffer (Bio-Rad; Hercules, CA, USA) at 140 V in an ice-water bath. After electrophoresis, the proteins were wet-transferred to a nitrocellulose membrane. The membranes were cut according to protein ladder (HiMark) for separate detection of respective proteins. All steps of immunodetection were performed in SNAP id Protein Detection System in buffer containing 0.25% non-fat dry milk in PBS/0.9% NaCl/0.1% Tween-20. For huntingtin and plectin detection, the blots were probed with the primary antibodies: rabbit anti-huntingtin (1:1000; ab109115 (EPR5526), Abcam) and rabbit anti-plectin-1 (1:1000; #12254 (D6A11), Cell Signaling) and then with anti-rabbit HRP-conjugated secondary antibodies (1:500; 711-035-152, Jackson ImmunoResearch). The immunoreaction was detected using ECL Western Blotting Substrate (ThermoScientific).

Results

CMBL4a binds to RNA CGG repeats *in vitro* and attenuates FMRpolyG production in cell culture

Based on our previous studies (40–42), we predicted binding of naphthyridine ligands to CGG repeats and their use as potential therapeutics in FXTAS. CGG repeats in RNA form a thermodynamically stable A-form hairpin structure with consecutive structural motifs containing two Watson-Crick G-C pairs and structure stabilizing G-G pair with hydrogen bonds between the Watson-Crick and Hoogsteen edges (54–56). The most recent study showed that if the CGG repeat tract reaches pathogenic length it may also adopt G-quadruplex structure (38). We investigated binding affinity of NCD, Z-NCTS and CMBL4a that encompass H bonding surface complementary to guanine (39–42, 44) in the RNA structure formed by CGG tandem repeats *in vitro* using surface plasmon resonance (SPR) single cycle kinetics and filter binding assays (Figure 1A-C). All compounds showed apparent binding to $r(CGG)_9$, and NCD had the weakest SPR response among them, considering its two-fold higher concentration used in the assay (Figure 1B). Specifically, apparent association (k_{on}) and dissociation (k_{off}) constants of NCD applying 1:1 fitting model indicated relatively slow association and rapid dissociation kinetics as well as high dissociation constant ($K_{d(app)}$ = 15.3 μM; Supplementary Table S3). $K_{d(app)}$ for Z-NCTS and CMBL4a were in a nanomolar range ($K_{d(app)}$ ~300 nM and $K_{d(app)}$ ~800 nM, respectively) characterized by much slower dissociation kinetics of Z-NCTS (Supplementary Table S3). Filter binding assays also revealed the lowest affinity of NCD to $r(CGG)_{20}$ while CMBL4a and Z-NCTS interacted with the CGG repeats in similar ligand concentrations (Figure 1C).

We hypothesized that binding of small molecules to CGG repeat hairpin *in vivo* can affect efficiency of RAN-translation from RNA containing expanded CGG triplets. Therefore, we followed the *in vitro* binding tests with cell culture experiments, delivering NCD, Z-NCTS and CMBL4a to the COS7 cell line transfected with the 5'$(CGG)_{99}$-*GFP*(+1) vector, containing the full length of *FMR1* 5'UTR with 99 CGGs fused with *GFP*, and from which a fusion protein of FMRpolyG and GFP is generated (FMRpolyG-GFP) (Figure 1D-F; (47)). Compound toxicity comparison revealed increased cell mortality starting with 0.2 μM, 30 μM and 50 μM for Z-NCTS, NCD and CMBL4a, respectively. Using these subtoxic concentrations, we did not observe any significant inhibitory effect on the FMRpolyG-GFP signal upon Z-NCTS and NCD

administration 48 h post plasmid delivery (Figure 1E). In contrast, addition of CMBL4a markedly reduced the fluorescence in cells transfected with 5'$(CGG)_{99}$-*GFP*(+1) vector but not in control cells, to which control p*EGFP*-C1 vector was delivered. Interestingly, we observed a variable size of FMRpolyG inclusions in untreated cells and while the larger co-localized with Hoechst staining (Figure 1F, red arrows), the smaller were often Hoechst-negative (Figure 1F, yellow arrows) indicating their cytoplasmic distribution.

CMBL4c shows high activity and low toxicity in cell culture assays and mice

We then focused on CMBL chemistry (Figure 2A) and compared the effect of 15 previously reported (1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 3c, 4a, 4c, 5c, 6a, 6b, 7b) and six newly synthesized (3aL, 3bL, 3cL, 4b, 5a, 5b) CMBL compounds ((41, 42); Supplementary Data), differing in molecular composition of the variable linker (Figure 2A, Supplementary Figure S1). We first screened CMBLs using SPR assays with immobilized $r(CGG)_9$ repeats (Figure 2B, Supplementary Figure S2). The compounds' responses differed greatly in intensity and kinetic profiles, with CMBL4a showing the highest binding affinity, indicating the significance of the linker structure in the interaction with the CGG repeats (Figure 2B, Supplementary Figure S1 and S2). A CMBL4a analogue, CMBL4c, also showed a notable SPR response in concentrations higher than 1 μM, with rapid association and very slow dissociation profiles (Figure 2B, Supplementary Figure S2). CMBL5a, CMBL5c and CMBL3aL also exhibited notable SPR responses; however, with faster dissociation kinetics than CMBL4c (Supplementary Figure S2). Interactions of selected CMBL compounds with longer $r(CGG)_{20}$ as well as control $r(AGG)_{20}$ and $r(CUG)_{20}$ RNA hairpin structure were also tested by *in vitro* filter binding and electrophoretic mobility shift assays (Figure 2C, Supplementary Figure S3 and S4). The results showed that CMBLs bound to $r(CUG)_{20}$ with several-fold higher K_d than to $r(AGG)_{20}$ and $r(CGG)_{20}$. Particularly, among the all tested CMBLs, we observed the highest affinity of CMBL4a to the CGG repeats, in agreement with the SPR data.

We next compared the effect of 21 CMBL compounds in their capacity to lessen the luciferase signal of FMRpolyG-Firefly fusion protein in ATG$(CGG)_{99}$-*luc2*(+1)-transfected COS7 cells ((47); Figure 3). CMBL1a-c, CMBL2c, CMBL3aL, CMBL3cL and CMBL5a showed no significant effect on translation of FMRpolyG-Firefly whereas other compounds induced a decrease in the luciferase activity in a concentration-dependent manner (Figure 3B). Unexpectedly, despite the highest affinity to tandem repeats *in vitro* (Figure 2B and C, Supplementary Figure S4), CMBL4a effect on FMRpolyG-Firefly inhibition was moderate

(Figure 3B). Particularly, top activities in COS7 cells were observed for CMBL3a, 3b, 4c, 5c and 6b. Most of them displayed associated cellular toxicity in higher concentrations (marked red in Figure 3C, Supplementary Figure S5, S6, S7 and S8). Therefore, for further studies we chose CMBL4c along with its analogues CMBL4a and CMBL4b (marked blue in Figure 3B), which showed relatively high activity and low toxicity. We first estimated their effect on attenuation of fluorescent signal of FMRpolyG-GFP and inclusion formation (Figure 4). Interestingly, in cells transfected with 5'$(CGG)_{99}$-*GFP*(+1), increasing concentrations of CMBL4a (up to 10 μM, dashed line in Figure 4A) were first correlated with higher amounts and then decreasing numbers of FMRpolyG-GFP inclusions, particularly those of larger areas (Figure 4A). In agreement with the luciferase activity data (Figure 3B), CMBL4c compared to 4a and 4b showed the highest activity in lower concentration range (Figure 4A-C) that was associated with FMRpolyG-GFP translation efficiency from 5'$(CGG)_{99}$-*GFP*(+1) and ATG$(CGG)_{99}$-*GFP*(+1) vectors (Figure 1D, 4B and 4C) and the time that the compound was administered following ATG$(CGG)_{99}$-*GFP*(+1) delivery (Figure 4D). We then tested CMBL4c toxicity *in vivo* in BALB/c mice. Importantly, no signs of toxicity were observed in mice treated every 48 hours for two weeks with the highest dose (500 mg/kg) via subcutaneous injection.

CMBL4c interacts with various repeat sequences but does not lower the level of mRNAs containing CGG, CUG and CAG repeats

To evaluate specificity and affinity of CMBL4c towards various disease-associated repeat DNA and RNA sequences, we performed SPR assays with immobilized $d(CGG)_9$, $d(CCG)_9$, $d(CAG)_9$, $d(CTG)_9$, $d(GAA)_9$, $d(CCTG)_9$, $d(G_4C_2)_6$ and corresponding RNAs, $r(CGG)_9$, $r(CCG)_9$, $r(CAG)_9$, $r(CTG)_9$, $r(GAA)_9$, $r(CCTG)_9$, $r(G_4C_2)_6$, as well as random single stranded (rSS30) and double stranded RNAs (rDS30) (Figure 5 and Supplementary Figure S9). No significant responses were observed for CMBL4c towards $d(CCG)_9$, $d(CTG)_9$, $d(GAA)_9$, $d(G_4C_2)_6$, $r(CCG)_9$, $r(CUG)_9$, $r(GAA)_9$, $r(CCUG)_9$, $r(G_4C_2)_6$, rSS30, and rDS30. In contrast, $d(CCTG)_9$ showed the highest response from all of the tested sequences and the responses of CMBL4c towards $d(CGG)_9$, $d(CAG)_9$, and $r(CAG)_9$ were comparable to that of $r(CGG)_9$. Nonetheless, the $d(CCTG)_9$, $d(CGG)_9$, $d(CAG)_9$, and $r(CAG)_9$ SPR curves revealed relatively fast dissociation kinetics of CMBL4c when compared to the $r(CGG)_9$. This data indicates the highest binding affinity of CMBL4c towards $r(CGG)_9$ among the tested RNA/DNA sequences.

As we observed CMBL4c binding to RNA as well as DNA $(CGG)_9$, we asked whether the effect of CMBL compounds on synthesis of FMRpolyG is related to their interaction with tandem repeats in target mRNA molecules and/or to transcription inhibition due to binding to the CGG repeat-rich DNA cassette. Endogenous *FMR1* and 5'$(CGG)_{99}$-*GFP*(+1) mRNA levels were quantified in COS7 cells following delivery of 5'$(CGG)_{99}$-*GFP*(+1) vector and either CMBL4c or 5c (ligand associated with cellular toxicity; Figure 3C, Supplementary Figure S7 and S8), relative to *GAPDH* transcript levels (Figure 6A and B). Importantly, we did not notice any reduction in *FMR1* and 5'$(CGG)_{99}$-*GFP*(+1) mRNAs following CMBL4c delivery in COS7 cells and rather slightly increased amounts of *FMR1* mRNA and other short and long CGG repeat-containing mRNAs, including *CARM1*, *QKI* and *VKORC1L1* in human fibroblasts (Figure 6C, Supplementary Figure S10). Likewise, addition of CMBL5c did not change endogenous *FMR1* mRNA content in COS7 cells; however, significantly reduced the level of vector-originated mRNA containing CGG_{exp} (Figure 6A and B). The latter presumably pertains to the toxic effect that this compound exerts in higher concentrations.

We also tested whether CMBLs could bind other disease-causing expanded triplet repeats, including CUG_{exp} in DM1 and CAG_{exp} in Huntington's disease (HD). In DM1, CMBLs could potentially act by releasing sequestered MBNL proteins from expanded CUG repeats in *DMPK* mRNA (57). As MBNLs play crucial roles in mRNA metabolism, particularly splicing regulation (58, 59), therapeutic intervention would increase the level of available MBNLs and the percentage of correctly spliced mRNAs. In accordance with the *in vitro* data (Figure 5B, Supplementary Figure S4), we observed no significant changes in inclusion of alternative exons in MBNL-target mRNAs upon CMBL4c delivery to either non-DM1 or DM1 fibroblasts (Supplementary Figure S11A). Likewise, in HD CMBLs could potentially bind to CAG_{exp} in the mutated *HTT* transcript and block translation of the toxic huntingtin protein containing long polyglutamine tracts (60, 61). We observed; however, no change in the level of either normal or mutant huntingtin in HD fibroblasts upon 48h treatment with 1 μM CMBL3a (Supplementary Figure S11B). These results indicate that CMBLs, while potentially therapeutic in FXTAS, are not suitable for therapy of DM1 and HD.

The effect of CMBL4c on inclusion growth is delivery-time dependent

We then tested if CMBL4c could be used to reverse the inclusion deposition in COS7 cells following FMRpolyG-GFP forced translation from 5'$(CGG)_{99}$-*GFP*(+1) (Figure 7A and B,

Supplementary Figure S12 and S13) or ATG(CGG)$_{99}$-*GFP*(+1) (Supplementary Figure S14 and S15) vectors. We delivered CMBL4c at 2 μM concentration 18 or 30 h post transfection and estimated inclusion sizes 54, 72, 90 and 108 h post transfection. While significant time-dependent increases in inclusion areas were observed in untreated cells following delivery of 5'(CGG)$_{99}$-*GFP*(+1) and ATG(CGG)$_{99}$-*GFP*(+1), either no increases (Figure 7B) or slight decreases (Figure S14B) were noted once CMBL4c was delivered 18 h post transfection. In contrast, delivery of CMBL4c in the later time point did not have such prominent effect on halting the inclusion growth.

As our previous results indicated that GFP-tagging increases protein stability (28, 47), we delivered to COS7 cells a construct encoding an untagged version of FMRpolyG containing the full native protein sequence (5'(CGG)$_{99}$; Figure 7C), including the C-terminal part of the protein shown to exert cellular toxicity (23). Indeed, confocal immunofluorescence microscopy analysis revealed abnormalities in nuclear envelope structure and death of cells containing FMRpolyG aggregates (Figure 7D). To evaluate whether the cell death could be diminished following inclusion formation, we co-delivered either a long or short CGG repeat containing construct with a native *FMR1* 5'UTR sequence (5'(CGG)$_{99}$ or 5'(CGG)$_{16}$; Figure 7C) along with a marker of transfected cells, *mCherry*-NLS vector, to COS7 cells and treated them 24 h post transfection with CMBL1 control compound or CMBL4c (Supplementary Figure S16A). Administration of 5'(CGG)$_{99}$ resulted in significantly reduced numbers of mCherry-positive cells and neither addition of CMBL1 nor 4c prevented the cell death.

To further assess therapeutic potential of CMBL4c, we measured viability, cytotoxicity and apoptosis of HEK293T and COS7 cells following delivery of 5'(CGG)$_{16}$ or 5'(CGG)$_{99}$ and either DMSO (as a control) or CMBL4c, at 6 or 24 hours post transfection (Figure 7E, Supplementary Figure S16B). In accordance with the microscopy analysis (Figure 7D, Supplementary Figure S16A), we observed decreased viability, high cytotoxicity and apoptosis of cells to which 5'(CGG)$_{99}$ was delivered. Importantly, CMBL4c addition early after transfection (6 h) significantly increased viability of HEK293T cells (Figure 7E). Moreover, CMBL4c lowered cytotoxicity and apoptosis in both COS7 and HEK293T cells, regardless whether it was provided 6 or 24 hours post 5'(CGG)$_{99}$ administration (Figure 7E, Supplementary Figure S16B). We noticed; however, slightly higher cytotoxicity in cells, to which CMBL4c was added in the later time point (Figure 7E). Overall, these results indicate that CMBL4c delivery can reduce FMRpolyG-mediated cytotoxicity and apoptosis, increasing

cell viability. Importantly, therapeutic potential of CMBL4c is also observed once the inclusions are already formed.

CMBL4c partially inhibits FMRP translation in FXTAS cells

As both FMRpolyG and FMRP are translated from the same *FMR1* transcripts carrying CGG_{exp}, we tested if FMRP production is also affected along with FMRpolyG following delivery of CMBL4c. First, we generated a number of constructs with varying numbers of CGG repeats in the 5'UTR of *FMR1* and carrying Firefly luciferase sequence (*luc2*) in the FMRP reading frame (Figure 8A). Following delivery of 5'$(CGG)_{16}$-*luc2*(0), 5'$(CGG)_{34}$-*luc2*(0), 5'$(CGG)_{44}$-*luc2*(0) and 5'$(CGG)_{99}$-*luc2*(0) plasmids and CMBL4c to COS7 cells, we observed the CGG repeat length- and CMBL4c concentration-associated decrease of Firefly luciferase contents (Figure 8B), indicative of inhibitory effect of CMBL4c on translation of both FMRpolyG and FMRP-initiated open reading frame.

We then delivered either DMSO, CMBL4a or 4c to one control fibroblast cell line (male 1028-07 with $(CGG)_{22}$ in *FMR1*) and two FXTAS cell lines (male 1022-07, $(CGG)_{81}$ and male 1044-07, $(CGG)_{97}$) (Figure 8C, Supplementary Figure S17A) or DMSO, CMBL1a or 4c to a control line (C0603 from male with $(CGG)_{31}$) and an FXTAS line (homozygotic WC26 with $(CGG)_{60}$/$(CGG)_{90}$) (Supplementary Figure S17B). Neither DMSO nor CMBL1a affected the content of FMRP in either cell line (Figure 8C, Supplementary Figure S17), while CMBL4a and CMBL4c lowered FMRP in FXTAS fibroblasts in a repeat length-dependent manner, below 50% of control levels in a line containing $(CGG)_{97}$ (Figure 8C). In contrast, CMBL4c did not affect the content of FMRP in non-FXTAS fibroblasts.

Discussion

Based on our previous studies (40–42), we selected three types of prospective naphthyridine ligands that could interact with CGG_{exp} and potentially act therapeutically in FXTAS: a naphthyridine carbamate dimer (NCD; (39)), a naphthyridine tetramer containing two NCD molecules (Z-NCTS; (40, 44)) and a cyclic naphthyridine dimer, CMBL4a (41, 42). Comparison of their activity *in vitro* and in cell culture showed that CMBL4a has the highest affinity to CGG repeats in RNA and that it effectively inhibits production of FMRpolyG while being the least toxic. Binding affinity of Z-NCTS to the repeat is higher than NCD and comparable to CMBL4a *in vitro* ($K_{d(app)}$ ~ 300 nM and K_d = 2 μM in SPR and filter binding assays, respectively (Supplementary Table S3, Figure 1C)). Nonetheless, high compound toxicity (50% cell viability at 0.3 μM concentration in COS7 cells; see also (44)) presumably prevented this compound from being effective in COS7 cell culture assays (Figure 1E).

In search for the most optimal CMBL chemistry, we compared 21 CMBL analogues in their binding activity *in vitro* and capacity to inhibit FMRpolyG synthesis in the FXTAS cell model. The compounds could be divided into several groups, dependent on the molecular composition of the variable linker (Figure 2A, Supplementary Figure S1). Unexpectedly, despite the highest affinity to CGG repeats *in vitro* (Figure 2B and C) and the ability to penetrate the cell membrane *in vivo* (Figure 8C; the effect on FMRP level), CMBL4a showed only moderate activity on biosynthesis of FMRpolyG in cell culture experiments. Specifically, an NH to CH_2 group switch between CMBL4a and 4c resulted in more than a five-fold higher reduction of FMRpolyG at 2 μM concentration (Figure 3B). Superior activities in cell culture could also be observed for compounds such as CMBL3b; however, we also noted its concentration-dependent elevated toxicity (Figure 3C, Supplementary Figure S5) that could stem from the overall higher binding capacity to sequences other than CGG triplet repeats (Supplementary Figure S4). Importantly, our studies revealed attenuated FMRpolyG synthesis independently of the translation initiation start (Figure 4 and 7, Supplementary Figure S14). This indicates that upon binding to CGG_{exp}, CMBL4c would block translation of FMRpolyG directly from the repeats (21, 22, 26) as well as from near-cognate codons located upstream to them (23, 27).

Our previous *in vitro* data indicated that naphthyridine ligands interact not only with RNA but also with DNA sequences containing mismatches (39, 40, 42), and thus could affect not only translation but also transcription of various targets (62). Likewise, we found that CMBL4a has comparable binding affinity towards both $r(CGG)_9$ and $d(CGG)_9$ (Supplementary

Table 3), characterized by relatively rapid association and dissociation kinetics (Figure 2B, Supplementary Figure S18). CMBL4c also interacted with various DNA and RNA sequences, including d(CGG)$_9$, d(CCTG)$_9$, d(CAG)$_9$, r(CGG)$_9$, and r(CAG)$_9$ (Figure 5, Supplementary Figure S9) as well as r(CUG)$_{20}$ (Supplementary Figure S4) that indicated therapeutic potential of CMBLs not only in FXTAS and various novel CGG/GGC-repeat related diseases, including neuronal intranuclear inclusion disease (NIID), oculopharyngeal myopathy with leukoencephalopathy (OPML) and oculopharyngodistal myopathy (OPDM) (63–65), but also in HD (61), DM1 (57) and myotonic dystrophy type 2 (DM2) defined by a CCTG expansion in the *CNBP* gene (61, 66). Our analyses revealed that CMBL4c does not alter DM1-related splicing aberrations of MBNL-dependent alternative exons (Supplementary Figure S11A), allowing to conclude that CMBL4c is unable to release sequestered MBNL molecules from toxic CUG$_{exp}$ (48) and has no therapeutic potential in DM1, at least in a range of concentrations used in this study (57). Likewise, we did not observe any therapeutic effect of CMBL3a in HD fibroblasts carrying CAG$_{exp}$ (Supplementary Figure S11B), despite the observed translation inhibition of FMRpolyG in an FXTAS COS7 cell model (Figure 3B, Supplementary Figure S5). Therapeutic potential of CMBLs in DM2 is still yet to be determined. Nevertheless, it is important to note that according to SPR profiles CMBL4c interaction with r(CGG)$_9$ was characterized by much slower dissociation kinetics than with d(CGG)$_9$, d(CCTG)$_9$ and r(CAG)$_9$, suggesting the highest affinity of CMBL4c towards RNA CGG repeats.

As CMBL4c interacted with d(CGG)$_9$ and r(CGG)$_9$, we tested if the CMBL-driven inhibition of FMRpolyG synthesis is solely due to binding to RNA fragments carrying CGG repeats in the cytoplasm or also to the template DNA motifs in the nucleus. In the latter case, such interaction would attenuate transcription and lower the level of *FMR1* mRNA. Our data indicate unaffected transcription in COS7 cells and rather increased mRNA levels in fibroblasts upon delivery of CMBL4c in concentrations sufficient to abolish FMRpolyG synthesis (Figure 6). The latter might suggest retention of mRNAs containing CGGs in the nucleus or an increase in the mRNA stability due to less efficient translation upon CMBL treatment. Indeed, increased mRNA levels were not correlated with higher protein amounts in fibroblasts (compare Figure 6C and 8C), and decreased QKI protein levels were observed in some; however, not in all tested fibroblast cell models (Figure 8C). In contrast, high concentrations of CMBL5c that induced cell toxicity (Figure 3C, Supplementary Figure S7 and S8) reduced the content of vector-derived mRNA particles containing CGG$_{exp}$ repeats but not the wild-type *FMR1* transcripts (Figure 6B). Overall, these data point to the CMBL-associated toxicity stemming not only from binding to sequences other than CGG triplet repeats in mRNA targets but also to DNA

sequences containing repeat cassettes. Furthermore, our results indicate that CMBL binding to RNA in the cytoplasm and DNA in the nucleus is dependent on the compound concentration used in particular cell assays as well as on the molecular composition of the variable linker.

Although intranuclear inclusions in the central nervous system and other tissues are a hallmark of FXTAS (3, 49, 67), there is an ongoing debate regarding accumulation of misfolded/toxic proteins and the role of inclusion bodies formation in the pathological alterations of neuronal cells (18, 68). Current data from FXTAS cell models indicate high toxicity of FMRpolyG inclusions (21–23). Particularly, Sellier et al. (23) showed that FMRpolyG proteins aggregate in the cytoplasm and then move to the nucleus where they form inclusions disrupting the nuclear envelope through the interaction with LAP2β. We used either GFP or Firefly luciferase tagged versions of FMRpolyG to simplify detection and quantification of the inclusions. Nonetheless, as tagging often affects protein properties such as stability (21, 28, 47), we followed these studies with immunofluorescent detection of cells co-transfected with either 5'$(CGG)_{99}$ or 5'$(CGG)_{16}$ from which only FMRpolyG are generated (Figure 7C) but while the expanded variant forms inclusions, the latter is quickly degraded (23). We showed that CMBL4c-induced attenuation of toxic FMRpolyG production is long-lasting (more than 100 h; Figure 4D, 7A and B, Supplementary Figure S12, S13 and S14) and can either reverse or stop pathological alterations, particularly when added early in the disease process (Figure 7B, Supplementary Figure S14B), resulting in enhanced cell viability (Figure 7E) and markedly reduced cytotoxicity and apoptosis in various cell types (Figure 7E, Supplementary Figure S16B). The therapeutic effect of CMBL4c could be thus observed despite the presence of the mRNA from premutation alleles and the pathology related to protein sequestration by CGG_{exp} (18, 38). Interestingly, reversibility of the inclusion formation was previously shown in transgenic mice, in which CGG_{exp} were synthesized in the brain upon doxycycline administration; however, only in a short window of time, up to 8 weeks post transcription induction (69). When CGG_{exp} transcription was stopped 10 or 12 weeks after doxycycline treatment, then neither the number nor the size of FMRpolyG could be reduced.

FMRpolyG and FMRP are translated from the same *FMR1* transcript carrying CGG_{exp} (23), which raises the question regarding the possible selective inhibition of the toxic protein with either small molecules or antisense oligonucleotides (34). In a perfect scenario, a therapeutic drug should not attenuate FMRP synthesis, as its absence has been linked to the development of fragile X syndrome (FXS) characterized by alterations in physical appearance, autism and intellectual disabilities in children. Indeed, Rodriguez et al. recently showed that

antisense oligonucleotides selectively blocking RAN translation can enhance FMRP synthesis (24). Our study revealed that CMBL4c administration attenuates synthesis of both FMRpolyG and FMRP in cell models of FXTAS; however, without the consecutive downregulation of CGG_{exp}-containing transcripts. The precise binding mode of CMBL4c to the CGG repeat as well as the binding mode of the previously reported molecules (33–35, 38) has not been fully elucidated yet. Structural studies of the CMBL4c:rCGG repeat complex will provide a deeper understanding of the molecular mechanisms underlying attenuation of the repeat-associated translations by small molecules. In future, this knowledge will be a valuable aspect in designing more effective molecules. As FXTAS affects individuals carrying pre-mutation over the age of 50, future experiments should aim at evaluation of the therapeutic potential of CMBL molecules, their appropriate route of delivery (e.g., ability to cross the blood-brain barrier) as well as toxicity *in vivo*, especially in the adult brain tissue, in which the benefits of FMRpolyG translation inhibition might outweigh adverse effects related to the FMRP partial loss.

Author contribution

PK, KS, SM, CD and KN designed the study and analyzed the data. Data contribution, PK: Figures 1D-F, 3, 4, 6A and B, 7A-C, 8A and B, Supplementary Figure S5, S6, S12, S13, S14, S15, S16A; SM: Figures 1A and B, 2A and B, 5, Supplementary Figure S1, S2, S7, S9, S18, Supplementary Table S1, S2, S3, Supplementary Data; ESK: Figure 1F, 4D, 6C, 7E, Supplementary Figure S10, S11A, S13, S16B; KT: Figures 1C, 2C, Supplementary Figure S3, S4; DN: Figure 6C, 8C, Supplementary Figure S10, S17A; AP: Supplementary Figure S8, S11B; AW: Figures 7C and D, Supplementary Figure S16A; AB: Supplementary Figure S17B. PK prepared the figures and wrote the manuscript with the help of other co-authors.

Funding

This work was supported by the National Centre for Research and Development grant ERA-NET-E-Rare-2/III/DRUG_FXSPREMUT/01/2016 (to KS), Foundation for Polish Science-TEAM program cofinanced by the European Union within the European Regional Development Fund (TEAM POIR.04.04.00-00-5C0C/17-00 to KS), by the National Science Centre grants: 2020/38/A/NZ3/00498 (to KS), 2018/30/E/NZ5/00065 (to PK), 2014/15/B/NZ5/00142 (to ESK), 2020/37/B/NZ5/01263 (to ESK) and 2017/24/C/NZ1/00112

(to KT), and JSPS KAKENHI Grant-in-Aid for Specially Promoted Research (26000007 to KN).

Acknowledgments

We thank Nicolas Charlet-Berguerand for 5′(CGG_{exp})-*GFP*(+1) and ATG(CGG_{exp})-*GFP*(+1) plasmids and 8FM mouse monoclonal antibodies. We also thank A. Bhattacharyya and P. Hagerman for FXTAS fibroblasts and the NCBS animal facility, especially Shefta-E-Moula who did *in vivo* toxicity tests.

References

1. Nelson,D.L.L., Orr,H.T.T. and Warren,S.T.T. (2013) The Unstable Repeats—Three Evolving Faces of Neurological Disease. *Neuron*, **77**, 825–843.
2. Hagerman,R.J. and Hagerman,P. (2016) Fragile X-associated tremor/ataxia syndrome — features, mechanisms and management. *Nat. Rev. Neurol.*, **12**, 403–412.
3. Greco,C.M., Hagerman,R.J., Tassone,F., Chudley,A.E., Del Bigio,M.R., Jacquemont,S., Leehey,M. and Hagerman,P.J. (2002) Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain*, **125**, 1760–1771.
4. Brunberg,J.A., Jacquemont,S., Hagerman,R.J., Berry-Kravis,E.M., Grigsby,J., Leehey,M.A., Tassone,F., Brown,W.T., Greco,C.M. and Hagerman,P.J. (2002) Fragile X premutation carriers: characteristic MR imaging findings of adult male patients with progressive cerebellar and cognitive dysfunction. *AJNR. Am. J. Neuroradiol.*, **23**, 1757–66.
5. Jacquemont,S., Hagerman,R.J., Leehey,M., Grigsby,J., Zhang,L., Brunberg,J.A., Greco,C., Des Portes,V., Jardini,T., Levine,R., *et al.* (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am. J. Hum. Genet.*, **72**, 869–78.
6. Greco,C.M., Berman,R.F., Martin,R.M., Tassone,F., Schwartz,P.H., Chang,A., Trapp,B.D., Iwahashi,C., Brunberg,J., Grigsby,J., *et al.* (2006) Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). *Brain*, **129**, 243–255.
7. Tassone,F., Hagerman,R.J., Garcia-Arocena,D., Khandjian,E.W., Greco,C.M. and Hagerman,P.J. (2004) Intranuclear inclusions in neural cells with premutation alleles in fragile X associated tremor/ataxia syndrome. *J. Med. Genet.*, **41**, e43.

8. Seltzer,M.M., Baker,M.W., Hong,J., Maenner,M., Greenberg,J. and Mandel,D. (2012) Prevalence of CGG expansions of the FMR1 gene in a US population-based sample. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.*, **159B**, 589–597.
9. Hunter,J., Rivero-Arias,O., Angelov,A., Kim,E., Fotheringham,I. and Leal,J. (2014) Epidemiology of fragile X syndrome: A systematic review and meta-analysis. *Am. J. Med. Genet. Part A*, **164**, 1648–1658.
10. Tassone,F., Iong,K.P., Tong,T.-H.H., Lo,J., Gane,L.W., Berry-Kravis,E., Nguyen,D., Mu,L.Y., Laffin,J., Bailey,D.B., *et al.* (2012) FMR1 CGG allele size and prevalence ascertained through newborn screening in the United States. *Genome Med.*, **4**, 100.
11. Sullivan,A.K.K., Marcus,M., Epstein,M.P.P., Allen,E.G.G., Anido,A.E.E., Paquin,J.J.J., Yadav-Shah,M. and Sherman,S.L.L. (2005) Association of FMR1 repeat size with ovarian dysfunction. *Hum. Reprod.*, **20**, 402–412.
12. Primerano,B., Amaldi,F., Bagni,C., Tassone,F., Hagerman,P., Hagerman,R.J., Hagerman,P., Amaldi,F. and Bagni,C. (2002) Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *RNA*, **8**, 1482–8.
13. Tassone,F., Hagerman,R.J., Taylor,A.K., Gane,L.W., Godfrey,T.E. and Hagerman,P.J. (2002) Elevated Levels of FMR1 mRNA in Carrier Males: A New Mechanism of Involvement in the Fragile-X Syndrome. *Am. J. Hum. Genet.*, **66**, 6–15.
14. Kenneson, A. Zhang, F. Hagedorn, CH. Warren,S., Kenneson,A., Zhang,F., Hagedorn,C.H. and Warren,S.T. (2001) Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum. Mol. Genet.*, **10**, 1449–1454.
15. Tassone,F., Iwahashi,C. and Hagerman,P.J. (2004) FMR1 RNA within the intranuclear inclusions of fragile X-associated tremor/ataxia syndrome (FXTAS). *RNA Biol.*, **1**, 103–105.
16. Sellier,C., Rau,F., Liu,Y., Tassone,F., Hukema,R.K., Gattoni,R., Schneider,A., Richard,S., Willemsen,R., Elliott,D.J., *et al.* (2010) Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. *EMBO J.*, **29**, 1248–1261.
17. Jin,P., Duan,R., Qurashi,A., Qin,Y., Tian,D., Rosser,T.C., Liu,H., Feng,Y. and Warren,S.T. (2007) Pur α Binds to rCGG Repeats and Modulates Repeat-Mediated Neurodegeneration in a Drosophila Model of Fragile X Tremor/Ataxia Syndrome. *Neuron*, **55**, 556–64.
18. Drozd,M., Delhaye,S., Maurin,T., Castagnola,S., Grossi,M., Brau,F., Jarjat,M., Willemsen,R., Capovilla,M., Hukema,R.K., *et al.* (2019) Reduction of Fmr1 mRNA

Levels Rescues Pathological Features in Cortical Neurons in a Model of FXTAS. *Mol. Ther. - Nucleic Acids*, **18**, 546-553.

19. Sofola,O.A., Jin,P., Qin,Y., Duan,R., Liu,H., de Haro,M., Nelson,D.L. and Botas,J. (2007) RNA-Binding Proteins hnRNP A2/B1 and CUGBP1 Suppress Fragile X CGG Premutation Repeat-Induced Neurodegeneration in a Drosophila Model of FXTAS. *Neuron*, **55**, 565–571.
20. Sellier,C., Freyermuth,F., Tabet,R., Tran,T., He,F., Ruffenach,F., Alunni,V., Moine,H., Thibault,C., Page,A., *et al.* (2013) Sequestration of DROSHA and DGCR8 by expanded CGG RNA Repeats Alters microRNA processing in fragile X-associated tremor/ataxia syndrome. *Cell Rep.*, **3**, 869–880.
21. Todd,P.K.K., Oh,S.Y.Y., Krans,A., He,F., Sellier,C., Frazer,M., Renoux,A.J.J., Chen,K. chun, Scaglione,K.M.M., Basrur,V., *et al.* (2013) CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. *Neuron*, **78**, 440–455.
22. Kearse,M.G., Green,K.M., Krans,A., Rodriguez,C.M., Linsalata,A.E., Goldstrohm,A.C. and Todd,P.K. (2016) CGG Repeat-Associated Non-AUG Translation Utilizes a Cap-Dependent Scanning Mechanism of Initiation to Produce Toxic Proteins. *Mol. Cell*, **62**, 314–322.
23. Sellier,C., Buijsen,R.A.M., He,F., Natla,S., Jung,L., Tropel,P., Gaucherot,A., Jacobs,H., Meziane,H., Vincent,A., *et al.* (2017) Translation of Expanded CGG Repeats into FMRpolyG Is Pathogenic and May Contribute to Fragile X Tremor Ataxia Syndrome. *Neuron*, **93**, 331–347.
24. Tassone,F., Beilina,A., Carosi,C., Albertosi,S., Bagni,C., Li,L., Glover,K., Bentley,D. and Hagerman,P.J. (2007) Elevated FMR1 mRNA in premutation carriers is due to increased transcription. *RNA*, **13**, 555-562.
25. Rodriguez,C.M., Wright,S.E., Kearse,M.G., Haenfler,J.M., Flores,B.N., Liu,Y., Ifrim,M.F., Glineburg,M.R., Krans,A., Jafar-Nejad,P., *et al.* (2020) A native function for RAN translation and CGG repeats in regulating fragile X protein synthesis. *Nat. Neurosci.*, **23**, 386-397.
26. Krans,A., Kearse,M.G. and Todd,P.K. (2016) Repeat-associated non-AUG translation from antisense CCG repeats in fragile X tremor/ataxia syndrome. *Ann. Neurol.*, **80**, 871–881.
27. Gao,F.-B.B. and Richter,J.D. (2017) Microsatellite Expansion Diseases: Repeat Toxicity Found in Translation. *Neuron*, **93**, 249-251.
28. Konieczny,P., Stepniak-Konieczna,E., Taylor,K., Sznajder,L.J. and Sobczak,K. (2017)

Autoregulation of MBNL1 function by exon 1 exclusion from MBNL1transcript. *Nucleic Acids Res.*, **45**, 1760–1775.

29. Holm,K.N., Herren,A.W., Taylor,S.L., Randol,J.L., Kim,K., Espinal,G., Martiínez-Cerdeño,V., Pessah,I.N., Hagerman,R.J. and Hagerman,P.J. (2020) Human Cerebral Cortex Proteome of Fragile X-Associated Tremor/Ataxia Syndrome. *Front. Mol. Biosci.*, **7**, 600840.
30. Verma,A.K., Khan,E., Bhagwat,S.R. and Kumar,A. (2020) Exploring the Potential of Small Molecule-Based Therapeutic Approaches for Targeting Trinucleotide Repeat Disorders. *Mol. Neurobiol.*, **57**, 566-584.
31. Gonzalez-Alegre,P. (2019) Recent advances in molecular therapies for neurological disease: Triplet repeat disorders. *Hum. Mol. Genet.*, **28**(R1), R80-R87.
32. Su,Z., Zhang,Y., Gendron,T.F., Bauer,P.O., Chew,J., Yang,W.Y., Fostvedt,E., Jansen-West,K., Belzil,V. V., Desaro,P., *et al.* (2014) Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in c9FTD/ALS. *Neuron*, **84**, 239.
33. Yang,W.Y., He,F., Strack,R.L., Oh,S.Y., Frazer,M., Jaffrey,S.R., Todd,P.K. and Disney,M.D. (2016) Small Molecule Recognition and Tools to Study Modulation of r(CGG)exp in Fragile X-Associated Tremor Ataxia Syndrome. *ACS Chem. Biol.*, **11**, 2456-2465.
34. Yang,W.Y., Wilson,H.D., Velagapudi,S.P. and Disney,M.D. (2015) Inhibition of non-ATG translational events in cells via covalent small molecules targeting RNA. *J. Am. Chem. Soc.*, **137**, 5336-5345.
35. Green,K.M., Sheth,U.J., Flores,B.N., Wright,S.E., Sutter,A.B., Kearse,M.G., Barmada,S.J., Ivanova,M.I. and Todd,P.K. (2019) High-throughput screening yields several small-molecule inhibitors of repeat-associated non-AUG translation. *J. Biol. Chem.*, **294**, 18624-18638.
36. Derbis,M., Kul,E., Niewiadomska,D., Sekrecki,M., Piasecka,A., Taylor,K., Hukema,R.K., Stork,O. and Sobczak,K. (2021) Short antisense oligonucleotides alleviate the pleiotropic toxicity of RNA harboring expanded CGG repeats. *Nat. Commun.*, **12**, 1265.
37. Hagerman,R.J., Berry-Kravis,E., Hazlett,H.C., Bailey,D.B., Moine,H., Kooy,R.F., Tassone,F., Gantois,I., Sonenberg,N., Mandel,J.L., *et al.* (2017) Fragile X syndrome. *Nat. Rev. Dis. Prim.*, **3**, 17065.
38. Asamitsu,S., Yabuki,Y., Ikenoshita,S., Kawakubo,K., Kawasaki,M., Usuki,S., Nakayama,Y., Adachi,K., Kugoh,H., Ishii,K., *et al.* (2021) CGG repeat RNA G-quadruplexes interact with FMRpolyG to cause neuronal dysfunction in fragile X-related

tremor/ataxia syndrome. *Sci. Adv.*, **7**.

39. Nakatani,K., He,H., Uno,S.N., Yamamoto,T. and Dohno,C. (2008) Synthesis of dimeric 2-amino-1,8-naphthyridine and related DNA-binding molecules. *Curr. Protoc. Nucleic Acid Chem.*, **32**, 8.6.1-8.6.21.
40. Dohno,C., Kohyama,I., Hong,C. and Nakatani,K. (2012) Naphthyridine tetramer with a pre-organized structure for 1:1 binding to a CGG/CGG sequence. *Nucleic Acids Res.*, **40**, 2771–2781.
41. Mukherjee,S., Dohno,C. and Nakatani,K. (2017) Design and Synthesis of Cyclic Mismatch-Binding Ligands (CMBLs) with Variable Linkers by Ring-Closing Metathesis and their Photophysical and DNA Repeat Binding Properties. *Chem. - A Eur. J.*, **23**, 11385–11396.
42. Mukherjee,S., Dohno,C., Asano,K. and Nakatani,K. (2016) Cyclic mismatch binding ligand CMBL4 binds to the 5′-T-3′/5′-GG-3′ site by inducing the flipping out of thymine base. *Nucleic Acids Res.*, **44**, gkw672.
43. Mukherjee,S., Błaszczyk,L., Rypniewski,W., Falschlunger,C., Micura,R., Murata,A., Dohno,C., Nakatani,K. and Kiliszek,A. (2019) Structural insights into synthetic ligands targeting A-A pairs in disease-related CAG RNA repeats. *Nucleic Acids Res.*, **47**, 10906-10913.
44. Dohno,C., Kimura,M. and Nakatani,K. (2018) Restoration of Ribozyme Tertiary Contact and Function by Using a Molecular Glue for RNA. *Angew. Chemie - Int. Ed.*, **57**, 506–510.
45. Garcia-Arocena,D., Yang,J.E., Brouwer,J.R., Tassone,F., Iwahashi,C., Berry-Kravis,E.M., Goetz,C.G., Sumis,A.M., Zhou,L., Nguyen,D. V., *et al.* (2010) Fibroblast phenotype in male carriers of FMR1 premutation alleles. *Hum. Mol. Genet.*, **19**, 299-312.
46. Rovozzo,R., Korza,G., Baker,M.W., Li,M., Bhattacharyya,A., Barbarese,E. and Carson,J.H. (2016) CGG repeats in the 5'UTR of FMR1 RNA regulate translation of other RNAs localized in the same RNA granules. *PLoS One*, **11**, e0168204.
47. Derbis,M., Konieczny,P., Walczak,A., Sekrecki,M. and Sobczak,K. (2018) Quantitative evaluation of toxic polyglycine biosynthesis and aggregation in cell models expressing expanded CGG repeats. *Front. Genet.*, **9**, 216.
48. Sznajder,L.J., Michalak,M., Taylor,K., Cywoniuk,P., Kabza,M., Wojtkowiak-Szlachcic,A., Matlłoka,M., Konieczny,P. and Sobczak,K. (2016) Mechanistic determinants of MBNL activity. *Nucleic Acids Res.*, **44**, 10326-10342.
49. Buijsen,R.A.M., Sellier,C., Severijnen,L.A.W.F.M., Oulad-Abdelghani,M.,

Verhagen,R.F.M., Berman,R.F., Charlet-Berguerand,N., Willemsen,R. and Hukema,R.K. (2014) FMRpolyG-positive inclusions in CNS and non-CNS organs of a fragile X premutation carrier with fragile X-associated tremor/ataxia syndrome. *Acta Neuropathol. Commun.*, **2**, 162.

50. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods*, **25**, 402-408.
51. Wojtkowiak-Szlachcic,A., Taylor,K., Stepniak-Konieczna,E., Sznajder,L.J., Mykowska,A., Sroka,J., Thornton,C.A. and Sobczak,K. (2015) Short antisense-locked nucleic acids (all-LNAs) correct alternative splicing abnormalities in myotonic dystrophy. *Nucleic Acids Res.*, **43**, 3318-3331.
52. Stepniak-Konieczna,E., Konieczny,P., Cywoniuk,P., Dluzewska,J. and Sobczak,K. (2020) AON-induced splice-switching and DMPK pre-mRNA degradation as potential therapeutic approaches for Myotonic Dystrophy type 1. *Nucleic Acids Res.*, **48**, 2531-2543.
53. Fiszer,A., Mykowska,A. and Krzyzosiak,W.J. (2011) Inhibition of mutant huntingtin expression by RNA duplex targeting expanded CAG repeats. *Nucleic Acids Res.*, **39**, 5578–5585.
54. Napierala,M., Michalowski,D., de Mezer,M. and Krzyzosiak,W.J. (2005) Facile FMR1 mRNA structure regulation by interruptions in CGG repeats. *Nucleic Acids Res.*, **33**, 451-463.
55. Kiliszek,A., Kierzek,R., Krzyzosiak,W. and Rypniewski,W. (2011) Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. *Nucleic Acids Res*, **39**, 7308–7315.
56. Zumwalt,M., Ludwig,A., Hagerman,P.J. and Dieckmann,T. (2007) Secondary structure and dynamics of the r(CGG) repeat in the mRNA of the Fragile X Mental Retardation 1 (FMR1) gene. *RNA Biol.*, **4**, 93-100.
57. Tomé,S. and Gourdon,G. (2020) DM1 phenotype variability and triplet repeat instability: Challenges in the development of new therapies. *Int. J. Mol. Sci.*, 10.3390/ijms21020457.
58. Konieczny,P., Stepniak-Konieczna,E. and Sobczak,K. (2018) MBNL expression in autoregulatory feedback loops. *RNA Biol*. **15**, 1-8.
59. Konieczny,P., Stepniak-Konieczna,E. and Sobczak,K. (2014) MBNL proteins and their target RNAs, interaction and splicing regulation. *Nucleic Acids Res.*, **42**, 10873-10887.
60. Konieczny,P., Stepniak-Konieczna,E. and Sobczak,K. (2016) Modified Antisense

Oligonucleotides and Their Analogs in Therapy of Neuromuscular Diseases. In. Jurga,S., Erdmann,V., Barciszewski,J. (eds), Modified Nucleic Acids in Biology and Medicine. RNA Technologies. Springer, Cham, pp. 243–271.

61. Pan,L. and Feigin,A. (2021) Huntington's Disease: New Frontiers in Therapeutics. *Curr. Neurol. Neurosci. Rep.*, **21**, 10.
62. Kozlowski,P., de Mezer,M. and Krzyzosiak,W.J. (2010) Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.*, **38**, 4027-4039.
63. Sone,J., Mitsuhashi,S., Fujita,A., Mizuguchi,T., Hamanaka,K., Mori,K., Koike,H., Hashiguchi,A., Takashima,H., Sugiyama,H., *et al.* (2019) Long-read sequencing identifies GGC repeat expansions in NOTCH2NLC associated with neuronal intranuclear inclusion disease. *Nat. Genet.*, **51**, 1215-1221.
64. Ishiura,H., Shibata,S., Yoshimura,J., Suzuki,Y., Qu,W., Doi,K., Almansour,M.A., Kikuchi,J.K., Taira,M., Mitsui,J., *et al.* (2019) Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nat. Genet.*, **51**, 1222-1232.
65. Deng,J., Yu,J., Li,P., Luan,X., Cao,L., Zhao,J., Yu,M., Zhang,W., Lv,H., Xie,Z., *et al.* (2020) Expansion of GGC Repeat in GIPC1 Is Associated with Oculopharyngodistal Myopathy. *Am. J. Hum. Genet.*, **106**, 793-804.
66. Sznajder,Ł.J. and Swanson,M.S. (2019) Short tandem repeat expansions and RNA-mediated pathogenesis in myotonic dystrophy. *Int. J. Mol. Sci.*, **20**, 3365.
67. Hunsaker,M.R., Greco,C.M., Spath,M.A., Smits,A.P.T., Navarro,C.S., Tassone,F., Kros,J.M., Severijnen,L.A., Berry-Kravis,E.M., Berman,R.F., *et al.* (2011) Widespread non-central nervous system organ pathology in fragile X premutation carriers with fragile X-associated tremor/ataxia syndrome and CGG knock-in mice. *Acta Neuropathol.*, **122**, 467-479.
68. Chung,C.G., Lee,H. and Lee,S.B. (2018) Mechanisms of protein toxicity in neurodegenerative diseases. *Cell. Mol. Life Sci.*, **75**, 3159-3180.
69. Hukema,R.K., Buijsen,R.A.M., Schonewille,M., Raske,C., Severijnen,L.A.W.F.M., Nieuwenhuizen-Bakker,I., Verhagen,R.F.M., van Dessel,L., Maas,A., Charlet-Berguerand,N., *et al.* (2015) Reversibility of neuropathology and motor deficits in an inducible mouse model for FXTAS. *Hum. Mol. Genet.*, **24**, 4948-4957.

Figure legends

Figure 1. CMBL4a binds with high affinity to rCGG triplet repeats *in vitro* and attenuates FMRpolyG signal in cell culture. (**A**) Schematic representations of NCD, Z-NCTS and CMBL4a structures. (**B**) SPR single cycle kinetic analysis of NCD, Z-NCTS and CMBL4a to the r(CGG)$_9$ immobilized sensor surface. NCD was added sequentially to 0.5, 1.0, 2.0, 4.0 and 8.0 μM while Z-NCTS and CMBL4a to 0.25, 0.5, 1.0, 2.0 and 4.0 μM. (**C**) Graphical representation of filter binding assay (FBA) results showing an *in vitro* interaction of NCD, Z-NCTS and CMBL4a compounds with r(CGG)$_{20}$. Mean values from two experiments ± standard deviation (SD) are shown on the graph. (**D**) Schematic representation of GFP and GFP-fused vectors used in the study. (**E**) Visualization of either GFP or FMRpolyG-GFP signals in COS7 cells 48 h following transfection with p*EGFP*-C1 or 5'(CGG)99-*GFP*(+1) vectors and delivery of NCD, Z-NCTS, CMBL4a or equivalent solvent amounts (H_2O; Ctrl) 18 h post transfection. The graph in the right panel shows quantification of FMRpolyG-GFP signals. Mean values from at least two photos, each from a different biological replicate, with standard deviation (SD) are shown on the graph. Statistical significance was determined by unpaired two-tailed Student's *t*-test (*** indicates $p < 0.001$). (**F**) Representative FMRpolyG, Hoechst or merged FMRpolyG/Hoechst images of COS7 cells 48 hours post transfection with 5'(CGG)$_{99}$-*GFP*(+1) that were treated with either H_2O or CMBL4a. Hoechst was added to the cells 10 minutes prior to the analysis. Representative large FMRpolyG inclusions stained with Hoechst and Hoechst-negative smaller inclusions are marked with red and yellow arrows, respectively. Bar, 50 μm.

Figure 2. Comparison of CMBL4a, 4b and 4c *in vitro* binding to CGG repeats. (**A**) General structure of CMBLs as well as detailed representations of CMBL4b and 4c used in the study. **(B)** SPR single cycle kinetic analysis of CMBL4a, 4b and 4c to the r(CGG)$_9$ immobilized sensor surface. CMBLs were added sequentially to 0.25, 0.5, 1.0, 2.0 and 4.0 μM. (**C**) Comparison of affinities of CMBL4a, 4b and 4c to r(CGG)$_{20}$ based on FBA. Mean values from three experiments ± standard deviation (SD) are shown on the graph.

Figure 3. Comparison of CMBLs' activities in COS7 cells using luciferase assay. (**A**) Schematic representation of ATG(CGG)$_{99}$-*luc2*(+1) dual luciferase vector used in the study. The vector contains 99 CGG repeats fused with *luc2*, from which a fusion protein of FMRpolyG and Firefly luciferase is generated (FMRpolyG-Firefly) as well as *hRluc-neo* sequence, from which Renilla luciferase is produced independently from FMRpolyG-Firefly. (**B**, **C**) Graphs showing luminescence signals from COS7 cells 56 h after transfection with ATG(CGG)$_{99}$-

luc2(+1) and delivery of CMBL compounds at 0.25, 0.5, 1 or 2 μM concentration 6 h post plasmid addition. FMRpolyG-Firefly luciferase signals normalized to Renilla luminescence are shown in **B** while Renilla values that represent the content of living cells and toxicity of the compounds in **C**. Mean values from three experiments with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test (* indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$). CMBL4a, 4b and 4c that inhibit synthesis of FMRpolyG-Firefly without causing cellular toxicity are marked in blue in **B**. CMBL3a, 3b, 5c, 6b that lower relative FMRpolyG-Firefly contents and affect cell survival are marked red in **C**.

Figure 4. CMBL4c efficiently reduces FMRpolyG content in COS7 cells. (**A**) Representative images (upper panel), graphical representations of FMRpolyG-GFP fluorescence signals (lower left panel), total inclusion numbers (lower middle panel) and percentages of inclusion size areas (lower right panel) following transfection of COS7 cells with 5'$(CGG)_{99}$-*GFP*(+1) and delivery of increasing concentrations of CMBL4A, as indicated in the figure. Mean values from two biological replicates ± standard deviation (SD) are shown on the graphs. (**B**, **C**) Representative images (**B**) and graphs (**C**) showing GFP fluorescence signals 80 h following delivery of either ATG$(CGG)_{99}$-*GFP*(+1) or 5'$(CGG)_{99}$-*GFP*(+1) plasmids. CMBL4b or 4c compounds at indicated concentrations were added 6 h after transfection. Mean values from three experiments ± standard deviation (SD) are shown on the graphs. (**D**) Fluorescence signals after transfection with ATG$(CGG)_{99}$-*GFP*(+1) and delivery of either DMSO (Ctrl) or CMBL4c at 2 μM concentration 12 or 24 h post transfection. Statistical significance was determined by unpaired two-tailed Student's *t*-test (fluorescence signals and inclusion numbers in **A**, **C** and **D**) or by Kruskal-Wallis test (inclusion areas in **A**). * indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$.

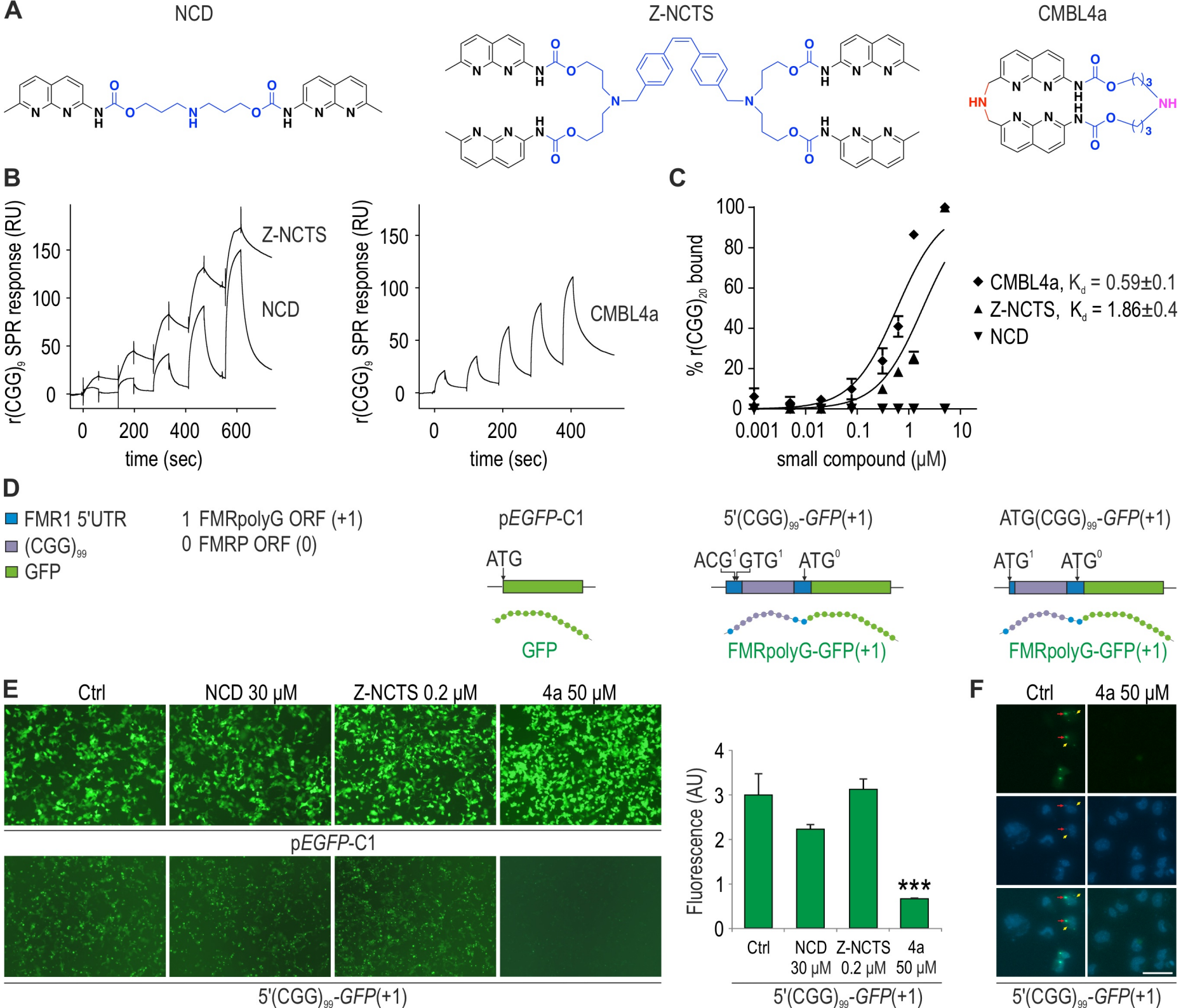
Figure 5. CMBL4c interacts with selected RNA and DNA repeat sequences. (**A**, **B**) SPR single cycle kinetic analysis of CMBL4c interaction with DNA sequences (**A**) and RNA sequences (**B**) as indicated in the figure. CMBL4c was added sequentially to 0.25, 0.5, 1.0, 2.0 and 4.0 μM.

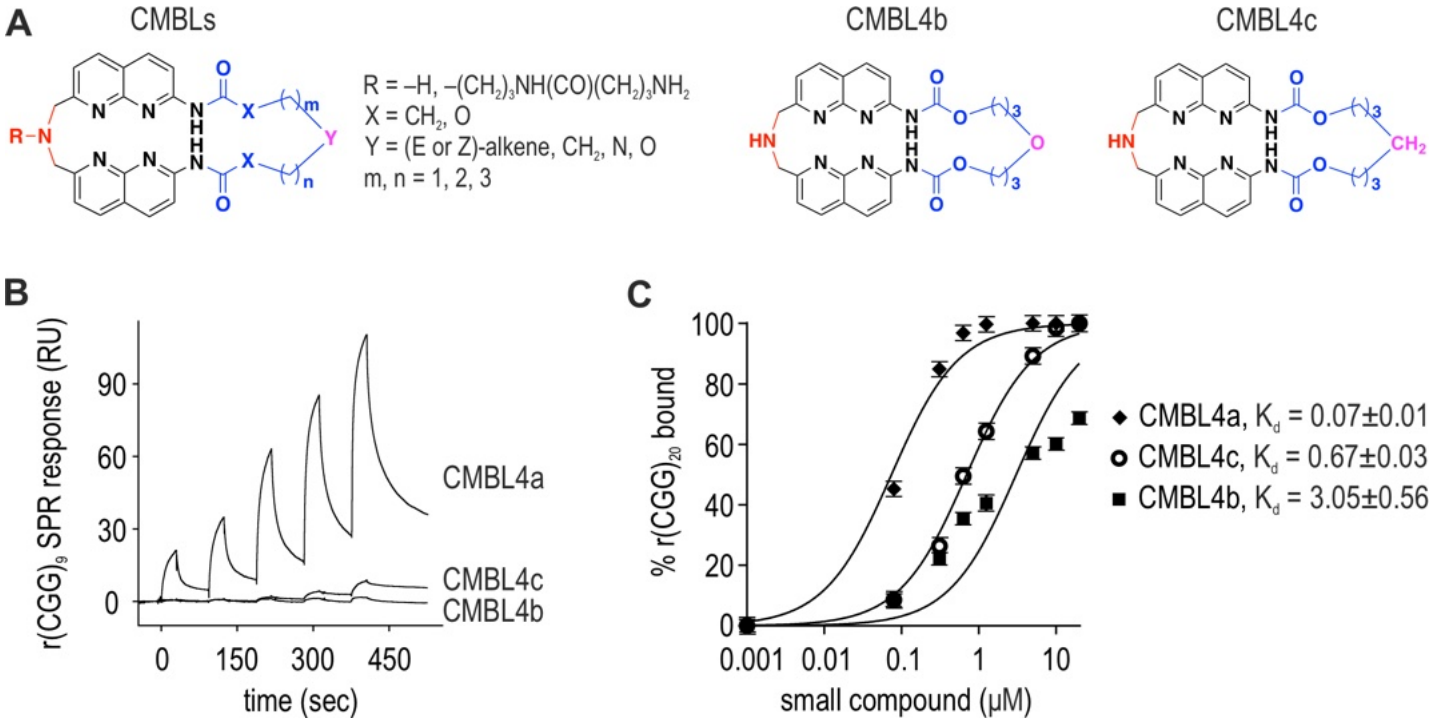
Figure 6. CMBL4c does not reduce the level of mRNAs containing CGG_{exp}, CAG_{exp} and CUG_{exp}. (**A**) Semi-quantitative multiplex RT-PCRs showing *FMR1*, 5'$(CGG)_{99}$-*GFP*(+1) and *GAPDH* mRNA levels in COS7 cells 40 h after transfection with 5'$(CGG)_{99}$-*GFP*(+1) vector and 22 h incubation with CMBL4c or 5c. Mock represents signals from control untransfected

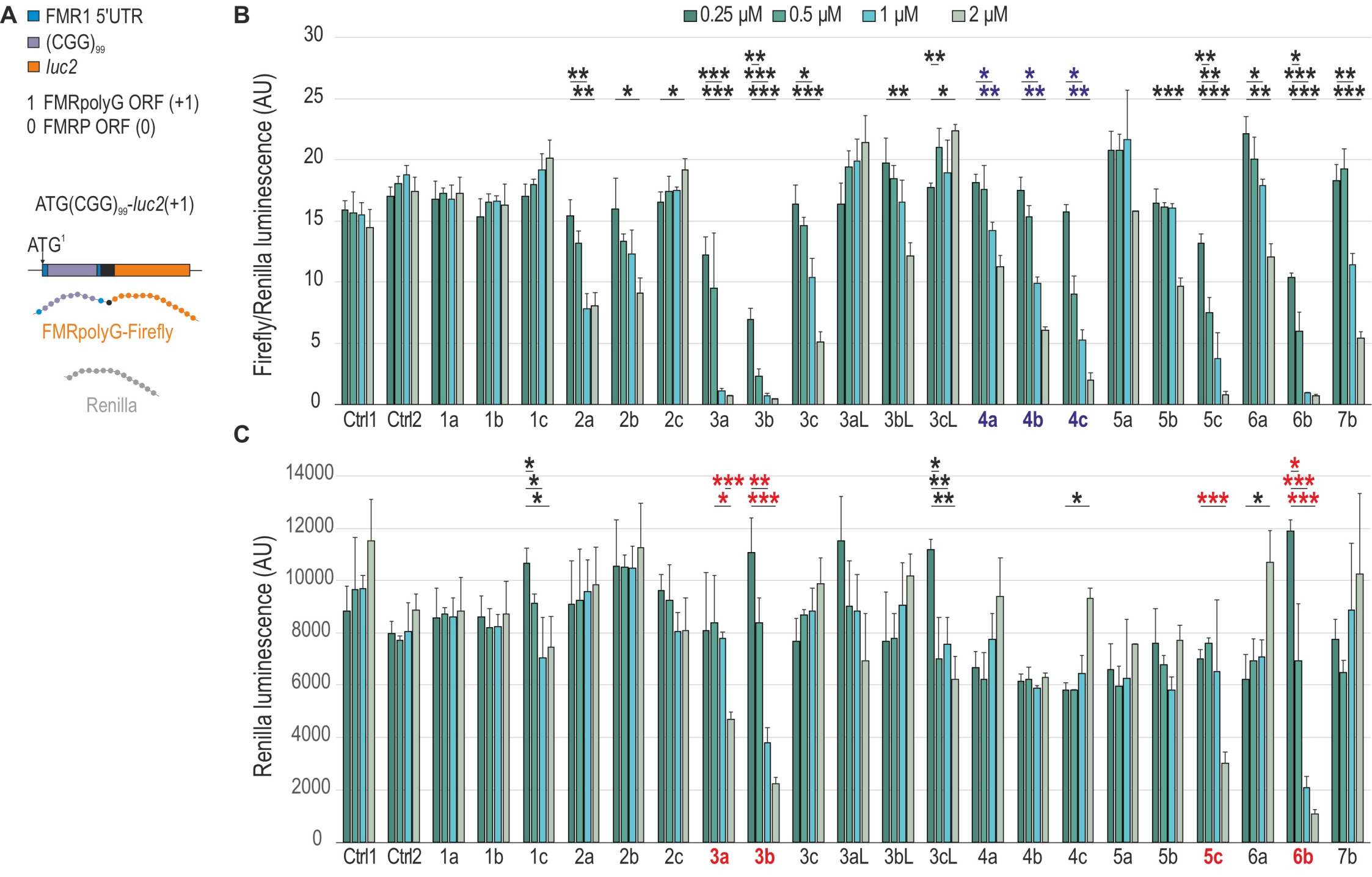
cells and DNA Ctrl lane shows RT-PCR signals from a sample in which 5'$(CGG)_{99}$-*GFP*(+1) plasmid was used as a template. (**B**) Quantitative analyses of multiplex RT-PCR signals shown in **A**. *FMR1* and 5'$(CGG)_{99}$-*GFP*(+1) mRNA levels were normalized to *GAPDH* content. (**C**) RT-qPCR analyses showing *AR*, *CARM1*, *DMPK*, *FMR1*, *QKI* and *VKORC1L1* transcript levels in control fibroblasts (GM07492) 48 h following delivery of DMSO, CMBL1(b and c) or 4c. Mean values normalized to *GAPDH* mRNA levels with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test (* indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$).

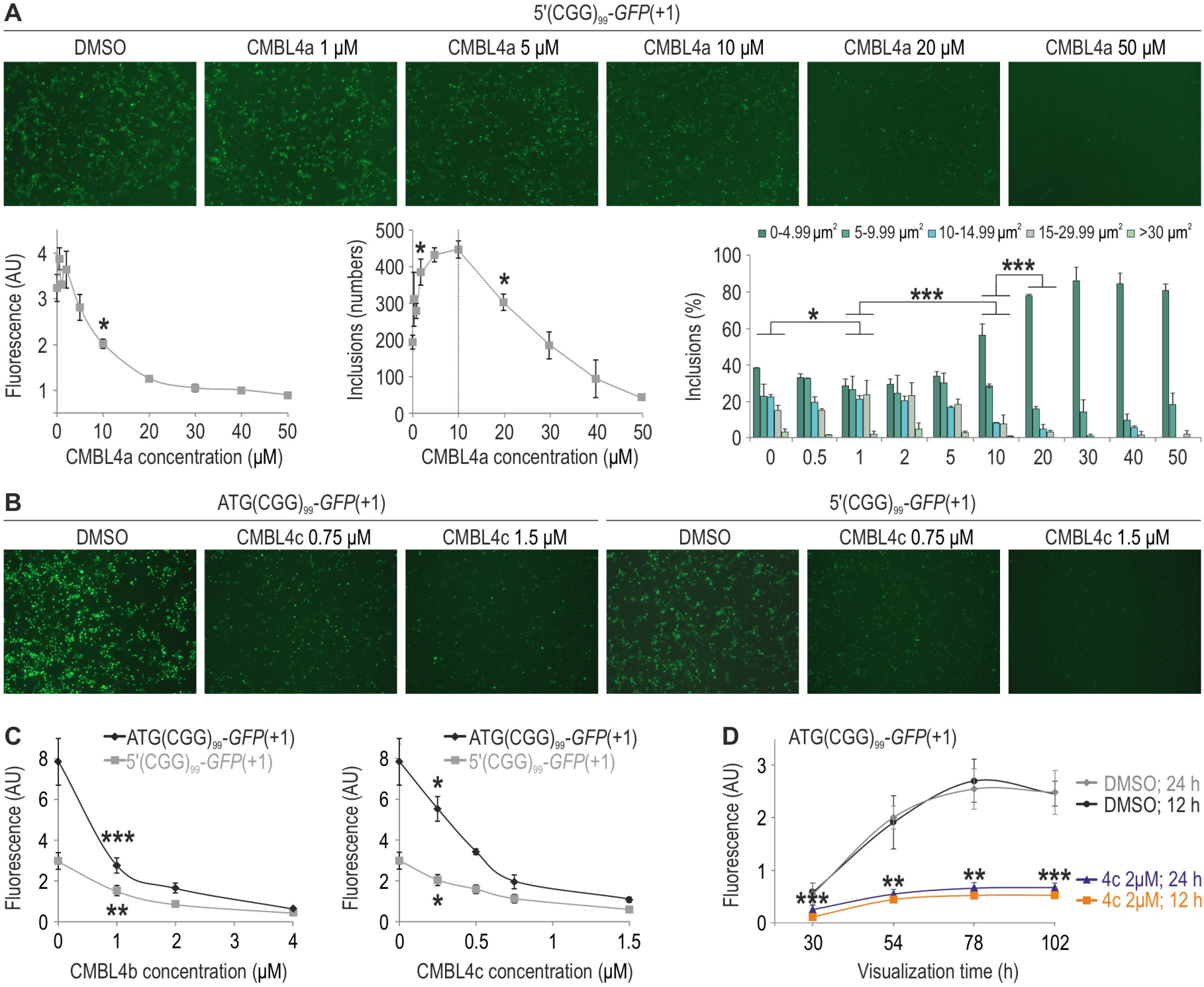
Figure 7. CMBL4c increases viability and reduces cytotoxicity and apoptosis in an FXTAS HEK293T cell model. (**A**, **B**) Visualization (**A**) and distribution area plots (**B**) of FMRpolyG-GFP inclusions after administration of 5'$(CGG)_{99}$-*GFP*(+1) and delivery of either DMSO or CMBL4c 18 or 30 h post transfection to COS7 cells. (**C**) Schematic representation of 5'$(CGG)_{16}$ and 5'$(CGG)_{99}$ vectors used in the study. (**D**) Fluorescence microscopy of COS7 cells immunostained for lamin (red), FMRpolyG (green) and DAPI (blue) 48 h after delivery of 5'$(CGG)_{99}$ plasmid. (**E**) Graphs showing viability, cytotoxicity and apoptosis of HEK293T cells following delivery of 5'$(CGG)_{16}$ or 5'$(CGG)_{99}$ and addition of DMSO or CMBL4c, 6 or 24 h post transfection. Mean values with standard deviation (SD) are shown on the graphs. Statistical significance was determined by Kruskal-Wallis test in **B** or unpaired two-tailed Student's *t*-test in **E**. * indicates $p< 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$.

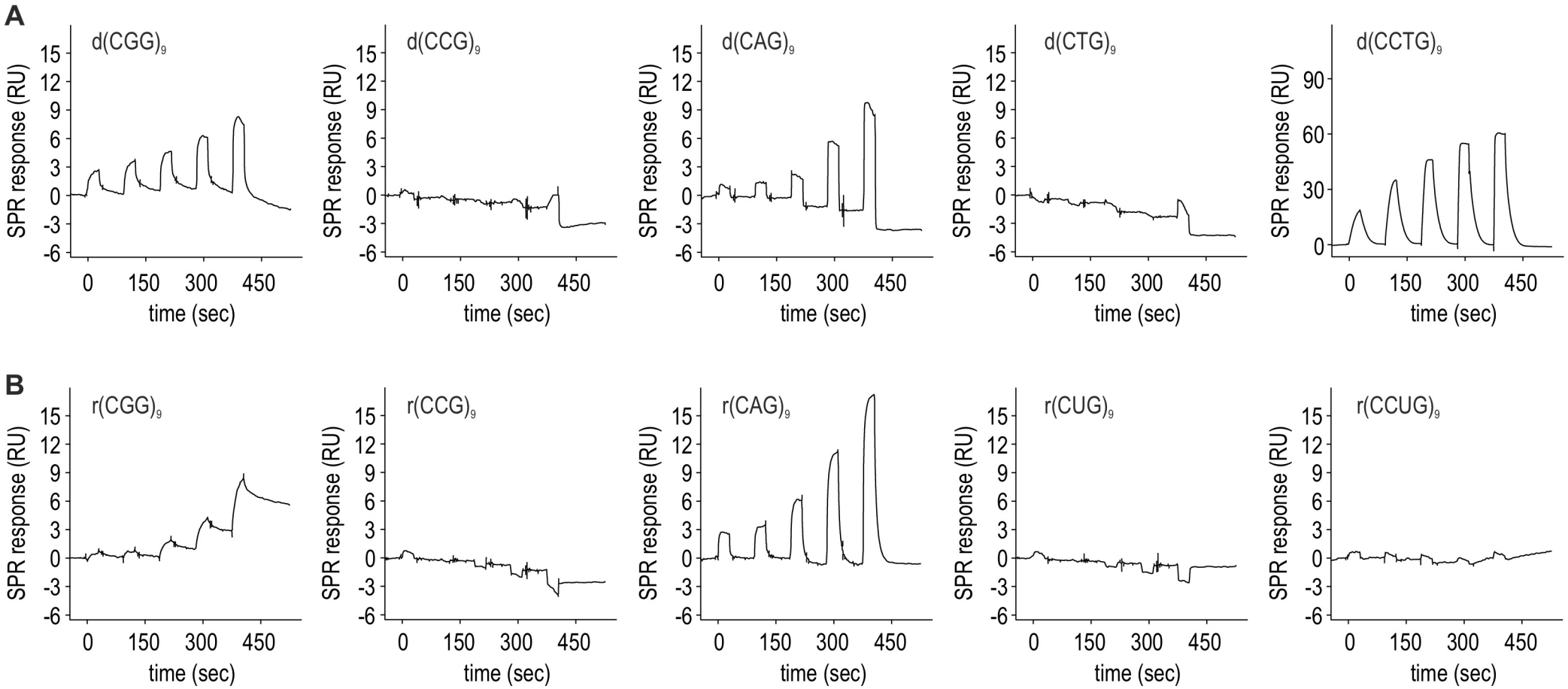
Figure 8. CMBL4c attenuates translation of FMRP. (**A**) Schematic representation of ATG-*luc2* and 5'$(CGG)_n$-*luc2*(0) dual luciferase vectors, differing in the number of CGG triplet repeats, used in **B**. (**B**) Luminescence signals of COS7 cells transfected with a control ATG-*luc2* or 5'$(CGG)_n$-*luc2*(0) plasmids and treated with CMBL4c at 0.25, 0.5 or 1.5 μM concentration. (**C**) Immunoblotting of control (Ctrl) and FXTAS fibroblast lysates with FMRP, GAPDH (normalization control), QKI (8 CGG repeats) and SLC40A1 (7 CGG repeats) antibodies. Fibroblasts were treated with DMSO, CMBL4a or 4c (2 μM) for 48 h. Mean values normalized to GAPDH protein levels with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test. * indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$.

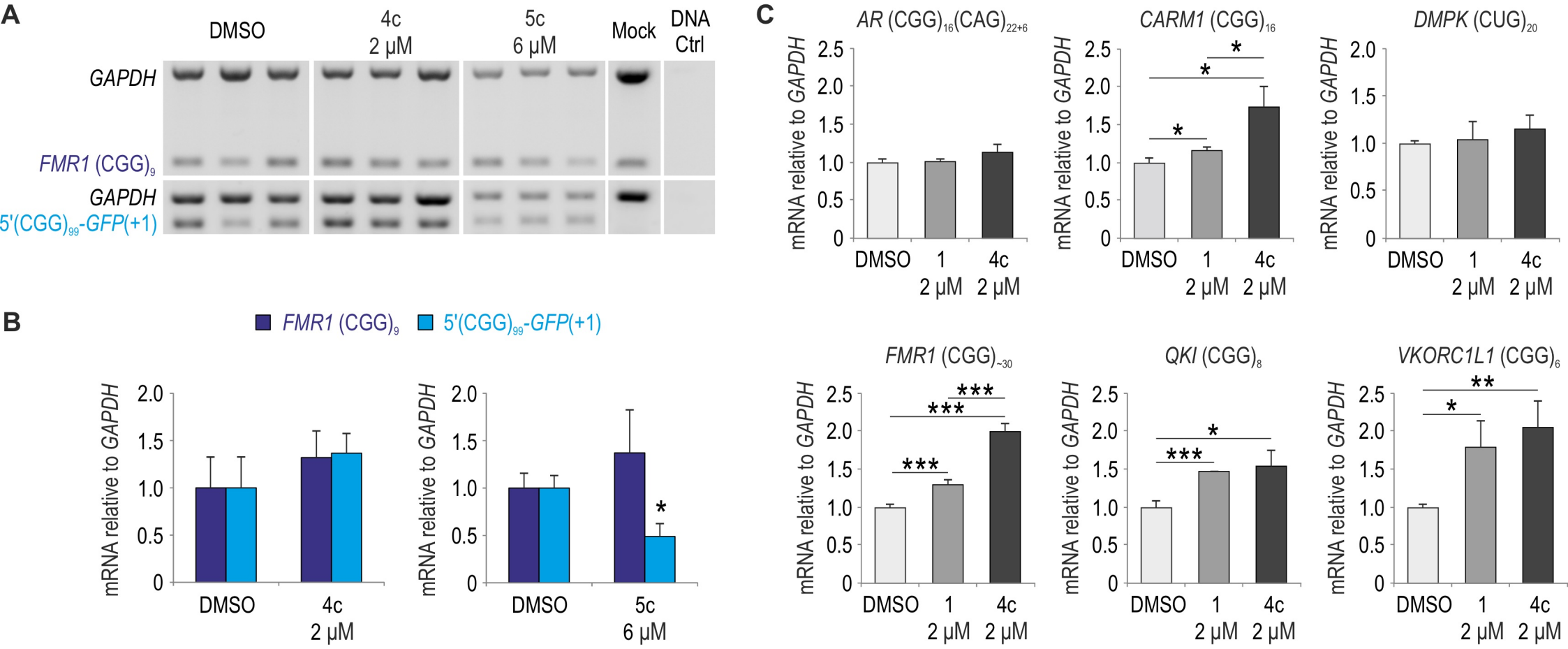

A
NCD
Z-NCTS
CMBL4a
B
r(CGG)9 SPR response (RU)
Z-NCTS
NCD
time (sec)
CMBL4a
C
% r(CGG)20 bound
small compound (μM)
CMBL4a, Kd = 0.59±0.1
Z-NCTS, Kd = 1.86±0.4
NCD
D
FMR1 5'UTR
(CGG)99
GFP
1 FMRpolyG ORF (+1)
0 FMRP ORF (0)
pEGFP-C1
5'(CGG)99-GFP(+1)
ATG(CGG)99-GFP(+1)
ATG
ACG GTG ATG
GFP
FMRpolyG-GFP(+1)
E
Ctrl
NCD 30 μM
Z-NCTS 0.2 μM
4a 50 μM
pEGFP-C1
5'(CGG)99-GFP(+1)
Fluorescence (AU)
***
F
Ctrl
4a 50 μM
5'(CGG)99-GFP(+1)

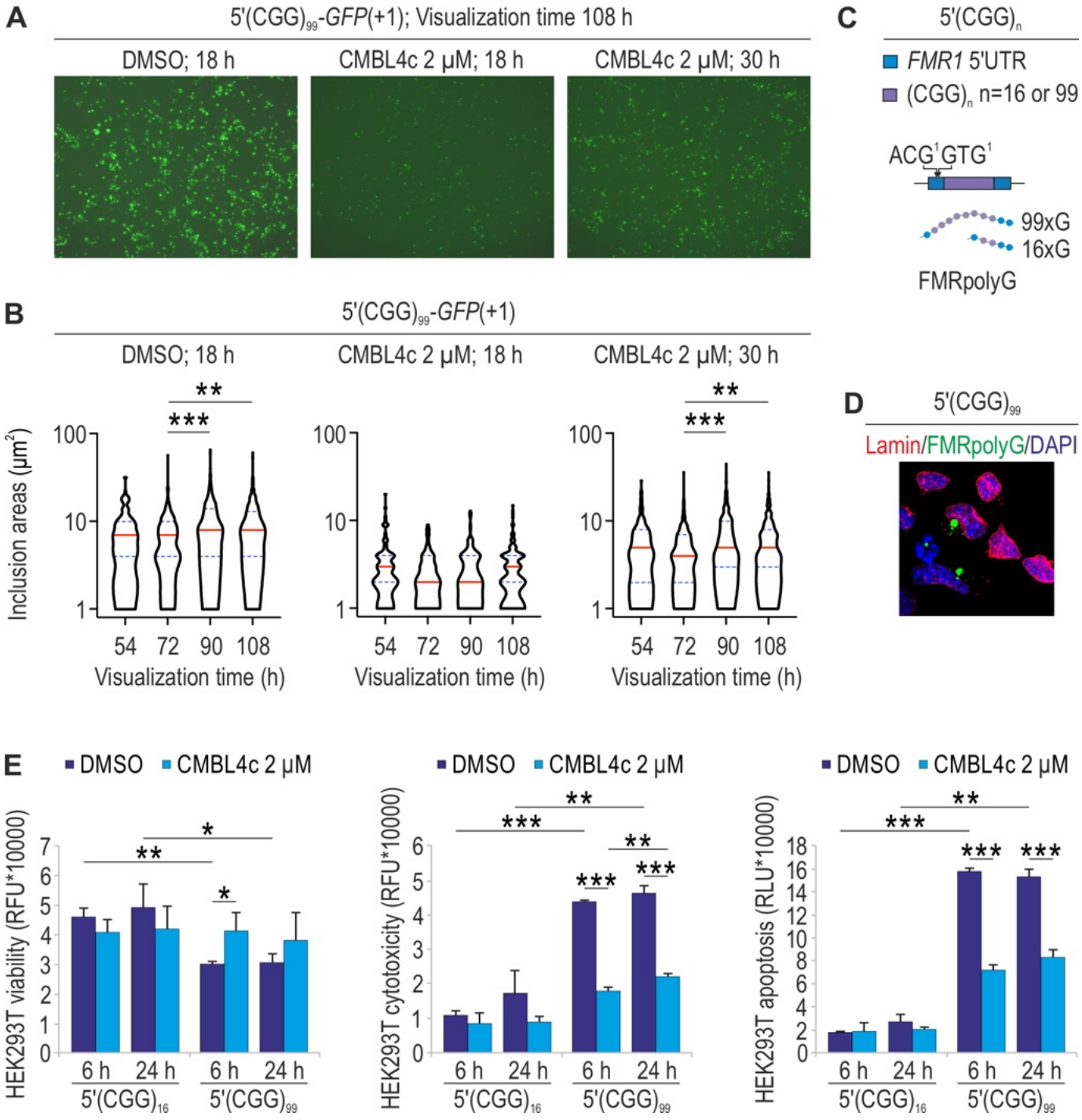

A
CMBLs
CMBL4b
CMBL4c
R = –H, –(CH2)3NH(CO)(CH2)3NH2
X = CH2, O
Y = (E or Z)-alkene, CH2, N, O
m, n = 1, 2, 3
B
r(CGG)9 SPR response (RU)
time (sec)
CMBL4a
CMBL4c
CMBL4b
C
% r(CGG)20 bound
small compound (μM)
CMBL4a, Kd = 0.07±0.01
CMBL4c, Kd = 0.67±0.03
CMBL4b, Kd = 3.05±0.56

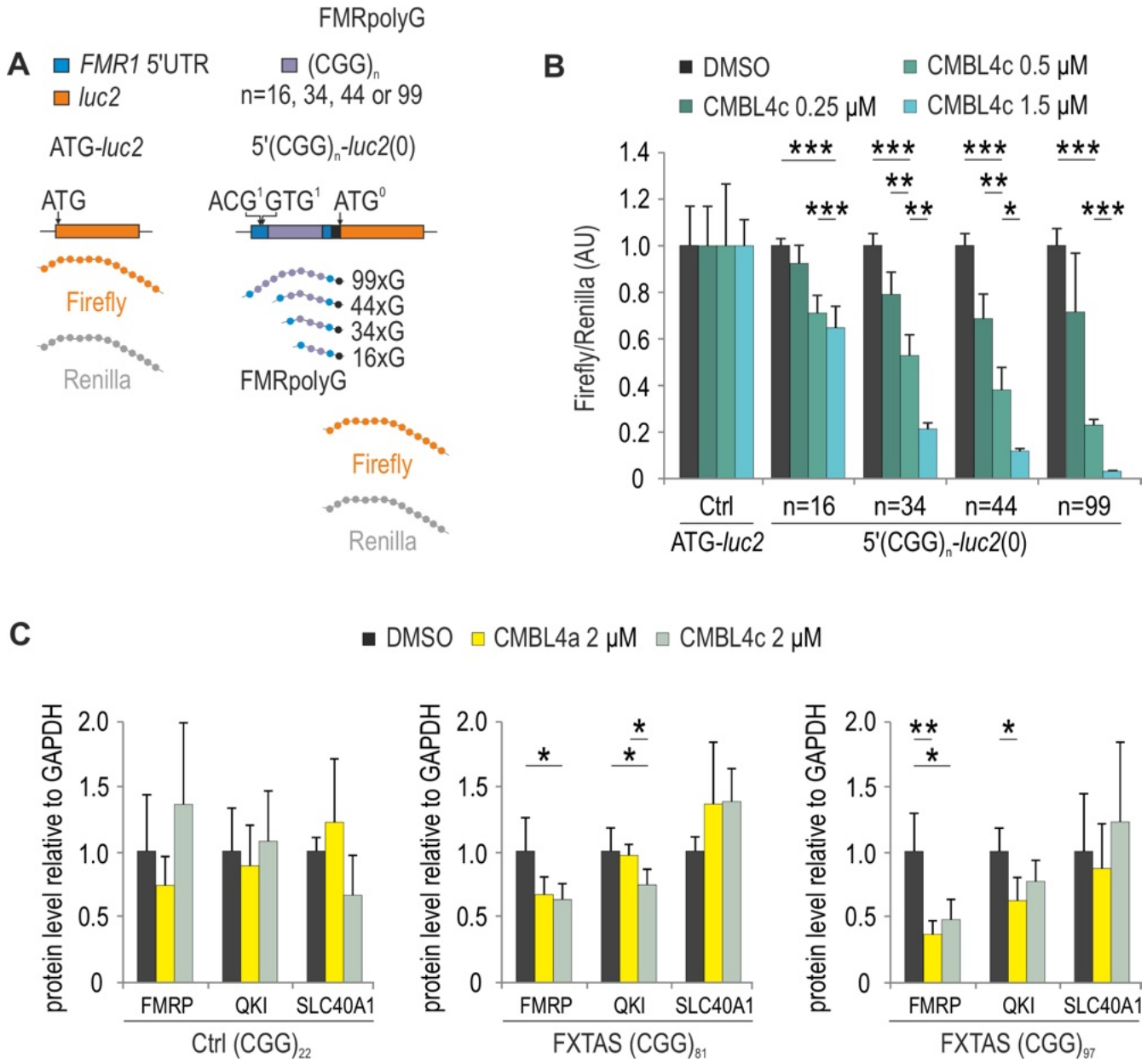

A
FMR1 5'UTR
(CGG)99
luc2
1 FMRpolyG ORF (+1)
0 FMRP ORF (0)
ATG(CGG)99-luc2(+1)
ATG1
FMRpolyG-Firefly
Renilla
B
0.25 µM
0.5 µM
1 µM
2 µM
Firefly/Renilla luminescence (AU)
0
5
10
15
20
25
30
Ctrl1 Ctrl2 1a 1b 1c 2a 2b 2c 3a 3b 3c 3aL 3bL 3cL 4a 4b 4c 5a 5b 5c 6a 6b 7b
C
Renilla luminescence (AU)
0
2000
4000
6000
8000
10000
12000
14000
Ctrl1 Ctrl2 1a 1b 1c 2a 2b 2c 3a 3b 3c 3aL 3bL 3cL 4a 4b 4c 5a 5b 5c 6a 6b 7b


A
5'(CGG)99-GFP(+1)
DMSO
CMBL4a 1 µM
CMBL4a 5 µM
CMBL4a 10 µM
CMBL4a 20 µM
CMBL4a 50 µM
Fluorescence (AU)
CMBL4a concentration (µM)
Inclusions (numbers)
CMBL4a concentration (µM)
0-4.99 µm²
5-9.99 µm²
10-14.99 µm²
15-29.99 µm²
>30 µm²
Inclusions (%)
B
ATG(CGG)99-GFP(+1)
5'(CGG)99-GFP(+1)
DMSO
CMBL4c 0.75 µM
CMBL4c 1.5 µM
C
ATG(CGG)99-GFP(+1)
5'(CGG)99-GFP(+1)
Fluorescence (AU)
CMBL4b concentration (µM)
CMBL4c concentration (µM)
D
ATG(CGG)99-GFP(+1)
DMSO; 24 h
DMSO; 12 h
4c 2µM; 24 h
4c 2µM; 12 h
Fluorescence (AU)
Visualization time (h)


A
d(CGG)9
d(CCG)9
d(CAG)9
d(CTG)9
d(CCTG)9
B
r(CGG)9
r(CCG)9
r(CAG)9
r(CUG)9
r(CCUG)9
SPR response (RU)
time (sec)
0
150
300
450


A
DMSO
4c
2 μM
5c
6 μM
Mock
DNA
Ctrl
GAPDH
FMR1 (CGG)9
GAPDH
5'(CGG)99-GFP(+1)
B
FMR1 (CGG)9
5'(CGG)99-GFP(+1)
mRNA relative to GAPDH
DMSO
4c
2 μM
5c
6 μM
C
AR (CGG)16(CAG)22+6
CARM1 (CGG)16
DMPK (CUG)20
FMR1 (CGG)~30
QKI (CGG)8
VKORC1L1 (CGG)6
DMSO
1
2 μM
4c
2 μM


A
5'(CGG)99-GFP(+1); Visualization time 108 h
DMSO; 18 h
CMBL4c 2 μM; 18 h
CMBL4c 2 μM; 30 h
B
5'(CGG)99-GFP(+1)
DMSO; 18 h
CMBL4c 2 μM; 18 h
CMBL4c 2 μM; 30 h
Inclusion areas (μm2)
100
10
1
**
***
54 72 90 108
Visualization time (h)
C
5'(CGG)n
FMR1 5'UTR
(CGG)n n=16 or 99
ACG1GTG1
99xG
16xG
FMRpolyG
D
5'(CGG)99
Lamin/FMRpolyG/DAPI
E
DMSO
CMBL4c 2 μM
HEK293T viability (RFU*10000)
HEK293T cytotoxicity (RFU*10000)
HEK293T apoptosis (RLU*10000)
*
**
***
6 h 24 h 6 h 24 h
5'(CGG)16
5'(CGG)99


A
FMR1 5'UTR
luc2
(CGG)n
n=16, 34, 44 or 99
ATG-luc2
5'(CGG)n-luc2(0)
ATG
ACG1GTG1 ATG0
Firefly
Renilla
99xG
44xG
34xG
16xG
FMRpolyG
FMRpolyG
Firefly
Renilla
B
DMSO
CMBL4c 0.5 µM
CMBL4c 0.25 µM
CMBL4c 1.5 µM
Firefly/Renilla (AU)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
***
**
*
Ctrl
n=16
n=34
n=44
n=99
ATG-luc2
5'(CGG)n-luc2(0)
C
DMSO
CMBL4a 2 µM
CMBL4c 2 µM
protein level relative to GAPDH
2.0
1.5
1.0
0.5
0
FMRP
QKI
SLC40A1
Ctrl (CGG)22
FXTAS (CGG)81
FXTAS (CGG)97

SUPPLEMENTARY MATERIAL

Cyclic mismatch binding ligands interact with disease-associated CGG trinucleotide repeats in RNA and suppress their translation

Patryk Konieczny[1,2]*, Sanjukta Mukherjee[3,4]*, Ewa Stepniak-Konieczna[1], Katarzyna Taylor[1], Daria Niewiadomska[1], Agnieszka Piasecka[1], Agnieszka Walczak[1], Anna Baud[1], Chikara Dohno[3], Kazuhiko Nakatani[3] and Krzysztof Sobczak[1]

[1] Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland

[2] Institute of Human Biology and Evolution, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland

[3] Department of Regulatory Bioorganic Chemistry, The Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, 567-0047, Japan

[4] National Centre for Biological Sciences (NCBS), Tata Institute of Fundamental Research (TIFR), Bellary Road, Bangalore 560065, Karnataka, India

* equally contributed authors

Supplementary Table S1

Amounts of immobilized RNA (r) or DNA (d) $(CGG)_9$ on sensor chips used for SPR assays

Series S sensor chip SA surface	**Immobilized r/d$(CGG)_9$**	**Amounts of immobilized r/d$(CGG)_9$ (RU)**	**Compounds**
SA Chip-1	r$(CGG)_9$	327.4	NCD, Z-NCTS
SA Chip-2	r$(CGG)_9$	481.3	CMBL1a, 1b, 1c, 2a, 2b, 2c, 3a, 4b,
SA Chip-3	r$(CGG)_9$	538	CMBL3b, 3c, 4a, 4c
SA Chip-4	r$(CGG)_9$	601.9	CMBL5a
SA Chip-5	d$(CGG)_9$	477.4	CMBL4a

Supplementary Table S2

Amounts of immobilized DNA (d) or RNA (r) on sensor chips used for SPR assays

Series S sensor chip SA surface	Immobilized DNA/RNA	Amounts of immobilized DNA/RNA (RU)	Compound
SA Chip-6	$d(CGG)_9$	477.4	CMBL4c
SA Chip-6	$d(CCG)_9$	481.8	CMBL4c
SA Chip-6	$d(CAG)_9$	500.6	CMBL4c
SA Chip-7	$d(CTG)_9$	483.1	CMBL4c
SA Chip-7	$d(GAA)_9$	519.3	CMBL4c
SA Chip-7	$d(CCTG)_9$	487	CMBL4c
SA Chip-8	$d(G_4C_2)_6$	488.2	CMBL4c
SA-Chip-8	rSS_{30}	674	CMBL4c
SA-Chip-8	rDS_{30}	767	CMBL4c
SA Chip-9	$r(CCG)_9$	497	CMBL4c
SA Chip-9	$r(CAG)_9$	493	CMBL4c
SA Chip9	$r(CUG)_9$	503	CMBL4c
SA Chip-10	$r(GAA)_9$	504.8	CMBL4c
SA Chip-10	$r(CCUG)_9$	494.5	CMBL4c
SA Chip-10	$r(G_4C_2)_6$	487.8	CMBL4c

Supplementary Table S3

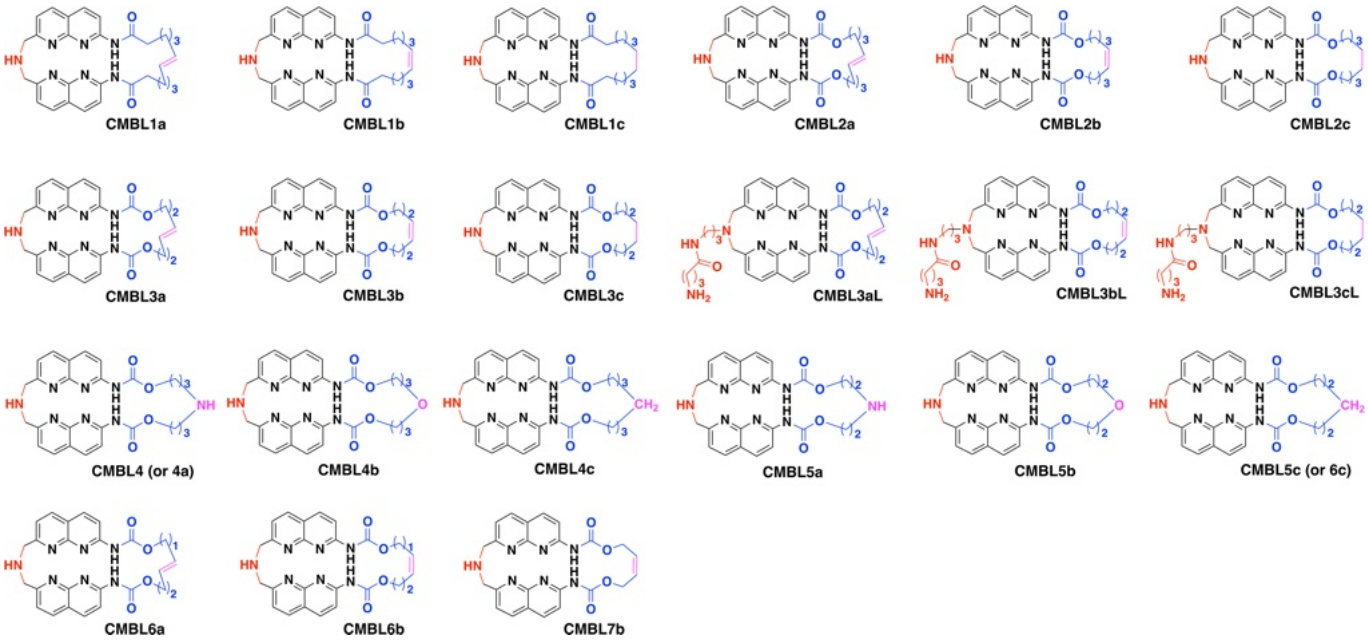
Apparent binding constants from SPR curves applying 1:1 fitting model

Ligand	**Immobilized DNA/RNA**	k_{on} **(× 10^3 $M^{-1}s^{-1}$)**	k_{off} **(× 10^{-3} s^{-1})**	$K_{d(app)}$ **(μM)**
NCD	$r(CGG)_9$	2.47	37.87	15.3
Z-NCTS	$r(CGG)_9$	6.00	1.71	0.285
CMBL4a	$r(CGG)_9$	11.60	9.71	0.837
CMBL4a	$d(CGG)_9$	12.5	17.17	1.37

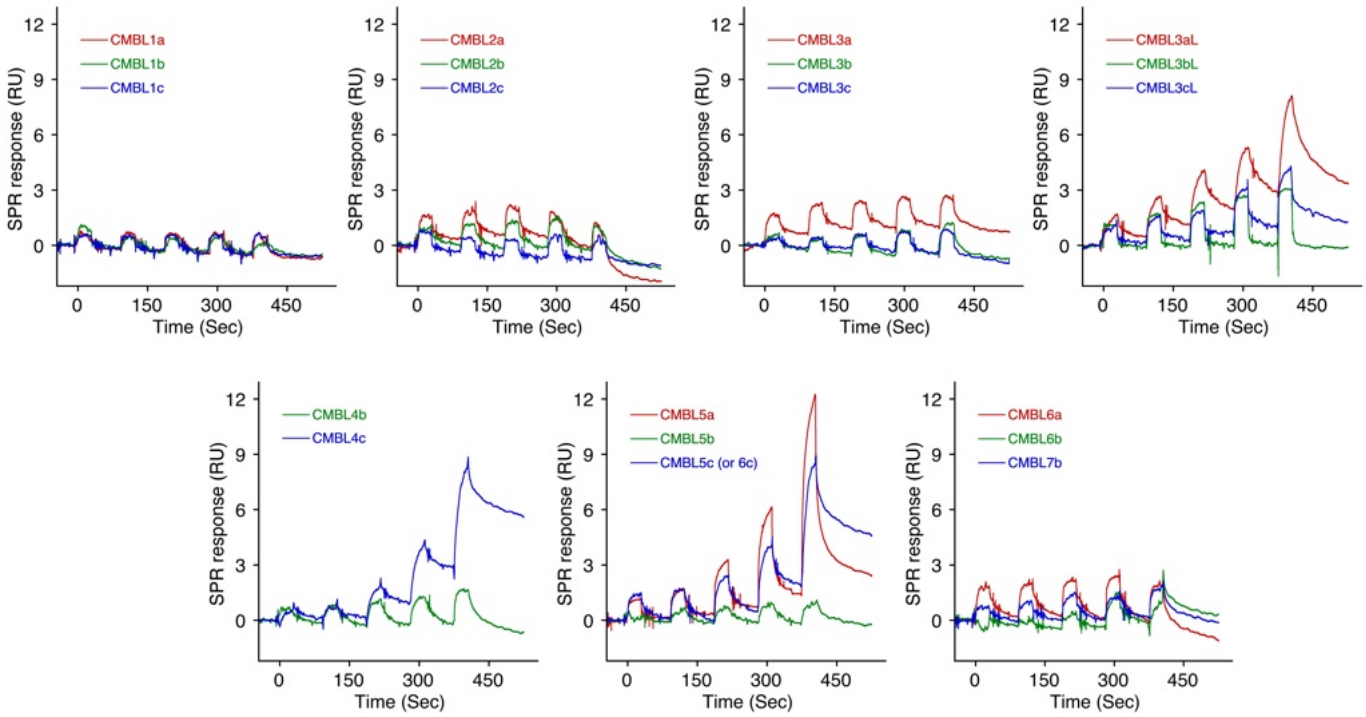
Supplementary Table S4

RT-PCR Primers

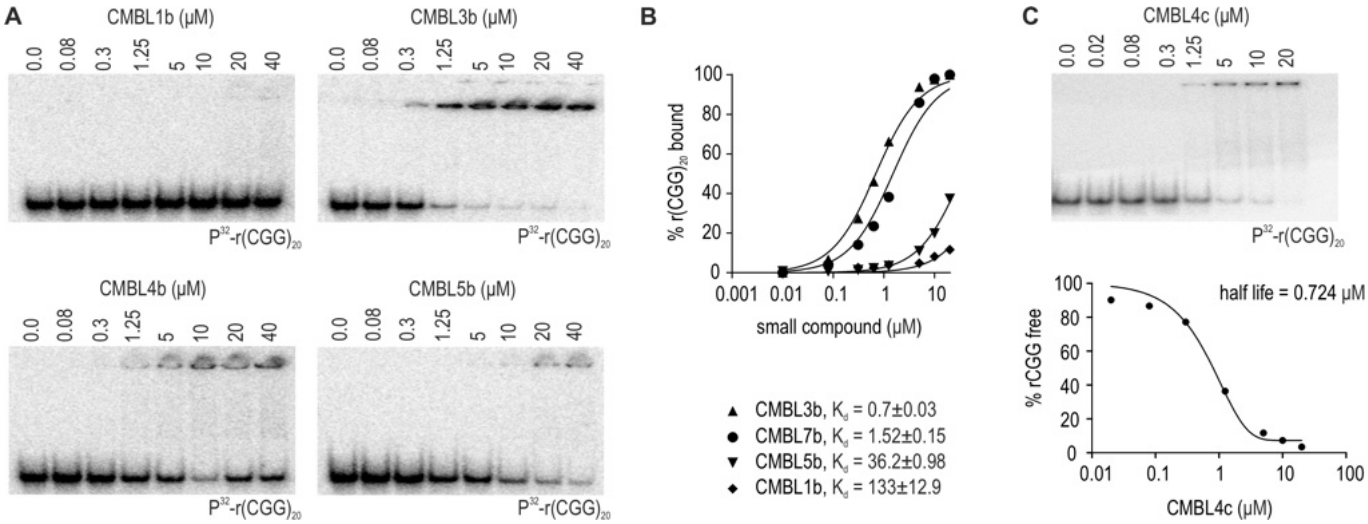
5′(CGG)$_{99}$	
F1	5’ Pho - CTAGGATAATAGTGACTAATCGTCC 3’
R1	5’ Pho - GGCCGGACGATTAGTCACTATTATC 3’
F2	5’ CTGGGCCTCGAGCGCCCGCAGCCCACCTC 3’
R2	5’ GAAGCGGCCGGACGATTAGTCACTATTATCATAGGTGCCTTGTAG 3’
AR	
5’ CTTGCCTGGCTTCCGC 3’	
5’ TGCGGTACTCATTGAAAACCA 3’	
CARM1	
5’ TGCCAACAAAAGACAGAGCT 3’	
5’ CGAGGGAGATGTGTAGTGGG 3’	
DMPK	
5’ GCGATCTCTGCCTGCTTACT 3’	
5’ GTCCTAGGTGGGGACAGACA 3’	
FMR1	
5’ ATCCCAACAAACCTGCCACA 3’	
5’ ATGTGCTCGCTTTGAGGTGA 3’	
GFP	
5’ CAGCCACAACGTCTATATCA 3’	
GAPDH	
5’ CATCAATGGAAATCCCATCAC 3’	
5’ GGTTTTTCTAGACGGCAGGTC 3’	
***GAPDH* (qPCR)**	
5’ GAGTCAACGGATTTGGTCGT 3’	
5’ TTGATTTTGGAGGGATCTCG 3’	
QKI	
5’ AGCTGATGGAGCTTGCGATT 3’	
5’ AAGGCAAGGGCTGGTGATTT 3’	
VKORC1L1	
5’ CCGGAATCCTGCTCTCCATC 3’	
5’ AATCCTCGACCCCATCTGGA 3’	



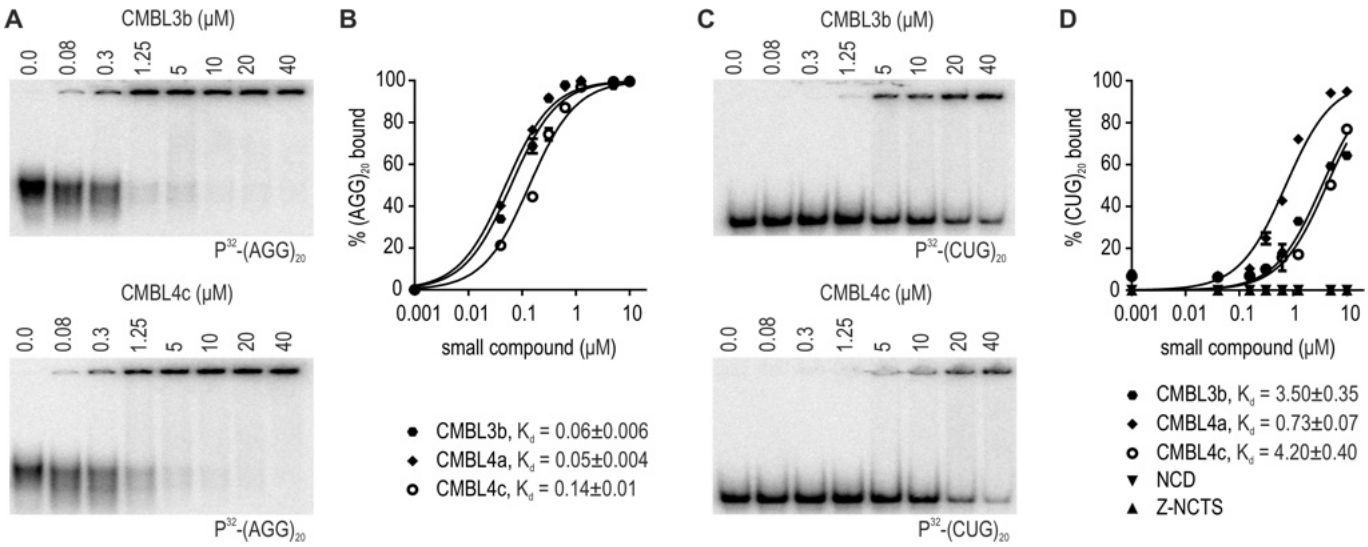
Supplementary Figure S1. Schematic chemical structures of CMBL compounds used in the study.



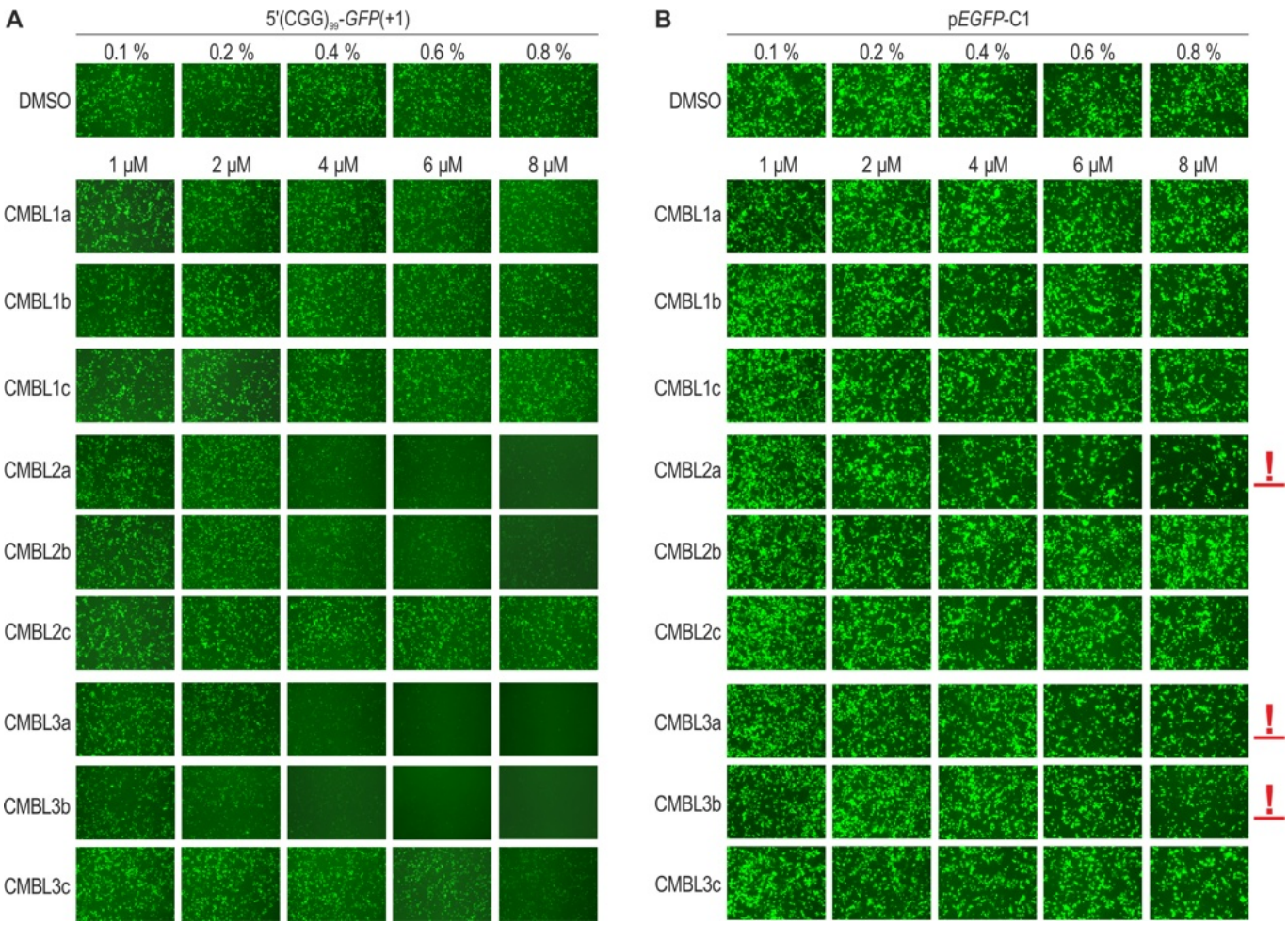
Supplementary Figure S2. SPR single cycle kinetic analysis of interaction of CMBL compounds with the r(CGG)$_9$ immobilized sensor surface. CMBLs were added sequentially to 0.25, 0.5, 1.0, 2.0 and 4.0 μM.



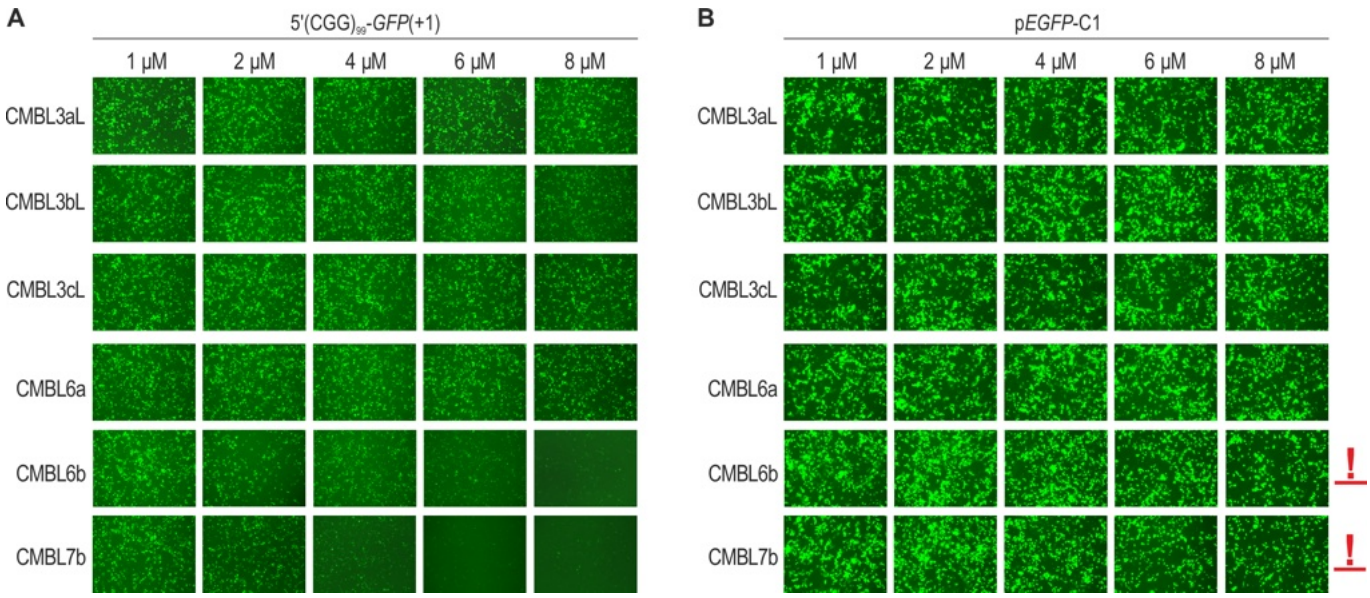
Supplementary Figure S3. The effect of CMBL compounds on the mobility of r(CGG)$_{20}$ repeats *in vitro*. **(A)** Electrophoretic mobility shift assay (EMSA) results showing binding of CMBL1b, 3b, 4b and 5b to r(CGG)$_{20}$ triplet repeats. (**B**) Binding curves and dissociation constant data (K_d) of CMBL1b, 3b, 5b, or 7b and r(CGG)$_{20}$ repeats quantified based on filter binding assays (FBAs). (**C**) EMSA data (upper panel) and graphical representation (lower panel) of CMBL4c interaction with r(CGG)$_{20}$ triplet repeats. Mean values from three experiments ± standard deviation (SD) are shown on the graph.



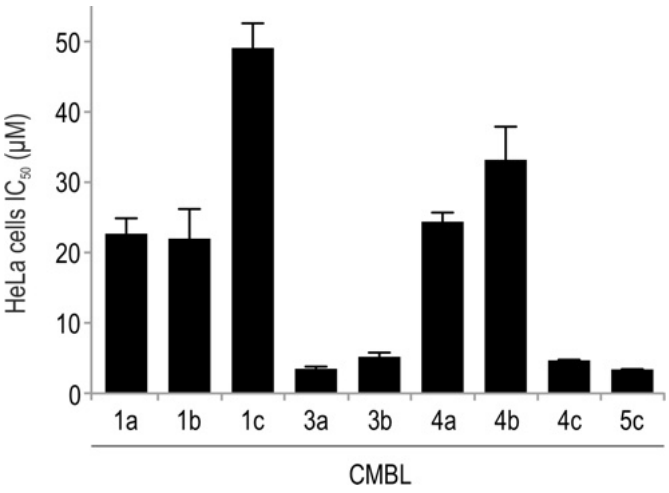
Supplementary Figure S4. Interaction of CMBL compounds with r(AGG)$_{20}$ and r(CUG)$_{20}$ repeats *in vitro*. (**A**) EMSA results showing binding of CMBL3b and 4c to r(AGG)$_{20}$ repeats. (**B**) Graphical representation and K_d quantification of CMBL3b, 4a and 4c interaction with r(AGG)$_{20}$ repeats *in vitro* based on FBAs. (**C**) EMSA results showing binding of CMBL3b and 4c to r(CUG)$_{20}$ repeats. (**D**) Graphical representation and K_d quantification of CMBL3b, 4a, 4c, NCD, and Z-NCTS interaction with r(CUG)$_{20}$ repeats *in vitro* based on FBAs. Mean values from two experiments ± standard deviation (SD) are shown on the graphs.



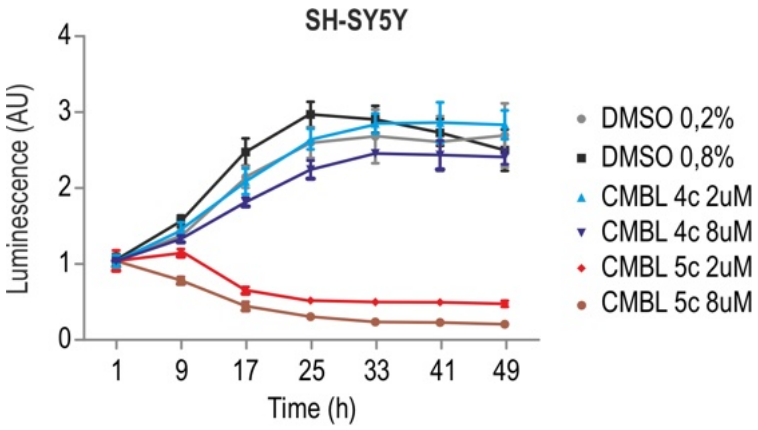
Supplementary Figure S5. (**A**, **B**) Fluorescence microscopy of COS7 cells transfected with either 5'(CGG)$_{99}$-*GFP*(+1) (**A**) or control p*EGFP*-C1 vector (**B**) and treated with DMSO or CMBL compounds 1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b and 3c. Red exclamation marks point to CMBL compounds that reduce GFP levels in p*EGFP*-C1-transfected cells.



Supplementary Figure S6. (**A**, **B**) Fluorescence microscopy of COS7 cells transfected with either 5'(CGG)$_{99}$-*GFP*(+1) (**A**) or a control p*EGFP*-C1 vector (**B**) and treated with CMBL compounds 3aL, 3bL, 3cL, 6a, 6b and 7b. Red exclamation marks point to CMBL compounds that reduce GFP levels in p*EGFP*-C1-transfected cells.

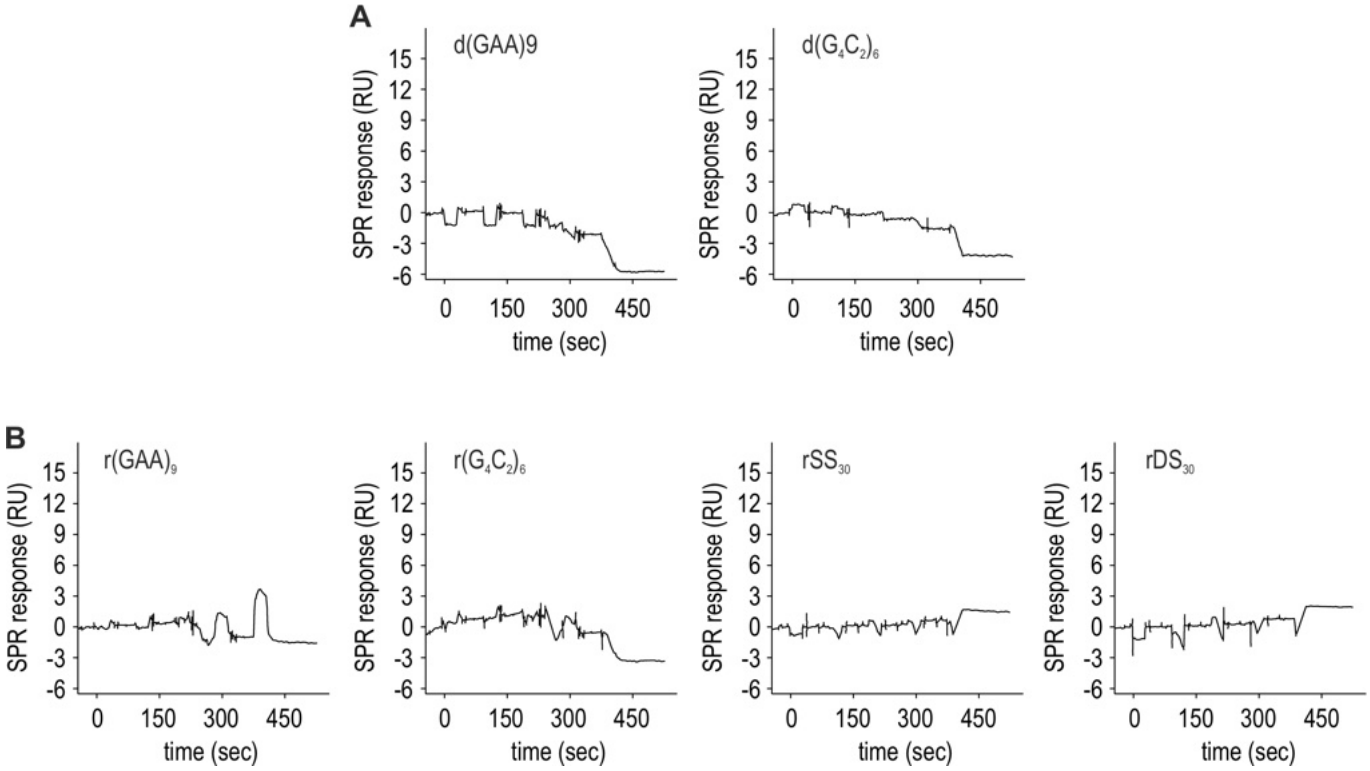


Supplementary Figure S7. The half maximal inhibitory concentration (IC_{50}; μM) of indicated CMBLs in inhibiting viability of HeLa cells.

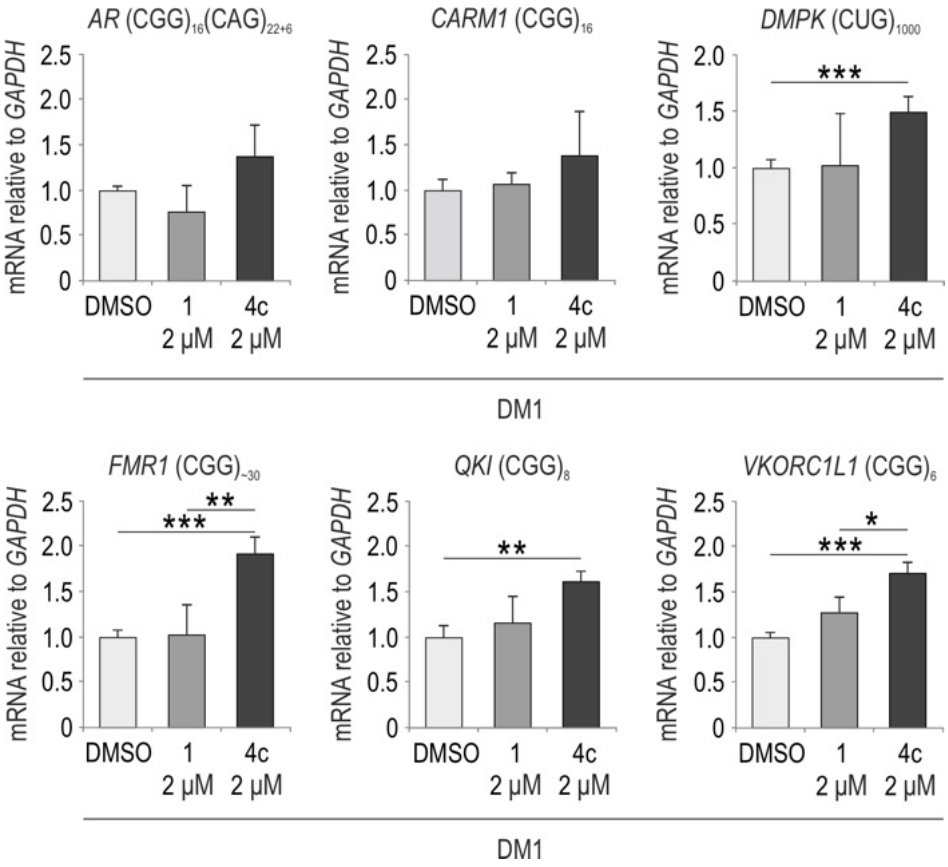


	CMBL 4c		CMBL 5c	
	2 µM	8 µM	2 µM	8 µM
9h	ns.	**	**	***
17h	ns.	***	***	***
25h	ns.	***	***	***
33h	ns.	**	***	***
41h	ns.	ns.	***	***
49h	ns.	ns.	***	***

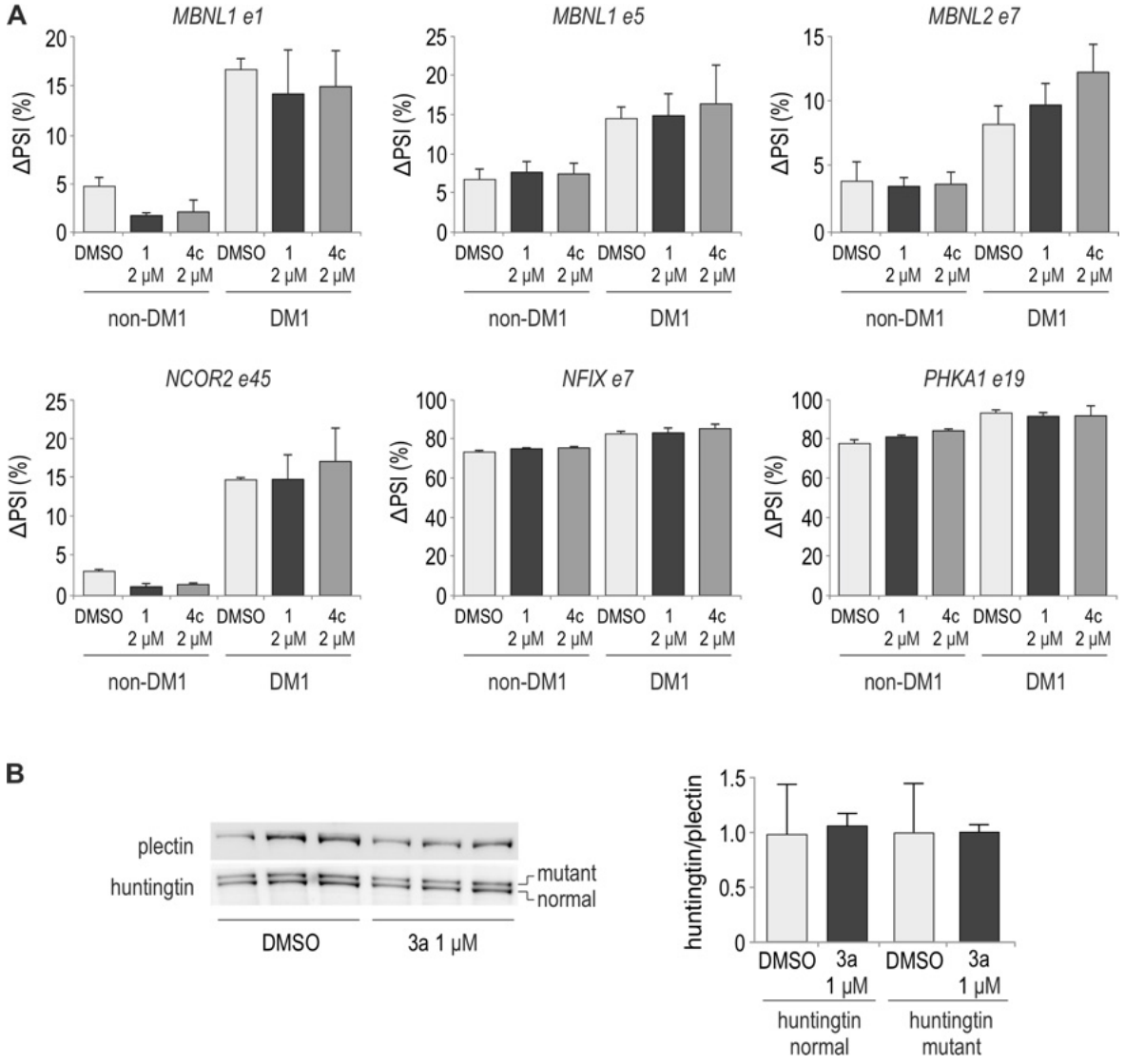
Supplementary Figure S8. Comparison of CMBL4c and 5c toxicity in SH-SY5Y cells. Luminescence was measured 1, 9, 17, 25, 33, 41 and 49 h following delivery of DMSO at 0.2% or 0.8% concentration and CMBL4c or 5c at 2 µM or 8 µM concentration (0.2% or 0.8% DMSO concentration, respectively). Statistical significance was determined by unpaired two tailed Student's t-test. * indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$.



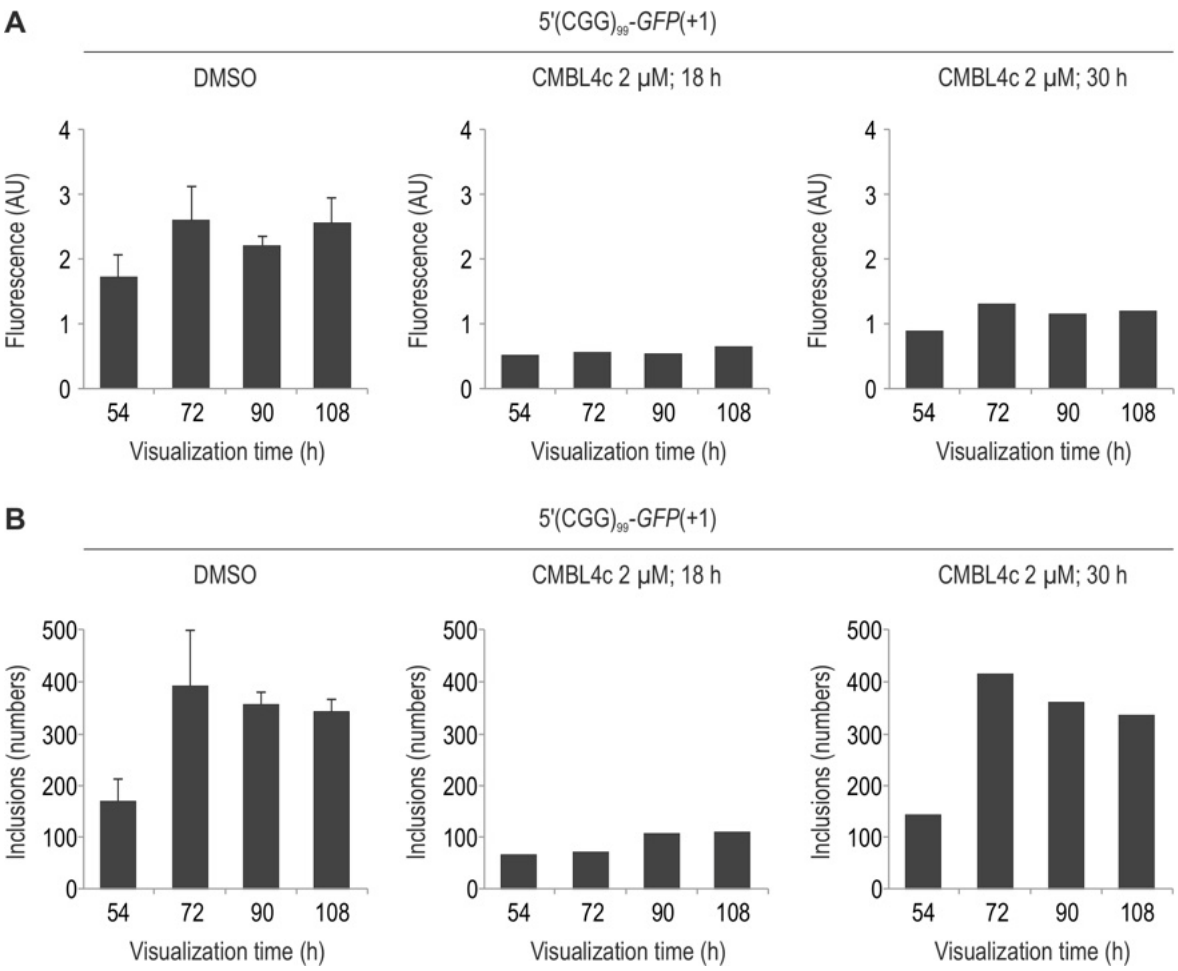
Supplementary Figure S9. CMBL4c interaction with DNA and RNA sequences. (**A**, **B**) SPR single cycle kinetic analysis of CMBL4c interaction with DNA (**A**) and RNA sequences (**B**) as indicated in the figure. CMBL4c was added sequentially to 0.25, 0.5, 1.0, 2.0 and 4.0 μM.



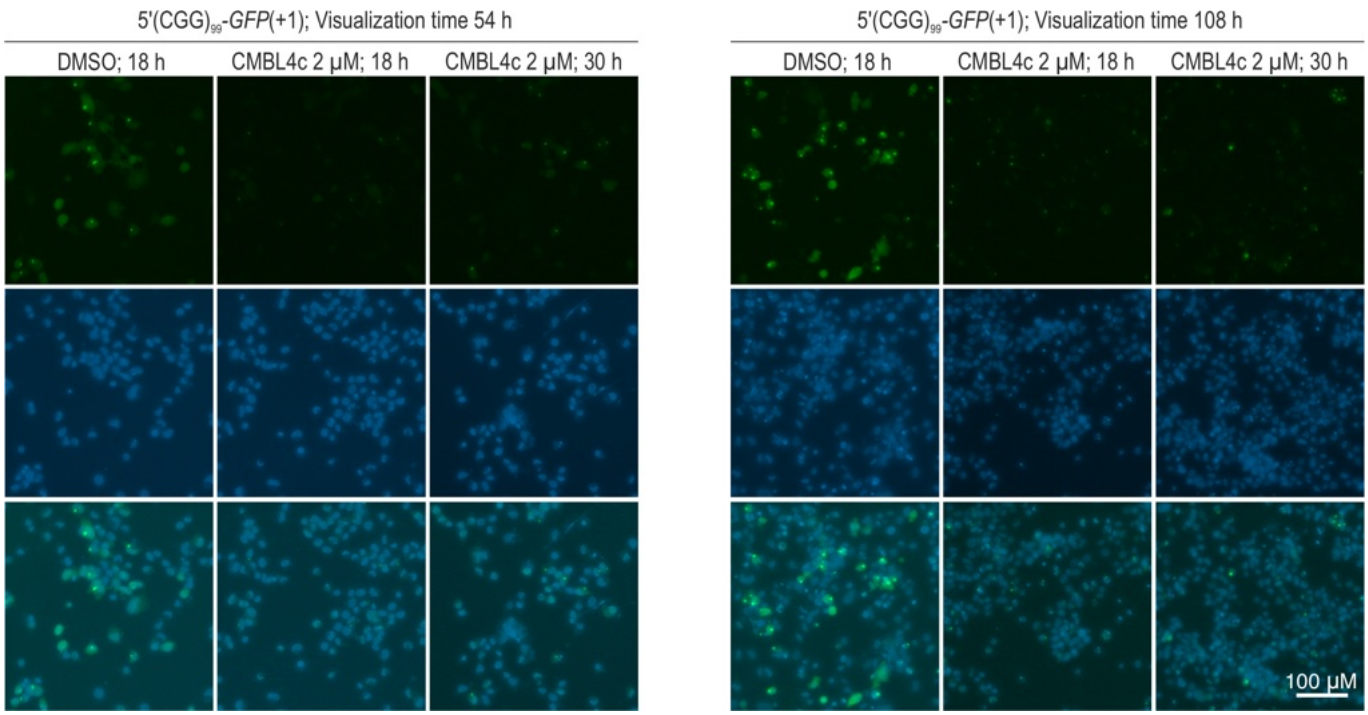
Supplementary Figure S10. RT-qPCR analyses showing *AR*, *CARM1*, *DMPK*, *FMR1*, *QKI* and *VKORC1L1* transcript levels in DM1 fibroblasts (GM04033) 48 h following delivery of DMSO, CMBL1(b and c) or 4c. Mean values with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test (* indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$).



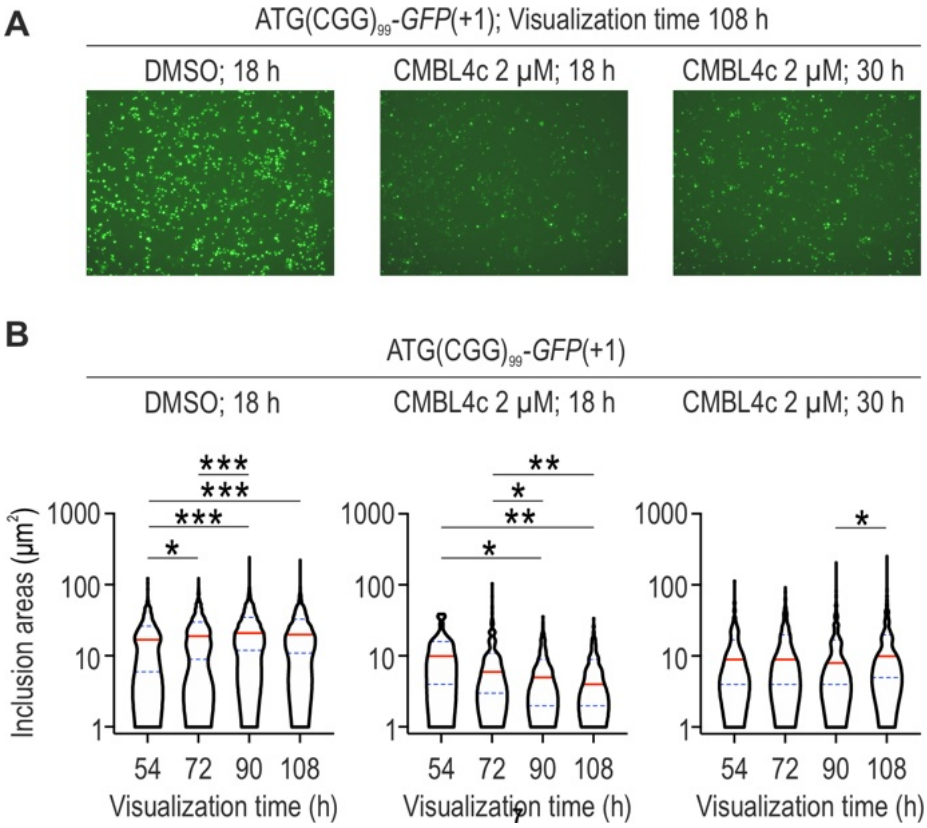
Supplementary Figure S11. CMBLs act therapeutically neither in DM1 nor HD. (**A**) CMBL4c does not affect splicing of MBNL-dependent alternative exons. Splicing analyses of *MBNL1* e1, *MBNL1* e5, *MBNL2* e7, *NCOR2* e45, *NFIX* e7 and *PHKA* e19 were performed in control (non-DM1; GM07492) and DM1 fibroblasts (GM04033) 48 h following delivery of DMSO (0.2%), CMBL1(b and c) or 4c (2 μM). (**B**) CMBL3a neither reduces mutant nor normal huntingtin levels. Western blot analysis of huntingtin levels in HD fibroblasts (17/68 Q) was performed 48 h after incubation with either DMSO (0.1%) or 1 μM CMBL3a. Signal intensities were normalized to plectin protein levels. Error bars represent standard deviations.



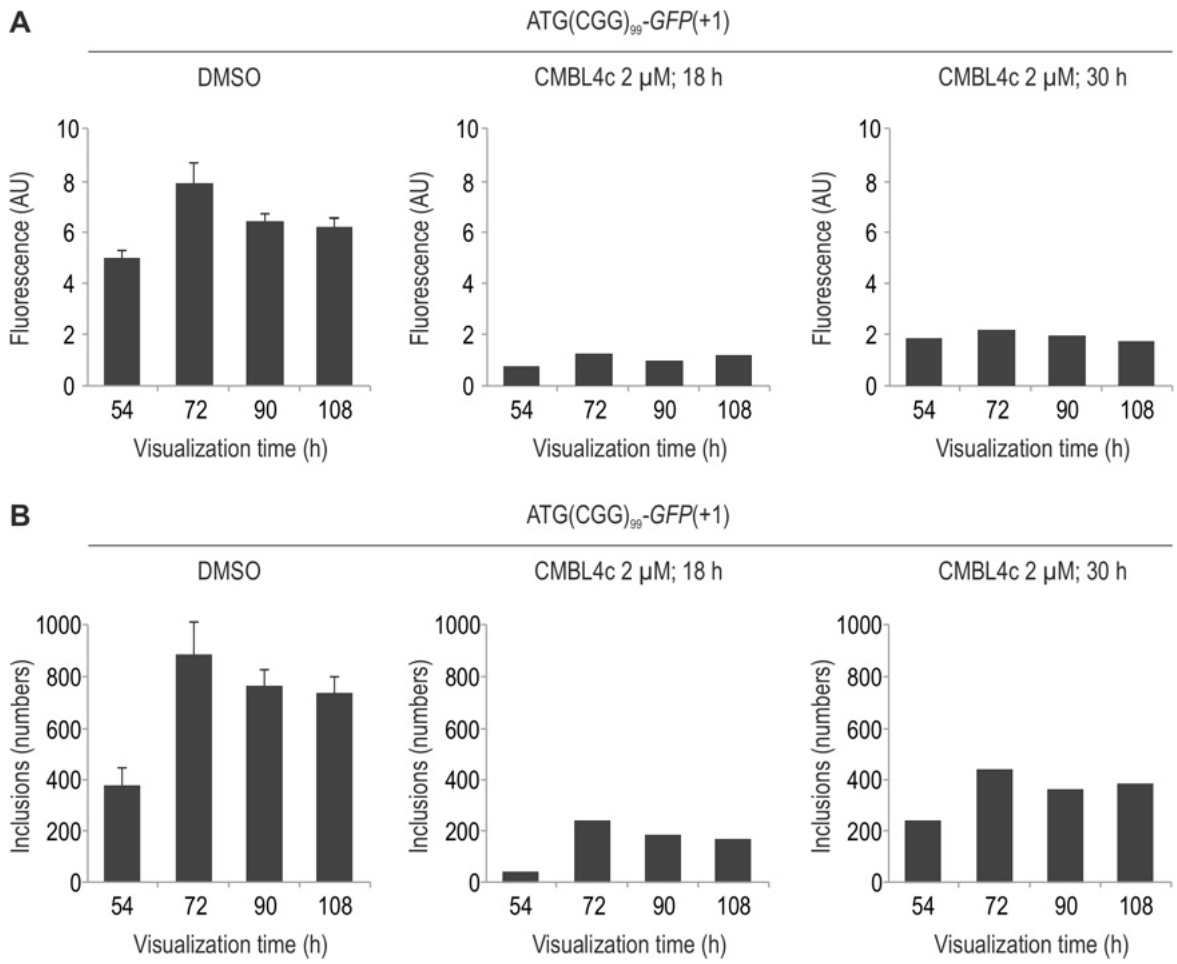
Supplementary Figure S12. The effect of CMBL4c on inclusion growth is delivery-time dependent. (**A**, **B**) Fluorescence signals (**A**) and inclusion numbers (**B**) were estimated 54, 72, 90 or 108 h after transfection with 5'(CGG)$_{99}$-*GFP*(+1). DMSO or CMBL4c were added either 18 or 30 h following the vector delivery. The analyses shown in Figure 7B and Supplementary Figure S12 were performed based on the same images.



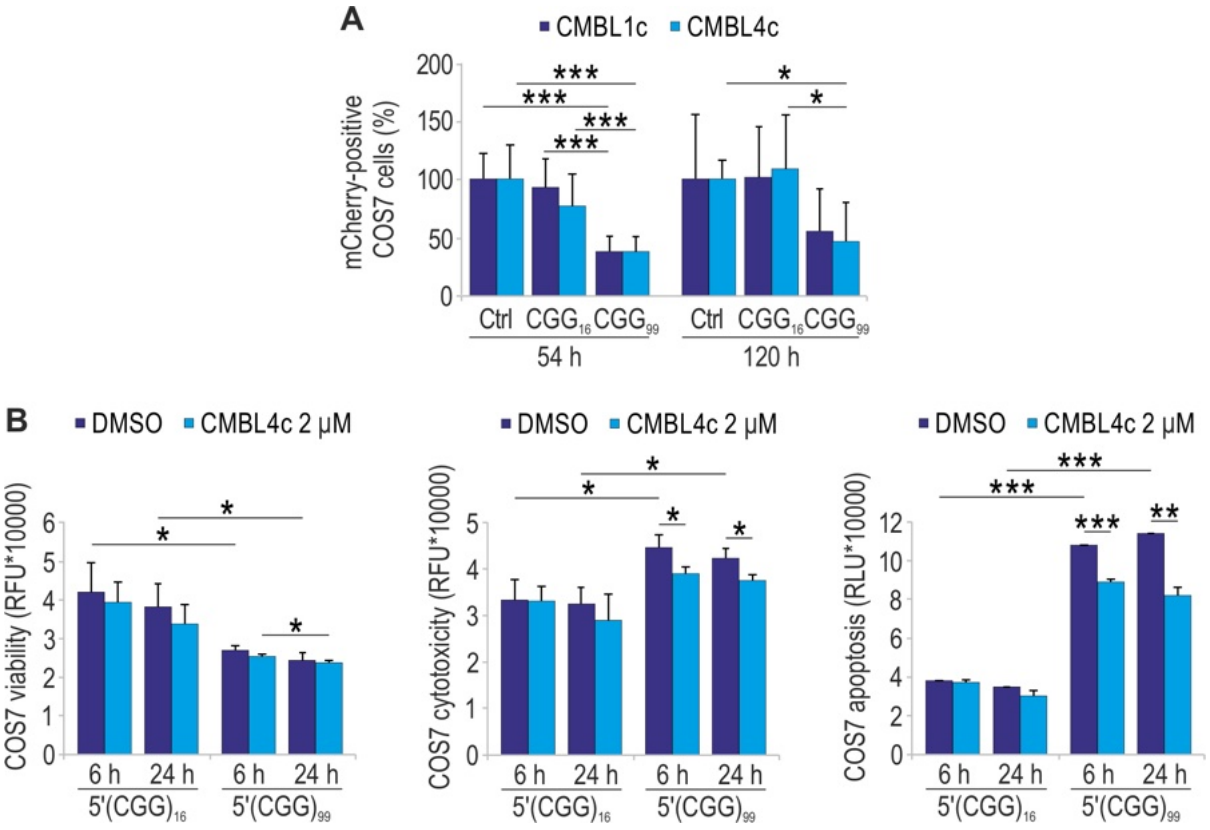
Supplementary Figure S13. The FMRpolyG inclusion size depends on the time of CMBL4c delivery. Microscopy analysis of FMRpolyG-GFP and nuclei (Hoechst staining) of COS7 cells was performed 54 or 108 h after transfection with 5'(CGG)$_{99}$-*GFP*(+1). DMSO or CMBL4c were added either 18 or 30 h following the vector delivery as indicated in the figure.



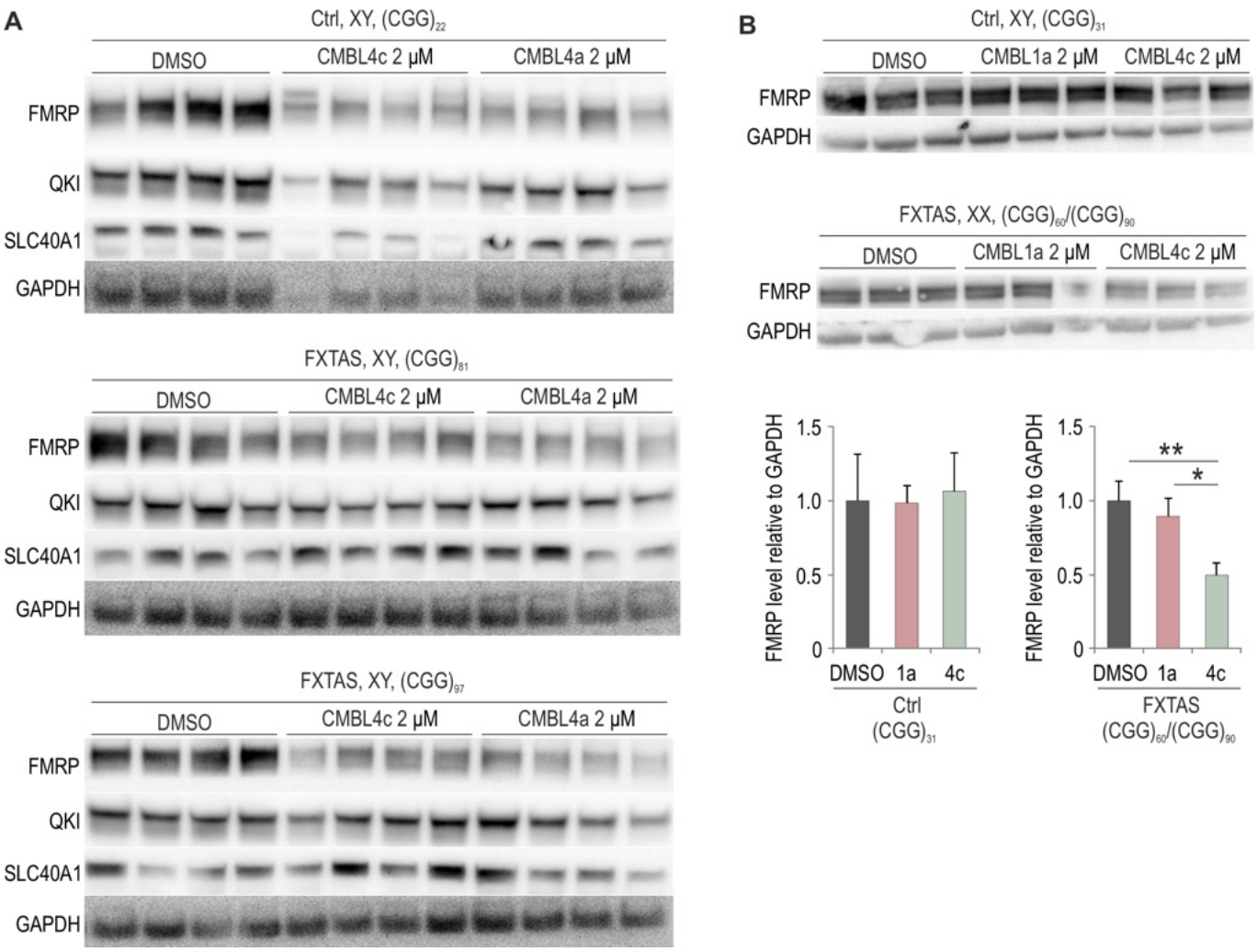
Supplementary Figure S14. The effect of CMBL4c on inclusion growth is delivery-time dependent. (**A**, **B**) Visualization (**A**) and distribution area plots (**B**) of FMRpolyG-GFP inclusions after administration of ATG(CGG)$_{99}$-*GFP*(+1) and delivery of either DMSO or CMBL4c 18 or 30 h post transfection as indicated in the figure. Statistical significance was determined by Kruskal-Wallis test. * indicates $p < 0.05$, ** indicates $p<0.01$ and *** indicates $p < 0.001$.



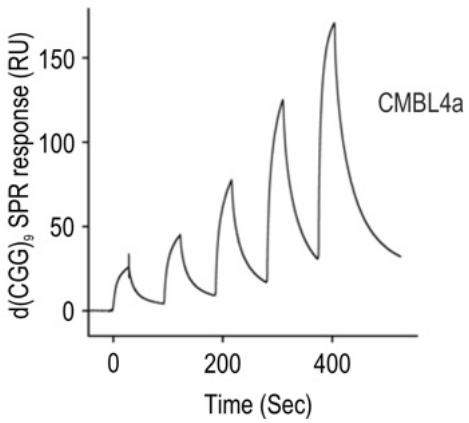
Supplementary Figure S15. The effect of CMBL4c on inclusion growth is delivery-time dependent. (**A**, **B**) Fluorescence signals (**A**) and inclusion numbers (**B**) were estimated 54, 72, 90 or 108 h after transfection with ATG(CGG)$_{99}$-*GFP*(+1). DMSO or CMBL4c were added either 18 or 30 h following the vector delivery as indicated in the figure. The analyses shown in Supplementary Figure S14B and S15 were performed based on the same images.



Supplementary Figure S16. CMBL4c reduces cytotoxicity and apoptosis in an FXTAS COS7 cell model. (**A**) Quantification of percentages of mCherry-positive cells 54 and 120 h after administration of *mCherry*-NLS (Ctrl) or co-delivery of *mCherry*-NLS and 5'(CGG)$_{16}$ or *mCherry*-NLS and 5'(CGG)$_{99}$. CMBL1c or CMBL4c were added 24 h post transfection. (**B**) Graphs showing viability, cytotoxicity and apoptosis of COS7 cells following delivery of 5'(CGG)$_{16}$ or 5'(CGG)$_{99}$ and addition of DMSO or CMBL4c, 6 or 24 h post transfection. Mean values with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test. * indicates $p < 0.05$, ** indicates $p < 0.01$ and *** indicates $p < 0.001$.



Supplementary Figure S17. CMBL4c attenuates translation of FMRP. (**A**) Photographs of membranes immunoblotted for FMRP, GAPDH, QKI and SLC40A1 used for quantitative analysis shown in Figure 8C. (**B**) Immunoblotting of control and FXTAS fibroblast lysates with FMRP and GAPDH antibodies. Fibroblasts were treated with DMSO, CMBL1a or 4c (2 μM) for 48 h. Mean values normalized to GAPDH levels with standard deviation (SD) are shown on the graphs. Statistical significance was determined by unpaired two-tailed Student's *t*-test. * indicates $p < 0.05$ and ** indicates $p < 0.01$.

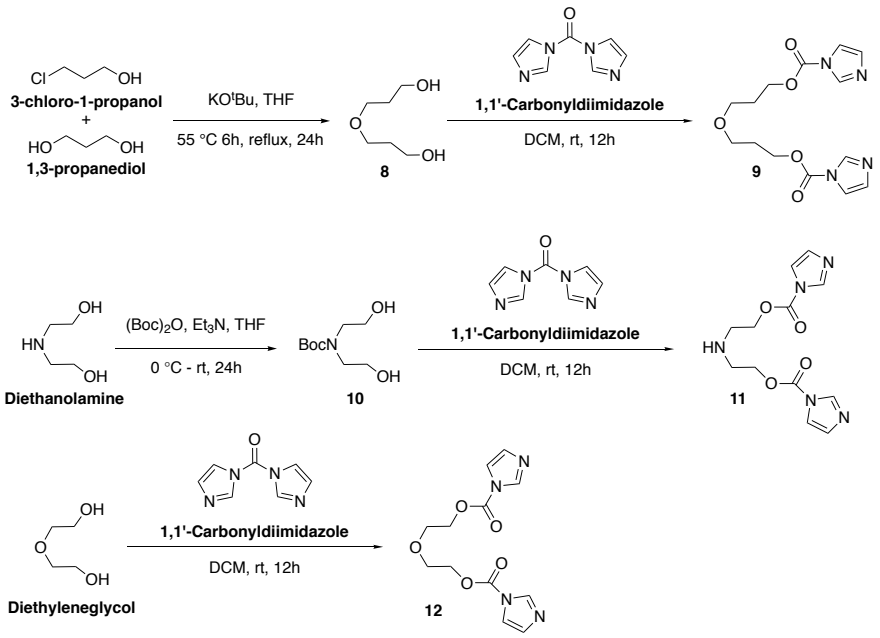


Supplementary Figure S18. SPR single cycle kinetic analysis of CMBL4a interaction with the $d(CGG)_9$ immobilized sensor surface.

Supplementary Data

Synthesis of CMBL 3aL, 3bL, 3cL, 4b, 5a and 5b

Synthesis of compounds 9, 11 and 12:



Scheme S1. Synthesis of linkers for amide coupling.

3,3'-Oxybis(propan-1-ol) (8): Compound **8** was prepared using previously reported synthetic procedure[1]. ^{1}H NMR (600 MHz, $CDCl_3$): δ 3.77 (t, J = 5.4 Hz, 4H), 3.63 (t, J = 6.0 Hz, 4H), 1.84 (quint, J = 5.85 Hz, 4H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 69.9, 61.6, 32.1; HRMS: calcd. for $C_6H_{15}O_3$ $[M+H]^+$ 135.1021, found 135.1017; calcd. for $C_6H_{14}NaO_3$ $[M+Na]^+$ 157.0841, found 157.0836.

Oxybis(propane-3,1-diyl) bis(1*H*-imidazole-1-carboxylate) (9): A mixture of compound **8** (154 mg, 1.15 mmol) and 1, 1'-carbodiimidazole (465.3 mg, 2.87 mmol) in dry DCM (10 ml) was stirred at room temperature for overnight under argon atmosphere. Reaction mixture was diluted with DCM and washed with large excess of water. Organic part was dried over sodium sulphate and evaporated to get pure product. White low melting solid (260 mg, 70%); ^{1}H NMR (600 MHz, $CDCl_3$): δ 8.12 (brs, 2H), 7.42-7.41 (m, 2H), 7.06 (brs, 2H), 4.51 (t, J = 6.3 Hz, 4H), 3.55 (t, J = 6.3 Hz, 4H), 2.03 (quint, J = 5.85 Hz, 4H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 148.6, 137.0, 130.6, 117.0, 66.9, 65.5, 28.8; HRMS: calcd. for $C_{14}H_{19}N_4O_5$ $[M+H]^+$ 323.1355, found 323.1350; calcd. for $C_{14}H_{18}N_4NaO_5$ $[M+Na]^+$ 345.1175, found 345.1169.

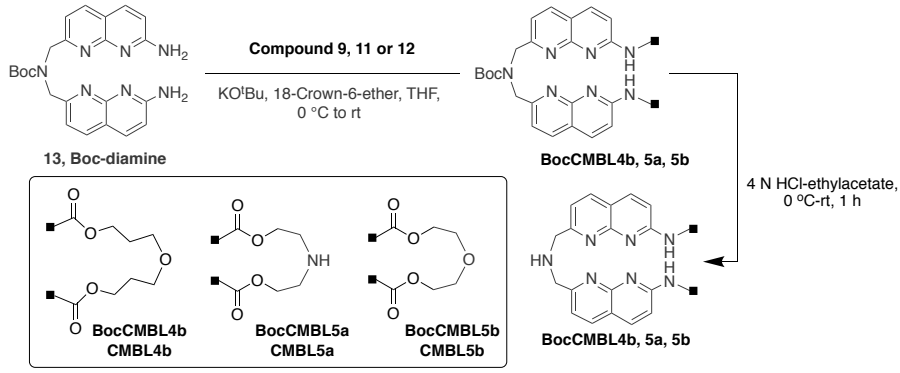
***N*-Boc-diethanolamine (10):** Boc-anhydride (655 mL, 2.85 mM) was added to a stirred ice-cold solution of diethanolamine (200 mg, 1.90 mM) in THF (5 mL). Reaction mixture was stirred at room temperature for 24h. All the solvent was evaporated off. Crude residue was purified by column chromatography using 25-60% ethyl acetate-hexane as running solvent. Sticky colorless liquid (300 mg, 76.8%); ^{1}H

NMR (400 MHz, $CDCl_3$): δ 3.80 (br, 4H), 3.44 (br, 4H), 1.46 (s, 9H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 156.3, 80.2, 61.6, 52.2, 52.1, 28.3; HRMS: calcd. for $C_9H_{19}NNaO_4$ $[M+Na]^+$ 228.1212, found 228.1207; calcd. for $C_{18}H_{38}N_2NaO_8$ $[2M+Na]^+$ 433.2526, found 433.2521.

((*tert*-butoxycarbonyl)azanediyl)bis(ethane-2,1-diyl) bis(1*H*-imidazole-1-carboxylate) (11): A mixture of compound **10** (300 mg, 1.46 mmol) and 1, 1'-carbodiimidazole (592.5 mg, 3.65 mmol) in dry DCM (15 ml) was stirred at room temperature for overnight. Reaction mixture was diluted with DCM and washed with large excess of water under argon atmosphere. Organic part was dried over sodium sulphate and evaporated to get pure product. White low melting solid (565 mg, 98%); ^1H NMR (400 MHz, $CDCl_3$): δ 8.12 (brs, 2H), 7.42-7.40 (m, 2H), 7.09-7.06 (m, 2H), 4.56-4.50 (m, 4H), 4.23-4.19 (m, 4H), 1.37 (s, 9H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 154.8, 148.2, 148.1, 136.8, 136.7, 130.4, 130.1, 116.9, 116.9, 80.8, 65.4, 65.3, 46.3, 45.9, 27.8. HRMS: calcd. for $C_{17}H_{23}N_5NaO_6$ $[M+Na]^+$ 416.1546, found 416.1543.

Oxybis(ethane-2,1-diyl) bis(1*H*-imidazole-1-carboxylate) (12): A mixture of diethylene-glycol (300 mg, 2.83 mmol) and 1, 1'-carbodiimidazole (1146 mg, 7.07 mmol) in dry DCM (28 ml) was stirred at room temperature for overnight under argon atmosphere. Reaction mixture was diluted with DCM and washed with large excess of water. Organic part dried over sodium sulfate and evaporated to get pure product. White low melting solid (800 mg, 96%); ^1H NMR (400 MHz, $CDCl_3$): δ 8.14-8.13 (m, 2H), 7.40-7.39 (m, 2H), 7.07-7.06 (m, 2H), 4.59-4.56 (m, 4H), 3.87-3.85 (m, 4H); ^{13}C NMR (100 MHz, $CDCl_3$): δ 148.3, 136.8, 130.4, 116.8, 68.3, 66.4; HRMS: $C_{12}H_{15}N_4O_5$ $[M+H]^+$ 295.1042, found 295.1038; calcd for $C_{12}H_{14}N_4NaO_5$ $[M+Na]^+$ 317.0856, found 317.0858.

Synthesis of CMBL4b, CMBL5a and CMBL5b:

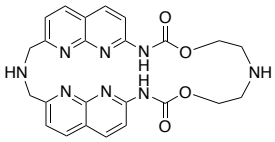


Scheme S2. Synthesis of **CMBL4b**, **CMBL5a** and **CMBL5b**

8,12,16-trioxa-3,6,18-triaza-1,5(2,7-bis(1,8-naphthyridina))cyclooctadecaphane-7,17-dione (CMBL4b): To a stirred solution of **13** (100 mg, 0.232 mmol)[2] in dry THF (1.5 ml) was added KO^{t}Bu (78.0 mg, 0.695 mmol) and 18-crown-6-ether (catalytic) at 0 ºC under argon atmosphere. Reaction mixture was stirred at ice cold condition for 15 min, followed by

drop wise addition of a solution of compound **9** (74.7 mg, 0.232 mmol) in dry THF (2.0 ml). Reaction mixture was then stirred at 0 ºC to room temperature during 1h (monitored by TLC). After completion, reaction mixture was quenched with crushed ice and all the THF was evaporated off. Resultant residue was diluted with water (20 mL) and extracted with chloroform (20 mL x 3). Combined organic part was washed with brine (20 mL) and dried over sodium sulfate and evaporated to get crude. Crude material was purified by silica-gel column chromatography. Desired product was eluted using 1.5-2.0% methanol-chloroform and unreacted starting (40 mg) was recovered using 5-6% methanol-chloroform contain 0.1% NH_3. Desired product was further purified by PLC using 4% methanol/chloroform as mobile phase to afford pure **BocCMBL4b** (40 mg, 28%)**.** To a stirred solution of pure **BocCMBL4b** (7.7 mg, 0.012 mmol) in chloroform (1 mL) was added 4N HCl-ethylacetate (1 mL) at 0 °C under argon atmosphere. Reaction mixture was then stirred at 0 °C to room temperature for 1h. All the solvent was evaporated to dryness. Crude product was basified by 7 N ammonia in methanol solution, followed by silica gel column chromatography (2-5% methanol-chloroform). Product was further purified by HPLC using acetonitrile-water (0.1% AcOH) as mobile phase to obtained pure **CMBL4b** as acetate salt. Off-white solid (12.5 mg, 87.1%); ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.64 (brs, 1H), 8.11 (d, *J* = 8.8 Hz, 2H), 7.85 (d, *J* = 9.2 Hz, 2H), 7.68 (d, *J* = 8.0 Hz, 2H), 7.06 (d, *J* = 8.0 Hz, 2H), 5.49 (brs, 1H), 4.37 (t, *J* = 5.6 Hz, 4H), 4.19 (s, 4H), 3.64 (t, *J* = 6.0 Hz, 4H), 3.49 (s, 1H), 2.02 (s, 3H; acetate CH_3), 1.96 (quint, *J* = 5.8 Hz, 4H). %); **HRMS (ESI)** calcd. for $C_{26}H_{28}N_7O_5[M+H]^+$ 518.2152, found 518.2145; calcd. for $C_{26}H_{27}N_7NaO_5[M+Na]^+$ 540.1971, found 540.1965.

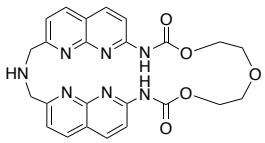
8,14-dioxa-3,6,11,16-tetraaza-1,5(2,7-bis(1,8-naphthyridina))cyclohexadecaphane-7,15-dione (CMBL5a):



To a stirred solution of **13** (50 mg, 0.116 mmol)[2] in dry THF (1 ml) was added KO^{t}Bu (39 mg, 0.348 mmol) and 18-crown-6-ether (catalytic) at 0 °C under argon atmosphere. Reaction mixture was then stirred at 0 ºC for 15 min, followed by drop wise addition of a solution of compound **11** 45.6 mg, 0.116 mmol) in dry THF (1.5 ml). Reaction mixture was then stirred at 0 ºC to room temperature during 1h (monitored by TLC). After completion, reaction mixture was quenched with crushed ice and all the THF was evaporated off. Resultant residue was diluted with water (20 mL) and extracted with chloroform (20 mL x 3). Combined organic part was washed with brine (20 mL) and dried over sodium sulfate and evaporated to get crude. Crude material was purified by silica-gel column chromatography using 1.5-2.0% methanol-chloroform followed by further purification by PLC using 4% methanol-chloroform as mobile phase to afford pure **BocCMBL5a**. White solid (29.5 mg, 37%); ^{1}H NMR (400 MHz, CD_3OD): δ 7.82 (br, 3H), 7.69 (br, 3H), 7.08-7.01 (m, 2H), 4.89 (brs, 4H), 4.40 (t, *J* = 5.4 Hz, 4H), 3.68 (br, 4H), 1.61 (s, 9H), 1.53 (s, 9H); ^{13}C NMR (175 MHz, CD_3OD): δ 160.2, 160.0, 155.1, 154.9, 154.5, 152.7, 152.3, 151.9, 136.9, 135.9, 118.3, 117.9, 116.9, 112.3, 80.3, 79.6, 78.7, 59.1, 58.8, 49.8, 49.6, 26.2. HRMS (ESI): calcd. for $C_{34}H_{40}N_8NaO_8$ $[M+Na]^+$ 711.2867, found 711.2866. To a stirred solution of **BocCMBL5a** (20 mg, 0.029 mmol) in chloroform (1-2 mL) was added 4N HCl-ethylacetate (1-2 mL) at 0 °C under argon atmosphere. Reaction mixture was then stirred at 0 ºC to room temperature for 1 h. Reaction was monitored by TLC. All the solvent was evaporated to dryness. Crude product was dissolved in methanol followed by basification using 6N ammonia in methanol solution at 0 ºC and purified by column chromatography using 2-6% methanol-chloroform as running solvent to afford pure **CMBL5a**. White solid (12mg, 85.7%); ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.11 (d, *J* = 8.8 Hz, 2H), 7.91 (d, *J* = 8.4 Hz, 2H), 7.77 (d, *J* = 8.4 Hz, 2H), 7.10 (d, *J* = 8.0 Hz, 2H), 4.36 (t, *J* = 4.8 Hz, 4H), 4.17 (s, 4H), 2.96 (t, *J* = 4.8 Hz, 4H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 164.1, 154.1, 153.4, 152.9, 138.4, 135.8, 119.8, 118.2, 112.7, 63.8, 56.2, 46.9;

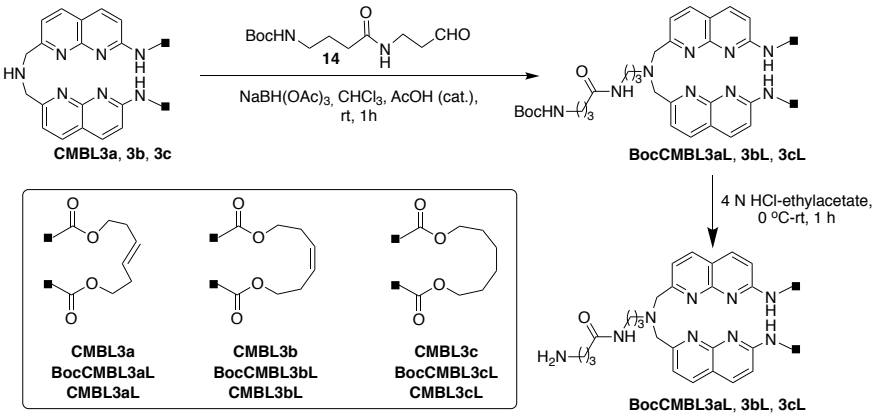
HRMS(ESI): calcd. for $C_{24}H_{25}N_8O_4$ $[M+H]^+$ 489.1999, found 489.1992; calcd. for $C_{24}H_{24}N_8NaO_4$ $[M+Na]^+$ 511.1818, found 511.1811.

8,11,14-trioxa-3,6,16-triaza-1,5(2,7-bis(1,8-naphthyridina))cyclohexadecaphane-7,15-dione (CMBL5b): To

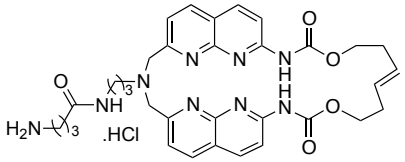


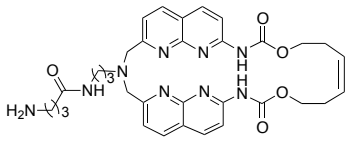
a stirred solution of **13** (60 mg, 0.139 mmol)[2] in dry THF (2 ml) was added KO^{t}Bu (46.8 mg, 0.417 mmol) and 18-crown-6-ether (catalytic) at 0 ºC under argon atmosphere. Reaction mixture was then stirred at 0 °C for 15 min, followed by drop wise addition of a solution of compound **12** (40.9 mg, 0.139 mmol) in dry THF (2.0 ml). Reaction mixture was then stirred at 0 ºC to room temperature during 1 h (monitored by TLC). After completion, reaction mixture was quenched with crushed ice and all the THF was evaporated off. Resultant residue was diluted with water (20 mL) and extracted with chloroform (20 mL x 3). Combined organic part was washed with brine (20 mL) and dried over sodium sulfate and evaporated to get crude. Crude material was purified by silica-gel column chromatography. Desired product was eluted at 1.5-2.0% methanol-chloroform and unreacted starting (30 mg) was eluted using 5-6% methanol-chloroform contain 0.1% NH_3)]. Desire product was further purified by PLC using 4% methanol-chloroform as mobile phase to afford pure **BocCMBL5b**. Off-white solid (32 mg, 39%). ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.30 (br, 2H), 8.04 (d, *J* = 9.2 Hz, 1H), 8.02 (d, *J* = 8.8 Hz, 1H), 7.66 (d, *J* = 10.8 Hz, 1H), 7.64 (d, *J* = 9.6 Hz, 1H), 7.48 (d, *J* = 8.4 Hz, 1H), 7.40 (d, *J* = 8.0 Hz, 1H), 7.08 (d, *J* = 8.0 Hz, 1H), 6.87 (d, *J* = 7.6 Hz, 1H), 4.94 (brs, 4H), 4.41-4.37 (m, 4H), 3.75 (t, J = 4.2 Hz, 4H), 1.62 (s, 9H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 161.8, 161.8, 155.3, 155.3, 153.2, 153.1, 137.7, 136.0, 135.8, 119.7, 118.4, 117.9, 117.7, 112.9, 112.9, 81.0, 68.65, 68.59, 63.6, 63.5, 55.5, 54.8, 28.5; HRMS: calcd. for $C_{29}H_{31}N_7NaO_7$[M+Na]$^+$ 612.2183, found 612.2173. To a stirred solution of **BocCMBL5b** (12.5 mg, 89.3%) in chloroform (1-2 mL) was added 4N HCl-ethylacetate (1-2 mL) at 0 ºC under argon atmosphere. Reaction mixture was then stirred at 0 ºC to room temperature for 1 h. Reaction was monitored by TLC. After completion all the solvent was evaporated to dryness. Crude product was dissolved in methanol followed by basification using 6N ammonia in methanol solution at 0 ºC. Crude product was purified by column chromatography (2-5% methanol-chloroform contain 0,.1% of ammonia) to afford pure **CMBL5b.** Off-white solid **(**12.5 mg, yield 89.3%); ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.46 (br, 1H), 8.06 (d, *J* = 8.8 Hz, 2H), 7.76 (d, *J* = 8.8 Hz, 2H), 7.56 (d, *J* = 8.4 Hz, 2H), 6.99 (d, *J* = 8.4 Hz, 2H), 4.41 (t, *J* = 4.2 Hz, 4H), 4.19 (s, 4H), 3.79 (d, *J* = 4.2 Hz, 4H); ^{13}C NMR (150 MHz, $CDCl_3$): δ 164.1, 154.0, 153.5, 153.2, 138.1, 135.7, 119.7, 118.0, 112.8, 68.8, 63.8, 56.3; HRMS: calcd. for $C_{24}H_{24}N_7O_5$ $[M+H]^+$ 490.1839, found 490.1833; calcd. for $C_{24}H_{23}N_7NaO_5$ $[M+Na]^+$ 512.1658, found 512.1651.

Synthesis of CMBL3aL, CMBL3bL and CMBL3cL:



Scheme S3. Synthesis of **CMBL3aL**, **CMBL3bL** and **CMBL3cL**

(*E*)-4-amino-*N*-(3-(7,16-dioxo-8,15-dioxa-3,6,17-triaza-1,5(2,7-bis(1,8-naphthyridina))cycloheptadecaphan-11-en-3-yl)propyl)butanamide (CMBL3aL): 
Sodium triacetoxyborohydride (7.6 mg, 0.036 mmol) was added to a solution of **CMBL3a** (12 mg, 0.024 mmol)[3] and **14** (9.3 mg, 0.036 mmol)[4,5] in dry chloroform (2.5 mL). The pH of the reaction mixture was adjusted to 6 with acetic acid. Reaction mixture was then stirred at room temperature for 1 h. Reaction was monitored by TLC. Reaction mixture was diluted with chloroform (20 ml) and washed with saturated $NaHCO_3$ (10 mL x 1), water (10 mL) and brine (10 mL). Organic part dried over sodium sulfate and evaporated to dryness. Crude product was purified by column chromatography (2.5-3.0% methanol-chloroform) to obtained pure **BocCMBL3aL** as off-white solid (8.2 mg, 43%). Resultant pure **BocCMBL3aL** (7.7 mg, 0.012 mmol) was dissolved in chloroform (1 mL) and cooled to 0 ºC followed by addition of 4N HCl-ethylacetate (1 mL). Reaction mixture was then stirred at 0 ºC to room temperature for 1 h under argon atmosphere. All the solvent was evaporated to dryness. Crude product was basified by 7 N ammonia solutions in methanol followed by basic silica gel column chromatography (2-5% methanol- chloroform) to obtained pure **CMBL3aL**. Off-white solid (2.7 mg, 40.1%); ^{1}H NMR (600 MHz, CD_3OD) δ 8.50 (d, *J* = 9.0 Hz, 2H), 8.33 (d, *J* = 7.8 Hz, 2H), 7.77 (d, *J* = 9.0 Hz, 2H), 7.65 (d, *J* = 8.4 Hz, 2H), 5.69-5.67 (m, 2H), 5.03 (s, 4H), 4.35 (t, *J* = 5.4 Hz, 4H), 3.67 (t, *J* = 7.2 Hz, 2H), 3.37 (t, *J* = 6.0 Hz, 2H), 3.13 (t, *J* = 6.6 Hz, 2H), 2.57 (t, *J* = 6.6 Hz, 2H), 2.47 (td, *J* = 5.4, 5.25 Hz, 4H), 2.11 (quint, *J* = 6.75 Hz, 2H), 2.02 (quint, *J* = 6.45 Hz, 2H); ^{13}C NMR (150 MHz, CD_3OD) δ 176.7, 158.3, 154.8, 141.3, 130.4, 123.6, 120.2, 116.1, 71.4, 67.6, 60.3, 56.3, 40.8, 37.1, 34.4, 33.0, 26.3, 24.0; HRMS (ESI) *m/z*: Calcd. for $C_{33}H_{40}N_9O_5$ $[M+H]^+$ 642.3152; Found: 642.3132; calcd. for $C_{33}H_{39}N_9NaO_5$ $[M+Na]^+$ 664.2972; Found: 664.2949.

(*Z*)-4-amino-*N*-(3-(7,16-dioxo-8,15-dioxa-3,6,17-triaza-1,5(2,7-bis(1,8-naphthyridina))cycloheptadecaphan-11-en-3-yl)propyl)butanamide (CMBL3bL): 
Sodium triacetoxyborohydride (9.5 mg, 0.045 mmol) was added to a solution of **CMBL3b** (15 mg, 0.030 mmol)[3] and **14** (12 mg, 0.045 mmol)[4,5] in dry chloroform (2.5 mL). The pH of the reaction mixture was adjusted to 6 with acetic acid. Reaction mixture was then stirred at room temperature for 1 h. Reaction was monitored by TLC. Reaction mixture was diluted with chloroform (20 ml) and washed with

saturated $NaHCO_3$ (10 mL x 1), water (10 mL) and brine (10 mL). Organic part dried over sodium sulfate and evaporated to dryness. Crude product was purified by column chromatography (2.5-3.0% methanol-chloroform) to obtained desire pure **BocCMBL3aL** as off-white solid (11.8 mg, 53.0%); To an ice cold stirred solution of resultant pure **BocCMBL3bL** (11.0 mg, 0.012 mmol) in chloroform (1 mL) was added 4N HCl-ethylacetate (1 mL) under argon atmosphere followed by stirring at 0 ºC to room temperature for 1 h. All the solvent was evaporated to dryness. Crude product was basified by 7 N ammonia solutions in methanol followed by basic silica gel column chromatography (2-5% methanol-chloroform). Desired product was further purified by HPLC using acetonitrile-water (0.1% AcOH) as mobile phase to obtained pure **CMBL3bL** as acetate salt. Off-white solid (5.8 mg, 61%); ^{1}H NMR (700 MHz, CD_3OD): δ 8.04 (d, J = 7.7 Hz, 2H), 7.92 (br, 2H), 7.32 (d, J = 8.4 Hz, 2H), 7.30 (d, J = 9.1 Hz, 2H), 5.69-5.65 (m, 2H), 4.28 (t, J = 7.4 Hz, 4H), 3.94 (s, 4H), 3.40 (t, J = 6.0 Hz, 2H), 3.14-3.31 (m, 2H), 2.84 (t, J = 6.0 Hz, 2H), 2.62 (q, J = 7.0 Hz, 4H), 2.50 (t, J = 6.7 Hz, 2H), 2.20 (quint, J = 6.8 Hz, 2H), 1.93 (s, 3H; acetate CH_3), 1.80 (quint, J = 6.0 Hz, 2H); ^{13}C NMR (175 MHz, CD_3OD) δ 178.5, 174.8, 164.1, 155.7, 155.3, 154.4, 140.1, 138.4, 128.9, 123.1, 119.7, 115.3, 65.9, 63.4, 57.0, 39.9, 34.1, 28.4, 27.5, 25.0, 22.9 (CH_3 from acetate salt); HRMS (ESI) *m/z*: Calcd. for $C_{33}H_{40}N_9O_5$ $[M+H]^+$ 642.3152; Found: 642.3146; calcd. for $C_{33}H_{39}N_9NaO_5$ $[M+Na]^+$ 664.2972; Found: 664.2964.

4-amino-*N*-(3-(7,16-dioxo-8,15-dioxa-3,6,17-triaza-1,5(2,7-bis(1,8-naphthyridina))cycloheptadecaphane-3-yl)propyl)butanamide (CMBL3cL): Sodium triacetoxyborohydride (16 mg, 0.075 mmol) was added to a solution of **CMBL3c** (25 mg, 0.050 mmol)[3] and **14** (19 mg, 0.075 mmol)[4,5] in dry chloroform (2.5 mL). The pH of the reaction mixture was adjusted to 6 with acetic acid. Reaction mixture was then stirred at room temperature for 1 h. Reaction was monitored by TLC. Reaction mixture was diluted with chloroform (30 ml) and washed with saturated $NaHCO_3$ (10 mL x 1), water (10 mL) and brine (10 mL). Organic part dried over sodium sulfate and evaporated to dryness. Crude product was purified by column chromatography (2.5-3.0% methanol-chloroform) to obtained desire product **BocCMBL3cL** as off-white solid (19.7 mg, 53.1%). To an ice-cold stirred solution of **BocCMBL3cL** (16 mg, 0.022 mmol) in chloroform (1 mL) was added 4N HCl-ethylacetate (1 mL) at 0 °C under argon atmosphere. Reaction mixture was then stirred at 0 °C to room temperature for 1 h. All the solvent was evaporated to dryness. Crude product was basified by 7 N ammonia solutions in methanol followed by basic silica gel column chromatography (2-5% methanol-chloroform) to obtained pure **CMBL3cL**. Off-white solid (8.5 mg, 61.4%); ^{1}H NMR (700 MHz, DMSO-d_6): δ 8.06 (t, J = 4.9 Hz, 1H), 7.94 (d, J = 9.1 Hz, 2H), 7.79 (d, J = 8.4 Hz, 2H), 7.75 (d, J = 8.4 Hz, 2H), 7.67-7.60 (br, 2H), 7.23 (d, J = 7.7 Hz, 2H), 4.19 (t, J = 4.9 Hz, 4H), 4.15-4.13 (br, 2H), 3.86 (s, 4H), 3.32 (q, J = 6.3 Hz, 2H), 2.95 (t, J = 6.3 Hz, 2H), 2.66 (t, J = 7.0 Hz, 2H), 2.17 (t, J = 7.35 Hz, 2H), 1.70-1.65 (m, 4H), 1.52-1.50 (br, 4H), 1.28-1.22 (m, 4H); ^{13}C NMR (175 MHz, DMSO-d_6): δ 172.0, 162.9, 153.8, 153.5, 153.4, 138.1, 135.5, 120.6, 117.5, 113.0, 69.8, 64.7, 62.6, 55.9, 40.0, 36.7, 33.0, 31.0, 28.5, 27.8, 22.4, 22.1; HRMS (ESI) *m/z*: Calcd. for $C_{33}H_{42}N_9O_5$ $[M+H]^+$ 644.3309; Found: 644.3311; calcd. for $C_{33}H_{41}N_9NaO_5$ $[M+Na]^+$ 666.3128; Found: 666.3128.

References for supporting data:

1. Bradshaw,J.S., Guynn,J.M., Wood,S.G., Wilson,B.E., Kent Dalley,N. and Izatt,R.M. (1987) Proton−ionizable crown compounds. 9. Synthesis and structural studies of new 14−crown−4 compounds containing a pyridine or 4−pyridone subcyclic unit. *J. Heterocycl. Chem.*, 10.1002/jhet.5570240222.

2. Mukherjee,S., Dohno,C., Asano,K. and Nakatani,K. (2016) Cyclic mismatch binding ligand CMBL4 binds to the 5′-T-3′/5′-GG-3′ site by inducing the flipping out of thymine base. *Nucleic Acids Res.*, **44**, gkw672.

3. Mukherjee,S., Dohno,C. and Nakatani,K. (2017) Design and Synthesis of Cyclic Mismatch-Binding Ligands (CMBLs) with Variable Linkers by Ring-Closing Metathesis and their Photophysical and DNA Repeat Binding Properties. *Chem. - A Eur. J.*, **23**, 11385–11396.

4. Hagihara,S., Kumasawa,H., Goto,Y., Hayashi,G., Kobori,A., Saito,I. and Nakatani,K. (2004) Detection of guanine-adenine mismatches by surface plasmon resonance sensor carrying naphthyridine-azaquinolone hybrid on the surface. *Nucleic Acids Res.*, 10.1093/nar/gkh171.

5. Nakatani,K., He,H., Uno,S.N., Yamamoto,T. and Dohno,C. (2008) Synthesis of dimeric 2-amino-1,8-naphthyridine and related DNA-binding molecules. *Curr. Protoc. Nucleic Acid Chem.*, 10.1002/0471142700.nc0806s32.