Supplemental Information for:

A small molecule that binds an RNA repeat expansion stimulates its decay via the exosome complex

Alicia J. Angelbello[1,5], Raphael I. Benhamou[1,5], Suzanne G. Rzuczek[1,5], Shruti Choudhary[1], Zhenzhi Tang[2], Jonathan L. Chen[1], Madhuparna Roy[3], Kye Won Wang[4], Ilyas Yildirim[4], Albert S. Jun[3], Charles A. Thornton[2], and Matthew D. Disney [1,6,*]

[1] Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458, United States of America

[2] Department of Neurology, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642, United States of America

[3] Wilmer Eye Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, United States of America

[4] Department of Chemistry, Florida Atlantic University, Jupiter, FL 33458 United States of America

[5] These authors contributed equally to this work

[6] Lead contact

* correspondence: Disney@scripps.edu

Table S1. Sequences of primers used for RT-PCR and RT-qPCR (related to STAR Methods).

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')	Purpose
MBNL1	GCTGCCCAATACCAGGTCAAC	TGGTGGGAGAAATGCTGTATGC	RT-PCR
MBNL1 exon 5	CTCAGTCGGCTGTCAAATCA	AGAGCAGGCCTCTTTGGTAA	qPCR
MAP4K4	CCTCATCCAGTGAGGAGTCG	ATCACAGGAAAATCCCACCA	RT-PCR
DMPK	CGTGCAAGCGCCCAG	CTCCACCAACTTACTGTTTCATCCT	qPCR
GAPDH	AAGGTGAAGGTCGGAGTCAA	AATGAAGGGGTCATTGATGG	qPCR
Clcn1	TGAAGGAATACCTCACACTCAAGG	CACGGAACACAAAGGCACTG	RT-PCR
Serca1	GCTCATGGTCCTCAAGATCTCAC	GGGTCAGTGCCTCAGCTTTG	RT-PCR
Itgb	CCTACTGGTCCCGACATCATC	CTTCGGATTGACCACAGTTGTC	RT-PCR
Capzb	GCACGCTGAATGAGATCTACTTTG	CCGGTTAGCGTGAAGCAGAG	RT-PCR
TCF4 mature mRNA	ACGATGAGGACCTGACAC	GTCTGGGGCTTGTCACTCTT	qPCR
TCF4 intron 3	GAGAGAGGGAGTGAAAGAGAGA	GGCAATGTCCATTTCCATCT	qPCR
EXOSC10	CTCTTTGGACCTCACGACTGCT	AAGAGGCTCGCCTGCTTCTGAA	qPCR
DIS3	ACCCTCACTTAAAATAGAAGATACAGT	CCATTAAGGTCCATGTTTGAAGT	qPCR
XRN1	CCAGCAAAGCAGTCGTGGAGAA	CCACGACTCTAGCTTCCTCAAG	qPCR

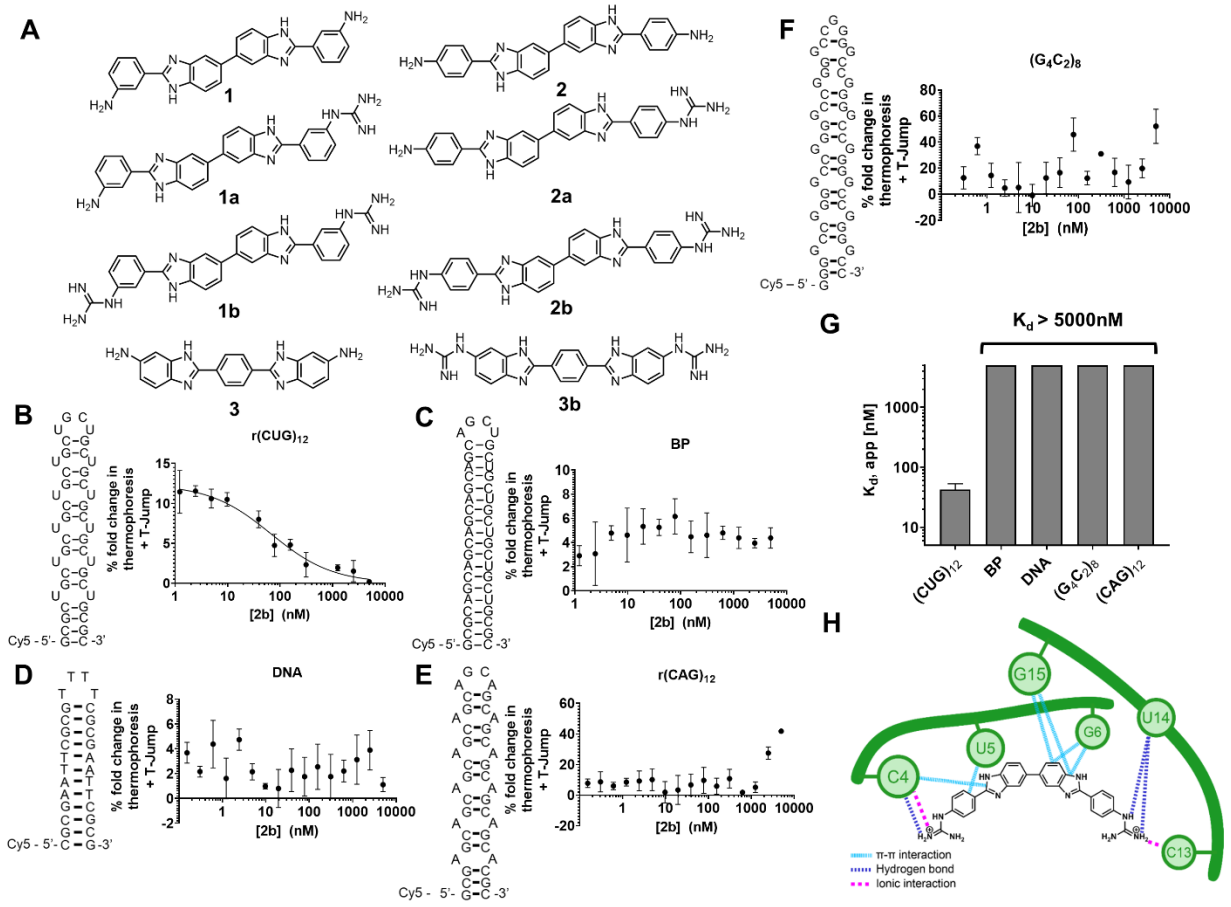


Figure S1. Chemical structures and binding analysis of **2b** (related to Figure 2). (A) Chemical structures of all compounds tested. (B) Thermophoresis analysis of **2b** bound to r(CUG)$_{12}$ (n = 3 independent experiments). (C) Thermophoresis analysis of **2b** with a base-paired control RNA (n = 3 independent experiments). (D) Thermophoresis analysis of **2b** with AT-rich (n = 3 independent experiments). (E) Thermophoresis analysis of **2b** with a r(CAG)$_{12}$ (n = 3 independent experiments). (F) Thermophoresis analysis of **2b** with a r(G$_4$C$_2$)$_8$ (n = 3 independent experiments). (G) Plot of the $K_{d,app}$ of **2b** and r(CUG)$_{12}$ (K_d = 42 ± 6 nM) and to a base-paired control , AT-rich DNA, r(G$_4$C$_2$)$_8$ and r(CAG)$_{12}$ (K_d > 5000 nM). Error bars indicate SD for all panels. (G) Schematic of interactions of 2b with r(CCGCUGCGG)$_2$. Pink lines are ionic interactions, dark blue lines are hydrogen bonding interactions, and light blue lines are stacking interactions.

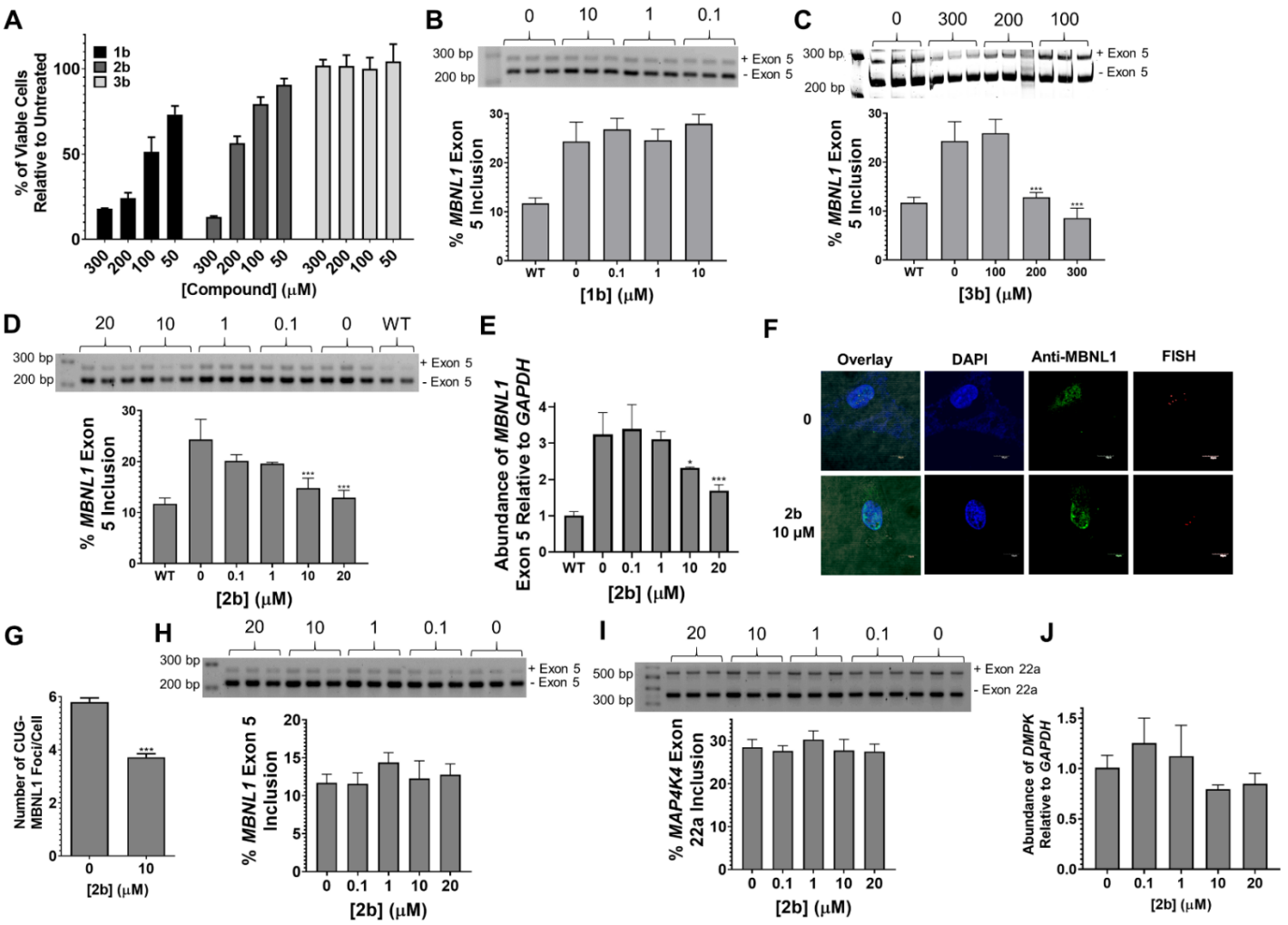


Figure S2. Activity of compounds in DM1 and wild-type fibroblasts (related to Figure 2). (A) Toxicity of **1b**, **2b**, and **3b** in DM1 fibroblasts as determined by WST-1 cell viability reagent (n = 5). (B) Representative gel image of *MBNL1* exon 5 splicing in DM1 fibroblasts treated with **1b** (top) and quantification of *MBNL1* exon 5 inclusion in DM1 fibroblasts treated with **1b** (bottom) (n = 3). (C) Representative gel image of *MBNL1* exon 5 splicing in DM1 fibroblasts treated with **3b** (top) and quantification of *MBNL1* exon 5 inclusion in DM1 fibroblasts treated with **3b** (bottom) (n = 3). (D) Representative gel image of *MBNL1* exon 5 splicing in DM1 fibroblasts treated with **2b** (top) and quantification of *MBNL1* exon 5 inclusion in DM1 fibroblasts treated with **2b** (bottom) (n = 3). (E) RT-qPCR analysis of *MBNL1* exon 5 inclusion in DM1 fibroblasts treated with **2b** (n = 3). (F) Representative images of r(CUG)^{exp}-MBNL1 foci as determined by MBNL1

immunostaining and RNA FISH. (G) Quantification of the number of nuclear foci/cell (n = 3, 40 cells counted/replicate). Error bars indicate SD; ***, $P < 0.001$, as determined by a two-tailed Student *t*-test. (H) Representative gel image of *MBNL1* exon 5 splicing in wild-type fibroblasts treated with **2b** (top) and quantification of *MBNL1* exon 5 splicing in wild-type fibroblasts treated with **2b** (bottom) (n = 3). (I) Representative gel image of *MAP4K4* exon 22a splicing in DM1 fibroblasts treated with **2b** (top) and quantification of *MAP4K4* exon 22a splicing in DM1 fibroblasts treated with **2b** (bottom) (n = 3). (J) Compound **2b** does not affect *DMPK* abundance in DM1 fibroblasts, as determined by RT-qPCR (n = 3). For all panels: error bars indicate SD; **, $P < 0.01$; ***, $P < 0.001$, as determined by a one-way ANOVA, unless otherwise noted.

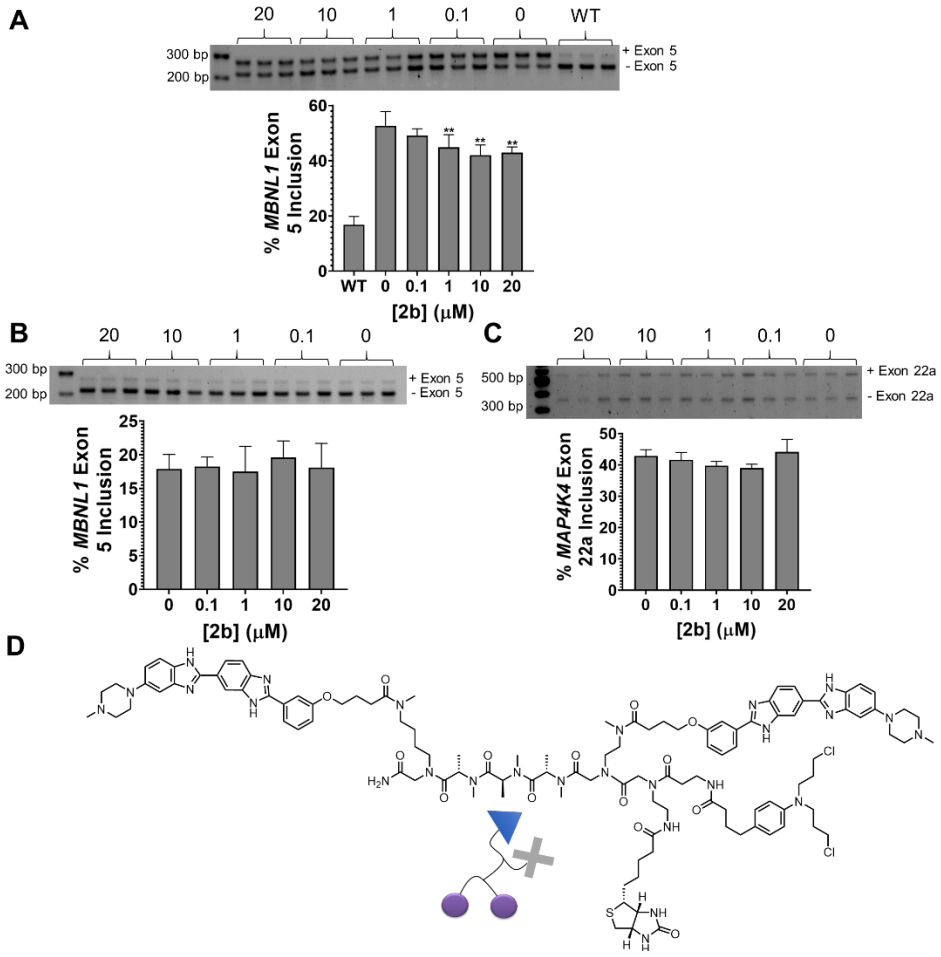


Figure S3. Activity of **2b** in DM1 and wild-type myotubes (related to Figures 3 and 4). (A) Representative gel image of *MBNL1* exon 5 splicing in DM1 myotubes treated with **2b** (top) and quantification of *MBNL1* exon 5 inclusion in DM1 myotubes treated with **2b** (bottom) (n = 3). (B) Representative gel image of *MBNL1* exon 5 splicing in wild-type myotubes treated with **2b** (top) and quantification of *MBNL1* exon 5 splicing in wild-type myotubes treated with **2b** (bottom) (n = 3). (C) Representative gel image of *MAP4K4* exon 22a splicing in DM1 myotubes treated with **2b** (top) and quantification of *MAP4K4* exon 22a splicing in DM1 myotubes treated with **2b** (bottom) (n = 3). (D) Chemical structure of Chem-CLIP probe **2H-K4NMeS-CA-Biotin** synthesized as previously described (Rzuczek et al., 2017). For all panels: n = 3; error bars indicate SD; **, $P < 0.01$; ***, $P < 0.001$, as determined by a one-way ANOVA.

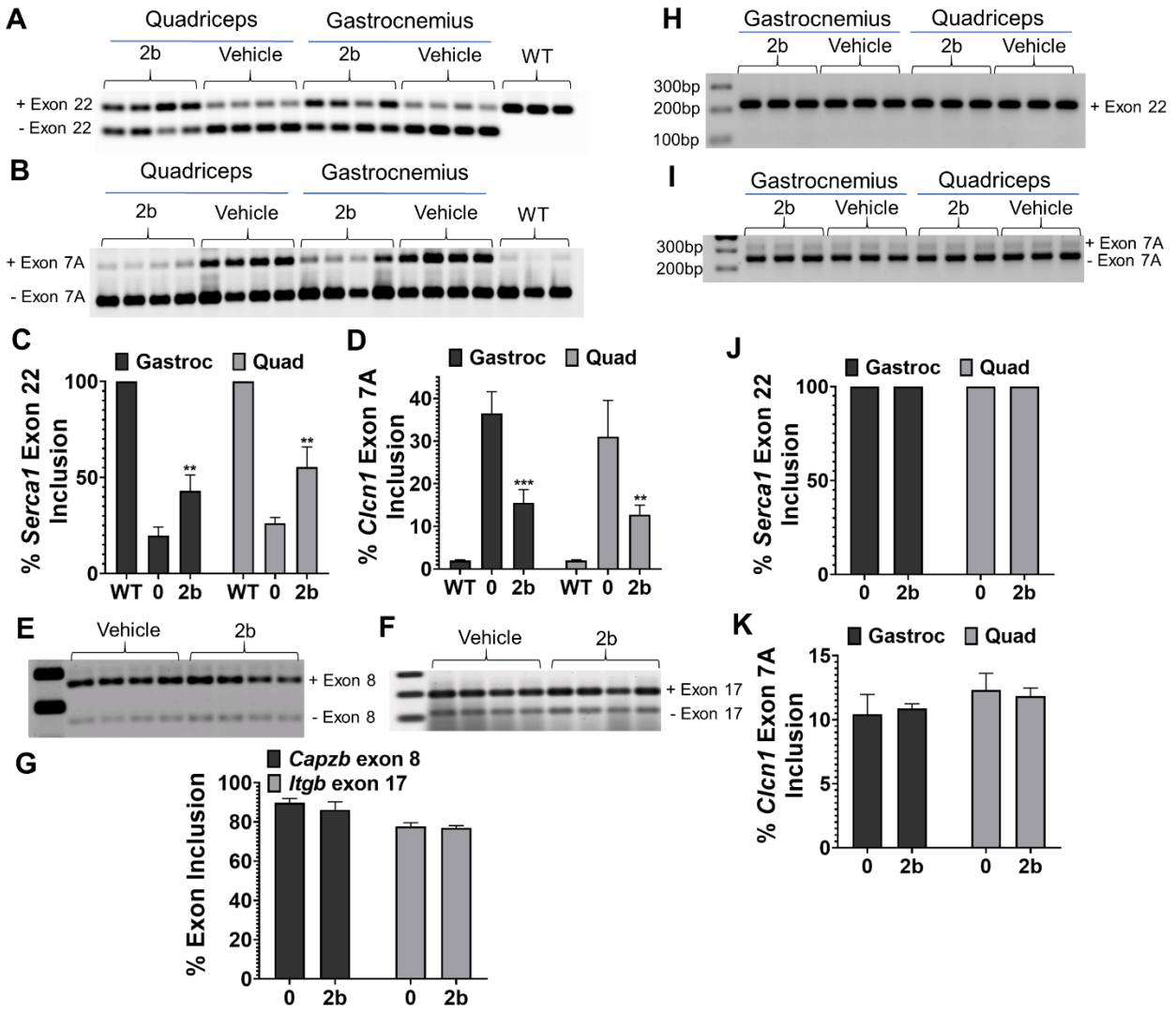


Figure S4. Activity of **2b** in *HSA*^{LR} mice (related to Figure 3). (A) Representative gel image of *Serca1* exon 22 splicing in the quadriceps and gastrocnemius muscles of *HSA*^{LR} mice treated with **2b** (40 mg/kg). (B) Representative gel image of *Clcn1* exon 7A splicing in the quadriceps and gastrocnemius muscles of *HSA*^{LR} mice treated with **2b**. (C) Quantification of *Serca1* exon 22 inclusion. (D) Quantification of *Clcn1* exon 7A inclusion. (E) Representative gel image of *Capzb* exon 8 splicing (non-MBNL1 regulated (Lin et al., 2006)) in the quadriceps muscle of *HSA*^{LR} mice treated with **2b** (40 mg/kg). (F) Representative gel image of *Itgb* exon 17 splicing [non-MBNL1 regulated (Lin et al., 2006)] in the quadriceps muscle of *HSA*^{LR} mice treated with **2b**. (G) Quantification of *Capzb* and *Itgb* exon inclusion. (H) Representative gel image of *Serca1* exon 22 splicing in the quadriceps and gastrocnemius muscles of wild-type (FVB) mice treated with **2b**.

(I) Representative gel image of *Clcn1* exon 7A splicing in the quadriceps and gastrocnemius muscles of wild-type (FVB) mice treated with **2b**. (J) Quantification of *Serca1* exon 22 inclusion. (K) Quantification of *Clcn1* exon 7A inclusion. For all panels: error bars represent SD, **, $P < 0.01$; ***, $P < 0.001$; as determined by a one-way ANOVA relative to vehicle treated (0). For HSA^{LR} experiments, n = 4 mice/group; for FVB experiments, n = 3 mice/group.

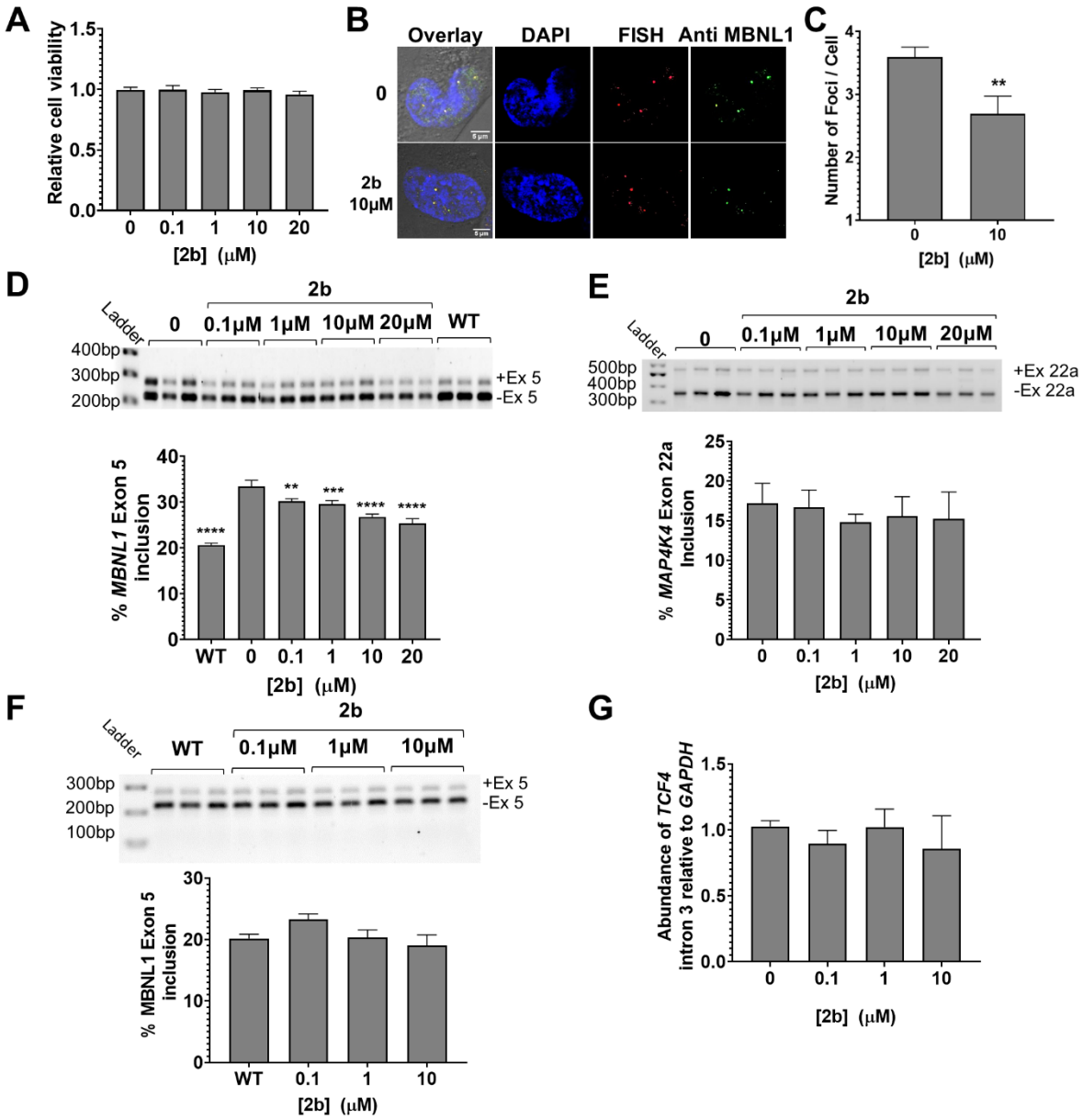


Figure S5. Activity of **2b** in FECD and wild-type corneal endothelial cells (related to Figure 5). (A) Toxicity of **2b** in FECD cells (F35T), as determined by CellTiter-Glo (n = 5). (B) Representative images of r(CUG)^{exp}-MBNL1 foci as determined by MBNL1 immunostaining and RNA FISH. (C) Quantification of the number of nuclear foci/cell (n = 3, 40 nuclei counted/replicate); **, $P < 0.01$, as determined by a two-tailed Student *t*-test. (D) Representative gel image of *MBNL1* exon 5 splicing in FECD cells treated with **2b** (top) and quantification of *MBNL1* exon 5 splicing (bottom) (n =3); *, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$, as determined by a one-way ANOVA. (E) Representative gel

image of *MAP4K4* exon 22a (non-MBNL1 regulated) splicing in FECD cells treated with **2b** (top) and quantification of *MAP4K4* exon 22a splicing (bottom) (n =3). (F) Representative gel image of *MBNL1* exon 5 splicing in wild-type (F20T) cells treated with **2b** (top) and quantification of *MBNL1* exon 5 splicing (n = 3). (G) Abundance of *TCF4* intron 3 levels in F20T cells treated with **2b** assessed by RT-qPCR (n = 3). For all panels error bars represent SD.

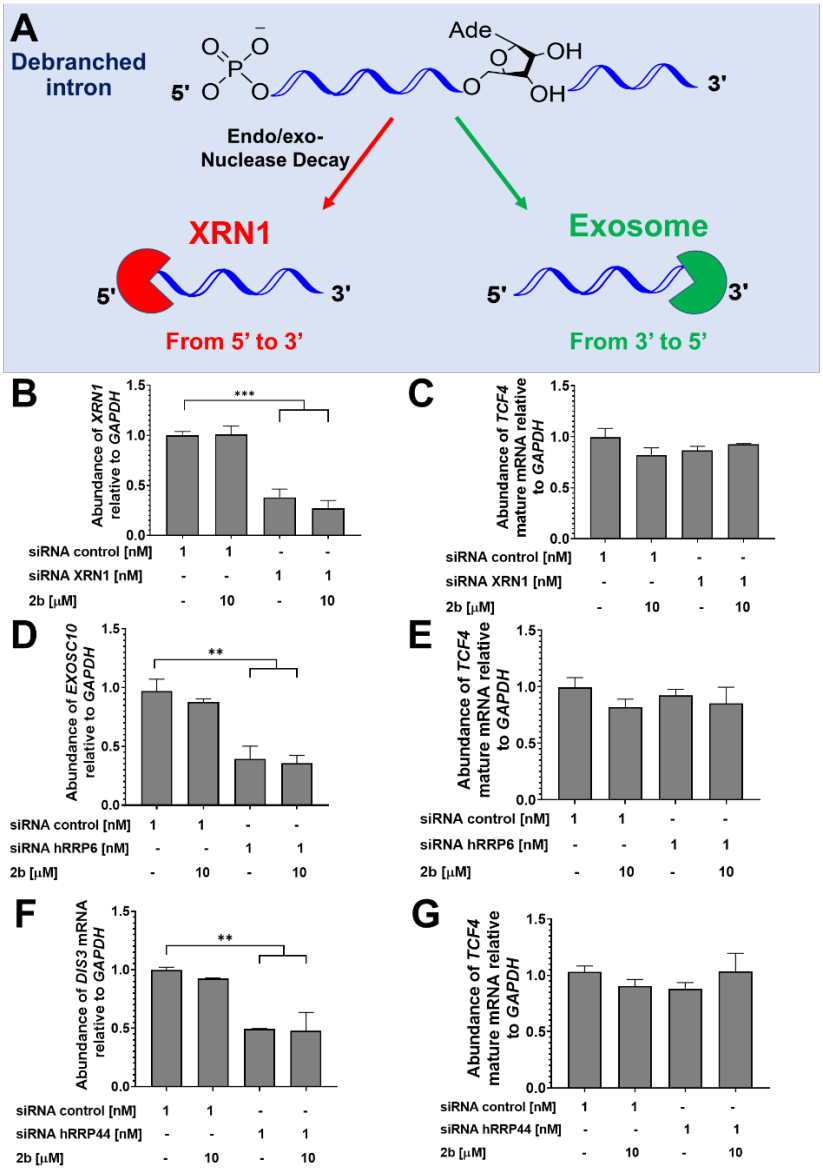
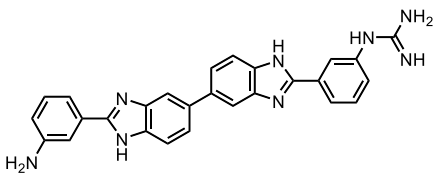


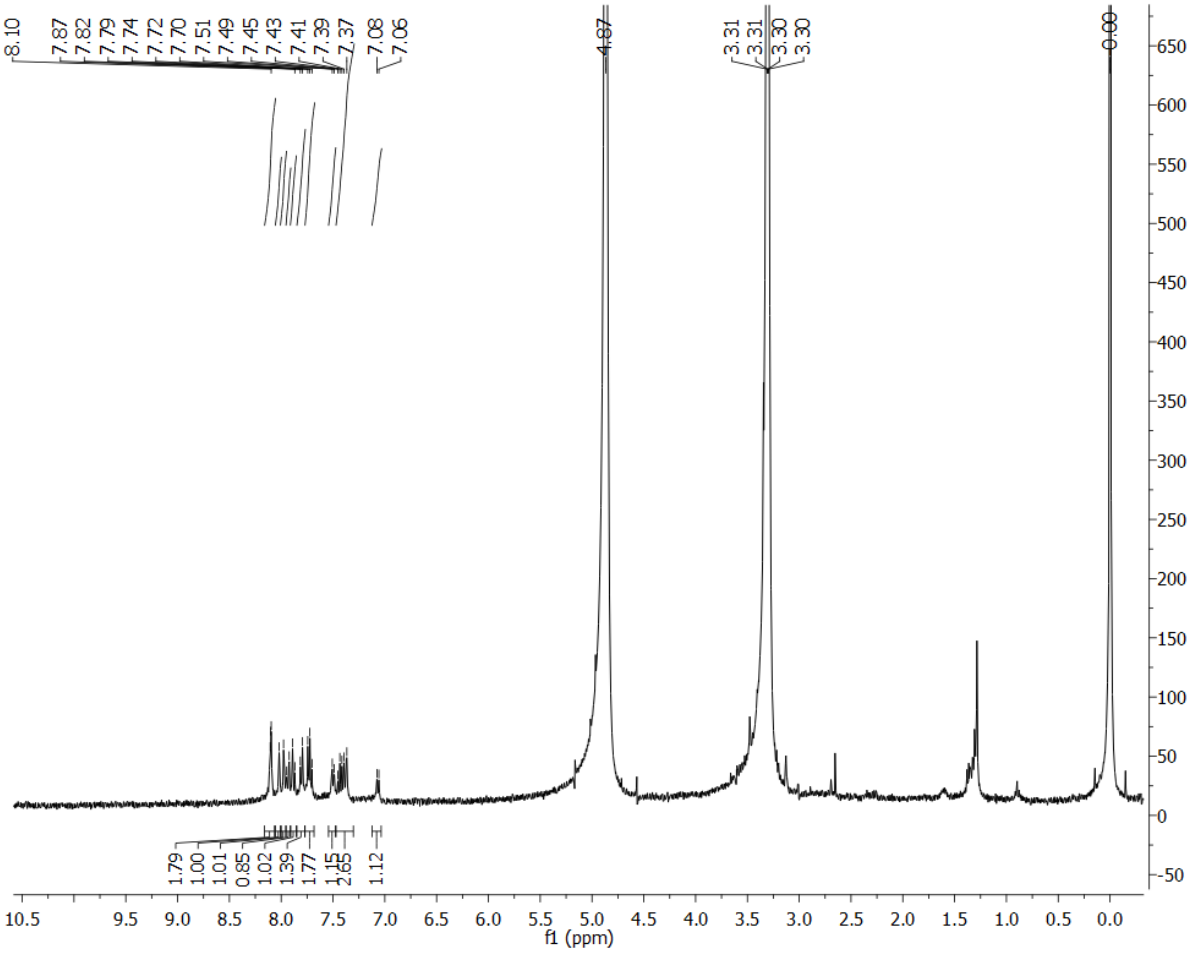
Figure S6. Studying the RNA quality control pathways responsible for intron degradation upon **2b** treatment in FECD cells (related to Figure 6). (A) Decay mechanisms for the exosome complex and XRN1 . (B, C) Impact of the siRNA knock down of *XRN1* on **2b**'s ability to reduce *TCF4* intron 3 levels in FECD cells. (B) *XRN1* mRNA levels, as determined by RT-qPCR. (C) *TCF4* mature mRNA levels, as determined by RT-qPCR. (D, E) Impact of the siRNA knock down of *hRRP6* on **2b**'s ability to reduce *TCF4* intron 3 levels in FECD cells. (D) *EXOSC10* (*hRRP6*) mRNA levels, as determined by RT-qPCR. (E) *TCF4* mature mRNA levels, as determined by RT-qPCR. (F, G) Effect of knock-down of *hRRP44* via an siRNA on **2b**'s ability to reduce *TCF4* intron 3 levels in FECD

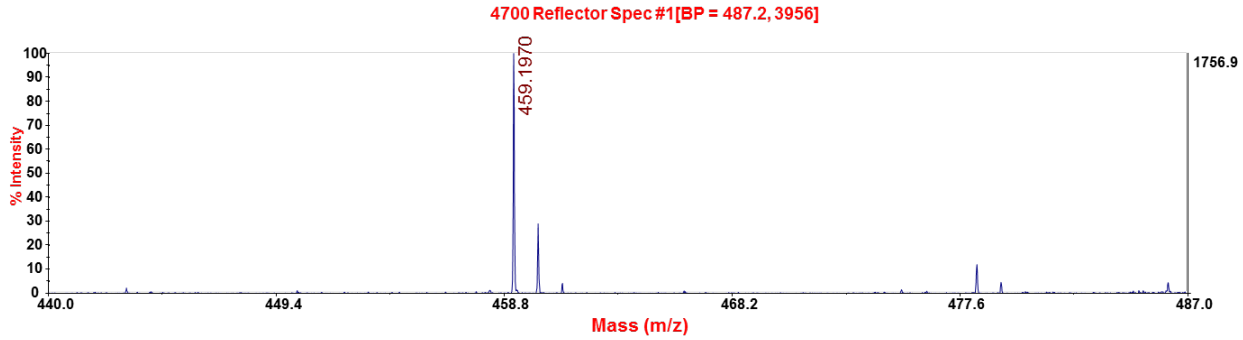
cells. (F) *DIS3* (hRRP44) mRNA levels, as determined by RT-qPCR. (G) *TCF4* mature mRNA levels, as determined by RT-qPCR. For all panels, n = 3, and error bars represent SD. **, $P < 0.01$; ***, $P < 0.001$, as determined by one-way ANOVA compared to untreated cells ("0").

Data S1. Compound characterization (NMR and HRMS spectra and analytical HPLC traces) (related to STAR Methods).

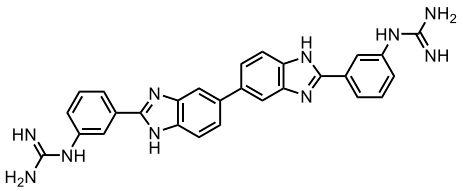
1a:

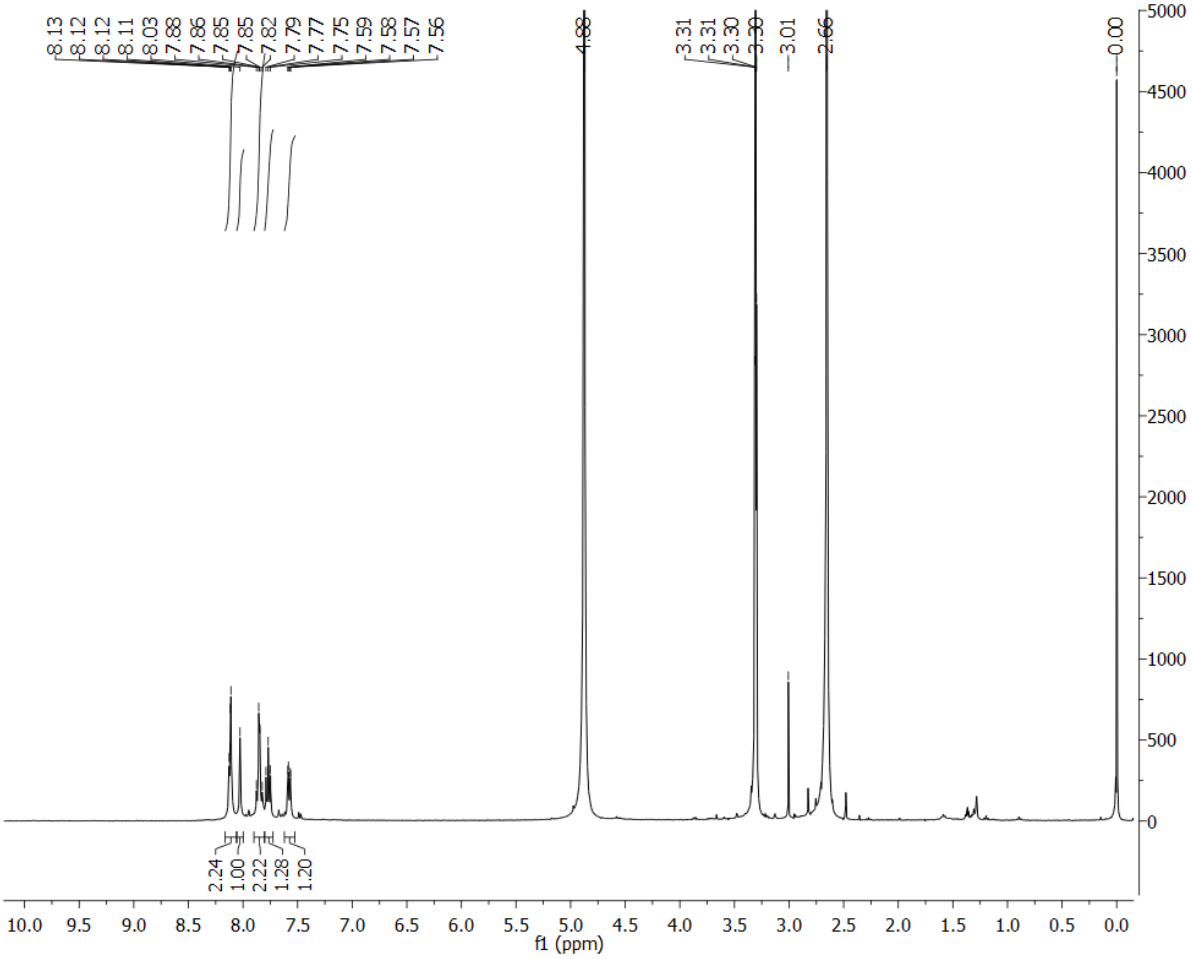


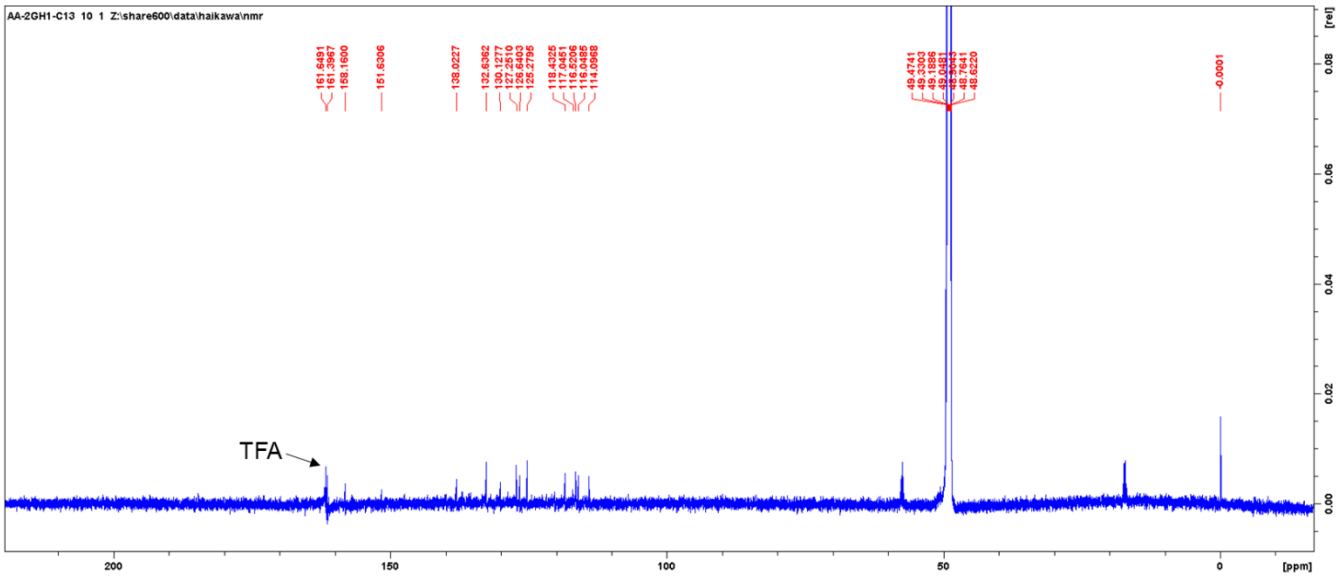


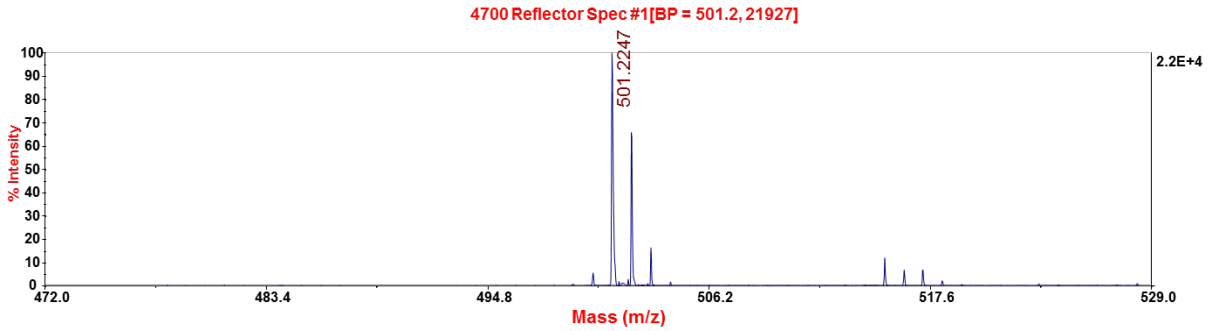

4700 Reflector Spec #1[BP = 487.2, 3956]
% Intensity
459.1970
1756.9
Mass (m/z)

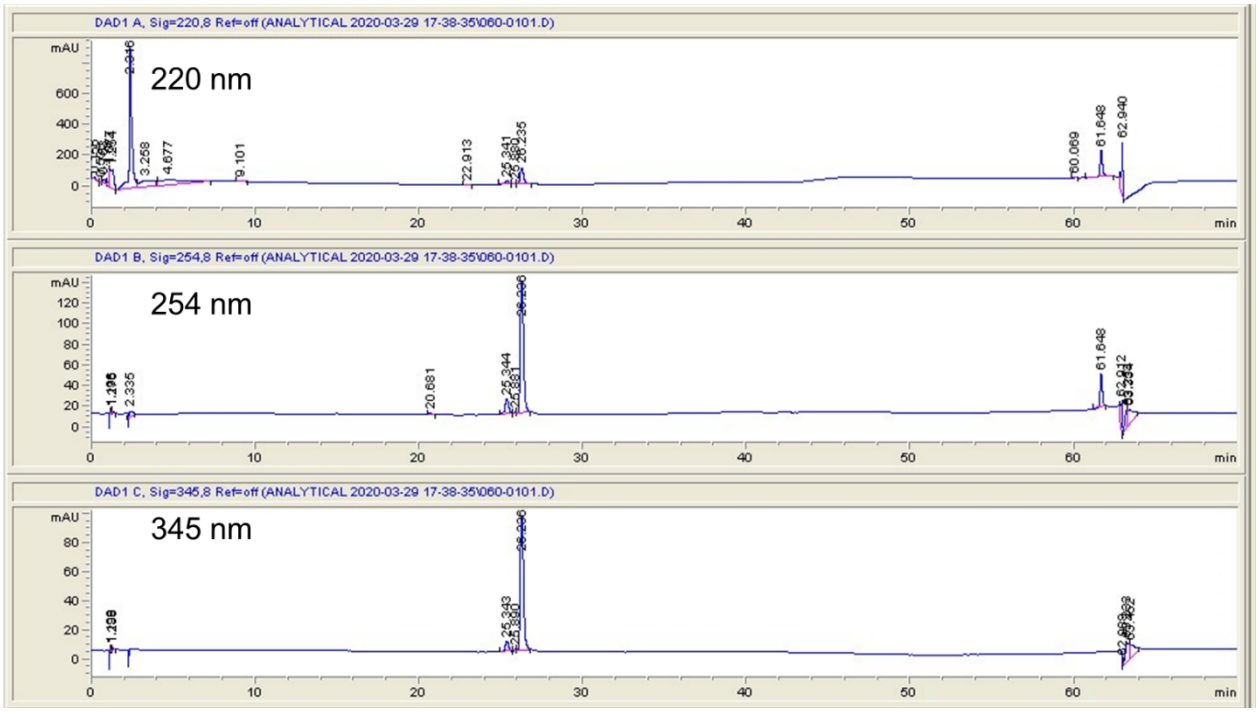
1b:



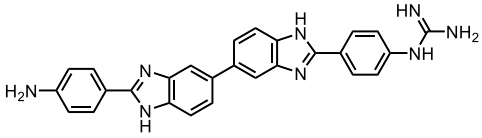

f1 (ppm)

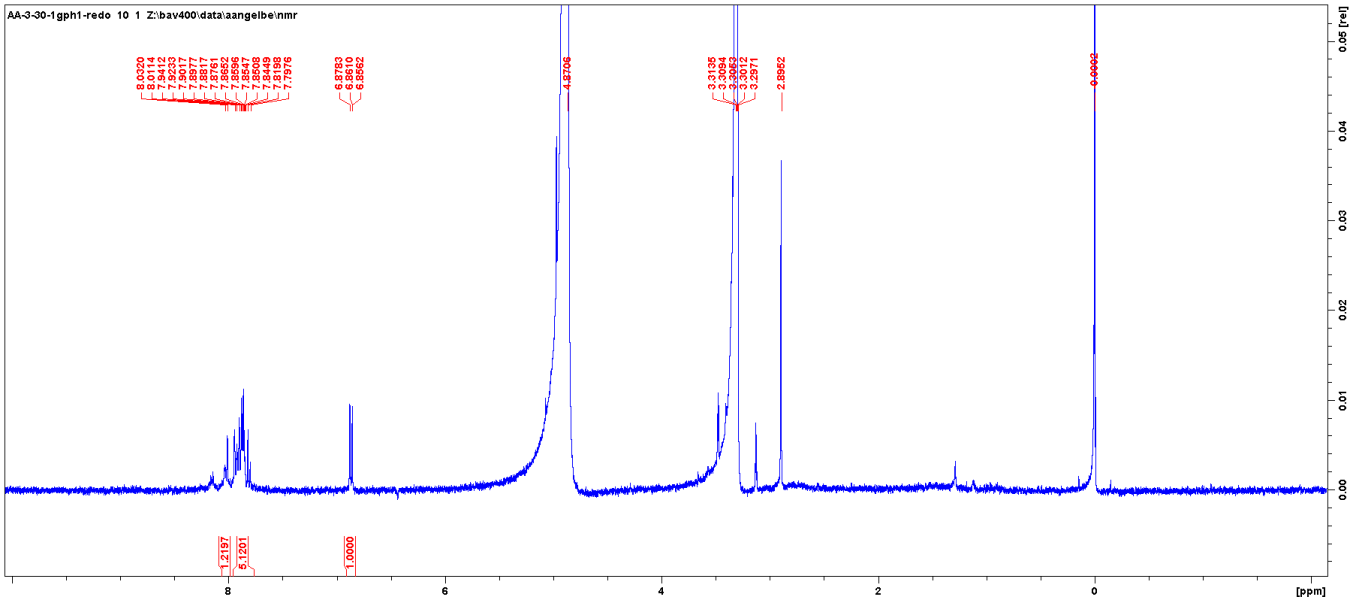

AA-2GH1-C13 10 1 Z:\share600\data\haikawa\nmr
TFA
[ppm]
[rel]

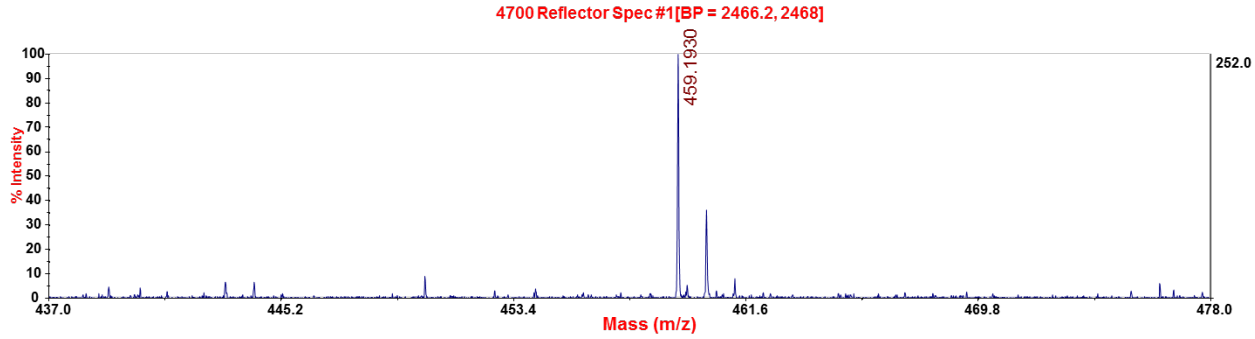

4700 Reflector Spec #1[BP = 501.2, 21927]
501.2247
2.2E+4
% Intensity
Mass (m/z)


DAD1 A, Sig=220,8 Ref=off (ANALYTICAL 2020-03-29 17-38-35\060-0101.D)
220 nm
DAD1 B, Sig=254,8 Ref=off (ANALYTICAL 2020-03-29 17-38-35\060-0101.D)
254 nm
DAD1 C, Sig=345,8 Ref=off (ANALYTICAL 2020-03-29 17-38-35\060-0101.D)
345 nm
mAU
min

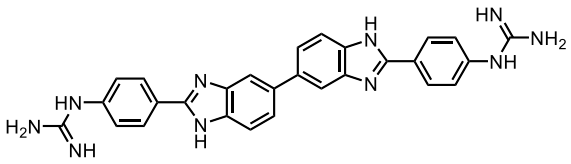
2a:

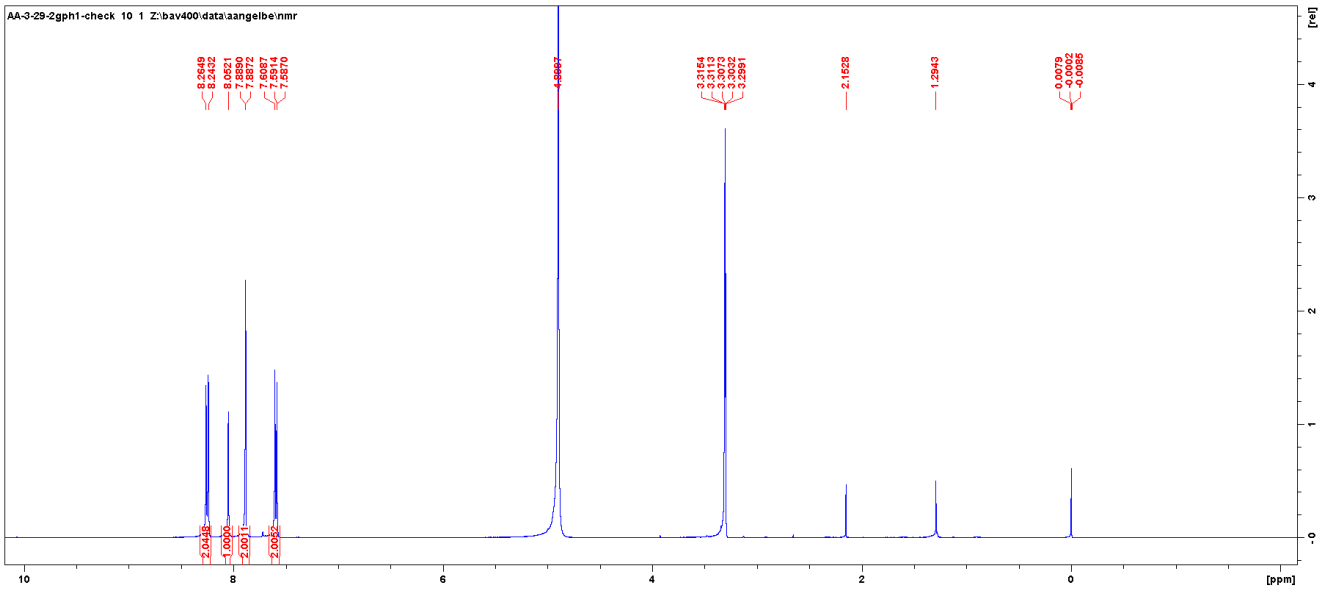

H2N
N
H
N
N
H
HN
NH2
NH

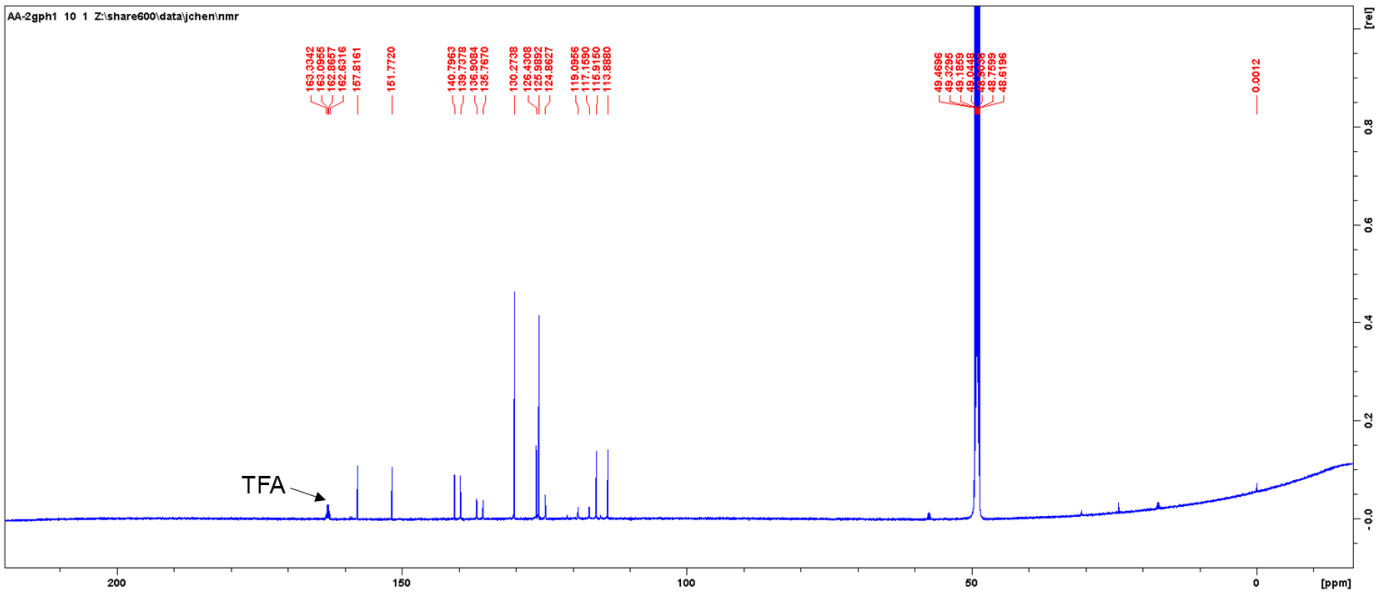

AA-3-30-1gph1-redo 10 1 Z:\bav400\data\aangelbe\nmr
8.0320
8.0114
7.9433
7.9233
7.9017
7.8977
7.8877
7.8761
7.8652
7.8596
7.8567
7.8508
7.8449
7.8198
7.7976
6.8783
6.8610
6.8562
4.8706
3.3135
3.3084
3.3033
3.3012
3.2971
2.8962
0.0000
1.2197
5.1201
1.0000
8
6
4
2
0
[ppm]
0.05 [rel]
0.04
0.03
0.02
0.01
0.00

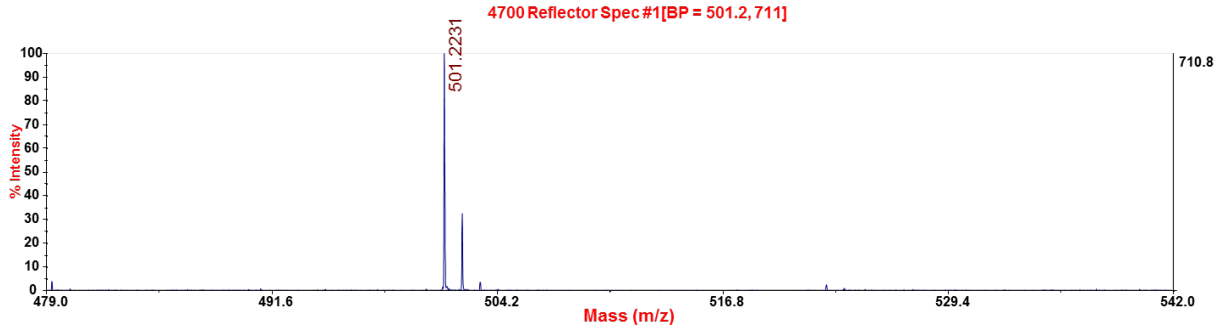

4700 Reflector Spec #1[BP = 2466.2, 2468]
% Intensity
100
90
80
70
60
50
40
30
20
10
0
459.1930
252.0
437.0
445.2
453.4
461.6
469.8
478.0
Mass (m/z)

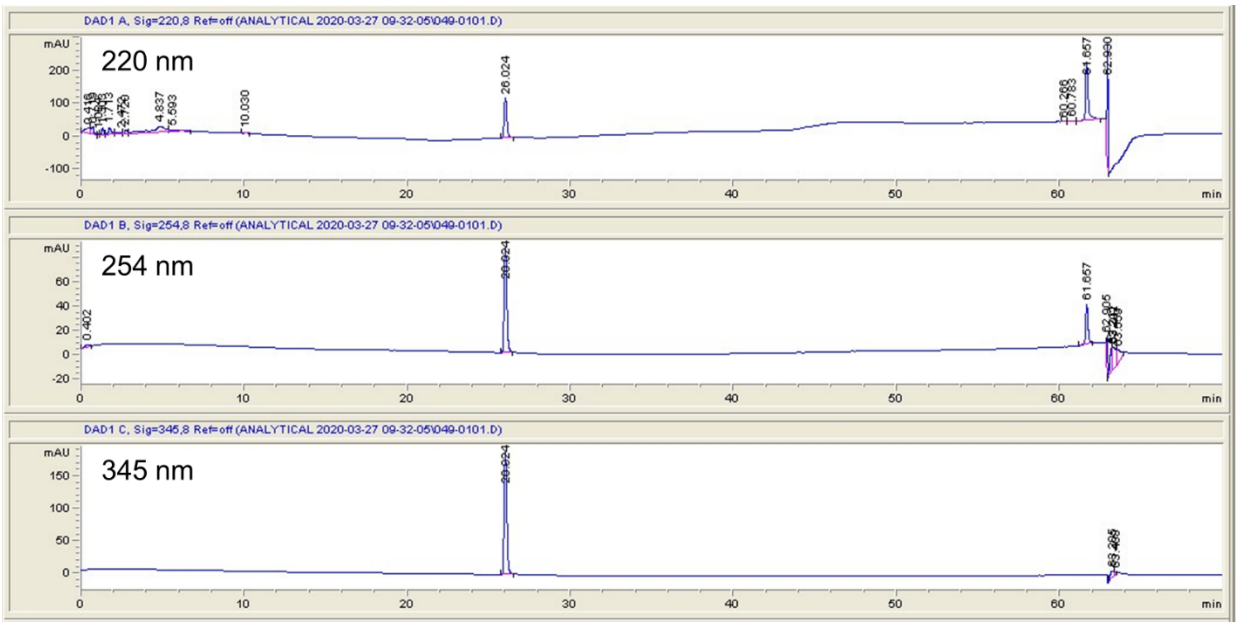
2b:


HN
NH2
NH
H
N
N
HN
H2N
NH
N
H

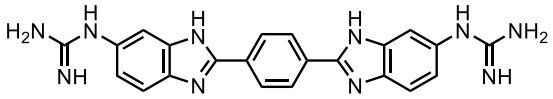

AA-3-29-2gph1-check 10 1 Z:\bav400\data\aangelbe\nmr
8.2649
8.2432
8.0521
7.8890
7.8872
7.6087
7.5914
7.5870
4.8867
3.3154
3.3113
3.3073
3.3032
3.2991
2.1628
1.2943
0.0079
0.0052
-0.0085
2.0448
1.0000
2.0311
2.0052
[rel]
[ppm]

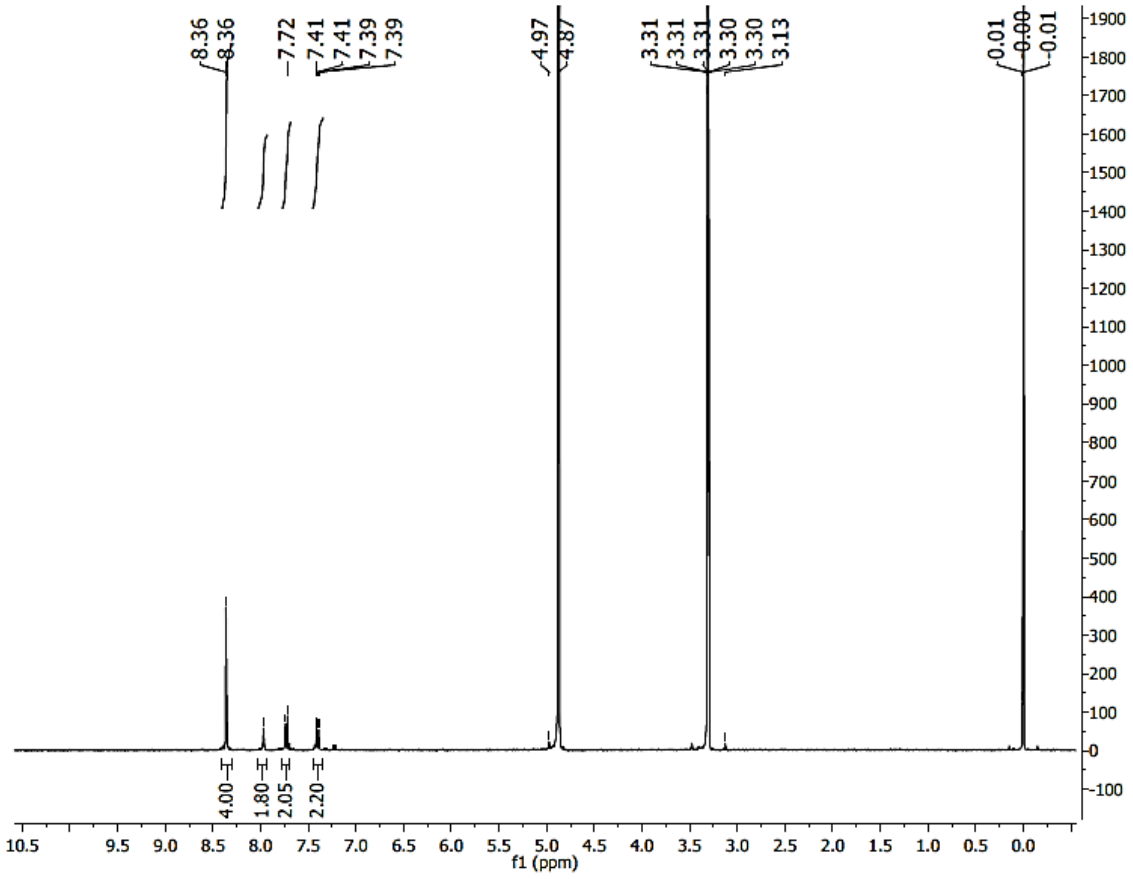

AA-2gph1 10 1 Z:\share600\data\jchen\nmr
163.3342
163.0955
162.8857
162.6759
157.8161
151.7720
140.7963
139.7378
136.9084
135.7670
130.2738
126.4308
125.8992
124.8627
119.0956
117.1590
115.8150
113.8860
49.4696
49.3295
49.1893
49.0492
48.9090
48.7689
48.6196
0.0012
TFA
[rel]
[ppm]

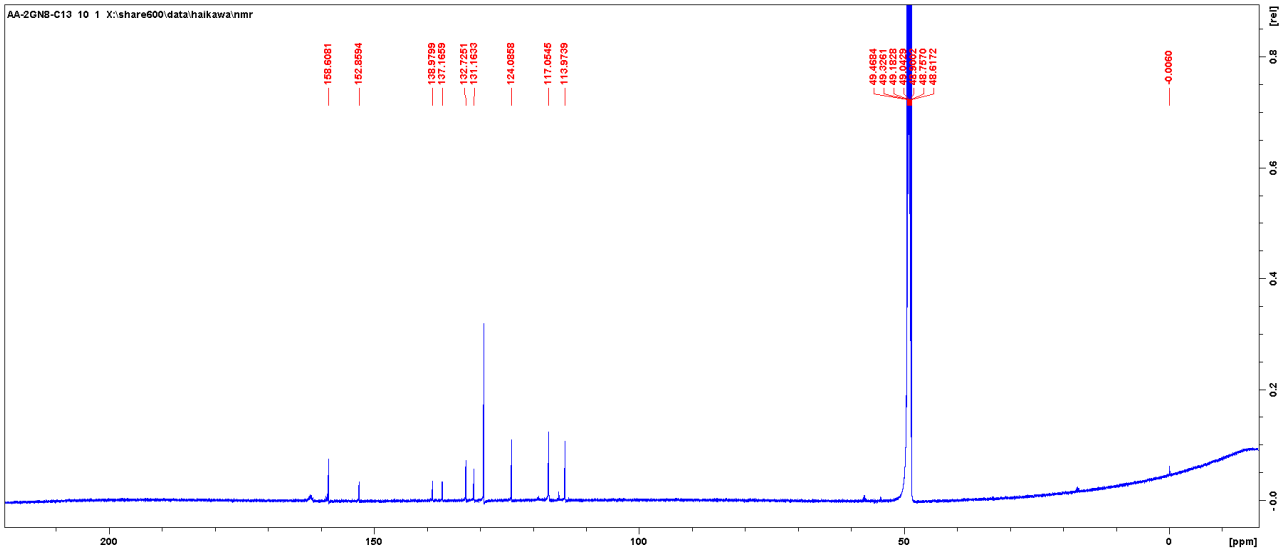

4700 Reflector Spec #1[BP = 501.2, 711]
501.2231
710.8
% Intensity
Mass (m/z)
479.0
491.6
504.2
516.8
529.4
542.0

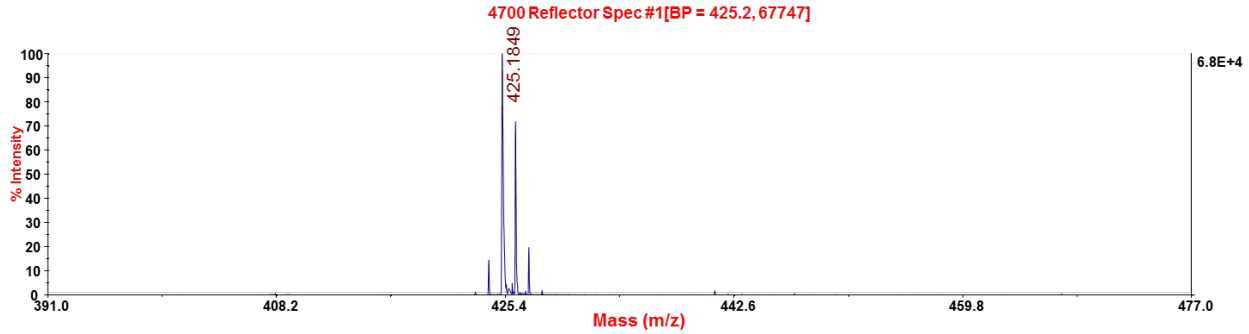

DAD1 A, Sig=220,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\049-0101.D)
220 nm
DAD1 B, Sig=254,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\049-0101.D)
254 nm
DAD1 C, Sig=345,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\049-0101.D)
345 nm
mAU
min

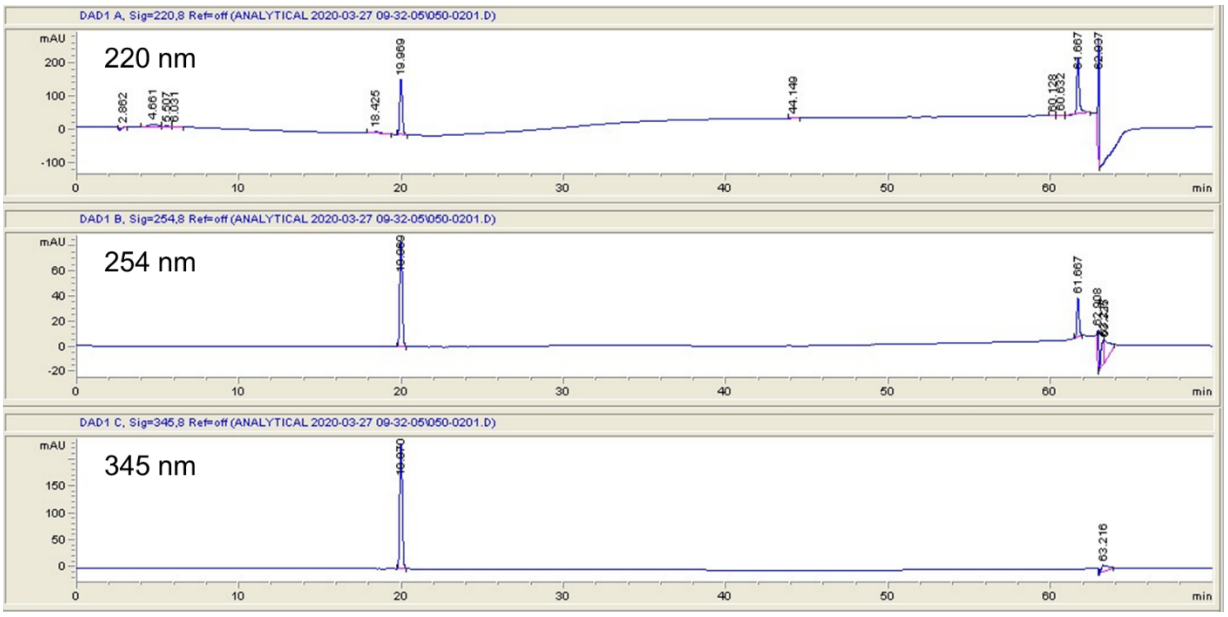
3b.




8.36
8.36
7.72
7.41
7.41
7.39
7.39
4.97
4.87
3.31
3.31
3.31
3.30
3.30
3.13
0.01
0.00
-0.01
4.00
1.80
2.05
2.20
f1 (ppm)


AA-2GN8-C13 10 1 X:\share600\data\haikawa\nmr
159.6081
152.8594
138.5799
137.1659
132.7251
131.1633
124.0868
117.0545
113.9739
49.4684
49.3261
49.1828
48.7570
48.6172
-0.0060
[ppm]


4700 Reflector Spec #1[BP = 425.2, 67747]
425.1849
6.8E+4
% Intensity
Mass (m/z)
391.0
408.2
425.4
442.6
459.8
477.0


DAD1 A, Sig=220,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\050-0201.D)
220 nm
DAD1 B, Sig=254,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\050-0201.D)
254 nm
DAD1 C, Sig=345,8 Ref=off (ANALYTICAL 2020-03-27 09-32-05\050-0201.D)
345 nm
mAU
min